



- (51) **International Patent Classification:**
C12N 15/86 (2006.01)
- (21) **International Application Number:**
PCT/GB2017/050481
- (22) **International Filing Date:**
24 February 2017 (24.02.2017)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
1603372.2 26 February 2016 (26.02.2016) GB
- (71) **Applicant:** UCL BUSINESS PLC [GB/GB]; The Network Building, 97 Tottenham Court Road, London W1T 4TP (GB).
- (72) **Inventors:** PULÉ, Martin; c/o UCL Business PLC, The Network Building, 97 Tottenham Court Road, London W1T 4TP (GB). MEKKAOUI, Leila; c/o UCL Business PLC, The Network Building, 97 Tottenham Court Road, London W1T 4TP (GB).
- (74) **Agent:** WILLIAMS, Aylsa; D Young & Co LLP, 120 Holborn, London EC1N 2DY (GB).
- (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, DJ, 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, KH, KN, KP, KR, KW, 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:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))



WO 2017/144892 A1

(54) **Title:** NUCLEIC ACID CONSTRUCTS FOR PRODUCING RETROVIRAL VECTORS

(57) **Abstract:** The present invention relates to a nucleic acid construct comprising: (i) a first nucleic acid sequence which either comprises a retroviral transfer vector or which encodes a retroviral protein; and (ii) a second nucleic acid sequence which encodes a detectable marker which is a cell surface protein comprising an extracellular domain and a membrane targeting domain.

NUCLEIC ACID CONSTRUCTS FOR PRODUCING RETROVIRAL VECTORS

FIELD OF THE INVENTION

5 The present invention relates to the field of retroviral vectors. In particular, the invention relates to nucleic acid constructs and methods for producing retroviral vectors and to packaging cells and producer cells for use in such methods.

BACKGROUND TO THE INVENTION

10

Retroviral vectors are relevant for a range of applications, including gene therapy. However, progress in lentiviral gene therapy, for example, has been hampered by the requirement for production of purified lentiviral vectors with high titre.

15 Methods for generating retroviral packaging cell lines are known in the art (see WO 92/05266, for example). Such packaging cell lines may be used to create producer cell lines for the production of retroviral vector particles.

Lentiviral production may be performed using transient transfection of 293T cells with four
20 plasmids: the viral genome vector plasmid and three helper plasmids which supply Gagpol, Rev and Env glycoprotein. However, transient transfection technology cannot be used to generate virus at an appropriate scale for clinical applications because very large scale transfections are not practicable.

25 Stable lentiviral packaging cell lines can be generated by stably integrating all four of the above mentioned plasmids into a cell line (e.g. a 293T cell line) by transient transfection, followed by a serendipitous integration event. Cells which have integrated the plasmid are selected by the co-expression of an antibiotic resistance gene from the integrated plasmid. This approach is limited for a number of reasons. Firstly, stable integration following
30 transient transfection is very inefficient. Second, antibiotic selection works well for a limited range of antibiotics, for example ZeocinTM and puromycin, but options are limited when attempting to select for the integration of four different genes and the available selection systems are not optimal. Thirdly, it is difficult to determine the relative expression obtained from each individual integration event.

35

Thus there is a need for retroviral, in particular lentiviral, production methods which are not associated with the disadvantages outlined above.

SUMMARY OF ASPECTS OF THE INVENTION

The present inventors provide a retroviral production system which is based on the co-expression of viral proteins and detectable marker proteins which can be quantified at the level of a single cell. In particular, the system enables the expression level of different retroviral proteins in packaging cells to be tracked and packaging cells with ideal expression ratios for different retroviral components to be selected. For example, the expression level of the detectable marker may be determined using flow cytometry to identify and select packaging cells. Because the marker protein is co-expressed with the viral protein, packaging cell which express a particular viral protein or a combination of viral proteins at levels or ratios which are advantageous for a retroviral production system can be identified and selected.

In a first aspect the present invention provides a nucleic acid construct comprising:

- (i) a first nucleic acid sequence which either comprises a retroviral transfer vector or which encodes a retroviral protein; and
- (ii) a second nucleic acid sequence which encodes a detectable marker which is a cell surface protein comprising an extracellular domain and a membrane targeting domain.

The nucleic acid construct may comprise the following structure:

A-X-B

in which

A is a nucleic acid sequence which either comprises a retroviral transfer vector or which encodes a retroviral protein,

B is a nucleic acid sequence encoding a detectable marker which is a cell surface protein comprising an extracellular domain and a membrane targeting domain and

X is a co-expression sequence which enables the retroviral protein and the detectable marker to be expressed as separate polypeptides.

The co-expression sequence may comprise an IRES sequence or a sequence encoding a self-cleaving peptide.

The co-expression sequence may be selected from a 2A self-cleaving peptide from an aphtho- or a cardiovirus, a 2A-like peptide, an internal ribosome entry site, a furin cleavage site or a Tobacco Etch Virus cleavage site.

The co-expression sequence may be a 2A self-cleaving peptide from an aphtho- or a cardiovirus or a 2A-like peptide.

The retroviral protein may be selected from gag-pol, env or rev.

5

The cell surface protein may be less than 200 amino acids.

The membrane targeting domain of the cell surface protein may comprise (i) a transmembrane domain and an endodomain; or (ii) a GPI anchor. The transmembrane domain may comprise a CD8 stalk or a part thereof.

10

The extracellular domain of the cell surface protein may comprise HA, V5, RQR8 or MYC

The cell surface protein may comprise an amino acid sequence shown as SEQ ID NO: 21-26 or a variant thereof which has at least 80% sequence identity.

15

The nucleic acid construct may further comprise transposon sequences flanking the A-X-B nucleic acid sequence. The transposon sequences may comprise the sequence shown as SEQ ID NO: 60 and 61 or a variant thereof sharing at least 80% sequence identity.

20

In a second aspect the present invention provides a kit comprising a plurality of nucleic acid constructs according to the present invention.

The kit may comprise:

(i) a first nucleic acid construct comprising a *gag-pol* sequence and a sequence encoding a first detectable marker;

25

(ii) a second nucleic acid construct comprising an *env* sequence and a sequence encoding a second detectable marker;

in which the first and second detectable markers are cell surface proteins comprising an extracellular domain and a membrane targeting domain and are different to each other.

30

The kit may further comprise:

(iii) a third nucleic acid construct comprising a *rev* sequence and a sequence encoding a third detectable marker which is a cell surface protein comprising an extracellular domain and a membrane targeting domain;

35

in which the first, second and third and detectable markers are all different to each other.

The kit may also comprise:

(iv) a fourth nucleic acid construct comprising a retroviral transfer vector and a sequence encoding a fourth detectable marker which is a cell surface protein comprising an extracellular domain and a membrane targeting domain;

5 in which the first, second, third (if present) and fourth detectable markers are all different to each other.

In a third aspect the present invention provides a plasmid comprising a nucleic acid construct of the present invention.

10

In a fourth aspect the present invention provides a packaging cell comprising a nucleic acid sequence of the invention which expresses at least one detectable marker which is a cell surface protein comprising an extracellular domain and a membrane targeting domain.

15 The term “co-expresses” means that the retroviral protein and the detectable marker are expressed together or at the same time. As a result, the expression level of the retroviral protein can be determined indirectly by assaying the expression level of the detectable marker with which it is co-expressed.

20 The packaging cell may co-express each of *gag-pol*, *env* and optionally *rev* with a different detectable marker.

The packaging cell may comprise:

25 (i) a first nucleic acid construct comprising a *gag-pol* sequence and a sequence encoding a first detectable marker;

(ii) a second nucleic acid construct comprising an *env* sequence and a sequence encoding a second detectable marker;

which co-expresses the first and second detectable markers at the cell surface.

30 The packaging cell may further comprise:

(iii) a third nucleic acid construct comprising a *rev* sequence and a sequence encoding a third detectable marker;

which co-expresses the first, second and third detectable markers at the cell surface.

35 In a fourth aspect the present invention provides a producer cell which comprises:

(i) a first nucleic acid construct comprising a *gag-pol* sequence and a sequence encoding a first detectable marker;

(ii) a second nucleic acid construct comprising an *env* sequence and a sequence encoding a second detectable marker;

(iii) optionally a third nucleic acid construct comprising a *rev* sequence and a sequence encoding a third detectable marker;

5 (iii) a fourth nucleic acid construct comprising a retroviral transfer vector and a sequence encoding a fourth detectable marker;

which co-expresses the first, second, third (if present) and fourth detectable markers at the cell surface.

10 The nucleic acid construct(s) may be stably integrated into the cell genome.

The packaging or producer cell may be a HEK293, HEK293-T, TE671 or HT1080, 3T3, or K562 cell.

15 In another aspect the present invention provides a method for making a packaging cell or producer cell of the invention which comprises the step of introducing one or more nucleic acid construct(s) of the invention, one or more plasmid(s) of the invention, a kit of nucleic acid constructs of the invention, or a kit of plasmids of the invention, into a cell.

20 In another aspect the present invention provides a method for selecting a packaging or producer cell of the invention by selecting for expression of the or each detectable marker encoded by the or each nucleic acid construct.

The cell may express a plurality of detectable markers and may be selected using multi-
25 parameter flow cytometry.

The method may comprise the following steps:

(a) introducing

30 (i) a first nucleic acid construct comprising a *gag-pol* sequence and a sequence encoding a first detectable marker;

(ii) a second nucleic acid construct comprising an *env* sequence and a sequence encoding a second detectable marker;

into a plurality of cells; and

(b) selecting a cell which co-expresses the first and second detectable markers.

35

Step (a) may further comprise introducing into the plurality of cells:

(iii) a third nucleic acid construct comprising a *rev* sequence and a sequence encoding a third detectable marker;

and step (b) further comprises selecting a cell which co-expresses the first, second and third detectable markers.

5

Step (a) may further comprise introducing into the plurality of cells:

(iv) a fourth nucleic acid construct comprising a retroviral transfer vector and a sequence encoding a fourth detectable marker;

and step (b) further comprises selecting a cell which co-expresses the first, second, fourth

10 and optionally third detectable markers.

Step (a) may also comprise introducing a nucleic acid sequence encoding piggyBAC transposase.

15 In another aspect the invention provides a method for producing a retroviral vector which comprises the steps of culturing a producer cell of the present invention and isolating the retroviral vector.

The retroviral vector may be a lentiviral vector.

20

DESCRIPTION OF THE FIGURES

Figure 1 shows a scheme for generation of a stable lentiviral producer line. Helper plasmids supplying gagpol, envelope (in this case RD-Pro) and *rev* also co-express surface expressed epitope tags (HA-tag, myc-tag and V5-tag respectively). One of the transfer vector transgenes is RQR8 which, like the tags, is easily detected on the surface of the packaging cell. In addition, all 4 of these expression cassettes are flanked by piggyBAC terminal repeats. Plasmids are transfected into 293T cells, along with an expression plasmid supplying the piggyBAC transposase. The transfer and helper cassettes are hence transposed into the 293T cells. After a period of culture, permanent expression of helper and transgene cassettes can easily be detected by flow cytometry. Large numbers of these “bulk” transposed cells are sorted by flow cytometric sorting, with precise population(s) of cells with high transgene and optimal relative expression being selected. These are directly sorted as single-cells by the flow-sorter. They are allowed to expand to fill the well of a 96 or 385 well plate and supernatant is titered. The highest titer clones are selected for virus production.

35

Figure 2 shows the type of compact surface marker genes generated. (a) V5, HA and MYC-tags are expressed on a CD8-stalk and transmembrane domain. A small portion of the endodomain of CD8 is retained which contains polar residues which anchor the transmembrane protein and prevent it from being secreted from the membrane. An
5 alternative strategy is to GPI-anchor the tags – with V5, HA and MYC-tags being shown attached to a GPI anchor. Here, the tag is attached to a GPI-anchor signal which results in surface enzymatic attachment of the tags to GPI anchors on the cell surface. Both strategies result in high levels of surface expression with a very compact coding sequence. (b) To test surface expression, the 6 different types of surface epitope-tags were co-expressed with
10 enhanced blue fluorescent protein 2 (eBFP2). (c) 293T cells were transfected with these plasmids and the cells stained with anti-tag antibodies and a fluorescent secondary antibody. The cells were then analysed by flow-cytometry and expression of the tag compared with fluorescent signal of eBFP2.

Figure 3 shows the permanent expression of a helper transgene up to 95 days. (a) 293T cells were co-transfected with a piggyBAC transposase expression plasmid and a lentiviral rev expression plasmid which also co-expressed V5-8 marker gene. Some 293T cells were also transfected with just the rev expression plasmid alone to determine contribution to expression from transient transfection as opposed to permanent insertion. At 10 days, a
20 stable population of V5-8 expression cells can be seen in the transposase co-transfected cells, while expression in the rev expression alone 293T cells is fading. (b) V5-8 expression for both these 293T cell population over time up to 95 days in culture.

Figure 4 shows the importance of selecting the correct co-expression mechanism with gagpol. (a) Three lentiviral gagpol expression cassettes were tested. All three were flanked
25 by piggyBAC terminal repeats (pB-TR). Expression was driven by a CMV promoter. The first plasmid (WT) had no epitope tag co-expression. The second plasmid (tag-2A) had an HA-8 tag coded 5' to the gagpol open-reading frame separated by a FMD-disease 2A like peptide coding sequence. In the third plasmid (IRES.tag), and internal ribosome entry sequence
30 (IRES) was inserted 3' to the gagpol frame and the sequence coding for HA-8 followed the IRES. (b) 293T cells transfected with these plasmids were stained with an anti-HA antibody and analysed by flow-cytometry. The tag could be readily detected in both tag-2A and IRES.tag constructs. (c) Lentiviral supernatant was generated by transiently transfecting helper plasmids and transfer vector where the gagpol was supplied by either one of the
35 above three plasmids. Tag-2A construct resulted in a lower virus titer.

Figure 5 shows design of optimal lentiviral helper expression cassettes. These cassettes supply rev, RD-PRO and gagpol. Each is flanked by piggyBAC terminal repeats (pB-TR). Promoters, polyadenylation sequences are different to reduce the probability of homologous recombination. Each lentiviral element is co-expressed with a surface epitope marker. Co-expression in the Rev and RD-PRO cassette is achieved with a FMD like 2A peptide. Co-expression in the gagpol cassette is achieved with an IRES. RD-PRO requires an intron for most efficient expression. This is provided by the mouse EFalpha 5' UTR. The gagpol cassette also contains a scaffold attachment region (SAR).

Figure 6 shows 293T cells stably transposed with lentiviral helper cassettes and a transfer vector. The helper cassettes co-express HA, V5 and MYC tags, while the transfer vector expresses RQR8. All 4 tags can be independently detected and populations of cells which express desired amounts of the different elements can be gated (an example gate is shown) and flow-sorted.

Figure 7 shows lentiviral vector production from cells transposed with a tagged gagpol expression cassette. (a) 293T cells were transposed. After expression stabilized, 293T cells were sorted into low, medium and high expressing populations using flow-sorting of tag expression; (b) Expression was confirmed immediately after sorting; (c) 293T cells were then transfected with RD114 envelop, transfer vector and REV. Supernatant was harvested and tittered on 293T cells. The lowest expressing cells have the highest titer.

Figure 8 shows lentiviral vector production from cells transposed with a tagged transfer vector. 293T cells were transposed with a lentiviral transfer vector which also expressed a surface marker. One expression had stabilized, 293T cells were flow sorted into high and low expressing populations using flow-sorting on the marker gene.

