TITLE PAGE

Article Title: **‘Therapeutic applications of the “NPGP” family of viral 2As’**

Running Head: **2A-mediated protein co-expression in biomedicine**

Key words: **Protein co-expression, 2A, Biomedical biotechnology.**

Authors: **Garry A. Luke* and Martin D. Ryan**

***Corresponding Author: gal@st-andrews.ac.uk**

Address: **Centre for Biomolecular Sciences, School of Biology,**
University of St Andrews, North Haugh,
St Andrews, KY16 9ST, U.K.

SUMMARY

Oligopeptide '2A' and '2A-like' sequences ('2As'; 18-25aa) are found in a range of RNA virus genomes controlling protein biogenesis through 'recoding' of the host-cell translational apparatus. Insertion of multiple 2As within a single open reading frame (ORF) produces multiple proteins, hence 2As have been used in a very wide range of biotechnological and biomedical applications. During translation, these 2A peptide sequences mediate a eukaryote-specific, self-"cleaving" event, termed "ribosome skipping" with very high efficiency. A particular advantage of using 2As is the ability to simultaneously translate a number of proteins at an equal level in all eukaryotic systems although, naturally, final steady-state levels depend upon other factors – notably protein stability. In contrast, the use of internal ribosome entry site (IRES) elements for co-expression results in an unbalanced expression due to the relative inefficiency of internal initiation. For example, a 1:1 ratio is of particular importance for the biosynthesis of the heavy-chain and light-chain components of antibodies: highly valuable as therapeutic proteins. Furthermore, each component of these 'artificial polyprotein' systems can be independently targeted to different sub-cellular sites. The potential of this system was vividly demonstrated by concatenating multiple gene sequences, linked *via* 2A sequences, into a single, long, ORF – a polycistronic construct. Here, ORFs comprising the biosynthetic pathways for violacein (5 gene sequences) and β-carotene (4 gene sequences) were concatenated into a single cistron such that all components were co-expressed in the yeast *Pichia pastoris*. In this review, we provide useful information on 2As to serve as a guide for future utilities of this co-expression technology in basic research, biotechnology and clinical applications.

List of Abbreviations: ACT, adoptive T-cell transfer; AAV, adeno-associated vector; BaMV, bamboo mosaic virus; BPMV, bean pod mottle virus; CAR, chimeric antigen receptor; CCR5, chemokine receptor 5; CEA, carcinoembryonic antigen; CIB1, calcium and integrin-binding protein 1; CP, coat protein; CRISPR, clustered regularly interspaced short palindromic repeat;

CRY2, cryptochrome 2; CTLA-4, cytotoxic T lymphocyte antigen-4; dsB, double-strand break; EDIII, envelope protein domain III; EGFR, epidermal growth factor receptor; ER, endoplasmic reticulum; eRF, eukaryotic release factor; FMDV, foot-and-mouth disease virus; GFP, green fluorescent protein; HDR, homology directed repair; HCV, hepatitis C virus; IDLV, integrase-defective lentiviral vector; Ig, immunoglobulin; IL, interleukin; 4-IPO, 4-ipomeanol; iPSC, induced pluripotent stem cell; IRES**,** internal ribosome entry site; JEV, japanese encephalitis virus; mAb, monoclonal antibody; LOV, light, oxygen and voltage; MAARGE, multiplexable activation of artificially repressed genes; MAPK, mitogen-activated protein kinase; MHC, major histocompatibility complex; MSTN, myostatin; MUC1, mucin-1; NHEJ, non-homologous end joining; NILV, non-integrating lentivirus; NLR, NOD-like receptor; NOD, nucleotide-binding oligomerization domain; ORF, open reading frame; PhyB, phytochrome B; PVX, potato virus X; RGN, RNA-guided nuclease; RNA, ribonucleic acid; scFv, single-chain variable fragments; TALEN, transcription activator-like effector nuclease; TCR, T-cell receptor; TIL, tumour infiltrating lymphocytes; TLR, traffic light reporter; UVR8, ultraviolet-resistance locus 8; ZFN, zinc finger nuclease.

1 – INTRODUCTION

The 2A/2Alike peptide sequences found in a number of RNA viruses share a highly conserved motif: -D[V/I]E(X)NPG$^{\downarrow}$P-, where "X" = any amino acid and $^{\downarrow}$ indicates the site of 'cleavage' between the C-terminal glycine of 2A and the N-terminal proline of the downstream protein. We have shown that 2A is not a proteinase, nor a substrate for a host-cell proteinase, but is an

autonomous element mediating a novel co-translational "recoding" event.[1-5] Briefly, our model proposes that during the elongation phase of translation, the nascent 2A sequence interacts with the ribosome exit tunnel to bring about a reorientation of the peptidyl-tRNA ester linkage in the peptidyl-transferase centre of the ribosome precluding it from nucleophilic attack by prolyl-tRNA. In this manner, the normal process of peptide bond formation between the glycine and the proline residues is prevented. Although no stop codon is involved, yeast genetic screens for the involvement of host-cell factors in the 'cleavage' mechanism showed that eukaryotic translation release factors 1 and 3 (eRF1/3) played a key role in the release of nascent protein from the ribosome at the C-terminus of 2A.[6-8] However, Machida and co-workers showed that processing of encephalomyocarditis virus (EMCV) proteins 2A and 2B in a reconstituted *in vitro* translation system does indeed occur in the elongation phase of translation – but without the involvement of eRFs.[9] Whilst the basis of this discrepancy remains unclear, this does not impact upon the utility of the system. Termed "ribosome skipping"[10] or "stop carry-on" translation,[6] our model of this translational recoding event predicts two outcomes subsequent to release of the nascent polypeptide: (i) translation of the downstream sequences resumes, or, (ii) ribosomes terminate translation. A third (minor) outcome is that 2A does not mediate a recoding event and the peptide bond between 2A and the down-stream sequence is formed in the normal manner – an outcome dependent upon the length of 2A and sequence variants. For an in-depth review of the model see reference 11. In support of our co-translational model of 2A-mediated 'cleavage', any signal sequences immediately downstream of 2A are recognised by signal recognition particle to target that protein to the exocytic pathway.[12] Highly efficient co-expression in association with the ability to independently target the individual protein 'cleavage' products to different subcellular compartments greatly enhances the versatility of the 2A system.[12-15]

The first 2A peptide sequence identified was from the picornavirus foot-and-mouth disease virus (FMDV: "F2A").[16] Measuring the activity of 2A/2A-like sequences usually involves artificial

self-processing polyproteins comprising reporter proteins flanking the 2A sequence in a single ORF.[17] Site-directed mutagenesis showed the conserved amino acid sequence, NPGP, is critical for activity.[2,4,18,19] Analyses of natural synonymous codon usage of this –NPGP- motif show preferential usage of AAC, CCT, GGG and CCC codons, respectively,[20] specific silent mutation of the codon encoding the glycine showed no effect[21]. Furthermore, manifold silent point mutations introduced within this motif for biotechnological purposes (to reduce the probability of deletion through recombination in the case where multiple 2As are used) have no effect on 2A 'cleavage'.

Caveats of the 2A system are that (i) proline forms the N-terminal residue of the protein downstream of 2A and (ii) 2A remains as a C-terminal extension on the protein upstream of 2A. These factors must be taken into consideration during construct design with regard to necessary post-translational modifications, trafficking or function - which may be overcome by correct order of the gene sequences.[22-24] For proteins passing through the exocytic pathway, inclusion of a furin proteinase cleavage site (-$^{\downarrow}$RX[K/R]R-) between the upstream protein and 2A results in efficient removal of the 2A peptide.[25,26] Similarly, in plants the 2A C-terminal extensions are removed by inclusion of a linker (-SN$^{\downarrow}$AADEVAT-) cleaved by an endogenous proteinase.[27,28] Antibodies have, however, been generated against 2A, serving as a useful probe for 2A "tags".[22]

F2A is widely used for co-expression, although other highly efficient "2A-like" peptides have been identified and are also widely used in biotechnology and biomedicine: T2A from *Thosea asigna* virus, P2A from porcine teschovirus-1 and E2A from equine rhinitis A virus (Figure 1).[29,30] Many researchers employ a 'spacer' sequence between the upstream protein and 2A: the V5 epitope tag: -GKPUPNPLLGLDST-,[31] a 3xFLAG epitope tag: -DYKDHDG-DYKDHDI-DYKDDDDK-,[32] or a glycine-serine linker: -GSG- or –SGSG-.[13,33]

This technology has been critical for expression of hetero-multimeric complexes and biochemical pathways in diverse areas such as human cancer gene therapies, production of induced pluripotent stem cells for regenerative medicine, creation of transgenic animals and plants and the production of high-value proteins for the pharmaceutical industry.[30,34] Below, we consider select studies to highlight current advances in this fast-moving technology.

2 - PLANT-BASED VACCINES AND ANTIBODIES

Plants are used as bioreactors to express a range of high-value proteins such as vaccine antigens and antibodies ('plantibodies').[35-37] Stable expression is accomplished by transformation of plant genomes, although plant viral vectors can produce high-level, transient, expression (without the creation of a genetically modified organism). Expression systems based on plant viruses include: epitope presentation systems in which the short antigenic peptide is fused to the viral coat protein (CP) and displayed on the surface of the virion – so-called 'overcoat' technology. Advantages of this approach are the relative ease with which modified viral particles can be purified from infected tissues and the presentation of multiple copies of an antigenic peptide on the surface of a macromolecular carrier, significantly increasing its immunogenicity.[38,39]

The helical rod-shaped potato virus X (PVX) CP as a presentation system was first demonstrated using green fluorescent protein (GFP). Expression of a simple [CP-GFP] fusion protein produced no virus particles. Linking the CP and GFP *via* a short, less efficient, version of F2A [CP-F2A$_{16}$-GFP] resulted in the production of free CP plus the [GFP-F2A$_{16}$-CP] fusion protein. Fully infectious virus particles 'decorated' with GFP were produced.[40] It is thought that the fusion protein (alone) cannot initiate the helical rod capsid structure, but having a mixture of the fusion protein plus the capsid protein allows the latter to initiate particle (rod) formation, which is then able to incorporate the fusion protein. Having established the utility of the approach, several types of protein have been expressed on the surface of these particles; rotavirus inner capsid

protein (VP6),[41] tuberculosis ESAT-6 protein,[42] classical swine fever virus E2 glycoprotein[43] and the R9 peptide from hepatitis C virus (HCV) envelope protein.[44] Other than these antigenic proteins, particles have been decorated with a single-chain antibody (scFv; Mr~25,000) or the enzyme lipase (*CalB*, Mr 33,000).[45,46]

In a recent study, Chen et al reported the use of a bamboo mosaic virus (BaMV)-based vector with the capacity to carry larger transgene loads than the PVX-based vector.[47] Using the BaMV as a plant viral display vector, the Japanese encephalitis virus (JEV) envelope protein domain III (EDIII) was genetically linked to the CP of BaMV *via* a $F2A_{18}$ sequence (Figure 2). Upon infection in the leaves of *Chenopodium quinoa*, the BaMV plant host, the EDIII peptide was displayed on virion surfaces. Most importantly, the chimeric virus was able to induce immune responses in mice and protection against JEV infection.

