VECTOR FOR THE PRODUCTION OF AAV PARTICLES

The present invention relates to the production of plasmids which are useful in the production of Adeno-Associated Virus (AAV) particles. In particular, the invention provides nucleic acid molecules comprising *cap* genes and *rep* genes, wherein the *cap* and *rep* genes are both operably-associated with the same promoter. The invention also provides host cells comprising nucleic acid molecules of the invention and methods for their use.

AAV vectors are developed from single-stranded DNA viruses that belong to the *Parvoviridae* family. This virus is capable of infecting a broad range of host cells, including both dividing and non-dividing cells. In addition, it is a non-pathogenic virus that generates only a limited immune response in most patients.

The AAV genome comprises two genes each encoding multiple open reading frames (ORFs): the *rep* gene encodes non-structural proteins that are required for the AAV life-cycle and site-specific integration of the viral genome; and the *cap* gene encodes the structural capsid proteins. In addition, these two genes are flanked by inverted terminal repeat (ITR) sequences consisting of 145 bases that have the ability to form hairpin structures. These hairpin sequences are required for the primase-independent synthesis of a second DNA strand and the integration of the viral DNA into the host cell genome.

In order to eliminate any integrative capacity of the virus, recombinant AAV vectors remove *rep* and *cap* from the DNA of the viral genome. To produce such vectors, the desired transgene(s), together with a promoter(s) to drive transcription of the transgene(s), is inserted between the inverted terminal repeats (ITRs); and the *rep* and *cap* genes are provided in *trans* in a second plasmid. A third plasmid, providing helper genes such as adenovirus E4, E2a and VA genes, is also used. All three plasmids are then transfected into cultured 'packaging' cells, such as HEK293.

Over the last few years, AAV vectors have emerged as an extremely useful and promising mode of gene delivery. This is owing to the following properties of these vectors:
- AAVs are small, non-enveloped viruses and they have only two native genes (*rep* and *cap*). Thus they can be easily manipulated to develop vectors for different gene therapies.

- AAV particles are not easily degraded by shear forces, enzymes or solvents. This facilitates easy purification and final formulation of these viral vectors.
- AAVs are non-pathogenic and have a low immunogenicity. The use of these vectors further reduces the risk of adverse inflammatory reactions. Unlike other viral vectors, such as lentivirus, herpes virus and adenovirus, AAVs are harmless and are not thought to be responsible for causing any human disease.
- Genetic sequences up to 4000 bp can be delivered into a patient using AAV vectors.
- Whilst wild-type AAV vectors have been shown to sometimes insert genetic material into human chromosome 19, this property is generally eliminated from most AAV vectors by removing *rep* and *cap* genes from the viral genome. In such cases, the virus remains in an episomal form within the host cells. These episomes remain intact in non-dividing cells, while in dividing cells they are lost during cell division.

The inventors have recognised, however, that methods for the production of AAV vectors can be improved by optimising the ratios and amounts of the Rep and Cap proteins present during the vector-production process.

It is an object of the invention, therefore, to provide a nucleic acid molecule which comprises *cap* and *rep* genes which are under the control of a single promoter; Cap and Rep polypeptides are thereby encoded within the same mRNA. The translation of the *cap* gene will be initiated by docking, at the ribosome, of a methylguanylate cap (m^7G) at the 5' terminal of the *cap* mRNA. Translation of the *rep* gene will be initiated by docking of a ribosome at an Internal Ribosome Entry Site (IRES) which is placed upstream of the *rep* gene.

Through the use of the nucleic acid molecules of the invention, higher virus titres may be obtained.

In some embodiments of the invention, the IRES replaces the wild-type p5 promoter. A further advantage of the removal of the p5 promoter is that, in the wild-type virus, the p5 promoter is bound by and is activated by the E2A DNA-binding protein (DBP). Hence the removal of the p5 promoter means that the E2A gene is not required (e.g. in a Helper Plasmid) to produce virus particles.

In one embodiment, the invention provides a nucleic acid molecule comprising:

(i) a promoter,

(ii) a *cap* gene, and

(iii) a *rep* gene,

in the above 5' - 3' order, wherein the *cap* gene and the *rep* gene are both operably-associated with the promoter, and wherein the *rep* gene is also operably-associated with an IRES.

The invention also provides a nucleic acid molecule comprising:

(i) a *cap* gene, and

(ii) a *rep* gene,

in the above 5' - 3' order, wherein the *rep* gene is operably-associated with an IRES.

The nucleic acid molecule may be DNA or RNA, preferably DNA. The nucleic acid molecule may be single- or double-stranded, preferably double-stranded.

The nucleic acid molecule of the invention comprises a *rep* gene. As used herein, the term "*rep* gene" refers to a gene that encodes one or more open reading frames (ORFs), wherein each of said ORFs encodes an AAV Rep non-structural protein, or variant or derivative thereof. These AAV Rep non-structural proteins (or variants or derivatives thereof) are involved in AAV genome replication and/or AAV genome packaging.

The structure of the wild-type AAV genome, illustrating the organisation of the wild-type *rep* and *cap* genes, is shown in Figure 1.

The wild-type *rep* gene comprises three promoters: p5, p19 and p40. Two overlapping messenger ribonucleic acids (mRNAs) of different lengths can be produced from p5 and from p19. Each of these mRNAs contains an intron which can be either spliced out or not using a single splice donor site and two different splice acceptor sites. Thus six different mRNAs can be formed, of which only four are functional. The two mRNAs that fail to remove the intron (one transcribed from p5 and one from p19) read through to a shared terminator sequence and encode Rep78 and Rep52, respectively. Removal of the intron and use of the 5'-most splice acceptor site does not result in production of any functional Rep protein – it cannot produce the correct Rep68 or Rep40 proteins as the frame of the remainder of the sequence is shifted, and it will also not produce the correct C-terminus of Rep78 or Rep52 because their terminator is spliced out. Conversely, removal of the intron and use of the 3' splice acceptor will include the correct C-terminus for Rep68 and Rep40, whilst splicing out the terminator of Rep78 and Rep52. Hence the only functional splicing either avoids splicing out the intron altogether (producing

Rep78 and Rep52) or uses the 3' splice acceptor (to produce Rep68 and Rep40). Consequently four different functional Rep proteins with overlapping sequences can be synthesized from these promoters.

In the wild-type *rep* gene, the p40 promoter is located at the 3' end. Transcription of the Cap proteins (VP1, VP2 and VP3) is initiated from this promoter in the wild-type AAV genome.

The four wild-type Rep proteins are Rep78, Rep68, Rep52 and Rep40. Hence the wild-type *rep* gene is one which encodes the four Rep proteins Rep78, Rep68, Rep52 and Rep40.

Rep78 and 68 can specifically bind the hairpin formed by the ITR and cleave it at a specific region (i.e. the terminal resolution site) within the hairpin. In the wild-type virus, they are also necessary for the AAV-specific integration of the AAV genome. Rep 78 and Rep68 are transcribed under control of the p5 promoter in the wild type virus, and the difference between them reflects removal (or not) of an intron by splicing, hence they have different C terminal protein composition.

Rep52 and Rep40 are involved in genome packaging. Rep52 and Rep40 are transcribed under control of the p19 promoter in the wild type virus, and the difference between them reflects removal (or not) of an intron by splicing, hence they have different C terminal protein composition.

All four Rep proteins bind ATP and possess helicase activity. They up-regulate transcription from the p40 promoter, but down-regulate both p5 and p19 promoters.

As used herein, the term "*rep* gene" includes wild-type *rep* genes and derivatives thereof; and artificial *rep* genes which have equivalent functions.

In one embodiment, the *rep* gene encodes functional Rep78, Rep68, Rep52 and Rep40 proteins.

In a preferred example of this embodiment, Rep78 and Rep 68 are translated by ribosomes docking 5' to the Rep78 and Rep68 ATG start codon, thus allowing production of both of these proteins. In this example, the Rep78 and Rep68 open reading frames contain an active p40 promoter that provides the expression of both Rep52 and Rep40.

In some embodiments of the invention, the function of one or more of the p5, p19 and p40 promoters is removed/disabled, for example by codon-changing and/or removal of the TATA box, in order to prevent unwanted initiation of transcription from that promoter.

Preferably, the p5 promoter is non-functional (i.e. it cannot be used to initiate transcription). More preferably, the p5 promoter is replaced with the IRES (thus removing the function of the p5 promoter). This allows Rep78 or Rep68 to be transcribed in the same mRNA as the *cap* genes, but translation of the Rep78 and Rep68 proteins will be under the control of the IRES.

A further advantage of the removal of the p5 promoter is that, in the wild-type virus, the p5 promoter is bound by and is activated by the E2A DNA-binding protein (DBP). Hence the removal of the p5 promoter means that the E2A gene is not required (e.g. in a Helper Plasmid) to produce virus particles.

In one embodiment, the *rep* gene does not have a p5 promoter upstream. In another embodiment, the p5 promoter is not used in AAV packaging.

Preferably, the p19 promoter within the *rep* gene is functional.

In some embodiments, the function of the p40 promoter is removed/disabled within the Rep gene by one or more codon changes.

The cap gene is preferably relocated and its transcription is placed under control of an alternative promoter (e.g. CMV immediate early promoter).

There is a degree of redundancy between the function of the different Rep proteins and hence, in some embodiments of the invention, not all of the Rep proteins are required.

In some embodiments, the *rep* gene only encodes one, two, three or four of Rep78, Rep68, Rep52 and Rep40, preferably one, two or four of Rep78, Rep68, Rep52 and Rep40.

In some embodiments, the *rep* gene does not encode one or more of Rep78, Rep68, Rep52 and Rep40.

In some embodiments, the *rep* gene encodes Rep78 and Rep52, but does not encode Rep68 or Rep40. In this embodiment, the splice donor site remains in the DNA but both the 5' and 3' splice acceptor sites are removed. Hence the intron cannot be removed by splicing and transcription continues through to the terminator sequence for Rep78 and Rep52 (which is common to both). The Rep78 protein is transcribed in the same mRNA as the *cap* gene (hence is driven by the same promoter), and translation of Rep78 is driven by the IRES. Transcription of Rep52 is driven by the p19 promoter; hence it forms a separate mRNA and is translated by 5' m^7G cap-dependent docking at the ribosome. Accordingly, Rep68 and Rep40 cannot be produced in this embodiment.

