Cleavage factor Im (CFIm) as a regulator of alternative polyadenylation

Jessica G Hardy and Chris J Norbury*

University of Oxford
Sir William Dunn School of Pathology
South Parks Road
Oxford
OX1 3RE
U.K.

* to whom correspondence should be addressed (chris.norbury@path.ox.ac.uk)

Keywords: polyadenylation / mRNA biogenesis / gene expression / post-transcriptional regulation of gene expression / cell proliferation / cancer

Abstract

Most mammalian protein coding genes are subject to alternative cleavage and polyadenylation (APA), which can generate distinct mRNA 3' untranslated regions (UTRs) with differing regulatory potential. While this process has been intensely studied in recent years, it remains unclear how and to what extent cleavage site selection is regulated under different physiological conditions. The cleavage factor I (CFIm) complex is a core component of the mammalian cleavage machinery, and the observation that its depletion causes transcriptome-wide changes in cleavage site use makes it a key candidate regulator of APA. This review aims to summarise current knowledge of the CFIm complex, and explores the evidence surrounding its potential contribution to regulation of APA.

Introduction

The closely coupled cleavage and polyadenylation reactions are key steps in the processing of most primary pre-mRNA transcripts into mature mRNA. Understanding of this process has improved considerably since the initial identification of a common AAUAAA polyadenylation signal (PAS) around 20 nt upstream from the cleavage site [1]. A large number of proteins and protein complexes co-operate to stimulate cleavage and subsequent polyadenylation in a manner that involves recognition of the PAS and other sequence elements [2,3]. Although a full understanding of how each of these components contributes to 3' end formation is lacking, the major factors involved have been characterised (Fig. 1).

Some genes have a single, constitutively used cleavage site, but the majority have at least two alternative sites – a phenomenon known as alternative polyadenylation (APA; Fig. 2). Alternative cleavage sites are usually present within the 3'UTR, generating transcripts with the same coding capacity but differing 3'UTR lengths. This 'UTR-APA' has regulatory potential, as the extended 3'UTRs generated through distal cleavage often contain additional regulatory elements, such as miRNA sites, which may influence transcript stability, localization and translation efficiency. Less commonly, in a situation known as 'coding region-APA' (CR-APA), alternative cleavage sites exist within the coding region, driving changes in protein coding capacity.

APA was first identified in the 1980s, with the demonstration that use of a proximal cleavage site in the immunoglobulin heavy chain coding region results in production of a secreted rather than membrane-bound IgM isoform [4,5]. Similar examples were soon identified in other transcripts [6], and in the past decade development of a range of 3' end-targeted sequencing techniques [7,8] has led to extensive annotation of cleavage sites transcriptome-wide. Such analyses have revealed APA to be widespread in yeast [9], plants [10] and animals [11-13], with recent estimates suggesting that around 70-80% of mammalian transcripts are subject to APA [12,14].

The finding that APA is so prevalent sparked efforts to characterize cleavage site usage in various physiological contexts. Intriguingly, a number of studies have revealed that cleavage site selection is subject to global, directional modulation under certain conditions. In particular, the onset of cell proliferation is associated with a widespread shift towards preferential usage of proximal cleavage sites (3'UTR shortening), as shown in T-cell activation [15], oncogenic transformation [16,17] and serum-induced fibroblast proliferation [18]. The potential for such global shifts to

radically alter gene expression has attracted intense interest, but several studies have suggested that the effects on RNA stability and translation are fairly modest at the global level [19-21]. Nevertheless, numerous individual transcripts show clear differential regulation of their alternative 3'UTR isoforms [16,21,22]. For example, certain membrane proteins, including CD47, change their subcellular localization depending on whether they are translated from the short or long 3'UTR isoform [23].

Thus, even in the absence of immediately obvious widespread changes in RNA stability or translation, global APA shifts may substantially alter the expression or dynamics of a subset of proteins. It is therefore of interest to understand how cleavage site choice is regulated, and how this differs between tissues and physiological states to orchestrate global changes in APA profiles.

