CFIm25 links alternative polyadenylation to glioblastoma tumour suppression

Chioniso P. Masamha1*, Zheng Xia2*, Jingxuan Yang3, Todd R. Albrecht1, Min Li1, Ann–Bin Shyu1, Wei Li2 & Eric J. Wagner1

The global shortening of messenger RNAs through alternative polyadenylation (APA) that occurs during enhanced cellular proliferation represents an important, yet poorly understood mechanism of regulated gene expression1–3. The 3’ untranslated region (UTR) truncation of growth-promoting mRNA transcripts that relieves intrinsic microRNA- and AU-rich-element-mediated repression has been observed to correlate with cellular transformation4; however, the importance to tumorigenicity of RNA 3’-end-processing factors that potentially govern APA is unknown. Here we identify CFIm25 as a broad repressor of proximal poly(A) site usage that, when depleted, increases cell proliferation. Applying a regression model on standard RNA-sequencing data for novel APA events, we identified at least 1,450 genes with shortened 3’ UTRs after CFIm25 knockdown, representing 11% of significantly expressed mRNAs in human cells. Marked increases in the expression of several known oncogenes, including cyclin D1, are observed as a consequence of CFIm25 depletion. Importantly, we identified a subset of CFIm25-regulated APA genes with shortened 3’ UTRs in glioblastoma tumours that have reduced CFIm25 expression. Downregulation of CFIm25 expression in glioblastoma cells enhances their tumorigenic properties and increases tumour size, whereas CFIm25 overexpression reduces these properties and inhibits tumour growth. These findings identify a pivotal role of CFIm25 in governing APA and reveal a previously unknown connection between CFIm25 and glioblastoma tumorigenicity.

Recently, it has become increasingly clear that mRNA 3’-end formation is subject to dynamic regulation under diverse physiological conditions1–3. Over 50% of human genes have multiple polyadenylation signals, thereby increasing the potential diversity in mRNA transcript length4. The formation of mRNA transcripts using these distinct poly(A) sites (PASs) is carried out by APA, with the most common form involving differential use of alternative PASs located within the same terminal exon (reviewed in ref. 7). Processing at the PAS most proximal to the stop codon (pPAS) removes negative regulatory elements that reduce mRNA stability or impair translation efficiency, such as AU-rich elements (AREs)5 and microRNA (miRNA) targeting sites6,7. It has been reported that both rapidly proliferating cells5,6 and transformed cells8,11 preferentially express mRNAs with shortened 3’ UTRs. Despite these observations, the mechanisms that control the extensive distal-to-proximal PAS switch observed in proliferative and/or transformed cells, the relationship between cause and effect, and the critical target genes subject to this regulation, are not well characterized.

To measure relative changes in endogenous APA events, we devised a quantitative polymerase chain reaction after reverse transcription (qRT–PCR) assay to monitor the transcript-specific use of the distal PAS (dPAS) while normalizing for total mRNA levels for three test transcripts, cyclin D1 (CCND1), DICER1 and TIMP2, known to undergo APA1,12. Using this approach, we readily detected appreciable usage of dPASs for all three genes in HeLa cells (Extended Data Fig. 1). This was somewhat surprising given their highly transformed state, but is consistent with previous reports that not all transformed cells tested exhibit appreciable 3’ UTR shortening1–3. Previous studies implicate multiple members of the cleavage and polyadenylation (CPA) machinery as potentially regulating poly(A) site selection12–14. To test the relative contribution of these factors to the APA of the three test genes, we used systematic RNA interference (RNAi) (Fig. 1a–c). We observed only small changes in the relative use of the dPAS after knockdown of members of the cleavage and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CSTF) and cleavage factor IIIm (CFIm) complexes (Fig. 1d–f). By contrast, we detected significant reduction in dPAS usage after knockdown of the members of the CFIm complex. These results are consistent with a recent report that CFIm68 depletion decreases 3’ UTR length15; however, the most notable PAS switching was found to occur after knockdown of CFIm25. We therefore focused all further analyses on CFIm25.

