

前体mRNA的选择性多聚腺苷酸化与人类疾病*

徐本锦^{**,***} 刘 玲^{**}

(山西医科大学汾阳学院医学检验系, 汾阳 032200)

摘要 真核细胞的前体 mRNA 必须经过复杂的加工过程才能成熟, 包括 5'端加帽、剪接和 3'端加工, 其中 3'端加工包括 3'端的切割和多聚腺苷酸化. 该过程由前体 mRNA 上的顺式作用元件以及多个蛋白质因子控制. 组成哺乳动物前体 mRNA 3'端加工机器的核心蛋白质复合体有切割和多聚腺苷酸化特异性因子、切割刺激因子、切割因子 I 和切割因子 II. 其他因子包括 poly(A)聚合酶、poly(A)结合蛋白、偶对蛋白 (symplekin) 等. 哺乳动物基因通常含有多个 poly(A)位点, 选择性多聚腺苷酸化不仅可产生具有不同长度 3'UTR 的 mRNA 异构体, 还可能改变基因的 CDS 区. 作为真核生物基因表达调控的关键机制, 选择性多聚腺苷酸化在细胞生长、增殖和分化中起着重要作用. 本文综述了哺乳动物前体 mRNA 的 3'端加工过程, 3'端加工机器的组成及功能, 探讨了选择性多聚腺苷酸化在多种人类疾病中的作用机制, 以期给读者带来一些新的见解.

关键词 多聚腺苷酸化, 3'UTR, 前体 mRNA, 选择性多聚腺苷酸化, 人类疾病

中图分类号 Q75

DOI: 10.16476/j.pibb.2019.0283

基因的表达包括转录和翻译两个过程. 当一个基因被 RNA 聚合酶 II 转录形成前体 mRNA 后, 要经历多个共转录加工步骤才能成为成熟的 mRNA. 前体 mRNA 的加工包括 5'端加帽、剪接和 3'端的多聚腺苷酸化^[1]. 真核生物 mRNA 3'端的形成包括两步反应, 第一步是切割, 即切割因子识别前体 mRNA 3'非翻译区 (3' UTR) 的多聚腺苷酸 (poly(A)) 信号后发生切割; 第二步是 poly(A) 尾巴的添加^[2], 即在 mRNA 的 3'端连续添加 60 nt (酵母) 或 250 nt (人类细胞) 左右的腺苷. 这两步反应是新生 RNA 成熟并发挥功能所必需的关键步骤^[3].

前体 mRNA 3'端的加工对于 mRNA 的成熟、核输出以及随后的翻译过程至关重要^[4], 因此切割和多聚腺苷酸化是一种复杂且重要的转录后基因表达调控机制. 据报道, 大部分哺乳动物基因有多个 poly(A)位点, 这些位点在不同情况下可以被选择性使用, 从而产生了多种 mRNA 变体^[5]. 这些 mRNA 变体在多个方面有着不同特点, 造成了转录后调控的复杂性. 人们把这种调控机制称为选择性多聚腺苷酸化 (alternative polyadenylation, APA).

mRNA 的剪接在真核基因表达调控中的重要作用已被认识多年. 然而, 在切割和多聚腺苷酸化水平上的基因调控直到近年来才被充分认识. 本文综述了哺乳动物前体 mRNA 3'末端加工过程、3'末端加工机器的组成及功能、以及 APA 在多种人类疾病中的作用机制, 以期给读者带来一些新的见解.

1 参与前体mRNA加工的核心顺式元件

参与前体 mRNA 加工的核心顺式元件有 poly(A)信号 (poly(A) signal, PAS)、下游序列元件 (downstream sequence element, DSE)、以及上游序列元件 (upstream sequence element, USE).

经典的 PAS 是富含 A 的六聚体 (AAUAAA), 是切割位点 (cleavage site, CS) 选择中最重要的元件, 通常位于 CS 上游 21 nt 处^[6]. 真核生物前体 mRNA 的切割和多聚腺苷酸化都需要 PAS, PAS 的

* 山西医科大学汾阳学院引进人才启动金资助项目 (2020A01).

** 并列第一作者.

*** 通讯联系人.

Tel: 0358-2100645, E-mail: benjin_ibprnalab@sina.com

收稿日期: 2019-11-25, 接受日期: 2020-02-17

缺失或突变会严重影响切割效率,阻止 mRNA 3'端的正常生成,因而也将阻止细胞核中 mRNA 的成熟^[3].

DSE 富含 U 或 G/U, 通常位于 CS 下游 14~30 nt 处^[7]. 与 PAS 不同, DSE 对于单核苷酸的突变具有一定程度的耐受性^[8]. 有临床证据显示, DSE 在某些情况下是调节切割效率的一个决定性因素^[9].

CS 位于 PAS 和 DSE 之间. 在哺乳动物中, CS 前最常见的二核苷酸是 CA 和 UA^[10]. 据报道, 脊椎动物前体 mRNAs 中 70% 的 CS 位于腺苷残基的 3' 端^[11]. 邻近 CS 的二核苷酸组成对于前体 mRNA 的

3' 末端加工有很大影响. 体外研究表明, 二核苷酸的突变对切割效率影响相对较小; 然而在高加索人群中, 他们的凝血酶原基因邻近 CS 的二核苷酸会发生从 CG 到 CA 的突变, 该突变显著提高了前体 mRNA 的切割效率, 也增大了血栓发生的可能性^[12].

USEs 位于 PAS 上游, 富含 U 但序列不保守^[8]. 在应激反应中, USE 控制凝血酶原基因 poly(A) 的使用. 此外, USE 还调控胶原基因的多聚腺苷酸化效率. 在不同物种之间, 这些核心顺式作用元件总体上保守, 但略有差异 (图 1).