Obvious candidates for involvement in selection of alternative cleavage sites include cleavage factors themselves, and indeed a general trend has been noted whereby increased expression of cleavage factors leads to increased use of weaker, proximal cleavage sites [18,24]. This is exemplified by the finding that simultaneous knockdown of CstF64 and its paralogue CstF64τ leads to widespread 3'UTR lengthening, through increased relative use of distal sites [25,26]. However, inhibition or depletion of other RNA-associated factors not obviously linked with the cleavage step, such as U1 snRNP [27] and PABPN1 [28], can also induce extensive, directional changes in cleavage site selection. It therefore seems likely that a range of RNA binding proteins (RBPs) and complexes, as well as other factors such as chromatin context and polymerase speed [29], may influence the choice of particular cleavage sites to some extent, probably in a transcript-specific manner. However, the challenge is to identify factors that act as 'master' global regulators of cleavage site choice *in vivo*.

One cleavage factor with a particularly well-established influence on cleavage site choice is cleavage factor I (CFIm) – a mammalian complex conserved in other multicellular organisms, but absent from yeasts [30]. CFIm is unusual, in that it goes against the observation described above that increasing cleavage factor expression favours the use of weaker, proximal sites. Rather, *knockdown* of CFIm causes pervasive 3'UTR shortening, through preferential use of proximal sites. This suggests that CFIm may have a more sophisticated regulatory role in cleavage site choice than other factors, actively promoting the use of distal sites or repressing the use of proximal sites. The rest of this review will therefore focus on CFIm and the evidence for its role as a key regulator of APA.

Cleavage factor Im (CFIm)

CFIm was first identified by biochemical fractionation as one of four nuclear components necessary and sufficient to induce cleavage of the SV40 pre-mRNA *in vitro* [31]. Further purification to near homogeneity showed that the CFIm fraction contains three polypeptides of 25, 59 and 68 kDa [32], which have since been named CFIm25 (or CPSF5/NUDT21), CFIm59 (or CPSF7) and CFIm68 (or CPSF6; Fig. 3). This purified complex was able to bind pre-mRNA independently of other cleavage factors, with all three subunits forming UV crosslinks to the substrate, and its presence stabilised the binding of a CPSF-CstF-PAP complex to the RNA [32]. Pre-incubation of CFIm with the pre-mRNA substrate also increased the rate and efficiency of cleavage in an *in vitro* assay, suggesting that binding of CFIm may be

one of the earliest steps in promoting formation of an active cleavage complex [33]. Following cloning and expression of the subunits, it was demonstrated that CFIm activity can be reconstituted by combining CFIm25 and CFIm68, suggesting some functional redundancy between the CFIm59 and CFIm68 subunits, which also show a high degree of sequence similarity [33].

These early experiments established CFIm as a factor with a central role in stabilizing the cleavage complex. Following this, a SELEX approach identified its preferred binding site on pre-mRNA as a UGUA motif [34], giving further mechanistic insight into its activity. UGUA motifs are often found around 20 nt upstream of the PAS, providing a platform to recruit CFIm to the cleavage site. UGUA-dependent CFIm recruitment generally stimulates cleavage, as supported by the observation that *PAPOLG* pre-mRNA cleavage was impaired when UGUA sites were mutated [35]. Intriguingly, it appears that CFIm binding also has the potential to suppress cleavage. The CFIm68 pre-mRNA itself has several UGUA sites, one of which overlaps the PAS. Pre-incubation of this particular pre-mRNA with high concentrations of CFIm inhibited cleavage, presumably due to steric hindrance preventing recognition of the PAS by CPSF [34]. This may be a rare example, representing an elegant CFIm autoregulatory mechanism, but highlights the possibility that CFIm activity may be suppressive as well as stimulatory, with the effect probably depending on the number and positions of UGUA sequences in the 3'UTR.

Analyses of CFIm mutants, and ultimately structural studies, further illuminated structure-function relationships in the complex (Fig. 3). The smallest subunit, CFIm25, contains an inactive NUDIX hydrolase domain and no known RNA binding motif. Despite this, it binds RNA more efficiently *in vitro* than the CFIm68 subunit, which contains an RRM [36]. A crystal structure of CFIm25 in complex with a UGUA-containing RNA molecule indicates that CFIm25 directly binds the UGUA sequence, with hydrogen bonds and stacking interactions providing sequence-specific recognition [37]. The RRM of CFIm68 and CFIm59, rather than mediating RNA binding, forms the CFIm25-interacting surface [36]. As well as the RRM, CFIm59 and CFIm68 both have a central proline-rich domain and a C-terminal RS domain, of the type usually seen in SR splicing proteins [33]. This RS domain appears to be necessary for nuclear localization and, in CFIm68, provides a platform for interaction with other SR proteins, potentially promoting links between cleavage and splicing [36].

