(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
12 January 2017 (12.01.2017)

WIPO | PCT

(10) International Publication Number
WO 2017/008065 A2

(51) **International Patent Classification**:
C12N 15/86 (2006.01) *C12N 15/85* (2006.01)

(21) **International Application Number**:
PCT/US2016/041659

(22) **International Filing Date**:
8 July 2016 (08.07.2016)

(25) **Filing Language**: English

(26) **Publication Language**: English

(30) **Priority Data**:
62/190,232 8 July 2015 (08.07.2015) US

(71) **Applicant**: **MEMORIAL SLOAN-KETTERING CANCER CENTER** [US/US]; 1275 York Avenue, New York, NY 10021 (US).

(72) **Inventors**: **MAYR, Christine**; c/o Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021 (US). **MA, Weirui**; c/o Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021 (US).

(74) **Agents**: **GOGORIS, Adda** et al.; Leason Ellis LLP, One Barker Avenue, Fifth Floor, White Plains, NY 10601 (US).

(81) **Designated States** *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

— *with sequence listing part of description (Rule 5.2(a))*

WO 2017/008065 A2

(54) **Title**: VECTOR CONSTRUCTS AND METHODS FOR ACHIEVING OR ENHANCING PROTEIN LOCALIZATION TO THE SURFACE OF CELLS

(57) **Abstract**: The present disclosure relates to fragments of 3 'UTRs that comprise an HuR binding site and that promote surface localization of proteins, even proteins not normally substantially localized on a cell surface. Also disclosed are vectors incorporating such fragments and methods for increasing localization of a polypeptide on the surface of a cell.

VECTOR CONSTRUCTS AND METHODS FOR ACHIEVING OR ENHANCING PROTEIN LOCALIZATION TO THE SURFACE OF CELLS

STATEMENT OR RELATED APPLICATIONS

This patent application claims the priority of U.S. Provisional Patent Application 62/190,232 filed July 8, 2015; the contents of this provisional application are hereby incorporated by reference in their entirety.

GOVERNMENT RIGHTS

The work described herein was funded in part by a grant from the National Institutes of Health U01-CA164190. The U.S. government may have certain rights in this disclosure.

SEQUENCE LISTING

The present application includes a sequence listing in electronic form as a txt file in ascii format titled "7704-0027-Z_ST25[5][4].txt" and having a size of 51.8 kb. The contents of this txt file are incorporated by reference herein.

BACKGROUND OF THE DISCLOSURE

Technical Field

[0001] This disclosure generally relates to membrane localization of proteins regulated by certain 3'UTR isoforms. More specifically, this disclosure relates to achieving or enhancing protein localization on a cell membrane using components of 3'UTRs.

Description of the Related Art

[0002] It was known that 3'UTRs can regulate mRNA localization. This was impressively shown in the Drosophila embryo where about 70% of mRNAs are localized (Lecuyer E et al. *Cell*. 5;131(1):174-87 (2007)). Furthermore, it was shown by this report, that in the Drosophila embryo the localization of many proteins is determined by RNA localization, because the protein is then generated locally.

[0003] But, it was never put forward that 3'UTRs may be involved in membrane protein localization because it was thought that the information necessary for protein localization is contained either in the amino acid sequence or in the secondary structure of the protein (Horton et al. Nucleic Acids Res. Jul;35 (2007)).

SUMMARY OF THE DISCLOSURE

[0004] In one aspect, the disclosure is directed to a recombinant vector comprising a promoter, the vector being adapted for introduction of a first polynucleotide encoding a first polypeptide of interest to be expressed in a target cell and localized on the surface of the target cell and a second polynucleotide segment comprising a fragment of a longer 3'UTR or nucleic acid encoding a fragment of a longer 3'UTR, the fragment derived from nucleic acid encoding a second polypeptide that is substantially localized to the surface of a cell in which it is expressed in nature, said fragment comprising an HuR binding site.

[0005] In some embodiments, the polynucleotide segment is flanked by restriction sites.

[0006] In some embodiments, a CAC motif is interposed between a first (5') restriction site and the fragment.

[0007] In some embodiments, the fragment is up to about 1000 nucleotides long and in some embodiments up to about 500 nucleotides long.

[0008] In some embodiments, the segment has a sequence selected from the group consisting of SEQ ID NO:3 ; SEQ ID NO: 56; SEQ ID NO: 57and SEQ ID NO: 58 or a sequence at least 90% homologous to any of the foregoing.

[0009] In some embodiments, the vector further comprises an insert containing an exogenous polynucleotide encoding the polypeptide of interest, wherein upon delivery of the vector to a target cell, the polypeptide is expressed and substantially localized on the target cell surface.

[0010] In some embodiments, the first polynucleotide encoding the first polypeptide of interest is or is encoded by an open reading frame (ORF) in turn encoding a polypeptide normally expressed on a cell surface; in other embodiments, the polynucleotide encodes a polypeptide that is not normally substantially localized on the target cell surface.

[0011] In some embodiments, the insert further comprises a polynucleotide encoding a signal peptide and a polynucleotide encoding a transmembrane domain derived from a polypeptide expressed on a cell surface.

[0012] In some embodiments, the insert also comprises a polynucleotide encoding a C-terminal of a polypeptide that is normally expressed on the target cell or a fragment thereof encoding a SET binding site.

[0013] In some embodiments, the transmembrane domain also includes a SET binding site.

[0014] In some embodiments, one or both of the signal peptide and transmembrane domain are derived from CD47.

[0015] In some embodiments, the transmembrane domain includes the entire transmembrane domain region of CD47.

[0016] In some embodiments, the polypeptide of interest is GFP.

[0017] In some embodiments, the first polypeptide of interest is selected from the group consisting of taste receptors and glutamate receptors.

[0018] In some embodiments, the first polypeptide of interest is selected from the group consisting of INSR, IGFR, TGFBR1, EPHB2, BMPR1A, EGFR, FZD5, CD44, and PDGFR.

[0019] In some embodiments, the insert comprises the polynucleotide encoding the signal peptide of CD47 followed by the polynucleotide encoding the polypeptide of interest followed by the transmembrane domain of CD47 including the C-terminal thereof and said insert is 5' to the 3'UTR.

[0020] In some embodiments, the promoter is selected from the group consisting of ubiquitous promoters or tissue-specific promoters.

[0021] In some embodiments, the vector is selected from the group consisting of plasmids and viral vectors.

[0022] In some embodiments, the vector is a viral vector selected from the group consisting of selected from the group consisting of an adenoviral vector, an adeno associated vector (AAV), a baculoviral vector and a retroviral vector.

[0023] In some embodiments, the promoter is part of an expression cassette comprised by the vector.

[0024] In some embodiments, the fragment is preceded by a CA-rich motif or a GA-rich motif or an AU-rich sequence interposed between the fragment and a first (5') restriction site.

[0025] In another aspect, the disclosure relates to a recombinant polynucleotide construct comprising (i) a first segment having a sequence encoding a polypeptide that is not normally substantially localized on the surface of a target cell, (ii) a second heterologous segment having a sequence that comprises a fragment of a 3'UTR comprising an HuR binding site or nucleic acid encoding said fragment, wherein said construct has the property of effecting

localization of at least a substantial fraction of said polypeptide to the cell surface upon expression of the polypeptide in a cell harboring the construct.

[0026] In some construct embodiments, the polypeptide comprises a transmembrane domain sequence and a peptide signal sequence of a membrane-associated polypeptide.

[0027] In some construct embodiments, the transmembrane domain includes the C-terminal of said membrane-associated polypeptide or at least a fragment of said C-terminal encoding a SET binding site.

[0028] In some construct embodiments, the 3'UTR fragment has a sequence comprising a nucleic acid having SEQ ID NO: 3 or SEQ ID NO: 56 or SEQ ID NO: 57 or SEQ ID NO: 58 or a sequence having at least 90% homology to any of the foregoing.

[0029] In another aspect the present disclosure relates to a polynucleotide construct comprising the sequence SEQ ID NO 3, SEQ ID NO 56, SEQ ID NO 57 and SEQ ID NO: 58 or a sequence having at least 90% homology to any of the foregoing flanked between two restriction sites.

[0030] In some embodiments, the construct comprises at least one of a CA-rich motif, a GA-rich motif and an AU-rich motif following the first restriction site.

[0031] In still another aspect, the present disclosure is directed to a method for increasing a fraction of a polypeptide of interest expressed within a cell that is localized on the surface of the cell, comprising transfecting the cell with one of the foregoing vectors under conditions that permit expression of the polypeptide, wherein HuR, SET and RAC1 are present within the cell, have access to the polypeptide at least post-translationally, and effect localization of the polypeptide to the cell surface.

[0032] In a further aspect, the present disclosure relates to methods for increasing a fraction of a polypeptide expressed within a cell that is transported to the surface of the cell, comprising introducing in the cell a construct described above in this section under conditions that permit expression of the polypeptide, wherein HuR, SET and RAC1 are present within the cell, have access to the polypeptide at least post-translationally, and effect localization of the polypeptide to the cell surface.

BRIEF DESCRIPTION OF THE DRAWINGS

[0033] **Figure 1. The long 3' UTR of *CD47* localizes GFP-TM protein to the plasma membrane, whereas the short 3' UTR localizes it to the endoplasmic reticulum. a**, Fluorescence confocal microscopy of endogenous CD47 protein in non-permeabilized (top)

and permeabilized (bottom) cells. IC, intracellular. **b**, 3'-seq analysis of naive B cells shows two 3'UTR isoforms of *CD47* mRNA (short 3'UTR (*SU*) and long 3'UTR (*LU*)). Shown is the last exon of the gene model CD47 with 5 transmembrane domains. Isoform abundance shown in transcripts per million (TPM). **c**, Northern blot analysis of human cell lines confirming *CD47* mRNA isoforms from **b**. The corresponding ethidium-bromide-stained RNA gel is shown as loading control. **d**, Staining of U2OS cells as in **a** after transfection of a control shRNA (sh Co) or an shRNA against the long *CD47* 3'UTR isoform. **e**, CD47 protein contains an N-terminal signal peptide, ECD, five TMDs and a cytoplasmic C terminus. In both constructs, the ECD was replaced with GFP and either fused with the long (GFP-TM-LU) or the short *CD47* 3'UTR (GFP-TM-SU). Constructs are drawn to scale. **f**, Fluorescence confocal microscopy of fixed U2OS cells after transfection of GFP-TM-LU or GFP-TM-SU. Bottom, with additional staining of the endoplasmic reticulum with anti-calnexin. **g**, FACS analysis of GFP expression in transfected U2OS cells with (black lines, detection of total expression) and without permeabilization (dark grey lines (top left and bottom right), detection of surface (surf.) expression). Values for mean fluorescence intensity (MFI) are shown in parentheses. Unstained cells are shown in light grey. Representative image from more than 20 experiments. **h**, RNA-fluorescence *in situ* hybridization (FISH) against *GFP* in permeabilized U2OS cells after transfection of GFP-TM-LU or GFP-TM-SU. Bottom panel also shows GPF-TM protein. **a**, **d**, **f** and **h** are representative images from hundreds of cells. Scale bars, 10 μm.

[0034] Figure 2. 3'UTR-dependent protein localization (UDPL) depends on HuR, SET and RAC1, and mediates surface localization of membrane proteins. a, Model of UDPL. HuR binds to the long 3' UTR and recruits SET. During translation of *CD47* mRNA, this protein complex is targeted to the endoplasmic reticulum (ER) surface where SET binds to the newly translated cytoplasmic domains of CD47. This step probably requires energy. SET interacts with RAC1 and active RAC1 translocates SET and CD47 to the plasma membrane. **b**, FACS analysis of endogenous CD47 protein expression in HEK293 cells. Left panel is after transfection of control shRNA (shCo) or shRNAs against HuR (shown are all GFP^+ cells). Middle and right panels depict cells stably expressing the indicated shRNAs. Surface CD47 (top) and total CD47 protein (bottom) were measured. **c**, 3'-seq analysis for *CD44* in HEK293 cells and for *TNFRSF13C* in B-LCL cells, as shown in Fig. 1b. FACS analysis of endogenous CD44 protein in HEK293 cells (left) and endogenous BAFFR protein in SHSY-5Y cells (right) shown as in **b** (left). **d**, Left, FACS analysis of GFP after

transfection of constructs containing a signal peptide and GFP fused to the TMD and C terminus of CD44 and either the long 3' UTR (black) or the short 3' UTR (light grey) in U251 cells. Right, as in left panel, but for BAFFR with transfection into HeLa cells. e, FACS analysis of GFP expression shows that HuR-BS is sufficient for surface localization of GFP-TM (black). Deletion of the binding sites from the HuR-BS construct (HuR-BSΔ) abrogates GFP surface expression (light grey). For **b**, **c**, **d** and **e**, surface and total protein expression were determined and shown as in Fig. 1g. Representative images from three (sh2 HuR, n=5) biological replicates.

[0035] Figure 3. Mechanism of UDPL. a, Left, RNA-immunoprecipitation after transfection of the indicated constructs into HEK293 cells. a. Protein–RNA complexes were pulled-down with anti-HuR antibody and *GFP* abundance was normalized to *GAPDH* and is shown as fraction of input. Right, RNA co-immunoprecipitation after transfection of the indicated shRNAs. Protein–RNA complexes were pulled down with anti-SET antibody and the abundance of *CD47-LU* was normalized to *GAPDH* and is shown as fraction of input. Shown is mean ± standard deviation (s.d.), n=3 biological replicates. ***P < 0.0003, **P < 0.002, NS, not significant (P>0.05), two-sided *t*-test for independent samples. **b,** MS2-binding sites (see Fig. 9) were added to GFP-TM-SU and co-transfected with MS2-m (left), MS2-mC-HuR (middle) or MS2-mC-SET (right). mC, mCherry. Fluorescence confocal microscopy of HEK293 cells after transfection of indicated constructs shows that recruitment of HuR or SET to the short 3'UTR redirects localization of GFP protein from the endoplasmic reticulum to the cell surface. Representative images from hundreds of cells. Scale bars, 10 μm. **c**, Co-immunoprecipitation of endogenous SET using anti-GFP antibody after transfection of CD47-SU or CD47-LU in HEK293 cells (for constructs, see Fig. 4a). Two percent of input was loaded. WB, western blot. **d**, FACS analysis of surface GFP expression after transfection of GFP-TM-LU (black), GFP-TM-LU with a C-terminal deletion (ΔC; light grey line; left) or with destruction of both SET-binding sites (ΔC and K163A, K166A, K175A; ΔCL; light grey line, right). Shown as in Fig. 1g. Representative image from more than four experiments. **e**, Co-immunoprecipitation of endogenous SET using anti-GFP antibody after transfection of the indicated constructs. Two percent of input was loaded. Asterisk indicates unspecific band.

[0036] Figure 4. CD47 protein has different functions depending on whether it was generated by the *SU* or *LU* isoform. a. To generate GFP–CD47, GFP was inserted in frame

between the signal peptide and the rest of the CD47 open reading frame. GFP–CD47 was fused with either the long or short *CD47* 3'UTR, called CD47-LU and CD47-SU, respectively. **b**, FACS analysis of surface (surf.; light grey) and total (black) GFP–CD47 expression in transfected JinB8 cells. Shown as in Fig. 1g. Representative images from four experiments. **c**, FACS analysis of GFP expression after transfection of CD47- LU or CD47-SU with or without co-transfection of dominant-negative RAC1 (N17RAC1). Shown as in Fig. 1g. Representative images from n=7 (LU) and n=2 (SU) experiments. **d**, Fraction of Mitomycin-C-treated cells that survived at day (d) 3 after co-culture with macrophages is displayed for Jurkat, JinB8 (*CD47*$^{-/-}$) and the GFP^{+} JinB8 cells after nucleofection of CD47-LU or CD47-SU. Shown is mean ± s.d., n=3 biological replicates. **P<0.005, *P<0.02, NS, not significant (P>0.05), two-sided *t*-test for independent samples. **e**, The fraction of surviving cells (TO-PRO3 negative) measured by FACS analysis at day 3 after γ-irradiation is shown for the same populations as in **d**. Shown is mean ± s.d., n=3 biological replicates of the 20% of cells with the highest GFP expression. Gy, Gray. **f,** Fluorescence confocal microscopy of permeabilized U251 cells after transfection of CD47-LU or CD47-SU co-stained with anti-RAC1 antibody. Representative images from hundreds of cells. Scale bars, 10 μm. **g,** Immunoprecipitation of endogenous RAC1–GTP (active RAC1) in HEK293 cells after transfection of CD47-LU, CD47-SU, or empty vector. Total RAC1 and GFP–CD47 were measured from input. n=3 biological replicates.

[0037] Figure 5. Expression of the long CD47 3' UTR isoform correlates with cell surface expression of CD47 protein. a, Fluorescence confocal microscopy of cells shown as in Fig. 1a. Representative images out of hundreds of cells are shown. Scale bars, 10 μm. **b**, FACS analysis of endogenous CD47 expression in cells shown in Fig. 1a and **a**. Permeabilized cells show total CD47 expression (black) and non-permeabilized cells show surface CD47 expression (dark grey). Representative histograms are shown (HEK293 cells, n=10; U2OS, Jurkat cells, n=5; Caov-3, n=3). Unstained cells are shown in light grey. **c**, Left, quantification of mean fluorescence intensity (MFI) values from **b**. Right, fraction of surface and intracellular CD47 levels in cells lines from **b**. Intracellular CD47 was calculated by subtracting CD47 surface values from total CD47 values. **d**, Northern blot of HEK293 cells stably expressing the indicated shRNAs and hybridized for *CD47*. The shRNAs against CD47-LU target only the long 3'UTR isoforms of *CD47*. The blot and corresponding RNA gel are shown as in Fig. 1c. **e**, Quantification of *CD47* total mRNA and 3'UTR isoform levels in U2OS cells by qRT–PCR. *GAPDH*-normalized values after transfection of sh2 CD47-LU

or sh Co are shown as the mean ± s.d., n=3 biological replicates. The total amount of *CD47* mRNA after transfection of sh Co was set to 1. **f**, FACS analysis of endogenous CD47 protein expression after stable expression of shRNAs against CD47-LU in HEK293 cells. Surface (top) and total (bottom) CD47 expression is shown. Representative histograms out of n=3 experiments are shown. Unstained cells are shown in dark grey. **g**, Quantification of MFI values from **f** is displayed. Intracellular CD47 was calculated as in **b**. **h**, Northern blot of HEK293 cells after transfection of indicated constructs and hybridized against *CD47*. Mutation of the proximal polyadenylation signal in *CD47-LU* abrogates production of short 3'UTR isoforms. Asterisk indicates cross-hybridization to ribosomal RNAs. **i**, Fluorescence confocal microscopy of endogenousCD47 and calnexin protein in permeabilized U2OS cells. Calnexin partially co-localizes with CD47. A representative image out of hundreds of cells is shown. Scale bars, 10 μm.

[0038] Figure 6. UDPL depends on HuR, SET and RAC1 and mediates surface localization of membrane proteins. a, Western blot of HEK293 cells transiently transfected (left, middle) or stably expressing (right) sh Co or shRNAs against HuR. The blot shows reduced HuR protein expression after HuR knockdown, but no change in protein expression of CD47, TSPAN13, CD44 or SET. Actin was used as loading control. **b**, Quantification of *CD47* total mRNA and 3'UTR isoform levels in HEK293 cells by qRT–PCR. GAPDH-normalized values after transfection of sh2 HuR or sh Co are shown. Shown is the mean ± s.d., n=3 biological replicates. The total amount of *CD47* mRNA after transfection of sh Co was set to 1. **c**, FACS analysis of HEK293 cells stably expressing the indicated shRNAs. Histograms are shown as in Fig. 2b. Representative histograms from n=3 experiments are shown. **d**, Western blot of HEK293 cells stably expressing shRNAs against SET. Actin was used as loading control. The marker is shown in kDa. **e**, As in **d**, but HEK293 cells stably expressing shRNAs against RAC1 are shown. **f**, 3'-seq analysis shows 3'UTR isoform expression of *ITGA1* in B-LCL and *TSPAN13* in HEK293 cells shown as in Fig. 1b. FACS analysis of endogenous protein levels is shown as in Fig. 2c. Left panel shows ITGA1 expression in HeLa cells and right panel shows TSPAN13 expression in HEK293 cells. Representative histograms from n=2 experiments are shown. **g**, FACS analysis of GFP after transfection of constructs containing a signal peptide and GFP fused to the TMD, C terminus and either the long 3'UTR (dark blue line) or the short 3'UTR (light blue line) of *ITGA1* in HEK293 cells. Representative histograms from n=3 experiments are shown as in Fig. 2d.

[0039] Figure 7. 3'UTR isoforms that encode proteins using UDPL contain uridine-rich elements. Shown are the 3'UTR sequences of *CD47*, *CD44*, HuR-BS and HuR-BSΔ. Dark grey, ApA signals. Light grey, uridine-rich elements with the potential to be HuR-binding sites.

[0040] Figure 8. 3'UTR isoforms that encode proteins using UDPL contain uridine-rich elements. Shown are the 3'UTR sequences of *ITGA1*, *TNFRSF13C* and *TSPAN13*. Dark grey, ApA signals. Light grey, uridine-rich elements with the potential to be HuR-binding sites.

[0041] Figure 9. Local recruitment of SET to the site of translation is required for UDPL. a, Western blot of cells used in Fig. 3b shows the amount of overexpression achieved by transfection of MS2-mC-SET or MS2-mC-HuR (for constructs, see **b**). Left, anti-SET detects endogenous expression of SET as well as overexpressed SET. Right, anti-HuR detects endogenous HuR and overexpressed HuR. Actin was used as loading control. Anti-HuR and anti-SET were used on the same blot. Actin as loading control was performed once. The marker is shown in kDa. Asterisk indicates unspecific band. mC, mCherry. **b**, The top construct depicts GFP-TM-SU (Fig. 1e) and the bottom construct shows a fusion of MS2 coat protein (MS2), mC (red) and HuR or SET, respectively. Overexpression of HuR or SET compared with expression of MS2- mC alone does not change surface or total GFP expression, when co-transfected with GFP-TM-SU (without the addition of MS2-binding sites to the *SU* isoform) as shown by FACS analysis. Surface expression (top) and total expression (bottom) in HEK293 cells are shown. Values for MFI are shown in parentheses. Unstained cells are shown in grey. Representative histograms from n=2 experiments are shown. **c**, FACS analysis of cells used in Fig. 3b. MS2-binding sites (MS2-BS, RNA stem loops) were added to GFP-TM-SU (and the proximal polyadenylation signal was mutated) to obtain GFP-TM-SU-MS2-BS. Transfection of MS2-mC-HuR (left, black line) or MS2-mC-SET (right, black line) increases surface GFP expression compared with transfection of MS2-mC (dark grey line), when GFP-TM-SU-MS2-BS is co-transfected. Thus, tethering of HuR or SET to the short 3'UTR of *GFPTM* localizes GFP to the cell surface without changing total GFP expression. Histograms are shown as in **b**. Representative histograms from n=5 experiments are shown. **d**, As in **c**, but tethering was impaired by omission of the MS2 coat protein. Histograms are shown as in **b**. Representative histograms from n=2 experiments are shown. Summary of the tethering experiment: To tether SET or HuR to the 3'UTR (which

brings it close to the site of translation through the scaffold function of the 3'UTR), we added MS2-binding sites to GFP-TM-SU **c**, MS2-binding sites are derived from the bacteriophage MS2 and form RNA stem loops. The capsid protein of MS2 (here, called MS2) specifically recognizes these MS2 stem loops. Constructs were generated containing MS2 fused to mC and then either HuR, SET or with no further coding sequence (Fig. 3b). Co-expression of these constructs with the construct containing the short UTR and MS2-binding sites results in recruitment of SET or HuR to the short 3'UTR of *GFP-TM*. The cells that express MS2 fused to only mC localize GFP to the endoplasmic reticulum, but constructs containing MS2 fusions to HuR or SET localize GFP primarily to the cell surface (Fig. 3b and Fig. 9c). Omitting either MS2 or the MS2-binding sites from the experiment abrogates surface localization (Fig. 9b, d).

[0042] Figure 10. CD47 contains at least two SET-binding sites in its cytoplasmic domains. a, FACS analysis of surface GFP expression after transfection of GFP-TM-LU (black line) and GFP-TM-LU constructs containing a single point mutation in the cytoplasmic C terminus (light grey line, adjacent to black line), K290A (left), K297A (middle), K304A (right) in HEK293 cells. Partial destruction of a single SET-binding site results in up to 37% reduction in GFP surface expression. Values for MFI are shown in parentheses. Unstained cells are shown in dark grey. Representative histograms from n=5 experiments are shown. **b**, FACS analysis of GFP expression after transfection of GFP-TM-LU (black line), GFP-TM-LU containing a mutation of the SET-binding site in the C terminus (K290A,K304A; 2Km; light grey middle line; left), containing a deletion of the C terminus (ΔC; light grey middle line; middle panel), or destruction of both SET-binding sites (ΔC combined with K163A, K166A, K175A; ΔCL; light grey middle line; right). Surface (top) and total (bottom) expression is shown in HEK293 cells. Values for MFI are shown in parentheses. Unstained cells are shown in dark grey. Representative histograms from several experiments are shown (2Km, n=3; ΔC, n=10; ΔCL, n=4).