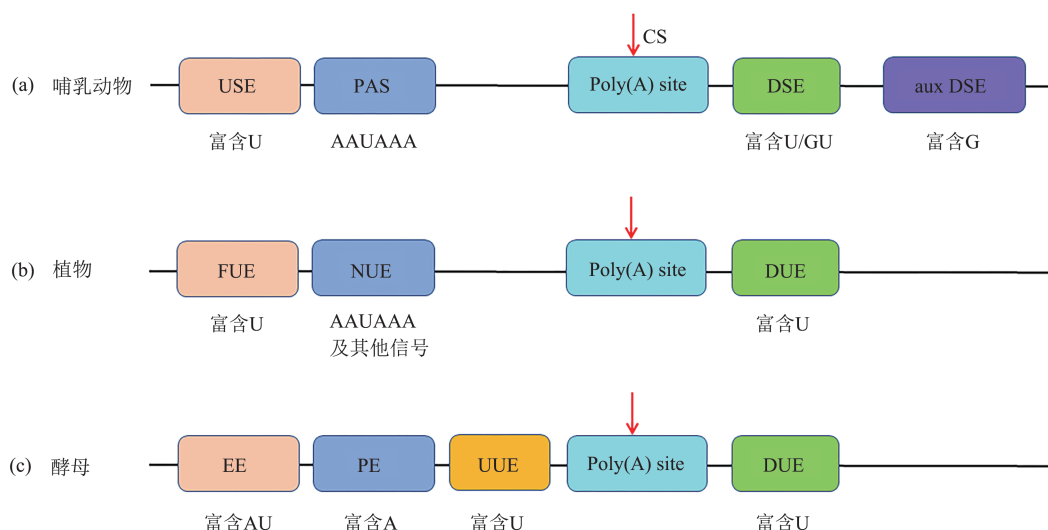


Fig. 1 Core cis elements involved in cleavage and polyadenylation site recognition

图1 参与切割和多聚腺苷酸位点识别的核心顺式元件

(a) 哺乳动物中的核心顺式元件. (b) 植物中的核心顺式元件. (c) 酵母的核心顺式元件. 如图所示, 一系列顺式元件引导切割和多聚腺苷酸化机器识别靶RNA. USE: 上游序列元件 (upstream sequence element); PAS: 多聚腺苷酸信号 (poly(A) signal); poly(A) site: 多聚腺苷酸位点; DSE: 下游序列元件 (downstream sequence element); aux DSE: 辅助下游序列元件 (auxiliary downstream sequence element); EE: 效率元件 (efficiency element); PE: 定位元件 (positioning element); FUE: 远端上游元件 (far-upstream element); NUE: 近端上游元件 (near-upstream element); CS: 切割位点 (cleavage site); UUE: 上游富含U元件 (upstream U-rich element); DUE: 下游富含U元件 (downstream U-rich element).

2 参与前体mRNA加工的核心因子

哺乳动物前体 mRNA 3' 端加工机器大约由 20 多种蛋白质组成. 核心因子有切割与多聚腺苷酸化特异性因子 (cleavage and polyadenylation specificity factors, CPSF)、切割刺激因子 (cleavage and stimulating factors, CstF)、切割因子 I (cleavage factor I, CFIm) 和切割因子 II (cleavage factor II, CFIIIm) 等 4 个多亚基蛋白质复合物, 以及一些辅助因子和 poly(A) 聚合酶 (poly(A) polymerase,

PAP)^[13].

2.1 切割和多聚腺苷酸化特异因子

切割和多聚腺苷酸化特异因子 (CPSF) 的主要功能是识别 PAS, 参与前体 mRNA 3' 末端加工. 哺乳动物 CPSF 复合体包含 6 个亚基, 分别是 CPSF160、CPSF100、CPSF73、CPSF30、WDR33 和 hFip1. hFip1 桥接着 PAP 和其他 CPSF 亚基^[14], 可结合到 PAS 的上游元件来调节 APA^[15-16]. 拟南芥 Fip1 具有刺激 PAP 活性和结合 RNA 的能力^[17], 还参与 poly(A) 位点的选择. CPSF30 和 WDR33 的

RNA 结合结构域与 PAS 协同结合^[18]. 之前的研究认为, CPSF160 结合并识别 PAS, 而新的研究证实, WDR33 和 CPSF30 才是真正的 PAS 识别因子^[19-20]. 最近的结构生物学研究表明, CPSF160 在协调其他 CPSF 亚基组装以促进 PAS 的结合上发挥着重要作用. CPSF73 具有核酸内切酶活性, 当其在 poly(A) 位点完成切割之后, PAP 在 CS 处添加 250 nt 左右的腺苷残基, 完成 3' 末端加尾过程.

2.2 切割刺激因子

切割刺激因子 (CstF) 由 CstF-64、CstF-77 和 CstF-50 3 个亚基组成. CstF-64 是多聚腺苷酸化复合体中最早发现的蛋白质之一^[21], 其氨基末端含有 RNA 识别基序, 能与 DSE 结合^[18]. 研究发现, CstF-64 的敲除导致细胞周期停滞和凋亡^[22]. CstF-77 是前体 mRNA 3' 末端正确切割所必需的, 在体内可能以二聚体的方式起作用. CstF-77 桥连 CstF-50 和 CstF-64, 形成有活性的异三聚体复合物^[23]. 此外, CstF-77 还与 CPSF160 相互作用, 有助于在前体 mRNA 上形成 CstF-CPSF 复合体^[24]. 体外实验证明, CstF-50 对于正确的切割是必需的^[23], 能够与 CstF-77 相互作用. 此外, CstF-50 还与 RNAPII 最大的亚基相互作用, 促进 RNAPII 介导的前体 mRNA 3' 端加工的激活^[25-26]. 人类 CstF 亚基的同源物已在其他物种里被发现, 例如果蝇和酵母, 提示 CstF 在前体 mRNA 多聚腺苷酸化中起着进化上保守的作用^[27]. 然而在酵母中, 目前还未见 CstF-50 序列同源物的报道^[14, 28-29]. CstF 不但有助于提高前体 mRNA 的 3' 末端加工效率, 还调节着多聚腺苷酸化与转录终止^[25] 和 DNA 损伤修复^[30] 之间的偶联.

2.3 切割因子 I

切割因子 I (CFIm) 是一种异源四聚体, 由 2 个相同的 25 ku 小亚基 (CFIm25, 基因名: *nudt21*) 和 2 个大亚基 (CFIm59、CFIm68 或 CFIm72) 组成. 其中 CFIm68 和 CFIm72 由 *Cpsf6* mRNA 的选择性剪接形成, 而 CFIm59 由 *Cpsf7* 基因编码^[31]. CFIm 在 APA 过程中发挥着重要作用, 提高了 CS 的准确性, 并影响 APA 过程^[32]. CFIm25 和 CFIm68 识别 PAS 上游的 UGUA 元件, 增强了 CPSF 向 poly(A) 位点的募集^[33]. CFIm 的敲降有利于近端 poly(A) 位点的使用^[16, 34]. Masamha 等^[35] 的研究表明, CFIm25 是近端 poly(A) 位点使用的广泛抑制因子, CFIm25 的敲除导致细胞增殖增加. 下调胶质母细胞瘤细胞中 CFIm25 的表达可增强其致癌特性并增加肿瘤大小, 而 CFIm25 的过表达可降低这些特性

并抑制肿瘤生长. 这些发现证实了 CFIm25 在调控 APA 中的关键作用, 并揭示了 CFIm25 和胶质母细胞瘤之间的内在联系.