While active CFIm was originally assumed to be a dimer of CFIm25 with one molecule of CFIm59 or CFIm68, reciprocal co-immunoprecipitations and size exclusion chromatography suggest it exists as a heterotetramer, including a homodimer of CFIm25 and two larger subunits, each of which may be CFIm59 or CFIm68 [38]. This finding has been supported by a number of crystal structures, all of which contain a CFIm25 homodimer [37,39-41]. The most informative crystal structures show a CFIm25 dimer in complex with two CFIm68 RRM domains and bound to a UGUA-containing RNA molecule [40,41]. Interestingly, these structures reveal that a CFIm heterotetramer can bind two UGUA sequences simultaneously in an antiparallel orientation, with one motif recognized by each CFIm25 molecule. Furthermore, binding assays indicate that the two UGUA sequences may be present within the same RNA, with looping out of the intervening sequence [40]. The CFIm68 RRM was shown to be important for increasing RNA binding efficiency of the complex, with mutagenesis analysis suggesting that it may directly contact the RNA and play a crucial role in directing the path of the RNA loop between UGUA

sequences [40,42].

CFIm in alternative polyadenylation

Although CFIm appears to be an essential, core component of the cleavage machinery, several observations suggest it may also have a regulatory role in selection of alternative cleavage sites. Transient siRNA-mediated knockdown of CFIm25 in HeLa cells caused a shift towards increased relative usage of proximal cleavage sites in a number of transcripts, identified by northern blotting [43]. This finding was extended to show that knockdown of CFIm68, although interestingly not its close relative CFIm59, causes the same phenotype [38]. These trends were ultimately validated at the transcriptome-wide level in HEK293 cells using a 3' end-biased sequencing approach, leading to the conclusion that the CFIm complex globally contributes to preferential selection of distal cleavage sites under normal conditions [44-46].

These observations surrounding the CFIm knockdown phenotype leave open two broad possibilities. Under normal conditions, CFIm may promote the use of distal sites more strongly than proximal sites, or alternatively it may repress the use of proximal sites. An elegant study which used CLIP to pinpoint the binding positions of CFIm within 3'UTRs, in conjunction with 3' end sequencing data, helped to test these two models [44]. This analysis demonstrated that where alternative sites are present, the position of CFIm68 binding is strongly predictive of the most frequently used cleavage site, usually the distal site [44]. Positioning of CFIm68 had the most predictive power of all factors analysed, followed closely by CstF64 and CstF64τ, and then by CFIm59 and CFIm25. This strongly supports a model whereby CFIm (and in particular CFIm containing CFIm68) preferentially binds at and promotes the use of distal sites, explaining the increased relative use of proximal sites in its absence. This idea is also supported by the finding that many transcripts subject to 3'UTR shortening upon CFIm knockdown are enriched for UGUA sequences in the region upstream of distal sites but depleted for the same motif around proximal sites [47]. However, this model is highly simplified, and it may well be that for some transcripts, there is an important repressive role of CFIm at proximal sites, which is not revealed when considering such transcriptome-wide data. Indeed, the demonstration that CFIm can bind two UGUA sequences simultaneously has led to suggestions that the complex may bind at both cleavage sites and loop out the intervening sequence, simultaneously obscuring the proximal site and promoting use of the distal site [40]. Further investigation, perhaps involving abrogating binding of CFIm at specific sites to probe the effect on APA, will be needed to clarify the stimulating versus suppressive roles of CFIm.

While the impact of knockdown of CFIm25 or CFIm68 on APA is clear, the observation that CFIm59 knockdown does not lead to the same 3'UTR shortening is surprising, and suggests that despite their structural similarity, CFIm59 and CFIm68 may have fundamentally different roles in cleavage. It has been suggested that this could result from differing protein interaction partners or post-translational modifications. For example, CFIm59 lacks the glycine-arginine rich (GAR) motif subject to arginine methylation in CFIm68 [48], and the RS region of CFIm68 mediates interactions with other SR proteins such as hTra2β, Srp20 and 9G8 [36], which have not been demonstrated for CFIm59. In addition, slight differences identified between the crystal structures of CFIm25 with the CFIm59 or CFIm68 RRM suggest that CFIm59 may have a narrower RNA exit cleft, resulting in a

different mode of RNA binding, which could explain its lack of influence on APA [42].