Traditional methods of global PAS profiling use mRNA partitioning and digestion to sequence poly(A) junctions within messages16,17. To identify global targets of CFIm25 with a more streamlined approach requiring less sample manipulation, we performed high-depth (>3 × 10⁶ reads) RNA sequencing (RNA-seq) after knocking down CFIm25 in parallel with a control knockdown. We determined that 23% of RNA-seq reads can be uniquely mapped to 3’ UTRs of expressed genes leading to approximately 200-fold sequence coverage (Extended Data Fig. 2a, b). We first analysed the three test genes and observed markedly reduced read density within the 3’ UTRs in response to CFIm25 depletion (Fig. 2a). These results not only confirm our qRT–PCR findings that HeLa cells robustly use the dPAS for all three test genes under basal conditions but also demonstrate that considerable 3’ UTR shortening induced by CFIm25 knockdown is readily visualized by analysing the read density of RNA-seq data.

On the basis of this promising observation, we applied a novel bioinformatics algorithm termed ‘dynamic analysis of alternative polyadenylation from RNA-seq’ (DaPars; see Methods) for the de novo identification of all instances of 3’ UTR alterations between control and CFIm25 knockdown cells, regardless of a pre-annotated dPAS within each RefSeq transcript. DaPars uses a linear regression model to identify the exact location of this novel proximal 3’ UTR as the optimal fitting point (Fig. 2b, red point) as well as the abundance of both novel and annotated UTRs. The degree of difference of 3’ UTR usage between the samples was then quantified as a change in percentage dPAS usage index (APDU1), which is capable of identifying lengthening (positive index) or shortening (negative index) within the 3’ UTR. When applied to the 12,273 RefSeq transcripts whose average terminal exon sequence coverage is more than 30-fold, DaPars identified 1,453 transcripts possessing a significant, reproducible shift in 3’ UTR usage in response to CFIm25 depletion (Fig. 2c and Extended Data Fig. 2c, d). Notably, among this group of transcripts, 1,450 are shifted to pPAS usage in CFIm25 knockdown cells. We found a significant enrichment of the CFIm25 UGUA binding motif and previously reported CFIm25 iCLIP sequence tags13 within 3’ UTRs that shortened after CFIm25 knockdown relative to transcripts exhibiting no length change (Extended Data Fig. 3).

1Department of Biochemistry and Molecular Biology, The University of Texas Medical School at Houston, Houston, Texas 77030, USA. 2Division of Biostatistics, Dan L Duncan Cancer Center and Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, Texas 77030 Texas, USA. 3The Vivian L. Smith Department of Neurosurgery, The University of Texas Medical School at Houston, Houston, Texas 77030, USA.

*These authors contributed equally to this work.
Research

PDUIs in control and CFIm25 knockdown cells where mRNAs significantly identified based on DaPars. The coverage.

knockdown (KD) in HeLa cells. Numbers on and upstream exon(s) after the control (Con.) siRNA treatment and CFIm25

Figure 2 | The DaPars algorithm identifies broad targets of CFIm25 in standard RNA-seq data. a, RNA-seq read density for 3’ UTR, terminal exon and upstream exon(s) after the control (Con.) siRNA treatment and CFIm25 knockdown (KD) in HeLa cells. Numbers on y-axis indicate RNA-seq read coverage. b, Diagram depicts how the differential alternative 3’ UTR usage was identified based on DaPars. The y-axis shows the fitting value of the DaPars regression model and the locus with minimum fitting value (red point) is the predicted alternative pPAS for the RNA-seq data (bottom). c, Scatterplot of PDUIs in control and CFIm25 knockdown cells where mRNAs significantly shortened (n = 1,450) or lengthened (n = 3) after CFIm25 knockdown (false discovery rate (FDR) ≤ 0.05, absolute ΔPDUI ≥ 0.2 and at least twofold change of PDUIs between CFIm25 knockdown and control cells) are coloured. The shifting towards pPAS is significant (P < 2.2 × 10^{-16}, binomial test). d, Correlation between dPAS site usage and gene expression levels of control and CFIm25 knockdown cells. The x-axis shows ΔPDUI; a negative value indicates that pPAS is prone to be used in CFIm25 knockdown cells. The y-axis shows the logarithm of the expression level of genes from the factors presented in the same order as isolated from cells represented in panels a–c with the factors presented in the same order as in western blots a–c. See Methods for quantification details.