[0043] Figure 11. CD47 protein has different functions depending on whether it was generated by the *SU* or *LU* isoform. a, Left, western blot of HEK293 cells after transfection of the indicated constructs shows GFP–CD47 expression using an anti-GFP antibody. Actin was used as loading control. Right, as in left panel after transfection of CD47-SU and CD47-LU into HEK293 (left) or JinB8 cells (right). GFP–CD47 expression was quantified after normalization with respect to actin using Image J. Shown is the fold change in GFP–CD47

expression of CD47-SU after setting CD47-LU to 1. **b**, The experiment is similar to Fig. 3b and Fig. 9c, but here the constructs containing the full open reading frame of CD47 were used. FACS analysis of GFP expression after transfection of CD47-SU-MS2-BS. Co-transfection of MS2-mC-HuR (left, black line) or MS2-mC-SET (right, black line) increases surface GFP expression compared to co-transfection of MS2-mC (grey line (middle peak)). Surface expression is shown in non-permeabilized HEK293 cells. Values for MFI are shown in parentheses. Representative histograms from n=3 experiments are shown. Unstained cells are shown in dark grey. **c**, Left, FACS analysis of GFP after transfection of constructs containing a signal peptide and GFP fused to the open reading frame of BAFFR and either the long 3'UTR (BAFFR-LU, black line) or the short 3'UTR (BAFFR-SU, light grey line adjacent to black line) in HEK293 cells. Surface (top) and total (bottom) GFP expression is shown. Values for MFI are shown in parentheses. Representative histograms from n=3 experiments are shown. Unstained cells are shown in grey. Right, as in left panel but forCD44. **d**, Table showing the fold increase in surface GFP expression mediated by the *LU* isoform compared with the *SU* isoform. Top row shows values of constructs without the ECD and bottom row shows values of constructs containing the full coding regions of the indicated proteins. The fold increase in surface GFP expression was calculated from MFI (LU)/MFI (SU). The contribution of the ECD domain for surface expression of BAFFR is 1.2-fold (3.8/3.1). **e**, FACS analysis of carboxyfluorescein succinimidyl ester (CFSE) uptake in macrophages. Macrophages were co-cultured without (grey) cells or with cells that were pretreated with CFSE and expressed high or low amounts of surface CD47 (data not shown). The experiment shows that the macrophages phagocytose the cells depending on their CD47 surface expression levels. A representative histogram from n=2 experiments is shown. **f**, The fraction of surviving cells (TO-PRO3 negative) as measured by FACS analysis at day 3 (d3) after increasing doses of γ-irradiation is shown for Jurkat, JinB8 (*CD47*$^{-/-}$) and the GFP^{+} fraction after nucleofection of JinB8 cells with either CD47-SU or CD47-LU. The values were obtained from the same experiment as shown in Fig. 4e, but here the values were calculated using all GFP-positive cells. Shown are the values for mean ± s.d., n=3 biological replicates. Gy, Gray. **g**, The fraction of surviving cells (TO-PRO3 negative) as measured by FACS analysis at day 3 (d3) after increasing doses of γ-irradiation is shown for Jurkat, JinB8 (*CD47*$^{-/-}$) and the GFP^{+} fraction after nucleofection of Jurkat cells with sh2 CD47-LU. Shown are the 20% of cells with the highest GFP expression (light grey middle line). Shown are the values for mean ± s.d., n=3 biological replicates. Gy, Gray.

[0044] Figure 12. HuR, SET and RAC1 are widely and highly expressed. a, The mRNAs of proteins necessary for UDPL are ubiquitously and highly expressed across cell lines (left) and tissues (right). Shown are values for transcripts per million (TPM). The median abundance levels of all expressed genes in the data sets are shown as dashed lines. *ELAVL1* encodes HuR. The data set from Lianoglou et al. was analyzed to obtain the TPM values (Lianoglou et al. *Genes Dev.*2013 Nov 1;27(21):2380-96). **b**, Here, 'HuR targets' consist of the union of HuR targets identified previously (Lebedeva et al. *Mol. Cell* 2011 Aug 5;43(3):340-52; Uren et al. *J Biol Chem.* 2011 Oct 28;286(43):37063-6). Membrane proteins consist of all the proteins that contain the tag ''membrane'' using gene ontology analysis. The fraction of membrane proteins found is consistent with the fraction of membrane proteins found in yeast (Stagljar et al. *Trends Biochem Sci.* 2002 Nov;27(11):559-63). Fisher's exact test shows no enrichment or depletion of membrane proteins among the HuR targets.

[0045] Figure 13. All tested UDPL candidates have potential SET binding sites in their cytoplasmic domains. Shown are the amino acid sequences of the TMDs and cytoplasmic domains of the membrane proteins studied. The TMDs are shown in grey.

[0046] Figure 14. UDPL sequence candidates to be tested. Shown are the nucleotide sequences of UTR fragments of indicated proteins that have or will be tested for their ability to facilitate the expression of proteins on the surface of the cell. 2x HuR-BS refers to a repeat of a 3'UTR fragment encompassing the HuR binding site and derived from CD47.

[0047] Figure 15. 3'UTR sequences that comprise an HuR- binding site of TNF alpha increase surface localization of proteins. Figure 15A shows the top three motifs that are enriched in the 3'UTRs of human membrane proteins. The motifs were determined using the search engine HOMER (homer.salk.edu). Figure 15B, 15C, and 15D show fold change in 3'UTR mediated increase in CD47 surface localization using UTR-1 (SEQ ID NO: 56, UTR-2 (SEQ ID NO: 57), and UTR-3 (SEQ ID NO: 58) sequences. Each 3'UTR was tested in HeLa (figure 15B), U2OS (figure 15C), and HEK293 (figure 15D) cells. Shown is a mean ± s.d. of biological replicates.

DETAILED DESCRIPTION

Definitions

[0048] As used herein, the following terms shall have the meanings ascribed to them below unless the context clearly indicates otherwise:

[0049] "Vector" generally is a DNA or RNA molecule, such as a plasmid, virus, phagemid, cosmid or other nucleic acid construct that (a) contains or is adapted to contain a heterologous nucleic acid segment and (b) serves as a vehicle for delivering a polynucleotide of interest to the interior of a target organism such as a target bacterium or target cell.

[0050] "Polypeptide" is used interchangeably with "protein" and means a polymer having a primary linear structure of about 100 or more amino acid residues linked one to the other in a chain by peptide bonds and forming part of (or the whole of) a molecule.

[0051] The term "fusion protein" refers to a polypeptide comprising a polypeptide or fragment thereof coupled via a peptide bond to one or more heterologous amino acid sequences.

[0052] "Transmembrane domain" or "TM" means a segment of a protein that spans the membrane bilayer of a cell and thus crosses the cell membrane. A protein may have more than one transmembrane domains, which together may constitute a transmembrane region. For example, CD47 has 5 transmembrane domains (Rebres et al. J Biol Chem. 2001;276:34607–34616). In the context of the present disclosure, "transmembrane domain addition" means the incorporation in a vector or polynucleotide construct of nucleic acid encoding one or more transmembrane domains of a membrane-associated polypeptide. The transmembrane domains may for example be separated by surface or cytoplasmic side segments of the original protein, such as for example a segment encompassing the binding site for SET.

[0053] "Operably linked" refers to a functional relationship between two or more polynucleotide (e.g., DNA) segments. Typically, the term refers to the functional relationship of a transcriptional regulatory sequence to a sequence to be transcribed. For example, a promoter or enhancer sequence is operably linked to a coding sequence if it stimulates or modulates the transcription of the coding sequence in an appropriate host cell or other expression system. Generally, promoter transcriptional regulatory sequences that are operably linked to a transcribable sequence are contiguous to the transcribable sequence, i.e., they are

cis-acting. However, some transcriptional regulatory sequences, such as enhancers, need not be physically contiguous or even located in close proximity to the coding sequences whose transcription they enhance.

[0054] "Promoter" refers to a sequence that regulates transcription of an operably- linked gene, or nucleotide sequence encoding a protein, etc. Promoters provide the sequence sufficient to direct transcription, as well as, the recognition sites for RNA polymerase and other transcription factors required for efficient transcription and can direct cell specific expression.

[0055] "Expression cassette" refers to a polynucleotide segment comprising a promoter and nucleic acid encoding a polypeptide to be expressed. The promoter is operably linked to the nucleic acid to be expressed. The cassette optionally also encompasses other elements that regulate transcription or translation such as enhancers, termination codons, polyadenylation site, IRES, selectable markers and other inducible elements.

[0056] "3'UTR" means the untranslated region of mRNA (or the transcribed corresponding region of DNA) that follows the translation termination codon. This region is known to contain regulatory segments that can influence polyadenylation, translation efficiency, as well as localization and stability of the mRNA. The 3'UTR is also known to contain binding sites for regulatory proteins and for microRNAs (miRNAs). Many 3'UTRs also contain AU-rich elements (AREs). Proteins bind AREs to affect the stability or decay rate of transcripts in a localized manner or to affect translation initiation. Furthermore, the 3'UTR contains the sequence AAUAAA that directs addition of several hundred adenine residues called the poly(A) tail to the end of the mRNA transcript. Poly(A) binding protein (PABP) binds to this tail, contributing to regulation of mRNA translation, stability, and export. The 3'UTR can also contain sequences that attract proteins to associate the mRNA with the cytoskeleton, transport it to or from the cell nucleus, or perform other types of RNA localization. In addition to sequences within the 3'UTR, the physical characteristics of the region, including its length and secondary structure, contribute to translation regulation.

[0057] "GFP" means green fluorescent protein, a protein originally isolated from the jellyfish Aequorea Victoria which fluoresces with a bright green color when exposed to light

in the blue to UV range, and similarly fluorescing proteins from other marine species such as the sea pansy.

[0058] "CD47" (Cluster of Differentiation 47) also known as integrin associated protein (IAP) is a transmembrane protein that in humans is encoded by the CD47 gene. CD47 is ubiquitously expressed in human cells but it is not normally particularly localized on the cell surface. CD47 belongs to the immunoglobulin superfamily, partners with membrane integrins and also binds the ligands thrombospondin-1 (TSP-1) and signal-regulatory protein alpha (SIRPα) which is expressed in phagocytic cells. CD47 is involved in a range of cellular processes, including apoptosis, proliferation, adhesion, and migration. It functions as a marker of self ("don't eat me signal") for phagocytic cells, and downregulates phagocytic activity. CD47 has also been implicated in processes such as, for example, apoptosis, survival, proliferation, adhesion, migration, and regulation of angiogenesis, blood pressure, tissue perfusion, and/or platelet homeostasis. CD47 inhibits dendritic cell (DC) maturation and activation. CD47 has been implicated in cancer (immune escape). CD47 is overexpressed in various hematological and solid malignancies. CD47 is a documented cancer stem cell/tumor initiating cell marker. It is thought that CD47 overexpression may help tumor cells to escape immune surveillance and killing by innate immune cells. High levels of CD47 are also associated with poor clinical outcome in cancers such as, for example, leukemias, lymphomas, breast cancer, colon cancer, ovarian cancer, bladder cancer, prostate cancer, and/or glioma. Because of its normally relatively poor expression on the cell surface, CD47 was arbitrarily chosen as a test protein by the present inventors to test the ability of the present methods to augment the proportion of a protein that is localized on the cell surface.

[0059] "HuR" or "human antigen R" denotes an RNA binding protein that modulates the stability and translational efficiency of mRNA.

[0060] "SET" or "protein SET" is a protein that in humans is encoded by the *SET* gene. SET contains a long acidic C-terminus that interacts with the RRM 3 (RNA recognition motif 3) of HuR. Nuclear SET binds to histone tails and prevents acetylation, but phosphorylated SET localizes to the cytoplasm and the surface of the endoplasmic reticulum. Furthermore, SET was shown to interact with Rac1, which results in translocation of SET to the plasma membrane upon Rac1 activation. Whereas Rac1-GDP is mostly localized to the cytoplasm, Rac1-GTP (active Rac1) localizes to the leading edge of the plasma membrane (ten Klooster

et al. EMBO J 26, 336-345 (2007)). In this disclosure, the inventors have demonstrated that in addition to HuR, SET and RAC1 are also necessary for 3' UTR-dependent protein localization (Example 3).

[0061] "RAC1" or "ras-related C3 botulinum toxin substrate 1," a protein encoded by the *RAC1* gene (a protooncogene), is a GTPase belonging to the RAS superfamily of small GTP-binding proteins controlling cell growth, cell motility, cytoskeletal organization and protein kinase activation. RaC1 regulates many cellular processes and is implicated in cancer metastasis. Certain activating RaC1 mutations have been associated with certain cancers. Orthologs exist in rodents, primates and many other species. There are differently spliced versions of this protein that have different functions.

[0062] "Substantially" when used in the context of protein localization on the cell surface means that a significant fraction of a protein being expressed is directed to the surface. Nonlimiting examples of such significant fraction are ½, ⅓, ¼, etc. Alternatively, substantiality of localization when using long 3'UTR or a fragment thereof containing an HuR binding site can be assessed by reference to the fraction of a polypeptide of interest that is directed to the cell surface without the use of such 3'UTR or fragment. Examples of a substantial increase in the fraction of a polypeptide directed to the cell surface include without limitation 1.2 or more the amount localized on the cell surface in the absence of the long 3'UTR or fragment.

EMBODIMENTS

[0063] The present invention is based in part on the discovery that different isoforms of 3'UTR act as scaffolds to regulate (enhance or curb) membrane protein cell surface localization.

Membrane Proteins and Other Proteins of Interest to be Expressed

[0064] The present disclosure provides compositions and methods for enhancing the cell surface localization of proteins, whether these are already membrane-associated proteins (in which case their localization is enhanced) or proteins which are not expressed on a cell's surface to any substantial extent but which are desired to be so expressed, such as G-protein

coupled receptor proteins (Dunham, J.H., et al, "Enhancement of the surface expression of G protein-coupled receptors" Trends in Biotechnology 27(9): 541 (2009)) and antigens to be presented on the surface of immune system cells.

[0065] "Membrane-associated polypeptide" or "membrane-associated protein" or "membrane protein" or "membrane polypeptide" means a polypeptide that is present on the membrane of a cell and includes without limitation membrane, transmembrane and cell surface polypeptides, receptor polypeptides, envelope polypeptides and the like. Membrane proteins are proteins which in their native state are associated with lipid membranes (e.g., nuclear membrane, cellular membrane, mitochondrial membranes, liposomal membranes, endoplasmic reticulum membranes, chloroplast membranes, etc.). The term "membrane proteins" includes transmembrane proteins, and proteins which are partially or fully embedded in membranes in their native state or simply residing on the surface of such membranes. The term "membrane proteins" includes both extrinsic and intrinsic proteins. Extrinsic membrane proteins are generally located entirely outside of the membrane, but are bound to the membrane by weak molecular attractions (such as, for example, ionic, hydrogen, and/or Van der Waals bonds). Intrinsic membrane proteins are, generally, embedded in the membrane. Intrinsic membrane proteins include, but are not limited to proteins which extend from one side of the membrane to the other, i.e., transmembrane proteins.

[0066] In one embodiment, incorporation of the 74 nt fragment of the long 3'UTR of CD47 (SEQ ID NO: 3) in a construct or vector according to the present disclosure comprising a polypeptide of interest will result in substantial localization of the polypeptide to the cell surface without needing to incorporate the entire long 3'UTR region of CD47.

[0067] In another embodiment, incorporation of the 53 nt artificial fragment (SEQ ID NO: 56) in a construct or vector according to the present disclosure comprising a polypeptide of interest will result in substantial localization of the polypeptide to the cell surface.

[0068] In another embodiment, incorporation of the 117 nt artificial fragment (SEQ ID NO: 57) in a construct or vector according to the present disclosure comprising a polypeptide of interest will result in substantial localization of the polypeptide to the cell surface.

[0069] In yet further embodiment, incorporation of the 384 nt artificial fragment (SEQ ID NO: 58) in a construct or vector according to the present disclosure comprising a polypeptide of interest will result in substantial localization of the polypeptide to the cell surface.

[0070] In one embodiment, the present disclosure relates to vectors that facilitate increased expression of a protein (more accurately, expression of a higher proportion of a protein) on the surface of a cell. Such proteins can include any protein, whether it can penetrate the lipid bilayer, or reside on a surface layer of the cell membrane. Nonlimiting examples include: receptors of enzymes, peptide hormones, and local hormones, sugar acceptors/carriers, and cell membrane antigens. Examples of the membrane protein include, but are not limited to, receptors for growth factors (e.g., vascular endothelial growth factor (VEGF), transforming growth factor (TGF), fibroblast growth factor (FF), platelet derived growth factor (PDGF), insulin-like growth factor), insulin receptor, MHC proteins (e.g. class I MHC and class II MHC protein), CD3 receptor, T cell receptors, glutamate receptors, taste receptors, cytokine receptors (e.g., interleukin (IL)-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15 receptors), tyrosine-kinase-associated receptors such as Src, Yes, Fgr, Lck, Flt, Lyn, Hck, and Blk, and G-protein coupled receptors such as receptors for the hormone relaxin (LGR7 and LGR8) and chemokine receptor (e.g., CCR1, CCR2, CCR3, CCR4, CCR5, CCR6, CCR7, CCR8, CCR9, CCR10, CCR11, CXCR1, CXCR2, CXCR3, CXCR4, CXCR5, CXCR6, CX3CR1, and XCR1), PSGR, beta1-adrenergic receptor, beta2-adrenergic receptor, parathyroid hormone (PTH) receptor, EDG1, G10d, opioid receptors, neutensin receptors, neuropeptide Y (NPY) receptors, melatonin receptor, adrenocorticotrophin (ACTH) receptor, leukotriene receptors, 5-hydroxytryptamine (5-HT, or serotonin) receptors (including 5-TH1a, 5-TH1b, 5-TH1d, 5-TH1e, 5-TH1f, 5-TH2a, 5-TH2b, 5-TH2c, 5-TH3, 5-TH4, 5-TH5 receptors), FSH receptor, LH/hCG receptor, TSH receptor, formyl-methionyl peptide (fMLP) receptors (FPR1 receptor, FPR2 receptor, FPR3 receptor), GABA receptors, endothelin (ET) receptors, sarafotoxin receptors, dopamine receptors (D1 receptor, D2 receptor, D3 receptor, D4 receptor, D5 receptor), cannabinoid receptor, CGRP1 receptor, CGRP2 receptor, amylin receptor, calcitonin receptor, C5a receptor, bradykinin receptors (B1, B2 and B3 receptors), bombesin Receptors (BB1, BB2 and BB3 receptors), angiotensin receptors (AT1 a, AT1 b and AT2 receptors), adrenaline and noradrenaline receptors (or adrenoceptors), acetylcholine Receptors, platelet activating factor (PAF) receptor, prostanoid receptors (DP, EP1, EP2, EP3, FP, IP and TP receptors), somatostatin receptors (SS1, SS2, SS3, SS4 receptors), tachykinin Receptors (NK1, NK2 and

NK3 receptors), thrombin receptor, vasoactive Intestinal polypeptide (VIP) receptor family (GRF, PACAP, secretin, VIP and helodermin-preferring receptors), stresscopin receptor, urocortin receptor, corticotropin releasing hormone receptors, LGR4, LGR5, LGR6, LGR7 and LGR8; and GPCR orphan receptors designated as GPR1-58 at GenBank.

[0071] In other more specific embodiments, the present disclosure relates to vectors that facilitate increased localization of G- protein coupled receptors (GPCRs) on the surface of the cell. GPCRs constitute the largest family of transmembrane proteins and play an important part in signal transduction by converting extracellular stimuli including light, smells, neurotransmitters and hormones, into intracellular signals. Given the widespread role of GPCRs in biological processes, a large fraction of therapeutic agents act on GPCRs. However, GPCRs exhibit low and hence not substantial natural expression on the cell surface. Thus, an expression system that permits higher localization of proteins of interest on the cell surface would facilitate the testing of putative drugs and hence the drug discovery process.

[0072] In other embodiments, the present disclosure relates to vectors that facilitate increased localization of receptor and other surface proteins such as INSR, IGFR, TGFBR1, EPHB2, BMPR1A, EGFR, FZD5, CD44, and PDGFR.

[0073] In yet other embodiments, the vectors, constructs and methods of the present disclosure facilitate the study of membrane proteins. For example, the vectors, constructs and methods of the present disclosure can be used to determine which membrane proteins completely depend on their 3'UTRs and which membrane proteins only partially, or not at all depend on their 3'UTRs for surface expression.

[0074] Additionally, the vectors, constructs, and methods of the present disclosure can be used in drug discovery for identifying agents that can modulate the activity of membrane proteins.

[0075] In one embodiment, the vectors, constructs, and methods of the present disclosure can be used in immunotherapy. For example, vectors, constructs, and methods of the present disclosure can be used for expressing a chimeric antigen receptor (CAR) on the surface of a T cell. CARs are artificial molecules that, when present at the surface of immune effector cells, enable them to recognize a desired protein, or antigen, and trigger the killing of cells

harboring this antigen at their surface (target cells). Primary T cells can be isolated and engineered to express CARs—receptors that combine an extracellular, single-chain antibody domain, which recognizes a specific tumor-associated antigen, with intracellular signaling domains from the T cell receptor (TCR) and costimulatory receptors. Upon antigen ligand engagement, CAR T cells execute multiple key therapeutic functions, including production of antitumor cytokines and killing of target tumor cells. Thus, use of specific 3'UTR sequences of the present disclosure can promote and/or enhance sustained surface expression and localization of CARs on T cells.

[0076] T cells require two signals to become fully activated. A first signal, which is antigen-specific, is provided through the T cell receptor which interacts with peptide-MHC molecules on the membrane of antigen presenting cells (APC). A second signal, the co-stimulatory signal, is antigen nonspecific and is provided by the interaction between co-stimulatory molecules expressed on the membrane of APC and the T cell. Co-stimulatory molecules are a heterogeneous group of cell surface molecules that act to amplify or counteract the initial activating signals provided to T cells from the T cell receptor following its interaction with an antigen/major histocompatibility complex. Thus, vectors, constructs, and methods of the present disclosure can be used to increase the cell surface expression of co-stimulatory molecules, which could lead to an overall increase of the immune response. In another embodiment, the vectors, constructs and methods of the present disclosure facilitate the localization on the cell surface of proteins fused to a fluorescent protein as a detectable marker. Fusion proteins are useful because they can be constructed to contain two or more desired functional elements that can be from two or more different proteins. Fusions that include larger polypeptides, such as an IgG Fc region, and even entire proteins, such as the green fluorescent protein ("GFP") chromophore-containing proteins, have particular utility. Fusion proteins can be produced recombinantly by constructing a nucleic acid sequence which encodes the polypeptide or a fragment thereof in frame with a nucleic acid sequence encoding a different protein or peptide and then expressing the fusion protein.

[0077] In one embodiment, the short 3'UTR appropriately linked (functionally tethered) to either nucleic acid encoding HuR or nucleic acid encoding SET serves as a surrogate for the long 3'UTR which by containing an HuR binding site recruits HuR in turn recruiting SET and drives the expressed polypeptide of interest to the cell surface. In one embodiment, a

polynucleotide sufficient to facilitate localization of membrane proteins to the cell surface is 74- nucleotide sequence SEQ ID NO 3.

[0078] In another embodiment, the sequence sufficient to facilitate localization of membrane proteins to the cell surface is at least 90% or at least 92% or at least 95% or at least 98% homologous (i.e., having the foregoing percent sequence identity) to the 74-nucleotide sequence SEQ ID NO 3. Percent sequence identity between nucleic acid sequences can be determined using FASTA with its default parameters (a word size of 6 and the NOP AM factor for the scoring matrix) or using Gap with its default parameters as provided in GCG Version 6.1. Alternatively, sequences can be compared using the computer program, BLAST (Altschul et al., J. Mol. Biol. 215:403-410 (1990); Gish and States, Nature Genet. 3:266-272 (1993); Madden et al, Meth. Enzymol. 266: 131-141 (1996); Altschul et al, Nucleic Acids Res. 25:3389-3402 (1997); Zhang and Madden, Genome Res. 7:649-656 (1997)), especially blastp or blastn (Altschul et al, Nucleic Acids Res. 25:3389-3402 (1997)).

[0079] In another embodiment, one or more uridine nucleotides within the 74-nucleotide (SEQ ID NO. 3) can be replaced by uridine analogs, wherein uridine analog refers to a compound which bares structural similarity with uridine, or a metabolic precursor thereof. Uridine analogs according to the invention include, but are not limited to: uracil, uridine monophosphate, uridine biphosphate, uridine triphosphate; thymine, thymidine, thymidine monophosphate, thymidine biphosphate, thymidine triphosphate; β-D-uridine-5'-bis(SATE)phosphodiester, a prodrug of β-D-uridine-5'-monophosphate; acylated uridine nucleosides, triacetyluridine (TAU), 5-(phenylthio)acyclouridine (PTAU), or any other uridine analog which can be used for substitution of uridine.