In other embodiments, the *rep* gene encodes Rep68 and Rep40, but does not encode Rep78 or Rep52. In this embodiment, the intronic sequence between the splice donor and 3' splice acceptor is removed at the DNA level, placing the C terminus of Rep68 and Rep40 in frame with the upstream coding sequence. Hence Rep68 and Rep40 (but not Rep78 and Rep52) are produced. For clarity, Rep68 is transcribed in the same mRNA as the Cap proteins and it is translated under control of the IRES. In contrast, Rep40 is transcribed into a separate mRNA by the p19 promoter and it is translated by 5' m^7G cap docking at the ribosome.

In some embodiments, the *rep* gene encodes Rep78 and Rep68, but does not encode Rep52 or Rep40. This may be achieved by mutating the p19 promoter (e.g. inserting a mutation at the p19 TATA box).

In some embodiments, the *rep* gene encodes Rep52 and Rep40, but does not encode Rep78 or Rep68. This may be achieved by including just the coding sequence from the ATG of Rep52/40.

As used above, the term "encodes" means that the *rep* gene encodes a functional form of that Rep protein. Similarly, the term "does not encode" means that the *rep* gene does not encode a functional form of that Rep protein.

In the absence of sufficient Rep proteins, lower titres (e.g. genome copies) would be observed (which could be determined by qPCR), due to the fact that there is less ITR plasmid to be packaged and that it would not be effectively packaged. The observation might also include an exaggerated empty:full particle ratio; this could be determined by ELISA or optical density measurement.

The wild-type AAV (serotype 2) *rep* gene nucleotide sequence is given in SEQ ID NO: 1. The wild-type AAV (serotype 2) Rep78, Rep68, Rep52 and Rep40 amino acid sequences are given in SEQ ID NOs: 2, 3, 4 and 5, respectively. The wild-type AAV (serotype 2) nucleotide sequence encoding Rep78 is given in SEQ ID NO: 6. The wild-type AAV (serotype 2) nucleotide sequence encoding Rep68 is given in SEQ ID NO: 7. The wild-type AAV (serotype 2) nucleotide sequence encoding Rep52 is given in SEQ ID NO: 8. The wild-type AAV (serotype 2) nucleotide sequence encoding Rep 40 is given in SEQ ID NO: 9.

In one embodiment, the term "*rep* gene" refers to a nucleotide sequence having at least 70%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 1 and which encodes one or more Rep78, Rep68, Rep52 and Rep40 polypeptides.

In another embodiment, the term "*rep* gene" refers to a nucleotide sequence having at least 70%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 6 and which encodes functional Rep78 and/or Rep52 polypeptides (and preferably does not encode functional Rep68 or Rep40 polypeptides).

In another embodiment, the term "*rep* gene" refers to a nucleotide sequence having at least 70%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 7 and which encodes functional Rep68 and/or Rep40 polypeptides (and preferably does not encode functional Rep78 or Rep52 polypeptides).

In another embodiment, the term "*rep* gene" refers to a nucleotide sequence having at least 70%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 8 and which encodes a functional Rep52 polypeptide (and preferably does not encode a functional Rep78 polypeptide).

In another embodiment, the term "*rep* gene" refers to a nucleotide sequence having at least 70%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 9 and which encodes a functional Rep40 polypeptide (and preferably does not encode functional Rep68 polypeptide).

In another embodiment, the term "*rep* gene" refers to a nucleotide sequence having at least 90%, 95%, 99% or 100% sequence identity to a nucleotide sequence which encodes SEQ ID

NO: 2 and which encodes functional Rep78 and/or Rep52 polypeptides (and preferably does not encode functional Rep68 or Rep40 polypeptides).

In another embodiment, the term "*rep* gene" refers to a nucleotide sequence having at least 90%, 95%, 99% or 100% sequence identity to a nucleotide sequence which encodes SEQ ID NO: 3 and which encodes functional Rep68 and/or Rep40 polypeptides (and preferably does not encode functional Rep78 or Rep52 polypeptides).

In another embodiment, the term "*rep* gene" refers to a nucleotide sequence having at least 90%, 95%, 99% or 100% sequence identity to a nucleotide sequence which encodes SEQ ID NO: 4 and which encodes a functional Rep52 polypeptide (and preferably does not encode a functional Rep78 polypeptide).

In another embodiment, the term "*rep* gene" refers to a nucleotide sequence having at least 90%, 95%, 99% or 100% sequence identity to a nucleotide sequence which encodes SEQ ID NO: 5 and which encodes a functional Rep40 polypeptide (and preferably does not encode functional Rep68 polypeptide).

In some embodiments, the nucleic acid molecule of the invention does not encode a functional Rep78 polypeptide. In some embodiments, the nucleic acid molecule of the invention does not encode a functional Rep68 polypeptide. In some embodiments, the nucleic acid molecule of the invention does not encode a functional Rep52 polypeptide. In some embodiments, the nucleic acid molecule of the invention does not encode a functional Rep40 polypeptide.

The nucleic acid molecule also comprises a *cap* gene. As used herein, the term "*cap* gene" refers to a gene that encodes one or more open reading frames (ORFs), wherein each of said ORFs encodes an AAV Cap structural protein, or variant or derivative thereof. These AAV Cap structural proteins (or variants or derivatives thereof) form the AAV capsid.

The three Cap proteins must function to enable the production of an infectious AAV virus particle which is capable of infecting a suitable cell. The three Cap proteins are VP1, VP2 and VP3, which are generally 87kDa, 72kDa and 62kDa in size, respectively. Hence the *cap* gene is one which encodes the three Cap proteins VP1, VP2 and VP3.

In the wild-type AAV, these three proteins are translated from the p40 promoter to form a single mRNA. After this mRNA is synthesized, either a long or a short intron can be excised, resulting in the formation of a 2.3 kb or a 2.6 kb mRNA.

Usually, especially in the presence of adenovirus, the long intron is excised. In this form the first AUG codon, from which the synthesis of VP1 protein starts, is cut out, resulting in a reduced overall level of VP1 protein synthesis. The first AUG codon that remains is the initiation codon for VP3 protein. However, upstream of that codon in the same open reading frame lies an ACG sequence (encoding threonine) which is surrounded by an optimal Kozak context. This contributes to a low level of synthesis of VP2 protein, which is actually VP3 protein with additional N terminal residues, as is VP1.

If the long intron is spliced out, and since in the major splice the ACG codon is a much weaker translation initiation signal, the ratio at which the AAV structural proteins are synthesized *in vivo* is about 1:1:10, which is the same as in the mature virus particle. The unique fragment at the N-terminus of VP1 protein has been shown to possess phospholipase A2 (PLA2) activity, which is probably required for the releasing of AAV particles from late endosomes.

The AAV capsid is composed of 60 capsid protein subunits (VP1, VP2, and VP3) that are arranged in an icosahedral symmetry in a ratio of 1:1:10, with an estimated size of 3.9 MDa.

As used herein, the term "*cap* gene" includes wild-type *cap* genes and derivatives thereof, and artificial *cap* genes which have equivalent functions. The AAV (serotype 2) *cap* gene nucleotide sequence and Cap polypeptide sequences are given in SEQ ID NOs: 10 and 11, respectively.

As used herein, the term "*cap* gene" refers preferably to a nucleotide sequence having the sequence given in SEQ ID NO: 10 or a nucleotide sequence encoding SEQ ID NO: 11; or a nucleotide sequence having at least 70%, 80%, 85%, 90%, 95% or 99% sequence identity to SEQ ID NO: 10 or at least 80%, 90%, 95% or 99% nucleotide sequence identity to a nucleotide sequence encoding SEQ ID NO: 11, and which encodes VP1, VP2 and VP3 polypeptides.

The *rep* and *cap* genes are preferably viral genes or derived from viral genes. More preferably, they are AAV genes or derived from AAV genes. In some embodiments, the AAV is an Adeno-associated dependoparvovirus A. In other embodiments, the AAV is an Adeno-associated dependoparvovirus B.

11 different AAV serotypes are known. All of the known serotypes can infect cells from multiple diverse tissue types. Tissue specificity is determined by the capsid serotype. The AAV may be from serotype 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or 11. Preferably, the AAV is serotype 1, 2, 5, 6, 7 or 8. Most preferably, the AAV serotype is 2 (i.e. AAV2).

The *rep* and *cap* genes (and each of the protein-encoding ORFs therein) may be from one or more different viruses (e.g. 2, 3 or 4 different viruses). For example, the *rep* gene may be from AAV2, whilst the *cap* gene may be from AAV5. It is recognised by those in the art that the *rep* and *cap* genes of AAV vary by clade and isolate. The sequences of these genes from all such clades and isolates are encompassed herein, as well as derivatives thereof.

The *cap* gene and *rep* gene are present in the nucleic acid in this 5'→ 3' order. However, since Rep52 and/or Rep40 may be transcribed from their own p19 promoter, the position of the coding sequence which encodes Rep52 and/or Rep40 may be varied. For example, the coding sequence which encodes Rep52 and/or Rep40 may be placed upstream or downstream of the *cap* genes and *rep* genes which encode Rep78/68; or indeed on the reverse strand of the nucleic acid of the invention or on a different nucleic acid.

The *cap* and *rep* genes are both operably-associated with the same promoter. The promoter is preferably 5' (i.e. upstream) of the *cap* and *rep* genes. In some embodiments, the promoter is a constitutive promoter. In other embodiments, the promoter is inducible or repressible.

Examples of constitutive promoters include the CMV, SV40, PGK (human or mouse), HSV TK, SFFV, Ubiquitin, Elongation Factor Alpha, CHEF-1, FerH, Grp78, RSV, Adenovirus E1A, CAG or CMV-Beta-Globin promoter, or a promoter derived therefrom. Preferably, the promoter is the cytomegalovirus immediate early (CMV) promoter, or a promoter which is derived therefrom, or a promoter of equal or increased strength compared to the CMV promoter in human cells and human cell lines (e.g. HEK-293 cells).

In some embodiments, the promoter is inducible or repressible by the inclusion of an inducible or repressible regulatory (promoter) element. For example, the promoter may one which is inducible with doxycycline, tetracycline, IPTG or lactose.

Preferably, the inducible promoter element comprises a plurality of Tet operator sequences to which the Tet repressor protein (TetR) is capable of binding. In the bound state, tight suppression of transcription is obtained. However, in the presence of doxycycline (or less preferably, tetracycline), suppression is alleviated, thus allowing the promoter to gain full transcriptional activity. Such an inducible promoter element is preferably placed downstream of another promoter, e.g. the CMV promoter.