2.4 切割因子 II

切割因子 II (CFIIm) 最初是在酵母中 (Pcf11p 和 Clp1p) 发现的, 属于酵母 CFIA 复合体^[36]. 该因子在真核生物中高度保守. 哺乳动物 CFIIm 包含 hPcf11 和 hClp1 两个亚单位. 研究发现, CFIIm 有助于提高前体 mRNA 的 3' 端加工效率^[37], 还参与调节多聚腺苷酸化与 RNA 聚合酶催化的转录终止之间的偶联^[25].

2.5 Poly(A) 聚合酶

Poly(A) 聚合酶 (PAP) 的主要功能是给 mRNA 的 3' 末端连续加上多聚腺苷酸尾巴. 在哺乳动物中, PAP 对于切割和多聚腺苷酸化这两步反应都是必要的, 而酵母的同源物 Pap1p 只在多聚腺苷酸化反应中是必需的. 在体内, PAP 通过与 3' 末端加工机器的其他蛋白质相互作用, 进而将 poly(A) 尾巴的长度控制在合适的范围^[38].

2.6 核 poly(A) 结合蛋白 1

核 poly(A) 结合蛋白 1 (poly(A) binding protein nuclear 1, PABPN1) 是真核生物中普遍存在的一种 RNA 结合蛋白, 进化上高度保守, 其主要功能是作为 poly(A) 因子参与前体 mRNA 的 3' 末端加工^[39]. PABPN1 在 mRNA 3' 末端加尾过程中发挥着重要作用^[40], 除了能够控制 poly(A) 尾巴的长度, PABPN1 还通过与较弱的近端 poly(A) 位点结合来抑制它们的切割^[41]. 因此, PABPN1 的敲除有助于近端 APA 位点的使用. 截至目前, 已确定的 PABPN1 在多聚腺苷酸化中的作用主要有两个, 首先是能够增加 PAP 的持续加尾能力^[42], 其次是控制 poly(A) 尾巴的长度在 250 nt 左右^[39].

3 前体 mRNA 3' 末端加工机器的组装

多聚腺苷酸化复合体在前体 mRNA 上的组装起始于上述因子与切割和多聚腺苷酸化位点的选择性结合. WDR33 和 CPSF30 与 CS 上游经典的 PAS 协同作用, 共同识别 PAS 序列. CstF 复合体通过 RNA 结合亚基 CstF-64 与 CS 下游的 U/GU-rich 序列结合. CPSF、CstF 和 CFIm 复合物分别识别并结合 PAS、DSE 和 USE^[43]. 根据最新的研究进展, 本文描绘了一个比较完整的前体 mRNA 3' 末端加工复合物模式图 (图 2).

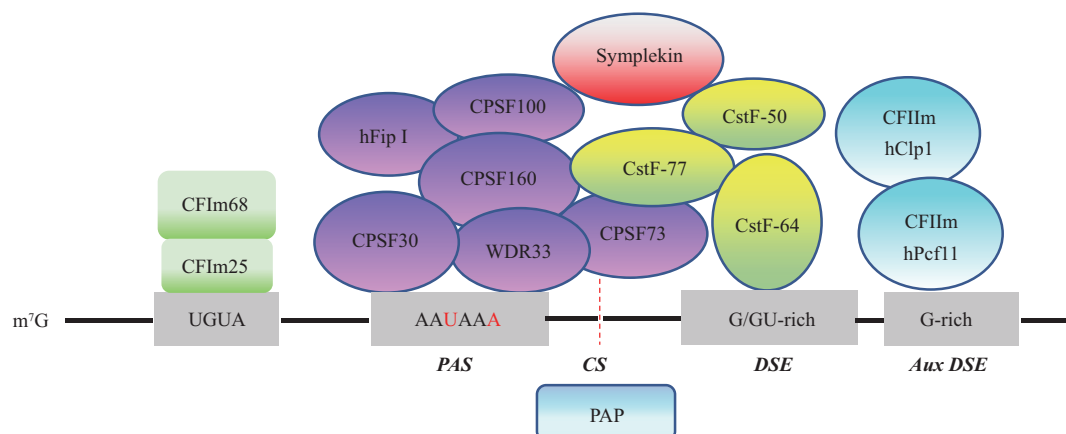


Fig. 2 Schematic drawing of the pre-mRNA 3' end processing complex in mammals

图2 哺乳动物前体mRNA 3'末端加工复合体示意图

前体mRNA (黑线) 由上游序列元件 (UGUA)、多聚腺苷酸化信号 (AAUAAA)、切割位点 (CS)、下游序列元件 (G/GU-rich) 和下游G-rich元件 (G-rich) 组成. 多聚腺苷酸化核心蛋白质由CPSF、CstF、CFIm和CFIIm组成. WDR33和CPSF30识别多聚腺苷酸化信号 (PAS), 而CstF-64识别下游序列元件 (DSE).

4 选择性多聚腺苷酸化

事实上, 真核基因往往含有多个poly(A)位点, 这些位点的选择性使用, 即选择性多聚腺苷酸化 (alternative polyadenylation, APA), 导致每个基因可产生多个转录本亚型^[44] (图3). 这些亚型的半衰期、稳定性和翻译效率等可能也不同^[45]. 在哺乳动物中, 大约70%的基因会发生APA^[5]. APA不仅可以改变3'UTR的长度, 还可能改变基因

的CDS区. 因此, APA是基因利用可选择性的poly(A)位点在转录后水平上调控基因表达的一种机制. 在人类中, 大约70%的基因通过使用多个APA位点来产生可变的3'UTR^[5]. 前体mRNA的3'UTR通常含有大量调控元件, 例如microRNA的结合位点或其他调控序列 (AU-rich元件)^[46], 能够为RNA结合蛋白或小RNA提供靶点, 进而调节转录本的稳定性、可翻译性或定位^[7, 47]. 由于3'UTR富含调控元件, 因此APA的使用会显著影响RNA的加工、稳定性、定位、翻译和降解^[48].