Figure 1 | CFIm25 depletion leads to consistent and robust 3’ UTR shortening of test genes. a–c, Western blot analysis of HeLa cell lysates treated with control siRNA (Con.) and siRNAs individually targeting each of members of the CPA machinery and Symplekin (Sym.). In all cases, tubulin (Tub.) was used as a loading control. d–f, Quantified results of three biologically independent qRT–PCR experiments on RNA isolated from cells represented in panels a–c with the factors presented in the same order as shown in western blots a–c. See Methods for quantification details.
Moreover, we determined that 70% of transcripts whose 3’ UTR is shortened after CFIm25 knockdown use a pPAS within the first one-third of their 3’ UTR. By contrast, only 29% of multi-PAS transcripts that did not alter 3’ UTR length in response to CFIm25 have an annotated pPAS in the first third of their 3’ UTR. This demonstrates that CFIm25 APA targets are enriched with pPASs positioned close to the stop codon to maximize their degree of 3’ UTR shortening. Collectively, these results clearly indicate that the function of CFIm25 is to broadly repress proximal poly(A) site choice, and consequently, the shortening of 3’ UTR length is considerable for the majority of CFIm25-regulated transcripts upon its depletion.

One potential consequence of 3’ UTR shortening in CFIm25 knockdown is the loss of miRNA-binding sites and/or AREs, resulting in truncated mRNA transcripts that evade negative regulation. Although the correlation between transcript expression change and APDUI was modest (Pearson correlation = −0.25), it does reveal that transcripts with shorter 3’ UTR in CFIm25 knockdown cells have overall higher expression levels (Fig. 2d). We observed that 64% of transcripts with shortened 3’ UTRs exhibited significantly increased steady-state levels, 34% were unchanged, and only 2% were significantly reduced (Extended Data Fig. 4).

We have also organized the list of CFIm25-regulated genes with respect to their APDUI score, change in relative levels of transcript, and predicted numbers of ARE motifs and miRNA target sites lost after APA (Supplementary Table 1) and observed that gene expression positively correlates with the number of lost ARE motifs and miRNA target sites (Extended Data Fig. 5). Several examples of novel genes whose APA is regulated by CFIm25 are shown in Fig. 2e and it is important to note that not all long 3’ UTRs were observed to shorten in response to CFIm25 knockdown, indicating that the CFIm complex regulates many, but not all genes capable of APA (Fig. 2f). Collectively, these data demonstrate the power and ease of the DaPars algorithm to identify APA within standard RNA-seq, and indicate that the major form of CFIm25 regulation is to repress pPAS choice at a global level.

To validate the APDUI results, we created qRT–PCR amplicons to monitor dPAS usage of six genes whose 3’ UTRs were found to be shortened after CFIm25 knockdown and two that were not altered. Using these amplicons, we analysed RNA isolated from cells effectively depleted of CFIm25 using two independent short interfering RNAs (siRNAs) (Fig. 3a, inset), and observed high congruence between qRT–PCR results and those obtained using RNA-seq and APDUI (Fig. 3a, graph). To test formally for the presence of de-repressed protein expression from mRNAs with shortened 3’ UTRs, we measured their levels in lysates from knockdown cells (Fig. 3b). We observed considerable increases in protein levels of CFIm25 target genes, including several that have a well-documented role in tumour growth, such as cyclin D1, glutaminase and methyl-CpG-binding protein 2 (MECP2)21–25. It is worth noting that the 3’ UTR of each of these genes has been shown to be subject to miRNA-mediated inhibition23–25. Consistent with this observation, we also noted enhanced cellular proliferation in response to knockdown of CFIm25 relative to control knockdown in HeLa cells (Fig. 3c). Finally, to determine whether the 3’ UTR is sufficient to elicit translational derepression of a heterologous protein in response to CFIm25 knockdown, we used reporters with the SMOC1 3’ UTR cloned downstream of luciferase or the GAPDH 3’ UTR, which was not found to alter its poly(A) site usage. We observed that only the luciferase activity specifically resulting from the luciferase–SMOC1 reporter was increased in response to knockdown of CFIm25 (Fig. 3d), supporting the idea that the increased expression of endogenous SMOC1 protein when CFIm25 is depleted is mediated through its 3’ UTR.

