bioRxiv preprint doi: https://doi.org/10.1101/373704; this version posted July 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

# The RNA-binding proteins CELF1 and ELAVL1 cooperatively control alternative splicing

Géraldine David[1], Stéphane Deschamps[1], Sergi Padilla-Parra[1,2], Marc Tramier[1,3], Yann Audic[1*], Luc Paillard[1*]

1 Univ Rennes, CNRS, IGDR-UMR 6290, F-35000 Rennes, France

2 Present address Division of Structural Biology, University of Oxford, The Henry Wellcome Building for Genomic Medicine, Headington, Oxford OX3 7BN, UK

3 Univ Rennes, CNRS, Inserm, BIOSIT-UMS 3480, US_S 018, F-35000 Rennes, France

* Corresponding authors: LP (Tel +33 22323 4473, luc.paillard@univ-rennes1.fr) and YA (Tel +33 22323 4475, yann.audic@univ-rennes1.fr)

**Character count 55,808**

**Running title** CELF1 and ELAVL1: alternative splicing partners

bioRxiv preprint doi: https://doi.org/10.1101/373704; this version posted July 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## ABSTRACT

ELAVL1 and CELF1 are two RNA-binding proteins involved in alternative splicing control. To address their functional relationships, we identify the differentially spliced mRNAs upon depletion of CELF1, ELAVL1, or both. These proteins control similar sets of genes with similar consequences on exon inclusion or skipping. The magnitude of the effect of the double depletion equals the sum of the magnitudes of the individual depletions, showing that CELF1 and ELAVL1 additively control their target RNAs. CELF1 and ELAVL1 regulated splicing events include *ACSL4*, *WNK1*, *CD44*, *MICAL3*, and *JDP2*. Using FRET, we find that CELF1 and ELAVL1 directly interact in cell nuclei. We demonstrate that the combined levels of *CELF1* and *ELAVL1* is a valuable biomarker in breast cancer, while their levels bring very limited information when taken individually. A “co-RNA splicing map” of CELF1 and ELAVL1 shows they repress alternative splice sites when bound nearby, but activate them when bound further away. Together, these data point to strong functional interactions between CELF1 and ELAVL1 to control alternative splicing with significant impacts in human pathology.

## KEYWORDS

Gene expression/Post-transcriptional control/Breast cancer/CD44

bioRxiv preprint doi: https://doi.org/10.1101/373704; this version posted July 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## INTRODUCTION

Alternative splicing is the process that allows the synthesis of several different species of mRNA molecules from a single gene. Deep sequencing approaches in the last decade revealed that virtually all human genes produce at least two alternative mRNA isoforms, largely in a tissue-specific manner (Wang *et al*, 2008; Pan *et al*, 2008; Merkin *et al*, 2012; Barbosa-Morais *et al*, 2012). Key splicing signals are the 5' and 3' splice sites, and the branch point. They are strictly required for splicing and their core sequences are highly conserved. Consequently, alternative splice control is seldom exerted through these sites themselves. Rather, cis-acting splicing regulatory elements (splice enhancers or silencers) are located in introns or exons. They regulate sites that may be tens or even hundreds of nucleotides apart. Their combinatory effects define a splice "code" that determines the mRNA to be produced from processing a pre-mRNA molecule in any given cell type or stage of differentiation (Wang & Burge, 2008; Fu, 2004).

Splicing regulatory elements essentially act through the binding of RNA-binding proteins (RBPs). CLIP (cross-linking and immunoprecipitation) approaches, combined with deep RNA sequencing, have made it possible to identify genome-wide the pre-mRNAs interacting with a given RBP. By integrating this information with transcriptome data, RNA splicing maps link the positions of the binding sites of an RBP with its regulatory functions (Park *et al*, 2016; Ule *et al*, 2006; Zhang & Darnell, 2011; Witten & Ule, 2011). They’re aimed at predicting splicing regulation by site-specific RBP binding. However, pre-mRNAs are bound by multiple RBPs, and the splicing output depends on the relationships between these RBPs. For example, QKI competes with PTBP1 to control the inclusion of *CAPZB* exon 9 (Hall *et al*, 2013), and HNRNPLL and HNRNPL antagonistically control the splicing of *CHRNA1* mRNA (Rahman *et al*, 2013). Conversely, RBFOX1 stimulates HNRNPM-mediated control of splicing of many exons (Damianov *et al*, 2016). Hence, considering the splicing regulations exerted by one RBP at a time is not sufficient to foresee splicing patterns. However, except for RBFOX1 and HNRNPM, we still miss a global view of how pairs of RBPs functionally interact, provided by experimental work in living cells and at a

bioRxiv preprint doi: https://doi.org/10.1101/373704; this version posted July 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

genome-wide scale (Dassi, 2017).

Here, we addressed the functional relationships of two RBPs in alternative splicing control, CELF1 (CUGBP Elav-Like Family Member 1, also named CUGBP1) and ELAVL1 (ELAV Like RNA Binding Protein 1, also named HuR). These two proteins share a number of structural and functional properties. Both interact with their RNA ligands through three specific domains named 'RNA recognition motifs'. The two first motifs share 31% identity between CELF1 and ELAVL1, and the third one 46%. Both proteins have been reported to control alternative splicing. For instance, CELF1 regulates the splicing of cardiac troponin T, insulin receptor, or ryanodine receptor 1 and this is important for myotonic dystrophy and muscle diseases (Kalsotra *et al*, 2008; Philips *et al*, 1998; Tang *et al*, 2015). ELAVL1 promotes the splicing of *EIF4ENIF1* pre-mRNA to an isoform producing an unstable encoded protein that is an inhibitor of cap-dependent translation, hence globally enhancing translation (Chang *et al*, 2014). Both proteins oligomerize. For CELF1, oligomerization is mediated by a short domain lying between the second and the third RNA recognition motif, and is required for high affinity RNA binding (Cosson *et al*, 2006). ELAVL1 oligomerization is mediated by the third RNA recognition motif itself, and is also required for RNA interaction (Fialcowitz-White *et al*, 2007). The capacity of CELF1 and ELAVL1 to interact with themselves questions their capacity to interact with each other. Indeed, they co-immuno-precipitate, indicating that they are associated in complexes on the same mRNAs (Le Tonquèze *et al*, 2010). Finally, the phenotypes caused by the disruption of their respective genes in mice are related. Inactivating either *Elavl1* or *Celf1* on a pure genetic background is embryonic lethal (Cibois *et al*, 2012; Katsanou *et al*, 2009). Conditional inactivation of *Elavl1* (Chi *et al*, 2011) and constitutive inactivation of *Celf1* on a mixed genetic background (Boulanger *et al*, 2015) led to the conclusion that ELAVL1 and CELF1 are both required for spermatogenesis.

Despite this resemblance, ELAVL1 and CELF1 are highly divergent. CELF1 is significantly longer than ELAVL1 (489 and 326 aminoacids for the human reference isoforms, respectively), due to a longer linker region between the second and the third RNA recognition motif likely to

bioRxiv preprint doi: https://doi.org/10.1101/373704; this version posted July 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

confer new properties. Most importantly, their RNA binding specificities differ, as CELF1 interacts with GU-rich elements and ELAVL1 with AU-rich elements. CELF1 binding to GU-rich elements most often leads to mRNA decay (Vlasova *et al*, 2008; Masuda *et al*, 2012) and controls translational efficiency (Cibois *et al*, 2010; Chaudhury *et al*, 2016), whereas ELAVL1 binding to AU-rich elements generally stabilizes bound mRNAs (Mukherjee *et al*, 2011; García-Domínguez *et al*, 2011; Lafarga *et al*, 2009; Chen *et al*, 2002; Peng *et al*, 1998). The limited similarity of CELF1 and ELAVL1 makes this pair of proteins attractive to understand RBP relationships in alternative splicing control. Here, we identify the differentially spliced mRNAs upon depletion of CELF1, ELAVL1, or both. We find that these two RBPs control similar sets of genes with similar consequences on exon inclusion or skipping. Together, they additively control the splicing of their target genes. They directly interact in cell nuclei, and belong therefore to the same regulatory complex. We demonstrate the joint prognostic value of *CELF1* and *ELAVL1* mRNA levels in breast cancer that is hidden when these levels are taken into account separately. Finally, a "co-RNA splice map" reveals the principles linking CELF1 and ELAVL1 binding to a pre-mRNA with their joint regulatory functions. These data point to strong functional interactions between CELF1 and ELAVL1 that were not expected given their weak conservation and their dissimilar molecular functions, and underline the clinical importance of considering the combinatorial molecular functions of RNA-binding proteins.

## RESULTS

### CELF1 and ELAVL1 similarly control pre-mRNA splicing.

To identify the genes controlled by CELF1 and/or ELAVL1, we depleted these proteins separately or simultaneously by RNAi in HeLa cells. Full knockdown of these proteins was toxic to the cells, so we used siRNA quantities that depleted the proteins by no more than 75%. Of note the two proteins were knocked down equally with single and double siRNAs (Figure S1A). Therefore there was no apparent cross regulation between the two proteins. Having established our

bioRxiv preprint doi: https://doi.org/10.1101/373704; this version posted July 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

experimental conditions, we extracted RNAs from control and depleted cells in order to analyze RNA splicing by exon array hybridization. In exon arrays, probes are designed to measure the inclusion of individual exons. Exon arrays are used to infer splicing patterns with performances comparable to deep RNA sequencing (Zhang *et al*, 2015; Raghavachari *et al*, 2012). For each of the 117,841 probes retained after filtration, we calculated a normalized exon value as the fluorescence intensity of the probe normalized by the expression of the gene the probe belongs to. Next, we calculated the splicing index (SI) as the log-ratio of normalized exon value in a test (depletion) to the control (control siRNA-transfected cells). Identical conditions cluster together (Figure S1B).

We identified 2,868 and 3,130 differentially regulated exons in 2,166 and 2,077 target genes respectively upon depletion of CELF1 and ELAVL1 (Table S1, $|SI|>0.2$, $p<0.05$). 664 genes were targets of both proteins, which is 2.6 higher than expected by chance ($p<2.2x10^{-16}$, Chi-square test). Hence, the depletion of CELF1 and ELAVL1 affects the splicing of largely overlapping sets of genes. We used the annotation developed in the MISO framework to classify the major classes of alternative RNA processing events (Katz *et al*, 2010). We found that the distributions of differential RNA processing events upon depletion of either protein are very similar (Figure S2, $p = 0.24$, Chi-squared test). The most enriched event in both cases is alternative last exon (ALE), and tandem UTR is also highly enriched (Figure 1A). Therefore, CELF1 and ELAVL1 control a similar wide range of splicing events, but their most striking shared feature is the control of 3' terminal exons.

We next carried out Gene Ontology (GO) term enrichment analysis to retrieve the functional profiles of the different sets of genes. Table S2 lists the GO terms and their enrichment *p*-values in the targets of CELF1 or ELAVL1. We selected the 15 GO terms that are most over-represented within the targets of CELF1 or ELAVL1. Figure 1B shows the enrichment *p*-values of these 18 terms, as 12 are shared between these two datasets. The GO terms that are highly enriched upon depleting one protein are also highly enriched upon depletion of the other protein. A treemap generated by REVIGO summarizes these data (Supek *et al*, 2011), by removing redundant GO terms and grouping related terms into “superclusters” (Figures S3A-B). Of note, 6 out of 7

bioRxiv preprint doi: https://doi.org/10.1101/373704; this version posted July 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

superclusters are shared between the targets of CELF1 and ELAVL1 (Figure S3C). Together, these data indicate that CELF1 and ELAVL1 regulate sets of genes with similar functional profiles.