[0080] In one embodiment, a polynucleotide sufficient to facilitate localization of membrane proteins to the cell surface is a 53-nucleotide artificial sequence SEQ ID NO: 56. In another embodiment, a sequence sufficient to facilitate localization of membrane proteins to the cell surface is at least 90% or at least 92% or at least 95% or at least 98% homologous (i.e., having the foregoing percent sequence identity) to the 53- nucleotide sequence SEQ ID NO 56.

[0081] In one embodiment, a polynucleotide sufficient to facilitate localization of membrane proteins to the cell surface is a 117-nucleotide artificial sequence SEQ ID NO: 57. In another embodiment, the sequence sufficient to facilitate localization of membrane

proteins to the cell surface is at least 90% or at least 92% or at least 95% or at least 98% homologous (i.e., having the foregoing percent sequence identity) to the 117- nucleotide sequence SEQ ID NO 57.

[0082] In one embodiment, a polynucleotide sufficient to facilitate localization of membrane proteins to the cell surface is a 384-nucleotide artificial sequence SEQ ID NO: 58. In another embodiment, the sequence sufficient to facilitate localization of membrane proteins to the cell surface is at least 90% or at least 92% or at least 95% or at least 98% homologous (i.e., having the foregoing percent sequence identity) to the 384- nucleotide sequence SEQ ID NO 58.

[0083] In one embodiment, the polynucleotide sufficient to facilitate localization of membrane proteins to the cell comprises a sequence comprising an HuR binding site from TNF-α: TTGTGATTATTTATTATTTATTTATTATTTATTTATTTA (SEQ ID NO: 59).

[0084] In another embodiment, the polynucleotide sufficient to facilitate localization of membrane proteins to the cell comprises a sequence comprising an HuR binding site from TNF-α (SEQ ID NO: 59) and one or more sequence elements, including but not limited to restriction site sequences, sequences found in the long 3'UTR of *CD47,* and sequence motifs enriched in the 3'UTRs of membrane proteins.

Nonmembrane-associated Proteins

[0085] In some embodiments, the compositions and methods of the present disclosure facilitate expression of non-membrane proteins on the surface of the cell. In one embodiment, non-membrane proteins can include proteins that are routinely used as markers in molecular biology. The most frequently used markers are fluorescent proteins and epitope tags. Fluorescent markers such as green fluorescent protein (GFP), yellow fluorescent protein (YFP), mCherry, and cyan fluorescent protein (CFP) enable the study of cellular processes by fluorescent microscopy.

[0086] In some embodiments, the compositions and methods of the present disclosure facilitate the expression of immunogenic portions of antigens from foreign organisms or other pathogenic agents (e.g., cancerous cells) on the cell surface. Antigens generally include, but are not limited to, proteins, polypeptides, peptides, polysaccharides such as glycans, polysaccharide conjugates, peptide and non-peptide mimics of polysaccharides and other

molecules, small molecules, lipids, glycolipids, and carbohydrates. Antigens which can be localized on the cell surface using the methods and compositions of the present disclosure will generally comprise a peptide or polypeptide component or moiety. The expression of antigens on the cell surface can directly or indirectly stimulate one or more components of the immune response.

[0087] In some embodiments, the compositions and methods of the present disclosure facilitate expression of secretory proteins on the surface of a cell. Among such protein hormones are cytokines, lymphokines, neurotrophic hormones and adenohypophyseal polypeptide hormones such as growth hormone, prolactin, placental lactogen, luteinizing hormone, follicle-stimulating hormone, thyrotropin, chorionic gonadotropin, corticotropin, α or β-melanocyte-stimulating hormone, β-lipotropin, γ-lipotropin and the endorphins; hypothalamicmic release-inhibiting hormones such as corticotropin-release factor, growth hormone release-inhibiting hormone, growth hormone-release factor; and other polypeptide hormones.

[0088] In another embodiment, the present disclosure is directed to targeting cell surface (or membrane or transmembrane) proteins to reduce or eliminate their localization on the cell surface. To that end, the present disclosure provides inhibitors of 3'UTR localization function (such as shRNA, RNAi, or CRISPR system against DNA encoding long 3'UTR) that will block access by HuR to the HuR binding site within 3'UTR (see, e.g. Figure 1, Figure 5 and Example 1).

[0089] In one embodiment, the present disclosure provides compositions and methods for enhancing the cell surface localization of proteins. Observed and expected cell surface localization enhancements obtained by use of the compositions and methods of the present disclosure commonly include about 1.2 fold, about 1.3 fold, about 1.4 fold, about 1.5 fold, about 1.6 fold, about 1.7 fold, about 1.8 fold, about 1.9 fold, or about 2.0 fold or more increase in the proportion of a protein expressed on a cell surface compared to negative control (expression on the surface of the same type of cell in the absence of the present compositions and methods).

[0090] Vectors according to the invention include plasmids and viruses. Many such vectors and expression systems are well known and documented in the art. Numerous vectors are known in the art including, but not limited to, linear polynucleotides, transposons,

polynucleotides associated with ionic or amphiphilic compounds, plasmids, bacteriophages and viruses. Examples of viral vectors include, but are not limited to, vaccinia and attenuated vaccinia vectors such as MVA, adenoviral vectors, adeno-associated virus vectors, retroviral vectors such as herpes or lentivirus derived vectors, and the like.

[0091] In some embodiments, the vector is a plasmid, including plasmids designed to integrate into the host chromosome, and plasmids that are autonomously replicating in the host cell. The plasmid may be maintained in the host cell at various levels, and therefore, the vector of the invention may be a low copy-number plasmid or high copy-number plasmid. Alternatively, the vector may be a viral vector (commercially available from Addgene). Commercially available plasmids can be purchased from numerous vendors, including Addgene and New England Biolabs.

[0092] In some embodiments, suitable vectors can be chosen or constructed for expression and localization of the polypeptides of interest of the present disclosure, containing the appropriate regulatory sequences, including promoter sequences, terminator sequences, polyadenylation sequences, enhancer sequences, marker genes and other sequences as appropriate. Many known techniques and protocols for manipulation of nucleic acid, for example, in the preparation of nucleic acid constructs, mutagenesis, sequencing, introduction of DNA into cells and gene expression, and analysis of proteins, are described in detail in Short Protocols in Molecular Biology: a compendium of methods from Current protocols in molecular biology, Volume 2, F. M. Ausubel et al, eds., John Wiley & Sons, Inc., fifth edition, 2002.

[0093] The choice of the vector will depend on several factors, including the compatibility of the vector with the cell into which the vector is to be introduced (e.g., a mammalian cell for expression, or another host cell such as a bacterial cell, useful for propagating or amplifying the vector), and the ability of the vector to integrate into the mammalian or host cell genome. The vector can be a viral vector, a phage, a phagemid, a cosmid, a fosmid, a bacteriophage, an artificial chromosome, a cloning vector, a shuttle vector, a plasmid (linear or closed circular), or the like. Vectors can include chromosomal, non-chromosomal and synthetic DNA sequences. Large numbers of suitable vectors are known to those of skill in the art, and are commercially available. Examples of suitable vectors are provided in Sambrook et al, eds., Molecular Cloning: A Laboratory Manual (3rd Ed.), Vols. 1-3, Cold Spring Harbor Laboratory (2001). In some embodiments, vectors of the

present disclosure are commercially available and can be obtained from sources such as Addgene, Sigma Aldrich, Promega, Invitrogen, etc.

[0094] Vectors of the present disclosure can comprise any of a number of promoters known to the art, wherein the promoter is constitutive, regulatable or inducible, cell type specific, tissue-specific, or species specific. Further specific examples include, e.g., tetracycline-responsive promoters (Gossen M, Bujard H, Proc Natl Acad Sci USA. 1992,15;89(12):5547-51). In addition to the sequence sufficient to direct transcription, a promoter sequence of the invention can also include sequences of other regulatory elements that are involved in modulating transcription (e.g.: enhancers, kozak sequences and introns). Many promoter/regulatory sequences useful for driving constitutive expression of a gene are available in the art and include, but are not limited to, for example, the cytomegalovirus immediate early promoter enhancer sequence, the SV40 early promoter, the immunoglobulin promoter, as well as the Rous sarcoma virus promoter, and the like. Moreover, inducible and tissue specific expression of an RNA, transmembrane proteins, or other proteins can be accomplished by placing the nucleic acid encoding such a molecule under the control of an inducible or tissue specific promoter/regulatory sequence. Examples of tissue specific or inducible promoter/regulatory sequences which are useful for this purpose include, but are not limited to, the MMTV LTR inducible promoter, the SV40 late enhancer/promoter, synapsin 1 promoter, ET hepatocyte promoter, GS glutamine synthase promoter and many others. Various commercially available ubiquitous as well as tissue-specific promoters can be found at http://www.invivogen.com/prom-a-list. In addition, promoters which are well known in the art can be induced in response to inducing agents such as metals, glucocorticoids, tetracycline, hormones, and the like, are also contemplated for use with the invention. Thus, it will be appreciated that the present disclosure includes the use of any promoter/regulatory sequence known in the art that is capable of driving expression of the desired protein operably linked thereto. Expression of vectors can be transient or stable.

[0095] In addition, standard techniques are known in the art for creating functional promoters by mixing and matching known regulatory elements. "Truncated promoters" may also be generated from promoter fragments or by mixing and matching fragments of known regulatory elements as is practiced in the art (Andersen et al. Mol Biotechnol. 2011 Jun;48(2):128-37).

[0096] Vectors of the present disclosure can be delivered to cells of any type, including but not limited to neural cells (including cells of the peripheral and central nervous systems, in particular, brain cells), lung cells, retinal cells, epithelial cells (e.g., gut and respiratory epithelial cells), muscle cells, pancreatic cells (including islet cells), hepatic cells, myocardial cells, bone cells (e.g., bone marrow stem cells), hematopoietic stem cells, spleen cells, keratinocytes, fibroblasts, endothelial cells, prostate cells, germ cells, immune system cells, for example T-cells, and the like. Alternatively, the cell may be any progenitor cell. As a further alternative, the cell can be a stem cell. Moreover, the cells can be from any species of origin.

[0097] Vectors according to the present disclosure can be transformed, transfected or otherwise introduced into a wide variety of host cells. Transfection refers to the taking up of a vector by a host cell whether or not any coding sequences are in fact expressed. Numerous methods of transfection are known to the ordinarily skilled artisan, for example, lipofectamine, calcium phosphate co-precipitation, electroporation, DEAE-dextran treatment, microinjection, viral infection, and other methods known in the art. Transduction refers to entry of a virus into the cell and expression (e.g., transcription and/or translation) of sequences delivered by the viral vector genome. In the case of a recombinant vector, "transduction" generally refers to entry of the recombinant viral vector into the cell and expression of a nucleic acid of interest delivered by the vector genome.

CD47 as an Example of a Cell Surface Protein Localized by Long 3'UTR

[0098] CD47 is best known as a ubiquitous cell surface molecule that acts as a marker of self and protects cells from phagocytosis by macrophages. In the present disclosure, the inventors show that CD47 protein is expressed on the cell surface, as well as intracellularly. As described here, the CD47 gene produces alternative 3'UTRs (SEQ ID NO 1), wherein the two isoforms differ in the length of the 3'UTR. Further, the association of two 3'UTR isoforms of different length with the CD47 molecule phenomenon appears to be universal for CD47, as the presence of two 3' UTR isoforms was detected in every cell line tested (Example 1) and in other membrane proteins (Example 4). Surprisingly, exclusive knockdown of long 3'UTR decreased only CD47 surface expression, without changing intracellular (or total) expression, indicating that the long 3'UTR facilitates cell surface localization (Example 1). Moreover, it appears that some of CD47 encoded by RNA having the short 3'UTR isoform is also able to be localized on the cell surface but the mechanism is

unknown and the surface fraction is much smaller than the surface localized fraction expressed in association with the long 3'UTR isoform of CD47 (Examples 1 and Example 8).

[0099] The present disclosure demonstrates the ability of the long 3' UTR of CD47 and fragments thereof (Example 5) to control cell surface localization of CD47 and to facilitate cell surface localization of proteins other than CD47. As described in Example 2, transcription of a construct containing nucleic acid encoding green fluorescent protein (GFP) containing the long 3'UTR, along with the signal peptide and the C-terminal of CD47 containing one or more transmembrane domains of CD47 including the SET binding site (named GFP-TM-LU), followed by translation results in GFP localized on the cell surface. On the contrary, GFP encoded by an mRNA containing the peptide signal and transmembrane domain of CD47 including its C-terminal, but only the short 3'UTR of CD47 (named GFP-TM-SU) localized mostly in the endoplasmic reticulum (ER). Notably, the long 3'UTR isoform of CD47 encodes information necessary for cell surface expression independently of RNA localization.

[00100] The foregoing results make clear that the surface localization function of the long 3'UTR of CD47 is not confined to CD47 but will operate to localize to the surface any protein that is being expressed under the influence of such a 3'UTR or fragment, even proteins that are not localized to the surface in their natural state, such as GFP. Indeed, the choice of GFP was also arbitrary, making the showing widely applicable to any polypeptide of interest. The susceptibility of other polypeptides to be directed to the cell surface through their own long 3'UTRs containing uridine rich regions underscores the general applicability of the present vectors, constructs and methods. For all tested genes, the longer 3'UTR increased surface localization of GFP-TM (Example 4).

The Role of HuR in Protein Surface Localization Mediated by a Long 3'UTR

[00101] According to the present disclosure, 3'UTR-dependent protein localization is dependent on an RNA-binding protein (RBP) Embryonic Lethal Abnormal Vision (ELAV) L1/Human antigen R (HuR) (Ma et al. *J Biol Chem. 1996 Apr 5; 271(14):8144-51).* HuR belongs to the ELAV/Hu family of proteins. The ELAV/Hu proteins (HuR, HuB, HuC, and HuD) possess three RNA-recognition motifs through which they bind with high affinity and specificity to target mRNAs containing AU- and U-rich sequence elements (ARE) found

within 3′-untranslated regions (3′-UTRs) and modify their expression by altering their stability, translation, or both. Since the long 3'UTR of CD47 contains many uridine-rich elements (illustrated in the printed version of SEQ ID NO: 1) which could potentially be bound by HuR, the inventors tested whether HuR was needed in 3'UTR-dependent protein localization. Indeed, as shown in the Example 3 of the present disclosure, HuR knockdown by shRNA reduced CD47 protein surface expression without affecting its overall mRNA or protein levels. Thus, in order to facilitate the localization of proteins to the cell surface, the long 3'UTR of *CD47* requires the activity of HuR; an effective segment of the 3'UTR should include the HuR binding site, which can be found within or immediately adjacent to and overlapping with one or more uridine-rich regions of the long 3'UTR.

[00102] In addition to its role as an RBP, HuR has been shown to interact with various proteins through protein-protein interactions, including SET, ANP32A, and ANP32B (Brennan et al. *J. Cell Biol.* 151, 1-14 (2000). SET contains a long acidic C-terminus that interacts with the RRM 3 (RNA recognition motif 3) of HuR. While nuclear SET binds to histone tails and prevents acetylation, phosphorylated SET localizes to the cytoplasm and the surface of the endoplasmic reticulum (ten Klooster et al. *EMBO J* 26, 336-345 (2007); Fan et al. *EMBO J* 17, 3448- 3460 (1998)). Furthermore, SET was shown to interact with Rac1, which results in translocation of SET to the plasma membrane upon Rac1 activation. Whereas Rac1-GDP is mostly localized to the cytoplasm, Rac1-GTP (active Rac1) localizes to the leading edge of the plasma membrane (ten Klooster et al. *EMBO J* 26, 336-345 (2007)). In this disclosure, the inventors have demonstrated that in addition to HuR, SET and RAC1 are also needed for 3'UTR-dependent protein localization (Example 3). Indeed, if the C- terminus of the membrane-associated protein is not included in the polynucleotide construct, the SET binding site will be necessarily omitted and localization will be substantially reduced. Alternatively, the entire C-terminus may not be necessary as long as the SET-binding site is included. Blazemarks for SET binding sites are provided in Fig. 13 and Table 1 and its description herein (Example 7) or various membrane-associated proteins: CD47 (SEQ ID NO 8), CD44 (SEQ ID NO 9), ITGA1 (SEQ ID NO 10), BAFFR (SEQ ID NO 11) and TSPAN13 (SEQ ID NO 12).

[00103] As shown in the present disclosure, 3'UTR-dependent protein localization is a widespread mechanism for surface expression of membrane proteins. Examination of numerous transmembrane proteins that are derived from mRNAs with 3'UTR isoforms that

can be bound by HuR (Kishore, S. *et al. Nat Methods* 8, 559-564 (2011); Lebedeva, S. *et al. Mol Cell* 43, 340-352 (2011); Mukherjee, N. *et al. Mol Cell* 43, 327-339 (2011)) showed that their cell surface localization is also dependent on HuR (Example 4). The inventors generated GFP-fused long UTR (LU) and short UTR (SU) constructs for transmembrane proteins, including transmembrane proteins CD44, ITGA1, and TNFRSF13C, as well as their respective transmembrane domains (TMDs), C-termini, and 3'UTR. For all tested genes, the longer 3'UTR increased surface localization of GFP-TM (Example 4). These experiments demonstrated that 3'UTR-dependent protein localization is a general phenomenon and is not confined to the 3'UTR of CD47. Furthermore, 3'UTR dependent protein localization can be used to direct an arbitrary expressed polypeptide like GFP, and hence any polypeptide, to the cell surface.

[00104] In some embodiments, the present disclosure provides a smaller 3'UTR segment of the longer 3'UTR sufficient for surface localization of transmembrane proteins. As shown in Example 5, a 3'UTR (74 nucleotide long) with only a few HuR-binding sites (HuR-BS) was adequate to mediate the cell surface localization. On the contrary, deletion of the 31-nucleotide uridine-rich sequence (HuR-BSΔ) abolished the ability of 74- nucleotide long sequence to mediate cell surface protein localization. Thus, a uridine-rich sequence embodying an HuR binding site is necessary and sufficient for surface localization by UDPL. The uridine-rich sequence can be much shorter than the long 3UTR itself as illustrated not only by the 74-nt segment but also by even shorter polynucleotides described below.

3' UTR and the AU Rich Elements

[00105] As previously mentioned, 3 ' UTRs are known to have stretches rich in adenosine and uridine codons embedded in them. These AU rich signatures are particularly prevalent in genes with high rates of turnover. Based on their sequence features and functional properties, the AU rich elements (AREs) can be separated into three classes (Chen et al, 1995): Class I AREs contain several dispersed copies of an AUUUA motif within U-rich regions. C-My c and MyoD contain class 1 AREs. Class II AREs possess two or more overlapping UUAUUUA(U/A)(U/A) nonamers. Molecules containing this type of AREs include GM-CSF and TNF-α. Class III ARES are less well defined (Peng et al. *Mol. Cell. Biol.* 16:1490–1499 (1996). These AU rich regions do not contain an AUUUA motif. c-Jun and Myogenin are two well-studied examples of this class. HuR binds to AREs of all the three classes. Thus,

according to the methods of present disclosure, engineering the HuR specific binding sites into the 3 ' UTR of nucleic acid molecules is anticipated to lead to increased surface localization of proteins.

[00106] The minimum number of uridines necessary to establish an HuR binding site has not been established. Site directed mutagenesis may be used to ascertain the minimum number of uridines needed for HuR binding that would be effective to increase surface localization of a protein. It is also possible that not all uridine stretches need to be contiguous.

[00107] In some embodiments, the minimal sequence containing an HuR binding site that is sufficient for surface localization of proteins comprises one stretch of two uridines. In some embodiments, the minimal sequence sufficient for surface localization of proteins comprises a stretch of three uridines. In some embodiments, the minimal sequence sufficient for surface localization of proteins comprises a stretch of five uridines.

[00108] In some embodiments, the minimal sequence containing an HuR binding site that is sufficient for surface localization of proteins comprises two or more stretches of 2-7 uridines within a stretch of 100 nt.

The Role of SET

[00109] According to the present disclosure, HuR binds to the HuR-BS and to the *LU* but not to the *SU* isoform of *CD47* (Example 6). SET also associates with the long 3' UTR of *CD47*, which is dependent on HuR. As shown in Example 6, SET or HuR overexpression in the cell was insufficient to localize GFP-TM-SU to the cell surface. However, tethering of SET or HuR to the short 3'UTR isoform of *CD47* was sufficient to redirect GFP-TM localization from the endoplasmic reticulum to the plasma membrane. Thus, local recruitment of SET to the site of translation, mediated by the scaffold function of the long 3'UTR which binds HuR, is required for 3'UTR-dependent protein localization. However, if SET can be forced to stay near the 3'UTR region, the requirement for an HuR binding site in the 3'UTR is attenuated, confirming the mechanism proposed herein.

[00110] In further support of the generality of 3'UTR-dependent protein localization, all components of the pathway described in this disclosure (HuR, SET and RAC1) are

ubiquitously and highly expressed (Fig. 12 a) (*Genes Dev.* 27, 2380–2396 (2013). All candidates tested in this disclosure required both the presence of an HuR-binding site, as well as SET-binding site in the cytoplasmic domain of membrane proteins for efficient 3'UTR-dependent protein localization (Figs 1f–h, 2d, 4b, 6g, and 11c). HuR-binding sites are highly abundant as HuR binds to thousands of mRNAs (Kishore et al., *Nature Methods* 8, 559–564 (2011); Lebedeva et al., *Mol. Cell* 43, 340–352 (2011); Mukherjee, N et al., *Mol. Cell* 43, 327–339 (2011)), with a third of them being membrane proteins (Fig. 12b). Although the precise SET-binding motif has not been elucidated, the inventors (Fig. 3d, e and Fig. 10) and others have shown that SET binds to positively charged amino acids in histone tails or cytoplasmic domains of membrane proteins (Schneider et al. *J. Biol. Chem.* 279, 23859–23862 (2004), Avet et al., *J. Biol. Chem.* 288, 2641–2654 (2013). Thus, the SET binding site comprises of one or more positively charged amino acids, such as lysine or arginine and has to be located in the cytoplasmic domain of a membrane protein. According to the positive-inside rule for integral membrane proteins, the cytoplasmic domains of membrane proteins are enriched in positively charged amino acids for topological reasons (Nillson et al. *Proteins* 60, 606–616 (2005)). Therefore, potential SET-binding sites in cytoplasmic domains of membrane proteins are very widespread. In terms of the vectors and methods of the present disclosure, inclusion of all or a portion of the cytoplasmic domain (C- terminal) of a membrane protein harboring the SET binding site increases surface localization of the polypeptide of interest. Figure 13 and Table 1illustrate fragments of cytoplasmic domains of various proteins harboring SET- binding sites and indicate lysine and arginine residues potentially contained within such SET binding sites. These could be used in vectors, constructs and methods of the present disclosure to increase surface localization of a polypeptide of interest.

Table 1. Positively charged amino acids in the cytoplasmic domains of UDPL candidates, indicating potential SET-binding sites.

Name of the Protein	Sequence and Potential SET-Binding Sites (in bold, underscored)
CD47	NILIVIFPIFAILLFWGQFGI**K**TL**K**Y**R**SGGMDE**K**TIALLVAG LVITVIVIVGAILFVPGEYSLKNATGLGLIVTSTGILILLHY

	YVFSTAIGLTSFVIAILVIQVIAYILAVVGLSLCIAACIPMH GPLLISGLSILALAQLLGLVYM**K**FVASNQKTIQPP**RK**AVE EPLNAF**K**ES**K**GMM NDE
CD44	LIILASLLALALILAVCIAVNS**RRR**CGQ**KKK**LVINSGNGA VED**RK**PSGLNGEAS**K**SQEMV HLVN**K**ESSETPDQFMTADET**R**NLQNVDM**K**I GV
ITGA1	LWVILLSAF AGLLLLMLLI LALW**K**IGFF**KR**PL**KKK**ME**K**
BAFFR	FGAPALLGLALVLALVLVGLVSW**RRR**Q**RR**L**R**GASSAEA PDGD**K**DAPEPLD**K**VIILSPGISDATAPAWPPPGEDPGTTPP GHSVPVPATELGS TELVTT**K**TAG PEQQ
TSPAN1 3	MVCGGFACS**K**NCLCALNLLYTLVSLLLIGI AAWGIGFGLI SSLRVVGVVIAVGIFLFLIALVGLIGAV**K**HHQVLLFFYMII LLLVFIVQFSVSCACLALNQEQQGQLLEVGWNNTASARN DIQRNLNCCG FRSVNPNDTCLASCVKSDHSCSPCAPIIGE YAGEVLRFVGGIGLFFSFTEILGVWLTY**R**Y **R**NQ**K**DP**R**ANPSAFL

[00111] As described in Example 10, the inventors addressed the role of CD47 expressed in the ER. The majority of known functions of CD47 involve its expression on the cell surface, where, in addition to protecting cells from phagocytosis, it also plays an important role in cell adhesion and migration (Jaiswal et al. *Cell* 138, 271-285 (2009); Ridley et al., *Cell*

70, 401-410 (1992); Oldenborg et al. *ISRN hematology* 2013, 614619 (2013)). To examine if the difference in surface expression of CD47-LU and CD47-SU protects cells to a different extent from phagocytosis, similar total amounts of CD47-LU or CD47-SU were expressed in CD47-deficient Jurkat cells (called JinB8 cells) (Fig. 11a). Co-culture of these cells with macrophages demonstrated that CD47-LU fully protected the cells, whereas CD47-SU only partially protected the cells from phagocytosis (Fig. 4d).

