Alternative polyadenylation expands the mRNA isoform repertoire of human *CD46*

Phuong Thao Ly[1,2], Sze Jing Tang[1,3], Xavier Roca[1,*]

Affiliations

1 School of Biological Sciences, Nanyang Technological University, 637551 Singapore

2 The Neuroscience and Behavioral Disorders Programme, Duke-NUS Graduate Medical School, Singapore

3 Cancer Science Institute of Singapore, National University of Singapore, Singapore

* Corresponding author: **Xavier Roca:** School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, 637551 Singapore; E-mail: xroca@ntu.edu.sg; Tel: +65 65927561; FAX: +65 67913856.

Keywords

Alternative polyadenylation; intronic polyadenylation; CD46; soluble isoforms; pseudogene

Abbreviations

5'ss: 5' Splice Site
3'RACE: 3' Rapid Amplification of cDNA Ends
3'UTR: 3' UnTranslated Region
APA: Alternative PolyAdenylation
ASO: AntiSense Oligonucleotide
CCP: Complement Control Protein
CD46: Cluster of Differentiation gene 46
CD46P1: CD46 Pseudogene 1
CR1L: Complement Receptor 1-Like
CYT1 / CYT2: CYtoplasmic Tail 1 / 2
FBS: Fetal Bovine Serum
IPA: Intronic PolyAdenylation
MCP: Membrane Cofactor Protein
NSD: Non-Stop mediated mRNA Decay
PAS: PolyAdenylation Site
RCA: Regulators of Complement Activation
RTK: Receptor Tyrosine Kinase
RT-PCR: Reverse Transcription - Polymerase Chain Reaction
snRNP: small nuclear RiboNucleoProtein
SP: Signal Peptide
STP: Serine Threonine Proline rich
TM: TransMembrane
U: Unknown

Highlights

1. Intronic PolyAdenylation potentially generates soluble or tail-less CD46
2. Alternative polyadenylation in *CD46* also shortens 3' UTRs of *CD46* mRNAs
3. *CD46P1* pseudogene is transcribed but terminated by polyadenylation in intron 2
4. Alternative PolyAdenylation may increase the functional repertoire of CD46

Abstract

Alternative polyadenylation is a prevalent mechanism regulating mammalian gene expression. While tandem 3'-Untranslated-Region (3'UTR) polyadenylation changes expression levels, intronic polyadenylation generates shorter transcripts encoding truncated proteins. Intronic polyadenylation regulates 20% of genes and is especially common in receptor tyrosine-kinase transcripts, generating soluble repressors. Here we report that human *CD46*, encoding a transmembrane repressor of complement and T-cell co-stimulator, expresses multiple isoforms by alternative polyadenylation. We provide evidence for polyadenylation at several introns by RT-PCR of 5' intronic fragments, and by increase in such isoforms via functional U1 knockdown. We mapped various intronic polyadenylation sites by 3' Rapid Amplification of cDNA Ends (3'RACE), which could generate soluble or membrane-bound but tail-less CD46. Intronic polyadenylation could add to the source of soluble CD46 isoforms in fluids and tissues, which increase in cancers and autoimmune syndromes. Furthermore, 3'RACE identified three polyadenylation sites within the last intron and exon, whose transcripts with shortened 3'UTRs could support higher CD46 expression. Finally, 3'RACE revealed that the *CD46* pseudogene only expresses short transcripts by early polyadenylation in intron 2. Overall, we report a wide variety of *CD46* mRNA isoforms which could generate new protein isoforms, adding to the diverse physiological and pathological roles of CD46.

(198 words)

1. Introduction

Alternative polyadenylation (APA) regulates expression of ~70% of human protein-coding genes [1, 2]. Hence, APA together with alternative splicing largely contribute to transcriptome and proteome diversity. The four basic APA types depend on the location of the alternative polyadenylation sites (PASs): (i) tandem PASs within the last exon, generating 3' Untranslated Regions (3'UTRs) of different lengths which negatively correlate with mRNA stability and expression; (ii) alternative terminal exons with their own PASs; (iii) PASs at exonic sequences, and (iv) Intronic Polyadenylation (IPA) by usage of an intronic PAS, thereby extending the immediately upstream exon and making it the terminal one, and also potentially changing the protein's C-terminal domain. IPA affects about ~20% of all genes [3], and is common in mRNAs encoding receptor tyrosine kinases (RTKs), in which shorter open reading frames lacking the exons encoding the transmembrane domain generate soluble receptors with repressive activity [4]. IPA is inhibited by U1 small nuclear ribonucleoprotein (snRNP) binding to the nearest upstream 5' splice site (5'ss), an essential step during splicing [5], thus revealing a competition between splicing and IPA. Consistently, IPA tends to happen in long introns with weak 5'ss [3], it is prevalent upon functional U1 snRNP knockdown by antisense oligonucleotides (ASOs) [6, 7], it is enhanced upon UV damage [8], and it can be artificially induced by other ASOs blocking U1 binding to a specific 5'ss [4], with potential therapeutic applications [9]. As the detailed information on IPA in genes outside RTKs is very limited [3, 10], here we report this phenomenon for *CD46*, which encodes an important type I transmembrane protein without enzymatic activity.

CD46 (or Membrane Cofactor Protein, MCP) is expressed in all nucleated human cells, and plays roles in both innate and adaptive immunity [11, 12], as well as in epithelial and sperm cells [13-15]. CD46 halts the complement attack by acting as a cofactor for Factor I mediated inactivation of C3b and C4b bound to the same host cell [16]. CD46 also functions as a costimulatory molecule in immune cells such as T lymphocytes [17-19]. CD46 deficiency causes a rare autoimmune disease termed atypical Hemolytic Uremic Syndrome [11, 12], and is associated with the much more common age-related macular degeneration [12, 20]. Furthermore, cancer cells overexpress CD46 for protection from the immune system [21], and its expression is altered in autoinflammatory syndromes [22, 23]. Finally, CD46 serves as an entry receptor for certain bacteria and viruses [11], further emphasizing the multipronged connections between CD46 and human disease.

CD46 belongs to the Regulators of Complement Activation (RCA) gene family cluster in chromosome 1q3.2 [12]. From its N-terminus, mature CD46 comprises four Complement Control Protein (CCP) regions which bind C3b/C4b, one to three Serine, Threonine, Proline-rich regions (STP A, B, and C), a region of unknown function (U), a transmembrane anchor (TM) and two alternative cytoplasmic tails (CYT1 or CYT2) (Figure 1A) [24]. The *CD46* gene has fourteen exons encoding the signal peptide (SP) in exon 1, the four CCPs in exons 2 to 6, the A-C STPs respectively encoded in exons 7 to 9, the U region in exon 10, the TM domain in exons 11 and 12, and the two cytoplasmic tails in exons 13 or 14. *CD46* undergoes extensive alternative splicing

affecting the STP region and cytoplasmic tail [11, 12, 25, 26]. For STP, exon 7 is mostly skipped, exon 8 is either skipped or included, and exon 9 is mostly included in mature transcripts, and the resulting isoforms differ in the preferential complement pathways and binding of pathogens. For the cytoplasmic tails, cassette exon 13 is either included or skipped to respectively generate either CYT1 encoded in exon 13, or CYT2 in exon 14. Overall, the four most common CD46 isoforms are BC-CYT1, BC-CYT2, C-CYT1 and C-CYT2, yet there are additional splice isoforms with uncertain function. The two cytoplasmic tails differ in length and sequence, and bind different kinases with different functional consequences, best described in activation of T helper 1 cells [27, 28]. Upon ligand binding, the extracellular and intracellular domains of CD46-CYT1 are proteolytically ejected to induce gamma interferon production and subsequent T regulatory 1 phenotype by Interleukin 10 secretion, while later CD46-CYT2 cleavage restores homeostasis. The extracellular domain shedding is catalyzed by Matrix Metalloproteinases, and in apoptotic or necrotic cells soluble CD46 promotes complement activation and inflammatory clearance [29]. Furthermore, the elevated soluble CD46 from cancer cells retains its role as cofactor for C3b cleavage [30], and its levels are also increased in autoimmune disorders [31, 32]. Soluble CD46 was early on detected in body fluids, especially from the reproductive tract, but not cell-line supernatants [33-35]. The heterogeneity of soluble CD46, with multiple isoforms with different molecular weights, might be generated by not only proteolysis but also other mechanisms, such as the proposed intron retention [36], yet here we show that IPA might likely contribute to it.

We recently elucidated regulatory mechanisms of *CD46* alternative splicing [26]. We established that the 5'ss of STP cassette exons 7 and 8 are recognized by noncanonical base-pairing to U1 snRNA (the RNA moiety of U1 snRNP) through asymmetric loop registers [26, 37, 38]. We found numerous *cis*-acting elements regulating exon 13 inclusion vs skipping, and several *trans*-acting factors like SRSF1, PTBP1 and TIA1/L1 [26]. In addition, we reported that exon 13 inclusion is influenced by the speed of RNA polymerase II transcription and by nonsense mediated mRNA decay. Here we expand the knowledge of the posttranscriptional regulation of *CD46*, by describing multiple APA isoforms at mRNA level, especially via IPA, with potential to alter the various physiological and/or pathological CD46 functions.