We next plotted for each of the 117,841 probes the splicing index measured upon ELAVL1 versus CELF1 depletion (Figure 1C). These values are significantly correlated ($R^2$ = 0.31). When we considered only the probes with splicing indices above 0.2 in absolute value, we found that 5,316 (2,885 + 2,431, 94.6 %) indices have the same sign while only 303 (144 + 159, 5.4%) have opposite signs. Interestingly, the dynamic range of splicing indices upon CELF1 depletion is higher than upon ELAVL1 depletion, and the slope of the regression line is well below 1 (0.49). This suggests that depleting ELAVL1 has a smaller effect on alternative splicing than depleting CELF1. Nevertheless, these data show that CELF1 and ELAVL1 control the inclusion or exclusion of exonic regions essentially in the same direction.

The differentially spliced RNAs include direct and indirect targets of CELF1 and ELAVL1. The proteins bind to their direct targets, while the abnormal splicing of indirect targets is a secondary consequence of direct target defective regulation. Two CLIP-seq experiments carried out in HeLa cells identified the RNA ligands of CELF1 (Le Tonquèze *et al*, 2016) and of ELAVL1 (Uren *et al*, 2011). Respectively 2,951 and 4,284 RNAs associate with CELF1 and ELAVL1, and 1,475 associate with both, which is significantly more than the 710 expected by chance (Tables S3-S4, $p<2.2x10^{-16}$, Chi-square test). This suggests that the two proteins tend to bind to the same RNAs. We examined for the presumptive direct targets of both CELF1 and ELAVL1 the relationships between the splicing indices measured upon ELAVL1 and CELF1 depletion (Figure 1D). The correlation is high ($R^2$ = 0.77), and only 3 of 99 splicing indices have opposite signs. Hence, CELF1 and ELAVL1 similarly control the splicing patterns of their ligand RNAs.

**The controls exerted by CELF1 and ELAVL1 on pre-mRNA processing are additive.**

We used concepts of genetic interactions to describe the relationships between the depletions of CELF1, ELAVL1, or both (Drees *et al*, 2005). Here, each depletion is taken as a single genetic

bioRxiv preprint doi: https://doi.org/10.1101/373704; this version posted July 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

perturbation leading to a phenotype (RNA splicing patterns). The distribution of MISO alternative processing events in doubly depleted cells is very similar to those in CELF1 or ELAVL1-depleted cells (Figure S4A). The most enriched events are alternative last exon (ALE) and tandem UTR (Figure S4B), like in individual depletions. At a first glance, this suggests that depleting CELF1, ELAVL1, or both, have the same effects, revealing an "asynthetic" interaction. However, 10,483 exonic regions are significantly differentially included upon depletion of both CELF1 and ELAVL1 with the same criteria as above (Table S1, $|SI|>0.2$, $p<0.05$). Three quarters of them (7,523 / 10,483) were not classified as differential in single knockdowns (Figure S4C). Hence, the transcriptome of doubly depleted cells strongly differs from the transcriptomes of uniquely or non-depleted cells (also see Figure S1B). We assessed if the consequences of depleting CELF1 and ELAVL1 on their presumptive direct targets are additive. To do so, we plotted their splicing indices upon double depletion versus the sum of the splicing indices from each single depletion (Figure 1E). These values are high correlated ($R^2 = 0.94$). Furthermore, the slope of the regression line is close to 1 (1.07). A slope above 1, meaning a more severe defect in the double depletion than expected from the defects in each individual depletion, would have revealed a synergy between the depletions of CELF1 and ELAVL1. Here, depleting CELF1 and ELAVL1 have strictly additive consequences on RNA splicing patterns.

**Validation of CELF1 and ELAVL1-mediated regulation of splicing**

Next we independently confirmed the microarray results by RT-PCR on a panel of presumptive direct targets of both CELF1 and ELAVL1 (see Figure 1D). *ACSL4* encodes an enzyme involved in lipid metabolism. Its first coding exon (d-e in Figure 2A, left) is bordered by three alternative 3' splice sites. A protein translated from a transcript containing the region labelled "d" in Figure 2A (green) contains an additional N-terminal hydrophobic region with a different sub-cellular localisation (Küch *et al*, 2014). Depending on the 3' splice site used and on the inclusion of a short exon in the 5' UTR ("b"), three main isoforms are present ("ae", "abde", "abcde", Figure 2A

bioRxiv preprint doi: https://doi.org/10.1101/373704; this version posted July 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

right). A probe that hybridizes to region "d" was identified as differential. Its positive splicing index upon depletion of both CELF1 and ELAVL1 is close to the sum of the splicing indices following individual depletions. Both proteins bind to *ACSL4* RNA close to that region. This suggests that the two proteins additively and directly favor the distal 3' splice site. In accordance with the microarray data, we found in RT-PCR that depleting CELF1, ELAVL1, or both, favors the long isoforms obtained with the proximal 3' splice sites (Figure 2A, right, lanes 3-5), as compared with all the negative controls (lanes 1, 2, 6). Hence, CELF1 and ELAVL1 favor a short, cytosolic, ACSL4 isoform. *WNK1* encodes a kinase with key functions in the kidney. It includes two consecutive cassette exons, referred to as 11 (b in Figure 2B, left) and 12 (c). These exons are regulated in a tissue-specific manner. They encode a binding site for the E3 ubiquitin ligase NEDD4-2 that regulates WNK1 protein levels (Vidal-Petiot *et al*, 2012; Roy *et al*, 2015). CELF1 and ELAVL1 bind close to exon 11 (b), and our microarray experiments reveal an increase of exon 11 usage upon depletion of the proteins that we confirmed in RT-PCR (Figure 2B). Finally, we investigated *PHACTR2* as a control for the specificity of the effect of the two depletions. *PHACTR2* mRNA encodes a phosphatase, and it contains one cassette exon (labelled b in Figure 2C, left) found to be differentially included upon depletion of CELF1, but not ELAVL1. Indeed, in RT-PCR, the long isoform containing exon b was increased at the expense of the short isoform when CELF1, but not when ELAVL1, was depleted (Figures 2C, S5). Hence, while the depletions of CELF1 and ELAVL1 have similar and additive effects on *ACSL4 a*nd *WNK1*, they have different effects on *PHACTR2*. The results obtained by RT-PCR on these three simple cases are thus fully consistent with the microarray data.

We next investigated *CD44* as an example of a more complex regulation (Figure 2D). It encodes a membrane receptor controlling cell motility. The involvement of *CD44* alternative splicing in cancer is well-known (see discussion) (Prochazka *et al*, 2014). The splicing indices of 4 consecutive exons (exons 12-15 in the longest isoform, labelled b-e in Figure 2D) are negative, revealing a repression of these 4 exons in CELF1, ELAVL1 and double depletions. Depending on

bioRxiv preprint doi: https://doi.org/10.1101/373704; this version posted July 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the combinatorial inclusion of each of these 4 exons, we detected 7 isoforms by RT-PCR using primers flanking these exons, whose identities were confirmed by sequencing (Figure 2D). Taken globally, the relative amounts of the 2 shortest isoforms ("af" or "abf") was increased by the depletion of CELF1 or ELAVL1, at the expense of the 2 longest isoforms ("abcdef" or "acdef"). The 3 other isoforms were not significantly modified. Hence, RT-PCR experiments confirm the microarray data showing that CELF1 and ELAVL1 stimulate the inclusion of the 4 alternative exons of *CD44* in a strikingly similar fashion.

Because we observed that 3' terminal exons are highly susceptible to CELF1 and ELAVL1 regulation, we also assessed the microarray results on those 3' terminal exons (Figure 3). *MICAL3* has a composite internal / 3' terminal exon. Both microarray results (Figure 3A, left) and RT-PCR (right) demonstrate that depleting CELF1 or ELAVL1 significantly favor its use as the terminal exon ("abc"). The last exon of *JDP2*, which encodes the transcription factor Jun dimerization protein 2, includes 2 alternative cleavage-polyadenylation sites, resulting in a short or long 3'UTR. Microarray and RT-PCR demonstrate that CELF1 and ELAVL1 favor the proximal cleavage, hence the short 3'UTR, as depleting CELF1 or ELAVL1 increases the amount of the long 3'UTR (Figure 3B). As a control, we also investigated the phosphatase-encoding *PPP3CB* mRNA. It contains a composite internal/3' terminal exon. In microarray, CELF1, but not ELAVL1, stimulates its use as an internal exon at the expense of the shorter isoform. The RT-PCR results confirm both the effect of CELF1 and the lack of effect of ELAVL1 (Figure 3C). Altogether, the RT-PCR experiments, in Figures 2 and 3, confirm the microarray results.

**CELF1 and ELAVL1 directly interact in cell nuclei**

The joint regulation of pre-mRNA processing exerted by CELF1 and ELAVL1 could be achieved by a macromolecular complex containing these two proteins. Indeed, they co-immuno-precipitate, hence physically interact (Le Tonquèze *et al*, 2010). We tested by FRET (Förster resonance energy transfer) if the interaction between CELF1 and ELAVL1 is direct. We used a

bioRxiv preprint doi: https://doi.org/10.1101/373704; this version posted July 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

FLIM (Fluorescence-lifetime imaging microscopy)-based FRET approach by measuring the fluorescence lifetime of the donor. When FRET occurs, the donor lifetime decreases, which reveals a close proximity between the donor (EGFP) and the acceptor (mCherry) (Padilla-Parra & Tramier, 2012). We expressed an EGFP-tagged version of CELF1 in HeLa cell. The nuclei of cells expressing EGFP-CELF1 were green, and the fluorescence lifetime of EGFP was around 2.5 ns (Figure 4). The nuclei of cells co-expressing a mCherry-tagged version of histone 2B (H2B) were yellow in the "merge" channel (Figure 4). This indicates a colocalisation of EGFP-CELF1 and mCherry-H2B, but the unchanged lifetime of EGFP is consistent with a lack of interaction between CELF1 and H2B. By contrast, EGFP lifetime was reduced with a co-expressed mCherry-tagged version of CELF1. This reveals a close proximity of EGFP and mCherry fluorophores, as expected based on the previously demonstrated capacity of CELF1 to oligomerize (Cosson *et al*, 2006). Importantly, mCherry-ELAVL1 also reduced the lifetime of EGFP-CELF1 fluorescence (Figure 4). Quantifying fluorescence lifetime in several nuclei demonstrated a highly significant decrease of EGFP fluorescence lifetime in the presence of mCherry-ELAVL1 (Figure 4, lower panel and S6, $p<2.2x10^{-16}$, Wilcoxon test). Similarly, the fluorescence lifetime of EGFP-ELAVL1 was reduced by mCherry-ELAVL1 as expected (Fialcowitz-White *et al*, 2007), but also by mCherry-CELF1 (Figure S7). Overall we conclude then that CELF1 and ELAVL1 are complexed with each other in the nucleus.