The collective observations that CFIm25 depletion leads to broad 3’ UTR shortening, enhanced expression of growth promoting genes and increased cell proliferation support the hypothesis that CFIm25 is a novel anti-proliferative gene whose levels may be reduced in human cancers. We focused our analysis on glioblastoma, as recent reports indicate that brain tissue possesses the longest 3’ UTRs26,27. We reasoned that tumours derived from these cells might be more sensitive to changes in CFIm25 levels than other cancers. To test this prediction, we downloaded archived patient RNA-seq data from The Cancer Genome Atlas (TCGA), stratified it according to CFIm25 expression, and analysed it using DaPars. Indeed, following the same cut-offs in our HeLa RNA-seq analysis, we identified 60 genes with altered 3’ UTRs, with 59 of those experiencing shortening in glioblastoma expressing lower levels of CFIm25 (Fig. 4a and Supplementary Table 2). Among those genes, a significant number of events (24 genes; P = 2.2 × 10−12 by hypergeometric testing) were also shortened in CFIm25 knockdown HeLa cells and this percentage of overlap increased markedly to 86% as the APDUI cut-off was increased from 0.2 to 0.4 (Extended Data Fig. 6). Two representative examples of genes, FOS-related antigen 2 (FRA2; also known as FOSL2) and MECP2, with shortened 3’ UTRs in low CFIm25-expressing glioblastoma tumours is shown in Fig. 4b, demonstrating a compelling similarity between the patient samples and HeLa cells before and after CFIm25 knockdown. Overexpression of either of these genes has been shown to enhance cell proliferation18,28.

To test formally whether altering CFIm25 expression can modulate glioblastoma tumorigenic properties, we screened a panel of glioblastoma cell lines and observed that U251 cells naturally express lower levels of CFIm25 compared with LN229 cells (Fig. 4c). To raise CFIm25 levels in U251 cells, we created cell lines stably expressing either Myc-tagged CFIm25 or green fluorescent protein (GFP) as a control. In parallel, we used RNAi to reduce CFIm25 levels in LN229 cells (Fig. 4c). We observed a significant reduction in anchorage-dependent growth and cellular invasion in U251 cells overexpressing CFIm25 compared with the GFP control, whereas reducing CFIm25 in LN229 cells caused an increase in both of these properties (Extended Data Fig. 7). To determine if the altered in vitro properties of glioblastoma cells affected tumour growth kinetics in vivo, we used a subcutaneous xenograft model. Increased expression of CFIm25 in U251 cells resulted in a marked reduction in tumour growth and decreased tumour cell proliferation (Fig. 4d and Extended Data Fig. 8). By contrast, depletion of CFIm25 in LN229 cells caused a profound increase in tumour size (Fig. 4e and Extended Data Fig. 9).
Data Fig. 9). Collectively, these results uncover a tumour suppressive property of CFIm25 in glioblastoma that is probably mediated through its broad repression of APA-dependent mRNA 3′ UTR shortening.

We identified CFIm25 among 15 cleavage and polyadenylation factors as a key factor that broadly regulates APA. Importantly, the data presented here also extend our understanding of APA in regulated gene expression through the demonstration that extensive shortening of 3′ UTRs causally leads to enhanced cellular proliferation and tumorigenicity, probably through the upregulation of growth promoting factors, such as cyclin D1. These results indicate the importance of 3′ UTR usage in cell growth control and underscore the need for further research into the mechanism and regulation of APA and its potential links to other human diseases.

**METHODS SUMMARY**

Human cell lines used were cultured using standard techniques. RNAi and western blot experiments were conducted as described previously32. For luciferase experiments, one day after the second siRNA hit, cells were transfected with 3′ UTR Renilla luciferase plasmids and activity was assayed after 24 h. Total RNA for qRT–PCR was reverse transcribed using MMLV-RT (Invitrogen). qRT–PCR reactions were performed using SYBRGREEN (Fermentas). Duplicate control and CFIm25 knockdown samples were sequenced by HiSeq 2000. RNA-seq reads were aligned (hg19) to pPAS in the low CFIm25 group is significant ($P < 2.2 \times 10^{-16}$; binomial test).