[00112] However, CD47 also functions in the regulation of apoptosis (Lamy et al., *J Biol Chem* 278, 23915-23921 (2003)), as CD47 deficient Jurkat cells (called JinB8 cells, Reinhold et al., *International immunology* 11, 707-718 (1999)) and cells or tissues from CD47 knock-out mice (Lindberg et al. *Science* 274, 795-798 (1996)) fail to undergo apoptosis after γ-irradiation (Isanberg et al., *The American journal of pathology* 173, 1100-1112 (2008); Soto-Pantoja et al. *Journal of genetic syndrome & gene therapy* 2 (2011). The mechanism by which CD47 promotes a pro-apoptotic response to γ-irradiation has remained unknown. Interestingly, expression of CD47-SU in JinB8 cells restored apoptosis, but expression of CD47-LU did not affect the loss- of-apoptosis phenotype (Fig. 4e and Fig. 11f, g). Thus, when a cell requires increased CD47 surface expression, or such expression is desired for a cell, transcriptional upregulation alone would be non-optimal as it would confer increased susceptibility to apoptosis. ApA-generated 3' UTR isoforms allow independent regulation of differentially localized and functionally distinct CD47 protein. This highlights another benefit conferred by the vectors, constructs and methods of the present disclosure: a polypeptide of interest can be expressed and directed to the cell surface where it might exert one type of functionality without undesirable effects that may accompany wholesale upregulation of transcription of all forms (and increase of all effects) of the polypeptide of interest.

[00113] As shown in Example 11, CD47 protein localized to the same cellular compartment can exert different functions depending on whether it was produced by the *SU* or *LU* mRNA isoform. Based on findings of the present disclosure, it is likely that other surface proteins derived from their LU or SU isoforms have distinct biological functions. This in turn permits vectors constructs and methods of the present disclosure to enhance one or the other function of a polypeptide of interest by transcribing it with all or a portion of the long 3'UTR including an HuR binding site thereof and in certain embodiments also a SET, or with all or a portion of the short 3'UTR tethered to Hur or SET.

UTR sequences comprising HuR binding site from the 3'UTR of TNF-α enhance surface localization of proteins

[00114] In some embodiments, the present disclosure provides smaller 3'UTR segments that comprise an HuR binding site from TNF-α. As shown in Example 12, several artificial UTRs (UTR-1, UTR-2, and UTR-3) were able to increase surface localization of CD47 (Figure 15B, 15C, and 15D). Each of the three artificial UTRs comprises a sequence comprising an HuR binding site from TNF-α: TTGTGATTATTTATTATTTATTTATTATTTATTTATTTA (SEQ ID NO: 59). In order to identify additional elements that may enhance the ability of 3'UTR to localize proteins to the surface, the inventors identified three motifs that are enriched in the 3'UTRs of human membrane proteins (Figure 15A). These motifs (such as CACACA and GAGAGA) were then used as additional sequences in the design and generation of artificial 3'UTR-2. Given that these motifs are enriched in the 3'UTRs of membrane proteins, it is likely that the presence of one or more of these motifs in the 3'UTR serves to facilitate and/or increase surface localization of proteins. Additionally, artificial 3'UTR sequences of the present disclosure can comprise restriction site sequences, such as GC-rich sequences shown in Example 12. While GC-rich sequences are restriction sites, they make sure that the AU-rich sequences in between are accessible. In addition to restriction site sequences and sequence motifs enriched in the 3'UTRs of membrane proteins, the sequences of artificial 3'UTRs of the present disclosure can also comprise sequences found in the long 3'UTR of *CD47*. Addition of 3'UTR sequences of CD47 to the artificial 3'UTR of the present disclosure was carried out in UTR-2 and UTR-3 (see example 12, Table 2). Thus, in addition to sequences comprising an HuR binding site from TNF-α, the artificial 3'UTR sequences of the present disclosure can comprise one or more of additional elements including, but not limiting to restriction site sequences, sequences found in the long 3'UTR of *CD47,* sequence motifs enriched in the 3'UTRs of membrane proteins, and additional sequences that are likely to further enhance surface localization of proteins in at least one cell type.

[00115] In the present disclosure, the inventors have shown that AU-rich HuR binding sites found within the 3'UTR of TNF-α are efficient in enhancing protein surface localization. Thus, based on the present disclosure, the inventors anticipate that sequences found in the 3'UTRs of genes that contain AU-rich elements can also be used to promote and/or enhance protein surface localization. Genes that contain AU-rich elements within their

3'UTR include, but are not limited to ELAVL1, HNRNPD, IL1B, IL3, IFNB1, TNF, CSF2RA, VEGFA, FOS, JUNB, JUN, MYC, TP53, CDKN1A, CCNA2, CCNB1 and CCND1.

Kits

[00116] The present disclosure provides a variety of kits for conveniently and/or effectively carrying out methods of the present disclosure. Typically kits will comprise sufficient amounts of components to allow a user to perform multiple experiments. Such experiments include enhancing cell surface localization of a protein or fragment thereof, as well as promoting cell surface localization of a protein or a fragment thereof that otherwise does not localize to cell surface naturally.

[00117] In one embodiment, the present disclosure provides kits comprising the polynucleotides, vectors and constructs comprising one or more polynucleotides of the disclosure. In some embodiments, the kits of the present disclosure comprise fragments of the 3'UTR sequences, wherein the fragments comprise an HuR binding site. The kits can further comprise a vector into which said fragments of the 3'UTR sequences can be inserted. In another embodiment, the kit further comprises reagents such as a transfection reagent that can be used to introduce the vector after a polynucleotide of interest has been incorporated in it into cells. Instructions for use may also be optionally included. In another embodiment, the kit comprises a vector and instructions for its use.

EXAMPLES

Materials and Methods

[00118] Cell lines. MCF7 (breast cancer), HeLa (cervical cancer), HEK293 (embryonic kidney), Caov-3 (ovarian carcinoma), NTERA2 (embryonic carcinoma) and THP-1 cells (monocytic leukaemia) were purchased from ATCC. B-LCL are Epstein Barr virus (EBV)-immortalized human B cells described earlier (Lianoglou et al. *Genes Dev.* **27**, 2380–2396 (2013)). U2OS cells (sarcoma) were a gift from T. Brummelkamp, Toledo (B-cell lymphoma) cells were a gift from M. Mueschen, U251 (glioblastoma) cells were a gift from I. Mellinghoff, SHSY-5Y (neuroblastoma) cells were a gift from T. Tuschl and Jurkat (T-cell leukaemia) and JinB8 (CD47-negative Jurkat, Reinhold et al., *International immunology* 11,

707-718 (1999)) cells were a gift from W. Frazier. U2OS (ATCC® HTB-96™, Toledo (ATCC® CRL-2631™), Jurkat ((ATCC® TIB-152™), Epstein-Barr virus (EBV) transformed human cells ((ATCC® CRL-8119™) can be purchased from ATCC; while U-251 MG (09063001 Sigma) and SHSY-5Y (94030304 Sigma) are commercially available from Sigma Aldrich.

Constructs

[00119] For some of the shRNA knockdown experiments, pSUPERretropuro was modified by cloning IRES::GFP (derived from pMSCVpig) (Mayr et al. *Cell* **138**, 673–684 (2009)) downstream of puromycin to obtain pSUPERretropuro containing enhanced (e) GFP (shRNA-GFP). The following DNA oligonucleotides served as shRNA precursors and were cloned into pSUPER-GFP or pSUPER.

SEQ ID NO 13:CD47-LU2F: 5'-GATCTCCCAGCTGT GTTACCGTTAAATTCAAGAGATTTAACGGTAACACAGCTGTTTTTC-3'; SEQ ID NO 14: CD47-LU3F: 5'-GATCCCCCAGCTGTGTTACCGTTAAATTCAAGAGATTTAACGGTAACACAGCTGT TTTTC-3'; SEQ ID NO15: HuR2F: 5'-GATCTCCGATCAGA CTACAGGTTTGTTTCAAGAGAACAAACCTGTAGTCTGATCTTTTTC-3'; SEQ ID NO 16: HuR3F: 5'- GATCCCCGAGGCAATTACCAGTTTCATTCAAGAGATGAAA CTGGTAATTGCCTCTTTTTC-3'; SEQ ID NO 17: HuR4F: 5'-GATCCCCTCTTAAGTTTCG TAAGTTATTCAAGAGATAACTTACGAAACTTAAGATTTTTC-3'; SEQ ID NO 18: SET1F: 5'-GATCCCCTGAAATAGACAGACTTAATTTCAAGAGAATTAAGTCT GTCTATTTCATTTTTC-3'; SEQ ID NO 19: SET2F: 5'- GATCTCCCTGGTTTACTGAC CATTCTTTCAAGAGAAGAATGGTCAGTAAACCAGTTTTTC-3'; SEQ ID NO 20: RAC2F: 5'- GATCCCCCTTGCCTACTGATCAGTTATTCAAGAGATAACTGATCA GTAGGCAAGTTTTTC-3'; SEQ ID NO 21: RAC3F: 5'-GATCCCCGTCCCTTGGAACCT TTGTATTCAAGAGATACAAAGGTTCCAAGGGACTTTTTC-3'; SEQ ID NO 22: ControlF: 5'- GATCTCCTTCTCCGAACGTGTCACGTTTCAAGAGAACGTGACACG TTCGGAGAATTTTTC-3'.

[00120] The sequence of shRNA1SET and of shRNA Control were published earlier (Neviani et al. *Cancer Cell* **8**, 355–368 (2005); Nho et al. *J. Biol. Chem.* **281**, 33291–33301 (2006)). The GFP fusion constructs were generated in pcDNA3.1 expression vector (Life Technologies). The short and long 3'UTRs of *CD47* were PCR-amplified from genomic DNA using Q5 High Fidelity DNA polymerase (NEB) and the primers listed later and inserted between the NotI and XbaI sites. To obtain expression of only the long 3'UTR isoform of *CD47*, the proximal polyadenylation site was mutated from AAUAAA to ACUCAA using the QuikChange Multi Site Directed Mutagenesis Kit (Agilent). The resulting plasmids were used to test qPCR primers for accuracy in measuring short to long 3'UTR isoform ratios (see later). The short 3'UTR of *CD47* used in the MS2-BS construct was cloned from the plasmid containing the long 3'UTR with the mutated proximal polyadenylation site. eGFP was PCR-amplified from pMSCV-pig and inserted upstream of each *CD47* 3'UTR (BamHI, NotI). The signal peptide of *CD47* was generated by annealing two DNA oligonucleotides that were inserted into the BamHI site. To generate the GFP-TM constructs, the sequence of the TMDs and C-terminal tail of *CD47* (the longest isoform, isoform 4; ref. Reinhold et al. *J. Cell Sci.* **108**, 3419–3425 (1995)) were cloned fromToledo cDNA and inserted downstream of eGFP (BsrgI, NotI). Isoform 4 was chosen as it is the most abundant isoform in Jurkat cells (data not shown). The two nucleotides of eGFP that occur after the BsrgI site were included in the forward primer. To generate the GFP-CD47 constructs, the ECD, TMDs and C terminus of *CD47* were PCR amplified from Toledo cDNA using the TM reverse primer and the CDS forward primer and inserted downstream of eGFP (BsrgI, NotI). The sequence of HuR-BS and HuR-BSΔ are shown in Fig. 7 and replaced the long 3'UTR in GFP-TM-LU. The GFP constructs containing the TMDs and C termini fused to either the long or short 3'UTRs of *CD44* (SEQ ID NO 2), *ITGA1* (SEQ ID NO 5) and *TNFRSF13C* (SEQ ID NO 6) were generated as follows. TMDs, C termini and short 3'UTRs of *CD44*, *ITGA1* and *TNFRSF13C* were cloned from Toledo, SHSY-5Y and B-LCL cDNA respectively and inserted downstream of eGFP (BsrgI, XbaI). Short 3'UTRs consisted of the first 122 nucleotides of *CD44*, the first 45 nucleotides of *ITGA1* and the first 337 nucleotides of *TNFRSF13C*. Long 3' UTRs: 3,068 nucleotides after the stop codon of *CD44* and 3,996 nucleotides after the stop codon of *ITGA1* were used. For *TNFRSF13C*, 1,221 nucleotides after the stop codon together with the last 600 nucleotides of the 3'UTR (nucleotides 2,712–3,311 after the stop codon containing the majority of HuR-binding sites) were used as long 3'UTR and were cloned from genomic DNA. To generate BAFFR-SU and -LU, the complete open reading frame of *TNFRSF13C* (without the start codon) was amplified from human B-cell cDNA and cloned

using Gibson Assembly Cloning (NEB) into pcDNA3.1 vector used above downstream of eGFP. To generate CD44-SU and -LU, the open reading frame of CD44 (without the start codon) was amplified from cDNA of the human breast cancer cell line MDA-MB231 and cloned using Gibson Assembly Cloning (NEB) into pcDNA3.1 vector used above downstream of eGFP. To generate the GFP-TM-LUΔC construct the sequence of just the TMDs of *CD47* was cloned from Toledo cDNA using the TM forward primer and TMΔC reverse primer and inserted downstream of eGFP (BsrgI, NotI). The 24 MS2-binding sites were cloned from a plasmid obtained from J.Gerst (Slobodin et al. *RNA* **16**, 2277–2290 (2010)) using the primers listed later (XbaI, ApaI). The constructs in which K163, K166, K175, K290, K297 and K304 were mutated to alanines were generated using the QuikChange Multi Site Directed Mutagenesis Kit (Agilent).

SEQ ID NO: 23 CD47UTRF: 5'-ATGCGCGGCCGCAGTGAAGTGATGGACTCCGATT-3'; SEQ ID NO 24: CD47shortUTRR: 5' -ATGCTCTAGATGGGCAAACAACATAGATCA-3'; SEQ ID NO 25: CD47longUTRR: 5'ATGCTCTAGAAACACATTGGACTGATTTAAAACTT-3'; SEQ ID NO 26: GFPF: 5'-ATGCGGATCCATGGTGAGCAAGGGCGA-3'; SEQ ID NO 27: GFPR: 5'-ATGCGCGGCCGCTTACTTGTACAGCTCGTCCATG-3'; SEQ ID NO 28: SPCD47F: 5'-GATCCATGTGGCCCCTGGTAGCGGCGCTGTTGCTGGGCTCGGCGTGCTGCGGATCAGCTG-3'; SEQ ID NO 29: SPCD47R: 5'-GATCCAGCTGATCCGCAGCACGCCGAGCCCAGCAACAGCGCCGCTACCAGGGGCCACATG-3'; SEQ ID NO 30: CD47TMF: 5'-ATGCTGTACAAGATTCTTATTGTTATTTTCCCAATT-3'; SEQ ID NO 31: CD47TMR: 5'-ATGCGCGGCCGCTTATTCATCATTCATCATTCCTT-3'; SEQ ID NO 32: CD47CDSF: 5'-ATGCTGTACAAGCAGCTACTATTTAATAAAACAA-3'; SEQ ID NO 33: TMDCR: 5'-ATGCGCGGCCGCTTATTTCATATAAACTAGTCCAAGTAA-3'; SEQ ID NO 34: MS2-BSF: 5'-ATGCTCTAGAGGGCCCTATATATCGATCCTAAG-3'; SEQ ID NO 35: MS2-BSR: 5'-ATGCGGGCCCTTTATTATGCTTGGTACCGAGCTCG-3'.

[00121] To create theMS2 fusion constructs (Bertrand et al. *Mol. Cell* **2**, 437–445 (1998), the pUG34-MS2-GFP-SBP plasmid was obtained from J. Gerst (Slobodin et al. *RNA* **16**, 2277–2290 (2010)). SBP was replaced by either HuR, SET or a stop codon. HuR and SET were PCR-amplified from U2OS cDNA using the primers listed below (BsrgI, XbaI). The stop codon was generated by annealing two DNA oligonucleotides (BsrgI, XbaI). mCherry

was PCR-amplified and replaced GFP (BamHI, BsrgI). After cloning was complete, all plasmids were sequenced to assure fidelity of the sequences.

SEQ ID NO 36: HuRF: 5'-ATGCCGTACGAGTCTAATGGTTATGAAGACCACA-3'; SEQ ID NO 37: HuRR: 5'-ATGCTCTAGATTATTTGTGGGACTTGTTGGT-3'; SEQ ID NO 38: SETF: 5'-ATGCTGTACAAGTCGGCGCCGGCGGCCAAA-3'; SEQ ID NO 39: SETR: 5' -ATGCTCTAGATTAGTCATCTTCTCCTTCATCCTC-3'; SEQ ID NO 40: StopF: 5'-GTACAAGTAATAATAAT-3'; SEQ ID NO 41: StopR: 5'-CTAGATTATTATTACTT-3'; SEQ ID NO 42: mCherryF: 5'-ATGCGGATCCGTGAGCAAGGGCGAGGAG-3'; SEQ ID NO 43: mCherryR: 5'-TTACTTGTACAGCTCGTCCATGC-3'.

The N17RAC1 construct was provided by A. Hall (Ridley et al. *Cell* **70**, 401–410 (1992)).

Transfections

[00122] For transfections into U2OS, HEK293, HeLa and U251 cells Lipofectamine 2000 (Invitrogen) and for SHSY-5Y cells Xtreme reagent (Roche) was used. JinB8 and Jurkat cells were transfected using the Neon Transfection System (Invitrogen) according to the manufacturer's protocol for transfecting Jurkat cells. To account for differences in the sizes of transfected plasmids the same molar amounts were transfected. When RNA was to be extracted, the cells were grown in the presence of high amounts of puromycin (4 mg ml^{-1}) for 3 days and FACS analysis was performed to ensure that >90% of the cells were GFP^+.

Generation of cell lines with stable expression of shRNAs.

[00123] Stable cell lines were generated as described previously (Mayr et al. *Cell* **138**, 673–684 (2009)).

FACS analysis

[00124] For surface FACS (in order to detect surface protein expression), cells were incubated with mouse anti-CD47-PerCy5.5 (BD Biosciences, 561261), mouse anti-CD44-PE (BD Biosciences, 561858), chicken anti-GFP (Abcam, ab13970), mouse anti-ITGA1-PE (BD Biosciences, 555749), mouse anti- BAFFR-PE (BD Biosciences, 554680) or rabbit anti-TSPAN13 (Genetex, GTX52155) in FACS buffer A (0.5% FBS in PBS) for 30 min at 4°C,

and then washed twice in FACS Buffer A. For detection of GFP and TSPAN13, cells were then incubated with goat anti-chicken Alexa Fluor 568 or 633 (Invitrogen, A11041 or A21103) and goat anti-rabbit Alexa Fluor 680 (Invitrogen, A-21076), respectively, for 30 min at 4°C, and then washed twice in FACS Buffer A. At least 30,000 cells were analyzed on a BD FACS Calibur cell analyser (BD Biosciences) and FACS data were computed using the FlowJo VX software.

[00125] For intracellular FACS (in order to detect total protein expression), cells were fixed for 15 min at room temperature in fixation buffer (4% PFA, 0.02% sodium azide, and 0.1% Tween 20 in PBS), washed in FACS buffer B (0.02% sodium azide, 0.1% Tween 20 in PBS), permeabilized for 10 min at 4°C in permeabilization buffer (0.02% sodium azide, 0.1% Tween 20 and 10% dimethyl sulfoxide in PBS), washed, re-fixed for 5 min at room temperature in fixation buffer, and washed again. Cells were incubated with the same primary and secondary antibodies as for surface FACS in FACS buffer B for 30 min at 4°C, and then washed twice in FACS Buffer B. For live/dead analysis by FACS, cells were incubated with TO-PRO3 in FACS buffer A (0.5% FBS in PBS) for 10 min at 4°C, and then washed twice in FACS Buffer A. Cells were analysed is the same manner as for surface FACS.

[00126] For all the GFP-expressing plasmids the 20% of cells with the highest GFP expression are shown.

Immunocytochemistry

[00127] For surface staining of CD47, cells were fixed for 15 min at room temperature in fixation buffer A (4% PFA and 0.02% sodium azide in PBS), washed with PBS, blocked for 15 min at 4°C in 5% Normal Goat Serum (Invitrogen, PCN5000) in PBS and then incubated with mouse anti-human CD47 (Santa Cruz, sc-59079) primary antibody for 1 h at 4°C in PBS. After washing in PBS, donkey anti-mouse Alexa Fluor 594 (Invitrogen, A-21203) secondary antibody was incubated for 1 h at 4°C in PBS, and during the last 10 min of incubation 4',6-diamidino-2-phenylindole (DAPI; Invitrogen, D1306) was added, followed by three washes of PBS. Mowiol mounting media (Sloan Kettering Institute, Molecular Cytology Core Facility) was used to mount the slides. Imaging was performed at the Sloan Kettering Institute Molecular Cytology Core Facility, on a Leica TCS SP5 confocal microscope, using a ×63, 1.4 numerical aperture oil objective.

[00128] For intracellular staining of CD47, cells were fixed for 15 min at room temperature in fixation buffer B (4% PFA, 0.02% sodium azide, and 0.1% Tween 20 in PBS), washed in wash buffer (0.02% sodium azide, 0.1% Tween 20 in PBS), permeabilized for 10 min at 4°C in permeabilization buffer (0.02% sodium azide, 0.1% Tween 20 and 10% dimethyl sulfoxide in PBS), washed, re-fixed for 5 min at room temperature in fixation buffer B, washed and blocked for 15 min at 4°C in 5% Normal Goat Serum in wash buffer. Mouse anti-CD47 (Santa Cruz, sc-59079) primary antibody was incubated overnight at 4°C in wash buffer. The strong permeabilization and overnight staining were necessary to visualize intracellular CD47, as the CD47 antibody recognizes an epitope in the ECD of CD47, which is located in the lumen of the endoplasmic reticulum. Owing to the extended treatment with a buffer containing Tween 20 the plasma membrane could no longer be visualized in these cells.

[00129] For co-staining of GFP with calnexin or RAC1, the cells were fixed for 15 min at room temperature in fixation buffer B (4% PFA, 0.02% sodium azide, and 0.1% Tween 20 in PBS), washed with wash buffer B, blocked for 15 min at 4°C in 5% Normal Goat Serum (Invitrogen, PCN5000) in wash buffer and then incubated with rabbit anti-calnexin (Santa Cruz, sc-11397) or mouse anti-RAC1 (Abcam, ab12048) primary antibodies for 1 h at 4°C in wash buffer. After washing, goat anti-rabbit Alexa Fluor 680 (Invitrogen, A-21076) or donkey anti-mouse Alexa Fluor 594 (Invitrogen, A-21203) secondary antibodies were incubated for 1 h at 4°C in wash buffer, and during the last 10 min of incubation DAPI (Invitrogen, D1306) was added, followed by three washes. Mounting and imaging was performed as for the surface staining. Owing to the lack of permeabilization and short period in Tween 20, the plasma membrane was still visible. When calnexin was costained with endogenous CD47, the protocol for intracellular staining of CD47 was used and the plasma membrane was again not visible.

[00130] Alexa Fluor 594 and 680 were pseudo-coloured red and endogenous GFP and mCherry were imaged as they appear, without any antibody.

RNA-FISH

[00131] Custom Stellaris FISH Probes (Biosearch Technologies) were designed for the open reading frame of eGFP using the Stellaris Probe designer website, and with the assistance of Biosearch Technologies staff. The probes were conjugated to the Quasar 670

fluorochrome. Staining was carried out according to the manufacturer's protocols. Briefly, 24 h after transfection of the GFP-TM constructs cells were trypsinized and plated on Millicell EZ glass slides (Millipore) and allowed to grow overnight. Cells were washed in PBS, fixed in 4% PFA at room temperature for 10 min and permeabilized in 70% ethanol at 4°C for 2 h. After washing, the probes were hybridized at 37°C for 4 h, washed, incubated with DAPI at 37°C for 30 min, washed and mounted in Mowiol mounting media. Imaging was performed as for the surface immunostaining. RNA was pseudo-coloured red, and GFP was imaged as it appears, without any antibody.

3'-seq. 3'-seq reads of naive B cells, B-LCL and HEK293 cells were analysed and visualized as described previously[3].

Northern blot analysis

[00132] Northern blots were performed as previously described (Mayr et al. *Cell* **138**, 673–684 (2009)).