CFIm regulation in physiological settings

Are the molecular phenotypes seen on CFIm depletion relevant to physiological changes in cleavage site selection? There are two clear examples where changes in CFIm25 levels directly influence APA and cellular phenotype. The first comes from a study of the U251 glioblastoma cell line, which was found to have unusually low levels of CFIm25 [46]. Restoring a 'normal' level of CFIm25 in this cell line resulted in 3'UTR lengthening and decreased tumourigenic properties, both *in vitro* and in a mouse xenograft model [46]. In this case, the decrease in CFIm25 levels appears to have been key in promoting the shift towards increased proliferation and transformation, and this may well apply to other cancers. This may be particularly pertinent to brain tumours, as many neuronal transcripts have very long 3'UTRs and are therefore especially susceptible to 3'UTR shortening [49].

The second example relates to a small number of patients with neuropsychiatric disease who have natural copy number variations in CFIm25 [50]. It was found that the altered CFIm25 levels in these patients significantly affect expression of *MECP2*, a gene that is mutated in Rett Syndrome and is essential for proper brain function. Specifically, single-copy deletion of CFIm25 leads to MECP2 overexpression through increased use of the proximal cleavage site, while CFIm25 duplication promotes use of the MECP2 distal site, giving more of the inefficiently translated longer isoform and decreased protein expression. Both abnormally high and low levels of MECP2 cause neuropsychiatric disease, and so in both CFIm25 deletion and duplication, the altered CFIm25 level seems to have an impact on cellular physiology via changes in APA.

Aside from these two examples of altered CFIm25 expression in brain disease, CFIm is expressed quite uniformly across different cell types and physiological settings. For example, there is no large 'knockdown-like' change in expression seen upon cell cycle entry in fibroblasts (J.H., unpublished data), and it appears unlikely that a simple reduction in CFIm expression orchestrates global, proliferation-linked 3'UTR shortening. It therefore remains unclear how CFIm may contribute to regulating global APA changes without undergoing changes in expression.

One possibility is that the function of CFIm is specifically regulated by post-translational modification (PTM), which may alter its affinity for distal cleavage sites, or its interaction with other cleavage factors. In support of this hypothesis, de-phosphorylation of a purified 'CFm' fraction with serine/threonine phosphatases inhibited cleavage *in vitro* [51]. This fraction included CFIm and another cleavage factor, CFIIm, suggesting that one of these complexes needs to be phosphorylated to function.

Multiple potential sites of phosphorylation, acetylation, methylation and ubiquitylation, have been identified on all three subunits of CFIm [52]. Among these, K23 acetylation on CFIm25 is one of the few to have been validated experimentally [53]; K23 acetylation is catalysed by CBP, which is recruited to the CFIm complex via binding to CFIm68 [53]. One effect of this acetylation is to reduce interaction with poly(A) polymerase, which may promote disassembly of the 3' end processing complex. This example demonstrates the regulatory potential of PTMs on CFIm, although an *in vivo* APA assay suggests that removal of this acetylation site through a

K23R substitution does not have any impact on cleavage site selection in HEK293T cells (J.H., unpublished data).

Another well-studied modification on CFIm is the methylation of CFIm68 at three residues – R202, R204 and R206 – within a GAR motif. *In vitro* studies demonstrated symmetric dimethylation by PRMT5 and asymmetric dimethylation by PRMT1 within this GAR motif [48], and methylation at R202 and R204 has since been identified *in vivo* [54]. However, no biological role of this methylation has yet been identified [48].

It is clear, however, that there are still many CFIm PTMs that have not been explored with respect to a potential role in APA, and this is an area warranting further investigation.

Conclusions and perspectives

While significant progress has been made in identifying factors that influence cleavage site choice and APA, an understanding is still lacking of how these factors co-operate to select a cleavage site *in vivo,* and in particular how they may bring about global changes in cleavage site use, for example in proliferation.

