



Review

Alternative polyadenylation: An untapped source for prostate cancer biomarkers and therapeutic targets?



Akira Kurozumi ^a, Shawn E. Lupold ^{a,b,*}

^a The James Buchanan Brady Urologic Institute and Department of Urology, Johns Hopkins School of Medicine, Baltimore, MD, USA

^b The Department of Oncology, Sidney Kimmel Comprehensive Cancer Center, School of Medicine, Johns Hopkins University, Baltimore, MD, USA

Received 10 November 2020; received in revised form 20 March 2021; accepted 5 May 2021
Available online 6 June 2021

KEYWORDS

Prostate cancer;
Polyadenylation;
Alternative polyadenylation;
3'-untranslated region;
microRNA;
Intronic polyadenylation

Abstract *Objective:* To review alternative polyadenylation (APA) as a mechanism of gene regulation and consider potential roles for APA in prostate cancer (PCa) biology and treatment. *Methods:* An extensive review of mRNA polyadenylation, APA, and PCa literature was performed. This review article introduces APA and its association with human disease, outlines the mechanisms and components of APA, reviews APA in cancer biology, and considers whether APA may contribute to PCa progression and/or produce novel biomarkers and therapeutic targets for PCa.

Results: Eukaryotic mRNA 3'-end cleavage and polyadenylation play a critical role in gene expression. Most human genes encode more than one polyadenylation signal, and produce more than one transcript isoform, through APA. Polyadenylation can occur throughout the gene body to generate transcripts with differing 3'-termini and coding sequence. Differences in 3'-untranslated regions length can modify post-transcriptional gene regulation by microRNAs and RNA binding proteins, and alter mRNA stability, translation efficiency, and subcellular localization. Distinctive APA patterns are associated with human diseases, tissue origins, and changes in cellular proliferation rate and differentiation state. APA events may therefore generate unique mRNA biomarkers or therapeutic targets in certain cancer types or phenotypic states.

Conclusions: The full extent of cancer-associated and tissue-specific APA events have yet to be defined, and the mechanisms and functional consequences of APA in cancer remain incompletely understood. There is evidence that APA is active in PCa, and that it may be an untapped resource for PCa biomarkers or therapeutic targets.

* Corresponding author. The James Buchanan Brady Urologic Institute and Department of Urology, Johns Hopkins School of Medicine, Baltimore, MD, USA.

E-mail address: slupold@jhmi.edu (S.E. Lupold).

Peer review under responsibility of Second Military Medical University.

1. Introduction

There is accumulating evidence that polyadenylation site selection is associated with tissue-type and cell state. Certain cells and tissue types tend to select more proximal polyadenylation signals (PAS), while others tend to select more distal. For example, the brain generally produces longer mRNA transcripts through preferential distal PAS usage, while cells in the ovary generate shorter transcripts through proximal PAS usage [1,2]. Proliferating cells and poorly differentiated tissues commonly express shorter, proximally polyadenylated mRNAs. Conversely, highly differentiated and quiescent cells express longer, distally polyadenylated mRNAs [3,4]. Cancer cells, which are characteristically proliferative and poorly differentiated, often produce shorter and proximally polyadenylated mRNAs [5–7]. While these polyadenylation patterns can be transcriptome wide, and involve thousands of mRNA transcripts, changes in PAS selection are not yet predictable for a given gene, cell type, tissue type, or biologic state. Therefore, alternative polyadenylation (APA) analyses are currently necessary to detect changes associated with each disease or condition. The extent and role of APA in prostate biology and prostate disease remains essentially undetermined. We hypothesize that APA may be of particular interest for prostate cancer (PCa) given the potential for links between APA and features associated with PCa aggressiveness, including differentiation state and Gleason grade, proliferation index and risk of biochemical recurrence, and alternative mRNA processing and therapeutic resistance. The unique prostatic transcriptome may produce novel transcript isoforms, through APA, that could serve as predictive biomarkers or therapeutic targets for PCa.

Cleavage and 3'-end polyadenylation is a requisite transcriptional process in eukaryotic messenger RNA mRNA maturation. It prepares newly transcribed mRNAs for transport to the cytoplasm and subsequent translation initiation [8]. The resulting polyadenosine (poly[A]) tail, approximately 200 nucleotides in length, enhances mRNA stability and protects the coding sequence (CDS) from 3'-exonuclease degradation. Nearly all human protein-coding transcripts undergo this critical process, with the exception of specific histone genes [9].

Recent 3'-focused RNA sequencing technologies have precisely mapped and quantified PAS usage for multiple organisms, cells, and tissue types. The majority of protein-coding genes produce multiple mRNA transcripts by differential PAS site selection [10]. This phenomenon, known as APA, often occurs within the 3'-untranslated region (UTR) and therefore does not affect the CDS or protein composition. The 3'-UTR encodes critical signals that have evolved to regulate mRNA stability, protein translation efficiency, sub-cellular localization, and downstream protein–protein interactions ([11]). For instance, distal versus proximal polyadenylation would produce an extended 3'-UTR with additional potential for microRNA (miRNA) binding, which

canonically leads to reduced protein translation and mRNA destabilization through RNA interference (RNAi) [12]. APA can also occur prematurely, within introns and coding exons, producing transcript isoforms with altered coding sequence and 3'-UTRs [13]. Several new tools have been developed to characterize the diverse mRNA isoforms produced by APA, and to illuminate their role in biology and human disease [6].

Abnormalities in mRNA 3'-end processing have been broadly linked to human disease. Numerous genes that harbor mutations within the predominant PAS sequence have been linked to hematologic, neurologic, endocrine, and immune diseases, as well as cancer [14]. Similarly, alterations within mRNA cleavage and polyadenylation genes can broadly transform mRNA polyadenylation patterns and cause disease and developmental disorders. For example, expansion of the triplet (GCG) repeats within the poly(A)-binding protein gene (*PABPN1*), causes oculopharyngeal muscular dystrophy (OPMD), and results in transcriptome-wide APA within skeletal muscle [15]. In another example, genetic ablation of the *Cstf2t* cleavage stimulation factor in mice disrupts spermatogenesis and causes infertility by altering germ cell mRNA polyadenylation [16]. In addition to genetic alterations, changes in the expression of specific mRNA cleavage and polyadenylation factors can drive APA events that significantly alter cell fate, stemness, and cancer progression. Similarly, epigenetic changes can alter chromatin and PAS accessibility, linking DNA methylation, APA, and cancer [17]. To begin to understand the potential role of APA in PCa, we begin this review by discussing the key factors and steps in canonical mRNA cleavage and polyadenylation.