The TetR binding site may have a wild-type sequence, many of which are known in the art. Preferably, the TetR binding site has been subject to one or more improvements by incorporating minor sequence changes. A preferred version that may be used in an embodiment of the invention has the sequence tccctatcagtgatagaga (SEQ ID NO: 12).

Alternative versions of the repressor element that bind the TetR protein or derivatives of the TetR protein may also be used in an embodiment of the invention provided that the TetR repressor protein binds to the TetR binding sequence variant used. Some repressor/binding site variants will have higher than wild-type affinity for each other; these are preferable in an embodiment of the invention.

The TetR gene will generally be integrated into the chromosome of a human (host) cell. The gene may or may not be integrated adjacent to, or in conjunction with, the *cap* or *rep* genes. In some embodiments, the *TetR* gene is co-expressed with the *cap* gene or *rep* gene.

In one embodiment of the invention, the nucleotide sequence of the TetR protein is as given in SEQ ID NO: 13 or a nucleotide sequence having at least 80%, more preferably at least 85%, 90% or 95% sequence identity thereto and which codes for a TetR protein.

In another embodiment of the invention, the amino acid sequence of the TetR protein is as given in SEQ ID NO: 14 or an amino acid sequence having at least 80%, more preferably at least 85%, 90% or 95% sequence identity thereto and which encodes a TetR protein.

Preferably, the promoter which is operably-associated with the *cap* and rep genes is the CMV immediate early promoter or a derivative thereof. In some particularly-preferred embodiments, the promoter is a promoter as defined in WO2017/149292 (more preferably, a promoter as defined therein as “p565”). Preferably, the promoter which is operably-associated with the *cap* and *rep* genes is not an AAV promoter, e.g. it is not an AAV p5, p19 or p40 promoter.

Translation of the *cap* gene is preferably initiated from the standard 5' m^7G-cap at the 5' end of the mRNA.

The *rep* gene is also operably-associated with an Internal Ribosome Entry Site (IRES). The IRES regulates the translation of the *rep* mRNA. IRESs are distinct regions of nucleic acid molecules that are able to recruit eukaryotic ribosomes to the mRNA in a process which is known as cap-independent translation. IRESs are commonly located in the 5'-UTRs of RNA viruses. They facilitate translation of the viral RNAs in a cap-independent manner.

Examples of viral IRESs include Picornavirus IRES (Encephalomyocarditis virus, EMCV IRES), Aphthovirus IRES (Foot-and-mouth disease virus, FMDV IRES), Kaposi's sarcoma-associated herpes virus IRES, Hepatitis A IRES, Hepatitis C IRES, Pestivirus IRES, Cripavirus internal ribosome entry site (IRES), *Rhopalosiphum padi* virus internal ribosome entry site (IRES) and 5'-Leader IRES and intercistronic IRES in the 1.8-kb family of immediate early transcripts (IRES)1.

The invention also encompasses non-natural derivatives of the above IRESs which retain the capacity to recruit eukaryotic ribosomes to the mRNA. In some preferred embodiments, the IRES is an encephalomyocarditis virus (EMCV) IRES. In one embodiment of the invention, the nucleotide sequence of the EMCV IRES is as given in SEQ ID NO: 15 or a nucleotide sequence having at least 80%, more preferably at least 85%, 90% or 95% sequence identity thereto and which encodes an IRES.

In other embodiments, the IRES is a Foot-and-mouth disease virus (FMDV) IRES. In one embodiment of the invention, the nucleotide sequence of the FMDV IRES is as given in SEQ ID NO: 16 or a nucleotide sequence having at least 80%, more preferably at least 85%, 90% or 95% sequence identity thereto and which encodes an IRES.

The *rep* gene is operably-associated with the IRES. Preferably, the IRES is located downstream of the *cap* gene and upstream of the translation start site for Rep 78/68.

The production of stable cell lines in mammalian culture typically requires a method of selection to promote the growth of cells containing any exogenously-added DNA.

Preferably, the nucleic acid molecules of the invention additionally comprise a selection gene or an antibiotic resistance gene. To this end, a range of genes are known that provide resistance to specific compounds when the DNA encoding them is inserted into a mammalian cell genome.

Preferably, the selection gene is puromycin N-acetyl-transferase (Puro), hygromycin phosphotransferase (Hygro), blasticidin s deaminase (Blast), Neomycin phosphotransferase (Neo), glutathione S-transferase (GS), zeocin resistance gene (Sh ble) or dihydrofolate reductase (DHFR). Each of these genes provides resistance to a small molecule known to be toxic to mammalian cells, or in the case of GS provides a method for cells to generate glutathione in the absence of glutathione in the growth media.

In a preferred embodiment of the invention, the resistance gene is Puro. This gene is particularly effective because many of the cell lines used in common tissue culture are not resistant to Puro; this cannot be said for Neo where many, particularly HEK 293 derivatives, are already Neo resistant due to previous genetic manipulations by researchers (e.g. HEK 293T cells). Puro selection also has the advantage of being toxic over a short time window (<72 hours), and hence it allows variables to be tested rapidly and cells that do not harbour the exogenous DNA to be inserted into the genome are rapidly removed from the culture systems. This cannot be said of some other selection methods such as Hygro, where toxicity is much slower onset.

The development of stable cell lines using selection genes (e.g. Puro) requires that the resistance gene must be expressed in the cells. This can be achieved through a variety of methods including, but not limited to, internal ribosome entry sites (IRES), 2A cleavage systems, alternative splicing, and dedicated promoters.

In a preferred embodiment of the invention, the selection gene will be expressed from a dedicated promoter. This promoter will preferably transcribe in human cells at lower levels than the dedicated promoters driving the *rep* or *cap* genes.

Each of the genes in the nucleic acid molecule which encode a polypeptide or RNA will preferably be operably-associated with one or more regulatory elements. This ensures that the polypeptide or RNA is expressed at the desired level and at the desired time. In this context, the term “regulatory elements” includes one or more of an enhancer, promoter, intron, polyA, insulator or terminator.

The genes used in the AAV plasmids or vectors disclosed herein are preferably separated by polyA signals and/or insulators in an effort to keep transcriptional read-through to other genes to a minimum.

While some advantages may be obtained by using copies of the same regulatory element (e.g. promoter sequence) with more than one polypeptide or RNA-encoding nucleotide sequence (in terms of their co-ordinated expression), in the context of this invention, it is highly desirable to use different regulatory elements with each polypeptide or RNA-encoding nucleotide sequence.

Preferably, therefore, the *rep* and *cap* genes are operably-associated with different regulatory elements, e.g. different promoter, different intron, different polyA, different insulator and/or different terminator sequences. More preferably, the degree of nucleotide sequence identity between the *rep* promoter and the *cap* promoter is less than 95% or less than 90%, more preferably less than 85%, 80%, 70% or 60%. More preferably, the degree of nucleotide sequence identity between the *rep* terminator and the *cap* terminator is less than 95% or less than 90%, more preferably less than 85%, 80%, 70% or 60%. In this way, the risk of homologous recombination between these regulatory elements is reduced.

The nucleic acid molecule of the invention will, most embodiments, be a plasmid or vector which is useful in the production of AAVs. In most embodiments, therefore, the nucleic acid molecule of the invention (or the vector or plasmid comprising it) will not comprise inverted terminal repeats (ITRs).

In some embodiments, the nucleic acid molecule of the invention (or the vector or plasmid comprising it) will not comprise one or more genes selected from Adenovirus E1A, E1B, E4, E2A or VA. In some preferred embodiments, the nucleic acid molecule of the invention (or the vector or plasmid or plasmid system comprising it) does not comprise the Adenovirus E2A gene. As used herein, the term “E2A” or “E2A gene” refers to a viral E2A gene or a variant or derivative thereof. Preferably, the E2A gene is from or derived from a human adenovirus, e.g. Ad5. In one embodiment of the invention, the nucleotide sequence of the Adenovirus E2A gene is as given in SEQ ID NO: 17 or a nucleotide sequence having at least 80%, more preferably at least 85%, 90% or 95% sequence identity thereto and which encodes a DNA-binding protein which aids elongation of viral DNA replication.

In another embodiment, there is provided a plasmid or vector comprising a nucleic acid molecule of the invention.

Examples of preferred embodiments of the invention include nucleic acid molecules comprising the following elements in this order:

CMV promoter - AAV2 *cap* gene - FMDV IRES - *rep* gene

p565 promoter - AAV2 *cap* gene - EMCV IRES - *rep* gene

CMV promoter - AAV2 *cap* gene - EMCV IRES - *rep* gene

In some preferred embodiments, the "*rep* gene" refers to a gene which encodes Rep78, Rep52, Rep68 and Rep40 polypeptides. In other preferred embodiments, the term "*rep* gene" refers to a gene which encodes Rep78 and Rep52 polypeptides (but preferably does not encode functional Rep68 or Rep40 polypeptides).

There are many established algorithms available to align two amino acid or nucleic acid sequences. Typically, one sequence acts as a reference sequence, to which test sequences may be compared. The sequence comparison algorithm calculates the percentage sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters. Alignment of amino acid or nucleic acid sequences for comparison may be conducted, for example, by computer-implemented algorithms (e.g. GAP, BESTFIT, FASTA or TFASTA), or BLAST and BLAST 2.0 algorithms.

Percentage amino acid sequence identities and nucleotide sequence identities may be obtained using the BLAST methods of alignment (Altschul *et al.* (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402; and http://www.ncbi.nlm.nih.gov/BLAST). Preferably the standard or default alignment parameters are used.

Standard protein-protein BLAST (blastp) may be used for finding similar sequences in protein databases. Like other BLAST programs, blastp is designed to find local regions of similarity. When sequence similarity spans the whole sequence, blastp will also report a global alignment, which is the preferred result for protein identification purposes. Preferably the standard or default alignment parameters are used. In some instances, the "low complexity filter" may be taken off.

BLAST protein searches may also be performed with the BLASTX program, score=50, wordlength=3. To obtain gapped alignments for comparison purposes, Gapped BLAST (in BLAST 2.0) can be utilized as described in Altschul *et al.* (1997) Nucleic Acids Res. 25: 3389. Alternatively, PSI-BLAST (in BLAST 2.0) can be used to perform an iterated search that detects distant relationships between molecules. (See Altschul *et al.* (1997) supra). When utilizing BLAST, Gapped BLAST, PSI-BLAST, the default parameters of the respective programs may be used.