2. Materials and methods

2.1. Cell culture and transfection

We cultured HEK293T and HeLa cell lines in DMEM/Hi-Glucose supplemented with 10% Fetal Bovine Serum (FBS) (Hyclone, United States) and 1% penicillin-streptomycin (Gibco, United States). We maintained Jurkat E6.1 and K562 cells in RPMI-1640 (Hyclone, Thermo Scientific) supplemented with 10% FBS (Hyclone, United States), 1% penicillin-streptomycin (Gibco, United States), and 500 mM β-mercaptoethanol. We incubated all cells at 37^{o}C with 5% CO_2.

24 hours prior to transfection, we seeded HEK 293T cells onto 6-well plates at 10^5 cells/ml. We transfected cells with 833ng of Mock or U1-decoy (D1) or mutated U1-decoy (D7) plasmid [4, 39] using X-tremeGENE 9 DNA transfection reagents (Roche, Switzerland) according to the manufacturer's protocol. Additionally, we included 167ng of Mock plasmids in each experimental sample to make up to the amount of DNA required (1µg) for transfection. We then cultured the transfected cells for another 48 hours before RNA extraction.

2.2. RNA extraction and DNase I treatment

We isolated total RNA from cells using RNeasy kit (Qiagen, Germany) as per manufacturer's recommendation. Then we treated RNA with RQ1 RNase-Free DNase (Promega, USA) to degrade contaminant genomic DNA.

2.3. Reverse Transcription – Polymerase Chain Reaction and gel electrophoresis

We used these RNAs for cDNA synthesis via Reverse Transcription (RT) with M-MuLV Reverse Transcriptase (New England BioLabs, USA) and Oligo-dT primer as before [26]. We carried out all PCRs using GoTag Polymerase (Promega, USA), for 35 cycles with 56-60^{o}C annealing temperature (~4^{o}C lower than Tm of primers) and 1 min elongation (72^{o}C) per kb of amplimer. We separated and visualized the PCR products in 1.8% (w/v) agarose gels. We gel-purified bands using QIAquick gel extraction kit (Qiagen Sciences Inc., Germany) and sequenced them (1st BASE holding, Singapore) to determine the identity of the bands. We obtained all primers from Integrated DNA Technology (IDT), with their sequences available upon request.

To detect intronic retention in mRNAs, we performed PCR with forward (F) primers located in exons upstream of the exon preceding each examined intron, to ensure that the amplified product derived from processed mRNAs and not from unspliced RNA or residual genomic DNA. We placed the reverse intronic primers upstream the potential PASs or within the first 100 nt of examined introns.

For differential 5'/3' intronic expression experiment, for each intron that is suspected to undergo IPA, we performed PCR with reverse primers annealing to the region either upstream or downstream the suspected or confirmed PASs, used together with a corresponding forward primer.

For U1 functional knockdown experiments with decoys [4, 39], we carried out three-primer PCR as follows: we paired a common forward (F) with two reverse (R) primers, one for a downstream exon for full-length mRNA, and one for the examined intron, at a 1:3 ratio to enhance amplification of the IPA isoform. We quantified band intensities on agarose gels by area under curve from the lane plot, under gel analysis tool in Image J software. We performed these RT-PCRs with samples from three independent transfections (triplicates) with consistent results.

To detect *CD46 Pseudogene 1 (CD46P1)* transcripts, we performed PCR with specific primers with mismatches to *CD46*, and under three different annealing temperatures

(56°C, 58°C and 60°C), to establish the amplification specificity in comparison to that of homologous genes (*CD46* and *CR1L*).

2.4. 3' Rapid Amplification of cDNA Ends

We used DNase-treated total RNA from K562 cells as template for cDNA synthesis with SuperScriptTM II reverse transcriptase (Invitrogen) and an oligo-dT adapter primer (5'-GCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTT-3') for 3' Rapid Amplification of cDNA Ends (3'RACE), according to the manufacturer's instructions. For each PAS, we carried out 3'RACE using the first and second (nested) forward primers that annealed to the exon upstream each poly(A) tail, together with reverse Abridge Universal Adapter primer (AUAP) 5'- GGCCACGCGTCGACTAGTAC -3'. We performed the nested PCR with moderately stringent annealing temperature (2°C lower than Tm) to limit spurious amplifications. Primer sequences and detailed amplification conditions are available upon request.

3. Results

We aimed to identify new APA mRNA isoforms in *CD46*, focusing on IPA and tandem 3'UTR APA. Expression of IPA-generated transcripts would be supported by the following evidence (Figure 1B): First, detection of 5' intronic fragments in mRNAs due to the usage of intronic PASs. Second, the ratio of these intron-retained vs full-length transcripts should increase upon functional U1 knockdown [4]. Third, IPA-derived mRNAs must be identified by 3'RACE. We also used 3'RACE to identify tandem 3'UTR APA in *CD46* 3'-terminal exon 14. As poly(A) tails are deposited 15-30 nucleotides downstream the poly(A) signal (consensus AAUAAA), the new PASs would only be considered as such if located at least 30 nt away from the well-known *CD46* PAS [1]. Concordant results generated by the three experiments would be considered as strong evidence for IPA.

3.1 Retention of 5' intronic regions in *CD46* mRNAs suggests IPA

We performed RT-PCR using a reverse primer at the 5' portion of several *CD46* introns, and a forward primer located two or more exons upstream (Figure 1B; Figure 2A). This design enhances the amplification of processed mRNAs with respect to unspliced RNA or genomic DNA which are much larger, but these mRNAs would retain an intronic fragment potentially as a result of IPA. We conducted a systematic screening on almost all *CD46* introns, only excluding introns 1 and 7 because exon 1 encodes the signal peptide that is removed from mature protein [24], and exon 7 is included at very low levels in all tissues [26]. To enhance detection of cell-type specific IPA events, we used the cDNA templates from HEK293T, HeLa, K562 and Jurkat cell lines. Our RT-PCR results showed retention of the 5' portion of introns 2, 4, 6 and 10 in all mRNAs, of intron 3 in only K562 and Jurkat, and of intron 12 in all except HEK293T (Figure 2B). These results suggest that these introns undergo IPA via at least one PAS.

Both splicing and 3'-end formation are largely co-transcriptional processes, so it is possible that the detected RNA species with intronic retention were derived from partially processed pre-mRNAs (Figure 2C). To further support that the above detected bands come from IPA, we performed RT-PCR with reverse primers placed downstream of the putative intronic PASs. We observed no amplification with these 'downstream primers' in introns 2, 4, and 10, in comparison to reverse primers upstream of the PASs (Figure 2D). Nevertheless, intron 12 might contain additional polyadenylation sites upstream of the 3'-most primer (I12-dR), and the faint band seen with the intron 6 'downstream primer' is consistent with mRNAs exhibiting intron 6 and 7 retention as previously reported [36]. Overall, the 5'-intronic retention strongly suggests that the detected transcripts (except perhaps for intron 6) are generated by IPA and not by partial processing.

3.2 Induction of IPA upon functional U1 knockdown

We transfected HEK293T cells with a U1 decoy plasmid to express an RNA molecule with a consensus 5'ss sequence (D1), which base pairs to U1 and prevents its binding to *bona-fide* 5'ss [39] (Figure 1B). In addition to repressing splicing, the U1 decoy also enhances IPA [4]. As negative controls we used a decoy plasmid with a mutation that compromises its binding to U1 (D7), as well as mock transfection. As a positive control for D1-mediated inhibition of U1, we tested its effect on alternative splicing. D1 but not D7 or mock decreased *CD46* exon 8 inclusion, consistent with this exon being recognized by U1 (Figure 3A) [26]. In addition, D1 decoy increased the ratio of intron-retained to full-length transcripts for introns 2, 4, 10 and 12 (Figure 3B), consistent with IPA. For introns 10 and 12, D1 increased both exon 13 skipping and IPA, but this assay with different primers at different ratios cannot directly reveal which of these two effects is more dominant. All these results further support IPA occurring in introns 2, 4, 10 and 12.

3.3 Several APA-derived *CD46* mRNAs identified by 3'RACE

To detect APA isoforms, we used 3'RACE on total RNA extracted from K562 myeloid cell line, with nested PCR to increase the amplification specificity (Figure 1B, Figures 4-6). For IPA, we identified mRNAs from the use of three intronic PASs in intron 6 (with one reported just by cDNA/EST analysis [3]), one in intron 10 (Figure 4) and one in intron 12 (Figure 5). We did not detect the mRNAs polyadenylated in introns 2 and 4. All but one of the detected IPA-derived isoforms have an in-frame stop codon, thus being capable of producing truncated CD46 proteins, as either soluble or tail-less membrane-bound isoforms. The isoform generated by IPA at the 12th nucleotide of intron 6 lacks an in-frame stop codon, thereby serving as a possible target for degradation by Non-Stop mediated mRNA Decay (NSD) [40]. We also identified one IPA event in intron 13 (Figure 6A) and two tandem 3'UTR APA events in exon 14 in addition to the mRNA with the reference poly(A) tail (Figure 6B), all generating mRNA isoforms with shortened 3'UTRs.