Figure 9 shows a map of a transposable tagged transfer vector cassette. The 5' end starts with the piggyBAC 5' terminal repeat (5TR). Next, a CMV promoter fused to the 5' lentiviral long-terminal repeat drives expression. Next, the lentiviral packaging signal (PSI) is followed by the lentiviral REV-response element (RRE). Following this is the mature lentiviral expression cassette which has an EF1alpha promoter driving expression of the transgene which is a fusion between the surface marker gene RQR8 and a CD19 chimeric antigen receptor (CAR). An optimized woodchuck pre-processing element follows (mtPRE) which leads onto the lentiviral 3'LTR truncated to be self-inactivating. Transcription termination is re-enforced by an SV40 polyadenylation signal (SV40pA). Finally, the mammalian cassette

ends with the PiggyBAC 3' terminal repeats. The rest of the plasmid contains elements for bacterial propagation.

Figure 10 shows a map of a transposable gagpol expression cassette. The 5' end starts with the piggyBAC 5' terminal repeat (5TR). This is followed by a ubiquitin opening element (UCOE). This is followed by the CMV promoter and an intron. Next, is the staggered lentiviral gagpol open reading frame. Cloned in frame with the end of this is a 2A peptide and an epitope tag on a CD8 stalk (CD8STK). Next is a scaffold attachment region (SAR), a pre-processing element (PRE) and the polyadenylation signal from the rabbit beta-globin gene (RGBpA). The mammalian cassette ends with the piggyBAC 3' terminal repeat (3TR). The rest of the plasmid contains elements for bacterial propagation.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides a nucleic acid construct comprising

- (i) a first nucleic acid sequence which either comprises a retroviral transfer vector or which encodes a retroviral protein; and
- (ii) a second nucleic acid sequence which encodes a detectable marker which is a cell surface protein comprising an extracellular domain and a membrane targeting domain.

20

The nucleic acid construct may comprise the structure: A-X-B; in which A is the first nucleic acid sequence, B is the second nucleic acid sequence and X is a co-expression sequence.

The nucleic acid sequence may be an RNA or DNA sequence or a variant thereof.

25

RETROVIRUS

The concept of using viral vectors for gene therapy is well known (Verma and Somia (1997) Nature 389:239-242). As used herein the term "retroviral vector", when referring to a retroviral vector system also includes a retroviral vector particle capable of transducing a recipient cell with a nucleotide of interest (NOI).

30

A retroviral vector particle includes the following components: a vector genome, which may contain one or more NOIs, a nucleocapsid encapsidating the nucleic acid, and a membrane surrounding the nucleocapsid.

The term "nucleocapsid" refers to at least the group specific viral core proteins (gag) and the viral polymerase (pol) of a retrovirus genome. These proteins encapsidate the packagable

35

sequences and are themselves further surrounded by a membrane containing an envelope glycoprotein.

The term “vector genome” refers to both to the RNA construct present in the retroviral vector particle and the integrated DNA construct. The term also embraces a separate or isolated DNA construct capable of encoding such an RNA genome. A retroviral genome should
5 comprise at least one component part derivable from a retrovirus. The term “derivable” is used in its normal sense as meaning a nucleotide sequence or a part thereof which need not necessarily be obtained from a virus such as a lentivirus but instead could be derived therefrom. By way of example, the sequence may be prepared synthetically or by use of
10 recombinant DNA techniques.

There are many retroviruses. For the present application, the term “retrovirus” includes, but is not limited to: murine leukemia virus (MLV), human immunodeficiency virus (HIV), equine infectious anaemia virus (EIAV), mouse mammary tumour virus (MMTV), Rous sarcoma virus (RSV), Fujinami sarcoma virus (FuSV), Moloney murine leukemia virus (Mo-MLV), FBR
15 murine osteosarcoma virus (FBR MSV), Moloney murine sarcoma virus (Mo-MSV), Abelson murine leukemia virus (A-MLV), Avian myelocytomatosis virus-29 (MC29), and Avian erythroblastosis virus (AEV) and all other retroviridae including lentiviruses.

A detailed list of retroviruses may be found in Coffin et al (“Retroviruses” 1997 Cold Spring Harbour Laboratory Press Eds: JM Coffin, SM Hughes, HE Varmus pp 758-763).

20 In a preferred embodiment, the retroviral vector is derivable from a lentivirus.

Lentiviral vectors are major tools for gene delivery, providing efficient transduction of a wide variety cell types such as hematopoietic stem cells, neurons and endothelial cells. The advantages of lentiviral vectors over other systems are the ability to infect both dividing and non-dividing cells *in vivo* and *in vitro* and their greater packaging capacity that enables the
25 expression of larger RNA transcripts.

The lentivirus group can be split into “primate” and “non-primate”. Examples of primate lentiviruses include the human immunodeficiency virus (HIV), the causative agent of human acquired immunodeficiency syndrome (AIDS), and the simian immunodeficiency virus (SIV). The non-primate lentiviral group includes the prototype “slow virus” visna/maedi virus (VMV),
30 as well as the related caprine arthritis-encephalitis virus (CAEV), equine infectious anaemia virus (EIAV) and the more recently described feline immunodeficiency virus (FIV) and bovine immunodeficiency virus (BIV).

Details on the genomic structure of some lentiviruses may be found in the art. By way of example, details on HIV and EIAV may be found from the NCBI Genbank database (i.e. Genome Accession Nos. AF033819 and AF033820 respectively). Details of HIV variants may also be found at <http://hiv-web.lanl.gov>. Details of EIAV variants may be found through
5 <http://www.ncbi.nlm.nih.gov>.

Lentiviruses have three main genes coding for the viral proteins in the order: 5'-gag-pol-env-3'. There are two regulatory genes, *tat* and *rev*. There are additional accessory genes depending on the virus (e.g., for HIV-1: *vif*, *vpr*, *vpu*, *nef*) whose products are involved in regulation of synthesis and processing viral RNA and other replicative functions. The Long
10 terminal repeat (LTR) is about 600nt long, of which the U3 region is 450, the R sequence 100 and the U5 region some 70 nt long.

Viral proteins involved in early stages of replication include Reverse Transcriptase and Integrase. Reverse Transcriptase is the virally encoded RNA-dependent DNA polymerase. The enzyme uses the viral RNA genome as a template for the synthesis of a complementary
15 DNA copy. Reverse transcriptase also has RNaseH activity for destruction of the RNA-template. Integrase binds both the viral cDNA generated by reverse transcriptase and the host DNA. Integrase processes the LTR before inserting the viral genome into the host DNA. Tat acts as a trans-activator during transcription to enhance initiation and elongation. The Rev responsive element acts post-transcriptionally, regulating mRNA splicing and transport
20 to the cytoplasm.

RETROVIRAL PROTEIN

The retroviral protein may be Gag, Pol, Env or Rev.

25 Group-specific antigen (*gag*) proteins are major components of the viral capsid, which are about 2000–4000 copies per virion. Pol proteins are responsible for synthesis of viral DNA and integration into host DNA after infection. Env proteins play a role in association and entry of virions into the host cell. A functional copy of an *env* gene is what makes retroviruses distinct from retroelements. The ability of the retrovirus to bind to its target host
30 cell using specific cell-surface receptors is given by the surface component (SU) of the Env protein, while the ability of the retrovirus to enter the cell via membrane fusion is imparted by the membrane-anchored trans-membrane component (TM). Thus the Env protein is what enables the retrovirus to be infectious.

The retroviral protein may be gagpol. The nucleic acid sequence which encodes gagpol may comprise the nucleic acid sequence shown as SEQ ID NO: 9. The nucleic acid sequence which encodes gagpol may comprise a variant of SEQ ID NO: 9 which shares at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% or at least 99% sequence identity to SEQ ID NO: 9 and encodes functional Gag and Pol proteins.

SEQ ID NO: 9 - *Gag.Pol* – DNA sequence

ATGGGAGCACGGGCTAGTGTCTTTCTGGAGGTGAGCTTGACAGGTGGGAGAAAGATCAGACTGCGCCCCGGCGGCAAAAAGAAGTACAAGCTGA
AGCACATCGTGTGGGCTCTCGCGAATTGGAGAGTGTTCGCCGTGAACCCCGGGCTCTGGAGACAAGCGAGGGGTGCCGGCAGATCTCGGCCA
ATTTCAGCCAGTTTGCAAACCGGCAGCGAGGAGTTGCGGAGCCTGTACAACACCGTGGCCACATTTACTCGCTCCACCAGCGCATCGAAATC
AAGGATACAAAAGAGGCCCTGGATAAAATCGAAGAGGAACAGAATAAGAGCAAAAAGAAGGCCCAACAAGCCGCGCTGATACCGGCCATTTCTA
ACCAAGTCTCTCAGAACTATCCCATCGTCCAAAAATTTCAAGGCCAGATGGTCCACCAAGCTATCAGCCCCGGACCCTGAACGCCCTGGGTGAA
GGTGGTGGAGGAAAAAGCCTTTTCTCCCGAGGTCAATCCCTATGTTTACGCGCCCTGAGCGAGGGCGCTACACCCAGGACCTGAATACAATGTTG
AATACCGTCCGGCGCCACAGGCCGCTATCGAGATGCTGAAGGAAACAATTAACGAAGAGCCCGCGAGTGGGACCGGGTCCACCCCGTCCAGG
CTGGCCCCATCGCACCCGGGCAATGCGGGAGCCGAGAGGCTCCGATATCGCCGGCACCACCTCCACATTGCAAGAGCAGATCGGCTGGATGAC
CAACAATCCCCCAATTCAGCGTGGCGAGATCTACAAGCGGTGGATCATTCTCGGCCGTAACAAGATCGTCCGGATGACTCTCCACATCTATC
CTCGATATCCGGCAGGGCCCCAAAGAGCCTTTCCGGGATTACGTGGATAGATTTTACAAGACCTTGGCGGCTGAACAGGCCAGCCAAGAAGTGA
AGAAGTGGATGACCGAGACTCTCGTGCAGAACGCCAATCCCGACTGCAAAACCATCTGAAGGCCTTGGGCCAGCCGCCACCTTGGAGGA
GATGATGACCGCTGCCAAGCGTGGGAGGCCCTGGGCACAAAGCCCGGTGCTCGCCGAGGCCATGTTCTCAGGTGACCAACAGCGCCACAATC
ATGATGCAACCGGGGAACTTCCGCAATCGCGCAAAATCGTGAATGCTTTAAGTCTGCGCAAAAGAGGGCACACAGCCCGCAACTCGGGCCC
CTAGGAAAAAGGGCTGTTGGAAATGTGAAAGGAAGGACACCAAAATGAAAGATTGACTGAGAGACAGGCTAATTTTTTAGGGAAGATCTGGCC
TTCCACAAGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAAAGCCACCAGAAAGAGAGCTTCAGGTTTGGGGAAGAGACAACA
ACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAAGTATCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCAACAATAA
GATAGGGGGGCGAGCTCAAGGAGGCTCTCCCTGGCCACCGGAGCAGACACCGCTGCTGGAGGAGATGTCACTCCCTGGTGGGAGGATGCACTG
ATGATTGGTGGTATAGGGGGCTTCATTAAGGTGCGGCAATACGACCAATCTTGATCGAGATTTGCGGCCACAAGGCCATCGGCACCGTGTGG
TGGGCCCCACCCCGTGAATATCATCGGCCGGAACCTCCTCACCCAAATCGGCTGTACCCTGAACCTCCCTATCTCTCCCATCGAAACCGTGCC
CGTGAAGCTGAAACCCGGCATGGACGGGCCCCAAGGTGAAGCAGTGGCCCTCACCGAGGAGAAGATCAAGGCCCTGGTGGAGATCTGCACCCGAA
ATGGAGAAAGAGGGCAAGATCAGCAAGATCGGCCCGGAGAACCCCTATAACACCCCGTGTTCGCTATCAAAAAGAGGATTCACACCAAGTGGC
GGAAGCTGGTGGACTTTCCGGAGTTGAACAACCGACCCAGGATTTTTGGGAGGTGCAGCTGGGCATCCCCACCCCTGCCGGCTGAAGAAAAA
GAAGAGCGTGACCGTGTGACGCTCGGCCAGCCCTACTTCAGCGTGCCTCTGGACGAGGATTTTCGCAAAATACACCGCCTTCACAATCCCTCC
ATCAATAACGAAACCCCGGCATCCGGTACCAATATAACGCTTTGCCCAAGGCTGGAAGGGCAGCCCGCCATCTTTCAGTCTCTATGACCA
AGATTTCTGGAACCCCTTCGGAGGAGCCGAATTTGGAGCTCGCCGAGAAATCGGGAATCTGAAAGTATATGACACCCCGTGTGGAGGAGGATC
ACACCGGACCAAGATCGAAGAATCCGGCAGCACCTCCTCCGCTGGGCTTGACAACCCCGGATAAGAAGCACCAAAAAGGAGCCTCCCTTTTTG
TGGATGGGCTACGAGTTGCACCCGACAAAGTGGACCGTGAACCCATCGTCTCCCGGAGAGGATTCCTGGACCGTGAACGATATCAAAAAC
TGGTCCGCAAGCTCAACTGGGCTCCCAATCTATCCCGGCATCAAGGTGCGCCAGCTGTGCAAGTTGTTGGGGGCACAAGGCCGTTGACCGA
GGTATCCCTTGACCGAGGAGCCGAATTTGGAGCTCGCCGAGAAATCGGGAATCTGAAAGTATATGACACCCCGTGTGGAGGAGGATCAGGATC
AAGGATCTGATCGCCGAGATCAAAAACAAGGCCAGGGGAGTGGACCTACCAGATCTACAGGAACCCCTCAAGAACCTCAAGACCGGCAAGT
ACGCCCGGATGAGAGGCGCCATACCAACGACGTGAAGCAGCTGACCGAAGCCGTCAGAGATCAACCGAGTCTATCGTGTATCGGGGAA
AACCCCAAGTTCAAGCTCCCTATCCAGAAGGAACGTGGGAAACCTGGTGGACCGAATACTGGCAGGCTACATGGATCCCGAATGGGAGTTC
GTGAACACACCCCTCTGGTCAAGCTGTGGTATCAACTCGTCAAGGATTTGCAAGTGTGTTTCTGGGCGCCGAGACATTTACGTGGAGCGGCTG
GCAAAACCAAGCTGGGCAAGGCCGCTACGTGACCAATCGGGGCGCCAGAAAGTGGTGCATTTGACCGATACCAACAACAGAAAACCGAACT
GCAGGCCATCTACTTGGCCCTCCAAGACAGCGGCTGGAGGTGAATATCGTGACCGATAGCCAGTACGCCCTGGGCATTATCCAGGCCAGCCC
GACCAGTCCGAGAGCGAAGCTGGTGAACAGATCAAGCAACTGATCAAGAAAGAGAAAGTGTACTCGCTGGGTGCCCGCCATAAGGGGA
TCGGCGGCAACGAGCAGGTGGACAAGCTGTGCTCGCCGATTTGCAAGTGTGTTTCTGGACGGCATCGACAAGCTCAGGACGAGCAGCAGCA
AAAGTACCATTCCAATCGGCGGCTGAGTCCAGGCTGAGTGGAGAGCATGAACAAGAGGCTGAAAAGATATCGGCCAAGTGGCGGCAAGTGGC
TTGAAAGGGCAGGCTATGCACGGCCAGGTGGATTGCTCCCGGCATCTGGCAGTTGACTGCACCCACCTGGAGGGCAAGGTGATTCTCGTGG
CCGTGCACGTTGGCTTCCGGCTACATCGAGGCTGAGGTGATCCCGGCCAAACCGGCCAAGAGACTGCCTACTTCTGCTGAAGCTGGCCGGCAG
GTGGCCGCTAAAGACCATCCACCGGATAACGGGTCTAATTTACATCCGCCACCGTGAAGCTGCTTGTGGTGGGCGAGGATTAACAAGAG
TTCCGCATCCCTTATAACCTCAGTCCAGGGCGTGGTGGAGAGCATGAACAAGAGGCTGAAAAGATATCGGCCAAGTGGCGGCAAGTGGC
AGCACTTGAAAACCGCCGTGCAGATGGCCGTGTTTATTCAACTCAAGCGGAAGGGCGGCATCGGCGGCTATTCCCGGCTGAGCGGATCGT
GGATATCATCGCCACCGATATCCAGACCAAGGAGCTGCAGAAGCAGATACCAAGATCCAGAACTTCAGAGTGTACTATCCGGATTCGGGAAC
CCCTGTGGAAGGGCCAGCCAAATTTGTTGGAAGGGGAGGGCGCCGTTGGTATCCAGGACAACCTCCGATATCAAGTGGTCCCGCGGAGGA
AGGCCAAAATATCCGCGACTACGGCAAGCAAAATGCGCGGACGACTGCTCGCTCCCGCAAGACGAGGACTGA

The retroviral env protein may be the RD114 SU or TM protein or the RDPRO SU or TM protein. In one embodiment, the env protein is VSV-G (glycoprotein G of the Vesicular stomatitis virus (VSV)). VSV-G is a commonly used env protein for lentiviral pseudotyping as it is capable of transducing all cell types.