2. Genes, complexes, and stages of mRNA cleavage and polyadenylation

The 3'-ends of nearly all eukaryotic mRNA transcripts are processed by endonucleolytic cleavage and untemplated polyadenylation. The first step, cleavage is mediated by three large and intricate cleavage and polyadenylation (CPA) complexes: The cleavage and polyadenylation specificity factor (CPSF) complex, the cleavage stimulatory factor (CSTF) complex, and the cleavage factor complexes (CFIm and CFIIIm) [18]. Differential expression of individual CPA complex subunits and 3'-UTR sequence composition, can determine mRNA polyadenylation site selection and therefore 3'-UTR length and composition.

Each CPA complex recognizes specific *cis*-regulatory sequences on the precursor-mRNA (pre-mRNA), as illustrated in Fig. 1. The CPSF complex recognizes and binds to the PAS sequence [19]. This complex consists of multiple protein subunits: CPSF160 (CPSF1), CPSF100 (CPSF2), CPSF73 (CPSF3), CPSF30 (CPSF4), hFip1 (FIP1L1), and WDR33. Among these, CPSF30 and WDR33 are reported to directly recognize and bind the PAS [20]. The canonical PAS sequence is the AAUAAA hexamer, although there are

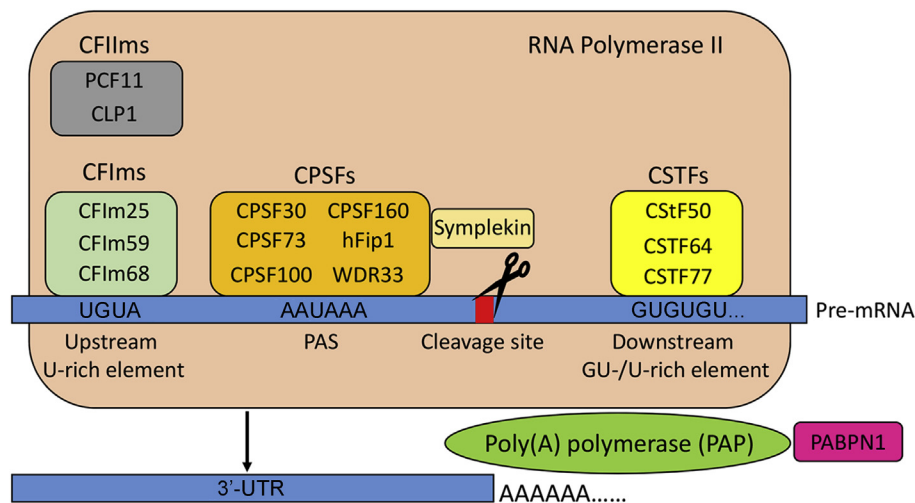


Figure 1 Schematic of the cleavage and polyadenylation complex machinery assembled onto a pre-mRNA polyadenylation site. The CPA complex proteins consist of three major factors: The CPSF complex, the CSTF complex, and the cleavage factor complexes (CFIm and CFIm), and other factors (Symplekin, PAP, and PABPN1). CPSFs consist of multiple protein subunits (CPSF30, CPSF73, CPSF100, CPSF160, hFip1, and WDR33) that collectively recognize and bind to the PAS specifically. CSTFs consist of three subunits (CSTF50, CSTF64, and CSTF77) that collectively recognize and bind to the GU-/U-rich elements downstream of the PAS. CFIm complexes consist of three subunits (CFIm25, CFIm59, and CFIm68) that collectively recognize and bind to the U-rich elements upstream of the PAS. PABPN1 and PAP are associated with the addition of untemplated adenosines. CFIm complexes (PCF11 and CLP1) contribute to transcription termination, and symplekin serves as a scaffold protein of CPA machinery. CPA, cleavage and polyadenylation; CPSF, cleavage and polyadenylation specificity factor; CSTF, cleavage stimulatory factor; PAS, polyadenylation signal; PABPN1, poly(A)-binding protein nuclear 1; PAPs, poly(A) polymerases.

numerous PAS variants that are frequently utilized in mammalian cells [21]. The CSTF complex then recognizes and binds to GU-/U-rich sequence elements located downstream of the PAS [22]. This heterotrimeric complex consists of CSTF50 (CSTF1), CSTF64 (CSTF2), and CSTF77 (CSTF3). The CSTF64 subunit interacts directly with the GU-/U-rich regions [23]. Finally, the CFIm complex recognizes and binds the UGUA motifs, or the so-called upstream U-rich element, approximately 10–35 nucleotides upstream of the PAS [24]. The CFIm complex is a heterotetramer consisting of a dimer of smaller CFIm25 (CPSF5/NUDT21) subunits, which directly bind two UGUA motifs, and two molecules of the larger CFIm68 (CPSF6) or CFIm59 (CPSF7) subunits [25]. Symplekin serves as a scaffold to hold together CPA machinery [26].

Once all complexes are bound, the CPSF-73 subunit (of the CPSF complex) cleaves the pre-mRNA approximately 15–30 nucleotides downstream of the PAS [27]. This most often occurs after a CA dinucleotide. Following 3'-cleavage, a non-templated poly(A) tail is created by poly(A) polymerases (PAPs). PAP alpha, encoded by the *PAPOLA* gene, is the canonical PAP. There are numerous additional PAP isoforms and genes that contribute to mRNA polyadenylation in different cell types and under different conditions [28].

3. Proliferation, differentiation, cancer, and alternative polyadenylation

Cancer is a disease characterized by rapid cell proliferation, loss of differentiation, apoptotic resistance, enhanced cellular motility, invasion, and metastasis. In the context of rapid cell division and altered differentiation state, cancer

cells generally have increased rates of proximal polyadenylation [5]. Proximal polyadenylation can occur within the coding region (CR-APA) or within the 3'-UTR (UTR-APA) (Fig. 2). Truncation through UTR-APA can activate proto-oncogenes, by removing miRNA binding sites and increasing mRNA stability and protein translation [5]. Premature polyadenylation within introns, or intronic polyadenylation (IPA), is also associated with cellular proliferation and cancer [29]. IPA is a form of CR-APA, because it can generate truncated protein isoforms with altered amino terminal peptides. This can be a double-edged-sword in cancer by either inactivating tumor suppressor genes or generating dominant-negative protein isoforms that act as oncogenes [29]. While transcript shortening by APA is generally associated with human cancer, it is not necessarily active in all cancer cells, and it does not affect all expressed genes [30]. It is therefore important to investigate APA in specific cell and tissue types, in order to identify mechanisms of activation and effects on gene expression and disease phenotype. These studies may identify new biomarkers of disease progression or treatment response, or new pathways and therapeutic targets. To date there is limited knowledge regarding APA prevalence in PCa, the genes most commonly regulated by APA in PCa, or the role of APA in specific PCa cells or phenotypes.