SEQ ID NO 44: CD47 probe F: 5'-TTGATGGAGCTCTAAACAAGTCC-3'; SEQ ID NO 45: CD47 probe R: 5'-GAATAACCAATATGGCAATGACG-3'; SEQ ID NO 46: GFP probe F: 5'-TAAACGGCCACAAGTTCAGC-3'; SEQ ID NO 47: GFP probe R: 5'-CTTGTACAGCTCGTCCATGC-3'.

Quantitative PCR

[00133] Runaways extracted using TRI Reagent (Ambion) according to the manufacturer's protocol. cDNA was synthesized using random hexamers and the TaqMan Reverse Transcription Kit (Applied Biosystems). qRT–PCR was performed using the Power SYBR Green master mix (Applied Biosystems) on a 7500 HT Fast Real-Time PCR System (Applied Biosystems). Each reaction was performed in triplicate. The experiments were performed at least three times to obtain at least three biological replicates. The following primers were used to quantify total *CD47* mRNA, the long 3' UTR isoform of *CD47* and *GAPDH* for normalization.

CD47TotalF: 5'-AGTGATGGACTCCGATTTGG-3'; CD47TotalR: 5'-GGGTCTCATAGGTGACAACCA-3'; CD47LongF: 5'-AAGAGAACTCCAGTGTTGCT-3'; CD47LongR: 5'-ACGGTAACACAGCTGTAAAACA-3'; GAPDHF: 5'-ACAACTTTGGTATCGTGGAAGG-3'; GAPDHR: 5'-

TATTTGGCAGGTTTTTCTAGACG-3'.

[00134] To measure 3'UTR isoform expression by qRT–PCR and to take into account different affinities of primers, we generated a standard using plasmids that contained either the short or the long 3'UTR of *CD47* (see earlier). We mixed together known quantities of the two plasmids ranging from 3:1 to 1:3 of short to long 3'UTR and performed qPCR on these mixtures to test the accuracy of our primer sets. The fraction of the long 3'UTR isoform was calculated by subtracting the CT value of the long isoform from the CT value for total *CD47* expression. The fraction of the long 3'UTR isoform corresponds to $2^{\text{CT difference}}$. The fraction of the short 3'UTR isoform was obtained by subtracting the fraction of the long 3'UTR isoform from the total *CD47* mRNA.

Immunoprecipitation of RNA complexes and RT–PCR

[00135] GFP-TM-LU, GFPTM- SU, GFP-TM-HuR-BS and GFP-TM-HuR-BSΔ were transfected into HEK293 cells and immunoprecipitation of protein–RNA complexes was carried out as previously described[37]. RNA-immunoprecipitations were performed with crosslinking to prevent re-association of HuR with mRNA after lysis[38]. Briefly, cells were harvested and washed twice in cold PBS. Formaldehyde was added to a final concentration of 1% (v/v) and the cells were incubated at room temperature for 10 min. The reaction was quenched by addition of glycine to a final concentration of 0.25M and incubated at room temperature for 5 min. After centrifugation the cell pellet was washed twice in cold PBS and resuspended in RIPA buffer (25mM Tris-HCl (pH 7.4), 150mM NaCl, 1% NP-40, 1% Na-deoxycholate, 0.1% SDS, 1mM EDTA, protease inhibitor cocktail (Roche, 04693124001)). The mRNPs were solubilized by three rounds of sonication for 15 s each in a Misonix Ultrasonic Processor S-4000 at an output of 8–9 W. Insoluble material was removed by centrifugation. Lysates were pre-cleared by addition of magnetic protein A beads (Millipore, LSKMAGA10) and incubation for 30 min at 4°C with constant mixing. The pre-cleared lysate was divided into three parts. One portion was retained for the input control and anti-HuR (Millipore, 07-1735) or IgG (Santa Cruz, sc-2025) were added to the two other portions and incubated at 4°C for 2 h. Magnetic protein A beads (Millipore, LSKMAGA10) were added and incubated at 4°C for 1 h. The beads were washed five times with high-stringency RIPA buffer (50mM Tris-HCl (pH 7.4), 1M NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS and 1mM EDTA). The crosslinking was reversed by resuspending the beads in 50mM Tris-HCl (pH 7.0), 5mM EDTA, 10mM dithiothreitol (DTT) and 1% SDS, followed

by incubation at 70°C for 45 min. RNA was extracted from the beads and buffer using TRI Reagent (Ambion) according to the manufacturer's protocol, and cDNA was synthesized as previously described. qRT–PCR was carried out using primers for *GAPDH* (see earlier) and *GFP*, so as not to amplify endogenous *CD47*. The primers used were as follows: SEQ ID NO 48: GFPF: 5'-TAAACGGCCACAAGTCAGC-3';SEQ ID NO 49: GFPR: 5'-AAGTCGTGCTGCTTCATGTG-3'.

[00136] HEK293 cells transiently transfected with sh2 HuR or sh Co were used to assess the presence of SET on endogenous *CD47* mRNA and the requirement of HuR for this association. Cells were transfected and treated with high dose puromycin for 3 days. FACS analysis was used to ensure greater than 90% of surviving cells expressed the shRNA. RNA-immunoprecipitation was carried out as described earlier, but using anti-SET (Abcam, ab181990) or IgG (Santa Cruz, sc-2025). The following primers were used for qRT PCR of *CD47-LU*: CD47F: 5'-AAGAGAACTCCAGTGTTGCT-3'; CD47R: 5'-ACGGTAACACAGCTGTAAAACA-3'.

[00137] The obtained CT values were first normalized to *GAPDH* and then the fraction of input mRNA was plotted.

Western blotting

[00138] HEK293 cells stably expressing sh3 HuR, sh4 HuR, sh1 SET, sh2 SET, sh2 Rac1, sh3 Rac1 or sh Co, as well as HEK293 cells transiently transfected with sh2 HuR, CD47-SU, CD47-LU, CD47-LU1N17Rac1, CD47-LU2Km, CD47-LUΔC or CD47-LUΔCL, were lysed in Laemmli buffer (Sigma, S3401), boiled for 7 min and then cooled on ice. Lysates were run on NuPAGE Novex 4–12% Bis-Tris Gel (Invitrogen, NP0322BOX) and transferred to PVDF membrane (Bio-rad, 162-0177). After blocking for 1 h at room temperature in Odyssey Blocking Buffer (Li-Cor, 927-40000) the following primary antibodies were used: rabbit anti-HuR (Millipore, 07-1735), rabbit anti-SET (Abcam, ab181990), mouse anti-RAC1 (Cell Signaling Technology, 8631S),mouse anti-ACTB (Sigma, A4700), rabbit anti-ACTB (Sigma, A2066), chicken anti-GFP (Abcam, ab13970), mouse anti-CD47 (SantaCruz, sc-59079),mouse anti-CD44 (BDBioscience, 561858), and rabbit anti-TSPAN13 (GeneTex, GTX52155). The antibodies were diluted in Odyssey Blocking Buffer containing 0.1% Tween 20 and the blots were incubated overnight at 4°C. After four washes in PBST (PBS plus 0.1% Tween 20) the blots were incubated for 1 h at room temperature in Odyssey

Blocking Buffer containing 0.1% Tween 20 and 0.01% SDS and the following secondary antibodies: donkey anti-mouse IRDye 700 (Rockland Immunochemicals, 610-730-002), donkey antirabbit IRDye 680 (Li-Cor Biosciences, 926-68073), donkey anti-rabbit IRDye 800 (Li-Cor Biosciences, 926-32213), donkey anti-mouse IRDye 800 (Li-Cor Biosciences, 926-32212) and rabbit anti-chicken IRDye 800 (Rockland Immunochemicals, 603-432-002). The blots were washed four times in PBST and then thoroughly soaked in PBS before imaging. Imaging was performed on an Odyssey CLx imaging system (Li-Cor). Quantification of western blots was performed using Image J.

Co-immunoprecipitation

[00139] CD47-SU, CD47-LU, CD47-LUΔC and CD47- LUΔCL were transfected into HEK293 cells and the cells were lysed in ice cold RIPA buffer ((25mMTris-HCl (pH 7.4), 150mM NaCl, 1% NP-40, 1% Na-deoxycholate, 0.1% SDS, 1 mM EDTA, protease inhibitor cocktail (Roche, 04693124001)) for 5 min on ice. After the cells were spun down at 20,000g for 20 min the supernatant was pre-cleared as described earlier. The lysate was divided in two equal parts (a small portion was removed to be used as the input control). Anti-GFP (Invitrogen, A-6455) or IgG (Santa Cruz, sc-2025) was added to the lysates and a 1 h rocking incubation at 4°C was performed, followed by addition of magnetic protein A beads (Millipore, LSKMAGA10) and another 1 h rocking incubation at 4°C. After seven washes in RIPA buffer the beads were then boiled in Laemmli buffer (Sigma, S3401) for 7 min and then cooled on ice. Western blotting was carried out as described earlier. Chicken anti-GFP (Abcam, ab13970) antibody was used to confirm immunoprecipitation of GFP constructs and rabbit anti-SET (Abcam, ab181990) antibody was used to assess co-immunoprecipitation of SET protein.

Immunoprecipitation of RAC1–GTP

[00140] CD47-SU, CD47-LU and pcDNA 3.1 vector alone were transfected into HEK293 cells. The levels of RAC1–GTP were assessed using an Active Rac1 Detection Kit (Cell Signaling Technology, 8815) following the manufacturer's protocol. Briefly, cells were lysed in ice cold lysis buffer and centrifuged at 16,000g for 15 min at 4°C. The supernatant was added to GST-PAK1-PBD and glutathione resin and incubated with rocking for 1 h at 4°C (a small portion was removed to be used as the input control). The resin was washed three times with lysis buffer and then SDS sample buffer was added to elute the bound proteins. Western

blotting was carried out as described earlier. Mouse anti-RAC1 (Cell signaling Technology, 8631S) antibody was used to detect RAC1– GTP as well as to assess total RAC1 in the samples.

Irradiation and cell survival assay

[00141] Jurkat, JinB8, and transfected JinB8 cells (24 h post-transfection) were irradiated for a total of 0, 5, 10, 20 or 40 Gy using a Shepherd Mark-1 caesium irradiator. For JinB8 or Jurkat cells transfected with CD47-LU, CD47-SU or with sh2 CD47-LU the percentage of GFP^+ cells was determined by FACS before irradiation. Three days after γ-irradiation, cells were stained with TO-PRO3 (Sloan Kettering Institute, Flow Cytometry Core Facility) in FACS buffer A (0.5% FBS in PBS) for 10 min at 4°C and analysed using a BD FACSCalibur cell analyser to evaluate cell survival. Percent survival was calculated as the TO-PRO3-negative cells divided by the total number of GFP^+ cells. Shown are mean and standard deviation of three biological replicates.

Phagocytosis assay

[00142] Human macrophages were obtained by differentiation of THP-1 cells with 25 $ngml^{-1}$ phorbol 12-myristate 13-acetate (PMA; Sigma) for 3 days. On day 3, Jurkat, JinB8, or transfected JinB8 cells were treated with 10 μg ml^{-1} Mitomycin C for 2.5 h at 37°C and washed three times in media. Mitocycin C treatment halts cell division and allows for a more accurate assessment of the percentage of cells that are phagocytosed. For JinB8 cells transfected with CD47-LU or CD47-SU the percentage of GFP^+ cells was determined by FACS before co-culture. The cells were either cultured alone or co-cultured with fully differentiated macrophages. After 3 days, cells were counted and the fraction of GFP^+ cells was determined by FACS analysis. The fraction of surviving cells after co-culture with macrophages was normalized by the number of surviving cells without co-culture and is show as mean and standard deviation of three independent experiments.

[00143] To demonstrate that the cells were phagocytosed, during the last 10 min of Mitomycin C treatment the cells were also labelled with carboxyfluorescein succinimidyl ester (CFSE; Invitrogen). Washing after Mitomycin C and CFSE treatment was carried out in cold media according to the manufacturer's protocol. After co-culture, CFSE uptake by

macrophages was measured by FACS analysis to demonstrate that a decrease in the number of surviving cells is due to phagocytosis by macrophages.

Fraction of membrane proteins among HuR target genes

[00144] The list of HuR target genes was obtained from previous publications (Lebedeva, S. *et al. Mol Cell* 43, 340-352 (2011)); Uren et al. *J. Biol. Chem.* **286**, 37063–37066 (2011)). The union of genes from both publications was analysed using gene ontology analysis (Huang et al. *Nature Protocols* **4**,44–57 (2009)) and all genes with the tag ''membrane'' were considered membrane proteins. This number is consistent with the number of membrane proteins obtained in yeast (Stagljar et al. *Trends Biochem. Sci.* **27**, 559–563 (2002)). Fisher's exact test was used to test for significance.

Statistical analysis

[00145] To test for significant differences between samples a two-sided *t*-test for independent samples was performed using SPSS.

Example 1

The long 3'UTR of CD47 localizes CD47 protein to the cell surface

[00146] Immunofluorescence confocal microscopy and FACS analysis were used to assess CD47 protein expression in several nonpermeabilized and permeabilized human cell types. As shown in Figure 1a, CD47 was found to be expressed not only on the cell surface but also intracellularly (Figure 1a and Figure 5a-c). 3' sequence analysis of naïve B cells led to identification of two alternative *CD47* mRNA isoforms generated by ApA, which differed in the length of the 3'UTR (Fig. 1b). The two 3'UTR isoforms were also detected in multiple additional cell lines (U2OS, MCF7, HeLa, HEK293, NTERA2, B-LCL, and Toledo) by Northern blot analysis (Fig. 1c). The shorter 3'UTR contained 221 nt, whereas the longer 3'UTR contained 4194 nt. shRNA-mediated knock-down (KD) of the longer 3'UTR isoform resulted in a decrease of CD47 surface expression without a change in intracellular expression when compared to the effects of a control shRNA (sh Co) (Fig. 1d and Fig. 5d-g). These results demonstrate that the long 3'UTR isoform primarily facilitates cell surface localization of CD47 protein.

Example 2

Long 3'UTR isoform of CD47 encodes information necessary for cell surface expression of proteins

[00147] Based on the findings shown in Example 1, it was hypothesized that 3'UTR of CD47 can facilitate cell surface localization of proteins other than CD47. To test this hypothesis, green fluorescent protein (GFP), which is not a membrane associated protein, encoded by an mRNA containing the long or the short 3'UTR of *CD47* was generated and tested for its ability to localize to the plasma membrane. Membrane proteins are translated by ribosomes associated with the ER and require a signal peptide (S) at the N- terminus and a transmembrane domain (TMD). In order to allow GFP to enter the secretory pathway, the extracellular domain (ECD) of CD47 was replaced with GFP, while CD47 signal peptide, TMDs and C-terminus were preserved. This newly generated GFP was referred to as GFP-TM (Fig. 1e). To express GFP-TM exclusively with a long 3'UTR, the proximal polyadenylation signal of *CD47* was mutated (Fig. 5h). As shown in Figure 1f, GFP-TM encoded by an mRNA containing the long 3'UTR of *CD47* (GFP-TM-LU) localized primarily to the cell surface in U2OS cells, whereas GFP-TM encoded by an mRNA with the short 3'UTR of *CD47* (GFP-TM-SU) was localized intracellularly.

[00148] Co-staining with the endoplasmic reticulum (ER) marker Calnexin (Seo et al. *Cell* **104**, 119–130 (2001)) revealed that intracellular GFP is located in the ER (Fig. 1f, lower panel). The localization results were confirmed by FACS analysis, using an anti- GFP antibody on permeabilized and non-permeabilized U2OS cells to measure total and surface GFP levels, respectively (Fig. 1g). RNA-fluorescence *in situ* hybridization (FISH) against GFP further showed that the GFP transcripts either produced by the LU or SU isoforms show a similar distribution near the perinuclear ER, indicating that the localization step occurs at the protein level (Fig. 1h). These findings demonstrated that the long 3'UTR of *CD47* encodes information that is necessary for cell surface expression of GFP-TM protein, in a manner independent of RNA localization.

Example 3

HuR, SET, and RAC1 mediate 3' UTR-dependent protein localization

[00149] In order to test whether HuR plays a role in 3'UTR-dependent protein localization (UDPL), shRNAs was used to knockdown (KD) HuR in HEK293 cells (Fig. 6a). HuR KD did not affect *CD47* mRNA or 3'UTR isoform levels (Fig. 5d, 6b), nor did it affect total CD47 protein levels (Fig. 2b, bottom panel, Fig. 6a, 6c). However, strikingly, KD of HuR reduced CD47 cell surface expression (Fig. 2b, top panel, Fig. 2c). Accordingly, these observations show that in the case of CD47, HuR neither affects mRNA stability nor translation, but instead mediates CD47 protein localization post-translationally.

[00150] HuR interacts with a number of proteins through protein-protein interaction, including SET. SET further interacts with Rac1, which results in translocation of SET to the plasma membrane upon Rac1 activation. To test the involvement of SET1 and Rac1 in UDPL, both proteins were knocked down (separately) in HEK293 cells. KD of SET or Rac1 by shRNAs reduced surface expression of CD47 without affecting overall CD47 levels (Fig. 2b, Fig. 6c, 6d).

[00151] Collectively, these results reveal the further requirement for HuR, SET1, and RAC1 in 3'UTR-dependent protein localization.

Example 4

3' UTR-dependent protein localization is a widespread phenomenon

[00152] To determine if UDPL is a widespread phenomenon, the inventors of the present disclosure examined the localization of additional HuR targets (Figs. 7 and 8) that are known to be expressed on the cell surface (Kishore, S. *et al. Nat Methods* 8, 559-564 (2011); Lebedeva, S. *et al. Mol Cell* 43, 340-352 (2011); Mukherjee, N. *et al. Mol Cell* 43, 327-339 (2011)). One of the four additional HuR targets (*TSPAN13* (SEQ ID NO 7)) only had one 3'UTR isoform, whereas three genes [*CD44, ITGA1, TNFRSF13C* (encoding BAFF receptor, BAFFR)] use ApA to generate alternative 3'UTR isoforms (Fig. 2c, top panel). As was the case for CD47, KD of HuR also decreased surface expression of all four proteins without changing their total protein levels (Fig. 2c, and Fig. 6a, f.). For the genes that generate alternative 3'UTR isoforms, GFP was fused to their corresponding TMDs and C-termini together with either their short or long 3'UTRs. The long 3'UTR of *CD44* contains a strong HuR-BS and enabled higher GFP surface expression than GFP generated from the short

3'UTR isoforms of *ITGA1* or *TNFRSF13C* which do not contain HuR-BS (Fig. 2d). For all four out of four tested cases, the longer 3'UTR increases surface localization of GFP (Fig. 1f, h, 2d). These experiments demonstrate that UDPL is a widespread phenomenon and that many surface proteins have a similar mechanism (3'UTR with HuR binding site plus HuR, SET and Rac1) to direct proteins to the cell surface.

Example 5

Uridine-rich Sequence of the HuR Binding Site is Sufficient for 3' UTR-dependent protein localization

The 3'UTR of *CD47* (SEQ ID NO 1) contains over 30 putative HuR-binding sites (HuR-BS) (Fig. 7). Next, a 3'UTR with only a few HuR-BS was tested for the ability to mediate surface localization (Fig. 7) and SEQ ID NO 3. As can be seen in Fig. 2e, indeed, a single HuR-BS containing uridine-rich sequence was sufficient for surface localization of GFP-TM. While a single HuR-BS containing uridine-rich sequence was less potent than the full-length 3'UTR of *CD47*, it still achieved 75% of the potency of the full-length 3'UTR. Additionally, deletion of the uridine-rich sequence of the HuR-BS (SEQ ID NO 4) abrogated surface expression of GFP-TM (Fig. 2e, Fig. 7). Collectively, these results demonstrate that the 74-nt region HuR-BS is sufficient for surface localization of proteins and is almost as efficient as the much longer sequenced long 3'UTR. Accordingly, incorporation of only the 74 nt fragment of the long 3'UTR of CD47 in a construct or vector according to the present disclosure comprising a polypeptide of interest will result in substantial localization of the polypeptide to the cell surface without needing to incorporate the entire long 3'UTR region of CD47. Based on the foregoing, it is believed that a segment having 90% homology with the 74-nt fragment tested would also direct a substantial fraction of an expressed polypeptide of interest to the cell surface.

Example 6

HuR/SET Containing Protein Complex is Necessary for Cell Surface Localization

[00153] Next, each step of UDPL was examined in more detail. CD47 is a known HuR target (Kishore, S. *et al. Nat Methods* 8, 559-564 (2011); Lebedeva, S. *et al. Mol Cell* 43,

340-352 (2011); Mukherjee, N. *et al. Mol Cell* 43, 327-339 (2011)). RNA immunoprecipitation (RNA-IP) showed that HuR binds to the long 3'UTR of *CD47*, but not to the short *CD47* 3'UTR (Fig. 3a). Furthermore, HuR was able to bind the single HuR-BS, which was abolished after mutation of the site (Fig. 3a). Previously, it has been shown that SET and HuR interact in the nucleus (Brennan et al., 2000). Here, using co-RNA-IP, the inventors show that SET also interacts with the long 3'UTR of *CD47* and that this interaction is dependent on HuR (Fig. 3c).

[00154] To test whether recruitment of SET to the short 3'UTR of *CD47* redirects GFP localization from the ER to the plasma membrane, MS2 binding sites (MS2-BS) were added to the short 3'UTR of GFP-TM (Fig. 3b). MS2-BS are derived from the bacteriophage MS2 and form RNA stem loops (Bertrand *et al. Mol Cell* 2, 437- 445 (1998)). The capsid protein of MS2 (here, called simply MS2) specifically recognizes these MS2 stem loops. Constructs were generated containing MS2 fused to mCherry and then either HuR, SET or with no further coding sequence (Fig. 3b). Co-expression of these constructs with the construct containing the short 3'UTR and MS2-BS resulted in recruitment of SET or HuR to the short 3'UTR of GFP-TM. As can be seen in Fig. 3b and 9c, HEK293 cells that expressed MS2 fused to only mCherry still localized GFP to the ER, but constructs containing MS2 fusions to HuR or SET localized GFP primarily to the cell surface. Thus, the short 3'UTR functionally tethered to either HuR or SET protein serves as a surrogate for the long 3'UTR which by containing an HuR binding site recruits HuR in turn recruiting SET and drives the expressed polypeptide of interest to the cell surface. Omitting either the MS2 or MS2-BS from the experiment abrogated surface localization (Fig. 9b) indicating that local recruitment of SET to the site of translation is necessary and sufficient for surface localization of GFP-TM. Overexpression of HuR or SET without functional tethering did not increase GFP surface expression (Fig. 9a, b). Thus, through a scaffold function of the long 3'UTR of *CD47,* the HuR/SET containing protein complex is recruited to the site of translation, which is necessary for GFP surface localization and predicts similar surface localization of any polypeptide of interest incorporated in an expression vector comprising the foregoing regulatory element(s). These results also elucidate the mechanism by which the long 3'UTR of CD47 and long 3'UTRs of other membrane proteins having a long and a short 3'UTR isoform exert their protein localization function: through binding HuR, and recruitment of SET which in turn interacts with Rac1.

Example 7

SET interacts with the CD47-LU protein but not with CD47-SU protein

[00155] Deletion of the cytoplasmic C-terminus (ΔC, amino acids 290-323; Fig. 13) of CD47 dramatically decreased GFP surface expression (Fig. 3d). Since SET was initially shown to bind to lysine residues in histone tails (Schneider et al. *J Biol Chem* 279, 23859-23862 (2004)), the influence of lysine mutations in the C-terminus of CD47 was tested. Whereas mutation of a single lysine residue decreased surface localization by up to 37% (Fig. 10a), mutation of 2/5 lysines (K297A and K304A; 2Km construct) decreased surface localization by more than 50% (Fig. 10b). Although deletion of the C-terminus decreased surface GFP expression by about 80%, it did not completely abrogate surface localization (Fig. 3d and Fig. 10). It was noticed that the first cytoplasmic loop also contains a potential SET-BS. Additional mutation of three lysines to alanines (K163A, K166A and K175A) completely abolished surface GFP expression mediated by GFP-TM with the long 3'UTR (ΔCL construct, Fig. 3d and Fig. 10b). The cytoplasmic domains of CD47 contain two separate SET-BS. In order to confirm that SET interacts with the cytoplasmic domains of CD47-LU but not CD47-SU, co-IP was performed using an anti-GFP antibody after transfection of either GFP-CD47-LU or GFP-CD47-SU into HEK293 cells (constructs are described in Fig. 4a). As shown in Fig. 3e, endogenous SET directly interacted with CD47-LU, but not with CD47-SU. The interaction was dependent on both SET-BS in the cytoplasmic domains of CD47, because impairment of both SET-BS prevented the interaction.

[00156] In conclusion, SET interacts with the CD47-LU protein to mediate 3' UTR-dependent protein localization.