The retroviral env protein may comprise the sequence shown as SEQ ID NO: 10. The retroviral env protein may comprise a variant of SEQ ID NO: 10 which retains ability to provide a functional env protein. The variant may share at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% or at least 99% sequence identity to SEQ ID NO: 10.

SEQ ID NO: 10 - VSVG env amino acid sequence

MKCLLYLAFLFIGVNCFTIVFPHNQKGNWKNVPSNYHYCPSSSDLNWHNDLIGTALQVKMPKSHKAI
 QADGWMCHASKWVTTCDFRWYGPKYITHSIRSFTPSVEQCKESIEQTKQGTWLNPGFPPQSCGYATVT
 10 DAEAVIVQVTPHHVLVDEYTGGEWVDSQFINGKCSNYICPTVHNSTTWHSYKVKGLCDSNLI SMDITF
 FSEDGELSSLGKEGTGFRSNYFAYETGGKACKMQYCKHWGVRLPSGVWFEMADKDLFAAARFPECPEG
 SSISAPSQTSVDVSLIQDVERILDYSLCQETWSKIRAGLPISPVDLSYLAPKNPGTGPAFTIINGTLK
 YFETRYIRVDIAAPILSRMVGMI SGT TTERELWDDWAPYEDVEIGPNGVLR TSSGYKFPLYMIGHGML
 15 DSDLHLSSKAQVFEHPHIQDAASQLPDDESLFFGDTGLSKNPIELVEGW FSSWKSSIASFFFII GLII
 GLFLVLRVGIHLCLIKLKH TKKRQIYTDIEMNRLGK

Suitably, 'Retroviral envelope protein' refers to the SU and/or TM proteins, as described above in relation to lentiviruses.

In one embodiment the packaging or producer cell is used to produce a lentivirus vector and retroviral protein may be Rev. Rev is specifically expressed by lentivirus and acts post-transcriptionally, regulating mRNA splicing and transport to the cytoplasm.

DETECTABLE MARKER

The detectable marker is a cell surface protein. The cell surface protein may be any cell surface protein which is not natively expressed on the surface of the cell type which is used as the packaging cell or producer cell.

The cell surface protein comprises a membrane targeting domain and an extracellular domain. When expressed at the cell surface at least one domain of the cell surface protein is exoplasmic (i.e. on the exterior of the cell). This domain of the cell surface protein is therefore accessible for antibody binding.

In one embodiment, the detectable marker is a cell surface protein where the expression level can be determined at the single cell level. In particular, the detectable marker may be detected without disrupting the cell. Because the expression level of the detectable marker can be determined at the single cell level, the co-expression of the detectable marker with a

retroviral protein means that the expression level of the retroviral protein within a packaging cell can also be determined.

This is in contrast to the selectable antibiotic resistance markers previously used to identify packaging cells expressing a retroviral protein. In particular, although the use of selectable antibiotic resistance markers enables cells which express the resistance marker - and therefore the retroviral protein - to be selected, it is difficult to determine the relative expression obtained from each individual integration event and thus the level of expression of each retroviral protein in a particular cell.

Accordingly, the detectable marker of the present invention is not an antibiotic resistance marker.

For example, the expression level of the detectable marker may be determined using flow cytometry. Methods for using techniques such as flow cytometry to determine the expression levels of proteins are well known in the art. As such, the detectable marker may be a cell surface protein which can be detected using flow cytometry by using a reagent(s) which allows expression of the cell surface protein to be determined. For example the reagent may be an antibody, for example a labelled antibody such as a fluorescently labelled antibody which specifically binds the cell surface protein. The detectable marker may be a fluorescent protein which is inherently detectable by flow cytometry due to its fluorescent characteristics.

In one embodiment the cell surface protein is less than 200 amino acids in length. The use of a cell surface protein of less than 200 amino acids is advantageous because the smaller nucleic acid sequences which encode such proteins are easier to transfect and stably integrate into cells.

Membrane targeting domain

A 'membrane targeting domain' is an entity which preferentially localises to the membrane and therefore anchors the cell surface protein to the membrane of, for example, a packaging cell.

A transmembrane domain is a hydrophobic alpha helix which spans a cell membrane and typically found in transmembrane proteins. It may comprise a hydrophobic alpha helix. The transmembrane domain may be derived from CD8, CD28 or human IgG.

The transmembrane domain may be derived from any type I transmembrane protein. The transmembrane domain may be a synthetic sequence predicted to form a hydrophobic helix.

- 5 The transmembrane domain may comprise the sequence shown as SEQ ID NO: 12.

SEQ ID NO: 12 (CD28 transmembrane domain)

FWWLVVGGVLACYSLLVTVAFIIFWW

- 10 The membrane targeting domain may be a transmembrane domain and an endodomain, which orientates to the interior of the packaging cell. The endodomain may comprise polar residues which anchor the cell surface protein to the membrane.

The membrane targeting domain may be a GPI anchor.

- 15 GPI anchoring is a post-translational modification which occurs in the endoplasmic reticulum. Preassembled GPI anchor precursors are transferred to proteins bearing a C-terminal GPI signal sequence (see Kinoshita et al.; J Biochem; 122, 251-257 (1997)). During processing, the GPI anchor replaces the GPI signal sequence and is linked to the target protein via an amide bond. The GPI anchor targets the mature protein to the membrane.

20

The present tagging protein may comprise a GPI signal sequence. For example, the tagging protein may comprise a sequence shown as SEQ ID NO: 13.

SEQ ID NO : 13

- 25 ASSNISGGIFLFFVANAIHLFCFS

Extracellular domain

- 30 The extracellular domain is the part of the cell surface protein that exoplasmic (i.e. on the exterior of the cell) when the cell surface protein is expressed at the cell surface.

The extracellular domain may be any polypeptide sequence which comprises an epitope that can be specifically bound by an antibody.

- 35 In one embodiment the extracellular domain is less than 200 amino acids in length.

The extracellular domain may comprise a HA, V5, RQR-8 or MYC tag.

Human influenza hemagglutinin (HA) is a surface glycoprotein required for the infectivity of the human virus. The HA tag is derived from the HA-molecule corresponding to amino acids 98-106. It has been extensively used as a general epitope tag in expression vectors. The HA tag may comprise the sequence shown as SEQ ID NO: 14.

YPYDVPDYA (SEQ ID NO: 14)

10 The V5 epitope tag (V5) is derived from a small epitope (Pk) present on the P and V proteins of the paramyxovirus of simian virus 5 (SV5). The V5 tag may comprise the sequence shown as SEQ ID NO: 15 or 16.

GKPIPPELLGLDST (SEQ ID NO: 15)

15 IPNPELLGLD (SEQ ID NO: 16)

A myc tag (MYC) is a polypeptide protein tag derived from the c-myc gene product. The MYC tag may comprise the sequence shown as SEQ ID NO: 17.

20 EQKLISEEDL (SEQ ID NO: 17)

The extracellular domain may comprise a domain which comprises a rituximab-binding epitope (R epitope) and/or a Qbend10 epitope (Q epitope). A rituximab-binding epitope refers to an epitope which specifically binds rituximab. For example, the rituximab-binding epitope may be based on the CD20 B-cell antigen.

The Rituximab-binding epitope sequence from CD20 is CEPANPSEKNSPSTQYC (SEQ ID NO: 18)

30 The CliniMACS CD34 selection system utilises the QBEnd10 monoclonal antibody to achieve cellular selection. The present inventors have previously mapped the QBEnd10-binding epitope from within the CD34 antigen (see WO 2013/153391) and determined it to have the amino acid sequence shown as SEQ ID NO: 19.

35 ELPTQGTFSNVSTNVS (SEQ ID NO: 19).

The binding domain of the cell surface protein may comprise a QBEnd10-binding epitope having the amino acid sequence shown as SEQ ID NO: 19 or a variant thereof which retains QBEnd10-binding activity.

- 5 The extracellular domain may comprise a binding domain which comprises or consists of 136 amino acid sequence shown as SEQ ID NO: 20.

SEQ ID NO: 20 (RQR8)

CPYSNP~~SLC~~SGGGGSELPTQGTFSNVSTNVSPAKPTTTACPY~~SNP~~SLC~~SGGGG~~SPAPRPPTPAPTIAS
 10 QPLSLRPEACRPAAGGAVHTRGLDFACDIYIWAPLAGTCGVLLLSLVI TLYCNHRNRRRVCKCPRPVV

The cell surface marker protein may comprise or consist of a sequence shown as SEQ ID NO: 21 to 26.

15 SEQ ID NO: 21 (V5-8)

<-----signal-----><----V5-----><link><-----CD8----->
 MGTSLLCW~~MAL~~C~~LL~~GADHADAGKPIP~~N~~PL~~L~~GLDSSGGGGSPAPRPPTPAPTIASQPLSLRPEACRPA
 -----CD8----->
 20 GGA~~VHTRGLDFACDIYIWAPLAGTCGVLLLSLVI~~TLYCNHRNRRRVCKCPRPVV

SEQ ID NO: 22 (HA-8)

<-----signal-----><--HA--><link><-----CD8----->
 25 MGTSLLCW~~MAL~~C~~LL~~GADHADAYPYDVPDYASGGGGSPAPRPPTPAPTIASQPLSLRPEACRPAAGGAV
 -----CD8----->
 HTRGLDFACDIYIWAPLAGTCGVLLLSLVI TLYCNHRNRRRVCKCPRPVV

30 SEQ ID NO: 23 (MYC-8)

<-----signal-----><--MYC--><link><-----CD8----->
 MGTSLLCW~~MAL~~C~~LL~~GADHADAEQKLI SEEDLSGGGGSPAPRPPTPAPTIASQPLSLRPEACRPAAGGA
 35 -----CD8----->
 VHTRGLDFACDIYIWAPLAGTCGVLLLSLVI TLYCNHRNRRRVCKCPRPVV

SEQ ID NO: 24 (V5-GPI)

40 <-----signal-----><----V5-----><-----GPI----->
 MGTSLLCW~~MAL~~C~~LL~~GADHADAGKPIP~~N~~PL~~L~~GLDSASSNISGGIFLFFVANAI IHLFCFS

SEQ ID NO: 25 (HA-GPI)

<-----signal-----><--HA---><-----GPI----->
 MGTSLLCWMLCCLLGADHADAYPYDVPDYAASSNISGGIFLFFVANAI IHLFCFS

5 SEQ ID NO: 26 (MYC-GPI)

<-----signal-----><--MYC---><-----GPI----->
 MGTSLLCWMLCCLLGADHADAEQKLI SEEDLASSNISGGIFLFFVANAI IHLFCFS

10 The cell surface protein may comprise a variant of an extracellular domain or a cell surface
 marker protein as described herein. For example, the extracellular domain may be based on
 a sequence provided herein but comprises one or more amino acid mutations, such as
 amino acid insertions, substitutions or deletions, provided that the epitope retains binding
 activity of the entity or epitope on which it is based. In particular, the sequence may be
 15 truncated at one or both terminal ends by, for example, one or two amino acids.

Deliberate amino acid substitutions may be made on the basis of similarity in polarity,
 charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the
 residues as long as binding activity of the epitope is retained. For example, negatively
 charged amino acids include aspartic acid and glutamic acid; positively charged amino acids
 20 include lysine and arginine; and amino acids with uncharged polar head groups having
 similar hydrophilicity values include leucine, isoleucine, valine, glycine, alanine, asparagine,
 glutamine, serine, threonine, phenylalanine, and tyrosine.

Conservative substitutions may be made, for example according to the Table below. Amino
 acids in the same block in the second column and in the same line in the third column may
 25 be substituted for each other.

ALIPHATIC	Non-polar	G A P
		I L V
	Polar – uncharged	C S T M
		N Q
	Polar – charged	D E
		K R
AROMATIC		H F W Y

The binding domain may, for example, contain 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer
 or 1 amino acid mutation(s) compared to a binding domain sequence as shown herein.

The extracellular domain may consist essentially of a domain sequence as shown herein or a variant thereof which retains the ability to bind an antibody which binds to the domain on which it is based. The extracellular domain may consist of a sequence as shown herein or a variant thereof which retains the ability to bind an antibody which binds to the domain on which it is based.

The cell surface protein may comprise the sequence shown as SEQ ID NO: 14 to 26 or a variant thereof having 80% identity thereto which retains the ability to be expressed at the cell surface and recognised by reagent, for example, an antibody which binds to one of SEQ ID NO: 14 to 26.

The cell surface protein may comprise the sequence shown as SEQ ID NO: 14 to 26 or a variant thereof having 85% identity thereto which retains the ability to be expressed at the cell surface and recognised by an antibody or a ligand which binds to one of SEQ ID NO: 14 to 26

The cell surface protein may comprise the sequence shown as SEQ ID NO: 14 to 26 or a variant thereof having 90% identity thereto which retains the ability to be expressed at the cell surface and recognised by an antibody or a ligand which binds to one of SEQ ID NO: 14 to 26

The cell surface protein may comprise the sequence shown as SEQ ID NO: 14 to 26 or a variant thereof having 95% identity thereto which retains the ability to be expressed at the cell surface and recognised by an antibody or a ligand which binds to one of SEQ ID NO: 14 to 26

The cell surface protein may comprise the sequence shown as SEQ ID NO: 14 to 26 or a variant thereof having 99% identity thereto which retains the ability to be expressed at the cell surface and recognised by an antibody or a ligand which binds to one of SEQ ID NO: 14 to 26

Signal sequence

The cell surface protein may also comprise a signal sequence so that when the protein is expressed inside a cell the nascent protein is directed to the endoplasmic reticulum (ER).

The core of the signal peptide may contain a long stretch of hydrophobic amino acids that has a tendency to form a single alpha-helix. The signal peptide may begin with a short positively charged stretch of amino acids, which helps to enforce proper topology of the polypeptide during translocation. At the end of the signal peptide there is typically a stretch
 5 of amino acids that is recognized and cleaved by signal peptidase. Signal peptidase may cleave either during or after completion of translocation to generate a free signal peptide and a mature protein. The free signal peptides are then digested by specific proteases.

The signal peptide may be at the amino terminus of the molecule.

10

The signal peptide may comprise the SEQ ID NO: 1 or 2 or a variant thereof having 5, 4, 3, 2 or 1 amino acid mutations (insertions, substitutions or additions) provided that the signal peptide still functions to cause secretion of the bi-specific molecule.

15 METDTLLLWVLLLWPGSTG (SEQ ID NO: 1)

MGTSLLCWMALCLLGADHADG (SEQ ID NO: 2)

The signal peptides of SEQ ID NO: 1 and 2 are compact and highly efficient. They are predicted to give about 95% cleavage after the terminal glycine, giving efficient removal by
 20 signal peptidase.

The cell surface marker protein as described herein may lack a signal sequence. For example, the cell surface marker shown as SEQ ID NO: 21-26 may not include the sequence annotated of the signal sequence. In particular, the cell surface marker protein
 25 will not include the signal sequence when the protein is present at the cell surface.