4. Oncogene activation through APA

Widespread APA in cancer can broadly affect gene expression. One major class of genes that are particularly affected by UTR-APA are oncogenes [7]. Below we briefly review selected oncogenes that can be activated by APA and consider their activity in PCa biology.

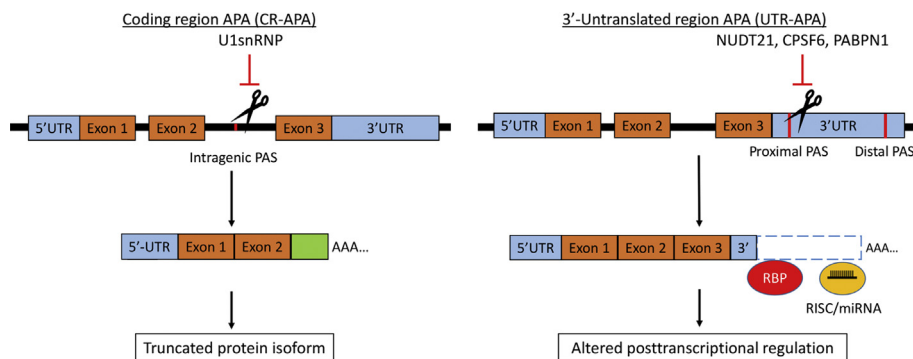


Figure 2 Schematic of CR and UTR APA. APA can be classified as CR-APA and UTR-APA. CR-APA occurs when cleavage and polyadenylation takes place within the coding region or within coding region introns, resulting in truncated protein isoforms. U1snRNP recognizes and blocks early polyadenylation signals to prevent premature cleavage and polyadenylation. UTR-APA occurs when mRNAs encode more than one polyadenylation signal within the 3'-UTR. Proximal polyadenylation generates transcript isoforms with shorter 3'-UTRs and potentially fewer miRNA and RNA binding protein binding sites. Several factors, including NUDT21, CPSF6, and PABPN1 have been reported to inhibit proximal polyadenylation. APA, alternative polyadenylation; CR, coding region; UTR, untranslated region; RBP, RNA binding proteins; PAS, polyadenylation signal.

4.1. Cyclin D1 (*CCND1*)

CCND1 is an established human proto-oncogene and critical regulator of the G1/S cell cycle transition [31]. The *CCND1* gene produces two isoforms through alternative splicing and polyadenylation, *Cyclin D1a* and *Cyclin D1b*. *Cyclin D1a* comprises exons 1–5, while *Cyclin D1b* comprises exons 1–4 due to premature polyadenylation within the fourth intron [32]. The *Cyclin D1a* transcript is also susceptible to UTR-APA, generating shorter and longer 3'-UTR isoforms [33]. *Cyclin D1b*, and the proximally polyadenylated isoform of *Cyclin D1a*, enhance proliferation and cell cycle progression [34]. *Cyclin D1* is notably susceptible to APA in a broad array of cancer cell lines and cell types [5]. PCa cell lines tend to express lower levels of *Cyclin D1* 3'-UTR, when compared to prostate epithelial cells, potentially making *Cyclin D1* expression less susceptible to miRNA regulation in PCa [35].

4.2. Cell division cycle 6 (*CDC6*)

CDC6 is an essential regulator of DNA replication and cell cycle progression [36]. In breast cancer cell lines, 17 β -estradiol (E2) induces proximal UTR-APA of *CDC6* transcripts, leading to shorter *CDC6* transcripts and increased *CDC6* protein expression in estrogen receptor positive (ER+) breast cancers [37]. This may be the first evidence that steroid signaling can regulate APA in cancer. In PCa, *CDC6* expression is induced by the activated androgen receptor (AR) through an androgen response element within the gene promoter [38]. *CDC6* expression increases with PCa progression, is correlated with AR expression, and promotes resistance to combined AR and Chk1/2 inhibition [39]. The prevalence of *CDC6* APA in PCa, and its contribution to disease progression and resistance, remains undefined.

4.3. High mobility group AT-hook 2 (*HMG2*)

HMG2 is a high mobility group, non-histone, regulator of chromatin structure [40]. In cancer, *HMG2* gene

expression is often elevated. One mechanism of increased expression is through the loss or inhibition *let-7* miRNAs, which have as many as eight binding sites in the *HMG2* 3'-UTR [41]. Genetic rearrangement also activates *HMG2* in cancer, by generating highly active and truncated *HMG2* protein isoforms, or by truncating the *HMG2* 3'-UTR and eliminating *let-7* binding sites and RNAi-mediated suppression [42]. In PCa, *HMG2* protein levels are elevated and *HMG2* over-expression leads to epithelial-to-mesenchymal transition (EMT) [43]. This is particularly apparent in treatment-induced neuroendocrine prostate cancer (t-NEPC), where elevated LIN28B expression represses *let-7d* maturation and causes elevated *HMG2* and SOX2 over-expression [44]. While *HMG2* could also escape *let-7* repression through APA, this does not appear to be a predominant mechanism in cancer. *HMG2* was investigated as part of an APA study that applied several non-transformed and cancer cell lines and focused on 23 different genes, each with multiple PASs and high potential for miRNA regulation [5]. No notable APA was reported for *HMG2* in this study, although PCa cells or tissues were not included in the analysis. Further studies are needed to determine whether the APA of *HMG2* is relevant to PCa biology.

4.4. AR

The AR is a steroid hormone receptor that is necessary for prostate development, and it is the primary therapeutic target for metastatic PCa. Castration resistant PCa (CRPC) frequently exhibits *AR* gene overexpression, amplification, mutation, or alternative splicing. *AR-V7* is a constitutively active and alternatively spliced isoform of *AR* that is associated with CRPC and resistance to abiraterone and enzalutamide therapies [45,46]. The *AR-V7* transcript includes exons 1–3 of *AR* and terminates at a single PAS within intron 3. This premature cleavage and polyadenylation (PCPA) event deletes the C-terminal ligand-binding domain of *AR*, resulting in ligand-independent *AR* activity. *AR-V7* polyadenylation can be inhibited by transfecting morpholinos

that target and block the PAS within the third intron. These morpholinos reduce *AR-V7* mRNA and protein expression, increase full-length AR protein expression, and inhibit androgen-independent growth of 22Rv1 PCa cells [47]. The targeted inhibition of CPSF1 or CPSF3, with siRNA transfection, also reduces the expression of AR-V7 and causes an increase in full-length AR transcription. These discoveries uncover PAS and CPA inhibitors as a potential new class of therapeutic targets for CRPC.