**The levels of *CELF1* and *ELAVL1* are correlated with prognosis and *CD44* splicing in human breast cancers**

We have shown above that CELF1 and ELAVL1 additively control mRNA splicing, and down regulating both proteins together has a much greater impact on the cellular transcriptome than either one alone. We thus used the TCGA (http://cancergenome.nih.gov/) breast invasive cancer panel to test the relevance of their additive function for human pathology. The levels of *CELF1* mRNA in tumors are more variable than in normal tissues but the mean is not significantly different

bioRxiv preprint doi: https://doi.org/10.1101/373704; this version posted July 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(Figure S8A). The survival rates of breast cancer patients with low or high *CELF1 m*RNA levels are also comparable (Figure 5A). Similarly, *ELAVL1 m*RNA level alone has no prognosis value (Figure 5B) although *ELAVL1* mRNA levels are higher in tumors than in normal tissues (Figure S8B). Importantly however, patients with low levels of both *CELF1* and *ELAVL1* have a better survival rate than patients with a high level of *CELF1* and / or *ELAVL1 (*Figure 5C). Hence, while *CELF1* and *ELAVL1* mRNA levels bring very limited information in breast cancers prognosis when taken individually, their combined levels may prove a valuable biomarker.

Among the mRNAs bound by CELF1 and ELAVL1 and whose splicing pattern is jointly controlled by both proteins, *CD44 i*s particularly relevant in human cancers. We show above in Figure 2D that CELF1 and ELAVL1 jointly stimulate the inclusion of 4 consecutive exons in *CD44* pre-mRNA. These exons are part of an array of ten cassette exons whose regulation plays key roles in metastasis and epithelial-mesenchymal transition (see discussion). Indeed, the ratio of inclusion of these 4 exons is higher in tumors than in normal tissues and a high rate of inclusion correlates with a bad prognosis (Figures S8C-D). Furthermore, patients with high *CELF1* and/or *ELAVL1* mRNA levels also have a significantly higher rate of inclusion of these exons (Figure 5D). Altogether these data demonstrate that, while CELF1 and ELAVL1 jointly stimulate the inclusion of 4 exons within *CD44* pre-mRNA, *CD44* exon inclusion may reduce survival of patients with high levels of *CELF1* and/or *ELAVL1*.

**A co-RNA map suggests principles of RNA splicing joint regulation by CELF1 and ELAVL1**

RNA splicing maps integrate binding and transcriptome data to link the position of the binding sites of an RBP with its regulatory functions (Witten & Ule, 2011). We produced "co-RNA maps" of CELF1 and ELAVL1 aimed at describing how they jointly regulate pre-mRNA maturation. Most RNA maps are centered on skipped exons, and RNA maps for cleavage sites were also published (Rot *et al*, 2017). As CELF1 and ELAVL1 regulate both internal and 3' terminal exons, we drew site-centric RNA maps. We focused on the 96 exonic regions that are controlled by

bioRxiv preprint doi: https://doi.org/10.1101/373704; this version posted July 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the two proteins in the same direction and belong to an mRNA bound by both proteins (see Figure 1D). We investigated how CELF1 and ELAVL1 bind around the 3' splice sites and 5' splice sites or cleavage-polyadenylation sites that border these regions (Figure 6). For 3' splice sites, the binding of CELF1 and ELAVL1 in a region overlapping the site generally represses it. CELF1 binding in the upstream intron is associated with 3' splice site activation rather than repression (Figure 6A). Repressed 5' splice sites often have CELF1 bound in the exon and ELAVL1 downstream in the intron. Activated 5' sites are essentially characterized by CELF1 binding in the intron (Figure 6B). Finally, CELF1 binds upstream of repressed cleavage sites and downstream of activated cleavage sites (Figure 6C). Overall these maps suggest general principles for CELF1 and ELAVL1-mediated control of alternative pre-mRNA processing (Figure 6D): (i) CELF1 binds downstream of or overlaps repressed 3' splice sites, and binds upstream of overlaps repressed 5' splice sites and cleavage sites. These correspond to the regions to be retained in the mature mRNA. Conversely, CELF1 binding in introns or downstream of a cleavage site activates the nearby sites. (ii) We did not detect preferential ELAVL1 binding around activated sites, but ELAVL1 assists CELF1 in repressing 3' or 5' splice sites by also binding close to the regulated sites.

## DISCUSSION

Previous studies demonstrated an antagonism between ELAVL1 and CELF1 in translational control. Namely, ELAVL1 stimulates the translation of *CDH1* (E-cadherin), *OCLN* (Occludin) and *MYC* in intestinal cells whereas CELF1 represses them (Yu *et al*, 2016; Liu *et al*, 2015; Yu *et al*, 2013). Here, we investigate the functional relationships between CELF1 and ELAVL1 at a global scale, for the first time to our knowledge. By contrast with translational control, we find that these two proteins strongly cooperate to dictate mRNA splicing patterns.

How is this cooperation between ELAVL1 and CELF1 achieved? Our FRET measurements show they are in direct contact. We can speculate that CELF1 and ELAVL1 are embedded in a large

bioRxiv preprint doi: https://doi.org/10.1101/373704; this version posted July 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

RBP complex, similar to the recently described LASR (large assembly of splicing regulators) complex (Damianov *et al*, 2016). The efficiency of that complex would be reduced by the depletion of either protein and more strongly reduced by the depletion of both proteins, which would account for the additive effects of depleting CELF1 and ELAVL1. If CELF1 and ELAVL1 are indeed part of a larger complex the next question becomes the identity of their partners. Both CELF1 and ELAVL1 form homo-oligomers, suggesting that the complex includes more than one molecule of each of these two RBPs (Cosson *et al*, 2006; Fialcowitz-White *et al*, 2007). CELF1 and ELAVL1 do both have RNA-dependent interactions with other proteins; CELF1 interacts with HNRNPH1 and H2 (Paul *et al*, 2006), while ELAVL1 interacts with HNRNPK (Hegele *et al*, 2012), HNRNPL (Matsui *et al*, 2008) and IGF2BP1 (Weidensdorfer *et al*, 2009). While CELF1 and ELAVL1 have multiple partners therefore it nevertheless remains to be demonstrated that they exist together in larger multiprotein complexes.

How then can this hypothetical large CELF1 and ELAVL1-containing complex control splicing? The co-RNA splicing maps shown in Figure 6 indicate that CELF1 and ELAVL1 binding close to or overlapping the splice sites generally inhibits their usage. This is in agreement with previous findings that CELF1 represses bound cassette alternative exons (Wang *et al*, 2015). Mechanistically, CELF1 and ELAVL1-containing complexes may inhibit nearby sites by sterically blocking spliceosome recruitment. It was reported that CELF1 binding to upstream and downstream intronic regions facilitates exon skipping and inclusion, respectively (Masuda *et al*, 2012), and we do not confirm this observation with RNAs interacting with both CELF1 and ELAVL1. Rather, we find that CELF1 binding to intronic regions at some distance from the regulated site tends to stimulate its usage. How the CELF1 and ELAVL1-containing complexes activate distant sites remains to be determined.

The finding that CELF1 and ELAVL1 cooperatively control alternative splicing has implications in human cancers. Several reports point to a deleterious role of CELF1 and ELAVL1 in cancer. In glioma and non-small cell lung cancers, the level of CELF1 protein is negatively

bioRxiv preprint doi: https://doi.org/10.1101/373704; this version posted July 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

355 correlated with patient survival (Xia *et al*, 2015; Jiao *et al*, 2013). It is also the case for ELAVL1 in ovarian high-grade serous carcinoma, colorectal cancer and hepatocellular carcinoma (Davidson *et al*, 2016; Yoo *et al*, 2009; Zhu *et al*, 2015). The major cause of death in cancers is metastasis, in which epithelial to mesenchymal transition (EMT) plays a key role. This cellular process involves loss of epithelial properties, including cell-cell and cell-matrix adhesion, and gain of mesenchymal 360 properties. CELF1 enhances the translation of transcription factors that are central in EMT in human breast cancer (Chaudhury *et al*, 2016), and ELAVL1 stabilizes the mRNA encoding the EMT driver SNAIL (Dong *et al*, 2007). We find here that CELF1 and ELAVL1 stimulate the inclusion of 4 consecutive exons of *CD44*, which are part of 10 cassette exons favoring EMT and metastasis in several human cancers (Prochazka *et al*, 2014). Therefore, there are multiple functions 365 of ELAVL1 and CELF1 in EMT. Importantly, although *CELF1* or *ELAVL1* levels have no significant impact on survival rate when considered independently in breast cancer, a low level of *CELF1* plus *ELAVL1* mRNA is correlated with a better prognosis. The cooperation between CELF1 and ELAVL1 in RNA splicing control suggests that simultaneously taking into account the abundance of these two proteins may considerably improve the stratification of a wide range of 370 tumors.

## MATERIALS AND METHODS

### Sequences and constructs

375 The sequences of the siRNAs and those of the primers are given in supplemental text. The cDNA encoding ELAVL1 or CELF1 were introduced in the pENTR directional TOPO vector (Invitrogen) and then inserted in the final vectors 223 pCS EGFP DEST or 362 pCS Cherry DEST (Addgene) by Gateway recombination (Invitrogen) to construct plasmids EGFP-ELAVL1, mCherry-ELAVL1, EGFP-CELF1 and mCherry-CELF1.

380

bioRxiv preprint doi: https://doi.org/10.1101/373704; this version posted July 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

**Cell manipulations and RNAi**

We cultured HeLa Kyoto cells at 37 °C, 5% $CO_2$ in DMEM (Gibco) with 10% FCS (PAA), 100 U / ml penicillin and 100 µg / ml streptomycin (Gibco). For RNAi, we transfected the cells (Jetprime, Polyplus) with the following siRNAs: 5 nM control siRNA plus 5 nM siRNA against ELAVL1, 5 nM control siRNA plus 5 nM siRNA against CELF1, 5 nM siRNA against ELAVL1 plus 5 nM siRNA against CELF1, 10 nM control siRNA, or 10 nM siRNA against luciferase. After 48 h, we recovered the cells for sampling in 500 µl of PBS by scraping after three washes with cold PBS and centrifuging them for 5 minutes at 3000 rpm at 4 ° C.

**FRET-FLIM**

HeLa cells Kyoto were seeded at 20% confluence in observation chambers on 4-well glass slides (Dustcher) and transfected (Jetprime, Polyplus) with 15 ng of expression vectors of EGFP-labeled proteins in the presence or absence of 150 ng of vector expressing mCherry-labeled proteins. After 24 h, the cells were washed with PBS and observed at 37°C.

FRET experiments were conducted with a fastFLIM system, using a plan APO 63x oil immersion objective with a 1.4 NA (Leica). For EGFP donor excitation, a narrow spectral band was selected (480/23 nm) from a white-light laser (Fianium) and sent to the microscope through a dichroic mirror (Di01-T 405/488/568/647, Semrock) inside the spinning head (CSUX1, Yokogawa). The emission of EGFP was then filtered using an emission filter on the spinning filter wheel (525/50 nm) and acquired with an intensified time-gated CCD camera (Picostar, LaVision and CoolSNAPHQ2, Roper). Five temporal gates of 2.25 ns were acquired sequentially by adjusting, step-by-step, the delay of the laser signal to trig the gated intensifier. The exposure time of the CCD camera was chosen between 500 ms and 2 s, depending on the brightness of the sample. The FLIM calculation to determine the mean fluorescence lifetime in a pixel-by-pixel basis was carried out online using flimager (Leray *et al*, 2013). For each region in the nucleus where lifetimes were measured, we checked the absence of photo-bleaching and the presence of both fluorophores for

bioRxiv preprint doi: https://doi.org/10.1101/373704; this version posted July 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

doubly transfected cells. Three independent series of transfection were carried out, and for each series 10 cells were analyzed. Thus, a total of 30 nuclei were analyzed per condition.