CO-EXPRESSION SEQUENCE

In one embodiment the retroviral protein and the detectable marker are transcribed as a
 30 nucleic acid which comprises a nucleic acid sequence which encodes the retroviral transfer vector or retroviral protein, a co-expression sequence and a nucleic acid sequence which encodes the detectable marker, wherein the co-expression sequence enables the retroviral transfer vector or retroviral protein and the detectable marker to be expressed as separate polypeptides.

35

The co-expression sequence may be a 2A self-cleaving peptide from an aphtho- or a cardiovirus, a 2A-like peptide, an internal ribosome entry site, a self-cleaving peptide, a furin cleavage site or a Tobacco Etch Virus cleavage site.

- 5 An internal ribosome entry site (IRES) is a nucleotide sequence that allows for translation initiation in the middle of a messenger RNA (mRNA).

The co-expression sequence may be a sequence encoding a cleavage site positioned between nucleic acid sequences which encode the retroviral protein and detectable marker,
10 such that the retroviral protein and detectable marker can be expressed as separate entities.

The cleavage site may be any sequence which enables the polypeptide comprising the retroviral protein and detectable marker to become separated.

- 15 The term "cleavage" is used herein for convenience, but the cleavage site may cause the retroviral protein and detectable marker to separate into individual entities by a mechanism other than classical cleavage. For example, for the Foot-and-Mouth disease virus (FMDV) 2A self-cleaving peptide (see below), various models have been proposed for to account for the "cleavage" activity: proteolysis by a host-cell proteinase, autoproteolysis or a
20 translational effect (Donnelly et al (2001) J. Gen. Virol. 82:1027-1041). The exact mechanism of such "cleavage" is not important for the purposes of the present invention, as long as the cleavage site, when positioned between nucleic acid sequences which encode retroviral protein and detectable marker, causes the retroviral protein and detectable marker to be expressed as separate entities.

25

The cleavage site may be a furin cleavage site.

- Furin is an enzyme which belongs to the subtilisin-like proprotein convertase family. The members of this family are proprotein convertases that process latent precursor proteins into
30 their biologically active products. Furin is a calcium-dependent serine endoprotease that can efficiently cleave precursor proteins at their paired basic amino acid processing sites. Examples of furin substrates include parathyroid hormone, transforming growth factor beta 1 precursor, proalbumin, pro-beta-secretase, membrane type-1 matrix metalloproteinase, beta subunit of pro-nerve growth factor and von Willebrand factor. Furin
35 cleaves proteins just downstream of a basic amino acid target sequence (canonically, Arg-X-(Arg/Lys)-Arg') (SEQ ID NO: 3) and is enriched in the Golgi apparatus.

The cleavage site may be a Tobacco Etch Virus (TEV) cleavage site.

TEV protease is a highly sequence-specific cysteine protease which is chymotrypsin-like proteases. It is very specific for its target cleavage site and is therefore frequently used for the controlled cleavage of fusion proteins both *in vitro* and *in vivo*. The consensus TEV cleavage site is ENLYFQ\S (SEQ ID NO: 4) (where '\ ' denotes the cleaved peptide bond). Mammalian cells, such as human cells, do not express TEV protease. Thus in embodiments in which the present nucleic acid construct comprises a TEV cleavage site and is expressed in a mammalian cell – exogenous TEV protease must also be expressed in the mammalian cell.

10

The cleavage site may encode a self-cleaving peptide.

A 'self-cleaving peptide' refers to a peptide which functions such that when the polypeptide comprising the retroviral protein and detectable marker and the self-cleaving peptide is produced, it is immediately "cleaved" or separated into distinct and a discrete retroviral protein and a detectable marker without the need for any external cleavage activity.

15

The self-cleaving peptide may be a 2A self-cleaving peptide from an aphtho- or a cardiovirus. The primary 2A/2B cleavage of the aphtho- and cardioviruses is mediated by 2A "cleaving" at its own C-terminus. In aphthoviruses, such as foot-and-mouth disease viruses (FMDV) and equine rhinitis A virus, the 2A region is a short section of about 18 amino acids, which, together with the N-terminal residue of protein 2B (a conserved proline residue) represents an autonomous element capable of mediating "cleavage" at its own C-terminus.

20

The C-terminal 19 amino acids of the longer cardiovirus protein, together with the N-terminal proline of 2B mediate "cleavage" with an efficiency approximately equal to the aphthovirus FMDV 2a sequence. Cardioviruses include encephalomyocarditis virus (EMCV) and Theiler's murine encephalitis virus (TMEV).

25

Mutational analysis of EMCV and FMDV 2A has revealed that the motif DxExNPGP (SEQ ID NO: 5) is intimately involved in "cleavage" activity (Donnelly et al (2001) as above).

30

The cleavage site of the present invention may comprise the amino acid sequence:

Dx_1Ex_2NPGP , where x_1 and x_2 are any amino acid (SEQ ID NO: 5). Suitably, x_1 may be selected from the following group: I, V, M and S. x_2 may be selected from the following group: T, M, S, L, E, Q and F (as shown in SEQ ID NO: 8).

35

For example, the cleavage site may comprise one of the amino acid sequences shown in Table 1.

Table 1

5

Motif	Present in:
DIETNPGP (SEQ ID NO: 6)	Picornaviruses EMCB, EMCD, EMCPV21
DVETNPGP (SEQ ID NO: 7)	Picornaviruses MENGO and TMEBEAN; Insect virus DCV, ABPV
DVEMNPGP (SEQ ID NO: 27)	Picornaviruses TMEGD7 and TMEBEAN
DVESNPGP (SEQ ID NO: 28)	Picornaviruses FMDA10, FMDA12, FMDC1, FMD01K, FMDSAT3, FMDVSAT2, ERAV; Insect virus CrPV
DMESNPGP (SEQ ID NO: 29)	Picornavirus FMDV01G
DVELNPGP (SEQ ID NO: 30)	Picornavirus ERBV; Porcine rotavirus
DVEENPGP (SEQ ID NO: 31)	Picornavirus PTV-1; Insect virus TaV; Trypanosoma TSR1
DIELNPGP (SEQ ID NO: 32)	Bovine Rotavirus, human rotavirus
DIEQNPGP (SEQ ID NO: 33)	Trypanosoma AP endonuclease
DSEFNPGP (SEQ ID NO: 34)	Bacterial sequence <i>T. maritima</i>

The cleavage site, based on a 2A sequence may be, for example 15-22 amino acids in length. The sequence may comprise the C-terminus of a 2A protein, followed by a proline residue (which corresponds to the N-terminal proline of 2B).

10

Mutational studies have also shown that, in addition to the naturally occurring 2A sequences, some variants are also active. The cleavage site may correspond to a variant sequence from a naturally occurring 2A polypeptide, have one, two or three amino acid substitutions, which retains the capacity to induce the "cleavage" of a polyprotein sequence into two or more separate proteins.

15

The cleavage sequence may be selected from the following which have all been shown to be active to a certain extent (Donnelly *et al* (2001) as above):

20

LLNFDLLKLAGDVESNPGP (SEQ ID NO: 35)

LLNFDLLKLAGDVQSNPGP (SEQ ID NO: 36)
 LLNFDLLKLAGDVEINPGP (SEQ ID NO: 37)
 LLNFDLLKLAGDVEFNPGP (SEQ ID NO: 38)
 LLNFDLLKLAGDVESHPPGP (SEQ ID NO: 39)
 5 LLNFDLLKLAGDVESEPPGP (SEQ ID NO: 40)
 LLNFDLLKLAGDVESQPPGP (SEQ ID NO: 41)
 LLNFDLLKLAGDVESNPPGP (SEQ ID NO: 42)

Based on the sequence of the DxExNPGP “a motif, “2A-like” sequences have been found in
 10 picornaviruses other than aptho- or cardioviruses, ‘picornavirus-like’ insect viruses, type C
 rotaviruses and repeated sequences within *Trypanosoma spp* and a bacterial sequence
 (Donnelly et al (2001) as above). The cleavage site may comprise one of these 2A-like
 sequences, such as:

15 YHADYYKQRLIHDVEMNPGP (SEQ ID NO: 43)
 HYAGYFADLLIHDVETNPGP (SEQ ID NO: 44)
 QCTNYALLKLAGDVESNPGP (SEQ ID NO: 45)
 ATNFSLLKQAGDVEENPGP (SEQ ID NO: 46)
 AARQMLLLSGDVETNPGP (SEQ ID NO: 47)
 20 RAEGRGSLTTCGDVEENPGP (SEQ ID NO: 48)
 TRAEIEDELIRAGIESNPGP (SEQ ID NO: 49)
 TRAEIEDELIRADIESNPGP (SEQ ID NO: 50)
 AKFQIDKILISGDVELNPGP (SEQ ID NO: 51)
 SSIIRTKMLVSGDVEENPGP (SEQ ID NO: 52)
 25 CDAQRQKLLLSGDIEQNPGP (SEQ ID NO: 53)
 YPIDFGGFLVKADSEFNPGP (SEQ ID NO: 54)

The cleavage site may comprise the 2A-like sequence shown as SEQ ID NO: 55
 (RAEGRGSLTTCGDVEENPGP).

30

It has been shown that including an N-terminal “extension” of between 5 and 39 amino acids
 can increase activity (Donnelly et al (2001) as above). In particular, the cleavage sequence
 may comprise one of the following sequences or a variant thereof having, for example, up to
 5 amino acid changes which retains cleavage site activity:

35

VTELLYRMKRAETYCPRLAIHPTEARHKQKIVAPVKQTLNFDLLKLAGDVESNPGP (SEQ
 ID NO: 56)

LLAIHPTEARHKQKIVAPVKQTLNFDLLKLAGDVESNPGP (SEQ ID NO: 57)

EARHKQKIVAPVKQTLNFDLLKLAGDVESNPGP (SEQ ID NO: 58)

APVKQTLNFDLLKLAGDVESNPGP (SEQ ID NO: 59)

5 TRANSPOSON ELEMENT

Various transposon elements are well known in the art. Any suitable transposon may be used in the present invention.

10 DNA transposons are genetic elements that can mobilize from one location to another in the host genome. Their use in engineering mammalian cells was initially hampered by the lack of active transposons in mammals. Over a decade ago, the first active DNA transposon in mammals, Sleeping Beauty, was reconstructed from fossilized transposon sequences found in the salmonid genome. This discovery made transposon mediated genetic modification of mammalian cells possible. This method is much more efficient at generating cell lines with modified genomes than simple transfection.

Since the generation of the Sleeping Beauty transposon, a number of transposons from different families have been reported to show active transposition in mammalian cells (Skipper et al, J of Biomed Sci 2013, 20:92). Currently, along with Sleeping Beauty, the piggyBac (PB) transposon isolated from cabbage looper moth *Trichoplusia ni* is most promising because of a variety of unique characteristics, namely exhibiting the most efficient transposition in mammalian cells, the ability of the transposase to form functional protein fusions, large cargo capacity, and traceless excision. More importantly, PB has been found 25 capable of mediating stable integration of up to 4 independent transposons simultaneously in human cells following a single transfection. A more active form of piggyback has been described (Yusa et al, PNAS, 2011, vol. 108, no.4).

In one embodiment of the invention, the transposon element or sequence comprises (or 30 consists of) a piggyBac transposon. During transposition, the PB transposase recognizes transposon-specific inverted terminal repeat sequences (ITRs) located on both ends of the transposon vector and efficiently moves the contents from the original sites and integrates them into TTAA chromosomal sites.

35 Unique features of PiggyBac transposons are that there is relatively no cargo limit as it can carry a transgene of up to 14kbp and it is also reversible: genomes containing an inserted

PiggyBac vector can be transiently re-transfected with the PB transposase expression vector and the transposase will remove the transposons from the genome, footprint-free.

SEQ ID NO: 60 – PiggyBAC 5' Terminal Repeat

5 CATTCTAGATTAACCCTAGAAAGATAGTCTGCGTAAAATTGACGCATGCATTCTTGAAATATTGCTCT
 CTCTTTCTAAATAGCGCGAATCCGTCGCTGTGCATTTAGGACATCTCAGTCGCCGCTTGGAGCTCCCG
 TGAGGCGTGCTTGTCAATGCGGTAAGTGCTCACTGATTTTGAACATAACGACCGCGTGAGTCAAAATG
 ACGCATGATTATCTTTTACGTGACTTTTAAAGATTTAACTCATACGATAATTATATTGTTATTTTCATGT
 10 TCTACTTACGTGATAACTTATTATATATATATTTTCTTGTATAGATATCAACTAGAATGCTAGCATG
 GGCCAT

SEQ ID NO: 61 – PiggyBAC 3' Terminal Repeat

AGTTTTGTTACTTTATAGAAGAAATTTTGTGTTTTTTTTTAAATAAATAAATAAACATAAATA
 AATTGTTTGTGAATTTATTATTAGTATGTAAGTGTAATATAATAAACTTAATATCTATTCAAAT
 15 AATAAATAAACCTCGATATACAGACCGATAAAACACATGCGTCAATTTTACGCATGATTATCTTTAAC
 GTACGTCACAATATGATTATCTTTCTAGGGTTAATCTAG

In one embodiment, the nucleic acid construct may comprise a PiggyBAC 5' terminal repeat and a PiggyBAC 3' terminal repeat such that the construct can insert into the host genome.

20 The PiggyBAC 5' and 3' terminal repeats may flank the sequence which is to be inserted into the host genome.

The PiggyBAC 5' Terminal Repeat may comprise or consist of a variant which shares at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% sequence identity with
 25 SEQ ID NO: 60 and retains the ability to recruit an RdRp. The PiggyBAC 5' Terminal Repeat may comprise or consist of SEQ ID NO: 60.

The PiggyBAC 5' Terminal Repeat may comprise or consist of a truncated version of SEQ ID NO: 60 or truncated version of a variant of SEQ ID NO: 60 which retains ability to facilitate
 30 insertion into the host genome.

The PiggyBAC 3' Terminal Repeat may comprise or consist of a variant which shares at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% sequence identity with
 35 SEQ ID NO: 61 and retains the ability to recruit an RdRp. The PiggyBAC 3' Terminal Repeat may comprise or consist of SEQ ID NO: 61.

The PiggyBAC 3' Terminal Repeat may comprise or consist of a truncated version of SEQ ID NO: 61 or truncated version of a variant of SEQ ID NO: 61 which retains ability to facilitate
 40 insertion into the host genome.

PLASMID

In one aspect the present invention provides a plasmid comprising a nucleic acid construct of the present invention.

5

The term plasmid covers any DNA transcription unit comprising a nucleic acid construct according to the invention and the elements necessary for its *in vivo* expression in a desired cell; and, in this regard, it is noted that a supercoiled or non-supercoiled, circular plasmid, as well as a linear form, are intended to be within the scope of the invention.

10

A plasmid may comprise the nucleic acid construct according to the invention, operably linked to a promoter or under the control of a promoter or dependent upon a promoter. In general, it is advantageous to employ a strong promoter functional in eukaryotic cells. The strong promoter may be, but not limited to, the immediate early cytomegalovirus promoter (CMV-IE) of human or murine origin, or optionally having another origin such as the rat or guinea pig.

15

In more general terms, the promoter has either a viral, or a cellular origin. A strong viral promoter other than CMV-IE that may be usefully employed in the practice of the invention is the early/late promoter of the SV40 virus or the LTR promoter of the Rous sarcoma virus. A strong cellular promoter that may be usefully employed in the practice of the invention is the promoter of a gene of the cytoskeleton, such as e.g. the desmin promoter (Kwissa et al., 2000), or the actin promoter (Miyazaki et al., 1989).

20

As to the polyadenylation signal (polyA) for the plasmids use can be made of the poly(A) signal of the bovine growth hormone (bGH) gene (see U.S. 5,122,458), or the poly(A) signal of the rabbit β -globin gene or the poly(A) signal of the SV40 virus.

25

KIT

30

The present invention further provides a kit comprising a plurality of nucleic acid constructs of the invention.