Clearly CFIm normally stimulates usage of distal sites over proximal sites and is a key player in pre-mRNA cleavage. How this preference is achieved, though, remains unclear – UGUA motifs are often found upstream of both proximal and distal sites, and CFIm can be recruited to some cleavage sites in the absence of an upstream UGUA [44]. The rules governing recruitment and activity of CFIm at proximal versus distal sites clearly need to be explored further, ideally *in vivo* and transcriptome-wide, to gain a fuller understanding.

Another area for future investigation is the manner in which CFIm is regulated in the absence of changes in its expression. PTM of CFIm remains an under-explored area, and more targeted proteomics studies, for example with immunopurified CFIm, will allow more extensive annotation of the PTMs present on the complex and, importantly, how these PTMs change in different physiological situations. This could provide important insights into how CFIm function is regulated to bring about specific changes in APA, in proliferating cells for example.

Finally, it should be noted that CFIm is unlikely to be the sole factor responsible for regulating all changes in APA patterns. Many factors have been shown to influence cleavage site choice, and it appears that E2F-co-ordinated up-regulation of several cleavage factors, including CstF64, could partly contribute to the APA shift in proliferation [18]. However, given its obvious potential as a major regulator of cleavage site choice, dissecting the role of CFIm is likely to be a crucial step in unravelling the complexity of APA regulation.

Figure Legends

Figure 1: The core factors involved in cleavage and polyadenylation. A typical, 'strong' mammalian cleavage site is preceded, around 20 nucleotides upstream, by a PAS of consensus sequence A(A/U)UAAA, which is recognised by the CPSF complex. G/U-rich downstream elements (DSEs) following the cleavage site promote recruitment of the CstF complex, while UGUA sequences, often found 40-50 nt upstream of the cleavage site, are recognised by the CFIm complex. This is described as a 'tripartite' recognition mechanism, with, for example, particularly prevalent UGUA sequences or downstream GU-rich motifs able to help define the cleavage site and potentially compensate for a suboptimal PAS [35]. These complexes, along with CFIIm and poly(A) polymerase, assemble around the cleavage site and stimulate endonucleolytic cleavage by CPSF73 at a site between the PAS and DSE (often a CA dinucleotide). The free 3' end of the transcript is then subject to addition of a string of non-templated adenosine monophosphates, the poly(A) tail, by poly(A) polymerase. This schematic is highly simplified, and does not depict the interactions between the cleavage factors, nor the full range of proteins that may contribute to cleavage and polyadenylation.

Figure 2: The two major classes of alternative polyadenylation. In UTR-APA there is more than one cleavage site present within the 3'UTR, such that use of alternative sites changes the 3'UTR length without affecting coding capacity. Use of such alternative sites may lead to exclusion/inclusion of regulatory elements, such as miRNA target sites, impacting upon protein expression. CR-APA involves the use of sites upstream of the terminal 3'UTR, which may be within an exon, or in an intron as illustrated here. This leads to differences in coding capacity, yielding different protein isoforms. In this example, use of the proximal site leads to exclusion of exon 3 and extension of the exon 2 coding sequence due to use of an intronic stop codon, with the remaining intronic sequence acting as a 3'UTR. Blue boxes: exons; black lines: introns; red boxes: 3'UTR; yellow boxes: miRNA sites.

Figure 3: The CFIm complex. The CFIm complex is a heterotetramer, containing a CFIm25 dimer and two larger subunits, each of which may be CFIm59 or CFIm68. Each CFIm25 subunit binds one larger subunit. The domain organization and key features of each subunit are illustrated here, with numbers representing amino acid position. For further details see text.