**Protein analysis**

Cell pellets (equivalent to one 10-cm Petri dish at 80-90% confluence) were lysed with RIPA buffer (50 mM Tris HCl pH7.4, 150 mM NaCl, 1% sodium deoxycholate, 1% Triton X-100; 0.1% SDS; P8340 Sigma protease inhibitor). After a 10-minute incubation at room temperature, the samples were sonicated with a micro-probe at 4 ° C, and centrifuged at 10,000 rpm for 5 minutes at 4 ° C. Proteins were separated by electrophoresis and electro-transferred (Hybond C plus, GE-Healthcare) following standard western blot procedures with anti-CELF1 (SantaCruz, 3B1, sc-20003), anti-PCNA (Sigma-Aldrich, p8825) or anti ELAV1 (SantaCruz, sc-5261) primary antibodies, and anti-mouse secondary antibodies for imaging with the Odyssey (Li-cor) imager.

**RNA analysis**

Total RNA was extracted with TRIzol (Invitrogen), treated with Turbo DNase (Ambion) and then sent to the Plateforme Génomique (Nantes) for hybridization on Sureprint G3 human Exon microarrays (Agilent). Raw signals were lowess-normalized before analysis. Alternatively, they were used as matrices to synthesize cDNA using random primers and Super Script II RT (Invitrogen). To assess alternative splicing, RT-PCR was performed with either unlabeled or $^{32}$P-labeled primers and analysed by Sybr staining or autoradiography after electrophoresis. The number of PCR cycles was empirically determined for each pair of primers to remain within the exponential phase of amplification.

**Data analysis**

The RNA maps were drawn as described (Haberman *et al*, 2017). We retrieved the TCGA

bioRxiv preprint doi: https://doi.org/10.1101/373704; this version posted July 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

data from Firebrowse (http://firebrowse.org/) and TCGA splice seq (Ryan *et al*, 2016). We analysed all the data with in-house R scripts, which are available as supplemental text.

435

**ACKNOWLEDGEMENTS**

We thank Laure Le Calvez for developing R scripts to draw RNA maps. The results shown here are in part based upon data generated by the TCGA Research Network http://cancergenome.nih.gov/. Microarray hybridizations were carried out on the Plateforme génomique de Nantes.

This research was funded by a grant from the Ligue Régionale contre le Cancer to LP. GD was supported by a joint doctoral fellowship from the Ligue Régionale contre le Cancer and the Région Bretagne (ARED).

445

**AUTHOR CONTRIBUTIONS**

GD performed and analysed the microarray experiments. GD, SPP and MT designed, performed and analysed the FRET experiments. SD performed the depletions and RT-PCR. YA conceived and designed experiments and analyzed data. LP conceived and designed experiments, analyzed data, and wrote the manuscript.

450

**CONFLICT OF INTEREST**

None.

455

bioRxiv preprint doi: https://doi.org/10.1101/373704; this version posted July 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## REFERENCES

Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, Slobodeniuc V, Kutter C, Watt S, Colak R, Kim T, Misquitta-Ali CM, Wilson MD, Kim PM, Odom DT, Frey BJ & Blencowe BJ (2012) The evolutionary landscape of alternative splicing in vertebrate species. *Science* **338:** 1587–1593

Boulanger G, Cibois M, Viet J, Fostier A, Deschamps S, Pastezeur S, Massart C, Gschloessl B, Gautier-Courteille C & Paillard L (2015) Hypogonadism Associated with Cyp19a1 (Aromatase) Posttranscriptional Upregulation in Celf1 Knockout Mice. *Mol. Cell. Biol.* **35:** 3244–3253

Chang S-H, Elemento O, Zhang J, Zhuang ZW, Simons M & Hla T (2014) ELAVL1 regulates alternative splicing of eIF4E transporter to promote postnatal angiogenesis. *Proc. Natl. Acad. Sci. U.S.A.* **111:** 18309–18314

Chaudhury A, Cheema S, Fachini JM, Kongchan N, Lu G, Simon LM, Wang T, Mao S, Rosen DG, Ittmann MM, Hilsenbeck SG, Shaw CA & Neilson JR (2016) CELF1 is a central node in post-transcriptional regulatory programmes underlying EMT. *Nat Commun* **7:** 13362

Chen C-YA, Xu N & Shyu A-B (2002) Highly selective actions of HuR in antagonizing AU-rich element-mediated mRNA destabilization. *Mol. Cell. Biol.* **22:** 7268–7278

Chi MN, Auriol J, Jégou B, Kontoyiannis DL, Turner JMA, de Rooij DG & Morello D (2011) The RNA-binding protein ELAVL1/HuR is essential for mouse spermatogenesis, acting both at meiotic and postmeiotic stages. *Mol. Biol. Cell* **22:** 2875–2885

Cibois M, Boulanger G, Audic Y, Paillard L & Gautier-Courteille C (2012) Inactivation of the Celf1 gene that encodes an RNA-binding protein delays the first wave of spermatogenesis in mice. *PLoS ONE* **7:** e46337

Cibois M, Gautier-Courteille C, Vallée A & Paillard L (2010) A strategy to analyze the phenotypic consequences of inhibiting the association of an RNA-binding protein with a specific RNA. *RNA* **16:** 10–15

Cosson B, Gautier-Courteille C, Maniey D, Aït-Ahmed O, Lesimple M, Osborne HB & Paillard L (2006) Oligomerization of EDEN-BP is required for specific mRNA deadenylation and binding. *Biol. Cell* **98:** 653–665

Damianov A, Ying Y, Lin C-H, Lee J-A, Tran D, Vashisht AA, Bahrami-Samani E, Xing Y, Martin KC, Wohlschlegel JA & Black DL (2016) Rbfox Proteins Regulate Splicing as Part of a Large Multiprotein Complex LASR. *Cell* **165:** 606–619

Dassi E (2017) Handshakes and Fights: The Regulatory Interplay of RNA-Binding Proteins. *Front Mol Biosci* **4:** 67

Davidson B, Holth A, Hellesylt E, Hadar R, Katz B, Tropé CG & Reich R (2016) HUR mRNA expression in ovarian high-grade serous carcinoma effusions is associated with poor survival. *Hum. Pathol.* **48:** 95–101

Dong R, Lu J-G, Wang Q, He X-L, Chu Y-K & Ma Q-J (2007) Stabilization of Snail by HuR in the process of hydrogen peroxide induced cell migration. *Biochem. Biophys. Res. Commun.* **356:** 318–321

bioRxiv preprint doi: https://doi.org/10.1101/373704; this version posted July 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Drees BL, Thorsson V, Carter GW, Rives AW, Raymond MZ, Avila-Campillo I, Shannon P & Galitski T (2005) Derivation of genetic interaction networks from quantitative phenotype data. *Genome Biol.* **6:** R38

Fialcowitz-White EJ, Brewer BY, Ballin JD, Willis CD, Toth EA & Wilson GM (2007) Specific protein domains mediate cooperative assembly of HuR oligomers on AU-rich mRNA-destabilizing sequences. *J. Biol. Chem.* **282:** 20948–20959

Fu X-D (2004) Towards a splicing code. *Cell* **119:** 736–738

García-Domínguez DJ, Morello D, Cisneros E, Kontoyiannis DL & Frade JM (2011) Stabilization of Dll1 mRNA by Elavl1/HuR in neuroepithelial cells undergoing mitosis. *Mol. Biol. Cell* **22:** 1227–1239

Haberman N, Huppertz I, Attig J, König J, Wang Z, Hauer C, Hentze MW, Kulozik AE, Le Hir H, Curk T, Sibley CR, Zarnack K & Ule J (2017) Insights into the design and interpretation of iCLIP experiments. *Genome Biol.* **18:** 7

Hall MP, Nagel RJ, Fagg WS, Shiue L, Cline MS, Perriman RJ, Donohue JP & Ares M (2013) Quaking and PTB control overlapping splicing regulatory networks during muscle cell differentiation. *RNA* **19:** 627–638

Hegele A, Kamburov A, Grossmann A, Sourlis C, Wowro S, Weimann M, Will CL, Pena V, Lührmann R & Stelzl U (2012) Dynamic protein-protein interaction wiring of the human spliceosome. *Mol. Cell* **45:** 567–580

Jiao W, Zhao J, Wang M, Wang Y, Luo Y, Zhao Y, Tang D & Shen Y (2013) CUG-binding protein 1 (CUGBP1) expression and prognosis of non-small cell lung cancer. *Clin Transl Oncol* **15:** 789–795

Kalsotra A, Xiao X, Ward AJ, Castle JC, Johnson JM, Burge CB & Cooper TA (2008) A postnatal switch of CELF and MBNL proteins reprograms alternative splicing in the developing heart. *Proc. Natl. Acad. Sci. U.S.A.* **105:** 20333–20338

Katsanou V, Milatos S, Yiakouvaki A, Sgantzis N, Kotsoni A, Alexiou M, Harokopos V, Aidinis V, Hemberger M & Kontoyiannis DL (2009) The RNA-binding protein Elavl1/HuR is essential for placental branching morphogenesis and embryonic development. *Mol. Cell. Biol.* **29:** 2762–2776

Katz Y, Wang ET, Airoldi EM & Burge CB (2010) Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods* **7:** 1009–1015

Küch E-M, Vellaramkalayil R, Zhang I, Lehnen D, Brügger B, Sreemmel W, Ehehalt R, Poppelreuther M & Füllekrug J (2014) Differentially localized acyl-CoA synthetase 4 isoenzymes mediate the metabolic channeling of fatty acids towards phosphatidylinositol. *Biochim. Biophys. Acta* **1841:** 227–239

Lafarga V, Cuadrado A, Lopez de Silanes I, Bengoechea R, Fernandez-Capetillo O & Nebreda AR (2009) p38 Mitogen-activated protein kinase- and HuR-dependent stabilization of p21(Cip1) mRNA mediates the G(1)/S checkpoint. *Mol. Cell. Biol.* **29:** 4341–4351

Le Tonquèze O, Gschloessl B, Legagneux V, Paillard L & Audic Y (2016) Identification of CELF1 RNA targets by CLIP-seq in human HeLa cells. *Genom Data* **8:** 97–103

bioRxiv preprint doi: https://doi.org/10.1101/373704; this version posted July 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Le Tonquèze O, Gschloessl B, Namanda-Vanderbeken A, Legagneux V, Paillard L & Audic Y (2010) Chromosome wide analysis of CUGBP1 binding sites identifies the tetraspanin CD9 mRNA as a target for CUGBP1-mediated down-regulation. *Biochem. Biophys. Res. Commun.* **394:** 884–889

Leray A, Padilla-Parra S, Roul J, Héliot L & Tramier M (2013) Spatio-Temporal Quantification of FRET in living cells by fast time-domain FLIM: a comparative study of non-fitting methods [corrected]. *PLoS ONE* **8:** e69335

Liu L, Ouyang M, Rao JN, Zou T, Xiao L, Chung HK, Wu J, Donahue JM, Gorospe M & Wang J-Y (2015) Competition between RNA-binding proteins CELF1 and HuR modulates MYC translation and intestinal epithelium renewal. *Mol. Biol. Cell* **26:** 1797–1810