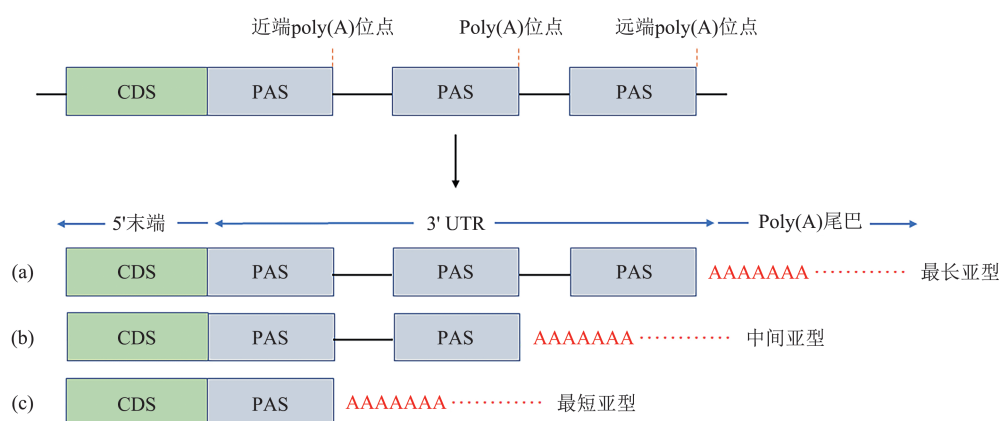


Fig. 3 Illustration of the 3'UTR length variation of a gene

图3 基因3'UTR长度变异示意图

(a) 最长的亚型. (b) 中间亚型. (c) 最短的亚型.

研究表明, 一条 mRNA 翻译的蛋白质数量取决于它的 3'UTR 长度, 3'UTR 更短的转录物产生更高水平的蛋白质^[49]. 当 APA 产生相同的蛋白质时, 由于 3'UTR 中调节元件的增加或减少, 蛋白质的产量也会受到影响^[44]. 当 APA 改变了开放阅读框时, 编码的蛋白质也就发生了改变^[50]. 总之, APA 促进了 RNA 结构的复杂性和功能的多样性, 从而使得遗传信息由基因组向表型组的传递变得复杂化. 在神经退行性疾病中, 例如阿尔茨海默病 (AD)、帕金森病 (PD) 和肌萎缩侧索硬化症 (ALS), 病人的 APA 谱发生了显著的改变^[51].

5 APA与人类疾病

研究发现, APA 通过在前体 mRNA 的特定位点优先添加 poly(A) 尾巴, 从而参与到人类多个生物学过程的调控, 例如细胞的激活、分化和增殖, 以及免疫防御等^[52-54]. APA 对肿瘤发生、癌症发展和表型的作用日益显现, 3'UTR 的整体截短会导致癌基因 3'UTR 的抑制元件丢失, 随后癌基因被激活^[55-56]. 截止目前, 已发现多种人类疾病与 APA 相关 (表 1). 本文选取几种代表性疾病, 探讨了 APA 在这些疾病中的作用机制.

Table 1 Summary of alternative polyadenylation (APA) related diseases
表1 选择性多聚腺苷酸化 (APA) 相关疾病总结

分类	疾病	靶基因	机制	引文
内分泌系统	糖尿病肾病	<i>HGRG-14</i>	高血糖水平导致远端PAS的使用	[57]
	2型糖尿病	<i>TCF7L2</i>	通过使用内含子PAS来增加不同的mRNA亚型	[58]
血液系统	β地中海贫血	<i>b-globin</i>	多聚腺苷酸化位点突变引起3'UTR延长和转录终止缺陷	[59]
癌症	乳腺癌 (MB231 细胞系)		近端PAS的使用导致3'UTR延长	[60]
	胶质母细胞瘤	<i>CCND1</i>	CPSF5的敲除诱导3'UTR缩短	[35]
感染与免疫	IPEX综合征	<i>FOXP3</i>	第一个PAS的突变 (AATAAA-AATGAA)	[61]
	系统性红斑狼疮	<i>GIMAP5</i>	近端PAS的突变 (AATAAA-AATAGA)	[62]
神经系统	帕金森病	<i>α-Synuclein</i>	多巴胺水平的增加导致远端PAS的使用	[63]
	阿尔茨海默病	<i>COX-2</i>	远端PAS使用增加	[64]

5.1 急性髓性白血病

急性髓性白血病 (AML) 是最常见的恶性髓样癌, 其特征是, 由于造血前体向成熟骨髓细胞分化受阻, 导致骨髓、血液和其他组织中的骨髓祖细胞 (“母细胞” 或白血病细胞) 增殖增加^[65]. 研究发现, APA 不但参与 AML 的调解, 还对许多癌细胞的发育起着调节作用^[66]. 据报道, 基因对近端 poly(A) 位点的偏好, 导致融合癌基因 *AML1-ETO* (AE) 的 3'UTR 缩短, 增强了 AE 基因的稳定性, 进而促进 T (8; 21) 型 AML 中白血病细胞的生长^[67]. Ye 等^[66] 利用健康和 AML 患者样本的单细胞 RNA-seq 数据, 分析了不同细胞类型中 APA 的动态使用情况. 结果显示, APA 在单细胞水平上广泛参与白血病的发展和红细胞生成. 此外, 他们还描绘了一个确定 AML 中分化细胞 mRNA 加工景观的高分辨率图谱.

5.2 肺纤维化

肺纤维化 (IPF) 是一种具有广泛细胞增殖和分化的疾病, 有着许多与癌症相似的特征, 预后

差、治疗方案少. 研究发现, APA 通过改变小 RNA 的敏感性来驱动纤维化因子表达, 进而在细胞应激反应中起重要作用. 此外, 在高度增殖或分化的细胞中, APA 位点的选择有一个从远端到近端的整体转移, 产生了具有较短 3'UTRs 的转录本^[13, 35]. 这些转录本能够逃避 miRNA 的识别和 AU-rich 元件的抑制, 通常情况下更稳定, 蛋白质翻译效率更高^[35]. Poly(A) 位点的选择不仅由邻近的序列元件决定, 还受多聚腺苷酸化蛋白复合物水平的影响^[68]. 据报道, CFIm25 的下调诱导 mRNAs 的整体缩短^[35], 增加了几种癌基因的表达并增强癌细胞的增殖, 这种调控方式被认为是一种新的机制. 与上述发现相一致, Weng 等^[68] 的研究表明, CFIm25 下调是提高促纤维化基因表达的重要机制.

5.3 乳腺癌

乳腺癌 (breast cancer) 是发生在乳腺上皮组织的恶性肿瘤. 乳腺癌患者中 99% 是女性, 男性仅占 1%. APA 作为一种转录后调控机制, 通过影响转录丰度、细胞定位以及与 microRNAs 的相互作用

来控制基因表达^[13]. 研究发现, 3'UTR长度的变化与免疫应答^[69]和肿瘤生长^[49]期间细胞的增殖和分化紧密相关. APA在90%左右的癌症中常常导致3'UTR变短^[56]. 此外, APA在特定基因和转录信号中的使用, 例如 *prelid1*^[70]、*usp9x*、*snx3* 和 *ymell1d*^[45], 被认为是乳腺癌临床结果的强有力预测因子. Kim等^[71]对乳腺癌患者细胞的RNA-seq数据进行APA事件分析后发现, 细胞类型和癌症类型特异性转录调控与APA信号相关, 单细胞的APA特征可作为早期乳腺癌的预后标志物. 三阴性乳腺癌(TNBC)占有乳腺癌病例的15%~20%, 在免疫组织学上表现为雌激素受体(ER)、孕酮受体(PR)和人类表皮生长因子受体2(HER2/neu)阴性, 具有高增殖指数、早期复发和较高死亡率等特点^[72]. Akman等^[45]揭示了TNBC患者3'UTR的长度变化模式, 以及APA与EGF信号之间的潜在联系. 总的来说, 检测不同基因3'UTR的长度变化可能有助于发现新的癌症相关基因, 这在传统的微阵列基因表达分析中可能被忽视.