3.4 *CD46 Pseudogene* only produces short transcripts via IPA

By 3'RACE with moderate stringency, we inadvertently identified one polyadenylated mRNA transcript from the *CD46 Pseudogene* (*CD46P1*). This gene is a partial duplicate of *CD46* exons 1 to 5 [41] within the RCA cluster (Figure 7A), yet no evidence of its expression was ever reported until now [12], even though active transcription marks such as H3K27 acetylation are found around *CR1L/CD46P1* first exon (UCSC genome browser), as well as RNA pol II footprints sit along the *CD46P1* gene (data not shown) [42]. The identified transcript was cleaved and polyadenylated at the 5' portion of *CD46P1* intron 2 (Figure 7B), with a point mutation 22 nucleotides upstream the poly(A) tail creating a consensus poly(A) signal (AAUAAA). To test whether *CD46P1* produces longer transcripts, we performed RT-PCR on total K562 RNA. We failed to detect any products from *CD46P1* downstream intron 2, but instead we found *CD46*-derived transcripts (Figure 7C). We also conducted 3'RACE with first and nested primers specific to *CD46P1* annotated exons, in order to identify longer polyadenylated transcripts from this pseudogene. We observed no *CD46P1* transcripts under very stringent 3'RACE with high annealing temperature to minimize cross-priming to *CD46*. The failure to detect any polyadenylated *CD46P1* transcripts was not likely due to primer inefficiency because we detected polyadenylated transcripts from the RCA gene Complement receptor 1-like (*CR1L*) (data not shown), whose first exon overlaps with that of *CD46P1*. We conclude that *CD46P1* is very likely a nonfunctional gene, from which only short transcripts are generated due to IPA in intron 2.

4. Discussion

We demonstrated that *CD46* undergoes APA at several sites, within some introns and in exon 14, generating alternative isoforms with potentially relevant functions (Figure 8).

4.1 Evidence for IPA-generated *CD46* transcripts

We provided three lines of evidence for the various *CD46* IPA isoforms, which include detection of 5' intronic fragments in mRNAs, U1-decoy enhancement of IPA, and 3'RACE. While 3'RACE provides the most solid proof, the first two criteria help rule out 3'RACE artifacts by oligo-dT spuriously binding to genome-encoded A-rich regions. Indeed, none of these three experiments stands on its own as absolute proof for IPA, but their combination result in strong evidence. In intron 6, we identified three IPA transcripts by 3'RACE (Figure 4) but we failed to detect their response to functional U1 knockdown (data not shown), possibly because the levels of such transcripts are very low due to the lack of an in-frame stop codon in one isoform which might mostly be degraded by NSD [40]. Finally, there might be other intron 12 IPA transcript(s) besides the one we found, because the intronic sequence downstream of the found intron 12 PAS was retained in mRNA (Figure 2D).

4.2 IPA may generate heterologous CD46 isoforms with distinct functions and therapeutic applications

Our study provides a means of generating soluble CD46 independent of proteolysis, and in addition to intron retention [36]. Proteolysis-generated soluble CD46 containing

all CCP and STP regions retains C3b/C4b binding capacity, as it can also facilitate fluid cleavage by Factor I [30] thus protecting CD46-nonbearing cells from complement damage, albeit at much lower efficiency than membrane-bound CD46 [35]. The potential CD46 protein isoform by intron 10 IPA, possessing all CCP- and STP-encoding exons but lacking the TM exons, might show a similar function. Similarly, the predicted tail-less CD46, encoded by intron 12 IPA transcript, is likely to retain its ligand affinity and cofactor activity but should be unable to induce intracellular signaling.

Many primary tumors and cancerous cell lines overexpress complement regulatory proteins, including CD46, which appear to be important for immune evasion [43, 44]. Therefore it is tempting to propose that, similar to the induction of soluble RTKs at the expense of the full-length counterpart via IPA, the same intervention for *CD46* would be beneficial in cancer treatment. However, unlike RTKs whose function entirely depends on intracellular signaling by the cytoplasmic tails, complement inhibition by CD46 is merely mediated by its ligand binding which can be partially maintained in soluble isoforms. Indeed, the increased levels of soluble CD46 generated by proteolytic cleavage after STP regions has been documented in sera from cancer patients [30], which, in this case, was suggested to limit inflammation in vicinity of solid tumors or to misdirect the immune effectors from attacking the tumor. In addition, the elevated soluble CD46 in multiple sclerosis [32] plus its association with Human Herpes Virus 6 genome in this disease [45] suggest that soluble CD46 might play a role in the etiology of this and perhaps other autoimmune disorders. More studies are required to establish the roles of soluble CD46 and the therapeutic value of *CD46* IPA manipulation in these diseases.

4.3 APA might regulate CD46 protein levels and offer opportunities for artificial manipulation

As tandem 3'UTR APA is very prevalent, it was not very surprising to find two PASs in addition to the major or annotated site at the *CD46* 3'-terminal exon 14, which is 1995 nt long. The two novel PASs within exon 14 and the site extending exon 13, all generate shorter 3'UTRs that lack the microRNA miR520b- and miR520e-binding site. As CD46 downregulation by these miRNAs caused insensitivity of breast cancer cell lines to complement-dependent cytotoxicity [46], tumors could upregulate CD46 by increasing use of these proximal PASs, and artificial repression of the proximal PASs (i.e. by ASOs) could restore normal CD46 levels. In turn, intron 6 IPA followed by NSD might be physiologically used under certain conditions to quickly downregulate CD46 before transcriptional shutdown.

4.4 Mutation-induced IPA severely shortens *CD46 Pseudogene 1* transcripts

We identified one mRNA from *CD46P1* as the result of IPA at the beginning of intron 2 (Figure 7B). Compared to *CD46* genomic sequence, *CD46P1* shows a G to A change at position +5 of intron 2 5'ss. This change both strongly weakens the 5'ss of *CD46P1* intron 2 [5], and simultaneously creates a perfect consensus poly(A) signal hexamer AAUAAA. Hence, the IPA in intron 2 of *CD46P1* is due to this single mutation which strongly tilts the competitive balance between splicing and polyadenylation in favor of the latter.

Whether the *CD46P1* intron 2 IPA transcript would be translated is unclear. The encoded protein would contain only the CCP1 region, which does not directly bind C3b/C4b but together with CCP2, is essential for the binding and entry of measles virus [47] and adenovirus [48]. Hence, this *CD46P1*-derived CCP1 might act as a decoy to prevent infection of these viruses, as recently shown for soluble CCP1-only CD46 preventing infection of bovine viral diarrhea virus in *Bos taurus* [49]. Most likely *CD46P1* IPA in intron 2 might serve as a means to prevent *CD46P1* expression by inducing early polyadenylation, as our lack of detection of *CD46P1* regions downstream exon 2 makes expression of full-length *CD46P1* (exons 1-5) very rare at best. Intron 2 IPA might explain the so-far elusive detection of *CD46P1* mRNA, despite possessing comparable promoter activity [50] and high homology to 5' portion of *CD46* [41].

5 Conclusion and future work

Future studies should test the abundance of these novel *CD46* APA transcripts and detection of the encoded proteins in human tissues. Based on the lack of reports on most *CD46* APA isoforms until now, we anticipate that their expression levels will be low or confined to specific cells and/or physiological or pathological conditions. The lack of conservation of *CD46* gene structure and expression between human and most mammals, including mice, and the very low detection of soluble CD46 in cell lines [33], restricts the search of such isoforms to human tissues or fluids. Nevertheless, this first description of the diversity of *CD46* APA isoforms identified PASs all along the *CD46* transcript (Figure 8). Whereas the intron 6, 10 and 12 IPA would generate soluble or membrane-bound but tail-less isoforms, intron 13 IPA and tandem APA in exon 14 would regulate CD46 abundance. This report further confirms IPA as a widespread mechanism regulating expression of transmembrane receptors beyond RTKs. Lastly, *CD46P1* is likely nonfunctional because it generates very short transcripts via intron 2 IPA. This and future studies could open therapeutic avenues by manipulating the balance of these APA isoforms in the context of CD46-related pathologies, such as cancer and autoimmunity.

ACKNOWLEDGEMENTS

We acknowledge members of Roca lab for reagents and advice.

FUNDING

XR acknowledges funding from Academic Research Fund Tier 1 RG 20/11 and Academic Research Fund Tier 2 MOE2013-T2-1-101 (ARC 45/13), both from Singapore's Ministry of Education. The funders played no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflicts of Interest

The authors have no conflicts of interest to declare.