The development of an anti-*Taenia solium* cysticercosis vaccine in transgenic plant tissue extracts by Monreal-Escalante et al. used a 2A-based polyprotein system for the simultaneous delivery of several *Taenia* spp. antigens (KETc1, KETc12, KETc7, GK1, TSOL18) *via Agrobacterium tuefaciens* transformation.[48] The plant derived vaccine was recognised by antibodies in the cerebrospinal fluid from neurocysticercosis patients and induced antigen-specific antibodies in the sera of BALB/c immunized mice. Here, the same F2A amino acid sequence was used as the linkers. To minimise the risk of homologous recombination and genetic instability, others have used either (i) different 2A-like sequences or (ii) different codon usages for the same 2A. For example, in soybean-infecting bean pod mottle virus (BPMV) vectors, non-identical nucleotide sequences encoding the same F2A were used.[49]

In the case of passive immunity, plantibodies have been expressed for therapeutic applications.[50] Since the expression and assembly of immunoglobulin (Ig) heavy and light chains into functional murine antibodies was first demonstrated in *Nicotiana tabacum*,[51] plant-based expression

systems have proved to be an attractive platform for the production of full-size antibodies and their derivatives.[52-55] In a biosimilar study, Chen and colleagues designed and expressed a biologically functional monoclonal antibody (mAb; bevacizumab) in transgenic rice callus using the 2A peptide.[56] This anti-angiogenic agent has been widely used in clinical practice for the treatment of different cancers, slowing the growth of new blood vessels in tumours.[57] To make a polyprotein the authors inserted a $F2A_{18}$ sequence between codon optimized heavy-chain and light-chain sequences of bevacizumab. In an attempt to avoid plant-specific glycosylation (an important issue for antibodies made in plant expression), an ER retention signal peptide (KDEL) was added to the C-termini of both gene sequences. A balanced heavy:light chain ratio allowed efficient assembly and folding of functional antibody, producing recombinant mAb at high levels. However, the C-terminal 2A extension on the protein upstream of 2A resulted in the loss of recognition of the KDEL signal – a furin proteinase cleavage site between the KDEL and 2A sequences (to remove the 2A) may have improved the ER retention problem, although furin is located primarily in the Golgi apparatus. This first plant-expressed bevacizumab might potentially be used as a cost-effective "biobetter" molecule in future cancer treatments.

3 - ADOPTIVE T-CELL ANTI-CANCER IMMUNOTHERAPY

T-cells play a central role in cell-mediated immunity. Adoptive T-cell transfer (ACT), based on the manipulation and transfusion of autologous or allogeneic immune cells into patients is currently being tested in numerous clinical trials. Approaches include the use of tumour-infiltrating lymphocytes (TILs) and *ex vivo*-engineered T-cells, such as T-cell receptor (TCR), chimeric antigen receptor (CAR) and synthetic Notch (synNotch) receptor transduced T-cells (reviewed in 58-60). Adoptive cell therapy with TILs in combination with lymphodepleting regimens and the use of interleukin-2 (IL-2) to sustain T-cell expansion has shown some success in the treatment of metastatic melanoma.[61] However, such a process is laborious and time-consuming and is limited to those patients for whom tumour-specific T-cells can be isolated and

amplified. One promising strategy to expand the range of ACT is to administer T-cells that have been genetically modified to express tumour-specific antigen receptors.

The TCR αβ heterodimer detects antigenic peptides in the context of major histocompatibility complex (MHC) proteins and signals through the non-covalently associated CD3εδ, CD3εγ and ζζ dimers. Both endogenous and transgenic TCRs require a complementary activating/inhibitory co-stimulatory signal, in effect, functioning as an on/off switch to regulate T-cell activity (reviewed in 62). 2A peptides have been widely adopted for TCR gene therapy because of their near-stoichiometric expression of both TCR α and β chains.[14,31,63] Further, exposure to the most widely used 2A sequences is unlikely to produce unwanted T-cell responses in immunocompetent subjects and supports their continued use for gene therapy studies.[64,65] A potential drawback to effective immunotherapy is mispairing between introduced and endogenous TCR chains, resulting in unexpected specificities and reduced cell surface levels of transferred TCRs.[66,67] A number of TCR modifications have been reported to minimise the extent of mispairing, including the introduction of a disulphide bond between cysteines[68] and codon optimisation of the transgenic TCR to enhance mRNA translation and protein synthesis.[69] To reduce the mismatch, knockdown or knockout of endogenous TCR-α/β gene expression has been achieved using small-interfering RNA (*siRNA*) technology (*siTCR* vector).[70,71] The 2A peptide-based siTCR vector simultaneously transduces codon-optimised therapeutic TCRs [α-T2A-β] and siRNA constant regions of intrinsic TCR-α/β gene sequences. Notably, the siTCR either alone or in combination with extra cysteines (introduced by site-directed mutagenesis) markedly increases the cell-surface expression of the introduced TCR.[71]

Another promising strategy to broaden the utility of ACT is to administer T-cells that have been genetically engineered to express CARs, which are antibody single-chain variable fragments (scFv) joined with TCR and T-cell co-stimulatory receptor signal domains targeting cell-surface

antigens in an MHC-independent fashion. Unlike TCRs, co-stimulation is provided *in cis* and in response to the same activating signal. Investigators have generated CARs targeted to various tumour antigens, such as carcinoembryonic antigen (CEA), epidermal growth factor receptor (EGFR), CD19 (leukaemia and lymphoma), or viral-associated cancer antigens (reviewed in 72). The modular design of the CAR, extracellular ligand-binding domain fused through a flexible hinge and transmembrane motif to intracellular signalling components, allows optimisation of T-cell function. The successful use of an αCEA-CAR 2A-based bicistronic vector in a mammalian cell-screening protocol for the isolation of functional scFvs highlights how this approach could also be used to improve the expression and binding of scFvs.[73] Empiric testing of hinge length and sequence, vis-à-vis epitope position, is required to determine the most favourable CAR design.[74,75] For example, modification of the IgG4 Fc "spacer" domain in a CD19CAR-T2A-EGFRt construct reduced "off-target" activation and improved *in vivo* persistence and anti-tumour strength of adoptively transferred T-cells.[76,77]

Compact combined sort-suicide gene strategies based on transgenic expression of a binding target for a therapeutic antibody have been proposed for easier and safer T-cell therapy. For instance, truncated EGFR paired with the IgG1 mAb cetuximab may be used as a non-immunogenic selection tool, an *in vivo* tracking marker and a suicide gene for transduced [CD19CAR-T2A-EGFRt] T-cells.[76-78] As an enhancement of this CAR, a small marker/suicide gene (RQR8) was created by combining epitopes from CD34 and CD20 enabling CD34 selection, cell tracking, as well as depletion after anti-CD20 mAb (rituximab) targeting [RQR8-T2A-αGD2CA].[79] For clinical applications, a novel suicide gene system based on bio-activation of 4-ipomeanol (4-IPO) was recently developed as a pro-drug with active *human* mutant CYP4B1 enzyme.[80] This enzyme converts the inert substrate into highly toxic DNA alkylating metabolites.[81] For the use of CYP4B1 in combination with a CAR, the lentiviral vector expressed a codon-optimised CD19 CAR upstream or downstream of the T2A site.

Although T-cells expressing CARs have shown exciting promise in the treatment of some refractory diseases, there is ample scope for improvement. For example, several cytokines (e.g. IL-2, IL-12 and IL-15) are currently being tested as combination partners that may improve the activity and/or survival of CAR-T cells.[82] Specifically, Wilkie at al. showed that IL-4 administration can be used to achieve rapid and selective *ex vivo* expansion and sustained anti-tumour activity of human T-cells engineered to co-express a chimeric IL-4 receptor (4αβ) fused to a CAR specific for tumour-associated MUC1 (HOX) *via* T2A (Figure 3).[83] A concern with ACT is the proliferation of endogenous T cells in relation to gene-modified cells following transfer *in vivo.* To improve *in vitro*/*in vivo* selection of therapeutic cells, target cells were transduced with a vector containing the hEGFRt marker gene along with two drug-resistance genes. One gene is a variant of the inosine monophosphate dehydrogenase 2 (*IMPDH*^{IY}; T333I, S351Y), conferring resistance to the immunosuppressive drug mycophenolate mofetil. The other, a human dihydrofolate reductase (*DHFR*^{FS}; L22F, F31S), confers resistance to methotrexate.[78] In this study, engraftment of [EGFRt-T2A-DHFR^{FS}-T2A-IMPDH^{IY}] T-cells supported the preferential expansion and selection of transduced over non-transduced human T-cells following drug administration to mice.

To increase the specificity of engineered T-cells researchers have used synNotch receptors, which consist of the small regulatory domain of Notch combined with bespoke extracellular and intracellular signalling domains.[84-86] The most prominent inhibitory co-stimulatory receptors are cytotoxic T-lymphocyte associated antigen 4 (CTLA-4) and programmed cell death protein 1 (PD-1). Strategies to overcome this "adaptive resistance" of infused T-cells boosts their responses by preventing exhaustion and anergy.[87] In one approach, synthetic T-cells were engineered with the α-GFP synNotch receptor controlling the expression of both anti-PD-1 and an scFv directed toward CTLA-4 – both antibodies were expressed as a single transcript linked *via* T2A.[60] A dual pathway blockade could have a synergistic effect, because CTLA-4 and PD-1

modulate different aspects of the T-cell response.[88] Additional complementary immune therapies include antigen-specific production of immunosuppressive agents in an autoimmune-associated setting. Here, the α-CD19 synNotch receptor, when stimulated, induced the expression of [PD-L1-T2A-IL-10].[60] As the main ligand for PD-1, PD-L1 induces a co-inhibitory signal in activated T-cells and also stimulates cytokine IL-10 production, further reducing the immune response.[89]

4 - GENE THERAPY BY ENGINEERED NUCLEASES

Three major programmable nuclease platforms exist: zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and RNA-guided nucleases (RGNs) (reviewed in 90-92) The double-strand break (dsB) made by these targetable nucleases is typically repaired through one of two major pathways – non-homologous end joining (NHEJ) or homology directed repair (HDR).[93] ZFNs and TALENs are similar and comprise a *Fok*I derived nuclease domain coupled to a DNA binding domain engineered to recognize a specific sequence.[94,95] The clustered regularly interspaced short palindromic repeat (CRISPR)/Cas RGNs rely on base-pairing between a "guide" RNA and the DNA target for recognition and a Cas nuclease for DNA cleavage.[90,96]

Regardless of the nuclease system used, it is essential to have methods that can select or enrich genome-modified cells. Several approaches using 2A have been applied to aid selective enrichment of transfected cells, including fusion of ZFNs/TALENs and Cas9 to a fluorescent or antibiotic protein,[97-101] or, co-transfection of the nucleases with a fluorescent or antibiotic resistance marker gene.[100] To recapitulate DNA nuclease activity, dual-reporter surrogate systems that contain the target sequence to be modified have been developed to detect cells with high repair efficiency.[102] The two-colour "Traffic Light Reporter" (TLR; *eGFP-T2A-mCherry*) developed by Scharenberg and colleagues uses a flow cytometric assay to simultaneously detect both gene repair and mutagenic NHEJ at a single targeted site (Figure 4).[103] HDR using a donor

template carrying an intact *eGFP* results in green fluorescent cells. If mutagenic NHEJ repairs the break, a shift in the reading frame of the *eGFP* reporter can result in translation of the T2A-mCherry, leading to red fluorescent cells.