5.4 神经退行性疾病

5.4.1 阿尔茨海默病

有人用RNA-seq技术研究了神经退行性疾病患者的转录组, 以评估基因表达的变化^[73]、剪接^[74]以及miRNA^[75]和lncRNAs表达的变化^[10]. 然而, 对神经退行性组织或细胞中APA谱的变化进行全面评估的研究却很少. Patel等^[51]的研究表明, 阿尔茨海默病(Alzheimer's disease, AD)患者不同脑区的不同基因表现出明显的APA, 但都表现为3'UTR的显著延长.

5.4.2 帕金森病

帕金森病(Parkinson's disease, PD)是一种与年龄相关的神经退行性疾病, 伴随着中脑多巴胺(DA)神经元的丢失和神经元内出现 α -突触核蛋白聚集体(称为Lewy体)^[76]. PD患者的特征是运动迟缓、震颤和运动僵硬. 此外, 患者体内参与多巴胺代谢的基因表达发生了改变^[77]. 截至目前, 只有一项研究评估了APA在PD中的作用. Rhinn等^[63]发现, 在PD患者的大脑中, *SNCA*基因(一个对PD发病至关重要的基因)远端poly(A)位点的使用增加, 导致产生更多具有较长3'UTR的转录本, 这些转录本翻译产生的蛋白质定位于线粒体, 并且更容易聚集. 表明APA在 α -突触核蛋白调控中起着关键作用.

5.4.3 肌萎缩侧索硬化

肌萎缩侧索硬化症(amyotrophic lateral sclerosis, ALS), 又称运动神经元疾病, 是一种使人衰弱的疾病, 导致上下运动神经元死亡、肌肉僵硬、虚弱, 最终丧失自主运动. 家族性(fALS)和散发性(sALS)病例均表现出相似的神经病理学特征, 并且患者样本都显示出类似的小RNA和基因表达失调^[75]. *FUS*和*TARDP*这两个基因编码与RNA加工相关的RNA结合蛋白, 它们的核定位序列(NLS)的突变与fALS有关. 突变后形成的蛋白质随后在细胞质内聚集形成包涵体, 导致运动神经元功能丧失^[78]. 有人用RNA-seq技术对一例*C9orf27*基因突变的ALS患者和一例sALS患者的3'UTR长度变化进行了全面评估, 结果发现与健康对照组相比, 这两名患者的APA发生了不同程度的改变, 并且大脑不同区域poly(A)位点的使用表现出广泛的不同^[79]. Patel等^[51]对阿尔茨海默病(AD)、帕金森病(PD)和肌萎缩侧索硬化症(ALS)患者和健康个体的RNA-seq数据进行了APA动态分析. 结果表明, 通过改变AD中*UBR1*和*OGDHL*、PD中的*LONP1*以及ALS中*UCHL1*等基因的表达, APA的异常可能在神经退行性疾病中发挥着重要作用.

5.5 IPEX综合征

IPEX是X染色体连锁的多内分泌疾病、肠病伴免疫功能障碍综合征. *Foxp3*是小鼠坏血病基因, 其在人类中的同源基因*FOXP3*定位于Xp11.23~Xq13.3, 编码一种新的叉头蛋白. Bennett等^[80]对IPEX综合征患者*FOXP3*基因的3'UTR分析发现, PAS序列发生了A→G转变, 即由AAUAAA→AAUGAA, 表明*FOXP3*基因信息不稳定. 系谱分析发现, 该突变与IPEX综合征共分离(co-segregate), 进一步表明人类IPEX综合征是一种由*FOXP3*基因PAS序列突变导致的罕见免疫系统遗传病. 目前, PAS序列突变导致疾病的报道并不多, 已知的有 α 2-珠蛋白基因突变(AAUAAA→AAUAAG)导致的贫血^[81], 以及*IL2RG*基因突变(AAUAAA→AAUAAG)导致的X-连锁重症联合免疫缺陷病(X-linked severe combined immunodeficiency syndrome, XSCID)^[82].

5.6 II型糖尿病

II型糖尿病是一种最常见的内分泌代谢疾病, 具有遗传易感性, 在环境因素的触发下发病. *TCF7L2*属于T细胞因子/淋巴增强因子(TCF/

LEF) 家族的转录因子, 能够利用选择性剪接产生具有独特时空特征和基因/环境依赖性反式激活能力的亚型^[83]. 研究表明, *TCF7L2* 的单核苷酸多态性(SNP)与II型糖尿病相关^[84]. Locke等^[58]对EST (expressed sequence tag)数据库进行分析后发现, *TCF7L2* 基因的内含子4中存在一个APA位点, 该位点的使用导致产生抑制TCF/LEF依赖性靶基因的关系提供了新的见解. 高血糖是糖尿病肾病的重要特征, 是I型和II型糖尿病的致命并发症. 另一个与II型糖尿病有关的基因是高血糖调节基因*HGRG-14*, 在正常情况下, 细胞只表达700 bp左右的短转录物, 但在高血糖条件下, *HGRG-14*受APA调节而差异表达最长的转录本(2 kb)以响应高浓度的葡萄糖^[85].

6 小结与展望

最新的研究进展使人们在原子水平上对前体mRNA的3'末端加工机制有了前所未有的认识. 近年来, 高通量测序技术的出现无疑对我们理解前体mRNA 3'末端的形成和APA做出了巨大贡献. 随着全基因组方法学的发展, 由APA引起的转录本数量将显著增加. 然而, 在一些病例中, mRNA的3'UTR缩短是由于近端PAS的使用还是因为存在沉默机制, 还需进一步阐明. APA不是由单因素控制的, 而是由多因素控制的相当复杂的过程. 高通量技术的发展无疑将加快APA相关疾病的发现. 深入阐明APA的分子调控机制, 对于了解某些疾病的发生发展具有重要意义.