References

1. Elkon R, Ugalde AP and Agami R. Alternative cleavage and polyadenylation: extent, regulation and function. Nat Rev Genet. 2013; 14(7):496-506.
2. Tian B and Manley JL. Alternative polyadenylation of mRNA precursors. Nat Rev Mol Cell Biol. 2017; 18(1):18-30.
3. Tian B, Pan Z and Lee JY. Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing. Genome Res. 2007; 17(2):156-165.
4. Vorlova S, Rocco G, Lefave CV, Jodelka FM, Hess K, Hastings ML, Henke E and Cartegni L. Induction of antagonistic soluble decoy receptor tyrosine kinases by intronic polyA activation. Mol Cell. 2011; 43(6):927-939.
5. Roca X, Krainer AR and Eperon IC. Pick one, but be quick: 5' splice sites and the problems of too many choices. Genes Dev. 2013; 27(2):129-144.
6. Berg MG, Singh LN, Younis I, Liu Q, Pinto AM, Kaida D, Zhang Z, Cho S, Sherrill-Mix S, Wan L and Dreyfuss G. U1 snRNP determines mRNA length and regulates isoform expression. Cell. 2012; 150(1):53-64.
7. Kaida D, Berg MG, Younis I, Kasim M, Singh LN, Wan L and Dreyfuss G. U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. Nature. 2010; 468(7324):664-668.
8. Devany E, Park JY, Murphy MR, Zakusilo G, Baquero J, Zhang X, Hoque M, Tian B and Kleiman FE. Intronic cleavage and polyadenylation regulates gene expression during DNA damage response through U1 snRNA. Cell Discov. 2016; 2:16013.
9. Spraggon L and Cartegni L. U1 snRNP-Dependent Suppression of Polyadenylation: Physiological Role and Therapeutic Opportunities in Cancer. Int J Cell Biol. 2013; 2013:846510.
10. Edwalds-Gilbert G, Veraldi KL and Milcarek C. Alternative poly(A) site selection in complex transcription units: means to an end? Nucleic Acids Res. 1997; 25(13):2547-2561.
11. Liszewski MK and Atkinson JP. Complement regulator CD46: genetic variants and disease associations. Hum Genomics. 2015; 9:7.
12. Liszewski MK, Java A, Schramm EC and Atkinson JP. Complement Dysregulation and Disease: Insights from Contemporary Genetics. Annu Rev Pathol. 2016.
13. Cardone J, Al-Shouli S and Kemper C. A novel role for CD46 in wound repair. Front Immunol. 2011; 2:28.
14. Ludford-Menting MJ, Thomas SJ, Crimeen B, Harris LJ, Loveland BE, Bills M, Ellis S and Russell SM. A functional interaction between CD46 and DLG4: a role for DLG4 in epithelial polarization. J Biol Chem. 2002; 277(6):4477-4484.
15. Riley RC, Kemper C, Leung M and Atkinson JP. Characterization of human membrane cofactor protein (MCP; CD46) on spermatozoa. Mol Reprod Dev. 2002; 62(4):534-546.
16. Seya T, Turner JR and Atkinson JP. Purification and characterization of a membrane protein (gp45-70) that is a cofactor for cleavage of C3b and C4b. J Exp Med. 1986; 163(4):837-855.
17. Arbore G and Kemper C. A novel "complement-metabolism-inflammasome axis" as a key regulator of immune cell effector function. Eur J Immunol. 2016; 46(7):1563-1573.
18. Astier A, Trescol-Biemont MC, Azocar O, Lamouille B and Rabourdin-Combe C. Cutting edge: CD46, a new costimulatory molecule for T cells, that induces p120CBL and LAT phosphorylation. J Immunol. 2000; 164(12):6091-6095.
19. Kemper C, Chan AC, Green JM, Brett KA, Murphy KM and Atkinson JP. Activation of human CD4+ cells with CD3 and CD46 induces a T-regulatory cell 1 phenotype. Nature. 2003; 421(6921):388-392.

20. Bora NS, Matta B, Lyzogubov VV and Bora PS. Relationship between the complement system, risk factors and prediction models in age-related macular degeneration. Mol Immunol. 2015; 63(2):176-183.
21. Fishelson Z, Donin N, Zell S, Schultz S and Kirschfink M. Obstacles to cancer immunotherapy: expression of membrane complement regulatory proteins (mCRPs) in tumors. Mol Immunol. 2003; 40(2-4):109-123.
22. Astier AL, Meiffren G, Freeman S and Hafler DA. Alterations in CD46-mediated Tr1 regulatory T cells in patients with multiple sclerosis. J Clin Invest. 2006; 116(12):3252-3257.
23. Tsai YG, Niu DM, Yang KD, Hung CH, Yeh YJ, Lee CY and Lin CY. Functional defects of CD46-induced regulatory T cells to suppress airway inflammation in mite allergic asthma. Lab Invest. 2012; 92(9):1260-1269.
24. Lublin DM, Liszewski MK, Post TW, Arce MA, Le Beau MM, Rebentisch MB, Lemons LS, Seya T and Atkinson JP. Molecular cloning and chromosomal localization of human membrane cofactor protein (MCP). Evidence for inclusion in the multigene family of complement-regulatory proteins. J Exp Med. 1988; 168(1):181-194.
25. Post TW, Liszewski MK, Adams EM, Tedja I, Miller EA and Atkinson JP. Membrane cofactor protein of the complement system: alternative splicing of serine/threonine/proline-rich exons and cytoplasmic tails produces multiple isoforms that correlate with protein phenotype. J Exp Med. 1991; 174(1):93-102.
26. Tang SJ, Luo S, Ho JX, Ly PT, Goh E and Roca X. Characterization of the Regulation of CD46 RNA Alternative Splicing. J Biol Chem. 2016; 291(27):14311-14323.
27. Cardone J, Le Friec G, Vantourout P, Roberts A, Fuchs A, Jackson I, Suddason T, Lord G, Atkinson JP, Cope A, Hayday A and Kemper C. Complement regulator CD46 temporally regulates cytokine production by conventional and unconventional T cells. Nat Immunol. 2010; 11(9):862-871.
28. Ni Choileain S, Weyand NJ, Neumann C, Thomas J, So M and Astier AL. The dynamic processing of CD46 intracellular domains provides a molecular rheostat for T cell activation. PLoS One. 2011; 6(1):e16287.
29. Cole DS, Hughes TR, Gasque P and Morgan BP. Complement regulator loss on apoptotic neuronal cells causes increased complement activation and promotes both phagocytosis and cell lysis. Mol Immunol. 2006; 43(12):1953-1964.
30. Hakulinen J, Junnikkala S, Sorsa T and Meri S. Complement inhibitor membrane cofactor protein (MCP; CD46) is constitutively shed from cancer cell membranes in vesicles and converted by a metalloproteinase to a functionally active soluble form. Eur J Immunol. 2004; 34(9):2620-2629.
31. Kawano M, Seya T, Koni I and Mabuchi H. Elevated serum levels of soluble membrane cofactor protein (CD46, MCP) in patients with systemic lupus erythematosus (SLE). Clin Exp Immunol. 1999; 116(3):542-546.
32. Soldan SS, Fogdell-Hahn A, Brennan MB, Mittleman BB, Ballerini C, Massacesi L, Seya T, McFarland HF and Jacobson S. Elevated serum and cerebrospinal fluid levels of soluble human herpesvirus type 6 cellular receptor, membrane cofactor protein, in patients with multiple sclerosis. Ann Neurol. 2001; 50(4):486-493.
33. Hara T, Kuriyama S, Kiyohara H, Nagase Y, Matsumoto M and Seya T. Soluble forms of membrane cofactor protein (CD46, MCP) are present in plasma, tears, and seminal fluid in normal subjects. Clin Exp Immunol. 1992; 89(3):490-494.
34. McLaughlin PJ, Holland SJ, Taylor CT, Olah KS, Lewis-Jones DI, Hara T, Seya T and Johnson PM. Soluble CD46 (membrane cofactor protein, MCP) in human reproductive tract fluids. J Reprod Immunol. 1996; 31(3):209-219.
35. Seya T, Hara T, Iwata K, Kuriyama S, Hasegawa T, Nagase Y, Miyagawa S, Matsumoto M, Hatanaka M, Atkinson JP and et al. Purification and functional properties of soluble forms of membrane

cofactor protein (CD46) of complement: identification of forms increased in cancer patients' sera. Int Immunol. 1995; 7(5):727-736.
36. Pollard AJ, Flanagan BF, Newton DJ and Johnson PM. A novel isoform of human membrane cofactor protein (CD46) mRNA generated by intron retention. Gene. 1998; 212(1):39-47.
37. Tan J, Ho JX, Zhong Z, Luo S, Chen G and Roca X. Noncanonical registers and base pairs in human 5' splice-site selection. Nucleic Acids Res. 2016; 44(8):3908-3921.
38. Tan J and Roca X. Informational Suppression to Probe RNA:RNA Interactions in the Context of Ribonucleoproteins: U1 and 5' Splice-Site Base-Pairing. Methods Mol Biol. 2016; 1421:243-268.
39. Roca X and Krainer AR. Recognition of atypical 5' splice sites by shifted base-pairing to U1 snRNA. Nat Struct Mol Biol. 2009; 16(2):176-182.
40. Klauer AA and van Hoof A. Degradation of mRNAs that lack a stop codon: a decade of nonstop progress. Wiley Interdiscip Rev RNA. 2012; 3(5):649-660.
41. Hourcade D, Garcia AD, Post TW, Taillon-Miller P, Holers VM, Wagner LM, Bora NS and Atkinson JP. Analysis of the human regulators of complement activation (RCA) gene cluster with yeast artificial chromosomes (YACs). Genomics. 1992; 12(2):289-300.
42. Scheidegger A, Burkholder A, Abbas A, Zarns K, Samarakkody A and Nechaev S. Analysis of paired end Pol II ChIP-seq and short capped RNA-seq in MCF-7 cells. Genom Data. 2015; 5:263-267.
43. Rushmere NK, Knowlden JM, Gee JM, Harper ME, Robertson JF, Morgan BP and Nicholson RI. Analysis of the level of mRNA expression of the membrane regulators of complement, CD59, CD55 and CD46, in breast cancer. Int J Cancer. 2004; 108(6):930-936.
44. Seya T, Hara T, Matsumoto M and Akedo H. Quantitative analysis of membrane cofactor protein (MCP) of complement. High expression of MCP on human leukemia cell lines, which is down-regulated during cell differentiation. J Immunol. 1990; 145(1):238-245.
45. Fogdell-Hahn A, Soldan SS, Shue S, Akhyani N, Refai H, Ahlqvist J and Jacobson S. Co-purification of soluble membrane cofactor protein (CD46) and human herpesvirus 6 variant A genome in serum from multiple sclerosis patients. Virus Res. 2005; 110(1-2):57-63.
46. Cui W, Zhang Y, Hu N, Shan C, Zhang S, Zhang W, Zhang X and Ye L. miRNA-520b and miR-520e sensitize breast cancer cells to complement attack via directly targeting 3'UTR of CD46. Cancer Biol Ther. 2010; 10(3):232-241.
47. Santiago C, Celma ML, Stehle T and Casasnovas JM. Structure of the measles virus hemagglutinin bound to the CD46 receptor. Nat Struct Mol Biol. 2010; 17(1):124-129.
48. Persson BD, Reiter DM, Marttila M, Mei YF, Casasnovas JM, Arnberg N and Stehle T. Adenovirus type 11 binding alters the conformation of its receptor CD46. Nat Struct Mol Biol. 2007; 14(2):164-166.
49. Alzamel N, Bayrou C, Decreux A and Desmecht D. Soluble forms of CD46 are detected in Bos taurus plasma and neutralize BVDV, the bovine pestivirus. Comp Immunol Microbiol Infect Dis. 2016; 49:39-46.
50. Cui W, Hourcade D, Post T, Greenlund AC, Atkinson JP and Kumar V. Characterization of the promoter region of the membrane cofactor protein (CD46) gene of the human complement system and comparison to a membrane cofactor protein-like genetic element. J Immunol. 1993; 151(8):4137-4146.