With regard to nucleotide sequence comparisons, MEGABLAST, discontiguous-megablast, and blastn may be used to accomplish this goal. Preferably the standard or default alignment parameters are used. MEGABLAST is specifically designed to efficiently find long alignments between very similar sequences. Discontiguous MEGABLAST may be used to find nucleotide sequences which are similar, but not identical, to the nucleic acids of the invention.

The BLAST nucleotide algorithm finds similar sequences by breaking the query into short subsequences called words. The program identifies the exact matches to the query words first (word hits). The BLAST program then extends these word hits in multiple steps to generate the final gapped alignments. In some embodiments, the BLAST nucleotide searches can be performed with the BLASTN program, score=100, wordlength=12.

One of the important parameters governing the sensitivity of BLAST searches is the word size. The most important reason that blastn is more sensitive than MEGABLAST is that it uses a shorter default word size (11). Because of this, blastn is better than MEGABLAST at finding alignments to related nucleotide sequences from other organisms. The word size is adjustable in blastn and can be reduced from the default value to a minimum of 7 to increase search sensitivity.

A more sensitive search can be achieved by using the newly-introduced discontiguous megablast page (www.ncbi.nlm.nih.gov/Web/Newsltr/FallWinter02/blastlab.html). This page uses an algorithm which is similar to that reported by Ma *et al.* (Bioinformatics. 2002 Mar; 18(3): 440-5). Rather than requiring exact word matches as seeds for alignment extension, discontiguous megablast uses non-contiguous word within a longer window of template. In coding mode, the third base wobbling is taken into consideration by focusing on finding matches at the first and second codon positions while ignoring the mismatches in the third position. Searching in discontiguous MEGABLAST using the same word size is more sensitive and

efficient than standard blastn using the same word size. Parameters unique for discontiguous megablast are: word size: 11 or 12; template: 16, 18, or 21; template type: coding (0), non-coding (1), or both (2).

In some embodiments, the BLASTP 2.5.0+ algorithm may be used (such as that available from the NCBI) using the default parameters. In other embodiments, a BLAST Global Alignment program may be used (such as that available from the NCBI) using a Needleman-Wunsch alignment of two protein sequences with the gap costs: Existence 11 and Extension 1.

One method for the production of recombinant AAVs is based on the transient transfection of all elements that are required for AAV production into host cells, such as HEK293 cells. This generally involves the co-transfection of AAV production cells with 3 plasmids:

(a) an AAV ITR-containing plasmid, carrying the gene of interest;

(b) a plasmid that carries the AAV *rep-cap* genes; and

(c) a plasmid that provides the necessary helper genes isolated from adenovirus.

In some instances, the helper genes are stably integrated into (and expressible from) the host cell genome; therefore plasmid (c) is not needed.

The invention therefore provides a kit comprising:

(a) a plasmid or vector comprising a nucleic acid molecule of the invention, together with one or more of the following -

(b) an AAV Transfer Plasmid comprising a transgene flanked by ITRs;

(c) a Helper Plasmid comprising one or more genes selected from Adenovirus E1A, E1B, E4 and VA.

In some embodiments of the invention, the Helper Plasmid additionally comprises an E2A gene. In other embodiments, the Helper Plasmid does not comprise an E2A gene. In the latter case, the omission of the E2A gene reduces considerably the amount of DNA which is needed in the Helper Plasmid.

The invention also provides a kit comprising:

(a) a plasmid or vector comprising a nucleic acid molecule of the invention, together with one or more of the following -

(b) an AAV Transfer Plasmid comprising a transgene flanked by ITRs;

(c) a mammalian host cell (e.g. HEK293) comprising one or more viral genes selected from E1A, E1B, E4 and VA expressible from the host cell genome.

In some embodiments of the invention, the mammalian host cell additionally comprises an E2A gene expressible from the host cell genome. In other embodiments, the mammalian host cell does not comprise an Adenovirus E2A gene.

The kit may also contain materials for purification of the AAV particles such as those involved in the density banding and purification of viral particles, e.g. one or more of centrifuge tubes, Iodixanol, dialysis buffers and dialysis cassettes.

The invention also provides a mammalian cell comprising a nucleic acid molecule, plasmid or vector of the invention. The nucleic acid molecule of the invention may be stably integrated into the nuclear genome of the mammalian cell or present within a vector or plasmid (e.g. episome) within the cell.

Preferably, the nucleic acid molecule of the invention is stably integrated into the nuclear genome of the mammalian cell (and wherein the *rep* and *cap* genes are expressible therefrom).

The cells may be isolated cells, e.g. they are not situated in a living animal or mammal. Examples of mammalian cells include those from any organ or tissue from humans, mice, rats, hamsters, monkeys, rabbits, donkeys, horses, sheep, cows and apes. Preferably, the cells are human cells. The cells may be primary or immortalised cells.

Preferred cells include HEK-293, HEK 293T, HEK-293E, HEK-293 FT, HEK-293S, HEK-293SG, HEK-293 FTM, HEK-293SGGD, HEK-293A, MDCK, C127, A549, HeLa, CHO, mouse myeloma, PerC6, 911 and Vero cell lines. HEK-293 cells have been modified to contain the E1A and E1B proteins and this obviates the need for these proteins to be supplied on a Helper Plasmid. Similarly, PerC6 and 911 cells contain a similar modification and can also be used. Most preferably, the human cells are HEK293, HEK293T, HEK293A, PerC6 or 911. Other preferred cells include CHO and VERO cells.

Preferably, the cells of the invention are capable of inducibly expressing the *rep* and *cap* genes.

The invention also provides an AAV packaging cell, preferably a mammalian cell, more preferably a human cell), comprising (a) a nucleic acid molecule of the invention, and optionally one or both of (b) an AAV Transfer Plasmid comprising a transgene flanked by ITRs, and (c) a Helper Plasmid comprising one or more genes selected from E1A, E1B, E4 and VA. In some embodiments of the invention, the Helper Plasmid additionally comprises an E2A gene. In other embodiments, the Helper Plasmid does not comprise an E2A gene. In the latter case, the omission of the E2A gene reduces considerably the amount of DNA which is needed in the Helper Plasmid.

The nucleic acid molecules, plasmids and vectors of the invention may be made by any suitable technique. Recombinant methods for the production of the nucleic acid molecules and packaging cells of the invention are well known in the art (e.g. "Molecular Cloning: A Laboratory Manual" (Fourth Edition), Green, MR and Sambrook, J., (updated 2014)).

The expression of the *rep* and *cap* genes from the nucleic acid molecules of the invention may be assayed in any suitable assay, e.g. by assaying for the number of genome copies per ml by qPCR (as described the Examples herein).

In a further embodiment, there is provided a process for producing an AAV packaging cell, the process comprising the steps:

(a) stably integrating a nucleic acid molecule of the invention into a mammalian cell, thereby producing a mammalian cell that expresses viral *rep* and *cap* genes.

The invention also provides the use of an AAV packaging cell of the invention in the production of an AAV particle.

The invention also provides a process for producing AAVs, the process comprising the steps:

(a) introducing a Transfer Plasmid comprising 5'- and 3'-AAV ITRs flanking a transgene into an AAV packaging cell, the AAV packaging cell comprising a nucleic acid molecule of the invention and sufficient helper genes (preferably selected from one or more of E1A, E1B, E4 and VA) for packaging the Transfer Plasmid, the helper genes either being present in an episomal Helper Plasmid within the cell or being integrated into the packaging cell genome;

(b) culturing the cell under conditions such that AAVs are assembled and secreted by the cell; and

(c) harvesting packaged AAVs from the supernatant.

In some embodiments of the invention, the helper genes additionally include an E2A gene. In other embodiments, the helper genes do not include an E2A gene.

Preferably, the harvested AAVs are subsequently purified.

As used herein, the term "introducing" one or more plasmids or vectors into the cell includes transformation, and any form of electroporation, conjugation, infection, transduction or transfection, *inter alia*.

In some preferred embodiments, the transgene encodes a CRISPR enzyme (e.g. Cas9, Cpf1) or a CRISPR sgRNA.

Processes for such introduction are well known in the art (e.g. Proc. Natl. Acad. Sci. USA. 1995 Aug 1;92 (16):7297-301).

The disclosure of each reference set forth herein is specifically incorporated herein by reference in its entirety.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 shows the organisation of the Rep and Cap protein genes in the wild-type AAV genome.

Figures 2A, 2B and 2C show three embodiments of the nucleic acid molecule of the invention. In Figure 2B, "OXGP3" refers to a CMV promoter variant with two Tet operator sites.

Figures 3-4 show the results of assays for the number of copies of the AAV genome per ml which were produced in cells transfected with various *rep-cap* plasmids. OxG = a standard RepCap configuration as found in the wild-type virus, including the p5 promoter which was placed distally; CMV-CMV = a configuration in which both Rep and Cap sequences were placed under CMV promoters, in the 5'-3' order CMV-Cap-CMV-Rep;CMV-PGK = a configuration in which Rep and Cap sequences were placed under PGK and CMV promoters respectively, in the 5'-3' order CMV-Cap-PGK-Rep; CMV-EMCV = a configuration in which Cap sequences were placed under the CMV promoter and Rep sequence placed under the control

of the IRES EMCV, in the 5'-3' order CMV-Cap-EMCV-Rep. In Figure 4, the cell lysate containing the virus was diluted 500-fold and quantified using qPCR. This demonstrates the physical titre.

Figure 5 shows the results following flow cytometry analysis of HEK293T cells 72 hours post infection with AAV particles. Data is given as a percentage of the GFP positive cells in P1. P1 corresponds to the viable cells in the sample.

Figure 6 shows the transducing units per millilitre of virus sample infected, as calculated by the results from Figure 5 and the number of cells infected. This demonstrates the infectious titre.

Figure 7 shows the results of assays for the number of copies of the AAV genome per ml which were produced in cells transfected with various *rep-cap* plasmids. For details of the plasmids, see the above for Figures 3-4. Clontech refers to the 3-plasmid system supplied by Clontech (pAAV-CMV-EGFP; pHelper; pRepCap-miR342).

Figure 8 shows the titres (GC, genome copies) obtained from virus produced from a) a 3-plasmid AAV system of the invention; b) the system of a) wherein pSF-helper plasmid is replaced with a plasmid containing CMV-E4orf6 (coding sequence) only; c) the system of a) wherein the pSF-helper plasmid is replaced by pSF-E4orf6-VAI; and d) the system of a) where the pSF-helper plasmid is removed and replaced with stuffer DNA (control).