尽管目前已经获得了单个poly(A)因子的结构, 但是对整个复合物三维结构的表征仍然未见报道. 因此, 我们仍难以全面理解前体mRNA 3'末端成熟机制这一基本过程, 例如PAS的识别机制、poly(A)位点的选择等问题. 此外, mRNA 3'末端加工因子与组蛋白mRNA或某些非编码RNA成熟之间的关系, 仍需进一步研究. 今后的一个目标是全面揭示mRNA 3'末端加工机器的三维结构, 进一步从原子水平阐明该复合体的工作机制. 目前发展迅速的冷冻电镜技术, 也许为这一目标的实现奠定了基础.

过去很长一段时间, 人们只是从整体上描绘mRNA的3'UTR在不同条件下“整体变短”或者“整体变长”, 将来的另一个目标是更多的关注与疾病或者生理相关的APA事件, 开发新的工具用于

潜在poly(A)位点的发掘. 此外, 针对APA的药物设计也是今后的一个方向.

参 考 文 献

- [1] Tudek A, Lloret-Llinares M, Jensen T H. The multitasking polyA tail: nuclear RNA maturation, degradation and export. *Philosophical Transactions of The Royal Society of London. Series B, Biological Sciences*, 2018, **373**(1762): 20180169
- [2] Proudfoot N J. Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science*, 2016, **352**(6291): aad9926
- [3] Wahle E, Keller W. The biochemistry of 3'-end cleavage and polyadenylation of messenger RNA precursors. *Annual Review of Biochemistry*, 1992, **61**: 419-440
- [4] Proudfoot N J, Furger A, Dye M J. Integrating mRNA processing with transcription. *Cell*, 2002, **108**(4): 501-512
- [5] Derti A, Garrett-Engle P, Macisaac K D, *et al.* A quantitative atlas of polyadenylation in five mammals. *Genome Research*, 2012, **22**(6): 1173-1183
- [6] Gruber A J, Schmidt R, Gruber A R, *et al.* A comprehensive analysis of 3' end sequencing data sets reveals novel polyadenylation signals and the repressive role of heterogeneous ribonucleoprotein C on cleavage and polyadenylation. *Genome Research*, 2016, **26**(8): 1145-1159
- [7] Neve J, Patel R, Wang Z, *et al.* Cleavage and polyadenylation: ending the message expands gene regulation. *RNA Biology*, 2017, **14**(7): 865-890
- [8] Zhao J, Hyman L, Moore C. Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiology and Molecular Biology Reviews*: MMBR, 1999, **63**(2): 405-445
- [9] Danckwardt S, Gehring N H, Neu-Yilik G, *et al.* The prothrombin 3' end formation signal reveals a unique architecture that is sensitive to thrombophilic gain-of-function mutations. *Blood*, 2004, **104**(2): 428-435
- [10] Li X Q, Du D. RNA polyadenylation sites on the genomes of microorganisms, animals, and plants. *Plos One*, 2013, **8**(11): e79511
- [11] Sheets M D, Ogg S C, Wickens M P. Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. *Nucleic Acids Research*, 1990, **18**(19): 5799-5805
- [12] Gehring N H, Frede U, Neu-Yilik G, *et al.* Increased efficiency of mRNA 3' end formation: a new genetic mechanism contributing to hereditary thrombophilia. *Nature Genetics*, 2001, **28**(4): 389-392
- [13] Di Giammartino D C, Nishida K, Manley J L. Mechanisms and consequences of alternative polyadenylation. *Mol Cell*, 2011, **43**(6): 853-866
- [14] Mandel C R, Bai Y, Tong L. Protein factors in pre-mRNA 3'-end processing. *Cellular and Molecular Life Sciences*, 2008, **65**(7-8): 1099-1122

- [15] Lackford B, Yao C, Charles G M, *et al.* Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal. *The Embo Journal*, 2014, **33**(8): 878-889
- [16] Li W, You B, Hoque M, *et al.* Systematic profiling of poly(A)+ transcripts modulated by core 3' end processing and splicing factors reveals regulatory rules of alternative cleavage and polyadenylation. *Plos Genetics*, 2015, **11**(4): e1005166
- [17] Forbes K P, Addepalli B, Hunt A G. An Arabidopsis Fip1 homolog interacts with RNA and provides conceptual links with a number of other polyadenylation factor subunits. *Journal of Biological Chemistry*, 2006, **281**(1): 176-186
- [18] Grozdanov P N, Masoumzadeh E, Latham M P, *et al.* The structural basis of CstF-77 modulation of cleavage and polyadenylation through stimulation of CstF-64 activity. *Nucleic Acids Research*, 2018, **46**(22): 12022-12039
- [19] Chan S L, Huppertz I, Yao C, *et al.* CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3' processing. *Genes & Development*, 2014, **28**(21): 2370-2380
- [20] Schonemann L, Kuhn U, Martin G, *et al.* Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33. *Genes & Development*, 2014, **28**(21): 2381-2393
- [21] Wilusz J, Shenk T. A 64 kd nuclear protein binds to RNA segments that include the AAUAAA polyadenylation motif. *Cell*, 1988, **52**(2): 221-228
- [22] Manley Y T a J L. Levels of polyadenylation factor CstF-64 control IgM heavy chain mRNA accumulation and other events associated with B cell differentiation. *Molecular Cell*, 1998, **2**(6): 761-771
- [23] Takagaki Y, Manley J L. A polyadenylation factor subunit is the human homologue of the *Drosophila* suppressor of forked protein. *Nature*, 1994, **372**(6505): 471-474
- [24] Murthy K G, Manley J L. The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation. *Genes & Development*, 1995, **9**(21): 2672-2683
- [25] McCracken S, Fong N, Yankulov K, *et al.* The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature*, 1997, **385**(23): 357-361
- [26] Manley Y H J L. RNA polymerase II is an essential mRNA polyadenylation factor. *Nature*, 1998, **395**(6697): 93-96
- [27] Manley J L, And Takagaki, Y. The end of the message-another link between yeast and mammals. *Science* 1996, **274**(5292): 1481-1482
- [28] Turner R E, Pattison A D, Beilharz T H. Alternative polyadenylation in the regulation and dysregulation of gene expression. *Seminars in Cell & Developmental Biology*, 2018, **75**: 61-69
- [29] Millevoi S, Vagner S. Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Research*, 2010, **38**(9): 2757-2774
- [30] Mirkin N, Fonseca D, Mohammed S, *et al.* The 3' processing factor CstF functions in the DNA repair response. *Nucleic Acids Research*, 2008, **36**(6): 1792-1804
- [31] Macdonald C C. Tissue-specific mechanisms of alternative polyadenylation: Testis, brain, and beyond (2018 update). *Wiley Interdisciplinary Reviews-RNA*, 2019, **10**(4): e1526
- [32] Thore S, Fribourg S. Structural insights into the 3'-end mRNA maturation machinery: snapshot on polyadenylation signal recognition. *Biochimie*, 2019, **164**: 105-110
- [33] Kubo T, Wada T, Yamaguchi Y, *et al.* Knock-down of 25 kDa subunit of cleavage factor Im in HeLa cells alters alternative polyadenylation within 3'-UTRs. *Nucleic Acids Research*, 2006, **34**(21): 6264-6271
- [34] Zhu Y, Wang X, Forouzmand E, *et al.* Molecular mechanisms for CFIm-mediated regulation of mRNA alternative polyadenylation. *Mol Cell*, 2018, **69**(1): 62-74
- [35] Masamha C P, Xia Z, Yang J, *et al.* CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. *Nature*, 2014, **510**(7505): 412-416
- [36] de Vries, Rügsegger U, Hübner W, *et al.* Human pre-mRNA cleavage factor IIm contains homologs of yeast proteins and bridges two other cleavage factors. *The Embo Journal*, 2000, **19**(21): 5895-5904
- [37] Schäfer P, Tüting C, Schönmann L, *et al.* Reconstitution of mammalian cleavage factor II involved in 3' processing of mRNA precursors. *RNA*, 2018, **24**(12): 1721-1737
- [38] Edmonds M. A history of poly A sequences: from formation to factors to function. *Progress in Nucleic Acid Research and Molecular Biology*, 2002, **71**: 285-389
- [39] Apponi L H, Leung S W, Williams K R, *et al.* Loss of nuclear poly(A)-binding protein 1 causes defects in myogenesis and mRNA biogenesis. *Human Molecular Genetics*, 2010, **19**(6): 1058-1065
- [40] Kuhn U, Gundel M, Knoth A, *et al.* Poly(A) tail length is controlled by the nuclear poly(A)-binding protein regulating the interaction between poly(A) polymerase and the cleavage and polyadenylation specificity factor. *Journal of Biological Chemistry*, 2009, **284**(34): 22803-22814
- [41] Jenal M, Elkon R, Loayza-Puch F, *et al.* The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. *Cell*, 2012, **149**(3): 538-553
- [42] Kerwitz Y, Kuhn U, Lilie H, *et al.* Stimulation of poly(A) polymerase through a direct interaction with the nuclear poly(A) binding protein allosterically regulated by RNA. *The Embo Journal*, 2003, **22**(14): 3705-3714
- [43] Chan S, Choi E A, Shi Y. Pre-mRNA 3'-end processing complex assembly and function. *Wiley Interdisciplinary Reviews-RNA*, 2011, **2**(3): 321-335
- [44] Tian B, Manley J L. Alternative cleavage and polyadenylation: the long and short of it. *Trends in Biochemical Sciences*, 2013, **38**(6): 312-320
- [45] Akman H B, Oyken M, Tuncer T, *et al.* 3'UTR shortening and EGF signaling: implications for breast cancer. *Human Molecular Genetics*, 2015, **24**(24): 6910-6920
- [46] Fabian M R, Sonenberg N, Filipowicz W. Regulation of mRNA translation and stability by microRNAs. *Annual Review of*