The advances in genome editing technologies have opened the door for an entirely new and promising approach to removing or correcting causative errors in therapeutic genome "surgery". One of the main hurdles to using ZFNs is the effective delivery of ZFNs and donor templates to target cells. Among viral vectors, integrase-defective lentiviral vectors (IDLVs, also called non-integrating lentiviral vectors),[104] adeno-associated vectors (AAVs),[105] and chimeric adenoviral vectors[106] have been used to safely deliver the gene targeting components. Gene delivery is a particularly significant issue for gene correction because it is necessary to introduce three components into cells (e.g. two ZFNs + donor). To improve the delivery of both ZFNs to the same cell in equal amounts, Joglekar et al. designed a single IDLV encoding both the left and right ZFN monomers linked by T2A and linked to mCherry *via* P2A, to monitor gene transfer and expression by flow cytometry.[107] By introducing DNA sequence changes to reduce homology between the two ZFNs, this study achieved delivery of both monomers by one IDLV. Independently, envelope pseudotyped non-integrating lentivirus (NILV) was used for co-expression of ZFNs to edit the chemokine receptor 5 (CCR5) gene as a therapeutic strategy to block HIV-1 infection.[108] CCR5 is the major co-receptor for HIV-1 entry and validated as a target for HIV therapy.[109] ZFN constructs targeting both strands of CCR5 were connected *via* F2A, and then inserted into a lentivirus cloning vector. The results showed that transduced activated and resting T-cells resist HIV infection *in vitro*. Likewise, endogenous virus replication was suppressed in humanized mice after adoptive transfer of NILV-transduced primary T-cells or $CD34^+$ stem cells. An alternative to using IDLV is to use AAV for delivery which, in comparison, is more efficient for transduction *in vitro* and *in vivo*. Although AAV has many advantages over other viral vectors, its relatively small packaging capacity (< 4.2kb) limits its

use for delivering large genes. To address this, a single AAV vector, which contains DNA encoding two GFP ZFNs separated by a 2A peptide, and includes the donor substrate was used for effective delivery into mouse and human cells.[110] Furthermore, the feasibility of autologous $CD4^+$ T-cell infusions modified by CCR5-specific 2A-linked ZFNs delivered with an AAV was demonstrated in the first bench to bedside trials in subjects infected with HIV-1.[111]

Because *Fok*1 functions as a dimer, TALENs, like ZFN's, are designed in pairs that bind opposing DNA target sites. Perhaps the most obvious approach for efficient genome editing is *via* 2A-linked TALEN monomers. To this end, Mariano and co-workers (2014) compared the gene editing performances of co-transfected TALEN-L and R versus 2A-linked TALENs targeting two different loci: myostatin (MSTN) and AAVS1.[99] MSTN is a potent inhibitor of skeletal muscle development and growth in several mammalian species; loss of this gene function leads to profound muscle 'over-growth'.[112,113] TALENs targeting the *MSTN* gene expressed from one plasmid exhibited higher gene editing activity compared to co-transfection with the same TALENS in two separate plasmids.[99,114] The well characterised human AAVS1 "safe harbor" locus has been previously shown to allow stable and long-term expression of transgenes in multiple cell types including human pluripotent stem cells.[115,116] Similarly, TALENs used for AAVS1 genome editing showed that expression from one plasmid induced higher gene editing activity than the co-transfected TALEN-L and R due to proper co-delivery of TALE monomers.

In contrast to protein-based tools such as zinc-fingers and TALE's, CRISPR technology is now the most popular genome editing tool. One of biggest advantages of the CRISPR/Cas9 system is its ability to target multiple sites by simply providing several distinct gRNAs along with the Cas9 protein. Multiplex editing offers a robust, efficient and designable approach for genome engineering including new therapeutic modalities.[117] For mammalian cells, a novel strategy using multiplexable activation of artificially repressed genes (MAARGE) allows the fast and convenient generation of stable cell lines with defined

expression of the gene(s) of interest, while avoiding off-target activity at different sites in the genome. Since the activation of gene expression occurs at low efficiency, the expression of each engineered gene should be coupled to the expression of a distinct fluorescent protein *via* a 2A peptide in order to facilitate enrichment of modified cells.[118]

5 - OPTOGENETIC/LIGHT CONTROL OF GENE EXPRESSION IN BIOLOGICAL SYSTEMS

The burgeoning field of "optogenetics" integrates optics and genome engineering approaches to control various protein-protein interactions with high spatial and temporal resolution. A variety of light-controlled dimerisation systems have been developed, such as LOV (light, oxygen, and voltage) domains,[119] phytochrome B (PhyB),[120,121] cryptochrome 2 (CRY2),[122] and UV-resistance Locus 8 (UVR8).[123,124] Among these, the blue light-dependent binding of *Arabidopsis* CRY2 to its partner cryptochrome interacting protein (CIB1), has been the most studied as this system offers rapid on- and off-kinetics of protein pair binding without the need for exogenous co-factors. CRY2-CIB1 heterodimerisation is based on the interaction between the N-terminal portion of CIB1 (CIBN) and the PHR domain of CRY2 in response to blue light.[122,125] So far, this approach has been used in a variety of cell lines and model systems to regulate transcription,[126] signalling pathways, [127,128] phosphoinositide metabolism,[129] and other cellular processes.

Despite its advantages, the CRY2-CIB system has an important practical limitation – the lack of ratiometric control between the fusion proteins. A solution was provided in the form of a 2A-based bicistronic optogenetic system, optimising light-activation of the Raf/MEK/ERK signalling cascade in PC12 cells (Figure 5).[130] The mitogen-activated protein kinase (MAPK) pathway regulates cell cycle progression, survival, senescence, and migration.[131] The 2A constructs contained a CRY2-mCherry-Raf1 expression cassette, the P2A peptide and CIBN-

GFP-CaaX cassette (with varying number of CIBNs, referred to as CRY2-2A-(n)CIBN). Light-induced binding between CIBN and CRY2 leads to membrane recruitment of CRY2-mCherry-Raf1, which activates the Raf/MEK/ERK signalling pathway and induces PC12 cell differentiation. The results suggest that a CIBN:CRY2 expression ratio of 2:1 is an optimal stoichiometry for light-induced PC12 differentiation. Additionally, the optimised [CRY2-2A-2CIBN] system allows reversible optogenetic control of kinase activity during *Xenopus* embryonic development and demonstrates the potential utility of this system to other *in vivo* models.

Recently, a photoactivatable Cre-*lox*P recombination system was developed *via* fusion of (i) truncated CR6Y2 to the N-terminal domain of Cre and (ii) fusion of CIB1 to the C-terminal domain of Cre. Blue light illumination causes dimerisation of CRY2 with CIB1 and reconstitution of split-Cre recombinase activity.[132] More recently, a highly efficient photoactivatable Cre recombinase (PA-Cre) that is based on reassembly of a pair of split Cre fragments, each appended to photoinducible dimerisation domains named "Magnets", has been validated *in vitro* and *in vivo* .[133] Both positive (pMag) and negative (nMag) "Magnets" are inducible dimerisation proteins (~150aa each) which heterodimerise in response to blue light irradiation[134] thereby unifying the split-Cre into an active enzyme. On the basis of recombination efficiency, CreN59 (residues 19-59) and CreC60 (residues 60-343) fragments were used for PA-Cre, which is different from the conventional CRY2-CIB1 split Cre system using the CreN104 (residues 19-104) and CreC106 (residues 106-343) pair. To package PA-Cre into a more compact format the CreN59-nMag ORF was followed by a P2A peptide sequence and then the coding sequence of pMag-CreC60.

6 – CONCLUDING REMARKS

Over 1.100 research publications and approximately 700 patent applications attest to the influence of 2As on co-expression technology. Previous studies have shown that 2A works in all eukaryotic cells tested to date: mammalian[135,136] plant[137], insect[138,139], filamentous fungi[140], the yeast *Saccharomyces cerevisiae*[8] and *Pichia pastoris*[141,142] Although eukaryotic ribosomes are highly conserved, they are not completely conserved. F2A, E2A and P2A have all evolved to cleave within mammalian ribosomes and yet they are not identical sequences. T2A has evolved to cleave within insect cells. We have proposed that 2A interacts with the ribosome exit tunnel, some 100Å long and between 10-20Å in width. The tunnel is thought to accommodate a 30aa tract (extended conformation) or up to 60aa (α-helical conformation). Indeed, it has been shown that nascent peptides may initiate folding while located well within the ribosome exit tunnel.[143] The length of 2As is, however, less than 30/60aa: gene sequences immediately upstream of 2A are also present within the exit tunnel: they may affect the interaction of 2A with the exit tunnel and, therefore, the cleavage activity. The cleavage activity of any specific 2A may, therefore, be somewhat different on a case-by-case basis. Interestingly, some investigators use flexible 'linker' peptides between the gene sequence and 2A (see below) – the extra (flexible) sequence may ameliorate any effect of gene sequences upstream of 2A. The majority of studies have primarily focused on the "cleavage" efficiencies of different 2As in bi-cistronic systems and obtained mixed results. This variability is due in part to the context of the experiment, including the, sequence of each specific construct, the organism and/or cell lines used: plus in the majority of cases the relative stability of the various translation products is unknown. We have shown $T2A_{20}$ has the highest cleavage efficiency (~100%), followed by $E2A_{20}$, $P2A_{20}$ and $F2A_{20}$, when compared using *in vitro* transcription/translation experiments[5]. On the other hand, Szymczak and colleagues demonstrated that $F2A_{22}$ and $T2A_{18}$ have higher efficiency (~100%) than $E2A_{20}$ in a similar experimental setup[14]. When evaluated in human cell lines, mice, zebrafish and *drosophila* using western blotting/fluorescence microscopy, $P2A_{19}$ exhibited the highest cleavage efficiency followed by $T2A_{18}$, $E2A_{20}$ and $F2A_{22}$. [135,138] In *S. cerevisiae*, these commonly used sequences all showed some level of cleavage, determined by western blotting: $P2A_{19}$(85%), $T2A_{18}$(56%), $E2A_{20}$(46%), $F2A_{22}$(43%) – interestingly, the best performance (~100%) was achieved using the less-used 2A peptide from Equine rhinitis B virus (ERBV-1).[144] The inclusion of –GSG- or –

SGS- 'linkers' between the first gene sequence and 2A has been shown to improve the efficiency of cleavage. [14,31,32,145]

Co-expression of multiple genes (more than two) is sometimes desired, e.g. 2A has played a pivotal role in the co-expression of the four transcription factors (Oct4, Sox2, c-Myc, Klf4) required to produce induced pluripotent stem cells (iPSCs).[146] In longer constructs protein expression is mostly affected by the position of individual gene sequences rather than the 2As and their order, possibly due to ribosome drop-off. [147] Notably, if multiple, identical, 2As are to be used in the same construct, the risk of deletion through recombination can be reduced by changing the codon usage for each 2A.