EXAMPLES

The present invention is further illustrated by the following Examples, in which parts and percentages are by weight and degrees are Celsius, unless otherwise stated. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions. Thus, various modifications of the invention in addition to those shown and described herein will be apparent to those skilled in the art from the foregoing description. Such modifications are also intended to fall within the scope of the appended claims.

Example 1: Production of AAV *cap-rep* plasmids

The following plasmids were produced having the genetic elements as shown in Figures 2A, 2B, and 2C:

pSF-CMV-AAV2Cap-FMDV-AAV2Rep

pSF-CMV-AAV2Cap-EMCV-AAV2Rep

pSF-p565_2xTetO-AAV2Cap-EMCV-AAV2Rep

The OxGP3 promoter is the p565 promoter as defined in WO2017/149292. FMDV is the Foot and Mouth Disease Virus IRES. CMV is the CytoMegaloVirus promoter. AAV2Cap is the Adeno-Associated Virus 2 *cap* gene. Rep is the Adeno-Associated Virus *rep* gene encoding Rep78 and Rep52 only.

pSF-AAV-CMV-EGFP

This plasmid encodes an EGFP protein driven by the CMV promoter, flanked by two AAV2 ITR sequences to allow the packaging of the ITR-CMV-EGFP-ITR sequence into the AAV capsid shell.

pSF-Helper

This plasmid contains Adenovirus 5 sequences E2A, E4orf6 and VAI RNA, to provide the helper functions required for AAV production in HEK293 cells.

pSF-RepCap

This plasmid contains the RepCap sequences in the wild-type configuration, with the p5 promoter removed and placed distally to lower the overall expression of Rep78/68.

pSF-CMV-Cap-CMV-Rep78/52

This plasmid contains the Cap sequence driven by a CMV and the Rep78/52 sequence separately driven by a CMV promoter. This gives equally strong expression of the two coding sequences.

pSF-CMV-Cap-PGK-Rep78/52

This plasmid contains the Cap sequence driven by a CMV and the Rep78/52 sequence separately driven by a PGK promoter, which gives lower expression than the CMV promoter.

pSF-CMV-Cap-EMCV-Rep78/52

This plasmid contains the Cap sequence driven by a CMV promoter, and the Rep78/52 protein produced from the IRES EMCV. The EMCV gives much lower expression levels than the CMV promoter.

Example 2: Assaying for genome copies (GC)

Plasmid vectors pSF-AAV-CMV-EGFP, pSF-Helper and one of: pSF-RepCap; pSF-CMV-Cap-CMV-Rep78/52; pSF-CMV-Cap-PGK-Rep78/52 or pSF-CMV-Cap-EMCV-Rep78/52 were transfected in a 1:1:1 molar ratio into >80% confluent HEK293T cells in a 6-well plate, to a total of 2.5 µg of DNA per well. Transfection reagent Lipofectamine 2000 was used in a 1:2.4 ratio of total DNA mass to Lipofectamine. Entire well contents were harvested at 48 hours for analysis by both flow cytometry and qPCR. Data is presented as both the Transducing Units (TU) per mL of lysate (Figure 6) and genome copies per mL of lysate (Figures 3-4).

By using the plasmid pSF-CMV-Cap-EMCV-Rep78/52, both the infectious and physical titre was improved compared to the wild-type configuration used in the OxG positive control.

Figure 5 shows the percentage of viable cells which are GFP positive after 72 hours incubation with AAV. The viral solution was diluted to 1 in 500, 1 in 2500 and 1 in 12500. The dilution which gives between 5 and 25% GFP positive cells was used to calculate the transducing units per millilitre of viral solution.

Example 3:

Plasmid vectors pSF-AAV-CMV-EGFP, pSF-Helper and one of: pSF-RepCap; pSF-CMV-Cap-EMCV-Rep78/52 or pSF-CMV-Cap-FMDV-Rep78/52 were transfected in a 1:1:1 molar ratio into 70-80% confluent HEK293T cells in a 6-well plate, to a total of 2.5 µg of DNA per well. This was run alongside the Clontech 3-plasmid system (www.clontech.com/GB/Products/Viral_Transduction/AAV_Vector_Systems/Helper_Free_Expression_System?sitex=10030:22372:US), also transfected as 1:1:1 molar ratio to total DNA mass 2.5 µg. Transfection reagent Lipofectamine 2000 used in a 1:2.4 ratio of total DNA mass to Lipofectamine. Entire well contents were harvested at 48 hours for analysis by qPCR. Data is presented (Figure 7) as genome copies per mL of lysate, with error bars representing the standard error of the mean.

This experiment shows that the pSF-CMV-Cap-EMCV-Rep78/52 reliably outperforms the standard wild-type configuration used in both pSF-RepCap and the Clontech pRepCap-miR342. It also demonstrates that using an alternative IRES, FMDV gives an increase in viral titre, compared to the wild-type configurations.

Example 4: Enhanced titres are obtained without using E2A

The effect of the presence or absence of E2A from an AAV production system of the invention was assessed using sets of plasmids which contained/did not contain the Ad5 E2A gene.

All experiments included the following plasmids:

(i) pSF-AAV-CMV-EGFP, as defined in Example 1.
(ii) pSF-CMV-Cap-EMCV-Rep. This plasmid contains the Cap sequence driven by a CMV promoter, and the Rep protein produced from the EMCV IRES.

In addition to the above (i) and (ii), the following plasmids were included in the following experiments:

a) pSF-Helper (this contains Ad5 regions E2A, E4orf6 and VA RNA I);
b) pSF-nano-CMV-E4orf6 (this contains the coding sequence for E4orf6 protein only);
c) pSF-E4orf6-VA I (this contains the full E4orf6 region and full VA RNA I sequence); and
d) OG10 (this is an empty pSF-CMV-Kan).

The results are shown in Figure 8, wherein the results of experiments a), b), c) and d) are shown (as genome copies/ml) labelled as “OxG Pro”, “E4Orf6/Pro”, “no E2A” and “no Ad5 Help”, respectively.

The results show that higher titres of virus may be obtained by using a Cap/Rep plasmid of the invention without using E2A in the AAV production system.

SEQUENCES

SEQ ID NO: 1 - Rep nucleotide sequence (AAV serotype 2)

atgccggggttttacgagattgtgattaaggtccccagcgaccttgacgagcatctgcccggcatttctg
acagctttgtgaactgggtggccgagaaggaatgggagttgccgccagattctgacatggatctgaatct
gattgagcaggcacccctgaccgtggccgagaagctgcagcgcgactttctgacggaatggcgccgtgtg
agtaaggccccggaggccctttttctttgtgcaatttgagaagggagagagctacttccacatgcacgtgc
tcgtggaaaccaccggggtgaaatccatggttttgggacgtttcctgagtcagattcgcgaaaaactgat
tcagagaatttaccgcgggatcgagccgactttgccaaactggttcgcggtcacaaagaccagaaatggc
gccggaggcgggaacaaggtggtggatgagtgctacatccccaattacttgctccccaaaacccagcctg
agctccagtgggcgtggactaatatggaacagtatttaagcgcctgtttgaatctcacggagcgtaaacg
gttggtggcgcagcatctgacgcacgtgtcgcagacgcaggagcagaacaaagagaatcagaatcccaat
tctgatgcgccggtgatcagatcaaaaacttcagccaggtacatggagctggtcgggtggctcgtggaca
aggggattacctcggagaagcagtggatccaggaggaccaggcctcatacatctccttcaatgcggcctc
caactcgcggtcccaaatcaaggctgccttggacaatgcgggaaagattatgagcctgactaaaaccgcc
cccgactacctggtgggccagcagcccgtggaggacatttccagcaatcggatttataaaattttggaac
taaacgggtacgatccccaatatgcggcttccgtctttctgggatgggccacgaaaaagttcggcaagag
gaacaccatctggctgtttgggcctgcaactaccgggaagaccaacatcgcggaggccatagcccacact
gtgcccttctacgggtgcgtaaactggaccaatgagaactttcccttcaacgactgtgtcgacaagatgg
tgatctggtgggaggaggggaagatgaccgccaaggtcgtggagtcggccaaagccattctcggaggaag
caaggtgcgcgtggaccagaaatgcaagtcctcggcccagatagacccgactcccgtgatcgtcacctcc
aacaccaacatgtgcgccgtgattgacgggaactcaacgaccttcgaacaccagcagccgttgcaagacc
ggatgttcaaatttgaactcacccgccgtctggatcatgactttgggaaggtcaccaagcaggaagtcaa
agacttttccggtgggcaaaggatcacgtggttgaggtggagcatgaattctacgtcaaaaagggtgga
gccaagaaaagacccgccccccagtgacgcagatataagtgagcccaaacgggtgcgcgagtcagttgcgc
agccatcgacgtcagacgcggaagcttcgatcaactacgcagacaggtaccaaaacaaatgttctcgtca
cgtgggcatgaatctgatgctgtttccctgcagacaatgcgagagaatgaatcagaattcaaatatctgc
ttcactcacggacagaaagactgtttagagtgctttcccgtgtcagaatctcaacccgtttctgtcgtca
aaaaggcgtatcagaaactgtgctacattcatcatatcatgggaaaggtgccagacgcttgcactgcctg
cgatctggtcaatgtggatttggatgactgcatctttgaacaaTAG

SEQ ID NO: 2 - Rep78 amino acid sequence (AAV serotype 2)

MPGFYEIVIKVPSDLDGHLPGISDSFVNWVAEKEWELPPDSDMDLNLIEQAPLTVAEKLQRDFLTEWRRV
SKAPEALFFVQFEKGESYFHMHVLVETTGVKSMVLGRFLSQIREKLIQRIYRGIEPTLPNWFAVTKTRNG
AGGGNKVVDECYIPNYLLPKTQPELQWAWTNMEQYLSACLNLTERKRLVAQHLTHVSQTQEQNKENQNPN

SDAPVIRSKTSARYMELVGWLVDKGITSEKQWIQEDQASYISFNAASNSRSQIKAALDNAGKIMSLTKTA
PDYLVGQQPVEDISSNRIYKILELNGYDPQYAASVFLGWATKKFGKRNTIWLFGPATTGKTNIAEAIAHT
VPFYGCVNWTNENFPFNDCVDKMVIWWEEGKMTAKVVESAKAILGGSKVRVDQKCKSSAQIDPTPVIVTS
NTNMCAVIDGNSTTFEHQQPLQDRMFKFELTRRLDHDFGKVTKQEVKDFFRWAKDHVVEVEHEFYVKKGG
AKKRPAPSDADISEPKRVRESVAQPSTSDAEASINYADRYQNKCSRHVGMNLMLFPCRQCERMNQNSNIC
FTHGQKDCLECFPVSESQPVSVVKKAYQKLCYIHHIMGKVPDACTACDLVNVDLDDCIFEQ*