- Biochemistry, 2010, **79**(1): 351-379
- [47] Neve J, Furger A. Alternative polyadenylation: less than meets the eye? Biochemical Society Transactions, 2014, **42**(4): 1190-1195
- [48] Matoulkova E, Michalova E, Vojtesek B, *et al.* The role of the 3' untranslated region in post-transcriptional regulation of protein expression in mammalian cells. RNA Biology, 2012, **9**(5): 563-576
- [49] Mayr C, Bartel D P. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. Cell, 2009, **138**(4): 673-684
- [50] Shi Y. Alternative polyadenylation: new insights from global analyses. RNA, 2012, **18**(12): 2105-2117
- [51] Patel R, Brophy C, Hickling M, *et al.* Alternative cleavage and polyadenylation of genes associated with protein turnover and mitochondrial function are deregulated in Parkinson's, Alzheimer's and ALS disease. BMC Medical Genomics, 2019, **12**(60): 1-14
- [52] Huang G, Huang S, Wang R, *et al.* Dynamic regulation of tandem 3' untranslated regions in zebrafish spleen cells during immune response. Journal of Immunology (Baltimore, Md : 1950), 2016, **196**(2): 715-725
- [53] Hwang H W, Saito Y, Park C Y, *et al.* cTag-PAPERCLIP reveals alternative polyadenylation promotes cell-type specific protein diversity and shifts *araf* isoforms with microglia activation. Neuron, 2017, **95**(6): 1334-1349
- [54] Qiu F, Fu Y, Lu C, *et al.* Small nuclear ribonucleoprotein polypeptide A-mediated alternative polyadenylation of STAT5B during Th1 cell differentiation. Journal of Immunology (Baltimore, Md : 1950), 2017, **199**(9): 3106-3115
- [55] Masamha C P, Wagner E J. The contribution of alternative polyadenylation to the cancer phenotype. Carcinogenesis, 2018, **39**(1): 2-10
- [56] Xia Z, Donehower L A, Cooper T A, *et al.* Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across seven tumour types. Nature Communications, 2014, **5**(5274): 1-13
- [57] Abdel Wahab N, Gibbs J, Mason R M. Regulation of gene expression by alternative polyadenylation and mRNA instability in hyperglycaemic mesangial cells. The Biochemical Journal, 1998, **336**(Pt 2): 405-411
- [58] Locke J M, Da Silva Xavier G, Rutter G A, *et al.* An alternative polyadenylation signal in *TCF7L2* generates isoforms that inhibit T cell factor/lymphoid-enhancer factor (TCF/LEF) -dependent target genes. Diabetologia, 2011, **54**(12): 3078-3082
- [59] Rund D, Dowling C, Najjar K, *et al.* Two mutations in the beta-globin polyadenylation signal reveal extended transcripts and new RNA polyadenylation sites. Proc Natl Acad Sci USA, 1992, **89**(10): 4324-4328
- [60] 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 Research, 2011, **21**(5): 741-747
- [61] Bennett C L, Christie J, Ramsdell F, *et al.* The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of *FOXP3*. Nature Genetics, 2001, **27**(1): 20-21
- [62] Hellquist A, Zucchelli M, Kivinen K, *et al.* The human *GIMAP5* gene has a common polyadenylation polymorphism increasing risk to systemic lupus erythematosus. Journal of Medical Genetics, 2007, **44**(5): 314-321
- [63] Rhinn H, Qiang L, Yamashita T, *et al.* Alternative alpha-synuclein transcript usage as a convergent mechanism in Parkinson's disease pathology. Nature Communications, 2012, **3**(1): 1084-1094
- [64] Lukiw W J, Bazan N G. Cyclooxygenase 2 RNA message abundance, stability, and hypervariability in sporadic Alzheimer neocortex. Journal of Neuroscience Research, 1997, **50**(6): 937-945
- [65] Dohner H, Weisdorf D J, Bloomfield C D. Acute myeloid leukemia. The New England Journal of Medicine, 2015, **373**(12): 1136-1152
- [66] Ye C, Zhou Q, Hong Y, *et al.* Role of alternative polyadenylation dynamics in acute myeloid leukaemia at single-cell resolution. RNA Biology, 2019, **16**(6): 785-797
- [67] Shima T, Davis A G, Miyauchi S, *et al.* CPSF1 regulates AML1-ETO fusion gene polyadenylation and stability in t(8; 21) acute myelogenous leukemia. Blood, 2017, **130** (Supplement 1): 2498
- [68] Weng T, Ko J, Masamha C P, *et al.* Cleavage factor 25 deregulation contributes to pulmonary fibrosis through alternative polyadenylation. The Journal of Clinical Investigation, 2019, **129**(5): 1984-1999
- [69] Sandberg R, Neilson J R, Sarma A, *et al.* Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. Science, 2008, **320**(5883): 1643-1647
- [70] Gillen A E, Brechbuhl H M, Yamamoto T M, *et al.* Alternative polyadenylation of *PRELID1* regulates mitochondrial ROS signaling and cancer outcomes. Molecular Cancer Research: MCR, 2017, **15**(12): 1741-1751
- [71] Kim N, Chung W, Eum H H, *et al.* Alternative polyadenylation of single cells delineates cell types and serves as a prognostic marker in early stage breast cancer. Plos One, 2019, **14**(5): e0217196
- [72] Griffiths C L, Olin J L. Triple negative breast cancer: a brief review of its characteristics and treatment options. Journal of Pharmacy Practice, 2012, **25**(3): 319-323
- [73] Dumitriu A, Golji J, Labadorf A T, *et al.* Integrative analyses of proteomics and RNA transcriptomics implicate mitochondrial processes, protein folding pathways and GWAS loci in Parkinson disease. BMC Medical Genomics, 2016, **9**(5): 1-17
- [74] Twine N A, Janitz K, Wilkins M R, *et al.* Whole transcriptome sequencing reveals gene expression and splicing differences in brain regions affected by Alzheimer's disease. Plos One, 2011, **6**(1): e16266
- [75] Butovsky O, Jedrychowski M P, Cialic R, *et al.* Targeting miR-155 restores abnormal microglia and attenuates disease in SOD1 mice. Annals of Neurology, 2015, **77**(1): 75-99
- [76] Chinta S J, Lieu C A, Demaria M, *et al.* Environmental stress,