Example 8

CD47-LU Localizes to the Cell Surface via UDPL mediated by active RAC1

[00157] To test if the difference in surface localization has phenotypic consequences, the ECD of CD47 was added to the GFP constructs (called CD47-LU or CD47-SU; Fig. 4a). Both constructs resulted in comparable overall CD47 protein levels (Fig. 11a). CD47-LU

efficiently localized to the cell surface via UDPL mediated by active RAC1, since expression of a dominant-negative Rac1 (N17Rac1, which has a higher affinity to Rac1-GDP) (Ridley et al. *Cell* 70, 401-410 (1992)) decreased surface localization of GFP (Fig. 4b, c, and Fig. 11b). Whereas GFP expressed from the GFP-TM-SU construct nearly completely localized to the endoplasmic reticulum (Fig. 1f–h), CD47-SU primarily localizes to the endoplasmic reticulum, but also localizes partially to the cell surface, but independently of active RAC1 (Fig. 4b, c).

[00158] These results merely indicate that protein localization is complex and involves more than one mechanism. The results do not detract from the conclusions drawn above about the role of the HuR binding site of the long 3'UTR or from the generality of the applicability of the approach of the present disclosure (through the vectors, constructs and methods disclosed herein) for directing proteins to the cell surface.

Example 9

Membrane proteins rely on 3'UTR-dependent protein localization for surface expression to varying degrees

[00159] Respective ECDs were added to CD44 and BAFFR (*TNFRSF13C* encodes the BAFF receptor, BAFFR), which increased surface expression of their SU isoforms compared with their GFP-TM isoforms, but to a lesser extent than was observed for CD47 (Fig. 11c, d). Di- or multimerization of cell surface receptor subunits, which often occurs through their ECDs, is a common strategy for overcoming endoplasmic reticulum retention, because it results in masking of endoplasmic reticulum retention signals (Zerangue et al., *Neuron,* 22, 537–548 (1999)). It is possible that CD47-SU, CD44-SU and BAFFR-SU might use such a mechanism (although the multimerization partners are unknown) for their partial surface expression. In the case of BAFFR the ECD only increased surface expression by 1.2-fold, indicating that BAFFR strongly depends on UDPL for surface expression (Fig. 11d). This is supported by the absence of BAFFR on B cells in Rac-deficient mice (Walmsley et al. *Science* 302, 459–462 (2003)). Taken together, the data presented in this example suggests that membrane proteins rely on UDPL for surface expression to varying degrees.

Example 10

CD47 protein has different functions regarding cell survival depending on whether it was generated by the short or long 3'UTR isoform

[00160] Next, the inventors tested whether CD47 is expressed in the ER ($CD47_{ER}$) due to mis-localization or if $CD47_{ER}$ actually has an independent intracellular function. All cell types analyzed expressed higher amounts of the short 3'UTR isoform of CD47 (Fig. 1c), which is associated with ER expression.

[00161] Cells with high CD47 surface levels are protected from phagocytosis by macrophages (Fig. 11e) (An et al. *Cell* 134, 175-187 (2008). To examine if the difference in surface expression of CD47-LU and CD47-SU protects cells to a different extent from phagocytosis, similar total amounts of CD47-LU or CD47-SU were expressed in CD47-deficient Jurkat cells (called JinB8 cells). Co-culture of these cells with macrophages demonstrated that CD47-LU fully protected the cells, whereas CD47-SU only partially protected the cells from phagocytosis (Fig. 4d).

[00162] CD47 also functions in the regulation of apoptosis (Ridley et al., *Cell* 70, 401-410 (1992)). Expression of GFP-CD47-SU (Fig. 4a) in JinB8 cells primarily leads to expression of $CD47_{ER}$ (Fig. 4b, left panel) and resulted in increased cell death after γ-irradiation, such that it rescued the loss-of-apoptosis phenotype of CD47 deficient cells (Fig. 4e,). GFP-CD47-LU (Fig. 4a) lead to comparable CD47 protein expression (Fig. 4b, Fig. 11a), but produced exclusively surface CD47 (Fig. 4b, right panel). GFP-CD47-LU did not rescue the loss-of-apoptosis phenotype of JinB8 cells (Fig. 4e, Fig. 11f).

This was also shown after KD of the long 3'UTR isoform of CD47 in Jurkat cells, where loss of surface CD47 did not reduce cell death after γ-irradiation (Fig. 11g). On the contrary, GFP-CD47-LU fully protected cells from phagocytosis by macrophages (Fig. 4d). GFPCD47-SU is to some extent expressed on the cell surface (Fig. 4b). As a result, GFP-CD47-SU is able to partially rescue cells from phagocytosis (Fig. 4d). These results confirm that the extent of phagocytosis depends on the expression level of surface CD47 (Fig. 11e). These results indicate that CD47 protein has opposite functions regarding cell survival depending on whether it was generated by the short or long 3'UTR isoform, with surface CD47 supporting cell survival and $CD47_{ER}$ being detrimental to cell survival.

Example 11

CD47 protein localized to the same cellular compartment, but produced by the SU or LU mRNA isoforms, can have different biological functions

[00163] As the surface localization of CD47-SU is RAC1-independent, it also did not co-localize with RAC1 at the plasma membrane (Fig. 4f). In contrast, CD47-LU showed strong co-localization with RAC1 at lamellipodia (Fig. 4f). Both activated RAC1 and CD47 are necessary for efficient cell migration and activated RAC1 localizes to the leading edge of migrating cells (ten Klooster et al. *EMBO J.* 26, 336–345 (2007); Lindberg et al. *Science* 274, 795–798 (1996); Frazier et al. *UCSD Nature Molecule Pages* (University of California, San Diego, 2010)). Here, the inventors showed that only the expression of CD47-LU resulted in changes in cell morphology with the generation of lamellipodia at the leading edge of cells. Furthermore, CD47-LU, but not CD47-SU, resulted in increased active RAC1 (Fig. 4g), which suggests that CD47-LU may cooperate with RAC1 during cell migration. Thus, CD47 protein localized to the same cellular compartment, but produced by the *SU* or *LU* mRNA isoforms, can exert different functions.

Example 12

Artificial 3'UTR Sequences Effective In Promoting Surface Localization

[00164] Next, the inventors designed artificial sequences that contain fewer nucleotides but still retain the regulatory elements that facilitate surface localization. Three artificial UTR sequences, which were able to mediate increased protein surface localization are listed in Table 2.

[00165] Artificial UTR-1 comprises the HuR-binding site from TNF-α. Previous studies have shown that HuR binds AU-rich elements, where the AU-rich element can have a core sequence of AUUUA or UUAUUUA(U/A)(U/A). Furthermore, it is known that binding of HuR to such an element regulates mRNA stability (in either a positive or negative manner, dependent on the context). In addition to HuR-binding site from TNF-α, the artificial UTR-1 sequence also comprises GC-rich sequence at each end of. The GC-rich sequences are restriction sites, but they make sure that the AU-rich sequence in between is accessible. UTR-

1 also contains a CAC sequence derived from TNF alpha which the inventors also considered important for surface localization.

[00166] Additionally, the inventors carried out motif analysis using the search engine HOMER (available from the Salk Institute for Biological Studies). The inventors searched for the 8-mers that are bound by RNA-binding proteins. The motif analysis demonstrated that CACACA and GAGAGA sequences may also be important for surface localization (Figure 15A). In consideration of the results of the motif analysis, the inventors designed artificial UTR-2, which contains restriction sites, but also endogenous elements from the *CD47* 3'UTR that resemble CACACA and GAGAGA, flanking the HuR binding element. Finally, the inventors also generated artificial UTR-3, which comprises the HuR binding element of TNF-α as well as a piece of the *CD47* 3'UTR. Table 2 provides UTR-1, UTR-2, and UTR-3 sequences, where each sequence element and/or fragment within UTR-1, UTR-2, and UTR-3 is marked using normal letters, underlined letters, capital letters, and bold capital letters (see below).

Table 2. Artificial UTR Sequences

Artificial UTR Name/ SEQ ID NO	Artificial UTR Sequence
UTR-1 (SEQ ID NO: 56)	ggccgc**CACTTGTGATTATTTATTATTTATTTATTATTTATT** **TATTTA**gggcc
UTR-2 (SEQ ID NO: 57)	ggccgcgggtgagcttgagagtCACA CA**TTGTGATTATTTATTATT** **TATTTATTATTTATTTATTTA**cttaaaagtgttgttatcatgactggtg agagaagaaaacaCCAACCTCt
UTR-3 (SEQ ID NO:	gcggccgcctcctgcatggcaacaaaatgtgtgtcaccatcaggccaa caggccagcccttgaatggggatttattactgttgtatctatgttgcat gataaacattcatcaccttcctcctgtagtcctgcctcgtactcccctt cccctatgattgaaaagtaaacaaaacccacatttcctatcctggtta

58)	gaagaaaattaatgttctgacagttgtgatcgcctggagtactttta gacttttagcattcgtttttacctgtttgtggatgtgtgtttgtatg tgcatacgtatgagataggcacatgcatcttctgtatggactctaga**CACT TGTGATTATTTATTATTTATTTATTATTTATTTATTTA**gggccc

[00167] Where sequences in bold capital letters indicate sequence comprising HuR-binding site from TNF-α; sequences in normal lower-case letters correspond to restriction sites; underlined lower-case sequences represent sequences found in the long 3'UTR of CD47 (UTR-2 comprises nucleotides 2798-2813 and 3558-3598 of long 3'UTR of CD47, while UTR-3 comprises nucleotides 996-1317 of long 3'UTR of CD47); and normal capital letters (non-bold capital letters) correspond to artificial CACA and CCAACCTC sequences.

[00168] Next, the inventors tested the effect of each UTR-1, UTR-2, and UTR-3 on CD47 surface localization. In order to confirm that UTR sequences of the present disclosure can be used for surface localization of various proteins in diverse cell types, the effect of each UTR sequence (UTR-1, UTR-2, and UTR-3) was tested in HeLa, U2OS, and HEK293 cells (Figure 15B). Briefly, HeLa, U2OS, or HEK293 cells were transfected with GFP-CD47 expression constructs contained either the short 3'UTR (negative control, NEG CO), the long 3'UTR (positive control, POS CO), or one of the three artificial 3'UTRs (UTR-1, UTR-2, and UTR-3). Surface GFP was measured using FACS analysis. As shown in Figures 15B (HeLa cells), 15C (U2OS), and 15D (HEK293), each artificial 3'UTR sequence (UTR-1, UTR-2, and UTR-3) was capable of increasing protein localization to the cell surface. Furthermore, while the effect of each UTR-1, UTR-2, and UTR-3 is comparable to that one of the long 3'UTR of *CD47*, it is possible that potencies for other membrane proteins or for different cell types can be further optimized for each of UTR-1, UTR-2, and UTR-3.

[00169] Taking into account that HuR is highly conserved across species, including drosophila, c. elegans, mouse, and human, and that it binds to AU rich regions, the use of polynucleotide sequences of the present disclosure is not limited to human cells, and can be used for surface localization of proteins in other organisms, as well as for localizing onto the surface non-human proteins. Specifically, the HuR-BS of TNF-α is completely conserved between mouse and human. Thus, the artificial UTR sequences of the present disclosure, which comprise the HuR-BS of TNF-α can be used in mouse cells or for surface localization of mouse proteins.

Example 13

Testing additional segments of the 3'UTR sequences for the ability to facilitate cell surface expression of proteins

[00170] As described in Example 5, the 74-nt region was identified containing HuR binding site (HuR-BS), which was successful in effecting surface localization of proteins. In this Example, additional 3'UTR sequences, multiple HuR-BS, were evaluated. Specifically, sequence comprising 2xHuR-BS (SEQ ID NO 50) was tested according to the experimental design explained in Example 5. Briefly, green fluorescent protein (GFP), encoded by an mRNA containing 2xHuR-BS (Figure 14) was generated. Furthermore, in order to allow GFP to enter the secretory pathway, a signal peptide, the transmembrane domain and C-terminus of a transmembrane protein was incorporated into the vector engineered to express GFP-TM-containing 2xHuR-BS. The resulting vector was tested for its ability to localize to the plasma membrane in numerous cell types, including HEK293 cells. The inventors did not observe an increase in surface localization of protein when two tandem HuR-BS (2xHuR-BS) were used instead of one. Similarly, when inventors tested the ability of more than two consecutive HuR-BS sequences (4xHuR-BS or 6xHuR-BS), no increase in surface localization was observed. As part of the ongoing experiments, the inventors will also test 2x HuR-BS with spacer (SEQ ID NO 51) for its ability to increase cell surface localization.

Prophetic Example

[00171] In this Example, additional 3'UTR sequences, including sequences found within other transmembrane proteins will be evaluated. For example, 3'UTR sequences found within various other transmembrane proteins, including CXCR4 (SEQ ID NO 52), FAS (SEQ ID NO 53), TNF (SEQ ID NO 54), and PDGFA (SEQ ID NO 55) will be evaluated in the similar manner (see Figure 14). Shorter fragments of 3'UTR sequences of transmembrane proteins listed in Figure 14 will also be evaluated for the ability to localize proteins to the plasma membrane. It is anticipated that fragments of these UTR sequences containing an HuR binding domain and optionally additional elements based on the foregoing motif analysis will be effective in achieving increased localization of proteins of interest on a cell surface.

All documents cited herein are incorporated by reference in their entirety for all purposes.

What is claimed is:

1. A recombinant vector comprising a promoter, the vector being adapted for introduction of a first polynucleotide encoding a first polypeptide of interest to be expressed in a target cell and localized on the surface of the target cell and a second polynucleotide segment comprising a fragment of a longer 3'UTR or nucleic acid encoding a fragment of a longer 3'UTR, the fragment derived from nucleic acid encoding a second polypeptide that is substantially localized to the surface of a cell in which it is expressed in nature, said fragment comprising an HuR binding site.

2. The vector of claim 1 wherein the polynucleotide segment is flanked by restriction sites.

3. The vector of claim 2 wherein a CAC motif is interposed between a first (5') restriction site and the fragment.

4. The vector of claim 1wherein the fragment is up to about 500 or up to about 1000 nucleotides long.

5. The vector of claim 1 wherein the segment has a sequence selected from the group consisting of SEQ ID NO:3 ; SEQ ID NO: 56; SEQ ID NO: 57and SEQ ID NO: 58 or a sequence at least 90% homologous to any of the foregoing.

6. The vector of claim 1 wherein the vector further comprises an insert containing an exogenous polynucleotide encoding the polypeptide of interest, wherein upon delivery of the vector to a target cell, the polypeptide is expressed and substantially localized on the target cell surface.

7. The vector of claim 6 wherein the first polynucleotide encoding the first polypeptide of interest is or is encoded by an open reading frame (ORF) in turn encoding a polypeptide normally expressed on a cell surface.

8. The vector of claim 6 wherein the polynucleotide encodes a polypeptide that is not normally substantially localized on the target cell surface.

9. The vector of claim 6, said insert further comprising a polynucleotide encoding a signal peptide and a polynucleotide encoding a transmembrane domain derived from a polypeptide expressed on a cell surface.

10. The vector of claim 9 the insert also comprising a polynucleotide encoding a C-terminal of a polypeptide that is normally expressed on the target cell or a fragment thereof encoding a SET binding site.

11. The vector of claim 9 wherein the transmembrane domain also includes a SET binding site.

12. The vector of claim 9 wherein one or both of the signal peptide and transmembrane domain are derived from CD47.

13. The vector of claim 9 wherein the transmembrane domain includes the entire transmembrane domain region of CD47.

14. The vector of claim 9 wherein said polypeptide of interest is GFP.

15. The vector of claim 1 wherein said first polypeptide of interest is selected from the group consisting of taste receptors and glutamate receptors.

16. The vector of claim 1wherein said first polypeptide of interest is selected from the group consisting of INSR, IGFR, TGFBR1, EPHB2, BMPR1A, EGFR, FZD5, CD44, and PDGFR.

17. The vector of claim 6 wherein the insert comprises the polynucleotide encoding the signal peptide of CD47 followed by the polynucleotide encoding the polypeptide of interest followed by the transmembrane domain of CD47 including the C-terminal thereof and said insert is 5' to the 3'UTR.

18. The vector of claim 1 wherein said promoter is selected from the group consisting of ubiquitous promoters or tissue-specific promoters.

19. The vector of claim 1 wherein said vector is selected from the group consisting of plasmids and viral vectors.

20. The vector of claim 19 wherein the vector is a viral vector selected from the group consisting of selected from the group consisting of an adenoviral vector, an adeno associated vector (AAV), a baculoviral vector and a retroviral vector.

21. The vector of claim1 wherein the promoter is part of an expression cassette comprised by the vector.

22. The vector of claim 2 wherein the fragment is preceded by a CA-rich motif or a GA-rich motif or an AU-rich sequence interposed between the fragment and a first (5') restriction site.

23. A recombinant polynucleotide construct comprising (i) a first segment having a sequence encoding a polypeptide that is not normally substantially localized on the surface of a target cell, (ii) a second heterologous segment having a sequence that comprises a fragment of a 3'UTR comprising an HuR binding site or nucleic acid encoding said fragment, wherein said construct has the property of effecting localization of at least a substantial fraction of said polypeptide to the cell surface upon expression of the polypeptide in a cell harboring the construct.

24. The construct of claim 23 wherein the polypeptide comprises a transmembrane domain sequence and a peptide signal sequence of a membrane-associated polypeptide.

25. The construct of claim 24 wherein the transmembrane domain includes the C-terminal of said membrane-associated polypeptide or at least a fragment of said C-terminal encoding a SET binding site.

26. The polynucleotide construct of claim 23 wherein the 3'UTR fragment has a sequence comprising a nucleic acid having SEQ ID NO: 3 or SEQ ID NO: 56 or SEQ ID NO: 57 or SEQ ID NO: 58 or a sequence having at least 90% homology to any of the foregoing.

27. A polynucleotide construct comprising the sequence SEQ ID NO 3, SEQ ID NO 56, SEQ ID NO 57 and SEQ ID NO: 58 or a sequence having at least 90% homology to any of the foregoing flanked between two restriction sites.

28. The construct of claim 27 further comprising at least one of a CA-rich motif, a GA-rich motif and an AU-rich motif following the first restriction site.

29. A method for increasing a fraction of a polypeptide of interest expressed within a cell that is localized on the surface of the cell, comprising transfecting the cell with the vector of any one of claims 1-22 under conditions that permit expression of the polypeptide, wherein HuR, SET and RAC1 are present within the cell, have access to the polypeptide at least post-translationally, and effect localization of the polypeptide to the cell surface.

30. A method for increasing a fraction of a polypeptide expressed within a cell that is transported to the surface of the cell, comprising introducing in the cell a the construct of any one of claims 23-28 under conditions that permit expression of the polypeptide, wherein HuR, SET and RAC1 are present within the cell, have access to the polypeptide at least post-translationally, and effect localization of the polypeptide to the cell surface.

Figure 1.

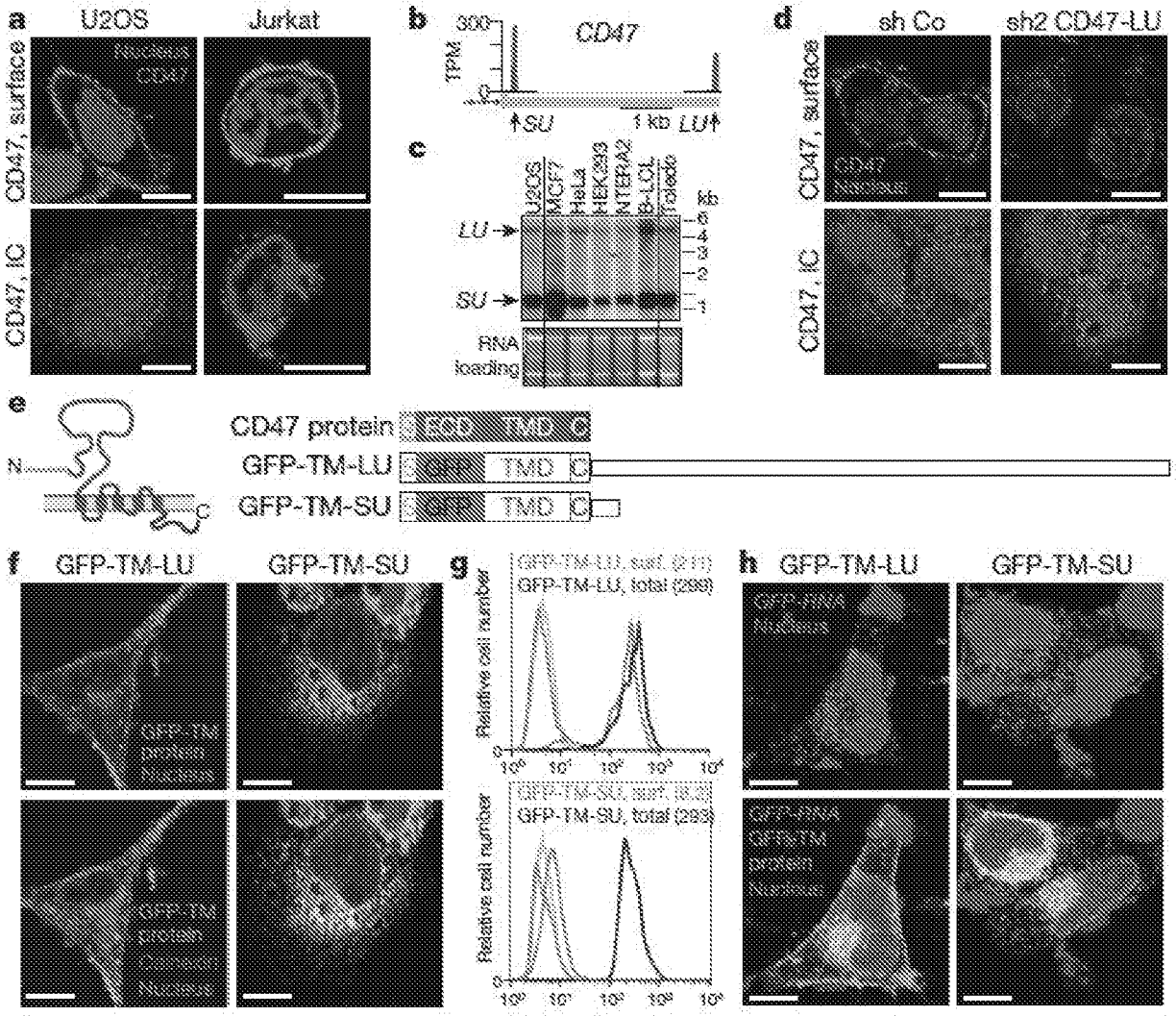

a
U2OS
Jurkat
CD47, surface
CD47, IC
b
CD47
TPM
SU
1 kb
LU
c
U2OS
MCF7
HeLa
HEK293
NTERA2
B-LCL
Raji
LU
SU
RNA loading
d
sh Co
sh2 CD47-LU
CD47, surface
CD47, IC
e
N
C
CD47 protein
GFP-TM-LU
GFP-TM-SU
TMD
f
GFP-TM-LU
GFP-TM-SU
g
GFP-TM-LU, total (299)
Relative cell number
GFP-TM-SU, total (283)
h
GFP-TM-LU
GFP-TM-SU

Figure 2.