In one embodiment the kit may comprise:

35

(i) a first nucleic acid construct comprising a *gag-pol* sequence and a sequence encoding a first detectable marker;

(ii) a second nucleic acid construct comprising an *env* sequence and a sequence encoding a second detectable marker;

in which the first and second detectable markers are cell surface proteins comprising an extracellular domain and a membrane targeting domain and are different to each other.

5

The kit may further comprise:

(iii) a third nucleic acid construct comprising a *rev* sequence and a sequence encoding a third detectable marker which is a cell surface protein comprising an extracellular domain and a membrane targeting domain;

10 in which the first, second and third and detectable markers are all different to each other.

The kit may also comprise:

(iv) a fourth nucleic acid construct comprising a retroviral transfer vector and a sequence encoding a fourth detectable marker which is a cell surface protein comprising an extracellular domain and a membrane targeting domain;

15

in which the first, second, third (if present) and fourth detectable markers are all different to each other.

In another aspect the present invention provides a kit comprising a plurality of plasmids according to the invention.

20

PACKAGING CELL AND PRODUCER CELL

As used herein, the term “packaging cell” refers to a cell which contains those elements necessary for production of infectious recombinant virus which are lacking in the RNA genome. Such packaging cells are capable of expressing viral structural proteins (such as gag-pol and env, which may be codon optimised) but they do not contain a packaging signal. The term “packaging signal” which is referred to interchangeably as “packaging sequence” or “psi” is used in reference to the non-coding, cis-acting sequence required for encapsidation of retroviral RNA strands during viral particle formation. In HIV-1, this sequence has been mapped to loci extending from upstream of the major splice donor site (SD) to at least the gag start codon.

25

30

As used herein, the term “producer cell” refers to a cell which contains all the elements necessary for production of retroviral vector particles.

35

The producer cells/packaging cells of the present invention may be any suitable cell type. Producer cells are generally mammalian cells but can be, for example, insect cells.

5 By using producer/packaging cell lines, it is possible to propagate and isolate quantities of retroviral vector particles (e.g. to prepare suitable titres of the retroviral vector particles) for subsequent transduction of a site of interest.

10 The packaging cell lines are useful for providing the gene products necessary to encapsidate and provide a membrane protein for a high titre vector particle production. The packaging cell may be a cell cultured *in vitro* such as a tissue culture cell line. Suitable cell lines include but are not limited to mammalian cells such as murine fibroblast derived cell lines or human cell lines. The packaging cell line may be a human cell line, such as for example: HEK293, 293-T, TE671, HT1080.

15 There are two common procedures for generating producer cells. In one, the sequences encoding retroviral Gag, Pol and Env proteins are introduced into the cell and stably integrated into the cell genome; a stable cell line is produced which is referred to as the packaging cell line. As used herein, the term "stably integrated" means that the foreign genes become integrated into the cell's genome. The packaging cell line produces the
20 proteins required for packaging retroviral RNA but it cannot bring about encapsidation due to the lack of a psi region.

25 However, when a vector genome (having a psi region) is introduced into the packaging cell line, the helper proteins can package the psi-positive recombinant vector RNA to produce the recombinant virus stock. This can be used to transduce recipient cells. The recombinant virus whose genome lacks all genes required to make viral proteins can infect only once and cannot propagate. Hence, a nucleic acid sequence can be introduced into a host cell genome without the generation of potentially harmful retrovirus.

30 The second approach is to introduce the three different DNA sequences that are required to produce a retroviral vector particle (i.e. the env coding sequences, the gag-pol coding sequence and the defective retroviral genome containing one or more NOIs) into the cell at the same time by transient transfection and the procedure is referred to as transient triple
35 transfection. WO 94/29438 describes the production of producer cells *in vitro* using this multiple DNA transient transfection method. WO 97/27310 describes a set of DNA sequences for creating retroviral producer cells either *in vivo* or *in vitro* for re-implantation.

The components of the viral system which are required to complement the vector genome may be present on one or more “producer plasmids” for transfecting into cells.

5 The present invention provides a packaging cell comprising a nucleic acid construct according to the invention, and which cell expresses at least one detectable marker which is a cell surface protein comprising an extracellular domain and a membrane targeting domain.

In one embodiment, the packaging cell comprises:

10 (i) a first nucleic acid construct comprising a *gag-pol* sequence and a sequence encoding a first detectable marker;

(ii) a second nucleic acid construct comprising an *env* sequence and a sequence encoding a second detectable marker;

which co-expresses the first and second detectable markers at the cell surface.

15 The packaging cell may further comprise:

(iii) a third nucleic acid construct comprising a *rev* sequence and a sequence encoding a third detectable marker;

which co-expresses the first, second and third detectable markers at the cell surface.

20 The invention further provides a producer cell which comprises:

(i) a first nucleic acid construct comprising a *gag-pol* sequence and a sequence encoding a first detectable marker;

(ii) a second nucleic acid construct comprising an *env* sequence and a sequence encoding a second detectable marker;

25 (iii) optionally a third nucleic acid construct comprising a *rev* sequence and a sequence encoding a third detectable marker;

(iii) a fourth nucleic acid construct comprising a retroviral transfer vector and a sequence encoding a fourth detectable marker;

30 which co-expresses the first, second, third (if present) and fourth detectable markers at the cell surface.

The nucleic acid construct(s) may be stably integrated into the cell genome.

35 The packaging or producer cell may be a HEK293, HEK293-T, TE671 or HT1080, 3T3, or K562 cell.

METHOD

In one aspect the present invention provides a method for making a packaging cell or producer cell of the invention which comprises the step of introducing one or more nucleic acid construct(s) of the invention, one or more plasmid(s) of the invention, a kit of nucleic acid constructs of the invention, or a kit of plasmids of the invention, into a cell.

5

The nucleic acid construct may be introduced into a cell by a variety of methods which are known in the art, for example using standard transfection methods such as electroporation or lipofection.

10 The nucleic acid constructs may be introduced into the cell during the same step of the method (e.g. transfected at the same time) or at different steps in the method (e.g. transfected at different times).

In another aspect the present invention provides a method for selecting a packaging or
15 producer cell of the invention by selecting for expression of the or each detectable marker encoded by the or each nucleic acid construct.

The method may comprise the following steps:

(a) introducing

20 (i) a first nucleic acid construct comprising a *gag-pol* sequence and a sequence encoding a first detectable marker; and

(ii) a second nucleic acid construct comprising an *env* sequence and a sequence encoding a second detectable marker;

into a plurality of cells; and

25 (b) selecting a cell which co-expresses the first and second detectable markers.

Step (a) may further comprise introducing into the plurality of cells:

(iii) a third nucleic acid construct comprising a *rev* sequence and a sequence encoding a third detectable marker;

30 and step (b) further comprises selecting a cell which co-expresses the first, second and third detectable markers.

Step (a) may further comprise introducing into the plurality of cells:

35 (iv) a fourth nucleic acid construct comprising a retroviral transfer vector and a sequence encoding a fourth detectable marker;

and step (b) further comprises selecting a cell which co-expresses the first, second, fourth and optionally third detectable markers.

Step (a) may also comprise introducing a nucleic acid sequence encoding piggyBAC transposase.

5 In another aspect, the present invention provides a method for selecting a packaging cell or producer cell according to the invention by selecting for expression of the or each detectable marker encoded by the or each nucleic acid construct.

10 In one embodiment, the screening is performed by fluorescence-activated cell sorting (FACS).

For example, by sorting vector, gagpol, env and rev expressing cells using flow cytometry, it is possible to monitor the expression levels of the different retroviral components and select cells with optimal ratios of expression of each component for more productive retroviral production.

15 The cell may express a plurality of detectable markers and may be selected using multi-parameter flow cytometry.

The method may comprise the following steps:

- 20 (a) introducing
- (i) a first nucleic acid construct comprising a *gag-pol* sequence and a sequence encoding a first detectable marker;
 - (ii) a second nucleic acid construct comprising an *env* sequence and a sequence encoding a second detectable marker;
- 25 into a plurality of cells; and
- (b) selecting a cell which co-expresses the first and second detectable markers.

Step (a) may further comprise introducing into the plurality of cells:

- 30 (iii) a third nucleic acid construct comprising a *rev* sequence and a sequence encoding a third detectable marker;
- and step (b) further comprises selecting a cell which co-expresses the first, second and third detectable markers.

Step (a) may further comprise introducing into the plurality of cells:

- 35 (iv) a fourth nucleic acid construct comprising a retroviral transfer vector and a sequence encoding a fourth detectable marker;

and step (b) further comprises selecting a cell which co-expresses the first, second, fourth and optionally third detectable markers.

5 Step (a) may also comprise introducing a nucleic acid sequence encoding piggyBAC transposase.

In another aspect the invention provides a method for producing a retroviral vector which comprises the steps of culturing a producer cell of the present invention and isolating the retroviral vector.

10

The producer cell may be cultured using standard culture conditions suitable for the particular cell type used to make the producer cell. During culture, retrovirus vector particles are released into the culture supernatant. The retrovirus vector may be isolated using standard techniques, for example, ultracentrifugation, ultrafiltration or affinity purification.

15

The retroviral vector may be a lentiviral vector.

Definitions of terms appear throughout the specification. Before the exemplary embodiments are described in more detail, it is to be understood that this disclosure is not limited to particular embodiments described, and as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting, since the scope of the present disclosure will be limited only by the appended claims.

25 Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure belongs. Singleton, *et al.*, DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY, 20 ED., John Wiley and Sons, New York (1994), and Hale & Marham, THE HARPER COLLINS DICTIONARY OF BIOLOGY, Harper Perennial, NY (1991) provide one of skill with a general dictionary of many of the terms used in this disclosure.

30 This disclosure is not limited by the exemplary methods and materials disclosed herein, and any methods and materials similar or equivalent to those described herein can be used in the practice or testing of embodiments of this disclosure. Numeric ranges are inclusive of the numbers defining the range.

35 Where a range of values is provided, it is understood that each intervening value, to the tenth of the unit of the lower limit unless the context clearly dictates otherwise, between the

upper and lower limits of that range is also specifically disclosed. Each smaller range between any stated value or intervening value in a stated range and any other stated or intervening value in that stated range is encompassed within this disclosure. The upper and lower limits of these smaller ranges may independently be included or excluded in the range, and each range where either, neither or both limits are included in the smaller ranges is also encompassed within this disclosure, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in this disclosure.

It must be noted that as used herein and in the appended claims, the singular forms "a", "an", and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "an enzyme" includes a plurality of such candidate agents and equivalents thereof known to those skilled in the art, and so forth.

The terms "comprising", "comprises" and "comprised of" as used herein are synonymous with "including", "includes" or "containing", "contains", and are inclusive or open-ended and do not exclude additional, non-recited members, elements or method steps. The terms "comprising", "comprises" and "comprised of" also include the term "consisting of".

The invention will now be further described by way of Examples, which are meant to serve to assist one of ordinary skill in the art in carrying out the invention and are not intended in any way to limit the scope of the invention.

EXAMPLES

Example 1 – Construction of cell surface marker proteins

Possible compact surface marker genes are detailed in Figure 2. Coding sequences for well-characterized epitope tags (V5, HA and MYC-tags) were cloned 3' to the coding sequence for a signal peptide. These in turn were cloned to the stalk, TM and a portion of the endodomain of CD8alpha. Alternatively, they were cloned in frame with the GPI-anchor signal of CD52. This resulted in a set of highly compact surface marker proteins which display the epitope on the surface of the cell. This allows easy detection by staining with a fluorescently conjugated antibody. The compact size of these markers keep the transcriptional burden of expressing a marker gene to a minimum. Surface expression means the cell does not have to be lysed to allow detection.

Example 2 – Expression of cell surface marker proteins in 293T cells

293T cells were co-transfected with a piggyBAC transposase expression plasmid and a lentiviral rev expression plasmid which also co-expressed V5-8 marker gene. Some 293T cells were also transfected with just the rev expression plasmid alone as a control. These latter cells allow determination of the contribution to expression from transient transfection as opposed to permanent insertion. At 10 days, both populations of cells as well as non-transfected controls were stained with a fluorescent antibody which recognizes the V5 epitope tag. These cells were then analysed by flow-cytometry. A clear population of V5-8 expression cells were seen in the transposase co-transfected cells, while expression in the rev expression alone 293T cells was lower (Figure 3). These cell cultures were maintained for 95 days and periodically analysed for v5 epitope expression. This demonstrates the stability of piggyBAC mediated transposition, and the convenience of marker gene in demonstrating stability of a lentiviral element.

Example 3 – Co-expression of different lentiviral proteins with cell surface marker proteins

Different lentiviral elements require different means to optimally co-express a marker gene. Lentiviral gagpol was tagged in two different ways: HA-8 tag marker gene was co-expressed with a FMD-2A like peptide, or with an IRES sequence (Figure 4). Three lentiviral gagpol expression cassettes were tested: a control without a tag, the 2A-tagged version and the IRES-tagged version. 293T cells transfected with these plasmids were stained with an anti-HA antibody and analysed by flow-cytometry. The tag could be readily detected in both tag-2A and IRES.tag constructs. Lentiviral supernatant was generated by transiently transfecting helper plasmids and transfer vector where the gagpol was supplied by either one of the above three plasmids. Use of the Tag-2A construct resulted in a lower virus titer. This demonstrated that the IRES was an optimal way of tagging gagpol in this experiment.

Example 4 – Optimal lentiviral helper expression cassette design

The helper cassettes must supply rev, RD-PRO and gagpol. Suggested optimal cassettes are shown in Figure 5. To assist insertion into the packaging cell genome for stable expression, each cassette is flanked by piggyBAC terminal repeats (pB-TR). Promoters, polyadenylation sequences and introns are selected to be as different as possible to reduce the probability of homologous recombination. In this case, rev cassette contains an RSV promoter, the RD-PRO cassette contains an SV40/Ferritin heavy-chain composite promoter and the gagpol cassette contains a CMV/ β -actin composite promoter. Each lentiviral element is co-expressed with a surface epitope marker. Co-expression in the Rev and RD-PRO cassette is achieved with a FMD like 2A peptide. Co-expression in the gagpol cassette is achieved with an IRES. RD-PRO requires an intron for efficient expression. This is provided

by the mouse EFalpha 5' UTR. The gagpol cassette contains an interon from rabbit β -globin as well a scaffold attachment region (SAR) to enhance expression.

Example 5 – Selection of optimal lentiviral producer cells

5 Figure 1 shows a scheme for generation of a stable lentiviral producer line. 293T cells are co-transfected with lentiviral helper plasmids as described above. A lentiviral transfer vector is also co-transfected. One of the transfer vector transgenes is RQR8 which, like the tags, is easily detected on the surface of the packaging cell. Like the helper plasmids, the transfer vector is flanked by piggyBAC terminal repeats. An expression plasmid supplying the
10 piggyBAC transposase is also co-transfected. After a period of culture, permanent expression of helper and transgene cassettes can easily be detected by flow cytometry. Figure 6 shows 293T stably transposed with lentiviral helper cassettes and a transfer vector. All 4 tags can be independently detected and populations of cells which express desired amounts of the different elements can be gated (an example gate is shown in Figure 6) and
15 flow-sorted. Large numbers of these “bulk” transposed cells are sorted by flow cytometric sorting, with precise population(s) of cells with high transgene and optimal relative expression being selected. These are directly sorted as single-cells by the flow-sorter. They are allowed to expand to fill the well of a 96 or 385 well plate and supernatant is titered. The highest titer clones are selected for virus production.