Figure 4 | Altered expression of CFIm25 modulates glioblastoma tumour growth. a. The global analysis of 3′ UTR changes in glioblastoma (GBM) patient samples with either high or low level of CFIm25. Scatterplot of PDUIs from both data sets using the same cut-offs as in Fig. 2c. The shifting of PDUIs from both data sets using the same cut-offs as in Fig. 2c. The shifting of PDUIs of high CFIm25 group to pPAS in the low CFIm25 group is significant ($P < 2.2 \times 10^{-16}$; binomial test).

b. Representative UCSC Genome Browser images of RNA-seq data, demonstrating 3′ UTR shortening after CFIm25 knockdown in HeLa cells and in glioblastoma patient samples with high (blue) or low CFIm25 expression (red). KD, knockdown.

c. Western blot analysis of lysates from two glioblastoma cell lines. Note that the overexpressed Myc–CFIm25 also increases endogenous CFIm25 levels in U251 cells. Tub, tubulin; Unt., untreated.

d. Growth comparison of U251 tumours overexpressing either GFP (control) or CFIm25. Data represent the average of ten mice per group. Right panel shows representative haematoxylin and eosin (H&E) and Ki67 staining of U251 GFP tumours (top) or U251 CFIm25 tumours (bottom). Scale bars, 200 μm.

e. Growth comparison of LN229 tumours derived from cells transduced with lentiviruses expressing a scrambled short hairpin RNA (shRNA) (control) or with lentiviruses expressing shRNA targeting CFIm25. Data represent the average of ten mice per group. Right panel shows representative H&E and Ki67 staining of LN229 tumours expressing shRNA targeting CFIm25 (top) or LN229 tumours expressing scrambled shRNA (bottom). Scale bars, 200 μm.
21. Liu, W. et al. Reprogramming of proline and glutamine metabolism contributes to the proliferative and metabolic responses regulated by oncogenic transcription factor c-MYC. Proc. Natl Acad. Sci. USA 109, 8983–8988 (2012).

22. Wang, J. B. et al. Targeting mitochondrial glutaminase activity inhibits oncogenic transformation. Cancer Cell 18, 207–219 (2010).

23. Klein, M. E. et al. Homeostatic regulation of MeCP2 expression by a CREB-induced microRNA. Nature Neurosci. 10, 1513–1514 (2007).

24. Deshpande, A. et al. 3′UTR mediated regulation of the cyclin D1 proto-oncogene. Cell Cycle 8, 3592–3600 (2009).

25. Gao, P. et al. c-Myc suppression of miR-23a/b enhances mitochondrial glutaminase expression and glutamine metabolism. Nature 458, 762–765 (2009).

26. Miura, P., Shenker, S., Andreu-Agullo, C., Westholm, J. O. & Lai, E. C. Widespread and extensive lengthening of 3′ UTRs in the mammalian brain. Genome Res. 23, 812–825 (2013).

27. Ulitsky, I. et al. Extensive alternative polyadenylation during zebrafish development. Genome Res. 22, 2054–2066 (2012).

28. Nakayama, T. et al. Aberrant expression of Fra-2 promotes CCR4 expression and cell proliferation in adult T-cell leukemia. Oncogene 27, 3221–3232 (2008).

29. Wagner, E. J. & Garcia-Blanco, M. A. RNAi-mediated PTB depletion leads to enhanced exon definition. Mol. Cell 10, 943–949 (2002).

30. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25, 1105–1111 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank members of the E.J.W., A.-B.S. and W.L. laboratories for helpful discussions, P. Carpenter for reviewing the manuscript, and Q. Zhu and T. Shan of LC Sciences for their efforts on the RNA-seq. This work was supported by a CPRIT grant to E.J.W. and A.-B.S. (RP100107), and in part by a Department of Defense grant to E.J.W. (W81XWH-11-1-0304), National Institutes of Health (NIH) grants to A.-B.S. (GM046454) and E.J.W. (CA167752 and CA166274), and an endowment from Houston Endowment (to A.-B.S.). Work in the W.L. laboratory is funded by grants from the Department of Defense (W81XWH-10-1-0501), CPRIT (RP110471-C3) and NIH (R01HG007538). Work in the M.L. laboratory is funded by grants from the Dr Marnie Rose Foundation and the William and Ella Owens Medical Research Foundation. C.P.M. acknowledges a Department of Defense Postdoctoral Visionary Award (W81XWH-12-1-0218).

Author Contributions A.-B.S., W.L. and E.J.W. designed the study. C.P.M., T.R.A. and J.Y. performed the described experiments with conceptual advice from M.L. W.L. and Z.X. conducted bioinformatic analyses and developed the DaPars algorithm. C.P.M., Z.X., W.L., A.-B.S. and E.J.W. wrote the manuscript.