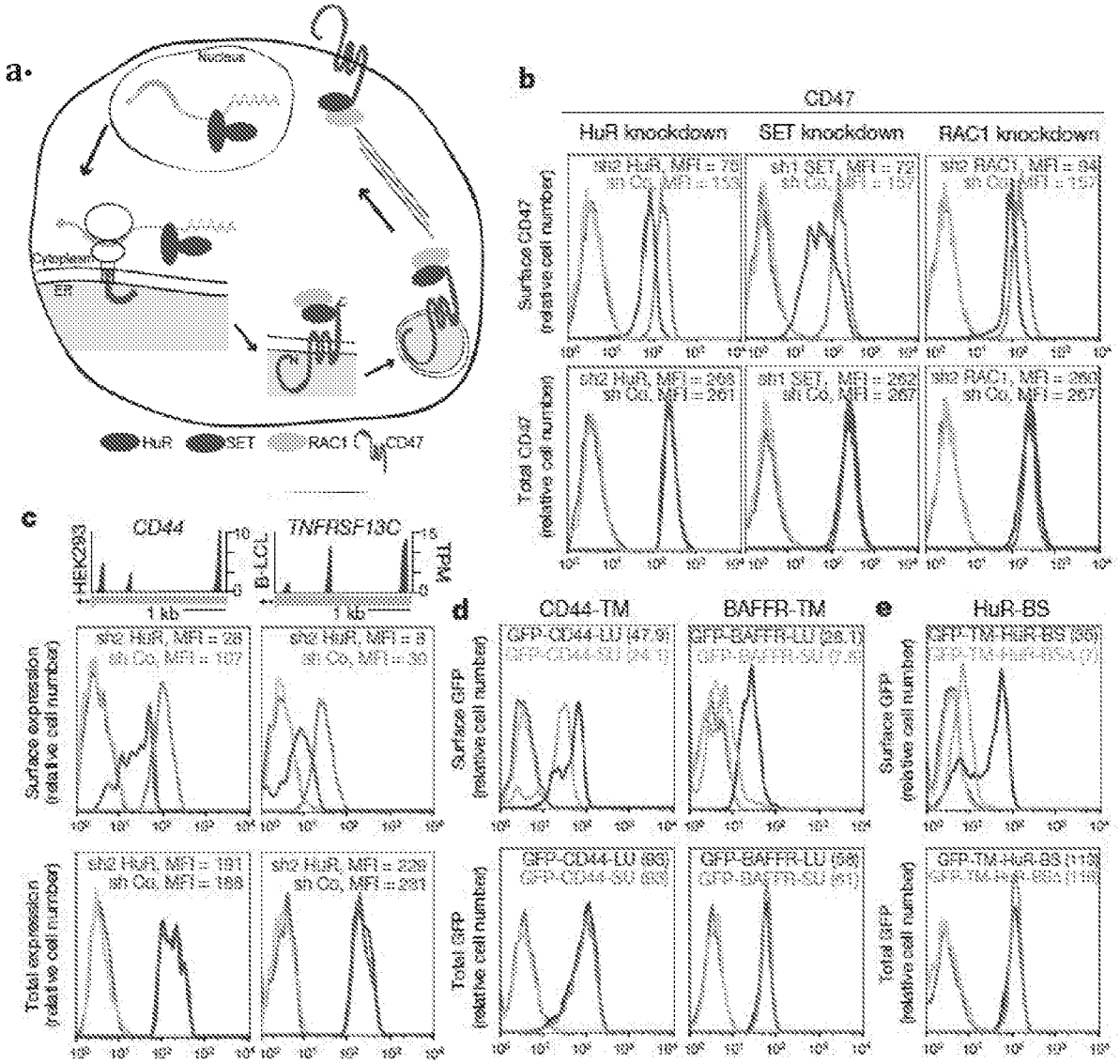

a.
Nucleus
HuR
SET
RAC1
CD47
b
CD47
HuR knockdown
SET knockdown
RAC1 knockdown
Surface CD47 (relative cell number)
Total CD47 (relative cell number)
c
CD44
TNFRSF13C
HEK293
B-LCL
1 kb
TPM
Surface expression (relative cell number)
Total expression (relative cell number)
d
CD44-TM
BAFFR-TM
Surface GFP (relative cell number)
Total GFP (relative cell number)
e
HuR-BS
Surface GFP (relative cell number)
Total GFP (relative cell number)

Figure 3.

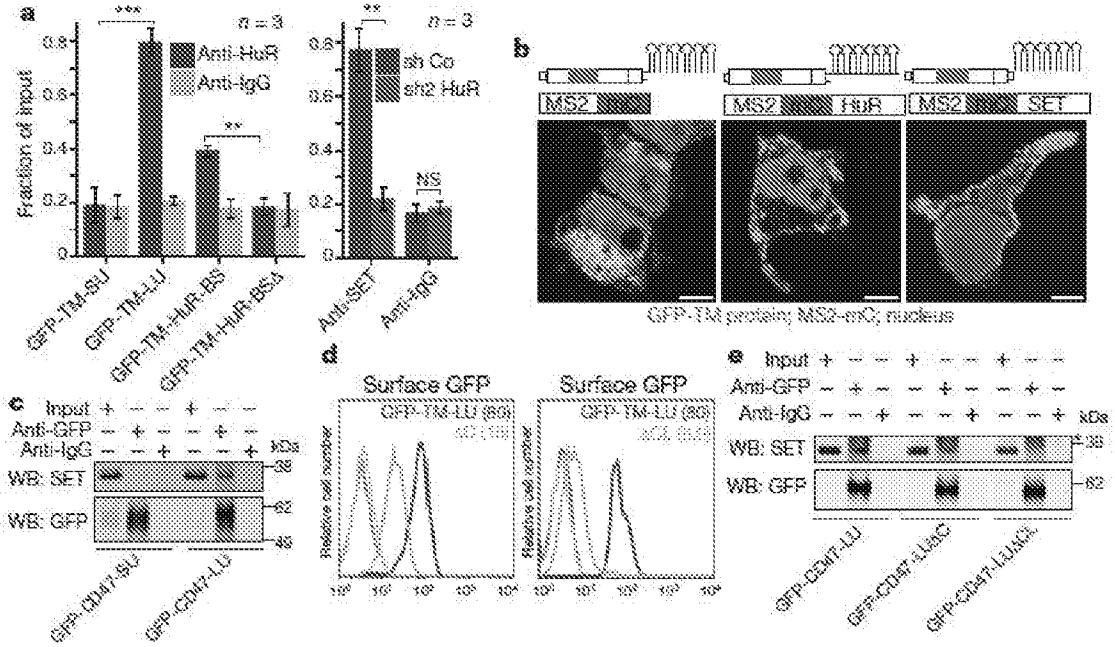

a
n = 3
Anti-HuR
Anti-IgG
Fraction of Input
GFP-TM-SU
GFP-TM-LU
GFP-TM-HuR-BS
GFP-TM-HuR-BSΔ
sh Co
sh2 HuR
Anti-SET
Anti-IgG
NS
b
MS2
HuR
SET
GFP-TM protein; MS2-mC; nucleus
c
Input
Anti-GFP
Anti-IgG
WB: SET
WB: GFP
kDa
GFP-CD47-SU
GFP-CD47-LU
d
Surface GFP
GFP-TM-LU (80)
Relative cell number
e
Input
Anti-GFP
Anti-IgG
WB: SET
WB: GFP
kDa
GFP-CD47-LU
GFP-CD47-LUΔC
GFP-CD47-LUΔCL

Figure 4.

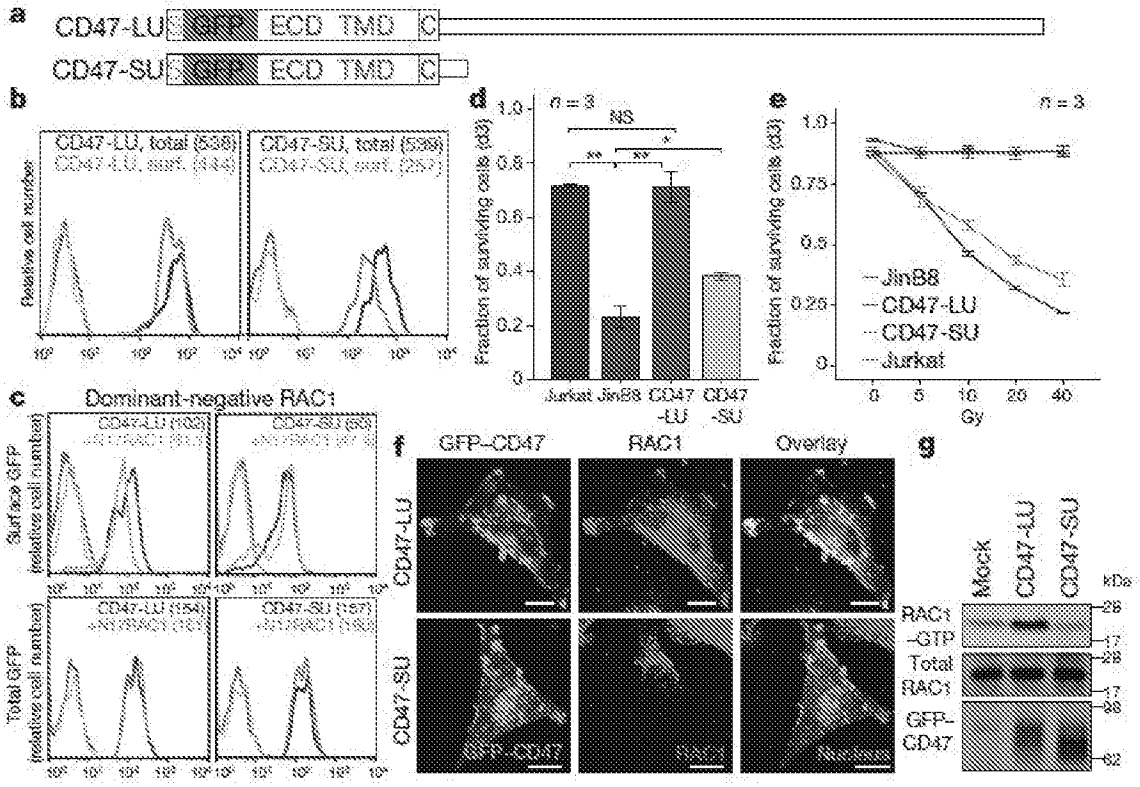

a
CD47-LU
CD47-SU
ECD
TMD
b
CD47-LU, total (538)
CD47-SU, total (539)
Relative cell number
c
Dominant-negative RAC1
Surface GFP (relative cell number)
Total GFP (relative cell number)
d
n = 3
NS
Fraction of surviving cells (d3)
Jurkat
JinB8
CD47 -LU
CD47 -SU
e
n = 3
Fraction of surviving cells (d3)
JinB8
CD47-LU
CD47-SU
Jurkat
Gy
f
GFP–CD47
RAC1
Overlay
CD47-LU
CD47-SU
g
Mock
CD47-LU
CD47-SU
kDa
RAC1 -GTP
Total RAC1
GFP-CD47

Figure 5.

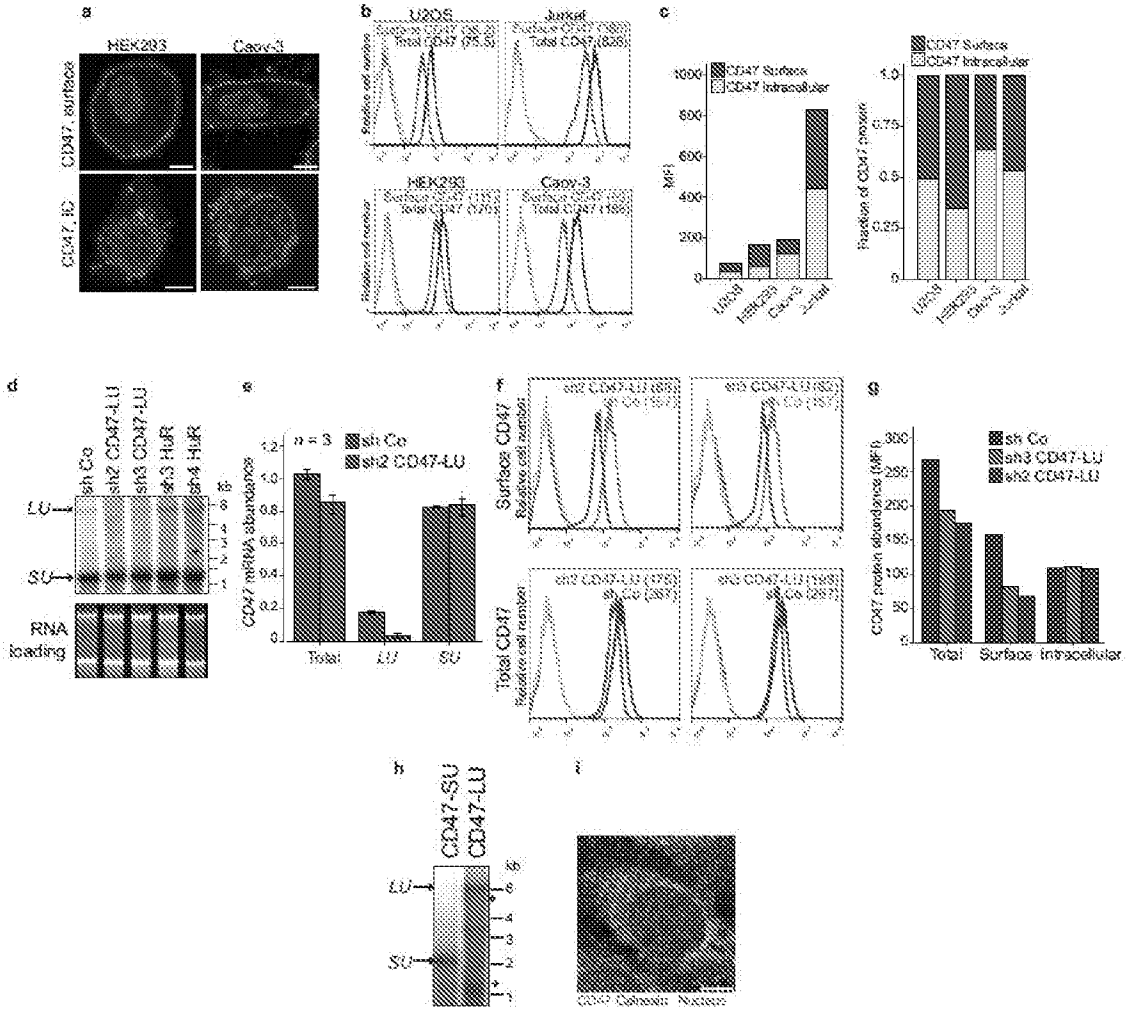

a
HEK293
Caov-3
CD47, surface
CD47, IC
b
U2OS
Jurkat
HEK293
Caov-3
c
CD47 Surface
CD47 Intracellular
MFI
Fraction of CD47 protein
U2OS
HEK293
Caov-3
Jurkat
d
sh Co
sh2 CD47-LU
sh3 CD47-LU
sh3 HuR
sh4 HuR
LU
SU
RNA loading
e
n = 3
sh Co
sh2 CD47-LU
CD47 mRNA abundance
Total
LU
SU
f
sh2 CD47-LU (63)
sh Co (167)
sh3 CD47-LU (55)
sh Co (167)
Surface CD47
Relative cell number
sh2 CD47-LU (178)
sh Co (287)
sh3 CD47-LU (158)
sh Co (287)
Total CD47
g
sh Co
sh3 CD47-LU
sh2 CD47-LU
CD47 protein abundance (MFI)
Total
Surface
Intracellular
h
CD47-SU
CD47-LU
LU
SU
i
CD47 Cytosol
Nucleus

Figure 6.

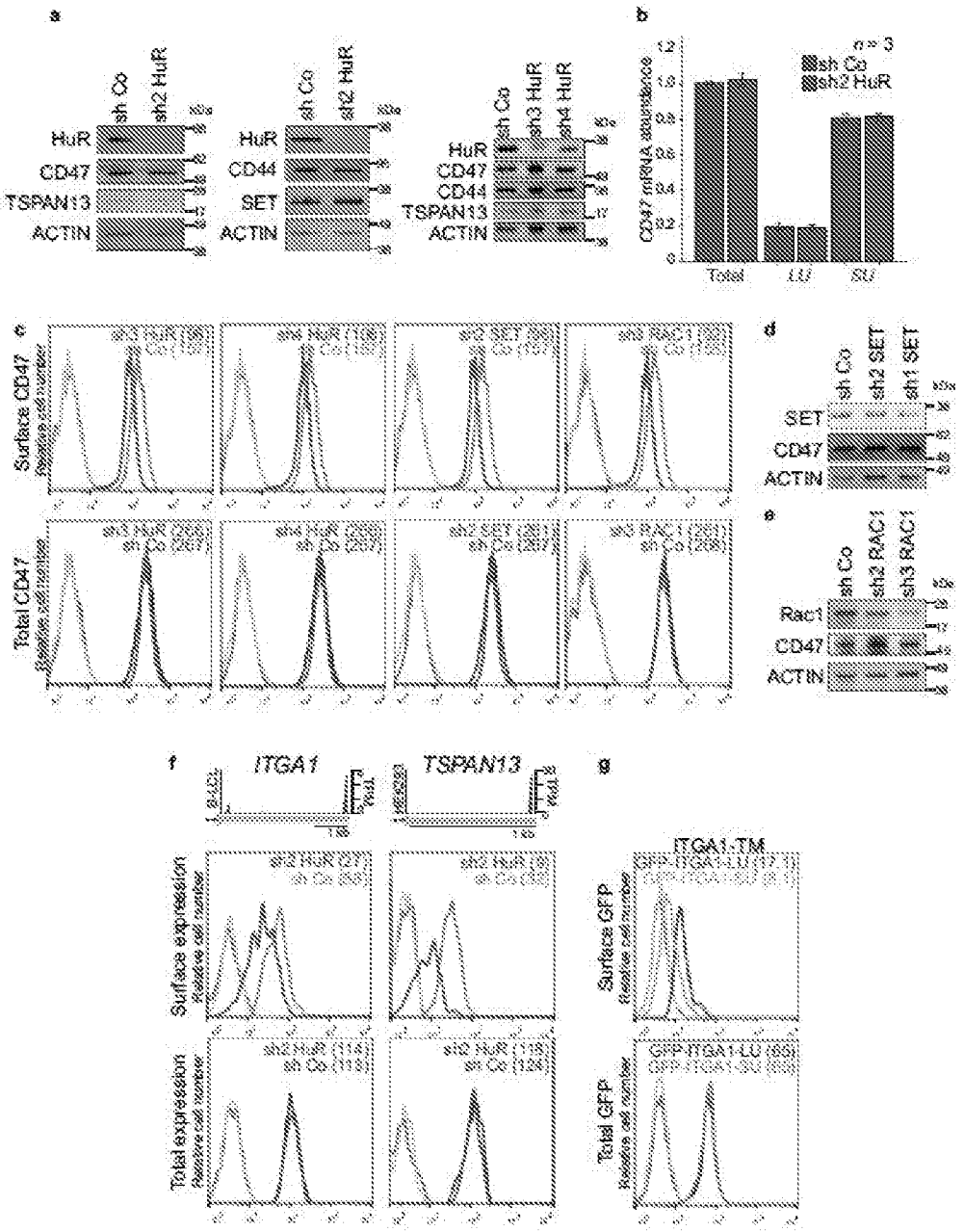
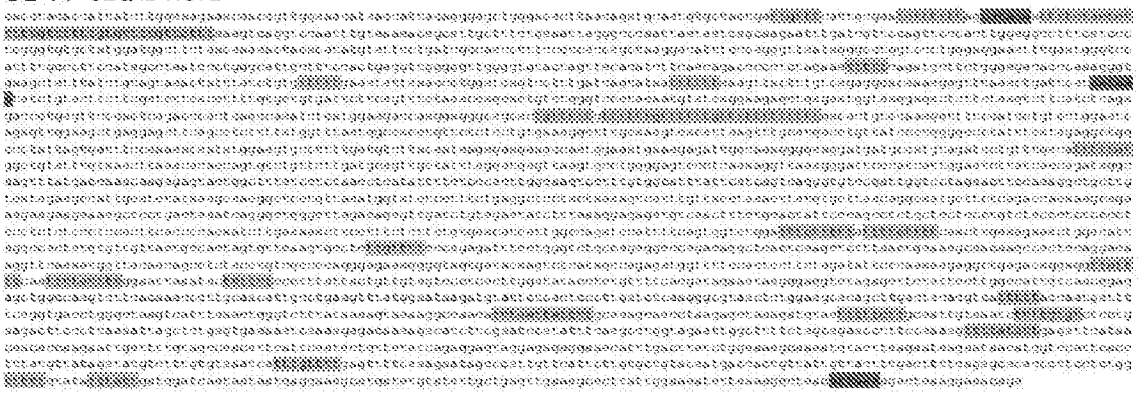

a
sh Co
sh2 HuR
HuR
CD47
TSPAN13
ACTIN
sh Co
sh2 HuR
HuR
CD44
SET
ACTIN
sh Co
sh3 HuR
sh4 HuR
HuR
CD47
CD44
TSPAN13
ACTIN
b
n = 3
sh Co
sh2 HuR
CD47 mRNA abundance
Total
LU
SU
c
Surface CD47
Total CD47
Relative cell number
d
sh Co
sh2 SET
sh1 SET
SET
CD47
ACTIN
e
sh Co
sh2 RAC1
sh3 RAC1
Rac1
CD47
ACTIN
f
ITGA1
TSPAN13
Surface expression
Total expression
Relative cell number
g
ITGA1-TM
Surface GFP
Total GFP
Relative cell number

Figure 7.

CD47 SEQ ID NO: 1

CD44 SEQ ID NO: 2



HuR-BS SEQ ID NO: 3

tttaatagggtgagcttgagag[redacted]gactaatttcacatgctctaa

HuR-BSΔ SEQ ID NO: 4

tttaatagggtgagcttgagag----------gactaatttcacatgctctaa

Figure 8.

ITGA1 SEQ ID NO: 5

TNFRSF13C SEQ ID NO: 6

TSPAN13 SEQ ID NO: 7

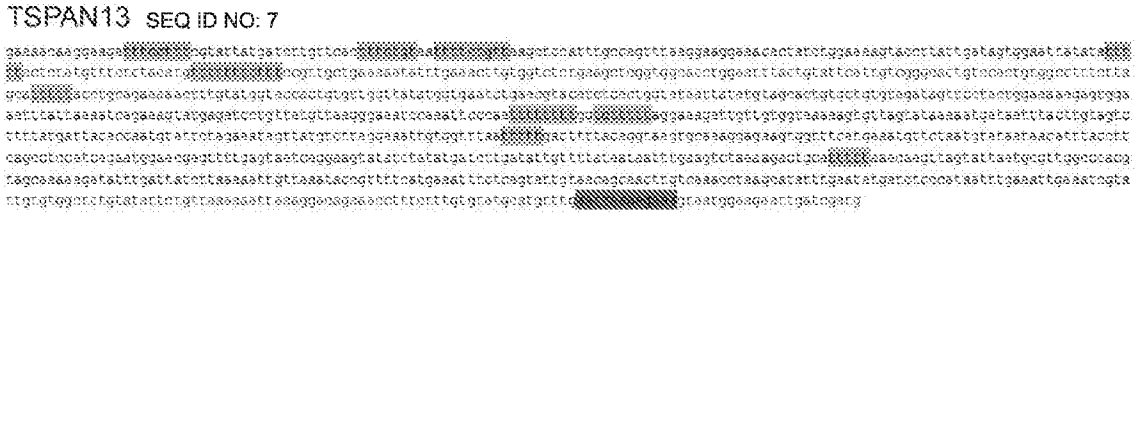


Figure 9.

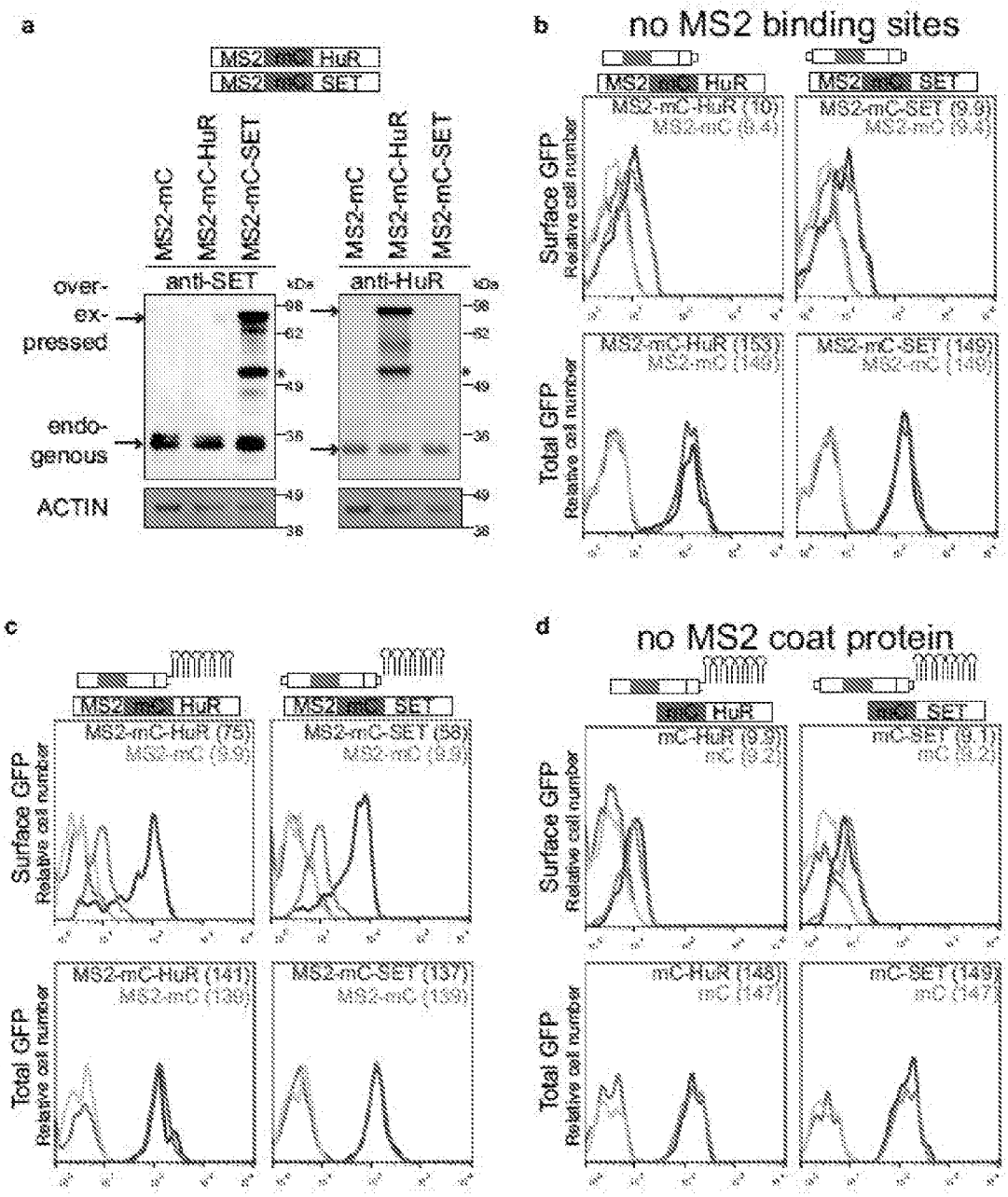

a
MS2 HuR
MS2 SET
MS2-mC
MS2-mC-HuR
MS2-mC-SET
anti-SET
anti-HuR
kDa
over-expressed
endogenous
ACTIN
b
no MS2 binding sites
MS2-mC-HuR (10)
MS2-mC (9.4)
MS2-mC-SET (9.9)
MS2-mC (9.4)
MS2-mC-HuR (153)
MS2-mC (140)
MS2-mC-SET (149)
MS2-mC (140)
Surface GFP
Total GFP
Relative cell number
c
MS2-mC-HuR (75)
MS2-mC (9.9)
MS2-mC-SET (58)
MS2-mC (9.5)
MS2-mC-HuR (141)
MS2-mC (139)
MS2-mC-SET (137)
MS2-mC (139)
d
no MS2 coat protein
mC-HuR (9.9)
mC (9.2)
mC-SET (9.1)
mC (9.2)
mC-HuR (148)
mC (147)
mC-SET (149)
mC (147)

Figure 10.