SEQ ID NO: 3 - Rep68 amino acid sequence (AAV serotype 2)

MPGFYEIVIKVPSDLDEHLPGISDSFVNWVAEKEWELPPDSDMDLNLIEQAPLTVAEKLQRDFLTEWRRV
SKAPEALFFVQFEKGESYFHMHVLVETTGVKSMVLGRFLSQIREKLIQRIYRGIEPTLPNWFAVTKTRNG
AGGGNKVVDECYIPNYLLPKTQPELQWAWTNMEQYLSACLNLTERKRLVAQHLTHVSQTQEQNKENQNPN
SDAPVIRSKTSARYMELVGWLVDKGITSEKQWIQEDQASYISFNAASNSRSQIKAALDNAGKIMSLTKTA
PDYLVGQQPVEDISSNRIYKILELNGYDPQYAASVFLGWATKKFGKRNTIWLFGPATTGKTNIAEAIAHT
VPFYGCVNWTNENFPFNDCVDKMVIWWEEGKMTAKVVESAKAILGGSKVRVDQKCKSSAQIDPTPVIVTS
NTNMCAVIDGNSTTFEHQQPLQDRMFKFELTRRLDHDFGKVTKQEVKDFFRWAKDHVVEVEHEFYVKKGG
AKKRPAPSDADISEPKRVRESVAQPSTSDAEASINYAD*

SEQ ID NO: 4 - Rep52 amino acid sequence (AAV serotype 2)

MELVGWLVDKGITSEKQWIQEDQASYISFNAASNSRSQIKAALDNAGKIMSLTKTAPDYLVGQQPVEDIS
SNRIYKILELNGYDPQYAASVFLGWATKKFGKRNTIWLFGPATTGKTNIAEAIAHTVPFYGCVNWTNENF
PFNDCVDKMVIWWEEGKMTAKVVESAKAILGGSKVRVDQKCKSSAQIDPTPVIVTSNTNMCAVIDGNSTT
FEHQQPLQDRMFKFELTRRLDHDFGKVTKQEVKDFFRWAKDHVVEVEHEFYVKKGGAKKRPAPSDADISE
PKRVRESVAQPSTSDAEASINYADRYQNKCSRHVGMNLMLFPCRQCERMNQNSNICFTHGQKDCLECFPV
SESQPVSVVKKAYQKLCYIHHIMGKVPDACTACDLVNVDLDDCIFEQ*

SEQ ID NO: 5 - Rep40 amino acid sequence (AAV serotype 2)

MELVGWLVDKGITSEKQWIQEDQASYISFNAASNSRSQIKAALDNAGKIMSLTKTAPDYLVGQQPVEDIS
SNRIYKILELNGYDPQYAASVFLGWATKKFGKRNTIWLFGPATTGKTNIAEAIAHTVPFYGCVNWTNENF
PFNDCVDKMVIWWEEGKMTAKVVESAKAILGGSKVRVDQKCKSSAQIDPTPVIVTSNTNMCAVIDGNSTT
FEHQQPLQDRMFKFELTRRLDHDFGKVTKQEVKDFFRWAKDHVVEVEHEFYVKKGGAKKRPAPSDADISE
PKRVRESVAQPSTSDAEASINYADRLARGHSL*

SEQ ID NO: 6 - Rep78 nucleotide sequence (AAV serotype 2)

atgccggggttttacgagattgtgattaaggtccccagcgaccttgacgggcatctgcccggcatttctg
acagctttgtgaactgggtggccgagaaggaatgggagttgccgccagattctgacatggatctgaatct

gattgagcaggcacccctgaccgtggccgagaagctgcagcgcgactttctgacggaatggcgccgtgtg
agtaaggccccggaggccctttttctttgtgcaatttgagaagggagagagctacttccacatgcacgtgc
tcgtggaaaccaccggggtgaaatccatggttttgggacgtttcctgagtcagattcgcgaaaaactgat
tcagagaatttaccgcgggatcgagccgactttgccaaactggttcgcggtcacaaagaccagaaatggc
gccggaggcgggaacaaggtggtggatgagtgctacatccccaattacttgctccccaaaacccagcctg
agctccagtgggcgtggactaatatggaacagtacctcagcgcctgtttgaatctcacggagcgtaaacg
gttggtggcgcagcatctgacgcacgtgtcgcagacgcaggagcagaacaaagagaatcagaatcccaat
tctgatgcgccggtgatcagatcaaaaacttcagccaggtacatggagctggtcgggtggctcgtggaca
aggggattacctcggagaagcagtggatccaggaggaccaggcctcatacatctccttcaatgcggcctc
caactcgcggtcccaaatcaaggctgccttggacaatgcgggaaagattatgagcctgactaaaaccgcc
cccgactacctggtgggccagcagcccgtggaggacatttccagcaatcggatttataaaattttggaac
taaacgggtacgatccccaatatgcggcttccgtctttctgggatgggccacgaaaaagttcggcaagag
gaacaccatctggctgtttgggcctgcaactaccgggaagaccaacatcgcggaggccatagcccacact
gtgcccttctacgggtgcgtaaactggaccaatgagaactttcccttcaacgactgtgtcgacaagatgg
tgatctggtgggaggaggggaagatgaccgccaaggtcgtggagtcggccaaagccattctcggaggaag
caaggtgcgcgtggaccagaaatgcaagtcctcggcccagatagacccgactcccgtgatcgtcacctcc
aacaccaacatgtgcgccgtgattgacgggaactcaacgaccttcgaacaccagcagccgttgcaagacc
ggatgttcaaatttgaactcacccgccgtctggatcatgactttgggaaggtcaccaagcaggaagtcaa
agacttttccggtgggcaaaggatcacgtggttgaggtggagcatgaattctacgtcaaaaagggtgga
gccaagaaaagacccgcccccagtgacgcagatataagtgagcccaaacgggtgcgcgagtcagttgcgc
agccatcgacgtcagacgcggaagcttcgatcaactacgcagacaggtaccaaaacaaatgttctcgtca
cgtgggcatgaatctgatgctgtttccctgcagacaatgcgagagaatgaatcagaattcaaatatctgc
ttcactcacggacagaaagactgtttagagtgctttcccgtgtcagaatctcaacccgtttctgtcgtca
aaaaggcgtatcagaaactgtgctacattcatcatatcatgggaaaggtgccagacgcttgcactgcctg
cgatctggtcaatgtggatttggatgactgcatctttgaacaaTAG

SEQ ID NO: 7 - Rep68 nucleotide sequence (AAV serotype 2)

ATGCCGGGGTTTTACGAGattgtgattaaggtccccagcgaccttgacgagcatctgcccggcatttctg
acagctttgtgaactgggtggccgagaaggaatgggagttgccgccagattctgacatggatctgaatct
gattgagcaggcacccctgaccgtggccgagaagctgcagcgcgactttctgacggaatggcgccgtgtg
agtaaggccccggaggccctttttctttgtgcaatttgagaagggagagagctacttccacatgcacgtgc
tcgtggaaaccaccggggtgaaatccatggttttgggacgtttcctgagtcagattcgcgaaaaactgat
tcagagaatttaccgcgggatcgagccgactttgccaaactggttcgcggtcacaaagaccagaaatggc
gccggaggcgggaacaaggtggtggatgagtgctacatccccaattacttgctccccaaaacccagcctg
agctccagtgggcgtGGACTAATATGGAACAGTACCTCAGCGCCTGTTTGAATCTCACGGagcgtaaacg

gttggtggcgcagcatctgacgcacgtgtcgcagacgcaggagcagaacaaagagaatcagaatcccaat
tctgatgcgccggtgatcagatcaaaaacttcagccaggtacatggagctggtcgggtggctcgtggaca
aggggattacctcggagaagcagtggatccaggaggaccaggcctcatacatctccttcaatgcggcctc
caactcgcggtcccaaatcaaggctgccttggacaatgcgggaaagattatgagcctgactaaaaccgcc
cccgactacctggtgggccagcagcccgtggaggacatttccagcaatcggatttataaaattttggaac
taaacgggtacgatccccaatatgcggcttccgtctttctgggatgggccacgaaaaagttcggcaagag
gaacaccatctggctgtttgggcctgcaactaccgggaagaccaacatcgcggaggccatagcccacact
gtgcccttctacgggtgcgtaaactggaccaatgagaactttcccttcaacgactgtgtcgacaagatgg
tgatctggtgggaggaggggaagatgaccgccaaggtcgtggagtcggccaaagccattctcggaggaag
caaggtgcgcgtggaccagaaatgcaagtcctcggcccagatagacccgactcccgtgatcgtcacctcc
aacaccaacatgtgcgccgtgattgacgggaactcaacgaccttcgaacaccagcagccgttgcaagacc
ggatgttcaaatttgaactcacccgccgtctggatcatgactttgggaaggtcaccaagcaggaagtcaa
agacttttccggtgggcaaaggatcacgtggttgaggtggagcatgaattctacgtcaaaaagggtgga
gccaagaaaagacccgccccccagtgacgcagatataagtgagcccaaacgggtgcgcgagtcagttgcgc
agccatcgacgtcagacgcggaagcttcgatcaactacgcagacagTAG

SEQ ID NO: 8 - Rep52 nucleotide sequence:

CATGGAGCTGGTCGGGTGGctcgtggacaaggggattacctcggagaagcagtggatccaggaggaccag
gcctcatacatctccttcaatgcggcctccaactcgcggtcccaaatcaaggctgccttggacaatgcgg
gaaagattatgagcctgactaaaaccgcccccgactacctggtgggccagcagcccgtggaggacatttc
cagcaatcggatttataaaattttggaactaaacgggtacgatccccaatatgcggcttccgtctttctg
ggatgggccacgaaaaagttcggcaagaggaacaccatctggctgtttgggcctgcaactaccgggaaga
ccaacatcgcggaggccatagcccacactgtgcccttctacgggtgcgtaaactggaccaatgagaactt
tcccttcaacgactgtgtcgacaagatggtgatctggtgggaggaggggaagatgaccgccaaggtcgtg
gagtcggccaaagccattctcggaggaagcaaggtgcgcgtggaccagaaatgcaagtcctcggcccaga
tagacccgactcccgtgatcgtcacctccaacaccaacatgtgcgccgtgattgacgggaactcaacgac
cttcgaacaccagcagccgttgcaagaccggatgttcaaatttgaactcacccgccgtctggatcatgac
tttgggaaggtcaccaagcaggaagtcaaagacttttccggtgggcaaaggatcacgtggttgaggtgg
agcatgaattctacgtcaaaaagggtggagccaagaaaagacccgcccccagtgacgcagatataagtga
gcccaaacgggtgcgcgagtcagttgcgcagccatcgacgtcagacgcggaagcttcgatcaactacgca
gacaggtaccaaaacaaatgttctcgtcacgtgggcatgaatctgatgctgtttccctgcagacaatgcg
agagaatgaatcagaattcaaatatctgcttcactcacggacagaaagactgtttagagtgctttcccgt
gtcagaatctcaacccgtttctgtcgtcaaaaaggcgtatcagaaactgtgctacattcatcatatcatg
ggaaaggtgccagacgcttgcactgcctgcgatctggtcaatgtggatttggatgaCTGCATCTTTGAAC
AATAG