Author Information Raw sequence data has been deposited in the Gene Expression Omnibus under accession number GSE42420. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.J.W. (Eric.J.Wagner@uth.tmc.edu), A.-B.S. (Ann-Bin.Shyu@uth.tmc.edu) or W.L. (WL1@bcm.edu).
METHODS

RNA-seq. We used whole transcriptome RNA-seq to investigate alternative PAS usage in a genome-wide fashion. Two control and two CFIM25 knockdown samples were sequenced by HiSeq2000 (LC Sciences). Paired-end RNA-seq reads with 101 bp in each end were aligned to the human genome (hg19) using TopHat 2.0.18. RefSeq gene expressions were quantified by RSEM10. A statistical summary of read alignments and average gene expressions can be found in Extended Data Fig. 2. More than 12,000 (~50%) human RefSeq genes can be detected through RNA-seq with expression levels more than 1 fragments per kilobase of transcript sequence per million mapped paired-end reads (FPKM)11. More importantly, the average of 23% of RNA-seq reads can be uniquely mapped to 3′ UTRs of expressed genes that renders around 200× coverage on UTRs. All the TCGA glioblastoma RNA-seq BAM files were downloaded from the UCSC Cancer Genomics Hub (CGHub: https://cghub.ucsc.edu/).

Analysis of APA from RNA-seq. We used a novel bioinformatics algorithm DaPars (Z.K. et al., unpublished observations; https://code.google.com/p/dapars/) for the de novo identification of APA from RNA-seq. The observed sequence coverage was represented as a linear combination of novel and annotated 3′ UTRs. For each RefSeq transcript with annotated PAS, we used a regression model to infer the end point of alternative novel PAS within this 3′ UTR at single nucleotide resolution, by minimizing the deviation between the observed read coverage and the expected read coverage based on a two-PAS model, in both control and CFIM25 knockdown samples simultaneously.

To quantify the relative PAS usage, we defined the percentage of dPAS usage for each sample as PDUI index. The greater the PDUI is, the more the dPAS of a transcript is used and vice versa.

APDUI. We used the following three criteria to detect the most significant shifted 3′ UTR events: First, given the expression levels of short and long 3′ UTRs in two samples in each condition, we compute the significance of the difference of mean PDUs using Fisher’s exact test, which is further adjusted by Benjamini–Hochberg regulatory sequences such as AREs reside in 3′ UTRs of each sample as PDUI index. The greater the PDUI is, the more the dPAS of a transcript is used and vice versa.

qRT–PCR. After appropriate transfections, total RNA was extracted using TRizol Reagent (Life Technologies) using the manufacturer’s protocol. For qRT–PCR the miRNA was reverse transcribed using MMLV–RT (Invitrogen) using the manufacturer’s protocol to generate cDNA. The qRT–PCR reactions were performed using Stratagene MxPro3000P (Agilent Technologies) and SYBRGREEN (Fermentas). Common primers were designed to target the open reading frame and normalize for total mRNA. The distal primers were designed to target sequences just before the dPAS and detect long transcripts that use the dPAS. All primers are used shown below. Data were calculated using a modified version of the 2−ΔΔCT method to show changes in dPAS usage, where CT is the threshold cycle. First, the CT values for the common and distal amplicons were normalized to the levels of 7SK, where ΔCT (common or distal) = CTcommon or distal − CT7SK. Then ΔACT = ΔCTdistal − ΔCTcommon (note that we applied the correction factor for difference in amplification efficiency calculated in Extended Data Fig. 1). To show fold changes normalized to the control siRNA-transfected samples the following equation was used: normalized ΔACT = ΔACTaverage target siRNA − ΔACTaverage of control siRNA. Then the decrease (−) or increase (+) in dPAS usage was calculated as:

**Differentially expressed gene expression analysis.** With two replicates in each group, we used edger29 to call differentially expressed genes with FDR < 0.05. To better quantify gene expression with shorter 3′ UTRs, we counted reads based on the coding regions of each transcript.

Cell culture and cell counts. All the cell lines used (HeLa, U251 and LN229) were cultured in DMEM supplemented with 10% FBS (~1% penicillin and streptomycin) and 1% penicillin and streptomycin.