5. Aberrant polyadenylation factor expression is associated with cancer

There is emerging evidence that differential CPA factor gene expression can alter PAS site selection and cellular phenotype. Several studies have identified the CFIm complex as a critical regulator of UTR-APA. Below we discuss some of these factors and consider their activity in PCa biology.

5.1. Nudix hydrolase 21 (*NUDT21*, *CFIm25*, and *CPSF5*)

Several RNAi library screening strategies have been applied to identify genes that contribute to APA. Masamha and colleagues [48] combined CPA factor knockdown by siRNA with an RT-PCR-based assay for APA in HeLa cells. The *NUDT21* gene, which encodes the CFIm25 protein, was identified as a key suppressor of proximal PAS site selection within the 3'-UTR. In this study, *NUDT21* knockdown resulted in the 3'-UTR truncation of at least 1450 gene transcripts. In a separate study, *NUDT21* was again identified as a top hit in an RNAi screen designed to identify genes that control cell fate during pluripotent stem cell induction [49]. *NUDT21* shRNAs were among the strongest and most consistent enhancers of cell fate reprogramming, largely by enhancing the proximal 3'-UTR polyadenylation of chromatin regulatory genes. *NUDT21* knockdown has also been shown to increase the growth of U-251 and LN-229 glioblastoma (GBM) tumors in mice, while *NUDT21* over-expression inhibited growth [48]. Significant shifts in proximal PAS use were also reported in GBM patient samples with low *NUDT21* gene expression. Additional studies have supported the tumor suppressive nature of *NUDT21*, while others have found *NUDT21* to promote cancer growth, indicating that this gene may have conflicting roles in different cancers or cell types [50].

The role of *NUDT21* in PCa biology and APA remains undefined. At the time of writing this article, PubMed searches for "*NUDT21* AND Prostate", "*CPSF5* AND Prostate", or "*CFIM25* AND Prostate" provided zero results. Publicly available data suggest that *NUDT21* is very rarely mutated in PCa, with only two mutations identified in over 4000 PCa cases (primary and metastatic combined) using cBioPortal [51,52]. However, analysis of the Cancer Genome Atlas PCa dataset (TCGA-PRAD) using the UALCAN portal indicates moderately reduced *NUDT21* expression in PCa, when compared to normal tissue (-1.34 fold, $p=5.6\times 10^{-7}$) [53]. Recent studies have identified a mechanism for the post-translational degradation of the *NUDT21* protein, CFIm25, through the cancer testes antigen and ubiquitin ligase, MAGEA11 [54]. Re-expression of MAGEA11

drives ubiquitination of the CFIm25 cofactor, PCF11, leading to loss of CFIm25 from the 3' end processing complex and broad 3'-UTR transcript shortening and enhanced tumor growth. It is notable that MAGEA11 is an established AR co-regulator, and its expression is increased in aggressive PCa [55,56]. However, direct links between MAGEA11, *NUDT21*, and APA are currently lacking in PCa. Further studies are needed to determine whether MAGEA11 or androgen signaling regulates APA in PCa cells through the CFIm complex.

5.2. Cleavage and specificity factor 6 (*CPSF6* and *CFIm68*)

Using a combination of 3'-targeted mRNA sequencing and PAR-CLIP mapping of CPA factor binding sites, Martin and colleagues [57] comprehensively defined the 3'-end processing and APA of the HEK293 cell transcriptome. The results found that CFIm proteins have much higher positional specificity within mRNA cleavage sites, when compared to CPSF components. CPA factor knockdown studies identified CFIm68, encoded by the *CPSF6* gene, as a key suppressor of proximal PAS usage within the 3'-UTR. The *CPSF6* gene product, CFIm68, interacts with a long noncoding RNA (lncRNA), *CCAT2*, to affect metabolism in colon cancer [58]. The *CCAT2* gene spans a single nucleotide polymorphism (SNP) associated with higher risk of colon cancer [59]. In an allele specific fashion, *CCAT2* binds to the CFIm68 and CFIm25 protein complex, leading to altered glutaminase (GLS) splicing and polyadenylation to preferentially generate a more catalytically active isoform of GLS, known as GLS isoform C (GAC). In breast cancer, *CPSF6* has tumor promoting functions in the aggressive luminal B, HER2-overexpressing, and triple negative subtypes. In these models, *CPSF6* knockdown inhibited proliferation, colony formation, and tumor growth [60]. Immunohistochemistry analysis of tissue microarrays further found that *CPSF6* protein expression was positively correlated with aggressive breast cancer and poor patient outcomes.

The expression and function of *CPSF6* in PCa remains relatively undefined. Interestingly, *CPSF6* was identified as one of several genes that had a greater proportion of unspliced RNA in CRPC when compared to normal prostate epithelium, untreated primary PCa, and PCa cells [61]. UALCAN portal gene expression analysis of the TCGA-PRAD dataset indicates that *CPSF6* expression is moderately, but significantly, elevated in prostate tumor tissue (1.13 fold, $p=2.3\times 10^{-7}$) [53], and cBioPortal analysis suggests that *CPSF6* is mutated in less than 1% of PCa. At the time of writing this article, there have been no other reports of *CPSF6* gene expression or activity in PCa. However, the *CCAT2* lncRNA, which interacts with the CFIm25/CFIm68 complex, is over-expressed in aggressive PCa and it correlates with tumor grade and metastasis [62]. The *CCAT2* SNP, rs698326, has also been associated with prostate cancer risk [63]. Functional studies found that *CCAT2* knockdown inhibited PCa cell line growth, migration, and invasion, although it has not been determined whether this involved CFIm25/CFIm68 complex regulation or APA [62]. Further studies are needed to examine this gene and pathway in PCa.

5.3. *PABPN1*

The canonical role of *PABPN1* involves direct binding to the nascent poly(A) tail (Fig. 1), stimulation of PAP, and control of poly(A) tail length [64]. *PABPN1* binding sites have been used to map the 3'-termini of mRNA through an approach called PAPERCLIP [65]. Recent studies suggest that *PABPN1* plays an important role in APA. Similar to *NUDT21* and *CPSF5*, *PABPN1* was identified as a potent suppressor of proximal PAS selection in an RNAi screen (Fig. 2, UTR-APA) [15]. Further supporting this discovery, reduced *PABPN1* expression is associated with 3'-UTR shortening and poor prognosis of non-small cell lung cancer [66]. In a separate bioinformatic analyses of APA across multiple cancer types in the cancer genome atlas (TCGA), *PABPN1* was identified as the factor most correlated with distal PAS usage [67]. These results are consistent with the functional role of *PABPN1* as a suppressor of proximal PAS site selection. However, the elevated level of *PABPN1* in many TCGA cancer types (when compared to normal) is inconsistent with the central dogma of proximal APA in cancer, because elevated *PABPN1* would lead to 3'-UTR lengthening rather than shortening in cancer. Further studies are needed to clarify the role of *PABPN1* in APA in cancer biology and management. The expression and role of *PABPN1* in prostate tissue and PCa remains relatively undefined.