Masuda A, Andersen HS, Doktor TK, Okamoto T, Ito M, Andresen BS & Ohno K (2012) CUGBP1 and MBNL1 preferentially bind to 3' UTRs and facilitate mRNA decay. *Sci Rep* **2:** 209

Matsui K, Nishizawa M, Ozaki T, Kimura T, Hashimoto I, Yamada M, Kaibori M, Kamiyama Y, Ito S & Okumura T (2008) Natural antisense transcript stabilizes inducible nitric oxide synthase messenger RNA in rat hepatocytes. *Hepatology* **47:** 686–697

Merkin J, Russell C, Chen P & Burge CB (2012) Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science* **338:** 1593–1599

Mukherjee N, Corcoran DL, Nusbaum JD, Reid DW, Georgiev S, Hafner M, Ascano M, Tuschl T, Ohler U & Keene JD (2011) Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability. *Mol. Cell* **43:** 327–339

Padilla-Parra S & Tramier M (2012) FRET microscopy in the living cell: different approaches, strengths and weaknesses. *Bioessays* **34:** 369–376

Pan Q, Shai O, Lee LJ, Frey BJ & Blencowe BJ (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **40:** 1413–1415

Park JW, Jung S, Rouchka EC, Tseng Y-T & Xing Y (2016) rMAPS: RNA map analysis and plotting server for alternative exon regulation. *Nucleic Acids Res.* **44:** W333-338

Paul S, Dansithong W, Kim D, Rossi J, Webster NJG, Comai L & Reddy S (2006) Interaction of muscleblind, CUG-BP1 and hnRNP H proteins in DM1-associated aberrant IR splicing. *EMBO J.* **25:** 4271–4283

Peng SS, Chen CY, Xu N & Shyu AB (1998) RNA stabilization by the AU-rich element binding protein, HuR, an ELAV protein. *EMBO J.* **17:** 3461–3470

Philips AV, Timchenko LT & Cooper TA (1998) Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. *Science* **280:** 737–741

Prochazka L, Tesarik R & Turanek J (2014) Regulation of alternative splicing of CD44 in cancer. *Cell. Signal.* **26:** 2234–2239

Raghavachari N, Barb J, Yang Y, Liu P, Woodhouse K, Levy D, O'Donnell CJ, Munson PJ & Kato GJ (2012) A systematic comparison and evaluation of high density exon arrays and RNA-seq technology used to unravel the peripheral blood transcriptome of sickle cell disease. *BMC Med Genomics* **5:** 28

bioRxiv preprint doi: https://doi.org/10.1101/373704; this version posted July 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Rahman MA, Masuda A, Ohe K, Ito M, Hutchinson DO, Mayeda A, Engel AG & Ohno K (2013) HnRNP L and hnRNP LL antagonistically modulate PTB-mediated splicing suppression of CHRNA1 pre-mRNA. *Sci Rep* **3:** 2931

Rot G, Wang Z, Huppertz I, Modic M, Lenče T, Hallegger M, Haberman N, Curk T, von Mering C & Ule J (2017) High-Resolution RNA Maps Suggest Common Principles of Splicing and Polyadenylation Regulation by TDP-43. *Cell Rep* **19:** 1056–1067

Roy A, Al-Qusairi L, Donnelly BF, Ronzaud C, Marciszyn AL, Gong F, Chang YPC, Butterworth MB, Pastor-Soler NM, Hallows KR, Staub O & Subramanya AR (2015) Alternatively spliced proline-rich cassettes link WNK1 to aldosterone action. *J. Clin. Invest.* **125:** 3433–3448

Ryan M, Wong WC, Brown R, Akbani R, Su X, Broom B, Melott J & Weinstein J (2016) TCGASpliceSeq a compendium of alternative mRNA splicing in cancer. *Nucleic Acids Res.* **44:** D1018-1022

Supek F, Bošnjak M, Škunca N & Šmuc T (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **6:** e21800

Tang Y, Wang H, Wei B, Guo Y, Gu L, Yang Z, Zhang Q, Wu Y, Yuan Q, Zhao G & Ji G (2015) CUG-BP1 regulates RyR1 ASI alternative splicing in skeletal muscle atrophy. *Sci Rep* **5:** 16083

Ule J, Stefani G, Mele A, Ruggiu M, Wang X, Taneri B, Gaasterland T, Blencowe BJ & Darnell RB (2006) An RNA map predicting Nova-dependent splicing regulation. *Nature* **444:** 580–586

Uren PJ, Burns SC, Ruan J, Singh KK, Smith AD & Penalva LOF (2011) Genomic analyses of the RNA-binding protein Hu antigen R (HuR) identify a complex network of target genes and novel characteristics of its binding sites. *J. Biol. Chem.* **286:** 37063–37066

Vidal-Petiot E, Cheval L, Faugeroux J, Malard T, Doucet A, Jeunemaitre X & Hadchouel J (2012) A new methodology for quantification of alternatively spliced exons reveals a highly tissue-specific expression pattern of WNK1 isoforms. *PLoS ONE* **7:** e37751

Vlasova IA, Tahoe NM, Fan D, Larsson O, Rattenbacher B, Sternjohn JR, Vasdewani J, Karypis G, Reilly CS, Bitterman PB & Bohjanen PR (2008) Conserved GU-rich elements mediate mRNA decay by binding to CUG-binding protein 1. *Mol. Cell* **29:** 263–270

Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP & Burge CB (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature* **456:** 470–476

Wang ET, Ward AJ, Cherone JM, Giudice J, Wang TT, Treacy DJ, Lambert NJ, Freese P, Saxena T, Cooper TA & Burge CB (2015) Antagonistic regulation of mRNA expression and splicing by CELF and MBNL proteins. *Genome Res.* **25:** 858–871

Wang Z & Burge CB (2008) Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* **14:** 802–813

Weidensdorfer D, Stöhr N, Baude A, Lederer M, Köhn M, Schierhorn A, Buchmeier S, Wahle E & Hüttelmaier S (2009) Control of c-myc mRNA stability by IGF2BP1-associated cytoplasmic RNPs. *RNA* **15:** 104–115

bioRxiv preprint doi: https://doi.org/10.1101/373704; this version posted July 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Witten JT & Ule J (2011) Understanding splicing regulation through RNA splicing maps. *Trends Genet.* **27:** 89–97

Xia L, Sun C, Li Q, Feng F, Qiao E, Jiang L, Wu B & Ge M (2015) CELF1 is Up-Regulated in Glioma and Promotes Glioma Cell Proliferation by Suppression of CDKN1B. *Int. J. Biol. Sci.* **11:** 1314–1324

Yoo PS, Sullivan CAW, Kiang S, Gao W, Uchio EM, Chung GG & Cha CH (2009) Tissue microarray analysis of 560 patients with colorectal adenocarcinoma: high expression of HuR predicts poor survival. *Ann. Surg. Oncol.* **16:** 200–207

Yu T-X, Gu B-L, Yan J-K, Zhu J, Yan W-H, Chen J, Qian L-X & Cai W (2016) CUGBP1 and HuR regulate E-cadherin translation by altering recruitment of E-cadherin mRNA to processing bodies and modulate epithelial barrier function. *Am. J. Physiol., Cell Physiol.* **310:** C54-65

Yu T-X, Rao JN, Zou T, Liu L, Xiao L, Ouyang M, Cao S, Gorospe M & Wang J-Y (2013) Competitive binding of CUGBP1 and HuR to occludin mRNA controls its translation and modulates epithelial barrier function. *Mol. Biol. Cell* **24:** 85–99

Zhang C & Darnell RB (2011) Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat. Biotechnol.* **29:** 607–614

Zhang W, Yu Y, Hertwig F, Thierry-Mieg J, Zhang W, Thierry-Mieg D, Wang J, Furlanello C, Devanarayan V, Cheng J, Deng Y, Hero B, Hong H, Jia M, Li L, Lin SM, Nikolsky Y, Oberthuer A, Qing T, Su Z, et al (2015) Comparison of RNA-seq and microarray-based models for clinical endpoint prediction. *Genome Biol.* **16:** 133

Zhu H, Berkova Z, Mathur R, Sehgal L, Khashab T, Tao R-H, Ao X, Feng L, Sabichi AL, Blechacz B, Rashid A & Samaniego F (2015) HuR Suppresses Fas Expression and Correlates with Patient Outcome in Liver Cancer. *Mol. Cancer Res.* **13:** 809–818

bioRxiv preprint doi: https://doi.org/10.1101/373704; this version posted July 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## FIGURE LEGENDS

### Figure 1. Alternative splicing in CELF1 and ELAVL1-depleted cells

**A**, Higher panel, enrichment *p*-values (-$\log_{10}$ scale) of hypergeometric tests carried out to appraise the enrichment of each MISO RNA processing event within the set of 2,868 differentially used exonic regions in CELF1-depleted cells. Lower panel, same as higher panel for the set of 3,130 differentially used exonic regions in ELAVL1-depleted cells. A3SS, alternative 3' splice site. A5SS, alternative 5' splice site. AFE, alternative first exon. ALE, alternative last exon. MXE, mutually exclusive exons. RI, retained intron. SE, skipped exon. Tandem, tandem UTR (alternative cleavage-polyadenylation sites). **B**, Enrichment *p*-values (-log10 scale) of the 15 most enriched GO terms in CELF1 and ELAVL1 targets. **C**, For the 117,841 retained probes in the exon array, splicing index of ELAVL1-depleted cells compared with controls against splicing index of CELF1-depleted cells compared with controls. The numbers of probes in each quadrant (with splicing indices above 0.2 in absolute value) are shown in red. **D**, same as **C**, restricted to the 99 probes in CELF1 and ELAVL1 ligands and targets. The genes analyzed by RT-PCR in the next figures are in red. **E**, For the 99 probes, splicing index of cells deprived of both CELF1 and ELAVL1 compared with control against sum of the splicing indices of CELF1- and ELAVL1-depleted cells. The numbers of probes in each quadrant are shown in red.

### Figure 2. Examples of alternative 3' splice sites and cassette exons controlled by CELF1 and ELAVL1

Four genomic regions of interest are depicted in the upper left part of each panel. Constitutive exons targeted by the RT-PCR primers are in black, and alternative exonic regions are colored. The splicing indices in cells treated by siRNAs against CELF1 (C), ELAVL1 (E) or both (C+E) were calculated from array experiments and are indicated with the same color code as the exonic regions. Previously published CLIPseq data are shown below the gene charts (CELF1 (Le Tonquèze *et al*,

bioRxiv preprint doi: https://doi.org/10.1101/373704; this version posted July 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2016), ELAVL1 (Uren *et al*, 2011)). The right panels show a representative RT-PCR experiment from HeLa cells left non-transfected (NT) or transfected with the indicated siRNA (C and E, siRNA against CELF1 and ELAVL1, respectively; Ctrl and luc, 2 control siRNAs, Ctrl being the same as in the microarray experiments and luc against luciferase). The quantification is the mean ± sd of at least 3 similar independent experiments. In **A**-**C**, FC is the percentage of the indicated amplimeres normalized by the same percentage in non-transfected conditions. In **D** the percentages of amplimeres 1+2, and 6+7, are given. **A**, *ACSL4* (alternative 3' splice sites). **B**, *WNK1* (2 consecutive cassette exons, one of which is regulated by CELF1 and ELAVL1). **C**, *PHACTR2* (one CELF1-controlled cassette exon). **D**, *CD44* (4 consecutive cassette exons, all regulated by CELF1 and ELAVL1).