Luciferase assays. One day after a second hit with siRNA (as described earlier), HeLa cells were transfected with 0.25 µg of gene-specific 3′ UTR Renilla luciferase plasmids (SMOCl and GAPDH from Switchgear Genomics) using Lipofectamine

©2014 Macmillan Publishers Limited. All rights reserved
Generation of stable cell lines. LN229 cells were transfected with CFIm25-specific shRNA or control shRNA using polybrene in 6-well plates. Two days after lentiviral transfection cells were transfected with a second hit of lentivirus. Selection was done using 1 μg ml⁻¹ of puromycin over 2 weeks. U251 cells were transfected with either GFP or CFIm25 expressing pcDNA3 plasmids using Lipofectamine 2000 (Invitrogen) according to the manufacturer’s protocol. Selection was performed over 1–2 weeks using 2.5 mg ml⁻¹ of G418.

Soft agar assay. Soft agar assays were used to determine anchorage-dependent growth. For the base layer, 1% of UltraPure low melting point agarose (Invitrogen) was mixed 1:1 with 2× DMEM media and plated in 6-well plates giving a 1.5 ml bottom layer of 0.5% agar. Then 3 × 10⁵ cells of LN229 shRNA stably transfected cells were triturated into 2× DMEM and mixed with an equal volume of 0.6% agar to give a 0.3% layer and 1.5 ml was dispensed into each well. The agar was covered with 1 ml of 1× DMEM and incubated in a humidified incubator at 37 °C (5% CO₂). Fresh media was added once a week. After 2 weeks, colonies formed were stained with 0.01% crystal violet, and the number of invading cells was counted at 20 magnification in 10 fields for each well.

Statistical tests. Unless otherwise specified, experiments were done using three biological replicates and data are shown as average ± s.d., and statistical analysis was done using a two-tailed student t-test.

Subcutaneous xenograft tumour model. Hsd:Athymic Nude-Foxn1nu nude mice at age 5–6 weeks were used. For each cell line (LN229 or U251), 20 male nude mice were randomly assigned into two groups (n = 10). Stably transfected LN229 and U251 cells were resuspended in pure culture medium with the concentration of 3 × 10⁶ cells ml⁻¹. One-hundred-microlitre cell suspensions (3 × 10⁶ cells) were inoculated subcutaneously into the lower right flank of the mice using a 27-gauge needle. Tumour diameters are measured with digital callipers, and the tumour volume in mm³ is calculated by the formula: volume = (width)² × length/2. The tumour size data were collected and processed blindly. The animal experiments were performed under the Institutional Review Board approved animal protocol AWC-13-115.
Extended Data Figure 1 | Design and optimization of the qRT–PCR assay to monitor APA of three test genes. a, Schematic denotes the relative location of the common and distal primer annealing sites in each test gene and the approximate locations of the annotated proximal and distal poly(A) sites, depicted as pPAS and dPAS, respectively. The numbers demarcate where the 3' UTR starts and ends according to ENSEMBL. b, Ethidium-stained agarose gel of RT–PCR products of equal cycle number from the different amplicons using HeLa cell mRNA. c, Both the common and distal cyclin D1 amplicons were cloned into the same pcDNA3 plasmid in tandem. Three dilutions of each plasmid were made and amplified individually with each amplicon in triplicate. The two lines on the graph depict the amplification curve for the common and distal amplicons. The expectation is that identical cycle threshold (CT) values should be attained for each, given that the PCR reactions were conducted using identical amounts of starting material. The average of three individual experiments is shown for each dilution and the average CT deviation of either amplicon at all of the dilutions was calculated as a correction factor. d, The experiment shown in c was repeated for DICER1 and TIMP2 to determine their respective correction factors, which was then applied to experiments shown in Fig. 1.
Extended Data Figure 2 | Summary of RNA-seq alignment and reproducibility of PDUI and CFIm25-depletion-induced 3' UTR shortening. a, RNA-seq read statistics of the four biologically independent experiments where HeLa cells were treated with either control siRNA (Control) or CFIm25 siRNA (CFIm25KD). Pie chart on the right represents genomic distribution of reads that were mapped to human genome hg19. The percentage was calculated by averaging all samples. CDS, coding region. b, Histogram of gene expression of RefSeq genes with fragments per kilobase of transcript sequence per million mapped paired-end reads (FPKM) no less than 1. c, Scatterplot of the two biological replicates for each condition with high Pearson correlation ($r \geq 0.9$) demonstrating a high level of reproducibility between sample PDUI scores. Each dot represents the PDUI of a transcript. d, Genome browser screen images from four independent RNA-seq experiments. Each represents an independent biological sample where HeLa cells were transfected with either the control siRNA (Con.) or an siRNA that knocked down CFIm25. Both VMA21 and SPCS3 were found to undergo 3' UTR shortening after CFIm25 knockdown whereas FHL1 was found not to change.
Extended Data Figure 3 | Shortened transcripts have more UGUA CFIm25-binding motifs than unaltered transcripts. 