These viral 2A peptides have near 100% cleavage efficiency in artificial polyprotein systems, but they can be made to cleave at lower efficiencies when key amino acid residues are mutated.[4,5] In some cases the presence of both cleaved and uncleaved forms can be useful or even essential: in the biogenesis of recombinant 'decorated' virus particles, for example, such a mixture is key.[47] Yu and colleagues developed a "Molecular Rheostat" system using appropriate mutant F2A peptides to select, or tune, the levels of membrane-bound and secreted IgG in B cells.[148] Subsequently, Cruz-Teran et al. demonstrated that inefficient ribosomal skipping can also be used for simultaneous cell surface display and secretion of proteins in *Saccharomyces cerevisiae.*[149] Recently, we identified "2A-like" sequences at the N-terminus of certain NOD-like receptors (NLRs) in the genome of the purple sea urchin *Strongylocentrotus purpuratus* - these innate immune receptors provide protection of a host against pathogens. We showed these cellular NLR-2As could function both as a signal sequence and as a translational recoding element: "uncleaved" forms enter the exocytic pathway while those translational products in which 2A has mediated "cleavage" become localised to the cytoplasm.[150,151] This represents a newly described form of dual protein targeting with obvious biotechnological and biomedical applications.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the long-term support of the Wellcome Trust and the UK Biotechnology and Biological Sciences Research Council (BBSRC). The University of St Andrews is a charity registered in Scotland no. SCO13532.

REFERENCES

1. Ryan MD, King AMQ, Thomas GP. Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence. *J Gen Virol.* 1991; 72: 2727-2732.
2. Ryan MD, Drew J. Foot-and-mouth disease virus 2A oligopeptide mediated cleavage of an artificial polyprotein. *Embo J*. 1994; 13: 928-933.
3. Ryan MD, Flint M. Virus-encoded proteinases of the picornavirus super-group. *J Gen Virol.* 1997; **78**: 699-723.
4. Donnelly MLL, Luke GA, Mehrotra A, *et al*. Analysis of the aphthovirus 2A/2B polyprotein "cleavage" mechanism indicates not a proteolytic reaction, but a novel translational effect : a putative ribosomal "skip". *J Gen Virol*. 2001; 82: 1013-1025.
5. Donnelly MLL, Hughes LE, Luke GA, *et al*. The "cleavage" activities of foot-and-mouth disease virus 2A site directed mutants and naturally occurring "2A-like" sequences. *J Gen Virol*. 2001; 82: 1027-1041.
6. Doronina VA, de Felipe P, Wu C, *et al*. Dissection of a co-translational nascent chain separation event. *Biochem Soc T*. 2008; 36(4): 712-716.
7. Doronina VA, Wu C, de Felipe P, Sachs MS, Ryan MD, Brown JD. site-specific release of nascent chains from ribosomes at a sense codon. *Mol Cell Biol*. 2008; 28(13): 4227-4239.

8. Sharma P, Yan F, Doronina V, Escuin-Ordinas H, Ryan MD, Brown JD. 2A peptides provide distinct solutions to driving stop-carry on translational recoding. *Nuc Acids Res.* 2012; 40(7): 1-9.
9. Machida K, Mikami S, Masutani M, Mishima K, Kobayashi T, Imataka H. A translation system reconstituted with human factors proves that processing of encephalomyocarditis virus proteins 2A and 2B occurs in the elongation phase of translation without eukaryotic release factors. *J Biol Chem*. 2014; 289(46): 31960-31971.
10. Ryan MD, Donnelly MLL, Lewis A, Mehrotra AP, Wilkie J, Gani D. A model for nonstoichiometric, co-translational protein scission in eukaryotic ribosomes. *Bioorg Chem*. 1999; 27: 55-79.
11. Brown JD, Ryan MD. Ribosome "Skipping": "Stop-Carry On" or "StopGo" Translation. In: Atkins JF, Gesteland RF, eds *Recoding: Expansion of Decoding Rules Enriches Gene Expression*. New York: Springer; 2010: 101-122.
12. de Felipe P, Ryan MD. Targeting of proteins derived from self-processing polyproteins containing multiple signal sequences, *Traffic*, 2004; 5: 616-626.
13. Lorens JB, Pearsall DM, Swift SE, *et al*. Stable, stoichiometric delivery of diverse protein functions. *J Biochem Bioph Meth*. 2004; 58: 101-110.
14. Szymczak AL, Workman CJ, Wang Y, *et al*. Correction of multi-gene deficiency *in vivo* using a single "selfcleaving" 2A peptide-based retroviral vector. *Nat Biotechnol*. 2004; 22(5): 589-594.
15. Provost E, Rhee J, Leach SD. Viral 2A peptides allow expression of multiple proteins from a single ORF in transgenic zebrafish embryos. *Genesis*. 2007; 45(10): 625-629.
16. Robertson BH, Grubman MJ, Weddell GN, *et al*. Nucleotide and amino acid sequence coding for polypeptides of foot-and-mouth disease virus type A12. *J Virol*. 1985; 54: 651-660.

17. Luke GA, Ryan MD. Using the 2A Protein Co-expression System: multi-cistronic 2A vectors expressing gene(s) of interest and reporter proteins. In: Daoiseaux R, Hasson S, eds. *Reporter Gene Assays: Methods and Protocols*. Springer New York, (in press).

18. Ryan MD, Luke GA, Hughes LE, *et al*. The aphtho- and cardiovirus "primary" 2A/2B polyprotein "cleavage". In: Semler BL, Wimmer E, eds *Molecular Biology of Picornaviruses*. Washington: ASM Press; 2002: 61-70.

19. Loughran G, Libbey JE, Uddowla S, *et al*. Theiler's murine encephalomyelitis virus contrasts with encephalomyocarditis and foot-and-mouth disease virus in its functional utilization of the StopGo non-standard translation mechanism. *J Gen Virol*. 2013; 94: 348-353.

20. Kjær J, Belsham GJ. Selection of functional 2A sequences within foot-and-mouth disease virus; requirements for the NPGP motif with a distinct codon bias. *RNA*. 2017; pii: rna.063339.117.doi: 10.1261/rna.063339.117

21. Gao ZL, Zhou JH, Zhang J, Ding YZ, Liu YS. (2014). The silent point mutations at the cleavage site of 2A/2B have no effect on the self-cleavage activity of 2A of foot-and-mouth disease virus. *Infect Genet Evol*. 2014; 28C: 101-106.

22. de Felipe P, Luke GA, Hughes LE, Gani D, Halpin C, Ryan MD. (2006). *E unum pluribus*: multiple proteins from a self-processing polyprotein. *Trends Biotechnol*. 2006; 24(2): 68-75.

23. Anderson RP, Voziyanova E, Voziyanov. YFlp and Cre expressed from Flp-2A-Cre and Flp-IRES-Cre transcription units mediate the highest level of dual recombinase-mediated cassette exchange. *Nucleic Acids Res*. 2012; 40:e62: PMID:22270085.

24. Ho SCL, Bardor M, Li B, *et al*. Comparison of internal ribosome entry site (IRES) and urin-2A (F2A) for monoclonal antibody expression level and quality in CHO cells. *Plos One*. 2013; 8:e63247; PMID:23704898.

25. Fang, J., Qian, J.J., Yi, S., *et al*. Stable antibody expression at therapeutic levels using the 2A peptide. *Nature Biotechnol*. 2005; 23(5): 584-590.

26. Fang J, Yi S, Simmons A, *et al*. An antibody delivery system for regulated expression of therapeutic levels of monoclonal antibodies *in vivo*. *Mol Ther*. 2007; 15(6): 1153-1159.

27. François IEJA, van Hemelrijck W, Aerts AM, *et al*. Processing in *Arabidopsis thaliana* of a heterologous polyprotein resulting in differential targeting of the individual plant defensins. *Plant Sci*. 2004; 166: 113-121.

28. Sun H, Lang Z, Zhu L, Huang D. Acquiring transgenic tobacco plants with insect resistance and glyphosate tolerance by fusion gene transformation. *Plant Cell Rep*. 2012; 31: 1877-1887.

29. Luke GA, de Felipe P, Lukashev A, Kallioinen SE, Bruno EA, Ryan MD. The occurrence, function, and evolutionary origins of "2A-like" sequences in virus genomes. *J Gen Virol*.2008; 89: 1036-1042.

30. Luke G.A, Ryan MD. The protein coexpression problem in biotechnology and biomedicine: virus 2A and 2A-like sequences provide a solution. *Future Virol*. *2013;* 8(10): 983-96.

31. Yang S, Cohen CJ, Peng PD, *et al*. Development of optimal bicistronic lentiviral vectors facilitates high-level TCR gene expression and robust tumor cell recognition. *Gene Ther*. 2008; 15(21): 1411-1423.

32. Tan Y, Liang H, Chen A, Guo X. Coexpression of double or triple copies of the rabies virus glycoprotein gene using a "self-cleaving" 2A peptide-based replication-defective human adenovirus serotype 5 vector. *Biologicals*. 2010; 38: 586-593.

33. Holst J, Vignali KM, Burton AR, *et al*. Rapid analysis of T-cell selection *in vivo* using t cell-receptor retrogenic mice. *Nat Methods*. 2006; 3(3): 191-197.

34. Luke, G.A. Translating 2A Research into Practice. In: Agbo EC (ed), *Innovations in Biotechnology*. Croatia: InTech; 2012: 161-186.

35. Rybicki EP. Plant-made vaccines for humans and animals. *Plant Biotechnol*. 2010; 8: 620-637.

36. Rybicki EP. Plant-based vaccines against viruses. *Virol J*. 2014; 11:205

37. Takeyama N, Kiyono H, Yuki Y. "Plant-based vaccines for animals and humans: recent advances in technology and clinical trials." *Ther Adv Vaccines*, 2015; 3(5-6): 139-154.

38. Lomonossoff GP, Johnson JE. Use of macromolecular assemblies as expression systems for peptides and synthetic vaccines. *Curr Opin Struc Biol.* 1996; 6: 176-182.

39. Lomonossoff GP, Evans DJ. Applications of plant viruses in bionanotechnology. *Curr Top Microbiol*. 2014; 375: 61-87.

40. Santa Cruz S, Chapman S, Roberts AG, Roberts IM, Prior DAM, Oparka KJ. Assembly and movement of a plant virus carrying a green fluorescent protein overcoat. *Proc Natl Acad Sci USA*. 1996; 93: 6286-6290.

41. O'Brien GJ, Bryant CJ, Voogd C, Greenberg HB, Gardner RC, Bellamy AR. Rotavirus VP6 expressed by PVX vectors in *Nicotiana benthamiana* coats PVX rods and also assembles into virus like particles. *Virology*. 2000; 270: 444-453.

42. Zelada AM, Calamante G, de la Paz Santangelo M, *et al*. Expression of tuberculosis antigen ESAT-6 in *Nicotiana tabacum* using a potato virus X-based vector. *Tuberculosis*. 2006; 86: 263-267.

43. Marconi G, Albertini E, Barone P, *et al*. In planta production of two peptides of the Classical Swine Fever Virus (CSFV) E2 glycoprotein fused to the coat protein of potato virus X. *BMC Biotechnol*. 2006; 6: 29.

44. Uhde-Holzem K, Schlösser V, Viazov S, Fischer R, Commandeur U. Immunogenic properties of chimeric potato virus X particles displaying the hepatitis C virus hypervariable region I peptide R9. *J Virol Methods*. 2010; 166: 12-20.

45. Smolenska L, Roberts IM, Learmonth D, *et al*. Production of a functional single chain antibody attached to the surface of a plant virus. *FEBS Lett*. 1998; 441: 379-382.

46. Carette N, Engelkamp H, Akpa E, *et al*. A virus-based biocatalyst. *Nat Nanotechnol*. 2007; 2: 226-229.

47. Chen TH, Hu CC, Liao JT, *et al*. Production of japanese encephalitis virus antigens in plants using bamboo mosaic virus-based vector. *Front Microbiol*. 2017; 8: 788.