**Figure 3. Examples of alternative 3' RNA processing controlled by CELF1 and ELAVL1**

Genomic regions of interest with splicing indices from our microarray experiment, published CLIPseq data (CELF1 (Le Tonquèze *et al*, 2016), ELAVL1 (Uren *et al*, 2011)), and a representative RT-PCR experiment with the quantification of at least 3 similar independent runs as in Figure 2. **A**, *MICAL3* pre-mRNA includes a composite exon used either as a terminal or internal exon. The maturation pattern was revealed by RT-PCR with one forward primer (in exon "a") and two competing reverse primers (in exonic regions "c" and "d"). **B**, The last exon of *JDP2* pre-mRNA contains two alternative cleavage-polyadenylation sites. The maturation pattern was revealed by separate RT-PCRs with one forward primer (in exon "a") and either a reverse primer in exonic region "b" for amplifications of both isoforms or a reverse primer in exonic region "c" for amplification of the longest isoform only. The PCR products were mixed before electrophoresis. **C**, *PPP3CB* pre-mRNA includes a CELF1-controlled composite internal-terminal exon. The maturation pattern was revealed by RT-PCR as in **A**.

bioRxiv preprint doi: https://doi.org/10.1101/373704; this version posted July 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

**Figure 4. CELF1 and ELAVL1 directly interact in the nucleus**

Representative images of HeLa cells co-transfected by a plasmid driving the expression of EGFP-tagged CELF1 and a plasmid driving the expression of mCherry-tagged histone 2B (H2B), CELF1 or ELAVL1. The merge (EGFP plus mCherry) and EGFP lifetime (short lifetimes reveal a close proximity of mCherry) channels are shown. Beneath: boxplot showing the distribution of EGFP fluorescence lifetime in all measured pixels from 30 nuclei.

**Figure 5. *CELF1* plus *ELAVL1* is a prognostic marker in breast cancers**

**A**, Kaplan-Meier curve of 1,079 breast invasive cancer patients (TCGA BRCA panel) according to low and high *CELF1 m*RNA level. p, log-rank test *p*-value. **B**, Same as **A** for *ELAVL1* mRNA level. **C**, Kaplan-Meier curve of the 1,079 patients according to low *CELF1* and *ELAVL1 m*RNA levels (n=749, intersect of n=833 low *CELF1* and n=958 low *ELAVL1*) or high *CELF1 o*r *ELAVL1 m*RNA levels (n=330, other patients with high *CELF1* and/or *ELAVL1* levels). **D**, Cumulative distribution function for the percentage of inclusion of the 4 consecutive cassette exons of *CD44* shown in Figure 2D in patients with low *CELF1* and *ELAVL1* mRNA levels, and with high *CELF1* or *ELAVL1* mRNA levels. The inset is a boxplot displaying the same data. p, Wilcoxon test *p*-value.

**Figure 6. co-RNA splice map of CELF1 and ELAVL1.**

**A-C**, RNA splice maps for the spatial distribution of binding of CELF1 (blue) and ELAVL1 (bronze) around alternative sites. The solid lines represent protein binding around repressed sites (positive splicing index upon protein depletion) and the dashed lines around activated sites. **A**, 3' splice sites, inferred from skipped exons, alternative 3' sites and alternative last exons. For alternative 3’ splice sites, we classified the downstream site as activated and the upstream site as repressed when the splicing index of the region between the sites was positive, and vice-versa. **B**, 5' splice sites, inferred from skipped exons and alternative 5' sites. For alternative 5’ splice sites, we classified the upstream site as activated and the downstream site as repressed when the splicing

bioRxiv preprint doi: https://doi.org/10.1101/373704; this version posted July 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

index of the region between the sites was positive. **C**, Cleavage-polyadenylation sites, inferred from alternative last exons and tandem cleavage-polyadenylation sites. For tandem sites, we classified the proximal site as activated and the distal site as repressed when the splicing index of the region between the sites was positive. **D**, Model for CELF1 and ELAVL1-mediated regulation of alternative splicing. The relative positions for CELF1 and ELAVL1 binding within the introns are not to scale.

bioRxiv preprint doi: https://doi.org/10.1101/373704; this version posted July 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



## SUPPLEMENTAL TABLES LEGEND

550 **Table S1**: A large file (41.5 Mo) of 117,841 rows and 20 columns. The 117,841 rows are the 117,841 exon array probes retained after initial filtration. The 20 columns are:

- GeneName
- ProbeName
- The normalized exon values for cells treated with 4 different combinations of siRNA

555 (control+control, control+CELF1, control+ELAVL1, CELF1+ELAVL1), each in triplicate

- The splicing indices and the *p*-values for cells treated with control+CELF1, control+ELAVL1, and CELF1+ELAVL1 (each compared with control+control).

**Table S2**: GO term enrichment: 571 lines and 31 columns. The 571 lines are the GO terms with an

560 enrichment *p*-value below 0.01 in at least one of five sets of genes among: the targets of CELF1, the targets of ELAVL1, the ligands of CELF1, the ligands of ELAVL1, and the genes that produce a transcript differentially spliced upon depletion of both proteins. The 31 columns are:

- The name of the GO term
- The reference list, the number of genes in this GO term found and expected in the targets of

565 CELF1, a binary stating if the number of retrieved genes is above ("+") or below ("-") the expected number, the enrichment fold and the *p*-value for enrichment.

- The same information for the 4 other datasets

**Table S3:** 2951 genes (GeneNames) that produce transcripts that interact with CELF1 in HeLa cells

570 in CLIP-seq experiments (ligands of CELF1) (Le Tonquèze *et al*, 2016).

bioRxiv preprint doi: https://doi.org/10.1101/373704; this version posted July 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

**Table S4:** 4284 genes (GeneNames) that produce transcripts that interact with ELAVL1 in HeLa cells in CLIP-seq experiments (ligands of ELAVL1) (Uren *et al*, 2011)**.**

bioRxiv preprint doi: https://doi.org/10.1101/373704; this version posted July 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## SUPPLEMENTAL FIGURES LEGEND

575

**Figure S1. Quality controls of microarray experiments**

**A**, Representative western blot showing the abundance of CELF1, ELAVL1 and PCNA (loading control) in HeLa cells transfected with the indicated siRNA. The numbers beneath the gels indicate the amounts of proteins after normalisation by PCNA and by the untransfected condition. **B,** 580 Hierarchical clustering dendrogram of exon array probe values. We hybridized RNA extracted from cells transfected with the indicated combinations of siRNAs (C, anti-CELF1; E, anti-ELAVL1; Ctrl, control) to exon array. After normalisation and filtration for low signal, we retained 117,841 probes corresponding to exonic regions. Arabic numbers 1-3 correspond to independent siRNA transfections. That identical combinations of siRNAs cluster together demonstrates the 585 reproducibility of the results.

**Figure S2. Distribution of differential RNA processing events with respect to the annotation developed in the MISO framework**

**A**, Distribution of the 2,868 differential events upon CELF1 depletion. **B**, Distribution of the 3,130 590 differential events upon ELAVL1 depletion. A3SS, alternative 3' splice site. A5SS, alternative 5' splice site. AFE, alternative first exon. ALE, alternative last exon. MXE, mutually exclusive exons. RI, retained intron. SE, skipped exon. Tandem, tandem UTR (alternative cleavage-polyadenylation sites).

595 **Figure S3. REVIGO treemap summarizing gene ontology biological process categories over-represented in the targets of CELF1 and ELAVL1**

We identified the Gene Ontology Biological Processes over-represented among the targets of CELF1 (**A**) or ELAVL1 (**B**). The over-represented categories with *p*-values below 0.01 (149 terms in **A**, 150 terms in **B**) were used to generate a treemap colored by functional category. The size of

bioRxiv preprint doi: https://doi.org/10.1101/373704; this version posted July 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

each rectangle is proportional to the absolute value of $\log_{10}$(*p*-value) for that category. **C**, Venn diagram showing the overlap of superclusters of related terms between the targets of CELF1 and ELAVL1.

**Figure S4. Differential RNA processing events upon CELF1 and ELAVL1 double depletion**

**A**, Distribution of the 10,483 exonic regions identified as differential upon CELF1 and ELAVL1 double depletion within the MISO annotation classes. **B**, Enrichment *p*-values (-$\log_{10}$ transformed) of hypergeometric tests carried out to appraise the enrichment of each MISO RNA processing event within the set of 10,483 differential exonic regions in cells depleted of both CELF1 and ELAVL1. A3SS, alternative 3' splice site. A5SS, alternative 5' splice site. AFE, alternative first exon. ALE, alternative last exon. MXE, mutually exclusive exons. RI, retained intron. SE, skipped exon. Tandem, tandem UTR (alternative cleavage-polyadenylation sites). **C**, Venn diagram showing the number of differential exonic regions upon depletion of CELF1, ELAVL1, or both.

**Figure S5. Analysis of *PHACTR2* splicing pattern**

Same representative RT-PCR experiment to analyze *PHACTR2* splicing pattern as shown in Figure 2C, but a reverse primer targeting the cassette exon (orange “b”) was included, giving a third amplimere labeled 3. The stronger signal of amplimere 3 upon CELF1 depletion (lanes 3 and 5) confirms that CELF1 inhibits the inclusion of exon b.

**Figure S6. Cumulative distribution fraction plot showing the quantification of EGFP-CELF1 fluorescence lifetime.**

All pixels from 30 nuclei resulting from 3 independent transfections were quantified (same data as Figure 5). The cumulative distribution (*y* value) is the fraction of EGFP fluorescence lifetimes that are less than or equal to the fluorescence lifetime (*x* value). A reduced EGFP fluorescence lifetime reveals that FRET occurs with mCherry, hence that the two fluorophores are in close proximity.

bioRxiv preprint doi: https://doi.org/10.1101/373704; this version posted July 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

**Figure S7. EGFP-ELAVL1 fluorescence lifetime.**

We co-transfected HeLa cells with a plasmid driving the expression of EGFP-tagged ELAVL1 and a plasmid driving the expression of mCherry-tagged histone 2B (H2B), CELF1 or ELAVL1. The quantifications of all measured pixels from 30 nuclei (3 independent transfections) are shown as a boxplot (**A**) and a cumulative distribution function (**B**). A reduced EGFP fluorescence lifetime reveals that FRET occurs with mCherry, hence that the two fluorophores are in close proximity.