a, CFIm25 is known to bind to the UGUA motif. The number of UGUA motifs within the 3' UTRs of genes with 3' UTR shortening after CFIm25 knockdown relative to genes with unaltered 3' UTRs was calculated and compared. Here we selected the genes without 3' UTR change according to them having a ΔPDU1 value ≤ 0.05. 

b, iCLIP tags from ref. 14 (Gene Expression Omnibus accession number GSE37398) were superimposed onto data derived from PDUI analysis of CFIm25 knockdown cells. The box plot demonstrates the enrichment of CFIm25 binding within 3' UTRs that are altered after CFIm25 knockdown ($P = 6.1 \times 10^{-10}$, t-test).
Extended Data Figure 4 | Gene expression changes of genes with shortened 3’ UTRs. Pie chart was calculated from the list of 1,450 genes exhibiting shortened 3’ UTRs due to CFIm25 knockdown (dn, down). Differentially expressed gene analysis was performed using edgeR with FDR ≤ 0.05 (see Methods).
Extended Data Figure 5 | The Pearson correlation between gene expression fold change and the number of lost negative regulatory elements. Left, the number of lost AREs (AU-rich elements) due to 3' UTR shortening was calculated using the ARE database and plotted against change in gene expression levels after CFIm25 knockdown (KD). Right, similar to the left except the number of lost patented miRNA target sites (Targetscan 6.2) was plotted.
Extended Data Figure 6 | Overlap between shortening events in glioblastoma with low CFIm25 and shortening events in HeLa cells after CFIm25 knockdown. Left, y-axis (red) represents the percentage of shortening events in low CFIm25 glioblastoma that are also shortened in HeLa cells after CFIm25 knockdown. Right, y-axis (blue) shows the number of shortening events in low CFIm25 glioblastoma (GBM) against different ΔPDUI cut-offs.
Extended Data Figure 7 | Overexpression of CFIm25 reduces invasion and colony formation whereas CFIm25 depletion increases invasion and colony formation. **a**, U251 cells were transfected with either GFP or CFIm25. Top left, Cells were replated in soft agar and the number of colonies/clusters formed were determined. Bottom left, Matrigel invasion assay for cells overexpressing CFIm25 or GFP. **b**, Top right, LN229 cells were transfected with either control or two different lentiviral plasmids targeting CFIm25 (KD1 and KD2). Stably transfected cells were plated on soft agar and the resulting colonies were counted for KD1 and KD2, respectively. Bottom right, LN229 cells were transfected with either control or two different siRNAs (KD1 and KD2) directed against CFIm25 and were replated for a Matrigel invasion assay. All the experiments were done in biological triplicates and shown is the mean ± s.d. All P values were from the two-tailed student t-test of the control versus sample. *P < 0.1, **P < 0.01, ***P < 0.001.
Extended Data Figure 8 | Overexpression of CFIm25 in U251 tumours reduces their size and weight. a, b, U251 subcutaneous (s.c.) xenograft tumours were isolated from nude mice on day 84 after implantation and measured for volume (a) and weight (b) \((n = 10)\). U251-GFP indicates control U251 cells expressing GFP and U251-CFIm25 indicates cells transduced with a lentivirus that overexpresses CFIm25.
Extended Data Figure 9 | Reduction in CFIm25 expression levels enhances LN229 tumour size and weight. a, b, LN229 subcutaneous (s.c.) xenograft tumours were isolated from nude mice on day 40 after implantation and measured for volume (a) and weight (b) (n = 10). LN229-shCon. indicates control lentiviral transduced cells and LN229-shCFIm25 indicates cells transduced with a lentivirus that expresses shRNA targeting CFIm25.