48. Monreal-Escalante E, Baňuelos-Hernández B, Hernández M, *et al*. Expression of multiple *Taenia Solium* immunogens in plant cells through a ribosomal skip mechanism. *Mol Biotechnol*. 2015; 57: 635-643.

49. Zhang C, Bradshaw JD, Whitham SA, Hill JH. The development of an efficient multipurpose bean pod mottle virus viral vector set for foreign gene expression and RNA silencing. *Plant Physiol*. 2010; 153: 52-65.

50. Ma JK-C, Barros E, Bock R, *et al*. Molecular farming for new drugs and vaccines. *EMBO Rep*. 2005; 6: 593-599.

51. Hiatt A, Cafferkey R, Bowdish K. Production of antibodies in transgenic plants. *Nature*. 1989; 342: 76-78.

52. Ma JK-C, Hiatt A, Hein M, et al. Generation and Assembly of Secretory Antibodies in Plants. *Science*. 1995; 268: 716-719.

53. Baum TJ, Hiatt A, Parrott WA, Pratt LH, Hussey RS. Expression in tobacco of a functional monoclonal antibody specific to stylet secretions of the root-knot nematode. *Mol Plant Microbe In*. 1996; 9: 382-387.

54. Fischer R, Schumann D, Zimmermann S, Drossard J, Sack M, Schillberg S. Expression and characterization of bispecific single chain Fv fragments produced in transgenic plants. *Eur J Biochem.* 1999; 262: 810-816.

55. Makvandi-Nejad S, McLean MD, Hirama T, Almquist KC, Mackenzie CR, Hall JC. Transgenic tobacco plants expressing a dimeric single-chain variable fragment (scfv) antibody against *Salmonella enterica* serotype Paratyphi B. *Transgenic Res*. 2005; 15(5): 785-792.

56. Chen L, Yang X, Luo D, Yu W. Efficient production of a bioactive bevacizumab monoclonal antibody using the 2A self-cleavage peptide in transgenic rice callus. *Front Plant Sci*. 2016; 7: 1156.

57. Wang Y, Fei D, Vanderlaan M, Song A. Biological activity of bevacizumab, a humanized anti-VEGF antibody *in vitro*. *Angiogenesis*. 2004; 7: 335-345.

58. Ngo MC, Rooney CM, Howard JM, Heslop HE. *Ex vivo* gene transfer for improved adoptive immunotherapy of cancer. *Hum Mol Genet*. 2011; 20(R1): R93-R99.

59. Wang M, Yin B, Wang HY, Wang R-F. Current advances in T-cell-based cancer immunotherapy. *Immunotherapy*. 2014; 6(12): 1265-1278.

60. Roybal KT, Williams WJ, Morsut, L, *et al*. Engineering T cells with customized therapeutic response programs using synthetic notch receptors. *Cell*. 2016; 167: 419-432.

61. Rosenberg SA, Yang JC, Sherry RM., *et al*. Durable complete responses in heavily pretreated patients with metastatic melanoma using T-cell transfer immunotherapy. *Clin Cancer Res*. 2011; 17(13): 4550-4557.

62. Call ME, Wucherpfennig KW. Common themes in the assembly and architecture of activating immune receptors. *Nat Rev Immunol*. 2007; 7: 841-850.

63. Chinnasamy N, Wargo JA, Yu Z, *et al*. A TCR targeting the HLA-A*0201-restricted epitope of MAGE-A3 recognizes multiple epitopes of the MAGE-A antigen superfamily in several types of cancer. *J Immunol*. 2011; 186: 685-696.

64. Arber C, Abhyankar H, Heslop HE, *et al*. The immunogenicity of virus-derived 2A sequences in immunocompetent individuals. *Gene Ther*. 2013; 20(9): 958-962.

65. Zhou X, Dotti G, Krance RA, *et al*. Inducible caspase-9 suicide gene controls adverse effects from alloreplete T cells after haploidentical stem cell transplantation. *Blood*. 2015; 125(26): 4103-4113.

66. Cohen CJ, Zhao Y, Zheng Z, Rosenberg SA, Morgan RA. Enhanced antitumor activity of murine-human hybrid T-cell receptor (TCR) in human lymphocytes is associated with improved pairing and TCR/CD3 stability. *Cancer Res*. 2006; 66(17): 8878-8886.

67. Heemskerk MH, Hagedoom RS, van der Hoom MA, *et al*. Efficiency of T-cell receptor expression in dual-specific T cells is controlled by intrinsic qualities of the TCR chains within the TCR-CD3 complex. *Blood*. 2007; 109(1): 235-243.

68. Kuball J, Dossett ML, Wolfl M, *et al*. Facilitating matched pairing and expression of TCR chains introduced into human T cells. *Blood*. 2007; 109: 2331-2338.

69. Scholten KBJ, Kramer D, Kueter EWM, *et al*. Codon modification of T cell receptors allows enhanced functional expression in transgenic human T cells. *Clin Immunol*. 2006; 119: 135-145.

70. Okamoto S, Mineno J, Ikeda H, *et al*. Improved expression and reactivity of transduced tumor-specific TCRs in human lymphocytes by specific silencing of endogenous TCR. *Cancer Res*. 2009; 69(23): 9003-9011.

71. Okamoto S, Amaishi Y, Goto Y, *et al*. A promising vector for TCR gene therapy; differential effect of siRNA, 2A peptide, and disulfide bond on the introduced TCR expression. *Mol Ther-Nucl Acids*. 2012; 1,e63; doi10.1038/mtna.2012.52

72. Fesnak AD, June CH, Levine, BL. Engineered T cells: the promise and challenges of cancer immunotherapy. *Nat Rev Cancer*. 2016; 16: 566-577.

73. Lipowska-Bhalla G, Gilham DE, Hawkins RE, Rothwell DG. Isolation of tumor antigen-specific single-chain variable fragments using a chimeric antigen receptor bicistronic retroviral vector in a mammalian screening protocol. *Hum Gene Ther Method*. 2013; 24: 381-391.

74. Guest RD, Hawkins RE, Kirillova N, *et al*. The role of extracellular spacer regions in the optimal design of chimeric immune receptors: evaluation of four different scFvs and antigens. *J Immunother*. 2005; 28: 203-211.

75. Hudecek M, Lupo Stanghellini MT, Kosasih PL, *et al*. Receptor affinity and extracellular domain modifications affect tumor recognition by RORI-specific chimeric antigen receptor T-cells. *Clin Cancer Res*. 2013; 19(12): 3153-3164.

76. Wang X, Chang WC, Wong CW, *et al*. A transgene-encoded cell surface polypeptide for selection, *in vivo* tracking, and ablation of engineered cells. *Blood*. 2011; 118(5): 1255-1263.

77. Jonnalagadda M, Mardiros A, Urak R, *et al*. Chimeric antigen receptors with mutated IgG4 Fc spacer avoid Fc receptor binding and improve T cell persistence and antitumor efficacy. *Mol Ther*. 2015; 23(4): 757-768.

78. Jonnalagadda M, Brown CE, Chang WC, Ostberg JR, Forman SJ, Jensen MC. Engineering human T cells for resistance to methotrexate and mycophenolate mofetil as an *in vivo* cell selection strategy. *PLoS ONE*. 2013; 8(6): e65519.doi:10.1371/journal.pone.0065519

79. Philip B, Kokaiaki E, Mekkaoui L, *et al*. A highly compact epitope-based marker/suicide gene for easier and safer T-cell therapy. *Blood*. 2014; 124(8): 1277-1287.

80. Roellecke K, Virts EL, Einholz R, *et al*. Optimized human CYP4B1 in combination with the alkylator prodrug 4-ipomeanol serves as a novel suicide gene system for adoptive T-cell therapies. *Gene Ther*. 2016; 23: 615-626.

81. Wiek C, Schmidt EM, Roellecke K, *et al*. Identification of amino acid determinants in CYP4B1 for optimal catalytic processing of 4-Ipomeanol. *Biochem J*. 2015; 465(1): 103-114.

82. Pegram HJ, Park JH, Brentjens RJ. CD28z CARs and Armored CARs. *Cancer J*. 2014; 20(2): 127-133.

83. Wilkie S, Burbridge SE, Chiapero-Stanke L, *et al*. Selective expansion of chimeric antigen receptor-targeted T-cells with potent effector function using Interleukin-4. *J Biol Chem*. 2010; 285(33): 25538-25544.

84. Bray SJ. (2006) Notch signalling: a simple pathway becomes complex. *Nat Rev Mol Cell Bio*. 2006; 7: 678-689.

85. Gordon WR, Zimmerman B, He L, *et al*. Mechanical allostery: evidence for a force requirement in the proteolytic activation of Notch. *Dev Cell*. 2015; 33: 729-736.

86. Morsut L, Roybal KT, Xiong X, *et al*. Engineering customized cell sensing and response behaviors using synthetic Notch receptors. *Cell*. 2016; 164: 780-791.

87. Buchbinder EI, Desai A. CTLA-4 and PD-1 pathways – similarities, differences, and implications of their inhibition. *Am J Clin Oncol*. 2016; 39(1): 98-106.

88. Ott PA, Hodi FS, Robert C. CTLA-4 and PD-1/PD-L1 blockade: new immunotherapeutic modalities with durable clinical benefit in melanoma patients. *Clin Cancer Res*. 2013; 19(19): 5300-5308.

89. Dong Y, Sun Q, Zhang X. PD-1 and its ligands are important immune checkpoints in cancer. *Oncotarget*. 2017; 8(2): 2171-2186.

90. Gaj T, Gersbach CA, Barbas CF 3rd. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol*. 2013; 31(7): 397-405.

91. Gupta RM, Musunuru K. Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. *J Clin Invest*. 2014; 124(10): 4154-4161.

92. Ain QU, Chung JY, Kim YH. Current and future delivery systems for engineered nucleases: ZFN, TALEN and RGEN. *J Control Release*. 2015; 205: 120-127.

93. Iliakis G, Wang H, Perrault AR, *et al*. Mechanisms of DNA double strand break repair and chromosome aberration formation. *Cytogenet Genome Res*. 2004; 104: 14-20.

94. Kim YG, Cha J, Chandrasegaran S. Hybrid restriction enzymes: zinc finger fusions to *FokI* cleavage domain. *Proc Natl Acad Sci USA*. 1996; 93: 1156-1160.

95. Miller JC, Tan S, Qiao G, *et al*. A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol*. 2011; 29: 143-148.

96. Carroll D. Genome engineering with zinc-finger nucleases. *Genetics*. 2012; 188: 773-782.

97. Ding Q, Lee Y-K, Schaefer EAK, *et al*. A TALEN genome editing system to generate human stem cell-based disease models. *Cell Stem Cell*. 2013; 12(2): 238-251.

98. Kim H, Kim M-S, Wee G, Lee C-I, Kim H, Kim J-S. Magnetic separation and antibiotics selection enable enrichment of cells with ZFN/TALEN-induced mutations. *Plos One*. 2013; 8(2): e56476, doi: 10.1371/journal.pone.0056476.

99. Mariano A, Xu L, Han R. Highly efficient genome editing via 2A-coupled co-expression of two TALEN monomers. *BMC Res Notes*. 2014; 7:628.