1. Proudfoot, N.J. and Brownlee, G.G. (1976) 3' non-coding region sequences in eukaryotic messenger RNA. Nature **263**, 211-214
2. Shi, Y., Di Giammartino, D.C., Taylor, D., Sarkeshik, A., Rice, W.J., Yates, J.R., 3rd, Frank, J. and Manley, J.L. (2009) Molecular architecture of the human pre-mRNA 3' processing complex. Mol. Cell **33**, 365-376
3. Proudfoot, N.J. (2011) Ending the message: poly(A) signals then and now. Genes Dev. **25**, 1770-1782
4. Alt, F.W., Bothwell, A.L.M., Knapp, M., Siden, E., Mather, E., Koshland, M. and Baltimore, D. (1980) Synthesis of Secreted and Membrane-Bound Immunoglobulin-Mu Heavy-Chains Is Directed by Messenger-Rnas That Differ at Their 3' Ends. Cell **20**, 293-301
5. Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R. and Hood, L. (1980) 2 Messenger-Rnas Can Be Produced from a Single Immunoglobulin-Mu Gene by Alternative Rna Processing Pathways. Cell **20**, 313-319
6. Edwalds-Gilbert, G., Veraldi, K.L. and Milcarek, C. (1997) Alternative poly(A) site selection in complex transcription units: Means to an end? Nucleic Acids Res. **25**, 2547-2561
7. Sun, Y., Fu, Y.G., Li, Y.X. and Xu, A.L. (2012) Genome-wide alternative polyadenylation in animals: insights from high-throughput technologies. J. Mol. Cell. Biol. **4**, 352-361
8. Mueller, A.A., Cheung, T.H. and Rando, T.A. (2013) All's well that ends well: alternative polyadenylation and its implications for stem cell biology. Curr. Opin. Cell Biol. **25**, 222-232
9. Mata, J. (2013) Genome-wide mapping of polyadenylation sites in fission yeast reveals widespread alternative polyadenylation. RNA Biol. **10**, 1407-1414
10. Wu, X.H., Liu, M., Downie, B., Liang, C., Ji, G.L., Li, Q.Q. and Hunt, A.G. (2011) Genome-wide landscape of polyadenylation in Arabidopsis provides evidence for extensive alternative polyadenylation. Proc. Natl. Acad. Sci. U. S. A. **108**, 12533-12538
11. You, L., Wu, J., Feng, Y., Fu, Y., Guo, Y., Long, L., Zhang, H., Luan, Y., Tian, P., Chen, L., Huang, G., Huang, S., Li, Y., Li, J., Chen, C., Zhang, Y., Chen, S. and Xu, A. (2015) APASdb: a database describing alternative poly(A) sites and selection of heterogeneous cleavage sites downstream of poly(A) signals. Nucleic Acids Res. **43**, D59-67
12. Derti, A., Garrett-Engele, P., Macisaac, K.D., Stevens, R.C., Sriram, S., Chen, R., Rohl, C.A., Johnson, J.M. and Babak, T. (2012) A quantitative atlas of polyadenylation in five mammals. Genome Res. **22**, 1173-1183
13. Tian, B., Hu, J., Zhang, H.B. and Lutz, C.S. (2005) A large-scale analysis of mRNA polyadenylation of human and mouse genes. Nucleic Acids Res. **33**, 201-212
14. Hoque, M., Ji, Z., Zheng, D., Luo, W., Li, W., You, B., Park, J.Y., Yehia, G. and Tian, B. (2013) Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. Nat. Methods **10**, 133-139
15. Sandberg, R., Neilson, J.R., Sarma, A., Sharp, P.A. and Burge, C.B. (2008) Proliferating cells express mRNAs with shortened 3 ' untranslated regions and fewer microRNA target sites. Science **320**, 1643-1647