SEQ ID NO: 9 - Rep40 nucleotide sequence

ATGGAGCTGGTCGGGTGGctcgtggacaaggggattacctcggagaagcagtggatccaggaggaccagg
cctcatacatctccttcaatgcggcctccaactcgcggtcccaaatcaaggctgccttggacaatgcggg
aaagattatgagcctgactaaaaccgcccccgactacctggtgggccagcagcccgtggaggacatttcc
agcaatcggatttataaaatttggaactaaacgggtacgatccccaatatgcggcttccgtctttctgg
gatgggccacgaaaaagttcggcaagaggaacaccatctggctgtttgggcctgcaactaccgggaagac
caacatcgcggaggccatagcccacactgtgcccttctacgggtgcgtaaactggaccaatgagaacttt
cccttcaacgactgtgtcgacaagatggtgatctggtgggaggaggggaagatgaccgccaaggtcgtgg
agtcggccaaagccattctcggaggaagcaaggtgcgcgtggaccagaaatgcaagtcctcggcccagat
agacccgactcccgtgatcgtcacctccaacaccaacatgtgcgccgtgattgacgggaactcaacgacc
ttcgaacaccagcagccgttgcaagaccggatgttcaaatttgaactcacccgccgtctggatcatgact
ttgggaaggtcaccaagcaggaagtcaaagacttttccggtgggcaaaggatcacgtggttgaggtgga
gcatgaattctacgtcaaaaagggtggagccaagaaaagacccgcccccagtgacgcagatataagtgag
cccaaacgggtgcgcgagtcagttgcgcagccatcgacgtcagacgcggaagcttcgatcaactacgcag
acagattggctcgaggacactctctcTAG

SEQ ID NO: 10 - Cap nucleotide sequence (AAV serotype 2)

Cagttgcgcagccatcgacgtcagacgcggaagcttcgatcaactacgcagacaggtaccaaaacaaatg
ttctcgtcacgtgggcatgaatctgatgctgtttccctgcagacaatgcgagagaatgaatcagaattca
aatatctgcttcactcacggacagaaagactgtttagagtgctttcccgtgtcagaatctcaacccgttt
ctgtcgtcaaaaaggcgtatcagaaactgtgctacattcatcatatcatgggaaaggtgccagacgcttg
cactgcctgcgatctggtcaatgtggatttggatgactgcatctttgaacaataaatgatttaaatcagg
tatggctgccgatggttatcttccagattggctcgaggacactctctctgaaggaataagacagtggtgg
aagctcaaacctggcccaccaccaccaaagcccgcagagcggcataaggacgacagcaggggtcttgtgc
ttcctgggtacaagtacctcggacccttcaacggactcgacaagggagagccggtcaacgaggcagacgc
cgcggccctcgagcacgacaaagcctacgaccggcagctcgacagcggagacaacccgtacctcaagtac
aaccacgccgacgcggagtttcaggagcgccttaaagaagatacgtcttttgggggcaacctcggacgag
cagtcttccaggcgaaaaagagggttcttgaacctctgggcctggttgaggaacctgttaagacggctcc
gggaaaaaagaggccggtagagcactctcctgtggagccagactcctcctcgggaaccggaaaggcgggc
cagcagcctgcaagaaaaagattgaattttggtcagactggagacgcagactcagtacctgacccccagc
ctctcggacagccaccagcagccccctctggtctgggaactaatacgatggctacaggcagtggcgcacc
aatggcagacaataacgagggcgccgacggagtgggtaattcctcgggaaattggcattgcgattccaca
tggatgggcgacagagtcatcaccaccagcacccgaacctgggccctgcccacctacaacaaccacctct
acaaacaaatttccagccaatcaggagcctcgaacgacaatcactactttggctacagcacccttgggg

gtattttgacttcaacagattccactgccacttttcaccacgtgactggcaaagactcatcaacaacaac
tggggattccgacccaagagactcaacttcaagctctttaacattcaagtcaaagaggtcacgcagaatg
acggtacgacgacgattgccaataaccttaccagcacggttcaggtgtttactgactcggagtaccagct
cccgtacgtcctcggctcggcgcatcaaggatgcctcccgccgttcccagcagacgtcttcatggtgcca
cagtatggatacctcaccctgaacaacgggagtcaggcagtaggacgctcttcattttactgcctggagt
actttccttctcagatgctgcgtaccggaaacaactttaccttcagctacacttttgaggacgttccttt
ccacagcagctacgctcacagccagagtctggaccgtctcatgaatcctctcatcgaccagtacctgtat
tacttgagcagaacaaacactccaagtggaaccaccacgcagtcaaggcttcagttttctcaggccggag
cgagtgacattcgggaccagtctaggaactggcttcctggaccctgttaccgccagcagcgagtatcaaa
gacatctgcggataacaacaacagtgaatactcgtggactggagctaccaagtaccacctcaatggcaga
gactctctggtgaatccgggcccggccatggcaagccacaaggacgatgaagaaaagtttttttcctcaga
gcggggttctcatctttgggaagcaaggctcagagaaaacaaatgtggacattgaaaaggtcatgattac
agacgaagaggaaatcaggacaaccaatcccgtggctacggagcagtatggttctgtatctaccaacctc
cagagaggcaacagacaagcagctaccgcagatgtcaacacacaaggcgttcttccaggcatggtctggc
aggacagagatgtgtaccttcaggggcccatctgggcaaagattccacacacggacggacattttcaccc
ctctcccctcatgggtggattcggacttaaacaccctcctccacagattctcatcaagaacaccccggta
cctgcgaatccttcgaccaccttcagtgcggcaaagtttgcttccttcatcacacagtactccacgggac
aggtcagcgtggagatcgagtgggagctgcagaaggaaaacagcaaacgctggaatcccgaaattcagta
cacttccaactacaacaagtctgttaatgtggactttactgtggacactaatggcgtgtattcagagcct
cgccccattggcaccagatacctgactcgtaatctgtaA

SEQ ID NO: 11 - Cap amino acid sequence (AAV serotype 2)

MAADGYLPDWLEDTLSEGIRQWWKLKPGPPPPKPAERHKDDSRGLVLPGYKYLGPFNGLDKGEPVNEADA
AALEHDKAYDRQLDSGDNPYLKYNHADAEFQERLKEDTSFGGNLGRAVFQAKKRVLEPLGLVEEPVKTAP
GKKRPVEHSPVEPDSSSGTGKAGQQPARKRLNFGQTGDADSVPDPQPLGQPPAAPSGLGTNTMATGSGAP
MADNNEGADGVGNSSGNWHCDSTWMGDRVITTSTRTWALPTYNNHLYKQISSQSGASNDNHYFGYSTPWG
YFDFNRFHCHFSPRDWQRLINNNWGFRPKRLNFKLFNIQVKEVTQNDGTTTIANNLTSTVQVFTDSEYQL
PYVLGSAHQGCLPPFPADVFMVPQYGYLTLNNGSQAVGRSSFYCLEYFPSQMLRTGNNFTFSYTFEDVPF
HSSYAHSQSLDRLMNPLIDQYLYYLSRTNTPSGTTTQSRLQFSQAGASDIRDQSRNWLPGPCYRQQRVSK
TSADNNNSEYSWTGATKYHLNGRDSLVNPGPAMASHKDDEEKFFPQSGVLIFGKQGSEKTNVDIEKVMIT
DEEEIRTTNPVATEQYGSVSTNLQRGNRQAATADVNTQGVLPGMVWQDRDVYLQGPIWAKIPHTDGHFHP
SPLMGGFGLKHPPPQILIKNTPVPANPSTTFSAAKFASFITQYSTGQVSVEIEWELQKENSKRWNPEIQY
TSNYNKSVNVDFTVDTNGVYSEPRPIGTRYLTRNL*

SEQ ID NO: 12 - TetR binding site

tccctatcagtgatagaga

SEQ ID NO: 13 - Nucleotide sequence of the TetR protein

Atgtcgcgcctggacaaaagcaaagtgattaactcagcgctggaactgttgaatgaggtgggaattgaag
gactcactactcgcaagctggcacagaagctgggcgtcgagcagccaacgctgtactggcatgtgaagaa
taaacgggcgctcctagacgcgcttgccatcgaaatgctggaccgccatcacacccacttttgccccctg
gagggcgaatcctggcaagattttctgcggaacaatgcaaagtcgttccggtgcgctctgctgtcccacc
gcgatggcgcaaaagtgcacctgggcactcggcccaccgagaaacaatacgaaaccctggaaaaccaact
ggctttcctttgccaacagggattttcactggagaatgccctgtacgcactatccgcggtcggccacttt
accctgggatgcgtcctcgaagatcaggagcaccaagtcgccaaggaggaaagagaaactcctaccactg
actcaatgcctccgctcctgagacaagccatcgagctgttcgaccaccagggtgctgaacctgcatttct
gttcgggcttgaactgattatctgcggcctggagaaacagttgaagtgcgagtcgggatcctag

SEQ ID NO: 14 - Amino acid sequence of the TetR protein

MSRLDKSKVINSALELLNEVGIEGLTTRKLAQKLGVEQPTLYWHVKNKRALLDALAIEMLDRHHTHFCPL
EGESWQDFLRNNAKSFRCALLSHRDGAKVHLGTRPTEKQYETLENQLAFLCQQGFSLENALYALSAVGHF
TLGCVLEDQEHQVAKEERETPTTDSMPPLLRQAIELFDHQGAEPAFLFGLELIICGLEKQLKCESGS