Figure Legends

Figure 1. A. Schematic of CD46 protein motifs and their coding sequence in *CD46* mRNA, with boxes as exons. Cassette exons regulated by alternative splicing are shown as black boxes, while constitutive exons are in gray. The two functional stop codons are indicated by the stop

signs. See text for details. **B.** Schematic of experimental approaches to test for APA in *CD46* transcripts.

Figure 2. Retention of 5' intronic fragments as evidence for *CD46* IPA. **A.** Schematic of RT-PCR design with primers depicted as arrows mapping to exons (boxes) or introns (lines). The forward primers in exons are paired in the same row with one or more reverse primers (one per reaction) mapping to 5' portion of introns. **B.** Agarose-gel electrophoresis of RT-PCR data for each combination of primers in four cell lines to test for IPA in each indicated intron. The identity of the PCR bands are schematically depicted on the right, as confirmed by sequencing. The position of the marker bands are indicated with their size in bp. **C.** Rationale of the 5' to 3' intronic retention as evidence of IPA. Right, partially-spliced transcripts would be amplified with all reverse primers along the intron, upstream (U) or downstream (D) the PAS (as long as the product is not too long). Conversely, IPA-derived transcripts on the left would only be amplified by reverse primers upstream but not downstream the PAS. **D.** Schematics depict primer design for detection of 5' to 3' intronic fragments by RT-PCR. Right of each schematic shows corresponding agarose-gel images of RT-PCR data for each intron, with the exonic or intronic reverse primer indicated on top of each lane.

Figure 3. U1 functional knockdown by decoys enhanced IPA in several *CD46* introns. **A.** Control experiment for U1 decoys. Left schematic indicates the position of the primers, and right agarose gel image of RT-PCR upon HEK293T transfection with decoy plasmids shows that exon 8 inclusion decreased by U1 decoy (D1) but not by mock transfection (M) or mutant U1 decoy (D7). RT-PCR bands are represented on the right of the gel. **B.** Representative gel images of RT-PCRs for each combination of forward with intronic and exonic reverse primers (upper schematics) upon HEK293T transfection with mock or decoy plasmids. In all cases in this figure, the ratio of intronic retention vs splicing increased by D1 decoy. Numbers below each lane indicate the IPA percentage in this experiment, while the increased IPA by D1 was seen in the other two experimental replicates (not shown). This result is consistent with the decoy de-repressing IPA by reducing U1 occupancy at the 5'ss upstream the PAS.

Figure 4. 3'RACE in K562 cells as definitive evidence for IPA isoforms encoding different CD46 proteins. Sequencing results of 3'RACE for different introns potentially generating soluble CD46. The sequencing chromatograms are aligned to the genomic sequence, with yellow boxes as putative poly(A) signals. For each IPA mRNA, the exon-intron structure and the predicted soluble protein are schematically shown, highlighting the position of the PAS (first A of the tail) and the intronic in-frame stop codon from the beginning of the intron as +1. The top intron 6 IPA transcript lacks a stop codon so it is likely a substrate for NSD.

Figure 5. Sequencing results of 3'RACE for intron 12 potentially generating a membrane-bound but tail-less CD46.

Figure 6. 3'RACE in K562 as definitive evidence for APA mRNA isoforms that might change the expression levels of CD46 proteins. **A.** Sequencing results of 3'RACE for intron 13 IPA, resulting in a very short 3'UTR which could enhance translation of CD46-CYT1 isoforms. The sequencing chromatograms are aligned to the genomic sequence, highlighting the putative poly(A) signal and the distance from the beginning of intron 13 to the PAS. **B.** Sequencing results of 3'RACE for exon 14 tandem 3'UTR APA, showing two upstream PASs at the indicated

positions counted from the first nucleotide of exon 14 as +1. Even though some of the 3'RACE results for alternative exon 14 PAS exhibited additional 'noisy' sequences downstream the poly(A) tail (not shown), the finding of these isoforms in both exon 13-containing or lacking *CD46* mRNAs, plus the proximity of these PASs to the yellow boxed putative poly(A) signals argues for these representing real APA isoforms.

Figure 7. CD46 Pseudogene (*CD46P1*) only generates short transcripts by IPA. **A.** Schematic representation of *CD46P1* compared with *CD46*, with boxes and lines respectively representing exons and introns. **B.** Sequencing results of 3'RACE with nested PCR in K562 for intron 2 IPA showed the *CD46P1* transcripts polyadenylated in intron 2. Top, genomic *CD46* and *CD46P1* regions aligned with the 3'RACE mRNA results, highlighting the G to A transition at position +5 of intron, weakening the 5'ss and creating a perfect consensus poly(A) signal (yellow box). The mRNA and potentially encoded soluble CD46 fragment are also shown. **C.** RT-PCR to detect other *CD46P1* mRNA isoforms in K562, instead only showed amplification of *CD46*. Primers are represented in the diagram with vertical lines depicting nucleotide differences between *CD46* and *CD46P1*. The annealing temperature for each PCR is shown on top. Note that noncontiguous lanes are placed next to one another, as highlighted by white space separating them. Amplification disappears at high temperature because the primers were designed for *CD46P1* and not *CD46*.

Figure 8. Confirmed alternative PASs along the *CD46* pre-mRNA found in this study, with their putative encoded protein isoforms. The PASs are numbered 1-3 relative to each intron or exon. See text for details.