20

Example 6 - Lentiviral vector production from a stably transposed tagged gagpol expression cassette

An expression cassette was generated where lentiviral gagpol could be transposed (the cassette was flanked by piggyBAC terminal repeats – see Figure 10). Further, gagpol was
25 co-expressed with a surface expressed marker gene. 293T cells were co-transfected with this plasmid and an expression plasmid which provided piggyBAC transposase. The 293T cells were cultured for 10 days to allow expression to stabilize (i.e. transient expression from transfection to be lost, leaving only expression from stably transposed cells). By staining for the marker gene, 293T cells were sorted by flow cytometry sorting into low, medium and
30 high expressing cells. These populations of 293T cells were expanded after the sort and were then transfected with RD114 envelop, REV and a transfer vector. Supernatant was harvested and titered on 293T cells.

Stable transposed gagpol could result in vector production. Low expressing cells generated
35 higher titer. This suggests that careful gating of cell populations possible during flow sorting may help to identify particularly productive cell populations or clones (Figure 7).

Example 7 - Lentiviral vector production from a stably transposed tagged transfer vector cassette

5 A lentiviral transfer vector was generated which was tagged with a surface-expressed epitope marker which was flanked by piggyBAC terminal repeats (see Figure 9). 293T cells were co-transfected with this cassette and a second expression cassette which supplied piggyBAC transposase. The 293T cells were cultured for 10 days to allow expression to stabilize (i.e. transient expression to dissipate). By staining for the marker gene, 293T cells could be sorted for cells which were high or low expressing. Cells which were not expressing the marker could be excluded. These 293T cells were cultivated and transfected with plasmids supplying gagpol, RD114 envelop and REV. Supernatant was tittered on 293T cells.

10 Titration showed that viral vector was produced. Productivity did not correlate with high expression of marker gene (see Figure 8).

15

All publications mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described methods and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in the art are intended to be within the scope of the following claims.

25

CLAIMS

1. A nucleic acid construct comprising:
(i) a first nucleic acid sequence which either comprises a retroviral transfer vector or
5 which encodes a retroviral protein; and
(ii) a second nucleic acid sequence which encodes a detectable marker which is a cell surface protein comprising an extracellular domain and a membrane targeting domain.
2. A nucleic acid construct according to claim 1 having the structure:
10 A-X-B
in which
A is the first nucleic acid sequence, B is the second nucleic acid sequence and X is a co-expression sequence.
3. A nucleic acid construct according to claim 2, wherein the co-expression sequence
15 comprises an IRES sequence or a sequence encoding a self-cleaving peptide.
4. A nucleic acid construct according to any preceding claim wherein the first nucleic acid sequence encodes a retroviral protein selected from gag-pol, env or rev.
20
5. A nucleic acid construct according to any preceding claim wherein the cell surface protein has less than 200 amino acids.
6. A nucleic acid construct according to any preceding claim wherein the membrane
25 targeting domain of the cell surface protein comprises (i) a transmembrane domain; or (ii) a GPI anchor.
7. A nucleic acid construct according to any preceding claim wherein the membrane targeting domain comprises a CD8 stalk or a part thereof.
30
8. A nucleic acid construct according to any preceding claim wherein the extracellular domain comprises HA, V5, RQR8 or MYC.
9. A nucleic acid construct according to any preceding claim wherein the cell surface
35 protein comprises an amino acid sequence as shown as SEQ ID NO: 21-26.

10. A nucleic acid construct according to any preceding claim which comprises transposon sequences flanking the first and second nucleic acid sequence sequences.
11. A nucleic acid construct according to claim 10, having the structure:
5 T1-A-X-B-T2
In which A, X and B are as defined in claim 2;
T1 is a first transposon sequence; and
T2 is a second transposon sequence.
- 10 12. A nucleic acid construct according to claim 11 wherein T1 is a PiggyBAC 5' Terminal Repeat comprising the sequence shown as SEQ ID NO: 60 and T2 is a PiggyBAC 3' Terminal repeat comprising the sequence shown as SEQ ID NO: 61.
13. A kit comprising a plurality of nucleic acid constructs according to any preceding
15 claim.
14. A kit according to claim 13, which comprises:
(i) a first nucleic acid construct comprising a *gag-pol* sequence and a sequence encoding a first detectable marker;
20 (ii) a second nucleic acid construct comprising an *env* sequence and a sequence encoding a second detectable marker;
in which the first and second detectable markers are cell surface proteins comprising an extracellular domain and a membrane targeting domain and are different to each other.
- 25 15. A kit according to claim 14 which further comprises:
(iii) a third nucleic acid construct comprising a *rev* sequence and a sequence encoding a third detectable marker which is a cell surface protein comprising an extracellular domain and a membrane targeting domain;
in which the first, second and third and detectable markers are all different to each other.
30
16. A kit according to claim 14 or 15, which further comprises:
(iv) a fourth nucleic acid construct comprising a retroviral transfer vector and a sequence encoding a fourth detectable marker which is a cell surface protein comprising an extracellular domain and a membrane targeting domain;
35 in which the first, second, third (if present) and fourth detectable markers are all different to each other.

17. A plasmid comprising a nucleic acid construct as defined in any of claims 1 to 12.
18. A kit as defined in any of claims 13 to 16 which comprises a plurality of plasmids according to claim 17.
- 5
19. A packaging cell comprising a nucleic acid construct according to any of claims 1 to 12 which expresses at least one detectable marker which is a cell surface protein comprising an extracellular domain and a membrane targeting domain.
- 10 20. A packaging cell according to claim 19 comprising:
- (i) a first nucleic acid construct comprising a *gag-pol* sequence and a sequence encoding a first detectable marker;
 - (ii) a second nucleic acid construct comprising an *env* sequence and a sequence encoding a second detectable marker;
- 15 which co-expresses the first and second detectable markers at the cell surface.
21. A packaging cell according to claim 20 which further comprises:
- (iii) a third nucleic acid construct comprising a *rev* sequence and a sequence encoding a third detectable marker;
- 20 which co-expresses the first, second and third detectable markers at the cell surface.
22. A producer cell which comprises first and second nucleic acid constructs as defined in claim 20 and optionally a third nucleic acid construct as defined in claim 21 and which further comprises:
- 25 (iii) a fourth nucleic acid construct comprising a retroviral transfer vector and a sequence encoding a fourth detectable marker;
- which co-expresses the first, second, third (if present) and fourth detectable markers at the cell surface.
- 30 23. A packaging cell according to any of claims 19 to 21 or a producer cell according to claim 22 wherein the or each nucleic acid construct is/are stably integrated into the cell genome.
24. A packaging cell or producer cell according to any claim of claims 19 to 23 which is a
- 35 HEK293, HEK293-T, TE671, HT1080, 3T3, or K562 cell.

25. A method for making a packaging cell or producer cell according to any of claims 19 to 24 which comprises the step of introducing one or more nucleic acid construct(s) as defined in any of claims 1 to 12, one or more plasmid(s) as defined in claim 17, a kit of nucleic acid constructs as defined in any of claims 13 to 16, or a kit of plasmids as defined in claim 18, into a cell.
26. A method for selecting a packaging cell or producer cell according to any of claims 19 to 24 by selecting for expression of the or each detectable marker encoded by the or each nucleic acid construct.
27. A method according to claim 26, wherein the cell expresses a plurality of detectable markers and the cell is selected using multi-parameter flow cytometry.
28. A method according to claim 26 or 27 which comprises the following steps:
- (a) introducing
 - (i) a first nucleic acid construct comprising a *gag-pol* sequence and a sequence encoding a first detectable marker;
 - (ii) a second nucleic acid construct comprising an *env* sequence and a sequence encoding a second detectable marker;
 - (b) selecting a cell which co-expresses the first and second detectable markers.
29. A method according to claim 28 in which step (a) further comprises introducing into the plurality of cells:
- (iii) a third nucleic acid construct comprising a *rev* sequence and a sequence encoding a third detectable marker;
- and step (b) further comprises selecting a cell which co-expresses the first, second and third detectable markers.
30. A method according to claim 28 or 29 in which step (a) further comprises introducing into the plurality of cells:
- (iv) a fourth nucleic acid construct comprising a retroviral transfer vector and a sequence encoding a fourth detectable marker;
- and step (b) further comprises selecting a cell which co-expresses the first, second, fourth and optionally third detectable markers.

31. A method according to any of claims 28 to 30 wherein step (a) further comprises introducing a nucleic acid sequence encoding piggyBAC transposase.

5 32. A method for producing a retroviral vector which comprises the steps of culturing a producer cell according to claim 22 and isolating the retroviral vector.

33. A method for producing a retroviral vector which comprises the steps of introducing a retroviral vector genome into a packaging cell according to any of claims 19 to 21, 23 or 24 and isolating the retroviral vector.