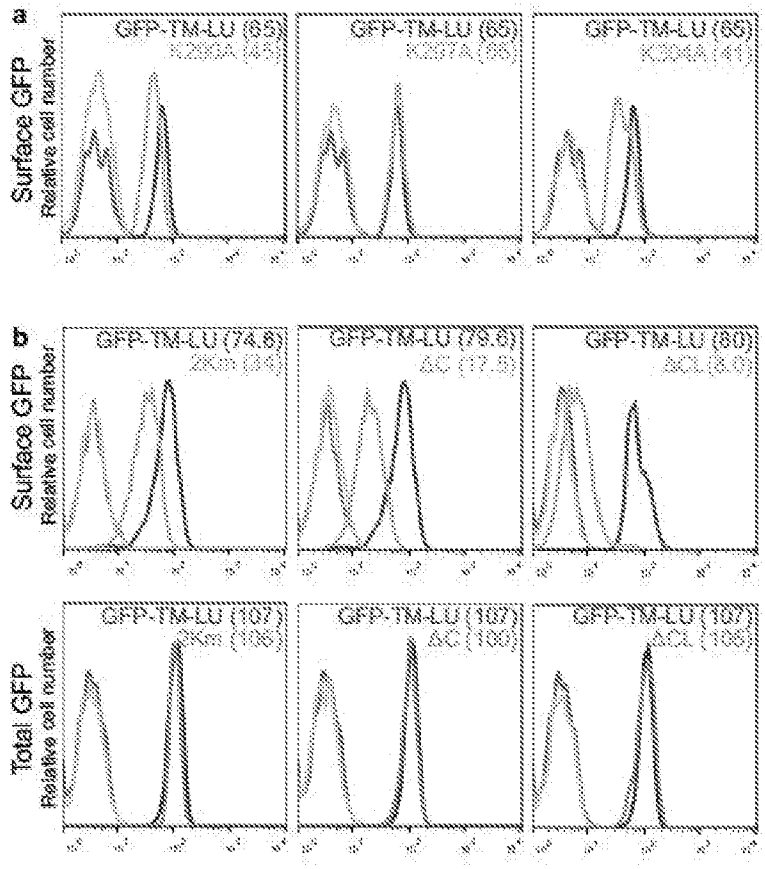

a
Surface GFP
Relative cell number
GFP-TM-LU (65)
K236A (45)
GFP-TM-LU (65)
K237A (65)
GFP-TM-LU (65)
K234A (41)
b
Surface GFP
Relative cell number
GFP-TM-LU (74.6)
ΔCL (3.4)
GFP-TM-LU (78.6)
ΔC (17.5)
GFP-TM-LU (80)
ΔCL (3.0)
Total GFP
Relative cell number
GFP-TM-LU (107)
ΔCL (106)
GFP-TM-LU (107)
ΔC (100)
GFP-TM-LU (107)
ΔCL (106)

Figure 11.

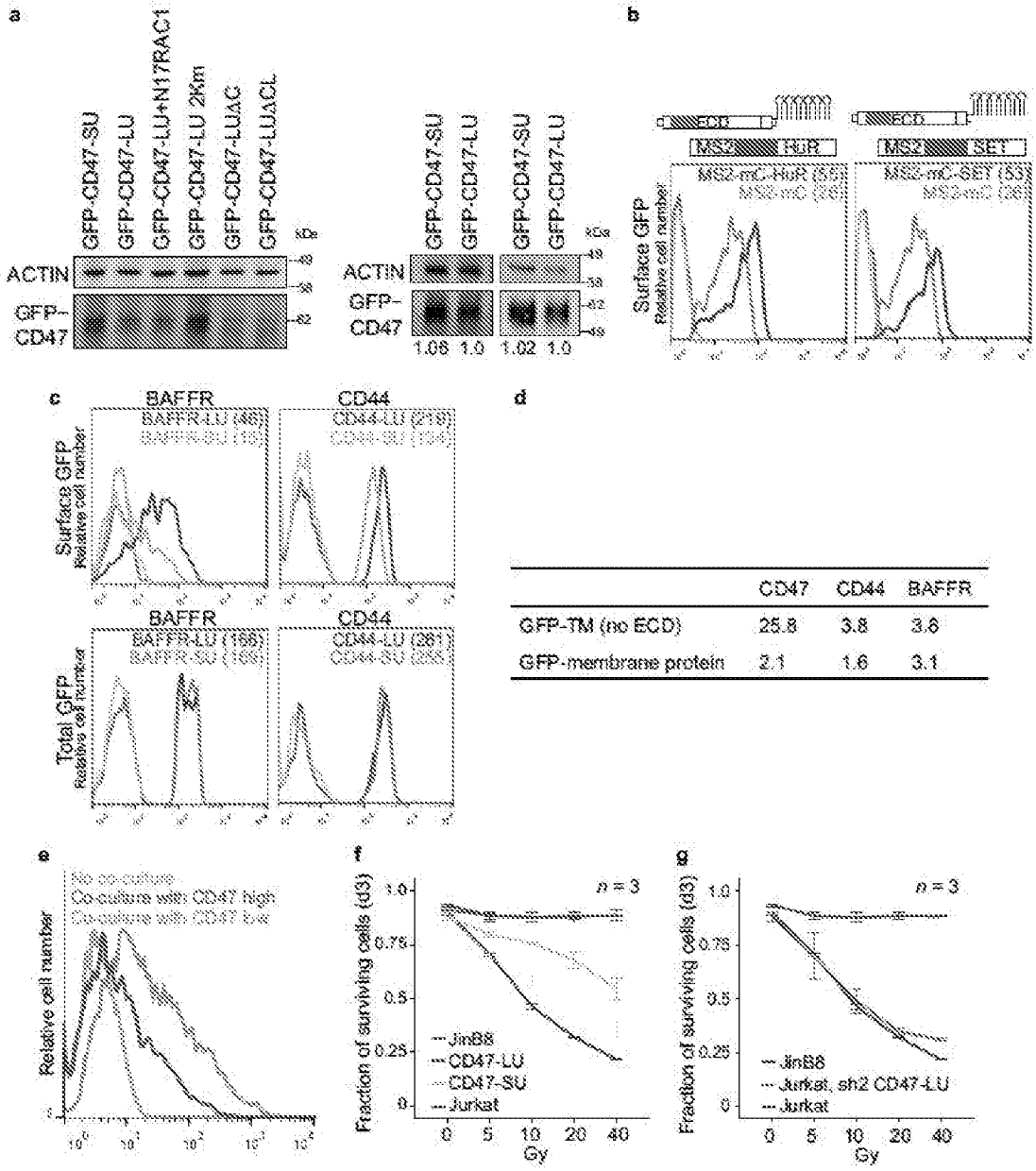
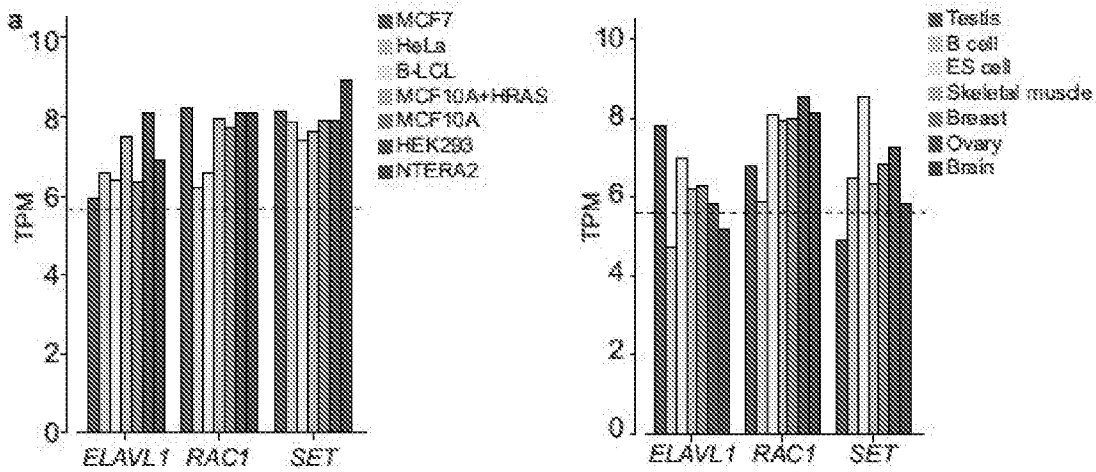

a
GFP-CD47-SU
GFP-CD47-LU
GFP-CD47-LU+N17RAC1
GFP-CD47-LU 2km
GFP-CD47-LUΔC
GFP-CD47-LUΔCL
ACTIN
GFP-CD47
kDa
1.06 1.0 1.02 1.0
b
ECD
MS2 HuR
MS2 SET
MS2-mC-HuR (55)
MS2-mC (26)
MS2-mC-SET (53)
MS2-mC (26)
Surface GFP
Relative cell number
c
BAFFR
CD44
BAFFR-LU (46)
BAFFR-SU (15)
CD44-LU (218)
CD44-SU (104)
Surface GFP
Relative cell number
BAFFR-LU (168)
BAFFR-SU (168)
CD44-LU (281)
CD44-SU (224)
Total GFP
Relative cell number
d
CD47 CD44 BAFFR
GFP-TM (no ECD) 25.8 3.8 3.6
GFP-membrane protein 2.1 1.6 3.1
e
No co-culture
Co-culture with CD47 high
Co-culture with CD47 low
Relative cell number
f
n = 3
Fraction of surviving cells (d3)
JinB8
CD47-LU
CD47-SU
Jurkat
Gy
g
n = 3
Fraction of surviving cells (d3)
JinB8
Jurkat, sh2 CD47-LU
Jurkat
Gy

Figure 12.



b

	n	%
HuR targets	2527	100
Membrane proteins	799	31.6

Figure 13.

CD47 SEQ ID NO: 8

```
                                                       150
                                                  NILIVIFPI
       160        170        180        190        200
FAILLFNGQF GIKTLKYRSG GMDEKTIALL VAGLVITVIV IVGAILFVPG
       210        220        230        240        250
EYSLKNATGL GLIVTSTGIL ILLHYYVFST AIGLTSFVIA ILVIQVIAYI
       260        270        280        290        300
LAVVGLSLCI AACIPMHGPL LISGLSILAL AQLLGLVYMK FVASNQKTIQ
       310        320
PPRKAVEEPL NAFKESKGMM NDE
```

CD44 SEQ ID NO: 9

```
       660        670        680        690        700
LIILASLLAL ALILAVCIAV NSRRRCGQKK KLVINSGNGA VEDRKPSGLN
       710        720        730        740
GEASKSQEMV HLVNKESSET PDQFMTADET RNLQNVDMKI GV
```

ITGA1 SEQ ID NO: 10

```
      1150       1160       1170
LWVILLSAF AGLLLLMLLI LALWKIGFFK RPLKKKMEK
```

BAFFR SEQ ID NO: 11

```
                           80         90         100
                          FG APALLGLALV LALVLVGLVS
       110        120        130        140        150
WRRRQRRLRG ASSAEAPDGD KDAPEPLDKV IILSPGISDA TAPAWPPPGE
       160        170        180
DPGTTPPGHS VPVPATELGS TELVTTKTAG PEQQ
```

TSPAN13 SEQ ID NO: 12

```
        10         20         30         40         50
MVCGGFACSK NCLCALNLLY TLVSLLLIGI AANGIGFGLI SSLRVVGVVI
        60         70         80         90         100
AVGIFLFLIA LVGLIGAVKN HQVLLFFYMI ILLLVFIVQF SVSCACLALN
       110        120        130        140        150
QEQQGQLLEV GWNNTASARN DIQRNLNCCG FRSVNPNDTC LASCVKSDHS
       160        170        180        190        200
CSPCAPIIGE YAGEVLRFVG GIGLFFSFTE ILGVWLTYRY RNQKDPPANP

SAFL
```

Figure 14.

2x HuR_BS SEQ ID NO: 50

tttaatagggtgagcttgagagttttctttctttctgttttttttttttttttgactaatttcac
atgctctaatttaatagggtgagcttgagagttttctttctttctgttttttttttttttttgac
taatttcacatgctctaa

2x HuR-BS with spacer SEQ ID NO: 51

tttaatagggtgagcttgagagttttctttctttctgttttttttttttttttgactaatttcac
atgctctaatgctcaccaataccctgtaaaaatacgtaattcttcctgtttgtgtaataagatat
tcatatttgtagttgcattaataatagttatttcttagtccatcagatgttcccgtgtgcctctt
ttatgccaaattg
attgtcatatttcatgttgggaccaagtagtttgcccatggcaaacctaaatttatgacctgctg
aggcctctcagaaaactgagcatactagcaagacagctcttcttgaaaaaaaaaatatgtataca
caaatttaatagggtgagcttgagagttttctttctttctgttttttttttttttttgactaatt
tcacatgctctaa

CXCR4 SEQ ID NO: 52

Cacagatgtaaaagacttttttttatacgataaataacttttttttaagttacacatttttcaga
tataaaagactgaccaatattgtacagtttttattgcttgttggatttttgtcttgtgtttcttt
agtttttgtgaagtttaattgacttatttatataaatttttttttgtttcatattgatgtgtgtct
aggcaggacctgtggccaagttcttagttgctgtatgtctcgtggtaggactgtagaaaagggaa
ctgaacattccagagcgtgtagtgaatcacgtaaagctagaaatgatccccagctgtttatgcat
agataatctctccattcccgtggaacgtttttcctgttcttaagacgtgattttgctgtagaaga
tggcacttataaccaaagcccaaagtggtatagaaatgctggtttttcagttttcaggagtgggt
tgatttcagcacctacagtgtacagtcttgtattaagttgttaataaaagtacatgttaaacttt

FAS SEQ ID NO: 53

agtgaaaaacaacaaattcagttctgagtatatgcaattagtgtttgaaaagattcttaatagct
ggctgtaaatactgcttggtttttactgggtacattttatcatttattagcgctgaagagccaa
catatttgtagatttttaatatctcatgattctgcctccaaggatgtttaaaatctagttgggaa
aacaaacttcatcaagagtaaatgcagtggcatgctaagtacccaaataggagtgtatgcagagg
atgaaagattaagattatgctctggcatctaacatatgattctgtagtatgaatgtaatcagtgt
atgttagtacaaatgtctatccacaggctaaccccactctatgaatcaatagaagaagctatgac
cttttgctgaaatatcagttactgaacaggcaggccactttgcctctaaattacctctgataatt
ctagagattttaccatatttctaaactttgtttataactctgagaagatcatatttatgtaaagt
atatgtatttgagtgcagaatttaaataaggctctacctcaaagacctttgcacagtttattggt
gtcatattatacaatatttcaattgtgaattcacatagaaaacattaaattataatgtttgacta
ttatatatgtgtatgcattttactggctcaaaactacctacttctttctcaggcatcaaaagcat
tttgagcaggagagtattactagagctttgccacctctccatttttgccttggtgctcatcttaa
tggcctaatgcacccccaaacatggaaatatcaccaaaaaatacttaatagtccaccaaaaggca
agactgcccttagaaattctagcctggtttggagatactaactgctctcagagaaagtagctttg
tgacatgtcatgaacccatgtttgcaatcaaagatgataaaatagattcttatttttcccccacc
cccgaaaatgttcaataatgtcccatgtaaaacctgctacaaatggcagcttatacatagcaatg
gtaaaatcatcatctggatttaggaattgctcttgtcatacccccaagtttctaagatttaagat
tctccttactactatcctacgtttaaatatctttgaaagtttgtattaaatgtgaatttaagaa
ataatatttatatttctgtaaatgtaaactgtgaagatagttataaactgaagcagatacctgga
accacctaaagaacttccatttatggaggatttttttgccccttgtgtttggaattataaaatat
aggtaaaagtacgtaattaaataatgtttttggtaa

TNF SEQ ID NO: 54

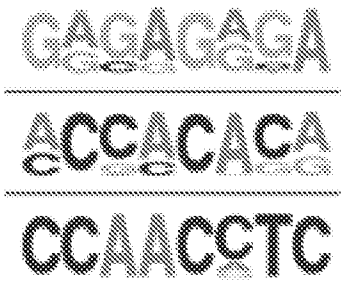
Ggaggacgaacatccaaccttcccaaacgcctcccctgccccaatccctttattacccccctcctt
cagacaccctcaacctcttctggctcaaaaagagaattgggggcttagggtcggaacccaagctt
agaactttaagcaacaagaccaccacttcgaaacctgggattcaggaatgtgtggcctgcacagt
gaagtgctggcaaccactaagaattcaaactggggcctccagaactcactggggcctacagcttt
gatccctgacatctggaatctggagaccagggagcctttggttctggccagaatgctgcaggact
tgagaagacctcacctagaaattgacacaagtggaccttaggccttcctctctccagatgtttcc
agacttccttgagacacggagcccagccctccccatggagccagctccctctatttatgtttgca
cttgtgattatttattatttatttattatttatttatttacagatgaatgtatttatttgggaga
ccggggtatcctgggggacccaatgtaggagctgccttggctcagacatgttttccgtgaaaacg
gagctgaacaataggctgttcccatgtagccccctggcctctgtgccttcttttgattatgtttt
ttaaaatatttatctgattaagttgtctaaacaatgctgatttggtgaccaactgtcactcattg
ctgagcctctgctccccaggggagttgtgtctgtaatcgccctactattcagtggcgagaaataa
agtttgcttag

PDGFA SEQ ID NO: 55

agcagccaaccagatgtgaggtgaggatgagccgcagccctttcctgggacatggatgtacatgg
cgtgttacattcctgaacctactatgtacggtgctttattgccagtgtgcggtctttgttctcct
ccgtgaaaaactgtgtccgagaacactcgggagaacaaagagacagtgcacatttgtttaatgtg
acatcaaagcaagtattgtagcactcggtgaagcagtaagaagcttccttgtcaaaaagagagag
agagaaagagagagagaaaacaaaaccacaaatgacaaaaacaaaacggactcacaaaaatatct
aaactcgatgagatggagggtcgccccgtgggatggaagtgcagaggtctcagcagactggattt
ctgtccgggtggtcacaggtgctttttgccgaggatgcagagcctgctttgggaacgactccag
aggggtgctggtgggctctgcaggggcccgcaggaagcaggaatgtcttggaaaccgccacgcga
actttagaaaccacacctcctcgctgtagtatttaagcccatacagaaaccttcctgagagcctt
aagtggtttttttttttgttttgttttgttttttttttttttgttttttttttttttttttttac
accataaagtgattattaagctttccttttactctttggctagcttttttttttttttttttttt
tttaattatctcttggatgacatttacaccgataacacacaggctgctgtaactgtcaggacagt
gcgacggtatttttcctagcaagatgcaaactaatgagatgtattaaaataaacatggtatacct
acctatgcatcatttcctaaatgtttctggctttgtgtttctcccttaccctgctttatttgtta
atttaagccattttgaaagaactatgcgtcaaccaatcgtacgccgtccctgcggcacctgcccc
agagcccgtttgtggctgagtgacaacttgttccccgcagtgcacacctagaatgctgtgttccc
acgcggcacgtgagatgcattgccgcttctgtctgtgttgttggtgtgccctggtgccgtggtgg
cggtcactccctctgctgccagtgtttggacagaacccaaattctttatttttggtaagatattg
tgctttacctgtattaacagaaatgtgtgtgtgtggtttgtttttttgtaaaggtgaagtttgta
tgtttacctaatattacctgttttgtatacctgagagcctgctatgttctttttttgttgatcca
aaattaaaaaaaaaaataccaccaacaaaa

Figure 15.

A.



B.

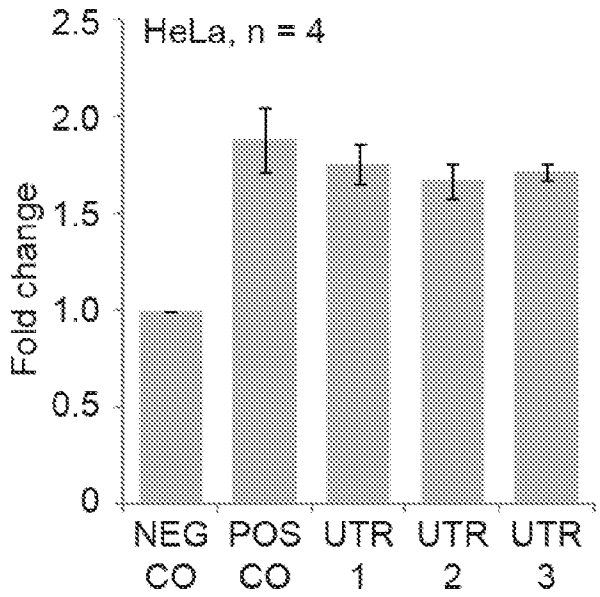
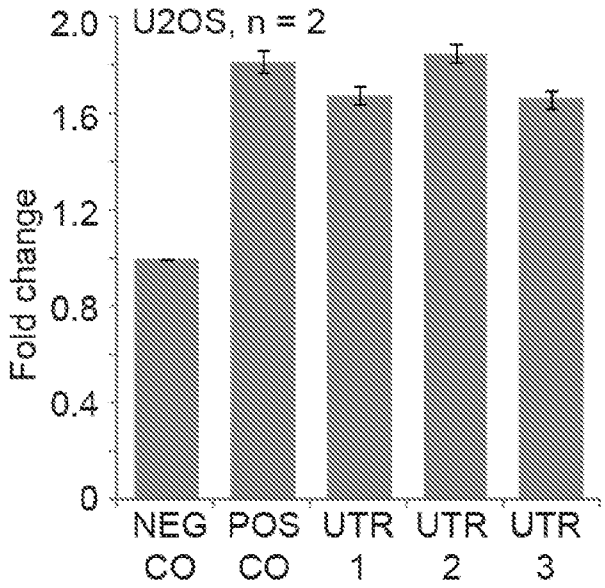
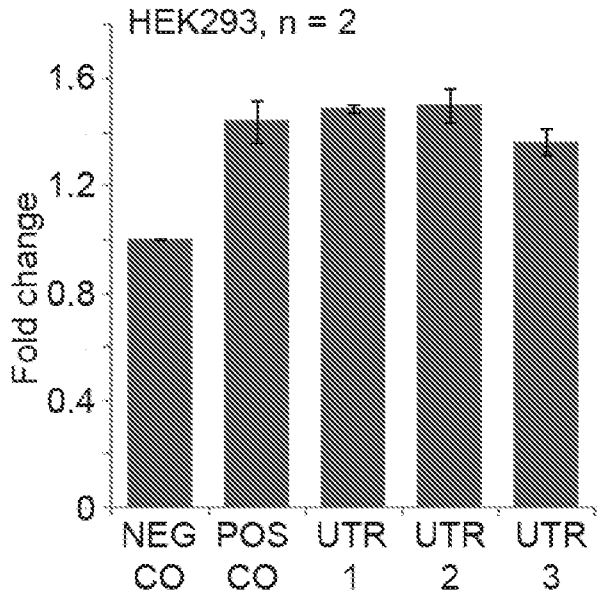


Figure 15.

C.


U2OS, n = 2
Fold change
2.0
1.6
1.2
0.8
0.4
0
NEG CO
POS CO
UTR 1
UTR 2
UTR 3

D.


HEK293, n = 2
Fold change
1.6
1.2
0.8
0.4
0
NEG CO
POS CO
UTR 1
UTR 2
UTR 3