Figure 1

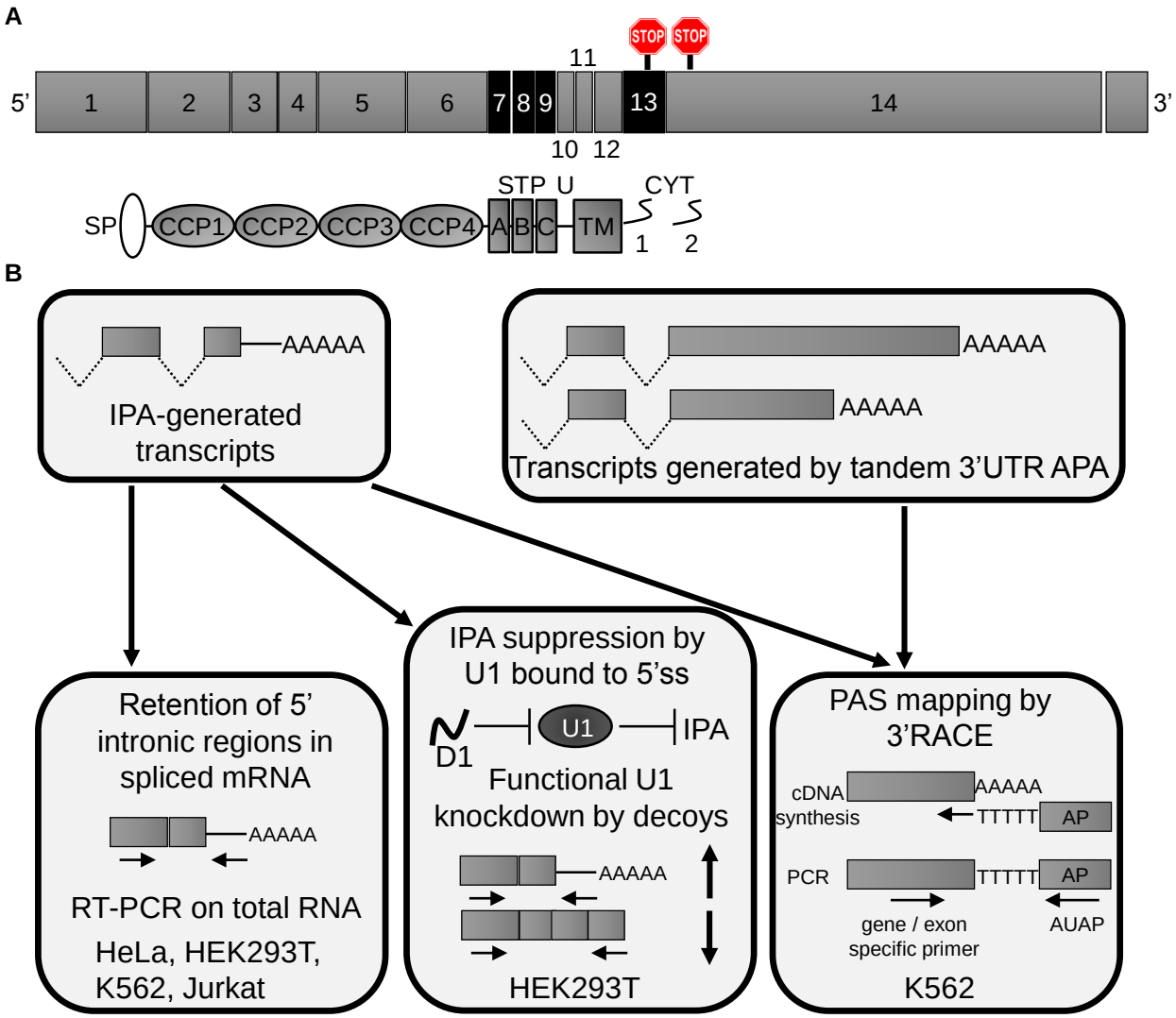

A
STOP
STOP
11
5'
1
2
3
4
5
6
7
8
9
13
14
3'
10
12
STP
U
CYT
SP
CCP1
CCP2
CCP3
CCP4
A
B
C
TM
1
2
B
AAAAA
IPA-generated transcripts
AAAAA
AAAAA
Transcripts generated by tandem 3'UTR APA
Retention of 5' intronic regions in spliced mRNA
AAAAA
RT-PCR on total RNA
HeLa, HEK293T, K562, Jurkat
IPA suppression by U1 bound to 5'ss
U1
IPA
D1
Functional U1 knockdown by decoys
AAAAA
HEK293T
PAS mapping by 3'RACE
cDNA synthesis
AAAAA
TTTTT
AP
PCR
TTTTT
AP
gene / exon specific primer
AUAP
K562

Figure 2

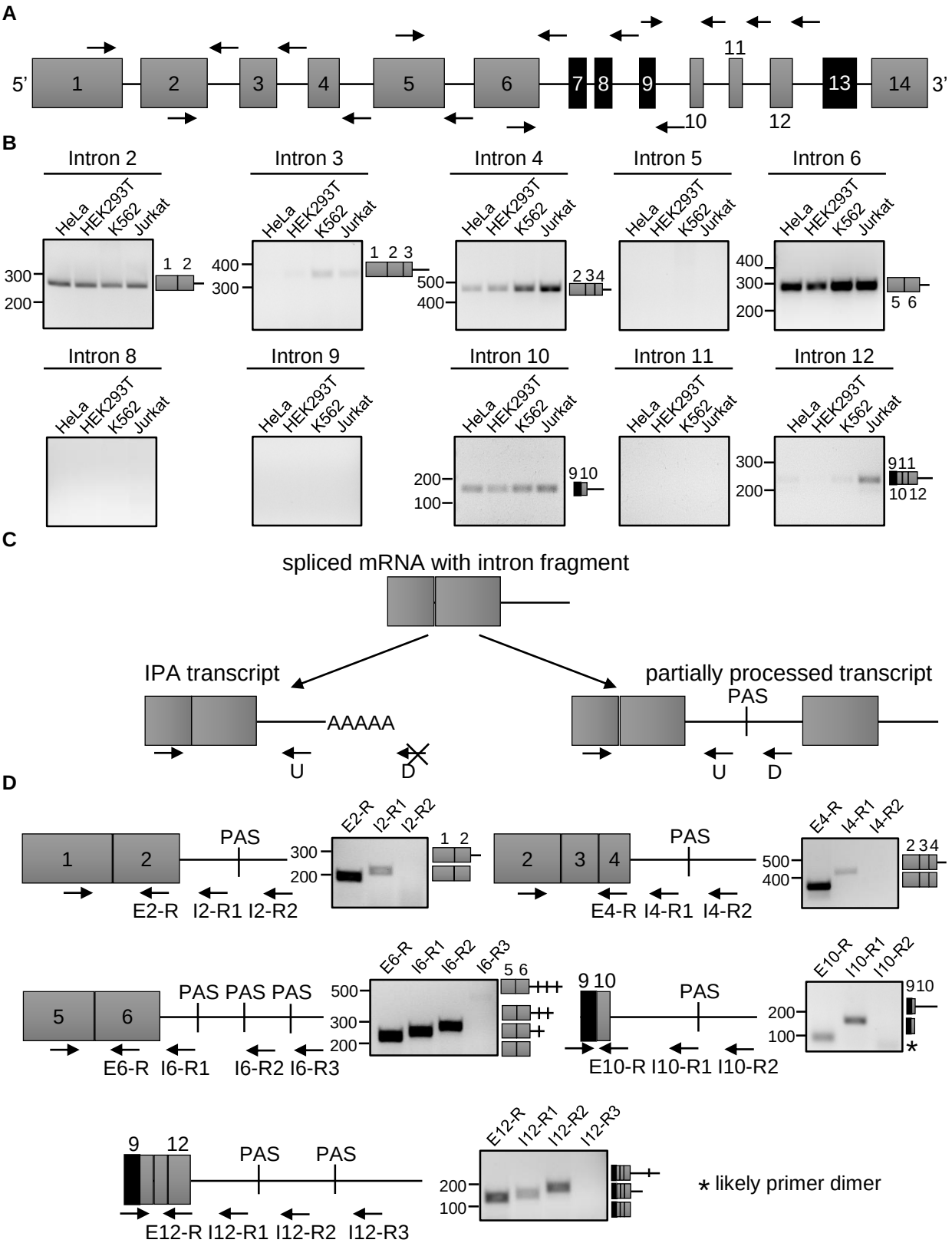

A
5'
1
2
3
4
5
6
7
8
9
10
11
12
13
14
3'
B
Intron 2
Intron 3
Intron 4
Intron 5
Intron 6
Intron 8
Intron 9
Intron 10
Intron 11
Intron 12
HeLa
HEK293T
K562
Jurkat
300
200
400
500
100
1 2
1 2 3
2 3 4
5 6
9 10
9 11
10 12
C
spliced mRNA with intron fragment
IPA transcript
partially processed transcript
AAAAA
PAS
U
D
D
E2-R
I2-R1
I2-R2
E4-R
I4-R1
I4-R2
PAS PAS PAS
E6-R
I6-R1
I6-R2
I6-R3
E10-R
I10-R1
I10-R2
E12-R
I12-R1
I12-R2
I12-R3
* likely primer dimer

Figure 3

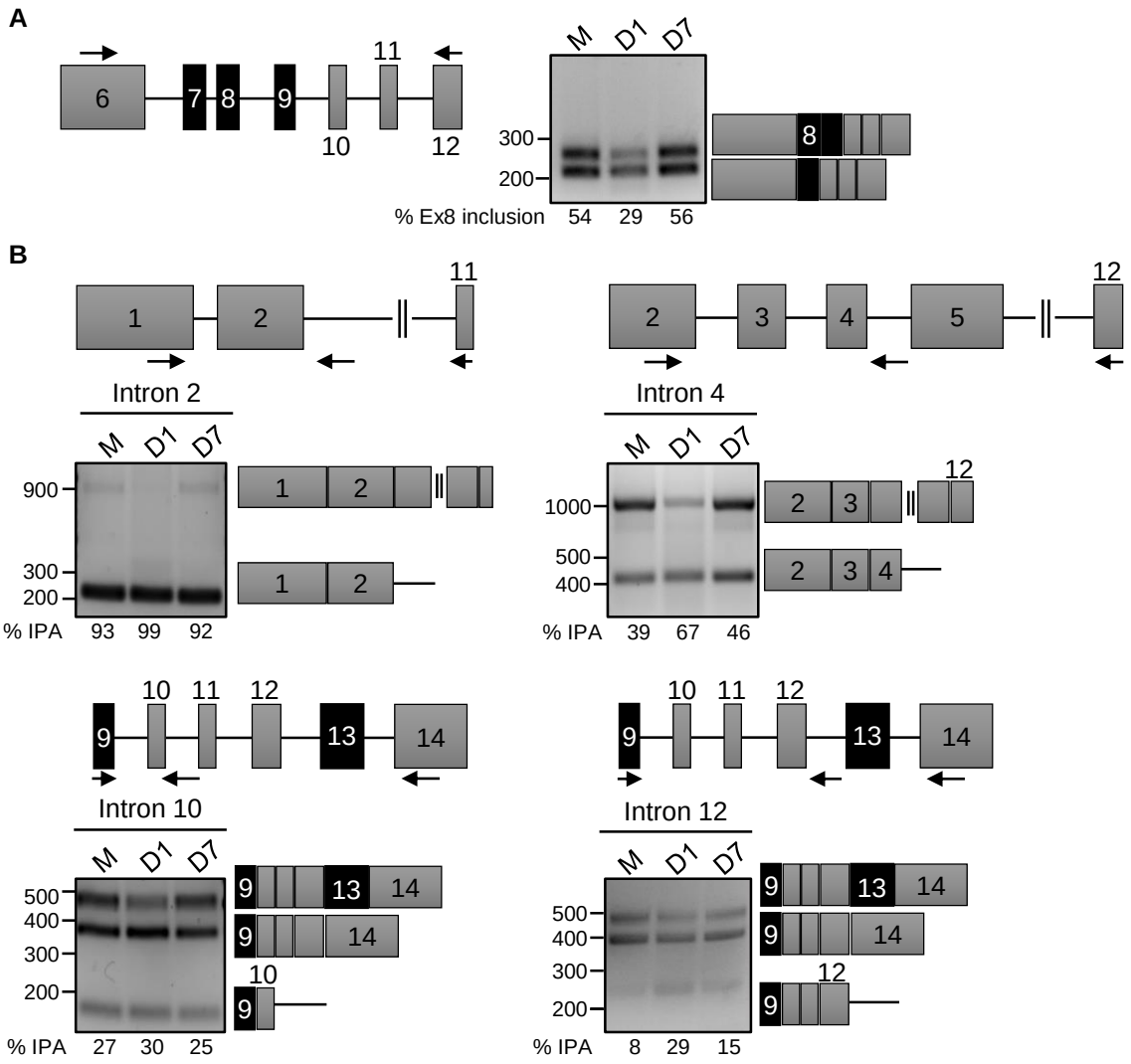

A
6
7
8
9
10
11
12
M
D1
D7
300
200
8
% Ex8 inclusion
54
29
56
B
1
2
11
12
2
3
4
5
Intron 2
Intron 4
900
300
200
% IPA
93
99
92
1000
500
400
39
67
46
10
13
14
Intron 10
Intron 12
27
30
25
8
29
15