16. Mayr, C. and Bartel, D.P. (2009) Widespread Shortening of 3 ' UTRs by Alternative Cleavage and Polyadenylation Activates Oncogenes in Cancer Cells. Cell **138**, 673-684
17. Morris, A.R., Bos, A., Diosdado, B., Rooijers, K., Elkon, R., Bolijn, A.S., Carvalho, B., Meijer, G.A. and Agami, R. (2012) Alternative Cleavage and Polyadenylation during Colorectal Cancer Development. Clin. Cancer Res. **18**, 5256-5266
18. Elkon, R., Drost, J., van Haaften, G., Jenal, M., Schrier, M., Vrielink, J.A.F.O. and Agami, R. (2012) E2F mediates enhanced alternative polyadenylation in proliferation. Genome Biol. **13**, R59
19. Spies, N., Burge, C.B. and Bartel, D.P. (2013) 3 ' UTR-isoform choice has limited influence on the stability and translational efficiency of most mRNAs in mouse fibroblasts. Genome Res. **23**, 2078-2090
20. Gruber, A.R., Martin, G., Muller, P., Schmidt, A., Gruber, A.J., Gumienny, R., Mittal, N., Jayachandran, R., Pieters, J., Keller, W., van Nimwegen, E. and Zavolan, M. (2014) Global 3' UTR shortening has a limited effect on protein abundance in proliferating T cells. Nat. Commun. **5**, 5465
21. Neve, J., Burger, K., Li, W., Hoque, M., Patel, R., Tian, B., Gullerova, M. and Furger, A. (2016) Subcellular RNA profiling links splicing and nuclear DICER1 to alternative cleavage and polyadenylation. Genome Res. **26**, 24-35
22. Lianoglou, S., Garg, V., Yang, J.L., Leslie, C.S. and Mayr, C. (2013) Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. Genes Dev. **27**, 2380-2396
23. Berkovits, B.D. and Mayr, C. (2015) Alternative 3' UTRs act as scaffolds to regulate membrane protein localization. Nature **522**, 363-367
24. Ji, Z., Lee, J.Y., Pan, Z., Jiang, B. and Tian, B. (2009) Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. Proc. Natl. Acad. Sci. U. S. A. **106**, 7028-7033
25. Yao, C.G., Biesinger, J., Wan, J., Weng, L.J., Xing, Y., Xie, X.H. and Shi, Y.S. (2012) Transcriptome-wide analyses of CstF64-RNA interactions in global regulation of mRNA alternative polyadenylation. Proc. Natl. Acad. Sci. U. S. A. **109**, 18773-18778
26. Yao, C., Choi, E.A., Weng, L., Xie, X., Wan, J., Xing, Y., Moresco, J.J., Tu, P.G., Yates, J.R., 3rd and Shi, Y. (2013) Overlapping and distinct functions of CstF64 and CstF64tau in mammalian mRNA 3' processing. RNA **19**, 1781-1790
27. Berg, M.G., Singh, L.N., Younis, I., Liu, Q., Pinto, A.M., Kaida, D., Zhang, Z., Cho, S., Sherrill-Mix, S., Wan, L. and Dreyfuss, G. (2012) U1 snRNP determines mRNA length and regulates isoform expression. Cell **150**, 53-64
28. Jenal, M., Elkon, R., Loayza-Puch, F., van Haaften, G., Kuhn, U., Menzies, F.M., Vrielink, J.A.F.O., Bos, A.J., Drost, J., Rooijers, K., Rubinsztein, D.C. and Agami, R. (2012) The Poly(A)-Binding Protein Nuclear 1 Suppresses Alternative Cleavage and Polyadenylation Sites. Cell **149**, 538-553
29. Elkon, R., Ugalde, A.P. and Agami, R. (2013) Alternative cleavage and polyadenylation: extent, regulation and function. Nat. Rev. Genet. **14**, 496-506
30. Darmon, S.K. and Lutz, C.S. (2012) mRNA 3' end processing factors: a phylogenetic comparison. Comp. Funct. Genomics **2012**, 876893
31. Takagaki, Y., Ryner, L.C. and Manley, J.L. (1989) Four factors are required for 3'-end cleavage of pre-mRNAs. Genes Dev. **3**, 1711-1724