5.4. *U1 snRNP*

The *U1 snRNP* is the first spliceosomal RNA-protein complex to bind to pre-mRNA, and to initiate assembly of the spliceosome [68]. Interestingly, *U1 snRNP* also plays critical roles in polyadenylation at two different locations in the gene body. At terminal exons, the *U1 snRNP* interacts with PAP to regulate polyadenylation [69]. The *U1 snRNP* also binds to cryptic PAS within introns, to prevent PCPA and CR-APA (Fig. 2) [70]. In a process call telescripting, the *U1 snRNP* binds to PAS sequences located within the initial introns of a gene transcript and prevents PCPA and termination [71]. Telescripting is particularly important for larger genes, where hexamer PAS sequences are more abundant [72].

Functional links between *U1 snRNP* and cancer are beginning to emerge. *U1 snRNP* inhibition with *U1* antisense morpholinos (AMOs) inhibits cancer cell proliferation, migration, and invasion, and alters mRNA splicing and polyadenylation [73]. Several studies are providing remarkable links between *U1 snRNP* gene expression, cancer, DNA repair, and cyclin dependent kinase 12 (*CDK12*) gene mutation or inhibition. DNA damage induction, through ultraviolet (UV) exposure, causes a widespread increase in intronic polyadenylation and reduced *U1 snRNP* expression in RKO colon cancer cells [74]. Widespread intronic polyadenylation is also observed following *CDK12* gene depletion, leading to loss of homologous recombination repair (*HRR*) gene transcript expression through PCPA [75]. *CDK12* is an important RNA polymerase II kinase that is frequently mutated in several cancer types, including PCa. In human tumors, *CDK12* mutation is associated with increased IPA of DNA repair gene transcripts, including *ATM*, *FANCD2*, and *WRN*. Moreover, two studies found that direct inhibition of *CDK12* activity with THZ531, a *CDK12/13*

inhibitor, drives PCPA of several *HRR* genes in multiple cancer cell lines, including 22RV1 and PC-3 [75,76]. Notably, longer gene length and lower *U1 snRNP*/PAS ratios, among other factors, are associated with THZ531-induced PCPA [76]. This pathway is clinically relevant to PCa, because approximately 5% of metastatic CRPCs harbor *CDK12* mutations [77]. PCa with *CDK12* mutations exhibit a unique genomic signature of focal tandem duplications (FTDs), increased gene fusions, neoantigen production, and increased T cell infiltration.

6. APA: An untapped source for PCa biomarkers?

The majority of men diagnosed with PCa may never require treatment for their disease. These very-low and low-risk patients undergo observation or active surveillance to treat symptoms rather than to cure the cancer, or to detect cancer progression that will require future treatment. While these men have a very low risk of death from PCa, a smaller but significant percentage of men are diagnosed with, or progress to, a much more aggressive form of the disease. In fact, PCa remains one of the leading causes of cancer death for men around the world. This deviation between indolent and lethal PCa underscores a major unmet need for this disease. There is a need for novel biomarkers that can detect clinically relevant or potentially lethal PCa early. These biomarkers could be useful for both diagnosis and patient management, to distinguish aggressive from indolent disease in newly diagnosed patients, and to monitor for disease progression in patients undergoing observation or active surveillance. This information is critical to guide decisions for treatment. We posit that APA may generate novel RNA biomarkers that correlate with PCa aggressiveness.

The Gleason grading system is one of the most powerful and most consistently proven indicators of PCa aggressiveness and disease outcome [78]. It is defined by histologic patterns of tissue and glandular shape and organization that range from well differentiated, to moderately differentiated and finally to poorly differentiated carcinoma. The Gleason score (GS) increases as the tissue becomes more poorly differentiated. For several decades countless biomarkers have been measured against the GS, but very few are able to provide significant additional or independent prognostic information. One of the few additional measurements that have been repeatedly informative is cellular proliferation, which can be measured by Ki-67 staining, mitotic index, or S-phase fraction [79,80]. Given the relationships between APA, transcript shortening, cancer, proliferation, and loss of differentiation (Fig. 3), we posit that prostate-specific or PCa-associated genes may undergo APA throughout disease progression. If sufficiently prostate-specific, these APA events could serve as biomarkers in tissue or bodily fluids, such as blood or urine. Tissue specificity is paramount, because other cell types with varying proliferation and differentiation states could also be present in blood and urine.

While the expression and contribution of APA to PCa development remains to be elucidated, there is evidence of APA in at least one human dataset. Xiang and colleagues [67] performed a comprehensive analysis of APA from

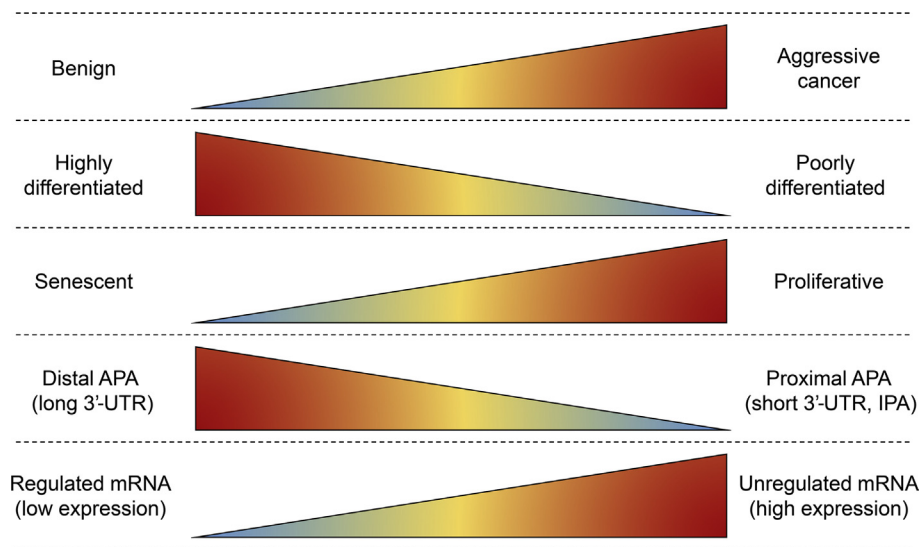


Figure 3 Alternative polyadenylation: An untapped source for prostate cancer biomarkers? Schematic to demonstrate shared relationships between APA and PCa progression. Carcinogenesis and cancer progression are associated with decreasing differentiation state and increasing cellular proliferation. Distal polyadenylation, IPA, and 3'-UTR shortening are similarly associated with proliferation, loss of differentiation, and cancer development. The mechanisms and relationships between APA and cancer remain incompletely defined. We hypothesize that these APA events could create unique transcript isoforms that could serve as diagnostic or prognostic biomarkers for prostate cancer and other urologic malignancies. APA, alternative polyadenylation; IPA, intronic polyadenylation; PCa, prostate cancer; UTR, untranslated region.