10

34. A method according to claim 32 or 33 wherein the retroviral vector is a lentiviral vector.

1/11

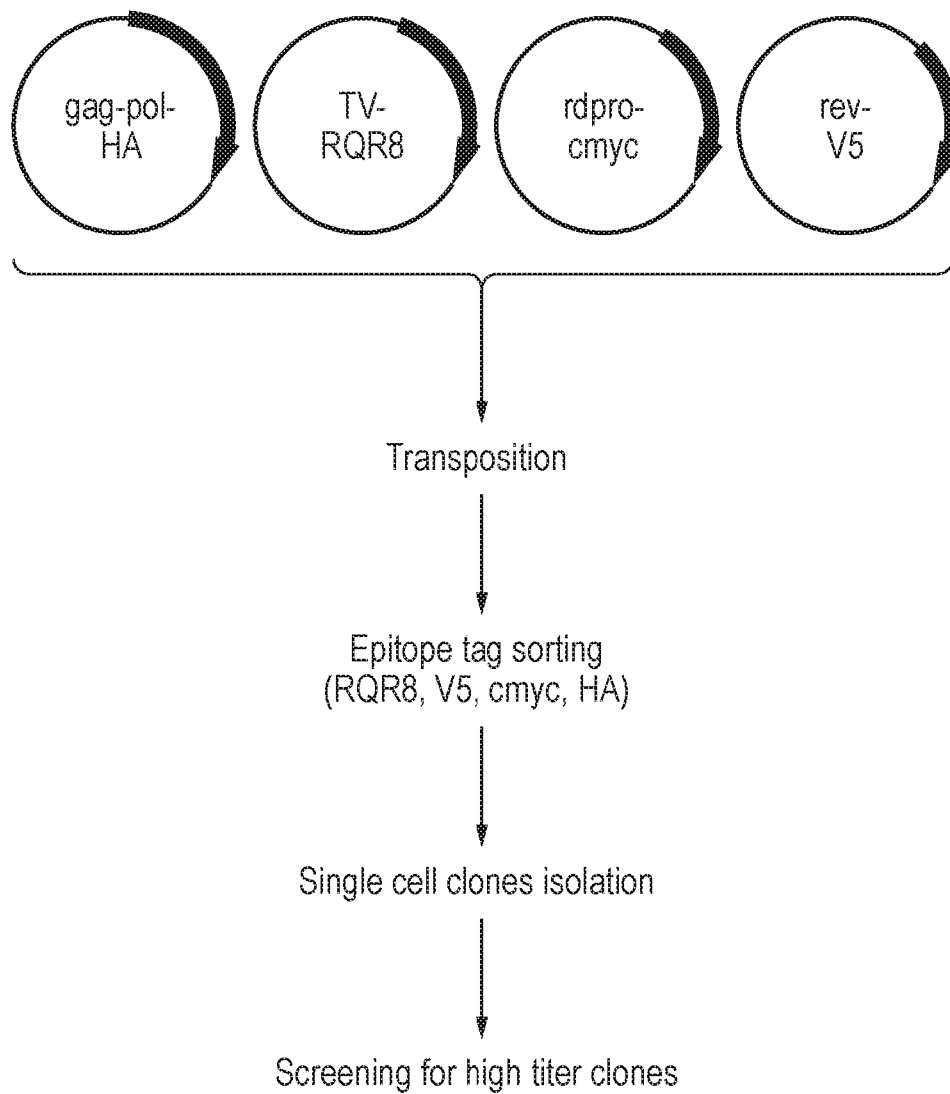


FIG. 1

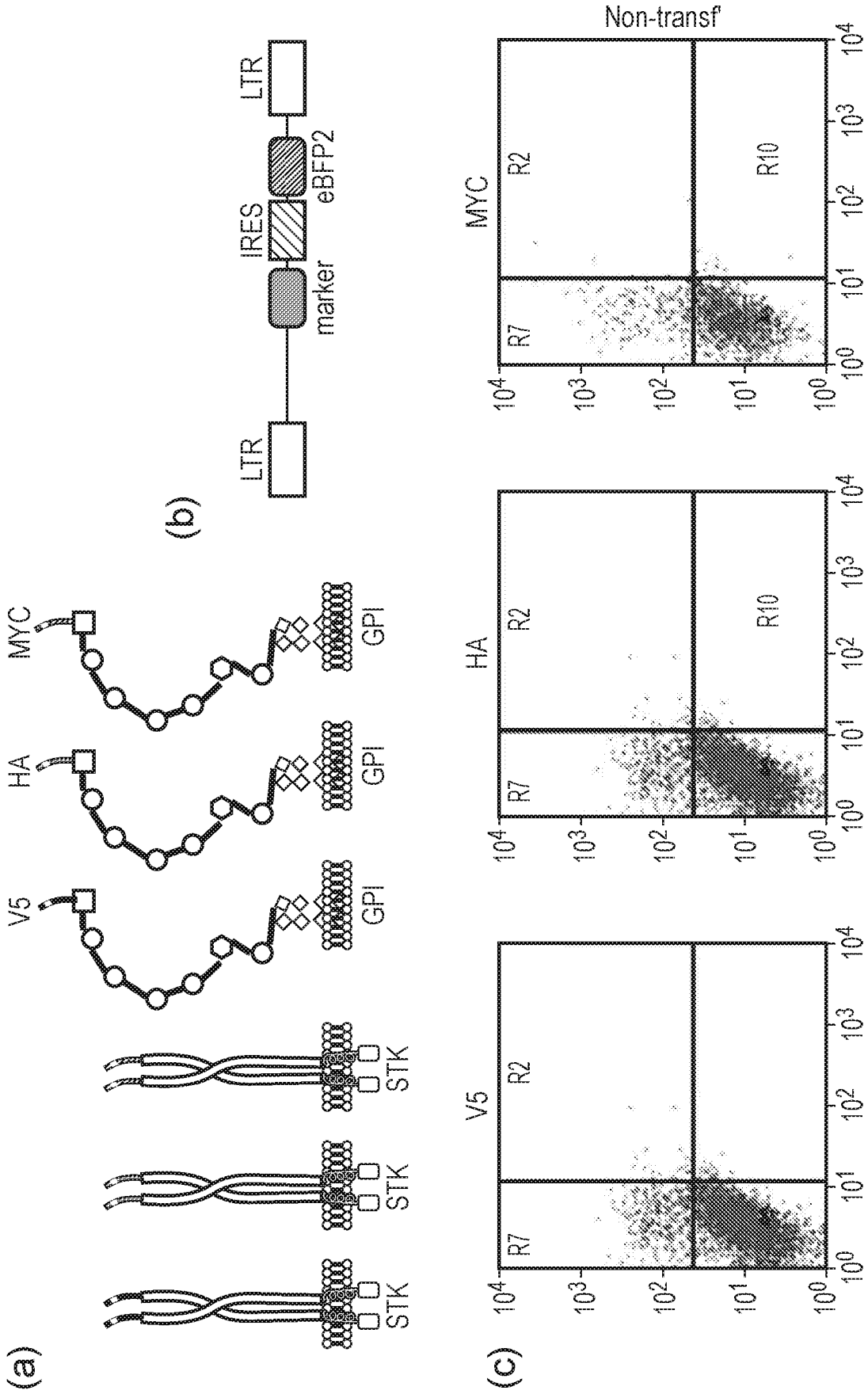


FIG. 2

3/11

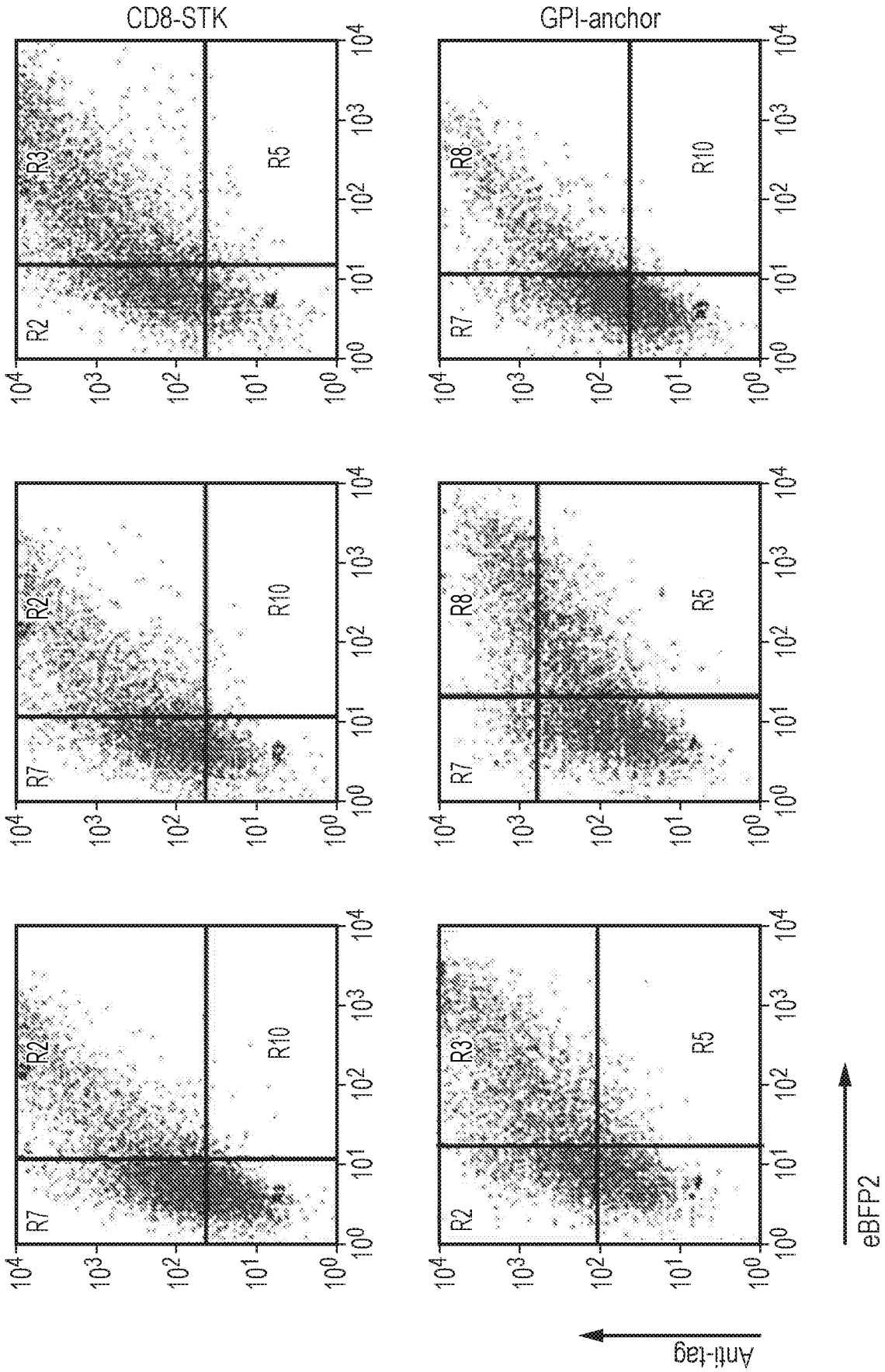


FIG. 2 (Continued)