Figure 4

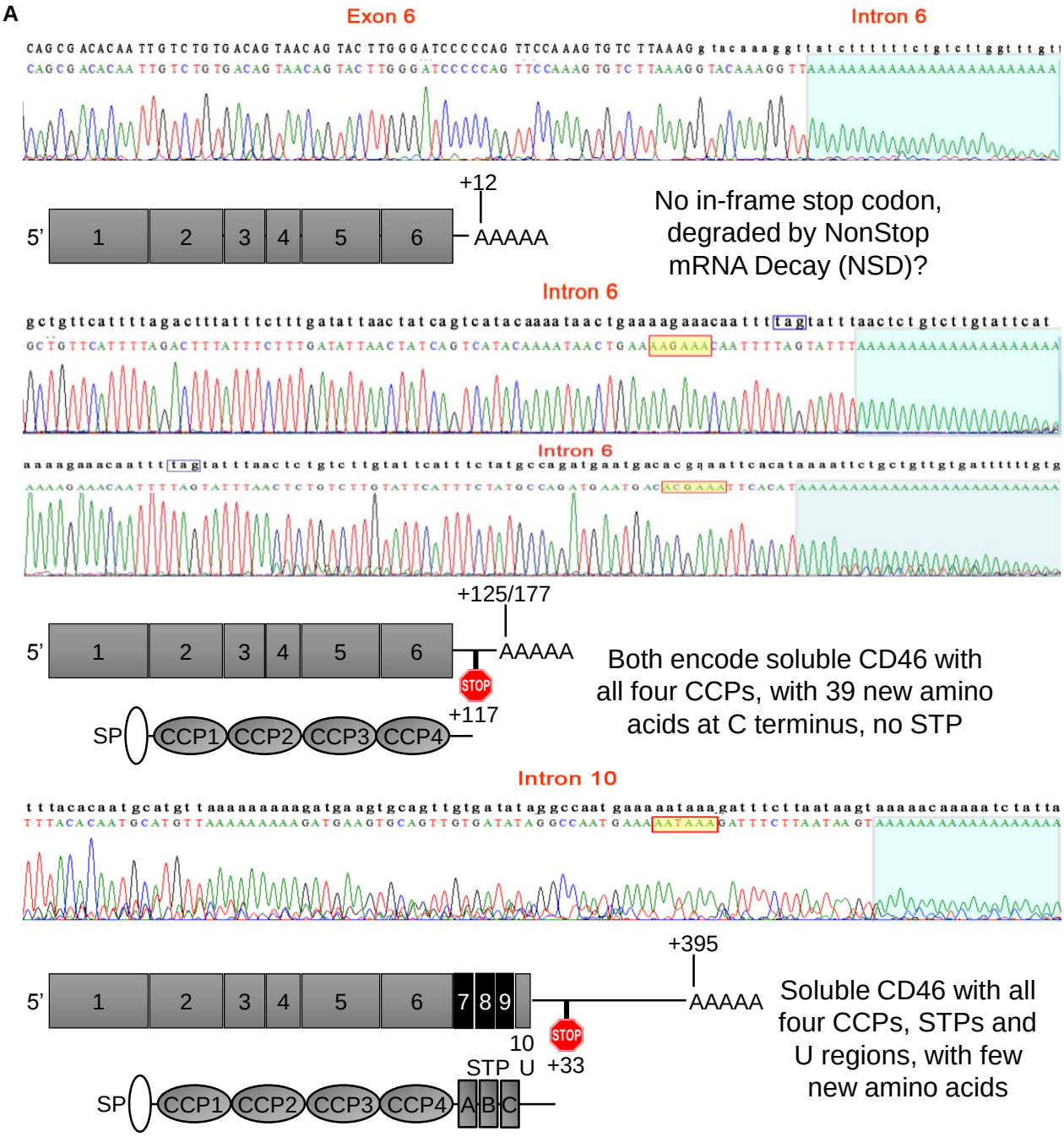
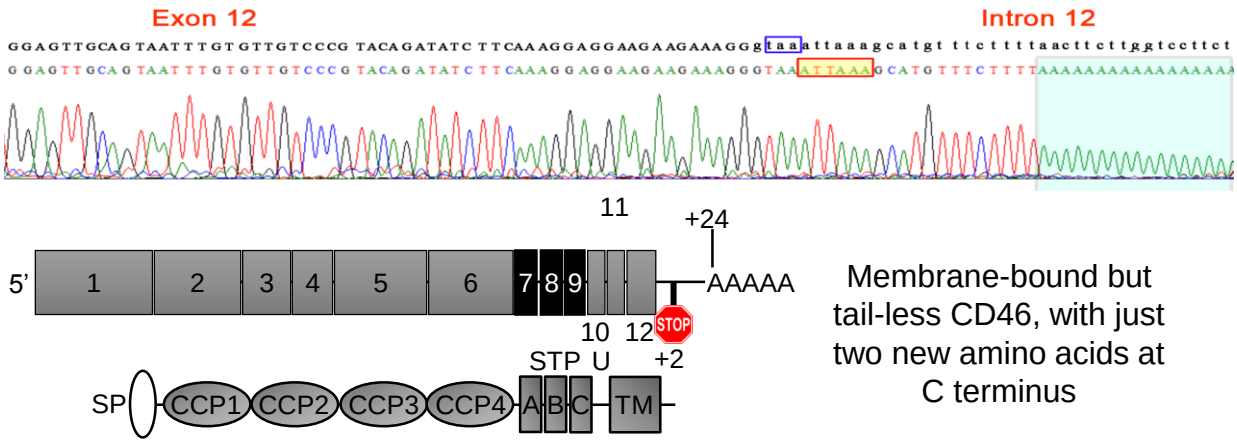

A
Exon 6
Intron 6
+12
5'
1
2
3
4
5
6
AAAAA
No in-frame stop codon, degraded by NonStop mRNA Decay (NSD)?
Intron 6
Intron 6
+125/177
5'
1
2
3
4
5
6
AAAAA
STOP
+117
SP
CCP1
CCP2
CCP3
CCP4
Both encode soluble CD46 with all four CCPs, with 39 new amino acids at C terminus, no STP
Intron 10
+395
5'
1
2
3
4
5
6
7
8
9
AAAAA
STOP
10
STP
U
+33
SP
CCP1
CCP2
CCP3
CCP4
A
B
C
Soluble CD46 with all four CCPs, STPs and U regions, with few new amino acids