100. Duda K, Lonowski LA, Kofoed-Nielsen M, *et al*. High-efficiency genome editing *via* 2A-coupled co-expression of fluorescent proteins and zinc finger nucleases or CRISPR/Cas9 nickase pairs. *Nucleic Acids Res*. 2014; 42(10), e84 doi: 10.109/nar/gku251.

101. Lonowski LA, Narimatsu Y, Riaz A, *et al*. Genome editing using FACS enrichment of nuclease-expressing cells and indel detection by amplicon analysis. *Nat Protoc*. 2017; 12(3):581-603.

102. Ren C, Xu K, Liu Z, *et al*. Dual-reporter surrogate systems for efficient enrichment of genetically modified cells. *Cell Mol Life Sci*. 2015; 72:2763-2772.

103. Certo MT, Ryu BY, Annis JE, *et al*. Tracking genome engineering outcome at individual DNA breakpoints. *Nat Methods*. 2011; 8(8): 671-676.

104. Provasi E, Genovese P, Lombardo A, *et al*. Editing T cell specificity towards leukemia by zinc finger nucleases and lentiviral gene transfer. *Nat Med*. 2012; 18: 807-815.

105. Li H, Haurigot V, Doyon Y, *et al*. *In vivo* genome editing restores haemostasis in a mouse model of haemophilia *Nature*. 2011; 475: 217-221.

106. Perez EE, Wang J, Miller JC, *et al*. Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat Biotechnol*. 2008; 26: 808-816.

107. Joglekar Jonnalagadda M, Brown CE, Chang WC, Ostberg JR, Forman SJ, Jensen MC. Engineering human T cells for resistance to methotrexate and mycophenolate mofetil as an *in vivo* cell selection strategy. *PLoS ONE*. 2013; 8(6): e65519.doi:10.1371/journal.pone.0065519

108. Yi G, Choi JG, Bharaj P, *et al*. CCR5 gene editing of resting CD4+T cells by transient ZFN expression from HIV envelope pseudotyped nonintegrating lentivirus confers HIV-1 resistance in humanized mice. *Mol Ther-Nucl Acids*. 2014; 3,e198.

109. Lederman MM, Veazey RS, Offord R, *et al*. Prevention of vaginal SHIV transmission in rhesus macaques through inhibition of CCR5. *Science*. 2004; 306(5695): 485-487.

110. Ellis BL, Hirsch ML, Porter SN, Samulski RJ, Porteus MH. Zinc-finger nuclease-mediated gene correction using single AAV vector transduction and enhancement by food and drug administration-approved drugs. *Gene Ther*. 2013; 20: 35-42.

111. Maier DA, Brennan AL, Jiang S, *et al*. Efficient clinical scale gene modification *via* zinc finger nuclease-targeted disruption of the HIV-1 co-receptor CCR5. *Hum Gene Ther*. 2013; 24: 245-258.

112. McPherron AC, Lawler AM, Lee SJ. Regulation of skeletal muscle mass in mice by a new TGF-beta superfamily member. *Nature*. 1997; 387: 83-90.

113. Lee SJ. Regualtion of muscle mass by myostatin. *Annu Rev Cell Dev Bi*. 2004; 20: 61-86.

114. Xu L, Zhao P, Mariano A, Han R. Targeted myostatin gene editing in multiple mammalian species directed by a single pair of TALE nucleases. *Mol Ther-Nucl Acids*. 2013; 2(7): e112.

115. Hockemeyer D, Soldner F, Beard C, *et al*. Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat Biotechnol*. 2009; 27(9): 851-857.

116. Dekelver RC, Choi VM, Moehle EA, *et al*. Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbour locus in the human genome. *Genome Res*. 2010; 20(8): 1133-1142.

117. Thompson DB, Aboulhouda S, Hysolli E, Smith CJ, Wang S, Castanon O, Church GM. The future of multiplexed eukaryotic genome engineering. *ACS Chem Biol*. 2018; 13(2):313-325.

118. Eisenhut P, Klanert G, Weinguny M, Baier L, Jadhav V, Ivansson D, Borth N. A CRISPR/Cas9 based engineering strategy for overexpression of multiple genes in chinese hamster ovary cells. *Metab Eng*. 2018; 48:72-81.

119. Renicke C, Schuster D, Usherenko S, Essen LO, Taxis CA. A LOV2 domain-based optogenetic tool to control protein degradation and cellular function. *Chem Biol*. 2013; 20: 619-626.

120. Shimizu-Sato S, Huq E, Tepperman JM, Quail PH. A light-switchable gene promoter system. *Nat Biotechnol*. 2002; 20: 1041–1044.

121. Levskaya A, Weiner OD, Lim WA, Voigt CA. Spatiotemporal control of cell signalling using a light-switchable protein interaction. *Nature*. 2009; 461: 997-1001.

122. Kennedy MJ, Hughes RM, Peteya LA, Schwartz JW, Ehlers MD, Tucker CL. Rapid blue-light-mediated induction of protein interactions in living cells. *Nat Methods*. 2010; 7: 973-975.

123. Chen D, Gibson ES, Kennedy MJ. A light-triggered protein secretion system. *J Cell Biol*. 2013; 201: 631-640.

124. Crefcoeur RP, Yin RH, Ulm R, Halazonetis TD. Ultraviolet-B-mediated induction of protein-protein interactions in mammalian cells. *Nat Commun*. 2013; 4:1779 doi: 10.1038/ ncomms2800.

125. Liu H, Yu X, Li K, *et al*. Photoexcited CRY2 interacts with CIB1 to regulate transcription and floral initiation in *Arabidopsis*. *Science*. 2008; 322: 1535-1539.

126. Konermann S, Brigham MD, Trevino AE., *et al*. Optical control of mammalian endogenous transcription and epigenetic states. *Nature*. 2013; 500: 472-476.

127. Kakumoto T, Nakata, T. Optogenetic control of PIP3: PIP3 is sufficient to induce the actin-based active part of growth cones and is regulated *via* endocytosis. *PLoS ONE*. 2013; 8: e70861.

128. Zhang K, Duan L, Ong Q, *et al*. Light-mediated kinetic control reveals the temporal effect of the Raf/MEK/ERK pathway in PC12 cell neurite outgrowth. *PLoS ONE*. 2014; 9: e92917.

129. Idevall-Hagren O, Dickson EJ, Hille B, Toomre DK, De Camilli P. Optogenetic control of phosphoinositide metabolism. *Proc Natl Acad Sci USA*. 2012; 109: E2316-E2323.

130. Krishnamurthy VV, Khamo JS, Mei W, *et al*. Reversible optogenetic control of kinase activity during differentiation and embryonic development. *Development*. 2016; 143: 4085-4094.

131. Niault TS, Baccarini M. Targets of Raf in tumorigenesis. *Carcinogenesis*. 2010; 31: 1165-1174.

132. Taslimi A, Zoltowski B, Miranda JG, Pathak GP, Hughes RM, Tucker CL. Optimized second-generation CRY2-CIB dimerizers and photoactivatable Cre recombinase. *Nat Chem Biol*. 2016; 12: 425-430.

133. Kawano F, Okazaki R, Yazawa M, Sato M. A photoactivatable Cre-loxP recombination system for optogenetic genome engineering. *Nat Chem Biol*. 2016; 12: 1059-1064.

134. Kawano F, Suzuki H, Furuya A, Sato M. Engineered pairs of distinct photoswitches for optogenetic control of cellular proteins. *Nat Commun*. 2015; 6: 6256.

135. Kim JH, Lee SR, Li LH, Park HJ, Lee KY. *et al*. High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines. zebrafish and mice. *PLos ONE*. 2011; 6(4): e18556.2011.

136. Chng J, Wang T, Nian R, Lau A, Hoi KM, Ho SCL. *et al*. Cleavage efficient 2A peptides for high level monoclonal antibody expression in CHO cells. *mAbs*. 2015; 7(2): 403-412.

137. Farré G, Blancquaert D, Capell T, van der Straeten D, Christou P, Zhu C. Engineering complex metabolic pathways in plants. *Annu Rev Plant Biol*. 2014; 65: 187-223.

138. Daniels RW, Rossano AJ, Macleod GT, Ganetzky B. Expression of multiple transgenes from a single construct using viral 2A peptides in *Drosophila*. *PLoS ONE*. 2014; 9: e100637.

139. Wang Y, Wang F, Wang R, Zhao P, Xia Q. 2A self-cleaving peptide-based multi-gene expression system in the silkworm *Bombyx mori*. *Sci Rep*. 2015; 5. DOI: 10.1038/srep16273.

140. Unkles SE, Valiante V, Mattern DJ, Brakhage AA. Synthetic biology tools for bioprospecting of natural products in eukaryotes. *Chem Biol*. 2014; 21: 502-508.

141. Amorim Araújo J, Ferreira TC, Rubini MR, Duran AGG, de Marco JL, de Moraes LMP. Coexpression of cellulases in *Pichia pastoris* as a self-processing protein fusion. *AMB Express*. 2015; 5: 84

142. Geier M, Fauland P, Vogl T, Glieder A. Compact multi-enzyme pathways in *P.pastoris*. *Chem Commun (Camb).* 2015; 51: 1643-1646.

143. Marino J, von Heijne G, Beckmann R. Small protein domains fold inside the ribosome exit tunnel. *FEBS Lett*. 2016; 590: 655-660.

144. Souza-Moreira T, Navarrete C, Chen X, Zanelli CF, Valentini SR, Furlan M. et al. Screening of 2A peptides for polycistronic gene expression in yeast. *FEMS Yeast Res*. 2018; 18(5), doi: 10.1093/femsyr/foy036.

. 145. Hofacre A, Yagiz K, Mendoza D, Lopez Espinoza F, Munday AW, Burrascano C, Singer O, Gruber HE, Jolly DJ, Lin AH. Efficient therapeutic protein expression using retroviral replicating vector with 2A peptide in cancer models. *Hum Gene Ther*. 2018; 29: 437-451.

146. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*. 2006; 126(4): 663-676.

147. Liu Z, Chen O, Wall JBJ, *et al*. Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. *Sci Rep-UK*. 2017; 7: 2193.

.148. Yu KK, Aguilar K, Tsai J, *et al*. Use of mutated self-cleaving 2A peptides as a molecular rheostat to direct simultaneous formation of membrane and secreted anti-HIV immunoglobulins. *PLoS ONE*. 2012; 7(110): e50438.

149. Cruz-Teran CA, Tiruthani K, Mischler A, Rao BM. Inefficient Ribosomal Skipping Enables Simultaneous Secretion and Display of Proteins in *Saccharomyces cerevisiae*. *ACS Synth Biol*. 2017. doi: 10.1021/acssynbio.7b00144.

150. Roulston C, Luke GA, de Felipe P, *et al*. "2A-Like" signal sequences mediating translational recoding: A novel form of dual protein targeting. *Traffic*. 2016; 17: 923-939.

151. Ryan MD, Roulston C, de Felipe P, Odon V, Tilsner J, Luke GA. The potential consequences for cell signaling by a class of NOD-like receptor proteins (NLRs) bearing an N-terminal signal sequence. *J Cell Signal*. 2017; 2: 2.

Figure 1. Stable antibody expression using the 2A peptide. The heavy chain (plus its signal sequence) and light chain (plus its signal sequence) are linked by 2A. Insertion of a furin recognition site upstream of 2A allows removal of 2A residues that would otherwise be attached to the heavy chain. Signal sequences direct the insertion of proteins into the membrane of the ER and are cleaved off by signal peptide peptidase (SPP). Examples of 2A/2A-like sequences used in biomedicine and biotechnology are shown below.