- ageing and glial cell senescence: a novel mechanistic link to Parkinson's disease?. *Journal of Internal Medicine*, 2013, **273**(5): 429-436
- [77] Poewe W, Seppi K, Tanner C M, *et al.* Parkinson disease. *Nature Reviews Disease Primers*, 2017, **3**(17013):1-21
- [78] Blokhuis A M, Groen E J, Koppers M, *et al.* Protein aggregation in amyotrophic lateral sclerosis. *Acta Neuropathologica*, 2013, **125**(6): 777-794
- [79] Prudencio M, Belzil V V, Batra R, *et al.* Distinct brain transcriptome profiles in *C9orf72*-associated and sporadic ALS. *Nature Neuroscience*, 2015, **18**(8): 1175-1182
- [80] Bennett C L, Brunkow M E, Ramsdell F, *et al.* A rare polyadenylation signal mutation of the *FOXP3* gene (AAUAAA->AAUGAA) leads to the IPEX syndrome. *Immunogenetics*, 2001, **53**(6): 435-439
- [81] Fei Y J, Oner R, Bozkurt G, *et al.* Hb H disease caused by a homozygosity for the AATAAA->AATAAG mutation in the polyadenylation site of the alpha 2-globin gene: hematological observations. *Acta Haematologica*, 1992, **88**(2-3): 82-85
- [82] Hsu A P, Tsai E J, Anderson S M, *et al.* Unusual X-linked SCID phenotype due to mutation of the poly-A addition signal of *IL2RG*. *American Journal of Human Genetics*, 2000, **67**(4): 50-50
- [83] Weise A, Bruser K, Elfert S, *et al.* Alternative splicing of *Tcf7l2* transcripts generates protein variants with differential promoter-binding and transcriptional activation properties at Wnt/beta-catenin targets. *Nucleic Acids Research*, 2010, **38**(6): 1964-1981
- [84] Cauchi S, El Achhab Y, Choquet H, *et al.* *TCF7L2* is reproducibly associated with type 2 diabetes in various ethnic groups: a global meta-analysis. *Journal of Molecular Medicine (Berlin, Germany)*, 2007, **85**(7): 777-782
- [85] Curinha A, Oliveira Braz S, Pereira-Castro I, *et al.* Implications of polyadenylation in health and disease. *Nucleus*, 2014, **5**(6): 508-519

Alternative Polyadenylation of Pre-mRNA and Related Human Diseases*

XU Ben-Jin^{*,***}, LIU Ling^{**}

(Laboratory Medicine Department in Fenyang College of Shanxi Medical University, Fenyang 032200, China)

Abstract The pre-mRNA of eukaryotic cells must undergo extensive and complex processes to produce mature mRNAs, including 5' end capping, splicing and 3' end processing. Among them, 3' end processing includes cleavage and polyadenylation, which is controlled by the *cis*-elements of pre-mRNA and a number of protein factors. Mammalian 3' end processing machinery consists of cleavage and polyadenylation factors, cleavage and stimulating factors, cleavage factor I and cleavage factor II. Other protein factors include poly(A) polymerase, poly(A) binding protein, symplekin and so on. Mammalian genes usually contain multiple polyadenylation sites, and alternative polyadenylation can not only produce mRNA variants with different length of 3'UTR, but also change the CDS region of genes. As a key mechanism for the regulation of gene expression in eukaryotes, alternative polyadenylation plays an important role in cell growth, proliferation and differentiation. This paper reviews the formation of the 3' end of mammalian pre-mRNA, the composition and function of the 3' end processing machinery, the mechanism of alternative polyadenylation in various human diseases, and hope to bring some new insights to readers.

Key words polyadenylation, 3'UTR, pre-mRNA, alternative polyadenylation, human diseases

DOI: 10.16476/j.pibb.2019.0283

* This work was supported by a grant from Starting Fund for Talent Introduction of Fenyang College of Shanxi Medical University (2020A01) .

** These authors contributed equally to this work.

*** Corresponding author.

Tel: 86-358-2100645, E-mail: benjin_ibprnalab@sina.com

Received: November 25, 2019 Accepted: February 17, 2020