multiple TCGA datasets using an algorithm called DaPars, which calculates the Distal PolyA site Usage Index (PDUI) from standard RNA sequencing data. The study reported 12 527 APA events in the TCGA-PRAD database of primary PCa samples. Comparable levels of APA events were found in other cancer types, ranging from 11 650 in hepatocellular carcinoma to 14 547 in breast cancer. Clinically relevant APA events were defined as those that were associated with changes in overall survival; however, this cannot be reliably determined for PCa due to the low level of patient death in the TCGA-PRAD cohort. Further studies are needed to identify and characterize alternatively polyadenylated transcript isoforms associated with clinical and pathologic features of PCa including stage, grade, biochemical recurrence, metastasis, and therapeutic resistance. These events, if sufficiently prostate-specific or PCa-specific, may serve as potential new biomarkers for PCa management.

7. Conclusion

A number of studies indicate that the 3'-UTR shortening and IPA are associated with cancer and aggressive features of cancer. The regulatory mechanisms of APA are complex, and the relationships between APA and cancer development and progression remain to be fully elucidated for multiple disease types. This developing field may be a new resource for diagnostic, prognostic, and therapeutic biomarkers for PCa.

Author contributions

Study concept and design: Akira Kurozumi, Shawn E. Lupold.

Data acquisition: Akira Kurozumi, Shawn E. Lupold.

Data analysis: Akira Kurozumi, Shawn E. Lupold.

Drafting of manuscript: Akira Kurozumi, Shawn E. Lupold.

Critical revision of the manuscript: Akira Kurozumi, Shawn E. Lupold.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgements

The authors would like to acknowledge funding from the Department of Defense Prostate Cancer Research Program W81XWH-17-1-0581 and W81XWH-19-1-0450 to Shawn EL, the National Cancer Institute P30CA006973 to William G. Nelson, and The Patrick C. Walsh Cancer Research Fund to Shawn EL, for support of this work.

References

- [1] Miura P, Shenker S, Andreu-Agullo C, Westholm JO, Lai EC. Widespread and extensive lengthening of 3' UTRs in the mammalian brain. *Genome Res* 2013;23:812–25.
- [2] Ulitsky I, Shkumatava A, Jan CH, Subtelny AO, Koppstein D, Bell GW, et al. Extensive alternative polyadenylation during zebrafish development. *Genome Res* 2012;22:2054–66.
- [3] Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science* 2008;320:1643–7.
- [4] Ji Z, Lee JY, Pan Z, Jiang B, Tian B. Progressive lengthening of 3' untranslated regions of mRNAs by alternative