Figure 2. Schematic representation of the recombinant constructs based on the BaMV cDNA infectious copy. Organization of the wild-type BaMV, BaMV-[EDIII-CP] and BaMV-[EDIII-F2A-CP]. Boxes represent coding sequences. The predicted Mr values of the four proteins common to all constructs are indicated. Inoculation of plants with a recombinant virus genome encoding EDIII fused to CP [EDIII-CP] showed this genomic form to be non-infectious, whereas incorporation of F2A between EDIII and CP [EDIII-F2A-CP] was infectious and generated both EDIII-2A-CP fusion protein and free CP in plant cells.

Figure 3. Co-expression of a chimeric IL-4 receptor (4αβ) and a CAR specific for tumour-associated MUC1 (HOX) in human T-cells. (A) Co-expression of 4αβ and a MUC1-specific CAR (HOX) was achieved using an intervening T2A peptide placed downstream of a furin cleavage site (-RRKR-) and SGSG linker. (B) Key features of the HOX fusion receptor are: an HMFG2 (human milk fat globule 2) scFv, that specifically binds to the cell surface glycoprotein Mucin 1(MUC1), an IgD hinge providing flexibility and a tripartite signaling domain.

Figure 4. Schematic outline of the Traffic Light Reporter assay. Arrow represents promoter and initial GFP start codon. The reading frame of each fluorescent protein is indicated in superscript. GFP expression can be recovered upon HDR-mediated repair following delivery of an appropriate homologous template and cells will fluoresce green. If the dsB undergoes NHEJ repair, GFP will be translated out of frame and the T2A and mCherry sequences are rendered in frame – leading to mCherry production.

Figure 5. Light controlled activation of the Raf/MEK/ERK signaling cascade. (A) Schematic of the CRY2-P2A-CIBN construct. Upon translation, one mRNA transcript generates two proteins: CRY2-mCherry-Raf1-P2A and CIBN-GFP-CaaX. (B) The CIBN domain is anchored

to the plasma membrane *via* a CaaX motif. Upon light stimulation, CIBN-CRY2PHR interaction recruits cytoplasmic CRY2PHR-mCherry-Raf1 to the plasma membrane which activates the Raf/MEK/ERK signaling pathway. In the absence of light, spontaneous dissociation of CIBN-CRY2PHR returns Raf1 to the cytoplasm and inactivates ERK.

CONFLICTS OF INTEREST.

The authors state that there are no conflicts of interest.

Figure 1

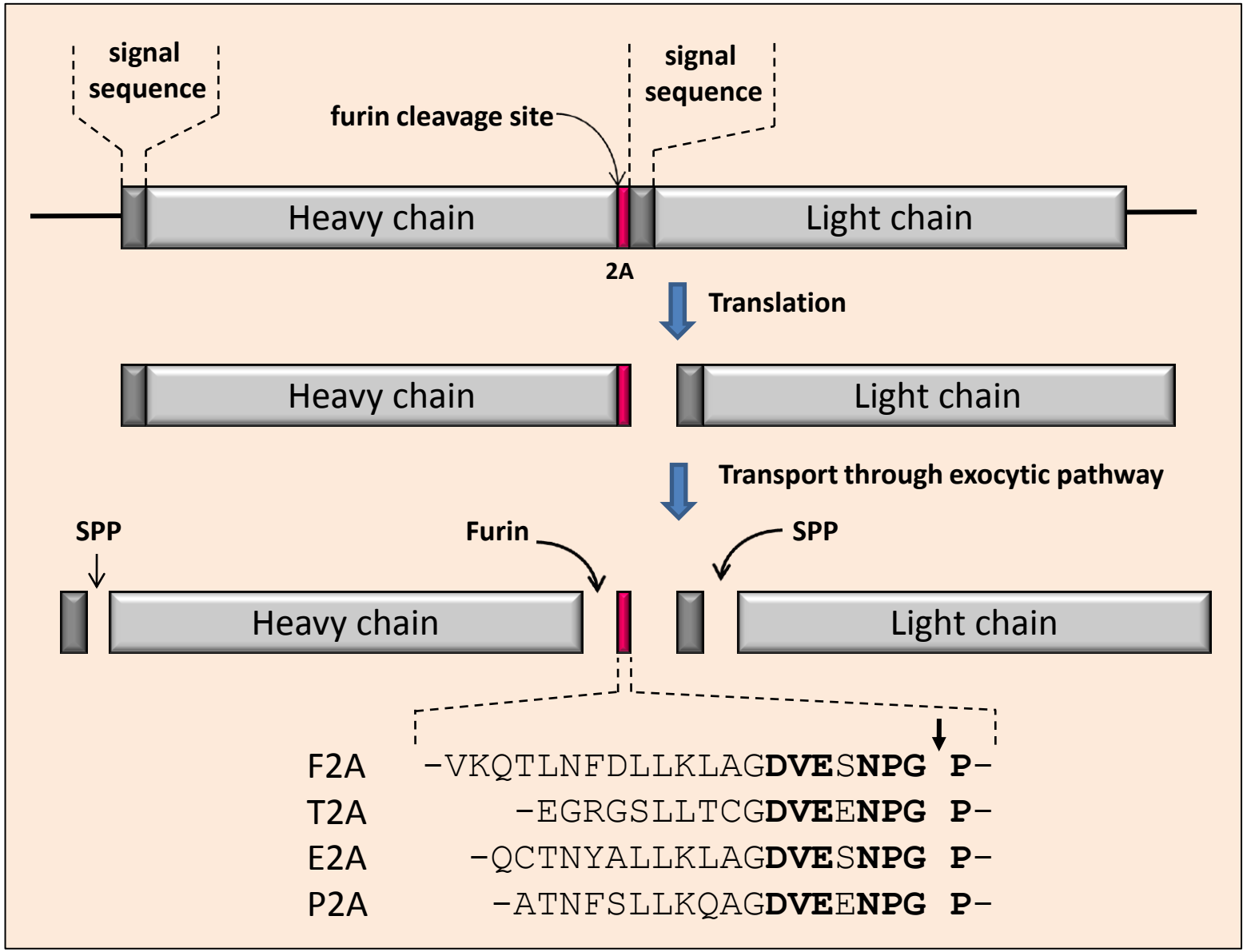

signal sequence
furin cleavage site
signal sequence
Heavy chain
Light chain
2A
Translation
Heavy chain
Light chain
Transport through exocytic pathway
SPP
Furin
SPP
Heavy chain
Light chain
F2A -VKQTLNFDLLKLAGDVESNPG P-
T2A -EGRGSLLTCGDVEENPG P-
E2A -QCTNYALLKLAGDVESNPG P-
P2A -ATNFSLLKQAGDVEENPG P-

Figure 2

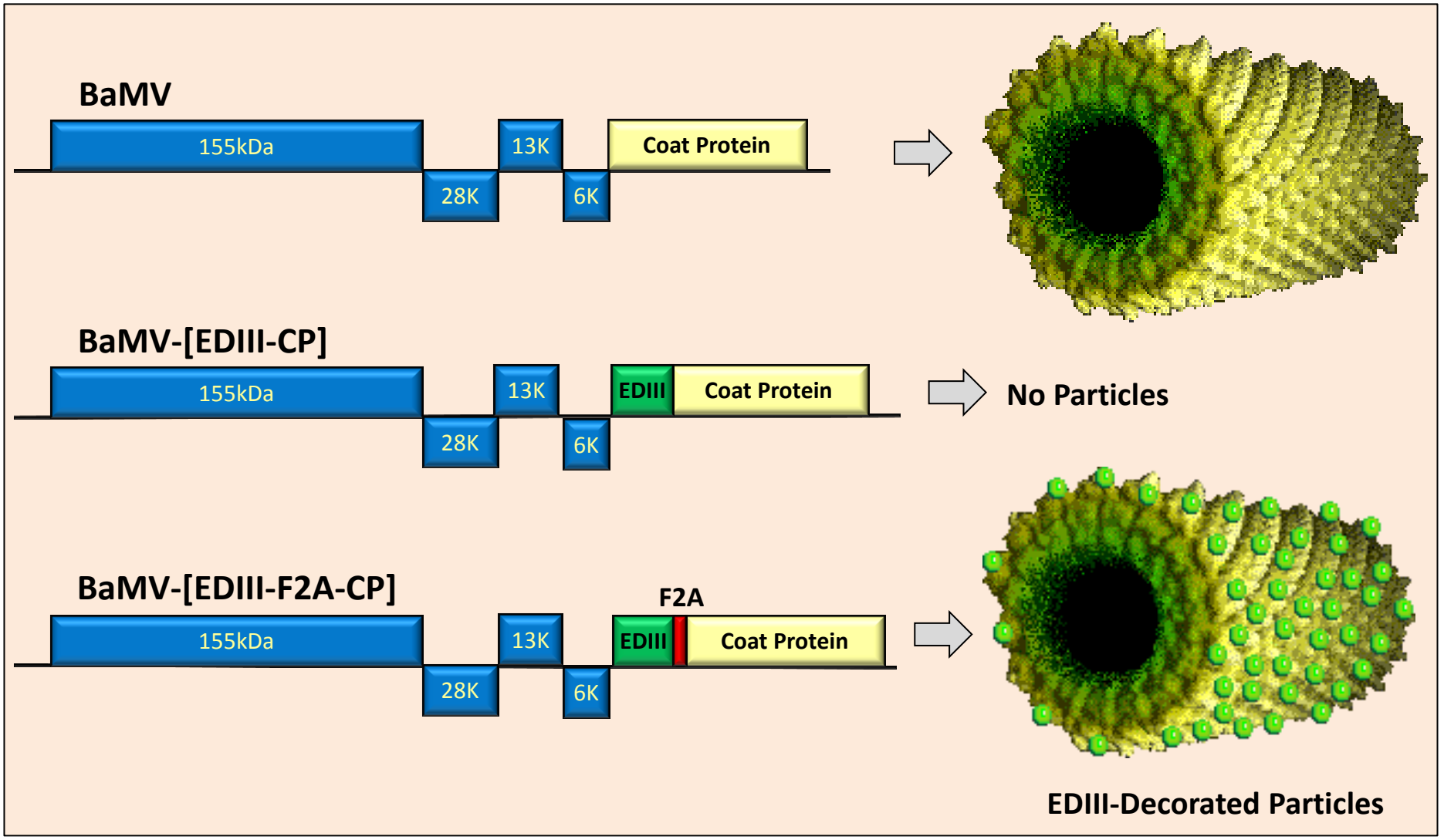

BaMV
155kDa
28K
13K
6K
Coat Protein
BaMV-[EDIII-CP]
155kDa
28K
13K
6K
EDIII
Coat Protein
No Particles
BaMV-[EDIII-F2A-CP]
155kDa
28K
13K
6K
F2A
EDIII
Coat Protein
EDIII-Decorated Particles