**Figure S8. Expression levels of *CELF1* and *ELAVL1* and *CD44* splicing in breast invasive cancers**

**A**, Left panel, cumulative distribution function for the normalized expression level of *CELF1 i*n 1079 breast invasive cancers and 98 normal tissues (TCGA BRCA panel). The cumulative distribution (*y* value) is the fraction of patients with *CELF1* levels that are less than or equal to the *CELF1* level (*x* value). Right panel, boxplot displaying the same data. p, Wilcoxon test *p*-value. **B**, Same as **A** for *ELAVL1* mRNA level. **C**, Same as **A** for the percentage of inclusion of the 4 consecutive cassette exons of *CD44* shown in Figure 2D. **D**, Kaplan-Meier curve of the 1,079 patients according to low or high inclusion of the same 4 consecutive cassette exons of *CD44*. p, log-rank test *p*-value.

bioRxiv preprint doi: https://doi.org/10.1101/373704; this version posted July 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## SUPPLEMENTAL TEXT

### Primer sequences

*ACSL4* CCCTCCGATTGAAATCACAG, GGGATGTCTATTACAGCTAGTGAG

*WNK1* CTCCTCAACAGACAGTGCAG, ACAGAGGAAGCCAAAGTGGT

*PHACTR2* GAAAATTCAAACGGGCACAT, CCGGTGTTTCAGAGTGGTTT, CTTTGAAGCTTTGGGACGAG

*CD44* CAGAAGGAACAGTGGTTTGG, GGGTGGAATGTGTCTTGGTC

*MICAL3* TCTACTGTAAGCCACACTACTG, CAGCCAAACAAGGAAGTGAG, GTAGTTCTCCAGCTCGATCC

JDP2 GGGAGAAGAACAAAGTCGCA, TGCTTCAGCTCCTCAATCTG, GCAGTTTCATTCTGAGTTCCA and CTCATCTGTGAAATAGGAGAGCTG

PPP3CB AAATTCGAGCAATTGGCAAG, AGACACTAGCTAATACAGTACCCTGTC, TGTGAGAGTCCCTGGGAAGT

### siRNA sequences

*CELF1*: a pool of GAGCCAACCUGUUCAUCUA, GCUCUUUAUUGGUAUGAUU, GCUGCAUUAGAAGCUCAGA

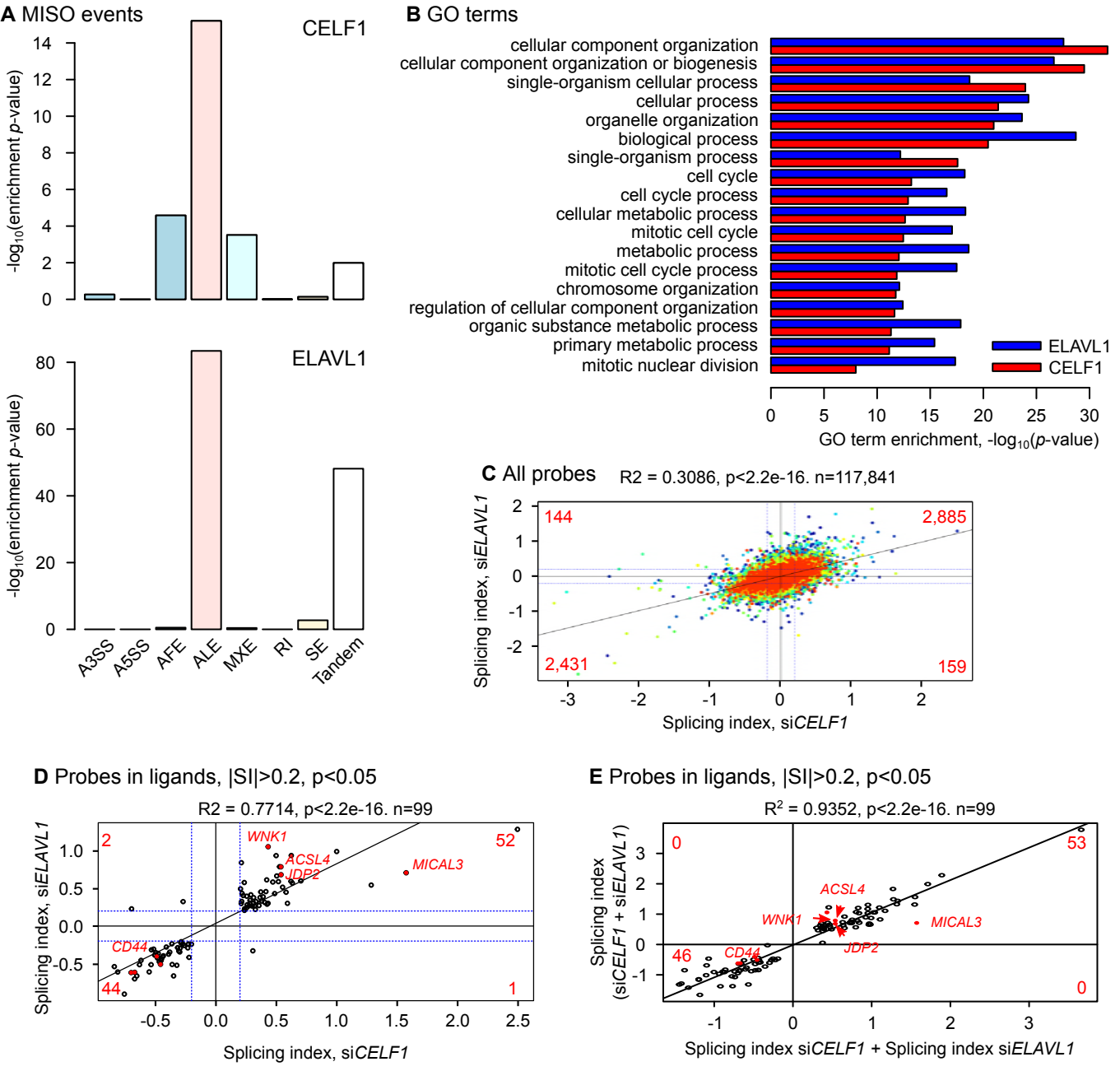
*ELAVL1*: a pool of GAGGCAAUUACCAGUUUCA, UCUUAAGUUUCGUAAGUUA