- polyadenylation during mouse embryonic development. *Proc Natl Acad Sci U S A* 2009;106:7028–33.
- [5] Mayr C, Bartel DP. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* 2009;138:673–84.
- [6] Gruber AJ, Zavolan M. Alternative cleavage and polyadenylation in health and disease. *Nat Rev Genet* 2019;20:599–614.
- [7] Masamha CP, Wagner EJ. The contribution of alternative polyadenylation to the cancer phenotype. *Carcinogenesis* 2018;39:2–10.
- [8] Eckner R, Ellmeier W, Birnstiel ML. Mature mRNA 3' end formation stimulates RNA export from the nucleus. *EMBO J* 1991;10:3513–22.
- [9] Marzluff WF, Wagner EJ, Duronio RJ. Metabolism and regulation of canonical histone mRNAs: Life without a poly(A) tail. *Nat Rev Genet* 2008;9:843–54.
- [10] Derti A, Garrett-Engle P, Macisaac KD, Stevens RC, Sriram S, Chen R, et al. A quantitative atlas of polyadenylation in five mammals. *Genome Res* 2012;22:1173–83.
- [11] Mayr C. What are 3' UTRs doing? *Cold Spring Harb Perspect Biol* 2019;11:a034728. <https://doi.org/10.1101/cshperspect.a034728>.
- [12] Bartel DP. MicroRNAs: Target recognition and regulatory functions. *Cell* 2009;136:215–33.
- [13] Kamieniarz-Gdula K, Proudfoot NJ. Transcriptional control by premature termination: A forgotten mechanism. *Trends Genet* 2019;35:553–64.
- [14] Curinha A, Oliveira Braz S, Pereira-Castro I, Cruz A, Moreira A. Implications of polyadenylation in health and disease. *Nucleus* 2014;5:508–19.
- [15] Jenal M, Elkon R, Loayza-Puch F, van Haaften G, Kuhn U, Menzies FM, et al. The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. *Cell* 2012;149:538–53.
- [16] Dass B, Tardif S, Park JY, Tian B, Weitlauf HM, Hess RA, et al. Loss of polyadenylation protein tauCstF-64 causes spermatogenic defects and male infertility. *Proc Natl Acad Sci U S A* 2007;104:20374–9.
- [17] Nanavaty V, Abrash EW, Hong C, Park S, Fink EE, Li Z, et al. DNA methylation regulates alternative polyadenylation via CTCF and the cohesin complex. *Mol Cell* 2020;78:752–64.e6. <https://doi.org/10.1016/j.molcel.2020.03.024>.
- [18] Colgan DF, Manley JL. Mechanism and regulation of mRNA polyadenylation. *Genes Dev* 1997;11:2755–66.
- [19] Sun Y, Zhang Y, Hamilton K, Manley JL, Shi Y, Walz T, et al. Molecular basis for the recognition of the human AAUAAA polyadenylation signal. *Proc Natl Acad Sci U S A* 2018;115:E1419–28. <https://doi.org/10.1073/pnas.1718723115>.
- [20] Chan SL, Huppertz I, Yao C, Weng L, Moresco JJ, Yates 3rd JR, et al. CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3' processing. *Genes Dev* 2014;28:2370–80.
- [21] Beaudoin E, Freier S, Wyatt JR, Claverie JM, Gautheret D. Patterns of variant polyadenylation signal usage in human genes. *Genome Res* 2000;10:1001–10.
- [22] Perez Canadillas JM, Varani G. Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein. *EMBO J* 2003;22:2821–30.
- [23] Beyer K, Dandekar T, Keller W. RNA ligands selected by cleavage stimulation factor contain distinct sequence motifs that function as downstream elements in 3'-end processing of pre-mRNA. *J Biol Chem* 1997;272:26769–79.
- [24] Yang Q, Gilmartin GM, Doublet S. The structure of human cleavage factor I(m) hints at functions beyond UGUA-specific RNA binding: A role in alternative polyadenylation and a potential link to 5' capping and splicing. *RNA Biol* 2011;8:748–53.
- [25] Yang Q, Coseno M, Gilmartin GM, Doublet S. Crystal structure of a human cleavage factor CFI(m)25/CFI(m)68/RNA complex provides an insight into poly(A) site recognition and RNA looping. *Structure* 2011;19:368–77.
- [26] Takagaki Y, Manley JL. Complex protein interactions within the human polyadenylation machinery identify a novel component. *Mol Cell Biol* 2000;20:1515–25.
- [27] Mandel CR, Kaneko S, Zhang H, Gebauer D, Vethantham V, Manley JL, et al. Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* 2006;444:953–6.
- [28] Thuresson AC, Astrom J, Astrom A, Gronvik KO, Virtanen A. Multiple forms of poly(A) polymerases in human cells. *Proc Natl Acad Sci U S A* 1994;91:979–83.
- [29] Lee SH, Singh I, Tisdale S, Abdel-Wahab O, Leslie CS, Mayr C. Widespread intronic polyadenylation inactivates tumour suppressor genes in leukaemia. *Nature* 2018;561:127–31.
- [30] Fu Y, Sun Y, Li Y, Li J, Rao X, Chen C, et al. Differential genome-wide profiling of tandem 3' UTRs among human breast cancer and normal cells by high-throughput sequencing. *Genome Res* 2011;21:741–7.
- [31] Musgrove EA, Caldon CE, Barraclough J, Stone A, Sutherland RL. Cyclin D as a therapeutic target in cancer. *Nat Rev Canc* 2011;11:558–72.
- [32] Sanchez G, Bittencourt D, Laud K, Barbier J, Delattre O, Auboeuf D, et al. Alteration of cyclin D1 transcript elongation by a mutated transcription factor up-regulates the oncogenic D1b splice isoform in cancer. *Proc Natl Acad Sci U S A* 2008;105:6004–9.
- [33] Wiestner A, Tehrani M, Chiorazzi M, Wright G, Gibellini F, Nakayama K, et al. Point mutations and genomic deletions in CCND1 create stable truncated cyclin D1 mRNAs that are associated with increased proliferation rate and shorter survival. *Blood* 2007;109:4599–606.
- [34] Wang Q, He G, Hou M, Chen L, Chen S, Xu A, et al. Cell cycle regulation by alternative polyadenylation of CCND1. *Sci Rep* 2018;8:6824. <https://doi.org/10.1038/s41598-018-25141-0>.
- [35] Elliman SJ, Howley BV, Mehta DS, Fearnhead HO, Kemp DM, Barkley LR. Selective repression of the oncogene cyclin D1 by the tumor suppressor miR-206 in cancers. *Oncogenesis* 2014;3:e113. <https://doi.org/10.1038/oncsis.2014.26>.
- [36] Borlado LR, Mendez J. CDC6: From DNA replication to cell cycle checkpoints and oncogenesis. *Carcinogenesis* 2008;29:237–43.
- [37] Akman BH, Can T, Erson-Bensan AE. Estrogen-induced upregulation and 3'-UTR shortening of CDC6. *Nucleic Acids Res* 2012;40:10679–88.
- [38] Jin F, Fondell JD. A novel androgen receptor-binding element modulates Cdc6 transcription in prostate cancer cells during cell-cycle progression. *Nucleic Acids Res* 2009;37:4826–38.
- [39] Karanika S, Karantanos T, Li L, Wang J, Park S, Yang G, et al. Targeting DNA damage response in prostate cancer by inhibiting androgen receptor-CDC6-ATR-Chk1 signaling. *Cell Rep* 2017;18:1970–81.
- [40] Cleynen I, Van de Ven WJ. The HMGA proteins: A myriad of functions (Review). *Int J Oncol* 2008;32:289–305.
- [41] Lee YS, Dutta A. The tumor suppressor microRNA let-7 represses the HMGA2 oncogene. *Genes Dev* 2007;21:1025–30.
- [42] Mayr C, Hemann MT, Bartel DP. Disrupting the pairing between let-7 and Hmga2 enhances oncogenic transformation. *Science* 2007;315:1576–9.
- [43] Hawsawi O, Henderson V, Burton LJ, Dougan J, Nagappan P, Odero-Marah V. High mobility group A2 (HMGA2) promotes EMT via MAPK pathway in prostate cancer. *Biochem Biophys Res Commun* 2018;504:196–202.
- [44] Lovnicki J, Gan Y, Feng T, Li Y, Xie N, Ho CH, et al. LIN28B promotes the development of neuroendocrine prostate cancer. *J Clin Invest* 2020;130:5338–48.