Figure 3

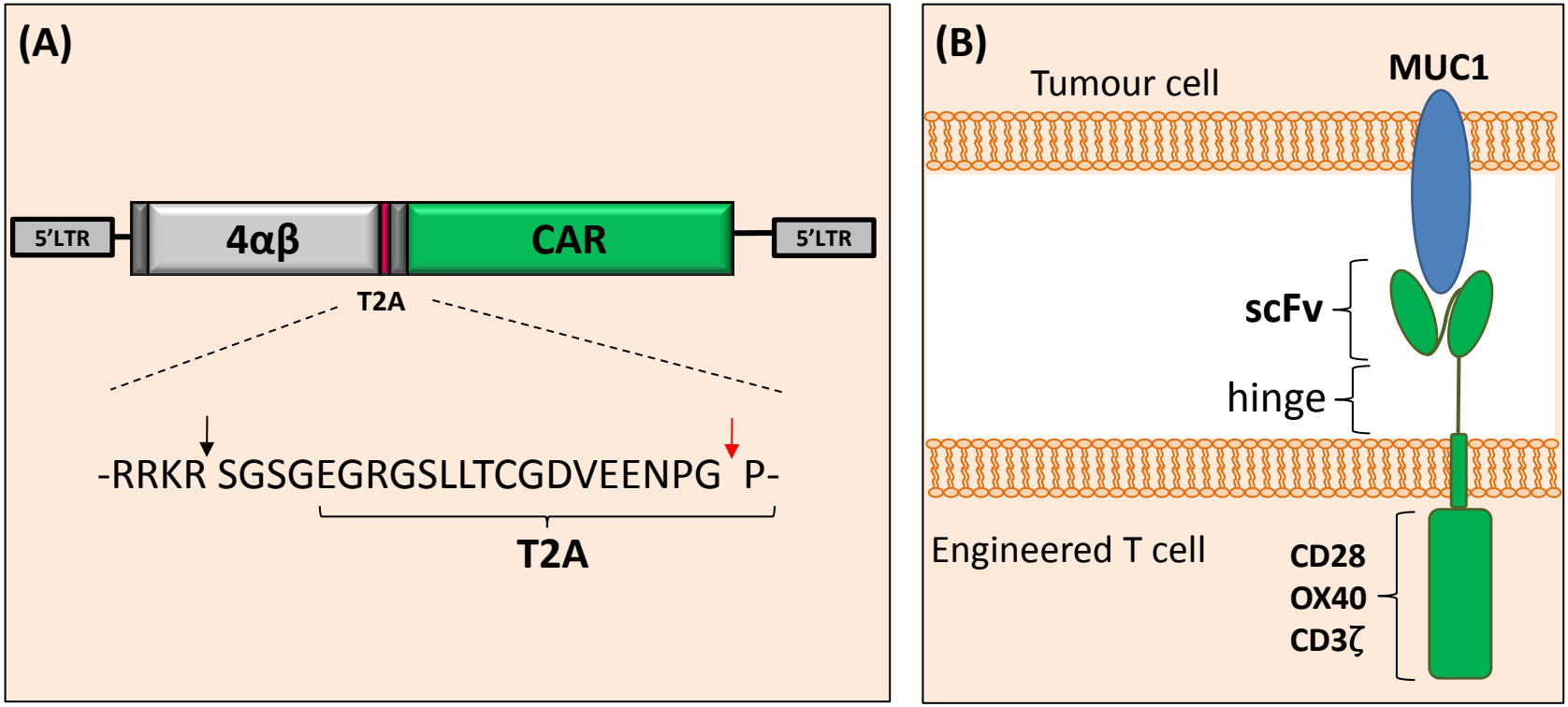

(A)
5'LTR
4αβ
CAR
5'LTR
T2A
-RRKR SGSGEGRGSLLTCGDVEENPG P-
T2A
(B)
Tumour cell
MUC1
scFv
hinge
Engineered T cell
CD28
OX40
CD3ζ

Figure 4

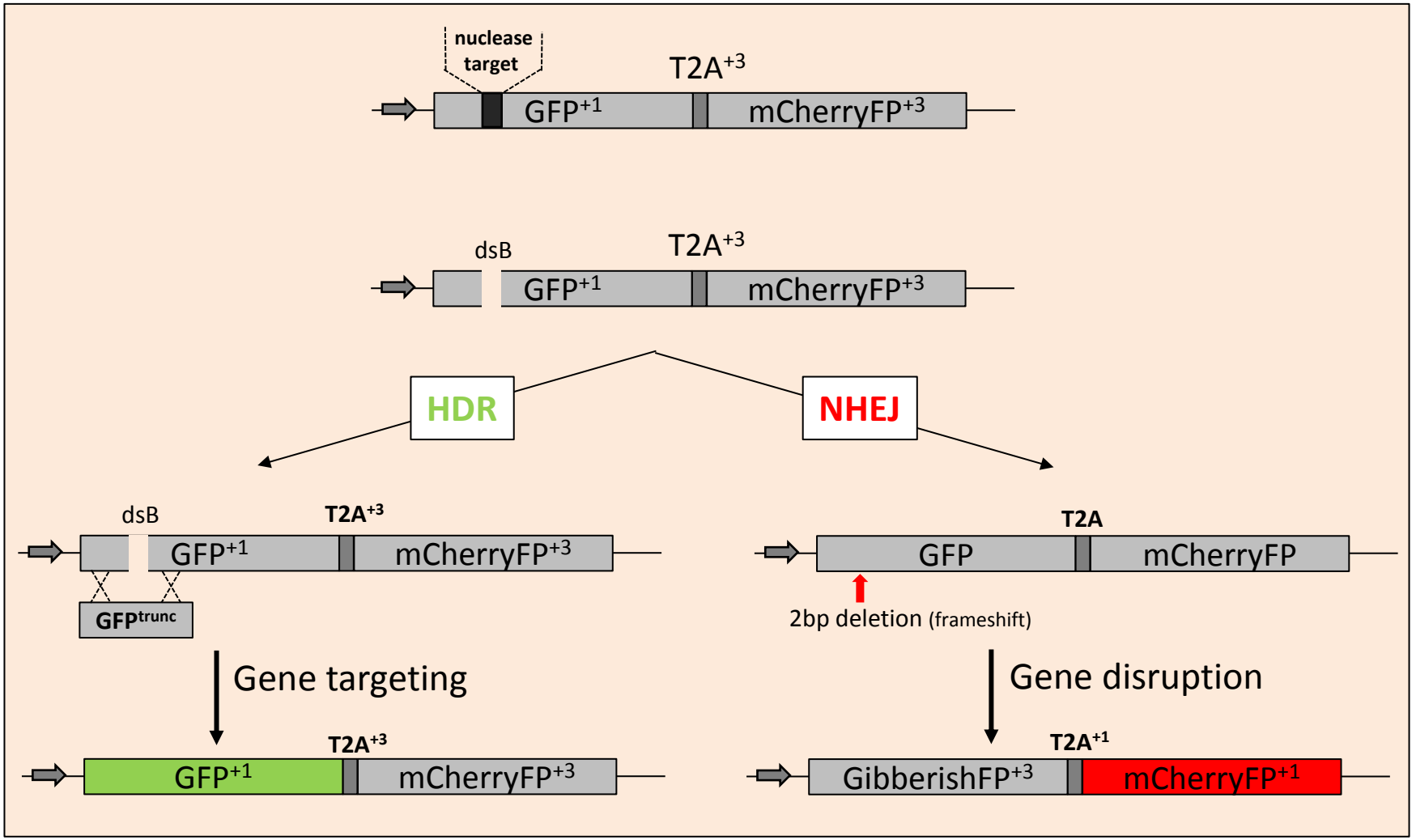
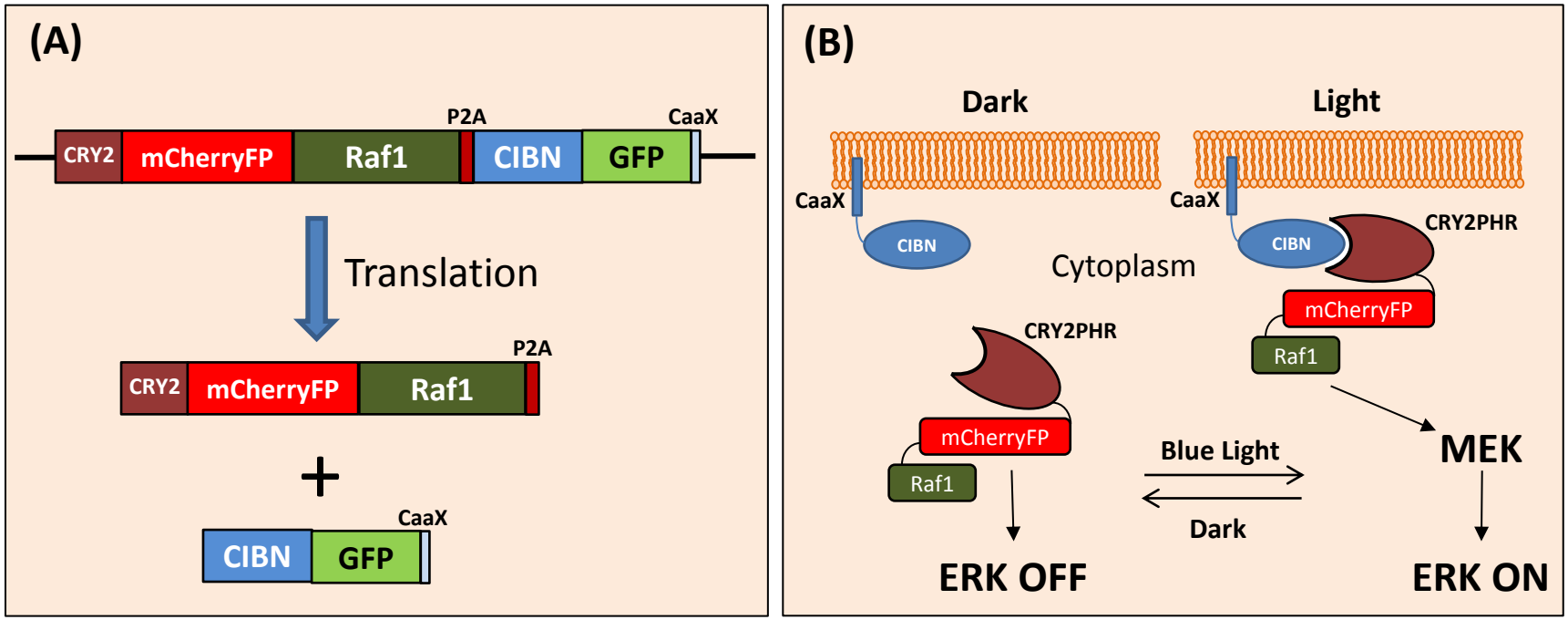

nuclease target
T2A+3
GFP+1
mCherryFP+3
dsB
T2A+3
GFP+1
mCherryFP+3
HDR
NHEJ
dsB
T2A+3
GFP+1
mCherryFP+3
GFPtrunc
T2A
GFP
mCherryFP
2bp deletion (frameshift)
Gene targeting
Gene disruption
T2A+3
GFP+1
mCherryFP+3
T2A+1
GibberishFP+3
mCherryFP+1

Figure 5


(A)
CRY2
mCherryFP
Raf1
P2A
CIBN
GFP
CaaX
Translation
CRY2
mCherryFP
Raf1
P2A
+
CIBN
GFP
CaaX
(B)
Dark
Light
CaaX
CIBN
Cytoplasm
CaaX
CIBN
CRY2PHR
mCherryFP
Raf1
CRY2PHR
mCherryFP
Raf1
Blue Light
Dark
MEK
ERK OFF
ERK ON