Ctrl: GUCUAGACGAGUGUGACAU

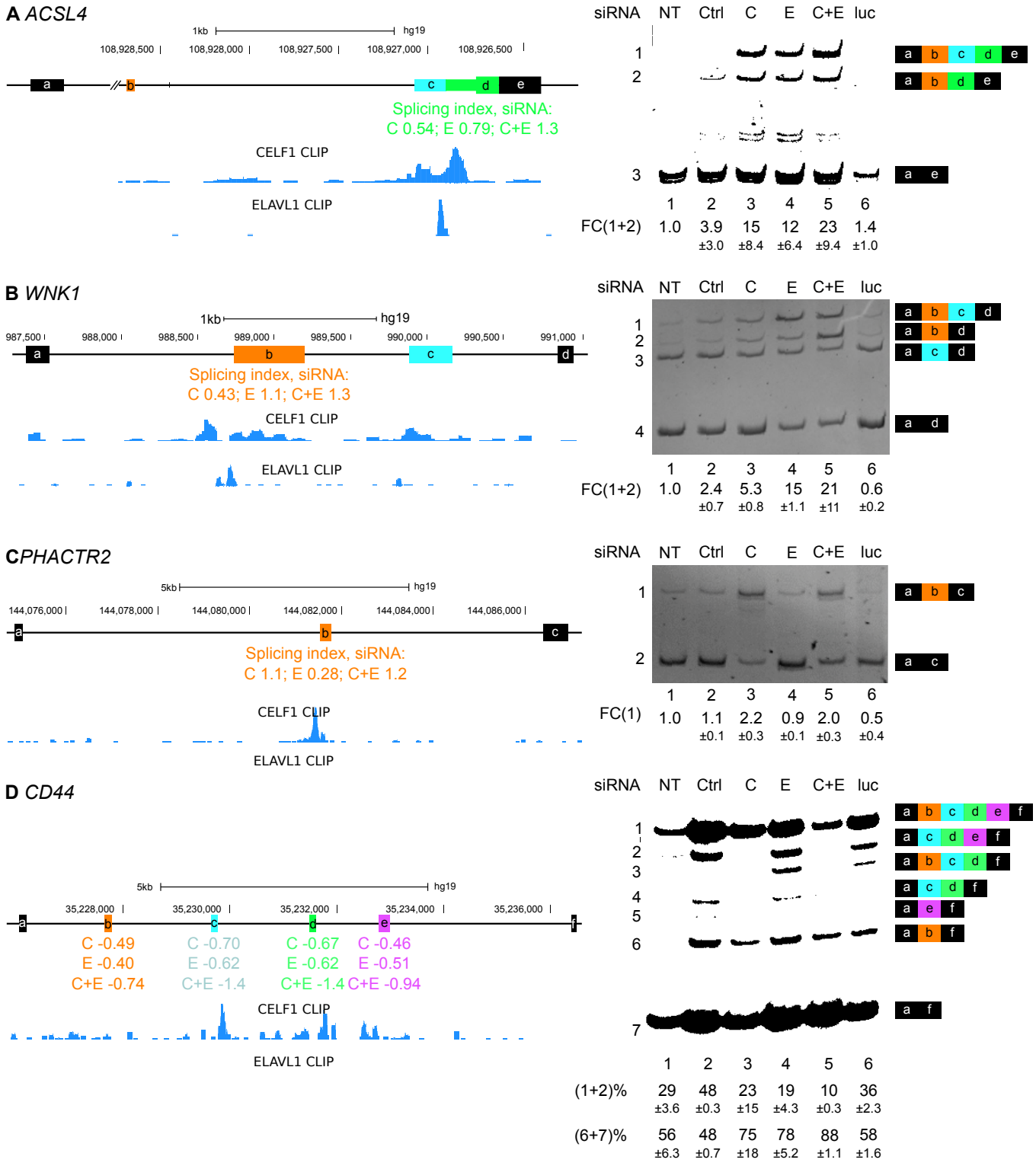
Luc: CAUUCUAUCCUCUAGAGGAUG

### Annotated R script

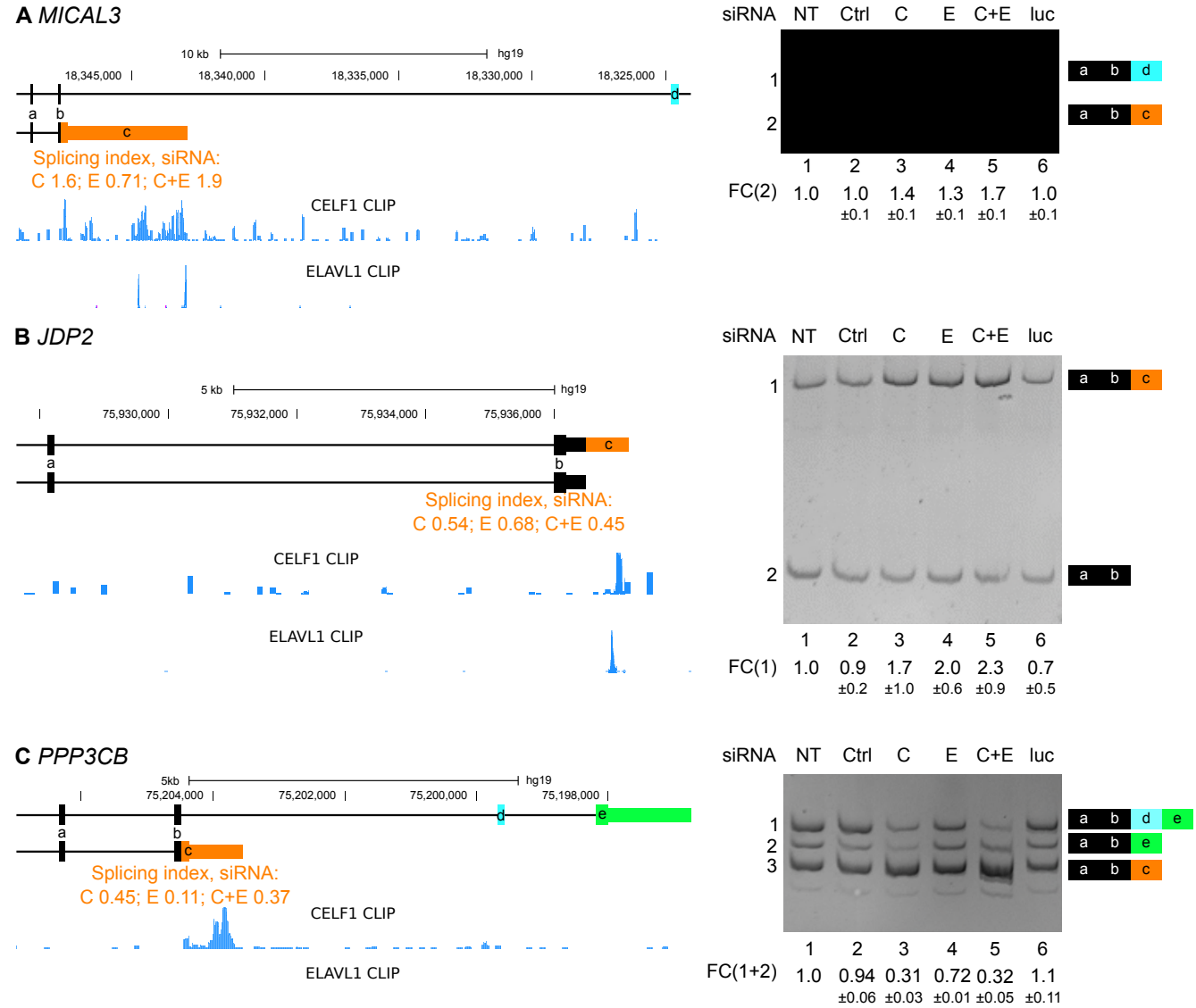
Supplemental file


A MISO events
CELF1
ELAVL1
-log10(enrichment p-value)
A3SS
A5SS
AFE
ALE
MXE
RI
SE
Tandem
B GO terms
cellular component organization
cellular component organization or biogenesis
single-organism cellular process
cellular process
organelle organization
biological process
single-organism process
cell cycle
cell cycle process
cellular metabolic process
mitotic cell cycle
metabolic process
mitotic cell cycle process
chromosome organization
regulation of cellular component organization
organic substance metabolic process
primary metabolic process
mitotic nuclear division
ELAVL1
CELF1
GO term enrichment, -log10(p-value)
C All probes
R2 = 0.3086, p<2.2e-16. n=117,841
144
2,885
2,431
159
Splicing index, siELAVL1
Splicing index, siCELF1
D Probes in ligands, |SI|>0.2, p<0.05
R2 = 0.7714, p<2.2e-16. n=99
WNK1
ACSL4
JDP2
MICAL3
CD44
2
52
44
1
Splicing index, siELAVL1
Splicing index, siCELF1
E Probes in ligands, |SI|>0.2, p<0.05
R2 = 0.9352, p<2.2e-16. n=99
0
53
46
0
Splicing index (siCELF1 + siELAVL1)
Splicing index siCELF1 + Splicing index siELAVL1

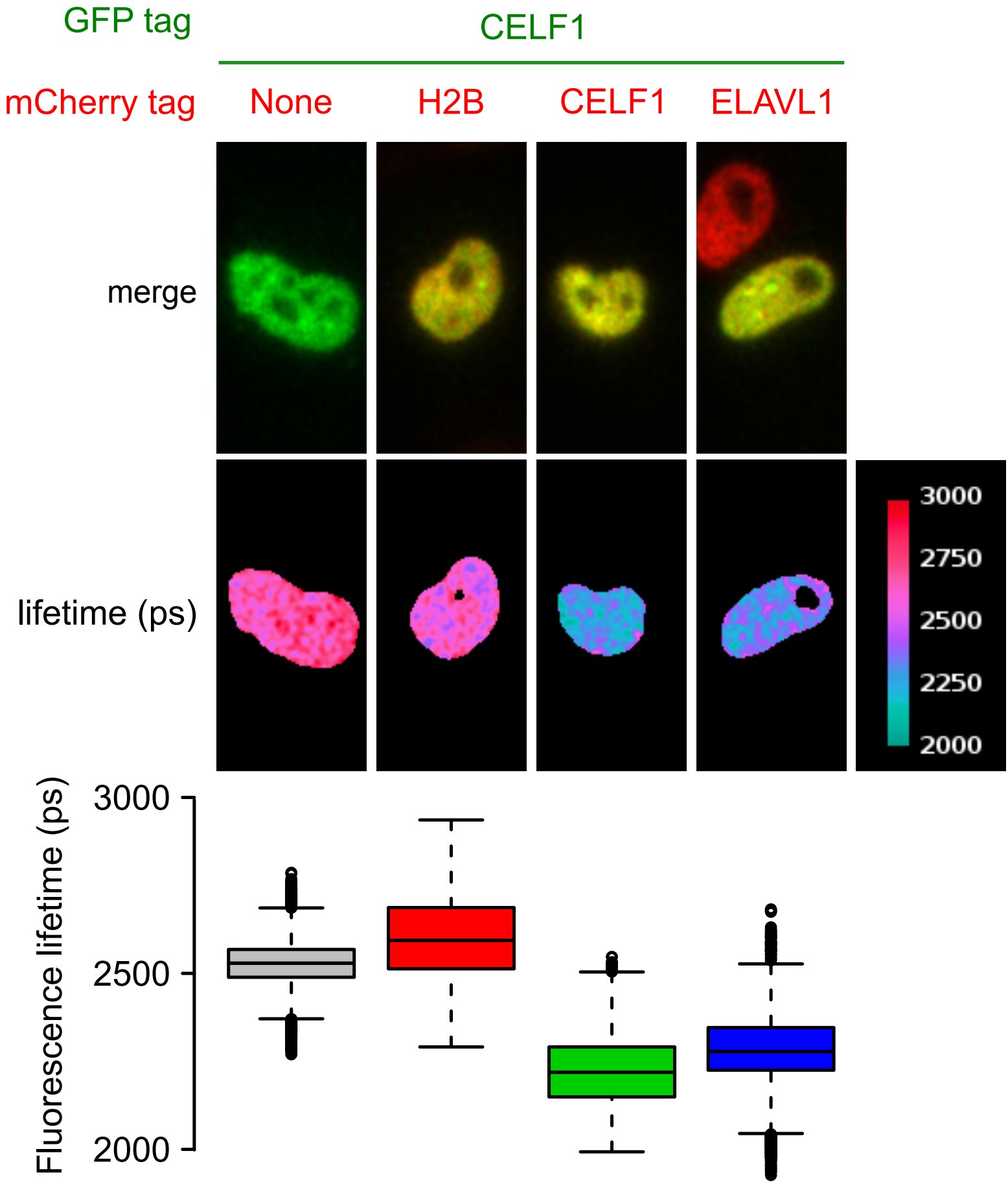
David et al., Figure 2



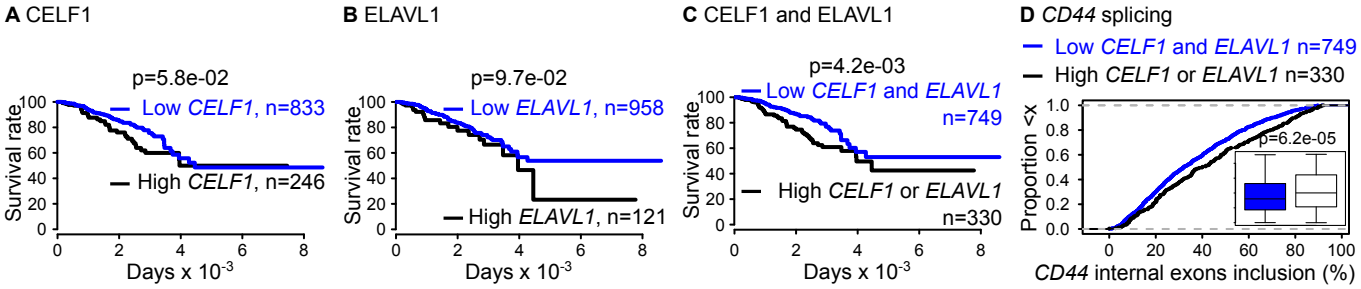
David et al., Figure 3

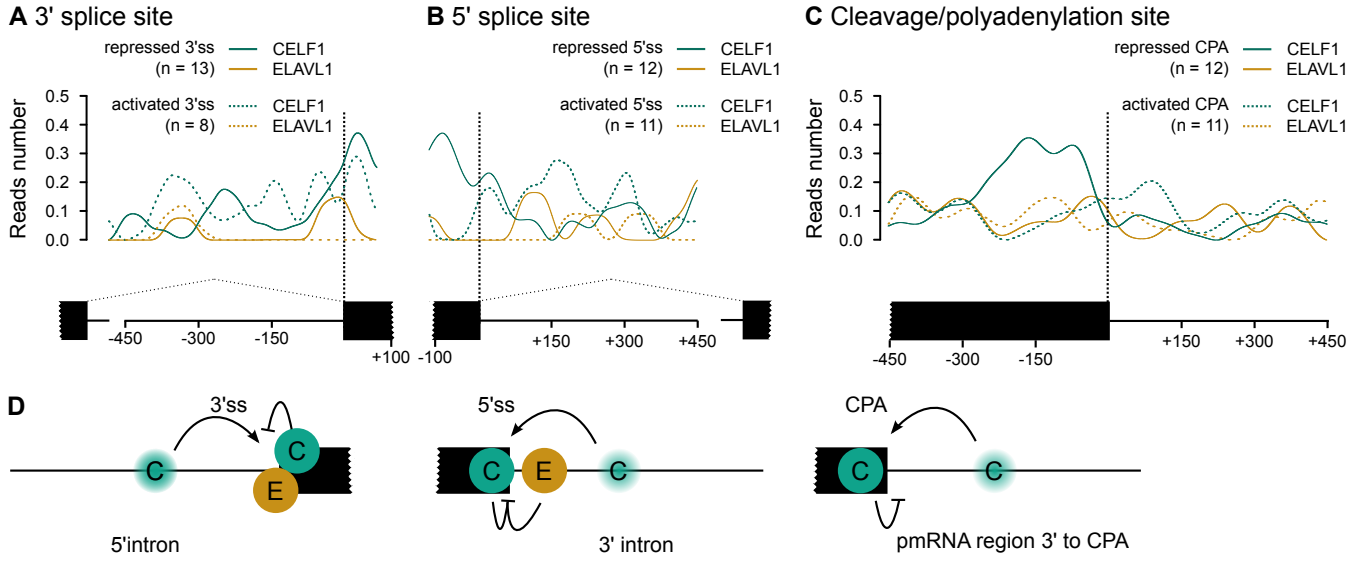


# David et al., Figure 4



David et al., Figure 5




A 3' splice site
repressed 3'ss (n = 13)
activated 3'ss (n = 8)
CELF1
ELAVL1
Reads number
0.5
0.4
0.3
0.2
0.1
0.0
-450
-300
-150
+100
B 5' splice site
repressed 5'ss (n = 12)
activated 5'ss (n = 11)
-100
+150
+300
+450
C Cleavage/polyadenylation site
repressed CPA (n = 12)
activated CPA (n = 11)
D
3'ss
5'ss
CPA
C
E
5'intron
3' intron
pmRNA region 3' to CPA