- [45] Hu R, Dunn TA, Wei S, Isharwal S, Veltri RW, Humphreys E, et al. Ligand-independent androgen receptor variants derived from splicing of cryptic exons signify hormone-refractory prostate cancer. *Canc Res* 2009;69:16–22.
- [46] Antonarakis ES, Lu C, Wang H, Luber B, Nakazawa M, Roeser JC, et al. AR-V7 and resistance to enzalutamide and abiraterone in prostate cancer. *N Engl J Med* 2014;371:1028–38.
- [47] Van Etten JL, Nyquist M, Li Y, Yang R, Ho Y, Johnson R, et al. Targeting a single alternative polyadenylation site coordinately blocks expression of androgen receptor mRNA splice variants in prostate cancer. *Canc Res* 2017;77:5228–35.
- [48] Masamha CP, Xia Z, Yang J, Albrecht TR, Li M, Shyu AB, et al. CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. *Nature* 2014;510:412–6.
- [49] Brumbaugh J, Di Stefano B, Wang X, Borkent M, Forouzmand E, Clowers KJ, et al. Nudt21 controls cell fate by connecting alternative polyadenylation to chromatin signaling. *Cell* 2018;172:106–20.e21. <https://doi.org/10.1016/j.cell.2017.11.023>.
- [50] Jafari Najaf Abadi MH, Shafabakhsh R, Asemi Z, Mirzaei HR, Sahebnasagh R, Mirzaei H, et al. CFIm25 and alternative polyadenylation: Conflicting roles in cancer. *Canc Lett* 2019;459:112–21.
- [51] Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Canc Discov* 2012;2:401–4.
- [52] Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 2013;6:pl1. <https://doi.org/10.1126/scisignal.2004088>.
- [53] Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi B, et al. UALCAN: A portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia* 2017;19:649–58.
- [54] Yang SW, Li L, Connelly JP, Porter SN, Kodali K, Gan H, et al. A cancer-specific ubiquitin ligase drives mRNA alternative polyadenylation by ubiquitinating the mRNA 3' end processing complex. *Mol Cell* 2020;77:1206–1221.e7. <https://doi.org/10.1016/j.molcel.2019.12.022>.
- [55] Bai S, He B, Wilson EM. Melanoma antigen gene protein MAGE-11 regulates androgen receptor function by modulating the interdomain interaction. *Mol Cell Biol* 2005;25:1238–57.
- [56] James SR, Cedeno CD, Sharma A, Zhang W, Mohler JL, Odunsi K, et al. DNA methylation and nucleosome occupancy regulate the cancer germline antigen gene MAGEA11. *Epigenetics* 2013;8:849–63.
- [57] Martin G, Gruber AR, Keller W, Zavolan M. Genome-wide analysis of pre-mRNA 3' end processing reveals a decisive role of human cleavage factor I in the regulation of 3' UTR length. *Cell Rep* 2012;1:753–63.
- [58] Redis RS, Vela LE, Lu W, Ferreira de Oliveira J, Ivan C, Rodriguez-Aguayo C, et al. Allele-specific reprogramming of cancer metabolism by the long non-coding RNA CCAT2. *Mol Cell* 2016;61:520–34.
- [59] Tomlinson I, Webb E, Carvajal-Carmona L, Broderick P, Kemp Z, Spain S, et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet* 2007;39:984–8.
- [60] Binohman N, Hachim IY, Lebrun JJ, Ali S. CPSF6 is a clinically relevant breast cancer vulnerability target: Role of CPSF6 in breast cancer. *EBioMedicine* 2017;21:65–78.
- [61] Sowalsky AG, Xia Z, Wang L, Zhao H, Chen S, Bubleby GJ, et al. Whole transcriptome sequencing reveals extensive unspliced mRNA in metastatic castration-resistant prostate cancer. *Mol Canc Res* : MCR 2015;13:98–106.
- [62] Zheng J, Zhao S, He X, Zheng Z, Bai W, Duan Y, et al. The up-regulation of long non-coding RNA CCAT2 indicates a poor prognosis for prostate cancer and promotes metastasis by affecting epithelial-mesenchymal transition. *Biochem Biophys Res Commun* 2016;480:508–14.
- [63] Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, Wacholder S, et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet* 2007;39:645–9.
- [64] Kuhn U, Gundel M, Knoth A, Kerwitz Y, Rudel S, Wahle E. 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. *J Biol Chem* 2009;284:22803–14.
- [65] Hwang HW, Park CY, Goodarzi H, Fak JJ, Mele A, Moore MJ, et al. PAPERCLIP identifies MicroRNA targets and a role of CstF64/64tau in promoting non-canonical poly(A) site usage. *Cell Rep* 2016;15:423–35.
- [66] Ichinose J, Watanabe K, Sano A, Nagase T, Nakajima J, Fukayama M, et al. Alternative polyadenylation is associated with lower expression of PABPN1 and poor prognosis in non-small cell lung cancer. *Canc Sci* 2014;105:1135–41.
- [67] Xiang Y, Ye Y, Lou Y, Yang Y, Cai C, Zhang Z, et al. Comprehensive characterization of alternative polyadenylation in human cancer. *J Natl Cancer Inst* 2018;110:379–89.
- [68] Stark H, Dube P, Luhrmann R, Kastner B. Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature* 2001;409:539–42.
- [69] Gunderson SI, Beyer K, Martin G, Keller W, Boelens WC, Mattaj LW. The human U1A snRNP protein regulates polyadenylation via a direct interaction with poly(A) polymerase. *Cell* 1994;76:531–41.
- [70] Kaida D, Berg MG, Younis I, Kasim M, Singh LN, Wan L, et al. U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* 2010;468:664–8.
- [71] Berg MG, Singh LN, Younis I, Liu Q, Pinto AM, Kaida D, et al. U1 snRNP determines mRNA length and regulates isoform expression. *Cell* 2012;150:53–64.
- [72] Oh JM, Di C, Venters CC, Guo J, Arai C, So BR, et al. U1 snRNP telecrypting regulates a size-function-stratified human genome. *Nat Struct Mol Biol* 2017;24:993–9.
- [73] Oh JM, Venters CC, Di C, Pinto AM, Wan L, Younis I, et al. U1 snRNP regulates cancer cell migration and invasion *in vitro*. *Nat Commun* 2020;11:1. <https://doi.org/10.1038/s41467-019-13993-7>.
- [74] Devany E, Park JY, Murphy MR, Zakusilo G, Baquero J, Zhang X, et al. Intronic cleavage and polyadenylation regulates gene expression during DNA damage response through U1 snRNA. *Cell Discov* 2016;2:16013. <https://doi.org/10.1038/celldisc.2016.13>.
- [75] Dubbury SJ, Boutz PL, Sharp PA. CDK12 regulates DNA repair genes by suppressing intronic polyadenylation. *Nature* 2018;564:141–5.
- [76] Krajewska M, Dries R, Grasseti AV, Dust S, Gao Y, Huang H, et al. CDK12 loss in cancer cells affects DNA damage response genes through premature cleavage and polyadenylation. *Nat Commun* 2019;10:1757. <https://doi.org/10.1038/s41467-019-09703-y>.
- [77] Wu YM, Cieslik M, Lonigro RJ, Vats P, Reimers MA, Cao X, et al. Inactivation of CDK12 delineates a distinct immunogenic class of advanced prostate cancer. *Cell* 2018;173:1770–82.e14. <https://doi.org/10.1016/j.cell.2018.04.034>.
- [78] Lotan TL, Epstein JI. Clinical implications of changing definitions within the Gleason grading system. *Nat Rev Urol* 2010;7:136–42.
- [79] Bubendorf L, Sauter G, Moch H, Schmid HP, Gasser TC, Jordan P, et al. Ki67 labelling index: An independent predictor of progression in prostate cancer treated by radical prostatectomy. *J Pathol* 1996;178:437–41.
- [80] Fisher G, Yang ZH, Kudahetti S, Moller H, Scardino P, Cuzick J, et al. Prognostic value of Ki-67 for prostate cancer death in a conservatively managed cohort. *Br J Canc* 2013;108:271–7.