Figure 5

A



Membrane-bound but tail-less CD46, with just two new amino acids at C terminus

Figure 6

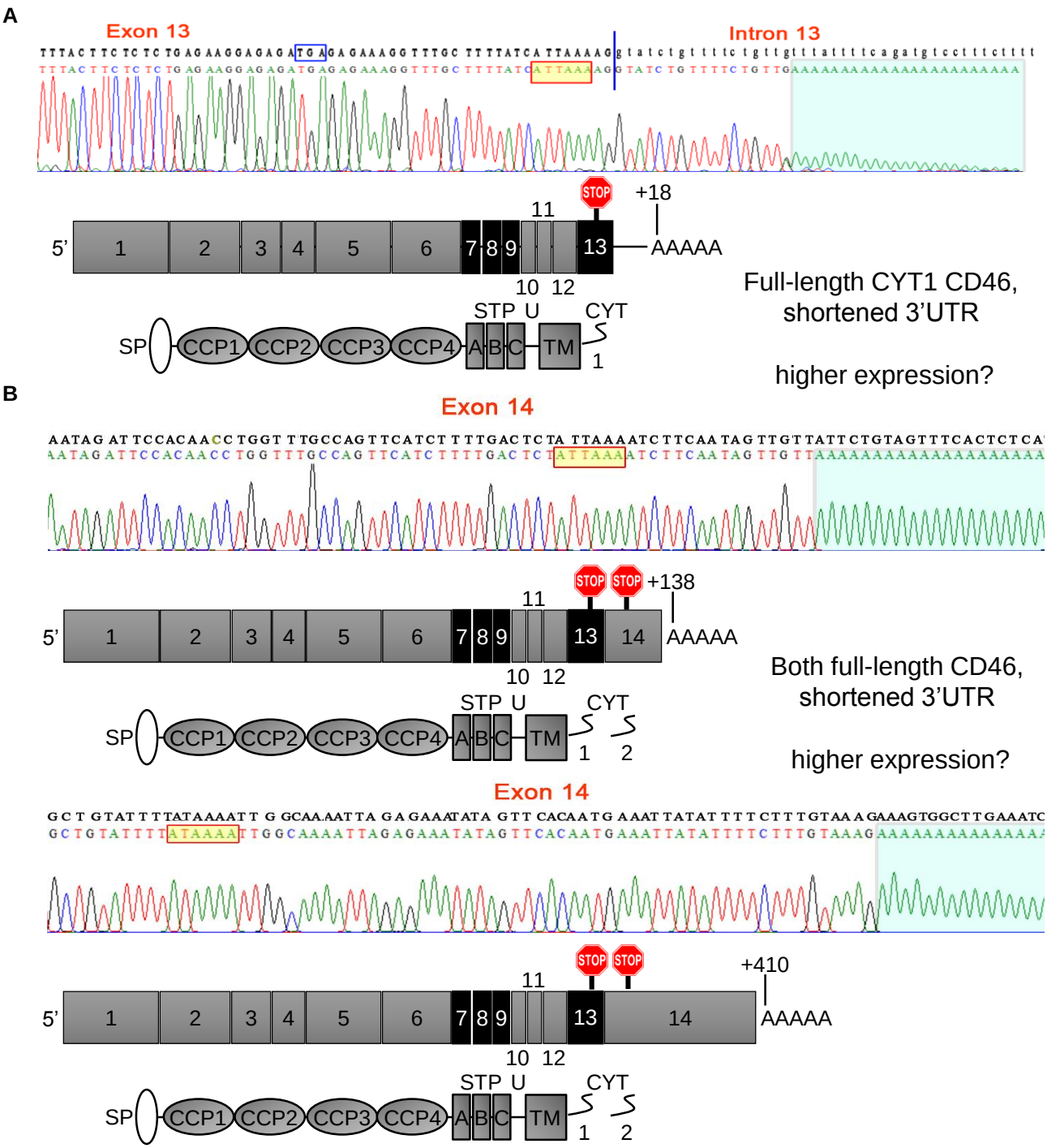

A
Exon 13
Intron 13
STOP
+18
11
5'
1
2
3
4
5
6
7
8
9
13
AAAAA
10
12
STP
U
CYT
SP
CCP1
CCP2
CCP3
CCP4
A
B
C
TM
1
Full-length CYT1 CD46, shortened 3'UTR
higher expression?
B
Exon 14
STOP
STOP
+138
14
Both full-length CD46, shortened 3'UTR
higher expression?
Exon 14
+410

Figure 7

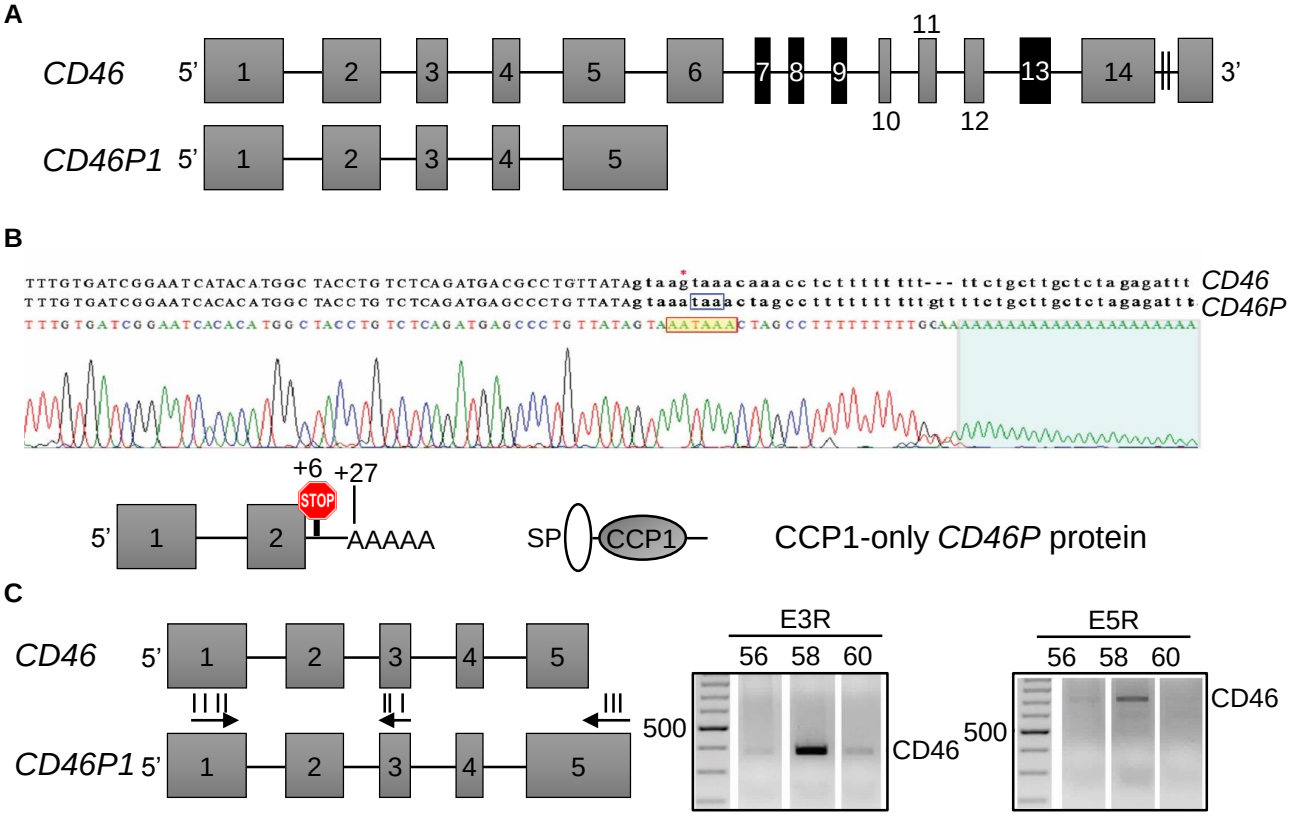
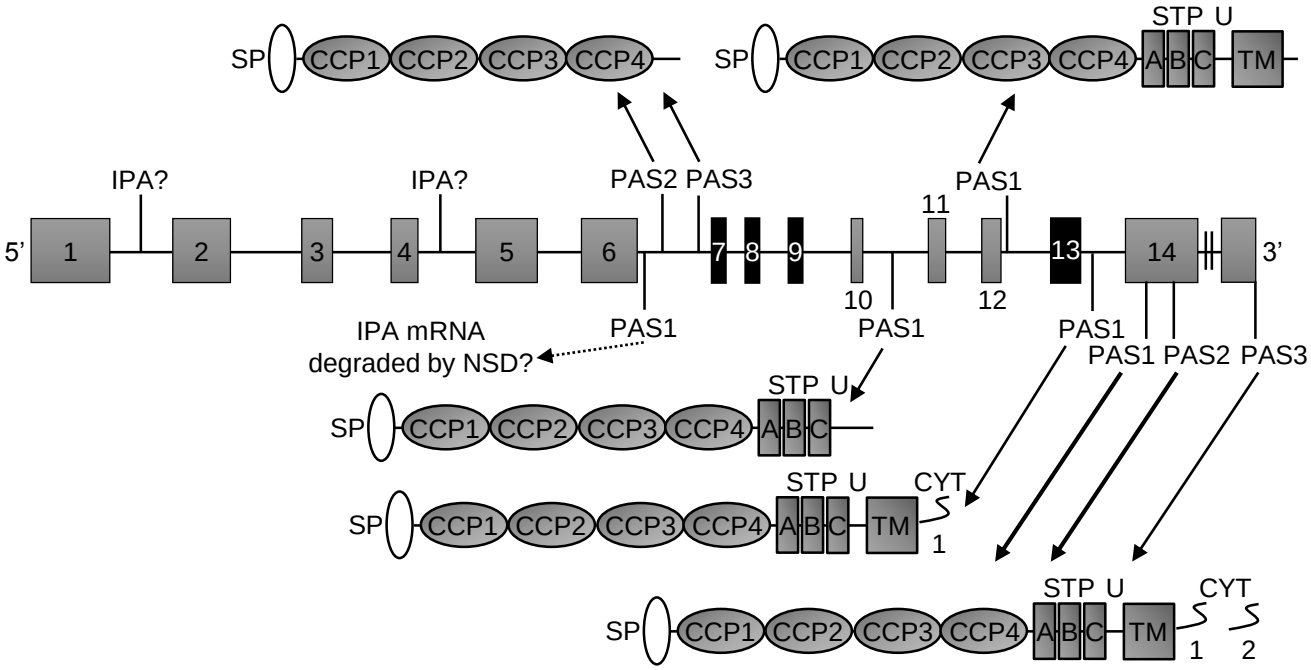

A
CD46
5'
1
2
3
4
5
6
7
8
9
10
11
12
13
14
3'
CD46P1
5'
1
2
3
4
5
B
CD46
CD46P
+6
+27
STOP
5'
1
2
AAAAA
SP
CCP1
CCP1-only CD46P protein
C
CD46
5'
1
2
3
4
5
CD46P1
5'
1
2
3
4
5
E3R
56
58
60
500
CD46
E5R
56
58
60
500
CD46

Figure 8


SP
CCP1
CCP2
CCP3
CCP4
STP
U
A
B
C
TM
IPA?
IPA?
PAS2
PAS3
PAS1
5'
1
2
3
4
5
6
7
8
9
10
11
12
13
14
3'
IPA mRNA
degraded by NSD?
PAS1
PAS1
PAS1
PAS1
PAS2
PAS3
CYT
1
2