32. Ruegsegger, U., Beyer, K. and Keller, W. (1996) Purification and characterization of human cleavage factor Im involved in the 3' end processing of messenger RNA precursors. J. Biol. Chem. **271**, 6107-6113
33. Ruegsegger, U., Blank, D. and Keller, W. (1998) Human pre-mRNA cleavage factor Im is related to spliceosomal SR proteins and can be reconstituted in vitro from recombinant subunits. Mol. Cell **1**, 243-253
34. Brown, K.M. and Gilmartin, G.M. (2003) A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor Im. Mol. Cell **12**, 1467-1476
35. Venkataraman, K., Brown, K.M. and Gilmartin, G.M. (2005) Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. Genes Dev. **19**, 1315-1327
36. Dettwiler, S., Aringhieri, C., Cardinale, S., Keller, W. and Barabino, S.M. (2004) Distinct sequence motifs within the 68-kDa subunit of cleavage factor Im mediate RNA binding, protein-protein interactions, and subcellular localization. J. Biol. Chem. **279**, 35788-35797
37. Yang, Q., Gilmartin, G.M. and Doublie, S. (2010) Structural basis of UGUA recognition by the Nudix protein CFI(m)25 and implications for a regulatory role in mRNA 3' processing. Proc. Natl. Acad. Sci. U. S. A. **107**, 10062-10067
38. Kim, S., Yamamoto, J., Chen, Y.X., Aida, M., Wada, T., Handa, H. and Yamaguchi, Y. (2010) Evidence that cleavage factor Im is a heterotetrameric protein complex controlling alternative polyadenylation. Genes Cells **15**, 1003-1013
39. Coseno, M., Martin, G., Berger, C., Gilmartin, G., Keller, W. and Doublie, S. (2008) Crystal structure of the 25 kDa subunit of human cleavage factor Im. Nucleic Acids Res. **36**, 3474-3483
40. Yang, Q., Coseno, M., Gilmartin, G.M. and Doublie, S. (2011) 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 **19**, 368-377
41. Li, H., Tong, S., Li, X., Shi, H., Ying, Z., Gao, Y., Ge, H., Niu, L. and Teng, M. (2011) Structural basis of pre-mRNA recognition by the human cleavage factor Im complex. Cell Res. **21**, 1039-1051
42. Yang, Q., Gilmartin, G.M. and Doublie, S. (2011) 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. **8**, 748-753
43. Kubo, T., Wada, T., Yamaguchi, Y., Shimizu, A. and Handa, H. (2006) Knock-down of 25 kDa subunit of cleavage factor Im in Hela cells alters alternative polyadenylation within 3 '-UTRs. Nucleic Acids Res. **34**, 6264-6271
44. Martin, G., Gruber, A.R., Keller, W. and Zavolan, M. (2012) 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 Reports **1**, 753-763
45. Gruber, A.R., Martin, G., Keller, W. and Zavolan, M. (2012) Cleavage factor I-m is a key regulator of 3 ' UTR length. RNA Biol. **9**, 1405-1412
46. Masamha, C.P., Xia, Z., Yang, J., Albrecht, T.R., Li, M., Shyu, A.B., Li, W. and Wagner, E.J. (2014) CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. Nature **510**, 412-416

47. Li, W., You, B., Hoque, M., Zheng, D., Luo, W., Ji, Z., Park, J.Y., Gunderson, S.I., Kalsotra, A., Manley, J.L. and Tian, B. (2015) Systematic profiling of poly(a)+ transcripts modulated by core 3' end processing and splicing factors reveals regulatory rules of alternative cleavage and polyadenylation. PLoS Genet. **11**, e1005166
48. Martin, G., Ostareck-Lederer, A., Chari, A., Neuenkirchen, N., Dettwiler, S., Blank, D., Ruegsegger, U., Fischer, U. and Keller, W. (2010) Arginine methylation in subunits of mammalian pre-mRNA cleavage factor I. RNA **16**, 1646-1659
49. Miura, P., Shenker, S., Andreu-Agullo, C., Westholm, J.O. and Lai, E.C. (2013) Widespread and extensive lengthening of 3 ' UTRs in the mammalian brain. Genome Res. **23**, 812-825
50. Gennarino, V.A., Alcott, C.E., Chen, C.A., Chaudhury, A., Gillentine, M.A., Rosenfeld, J.A., Parikh, S., Wheless, J.W., Roeder, E.R., Horovitz, D.D., Roney, E.K., Smith, J.L., Cheung, S.W., Li, W., Neilson, J.R., Schaaf, C.P. and Zoghbi, H.Y. (2015) NUDT21-spanning CNVs lead to neuropsychiatric disease and altered MeCP2 abundance via alternative polyadenylation. Elife **4**
51. Ryan, K. (2007) Pre-mRNA 3' cleavage is reversibly inhibited in vitro by cleavage factor dephosphorylation. RNA Biol. **4**, 26-33
52. Hornbeck, P.V., Zhang, B., Murray, B., Kornhauser, J.M., Latham, V. and Skrzypek, E. (2015) PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Res. **43**, D512-520
53. Shimazu, T., Horinouchi, S. and Yoshida, M. (2007) Multiple histone deacetylases and the CREB-binding protein regulate pre-mRNA 3'-end processing. J. Biol. Chem. **282**, 4470-4478
54. Guo, A., Gu, H., Zhou, J., Mulhern, D., Wang, Y., Lee, K.A., Yang, V., Aguiar, M., Kornhauser, J., Jia, X., Ren, J., Beausoleil, S.A., Silva, J.C., Vemulapalli, V., Bedford, M.T. and Comb, M.J. (2014) Immunoaffinity enrichment and mass spectrometry analysis of protein methylation. Mol. Cell. Proteomics **13**, 372-387