4/11

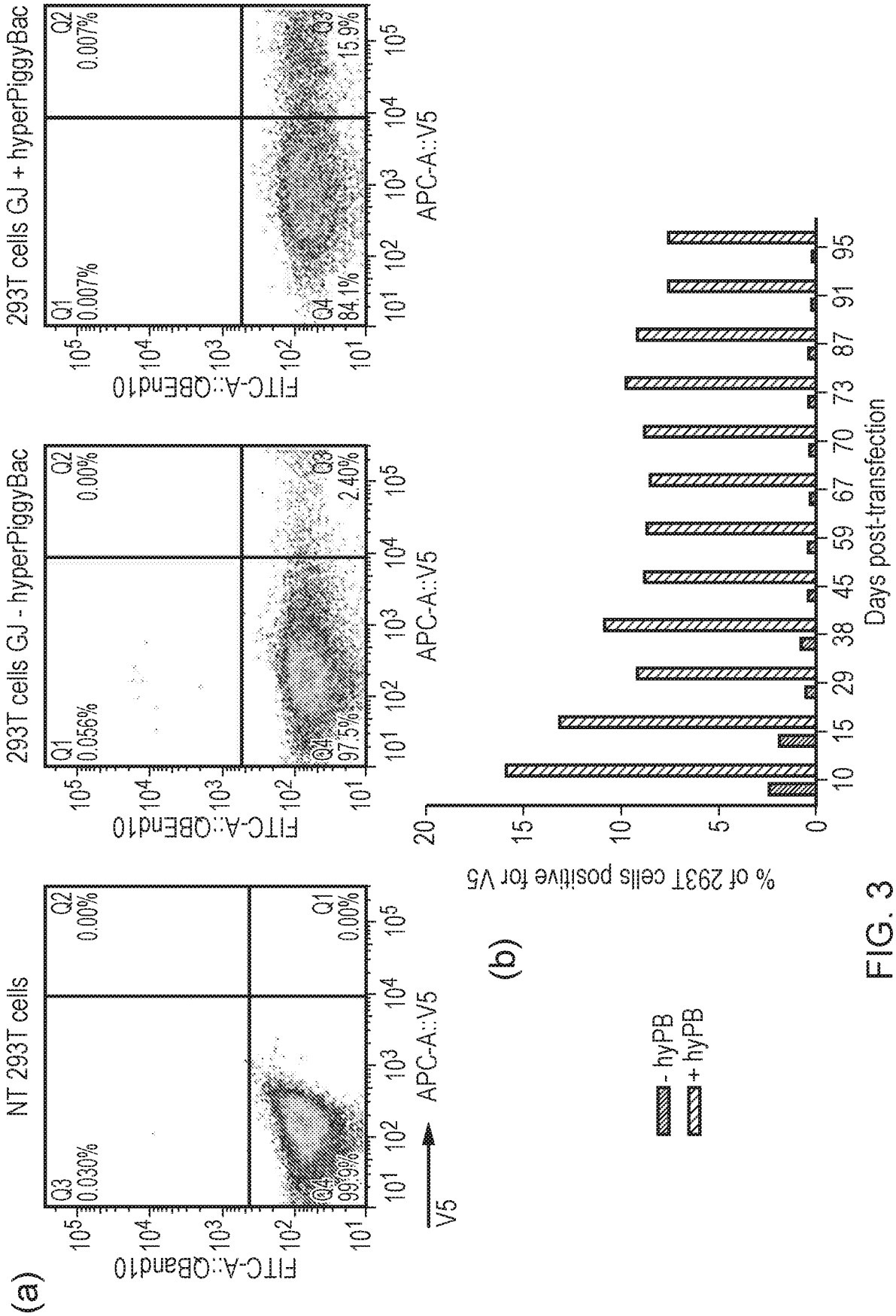


FIG. 3

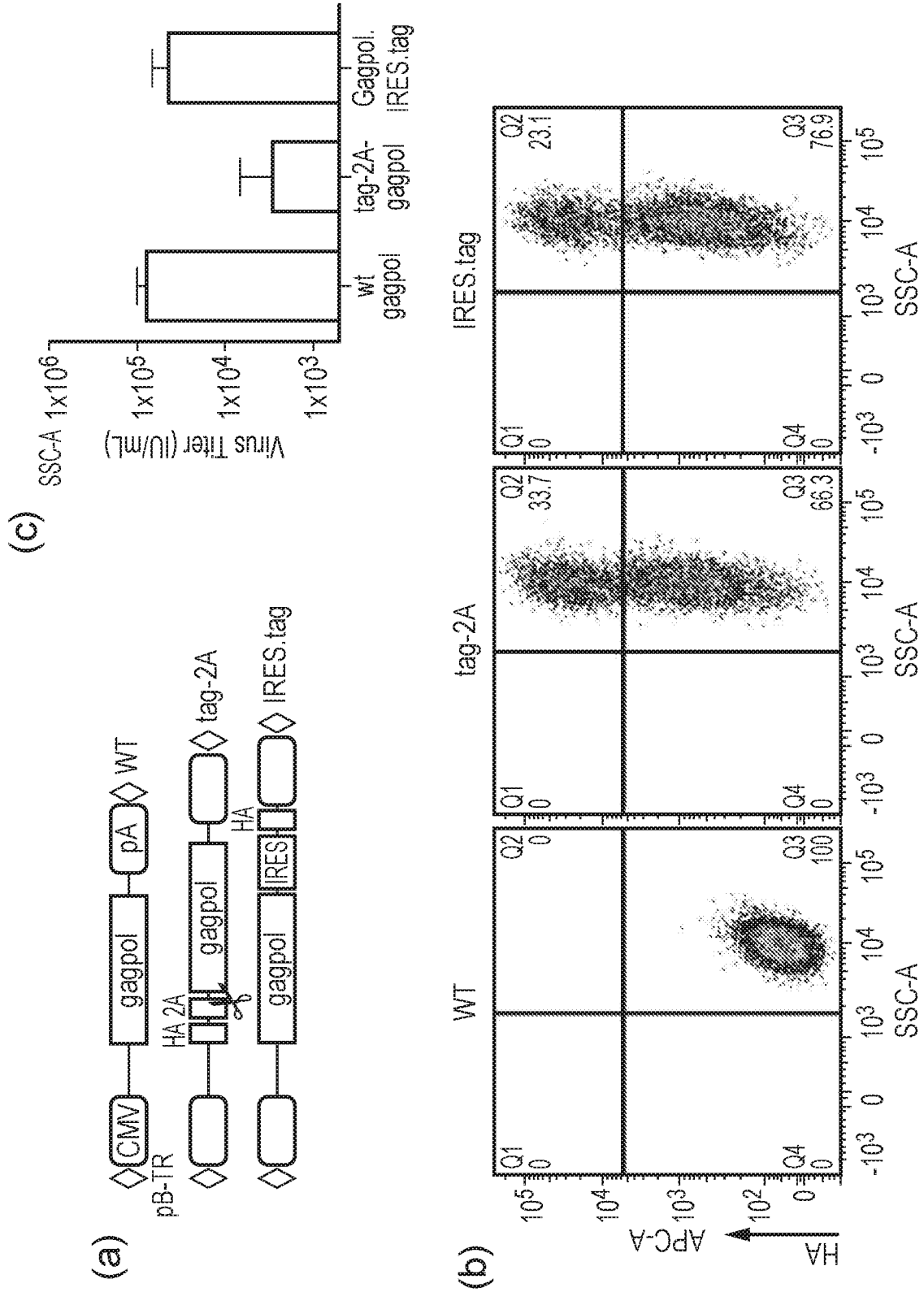


FIG. 4

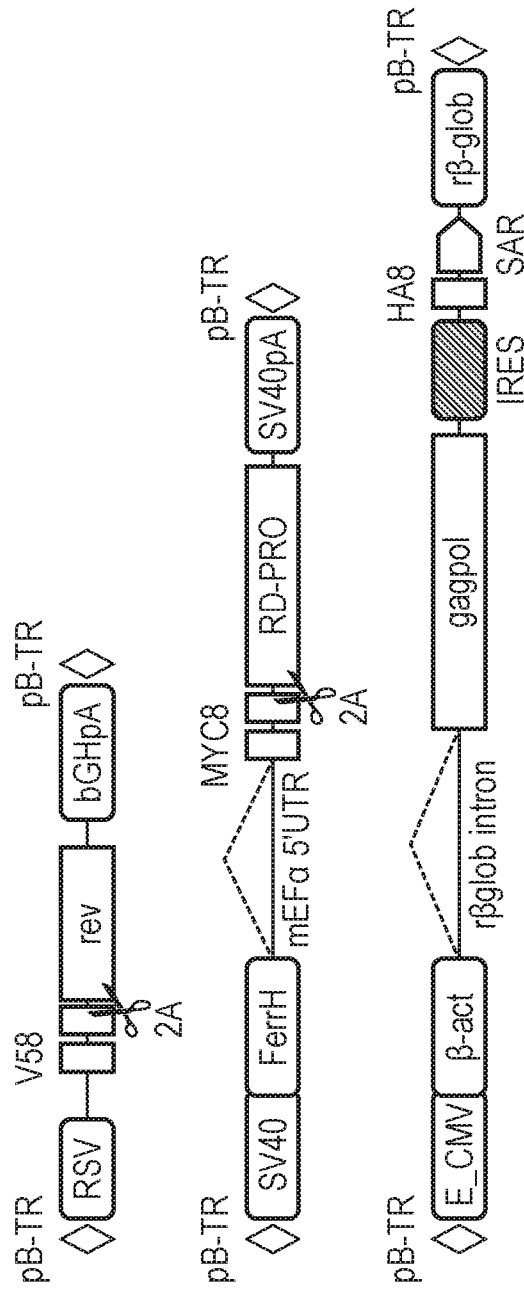


FIG. 5

7/11

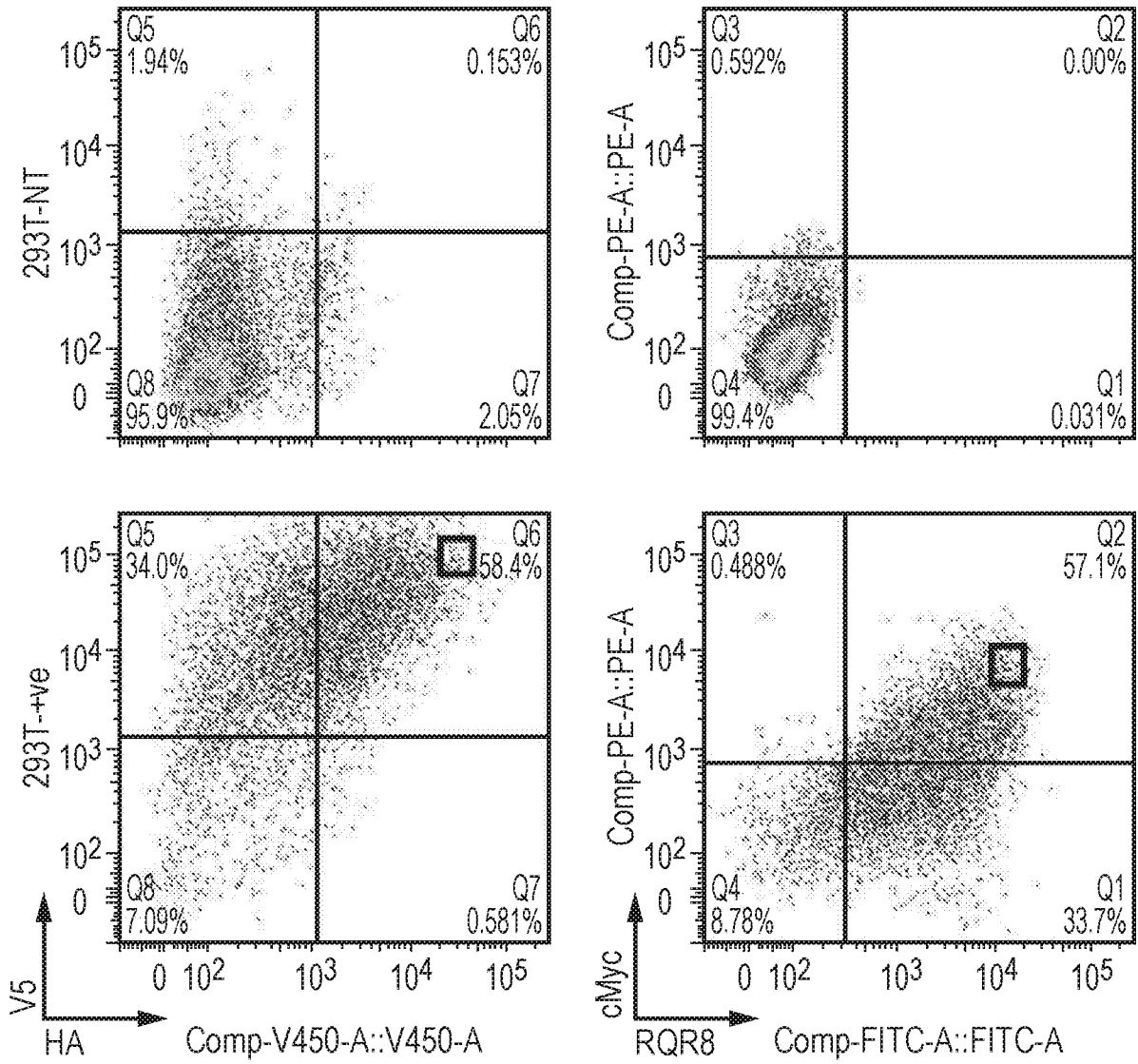
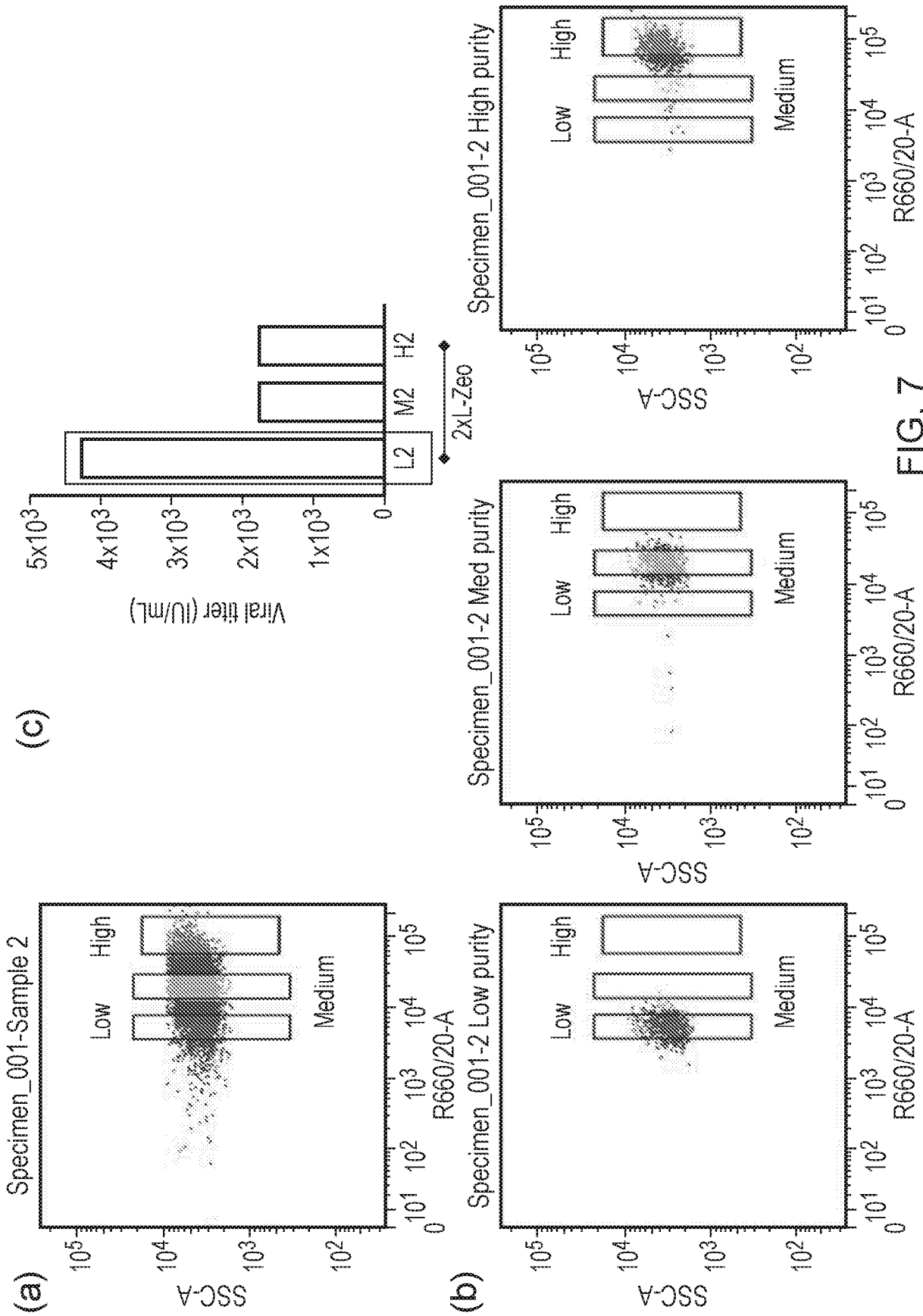


FIG. 6

8/11



9/11

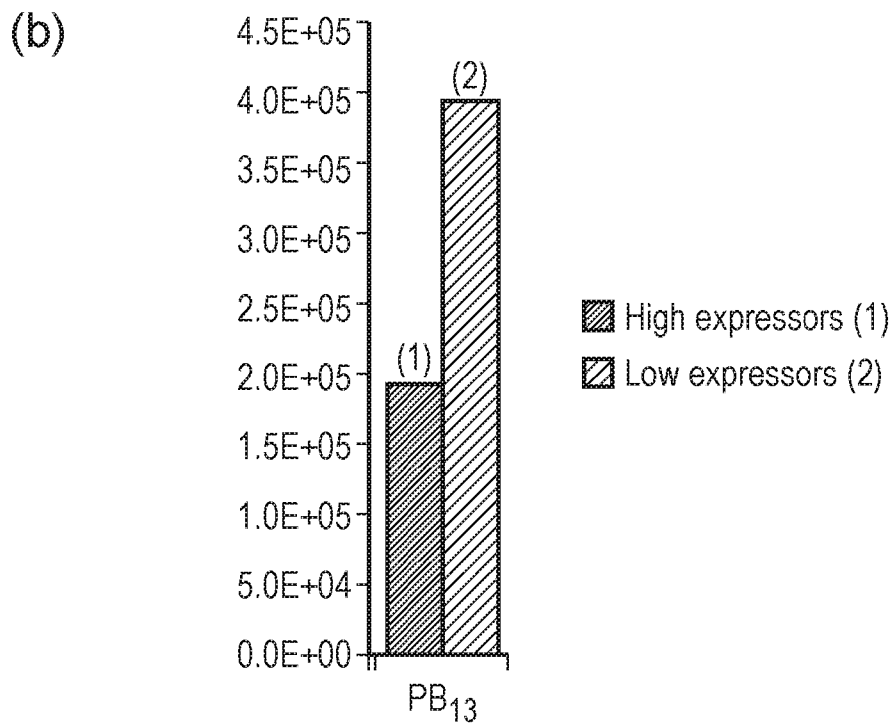
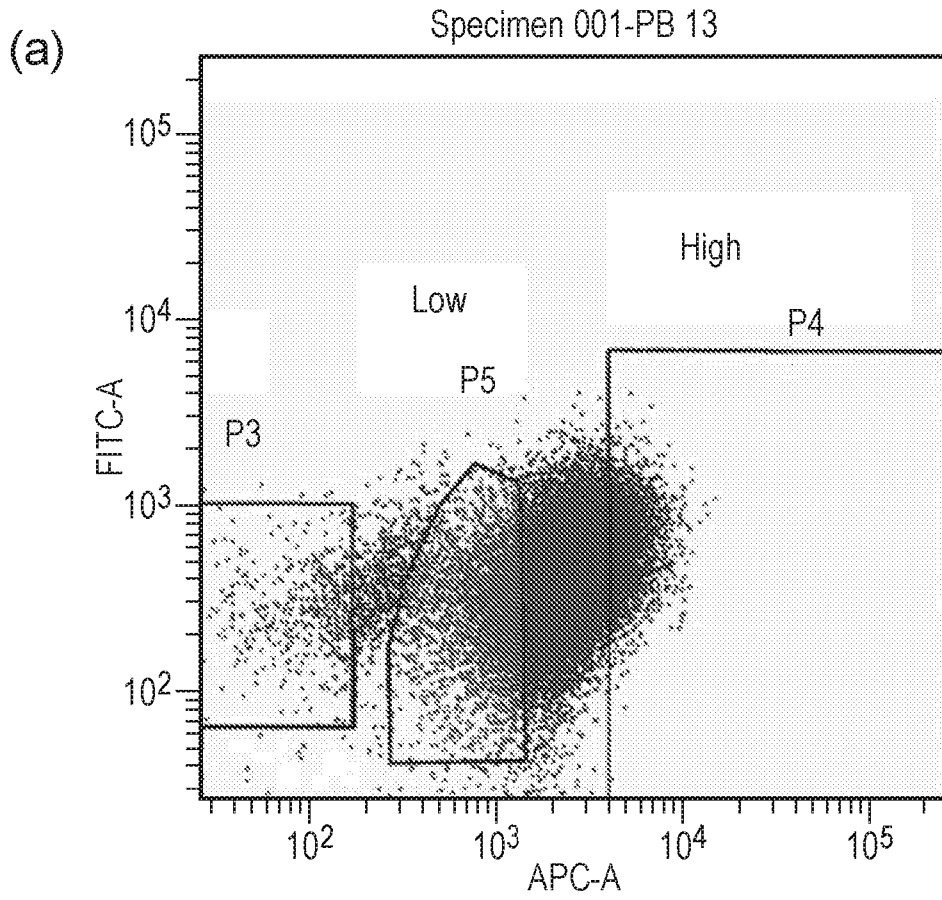


FIG. 8

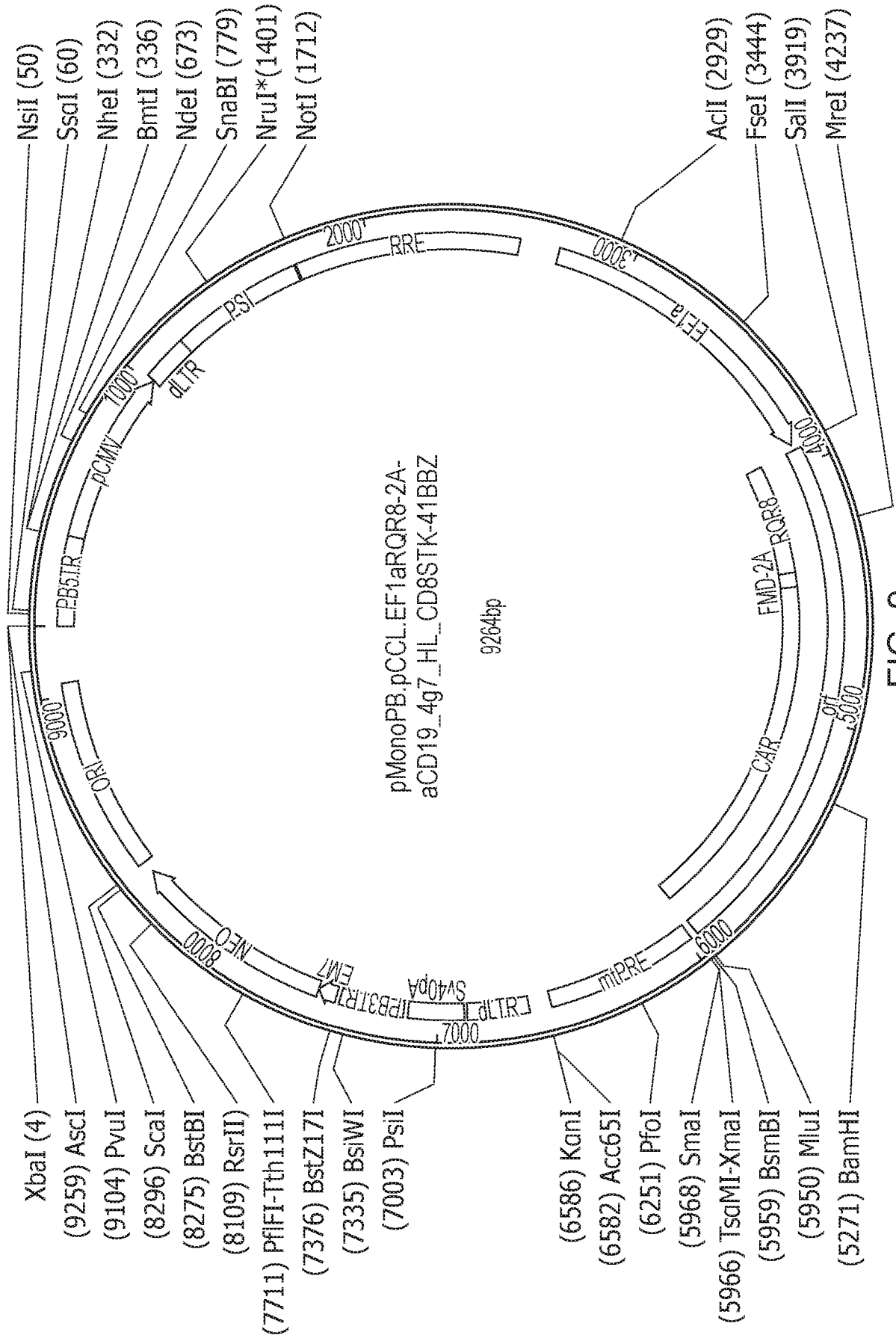


FIG. 9

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2017/050481

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N15/86
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED
Minimum documentation searched (classification system followed by classification symbols)
C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, BIOSIS, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2014/121005 A1 (US HEALTH [US]) 7 August 2014 (2014-08-07) paragraphs [0016], [0075]; figure 1 paragraph [0017]	1-34
A	----- WO 92/05266 A2 (VIAGENE INC [US]) 2 April 1992 (1992-04-02) cited in the application the whole document	1-34
X,P	----- WO 2016/030690 A1 (UCL BUSINESS PLC [GB]) 3 March 2016 (2016-03-03) claims 1, 24 examples 3, 4 -----	1-34

Further documents are listed in the continuation of Box C.

See patent family annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier application or patent but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- "&" document member of the same patent family

Date of the actual completion of the international search

18 April 2017

Date of mailing of the international search report

28/04/2017

Name and mailing address of the ISA/
European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Herrmann, Klaus

INTERNATIONAL SEARCH REPORT

International application No.

PCT/GB2017/050481

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
 - a. forming part of the international application as filed:
 - in the form of an Annex C/ST.25 text file.
 - on paper or in the form of an image file.
 - b. furnished together with the international application under PCT Rule 13~~ter~~.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.
 - c. furnished subsequent to the international filing date for the purposes of international search only:
 - in the form of an Annex C/ST.25 text file (Rule 13~~ter~~.1(a)).
 - on paper or in the form of an image file (Rule 13~~ter~~.1(b) and Administrative Instructions, Section 713).
2. In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.
3. Additional comments:

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/GB2017/050481

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2014121005	A1	07-08-2014	NONE

WO 9205266	A2	02-04-1992	AU 665176 B2 21-12-1995
			AU 690427 B2 23-04-1998
			CA 2092195 A1 22-03-1992
			EP 0549723 A1 07-07-1993
			EP 0854193 A1 22-07-1998
			EP 1285967 A2 26-02-2003
			JP 4092278 B2 28-05-2008
			JP H06500923 A 27-01-1994
			JP 2003159062 A 03-06-2003
			JP 2004041234 A 12-02-2004
			WO 9205266 A2 02-04-1992

WO 2016030690	A1	03-03-2016	NONE