SEQ ID NO: 15 - EMCV IRES

CGTTACTGGCCGAAGCCGCTTGGAATAAGGCCGGTGTGCGTTTGTCTATATGTTATTTTCCACCATATTG
CCGTCTTTTGGCAATGTGAGGGCCCGGAAACCTGGCCCTGTCTTCTTGACGAGCATTCCTAGGGGTCTTT
CCCCTCTCGCCAAAGGAATGCAAGGTCTGTTGAATGTCGTGAAGGAAGCAGTTCCTCTGGAAGCTTCTTG
AAGACAAACAACGTCTGTAGCGACCCTTTGCAGGCAGCGGAACCCCCCACCTGGCGACAGGTGCCTCTGC
GGCCAAAAGCCACGTGTATAAGATACACCTGCAAAGGCGGCACAACCCCAGTGCCACGTTGTGAGTTGGA
TAGTTGTGGAAAGAGTCAAATGGCTCCCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCCAGAAGGT
ACCCCATTGTATGGGATCTGATCTGGGGCCTCGGTGCACATGCTTTACATGTGTTTAGTCGAGGTTAAAA
AACGTCTAGGCCCCCCGAACCACGGGGAC

SEQ ID NO: 16 - FMDV IRES

AGCAGGTTTCCCCAACTGACACAAAACGTGCAACTTGAAACTCCGCCTGGTCTTTCCAGGTCTAGAGGGG
TAACACTTTGTACTGCGTTTGGCTCCACGCTCGATCCACTGGCGAGTGTTAGTAACAGCACTGTTGCTTC
GTAGCGGAGCATGACGGCCGTGGGAACTCCTCCTTGGTAACAAGGACCCACGGGGCCAAAAGCCACGCCC
ACACGGGCCCGTCATGTGTGCAACCCCAGCACGGCGACTTTACTGCGAAACCCACTTTAAAGTGACATTG
AAACTGGTACCCACACACTGGTGACAGGCTAAGGATGCCCTTCAGGTACCCCGAGGTAACACGCGACACT

CGGGATCTGAGAAGGGGACTGGGGCTTCTATAAAAGCGCTCGGTTTAAAAAGCTTCTATGCCTGAATAGG
TGACCGGAGGTCGGCACCTTTCCTTTGCAATTACTGACCAC

SEQ ID NO: 17 - Ad5 E2A

GGTACCCAACTCCATGCTCAACAGTCCCCAGGTACAGCCCACCCTGCGTCGCAACCAGGAACAGCTCTAC
AGCTTCCTGGAGCGCCACTCGCCCTACTTCCGCAGCCACAGTGCGCAGATTAGGAGCGCCACTTCTTTTT
GTCACTTGAAAAACATGTAAAAATAATGTACTAGAGACACTTTCAATAAAGGCAAATGCTTTTATTTGTA
CACTCTCGGGTGATTATTTACCCCCACCCTTGCCGTCTGCGCCGTTTAAAAATCAAAGGGGTTCTGCCGC
GCATCGCTATGCGCCACTGGCAGGGACACGTTGCGATACTGGTGTTTAGTGCTCCACTTAAACTCAGGCA
CAACCATCCGCGGCAGCTCGGTGAAGTTTTCACTCCACAGGCTGCGCACCATCACCAACGCGTTTAGCAG
GTCGGGCGCCGATATCTTGAAGTCGCAGTTGGGGCCTCCGCCCTGCGCGCGCGAGTTGCGATACACAGGG
TTGCAGCACTGGAACACTATCAGCGCCGGGTGGTGCACGCTGGCCAGCACGCTCTTGTCGGAGATCAGAT
CCGCGTCCAGGTCCTCCGCGTTGCTCAGGGCGAACGGAGTCAACTTTGGTAGCTGCCTTCCCAAAAAGGG
CGCGTGCCCAGGCTTTGAGTTGCACTCGCACCGTAGTGGCATCAAAAGGTGACCGTGCCCGGTCTGGGCG
TTAGGATACAGCGCCTGCATAAAAGCCTTGATCTGCTTAAAAGCCACCTGAGCCTTTGCGCCTTCAGAGA
AGAACATGCCGCAAGACTTGCCGGAAAACTGATTGGCCGGACAGGCCGCGTCGTGCACGCAGCACCTTGC
GTCGGTGTTGGAGATCTGCACCACATTTCGGCCCCACCGGTTCTTCACGATCTTGGCCTTGCTAGACTGC
TCCTTCAGCGCGCGCTGCCCGTTTTCGCTCGTCACATCCATTTCAATCACGTGCTCCTTATTTATCATAA
TGCTTCCGTGTAGACACTTAAGCTCGCCTTCGATCTCAGCGCAGCGGTGCAGCCACAACGCGCAGCCCGT
GGGCTCGTGATGCTTGTAGGTCACCTCTGCAAACGACTGCAGGTACGCCTGCAGGAATCGCCCCATCATC
GTCACAAAGGTCTTGTTGCTGGTGAAGGTCAGCTGCAACCCGCGGTGCTCCTCGTTCAGCCAGGTCTTGC
ATACGGCCGCCAGAGCTTCCACTTGGTCAGGCAGTAGTTTGAAGTTCGCCTTTAGATCGTTATCCACGTG
GTACTTGTCCATCAGCGCGCGCGCAGCCTCCATGCCCTTCTCCCACGCAGACACGATCGGCACACTCAGC
GGGTTCATCACCGTAATTTCACTTTCCGCTTCGCTGGGCTCTTCCTCTTCCTCTTGCGTCCGCATACCAC
GCGCCACTGGGTCGTCTTCATTCAGCCGCCGCACTGTGCGCTTACCTCCTTTGCCATGCTTGATTAGCAC
CGGTGGGTTGCTGAAACCCACCATTTGTAGCGCCACATCTTCTCTTTCTTCCTCGCTGTCCACGATTACC
TCTGGTGATGGCGGGCGCTCGGGCTTGGGAGAAGGGCGCTTCTTTTTCTTCTTGGGCGCAATGGCCAAAT
CCGCCGCCGAGGTCGATGGCCGCGGGCTGGGTGTGCGCGGCACCAGCGCGTCTTGTGATGAGTCTTCCTC
GTCCTCGGACTCGATACGCCGCCTCATCCGCTTTTTTGGGGGCGCCCGGGGAGGCGGCGGCGACGGGGAC
GGGGACGACACGTCCTCCATGGTTGGGGGACGTCGCGCCGCACCGCGTCCGCGCTCGGGGGTGGTTTCGC
GCTGCTCCTCTTCCCGACTGGCCATTTCCTTCTCCTATAGGCAGAAAAAGATCATGGAGTCAGTCGAGAA
GAAGGACAGCCTAACCGCCCCCTCTGAGTTCGCCACCACCGCCTCCACCGATGCCGCCAACGCGCCTACC
ACCTTCCCCGTCGAGGCACCCCCGCTTGAGGAGGAGGAAGTGATTATCGAGCAGGACCCAGGTTTTGTAA
GCGAAGACGACGAGGACCGCTCAGTACCAACAGAGGATAAAAAGCAAGACCAGGACAACGCAGAGGCAAA
CGAGGAACAAGTCGGGCGGGGGGACGAAAGGCATGGCGACTACCTAGATGTGGGAGACGACGTGCTGTTG

AAGCATCTGCAGCGCCAGTGCGCCATTATCTGCGACGCGTTGCAAGAGCGCAGCGATGTGCCCCTCGCCA
TAGCGGATGTCAGCCTTGCCTACGAACGCCACCTATTCTCACCGCGCGTACCCCCCAAACGCCAAGAAAA
CGGCACATGCGAGCCCAACCCGCGCCTCAACTTCTACCCCGTATTTGCCGTGCCAGAGGTGCTTGCCACC
TATCACATCTTTTTCCAAAACTGCAAGATACCCCTATCCTGCCGTGCCAACCGCAGCCGAGCGGACAAGC
AGCTGGCCTTGCGGCAGGGCGCTGTCATACCTGATATCGCCTCGCTCAACGAAGTGCCAAAAATCTTTGA
GGGTCTTGGACGCGACGAGAAGCGCGCGGCAAACGCTCTGCAACAGGAAAACAGCGAAAATGAAAGTCAC
TCTGGAGTGTTGGTGGAACTCGAGGGTGACAACGCGCGCCTAGCCGTACTAAAACGCAGCATCGAGGTCA
CCCACTTTGCCTACCCGGCACTTAACCTACCCCCCAAGGTCATGAGCACAGTCATGAGTGAGCTGATCGT
GCGCCGTGCGCAGCCCCTGGAGAGGGATGCAAATTTGCAAGAACAAACAGAGGAGGGCCTACCCGCAGTT
GGCGACGAGCAGCTAGCGCGCTGGCTTCAAACGCGCGAGCCTGCCGACTTGGAGGAGCGACGCAAACTAA
TGATGGCCGCAGTGCTCGTTACCGTGGAGCTTGAGTGCATGCAGCGGTTCTTTGCTGACCCGGAGATGCA
GCGCAAGCTAGAGGAAACATTGCACTACACCTTTCGACAGGGCTACGTACGCCAGGCCTGCAAGATCTCC
AACGTGGAGCTCTGCAACCTGGTCTCCTACCTTGGAATTTTGCACGAAAACCGCCTTGGGCAAAACGTGC
TTCATTCCACGCTCAAGGGCGAGGCGCGCCGCGACTACGTCCGCGACTGCGTTTACTTATTTCTATGCTA
CACCTGGCAGACGGCCATGGGCGTTTGGCAGCAGTGCTTGGAGGAGTGCAACCTCAAGGAGCTGCAGAAA
CTGCTAAAGCAAAACTTGAAGGACCTATGGACGGCCTTCAACGAGCGCTCCGTGGCCGCGCACCTGGCGG
ACATCATTTTCCCCGAACGCCTGCTTAAAACCCTGCAACAGGGTCTGCCAGACTTCACCAGTCAAAGCAT
GTTGCAGAACTTTAGGAACTTTATCCTAGAGCGCTCAGGAATCTTGCCCGCCACCTGCTGTGCACTTCCT
AGCGACTTTGTGCCCATTAAGTACCGCGAATGCCCTCCGCCGCTTTGGGGCCACTGCTACCTTCTGCAGC
TAGCCAACTACCTTGCCTACCACTCTGACATAATGGAAGACGTGAGCGGTGACGGTCTACTGGAGTGTCA
CTGTCGCTGCAACCTATGCACCCCGCACCGCTCCCTGGTTTGCAATTCGCAGCTGCTTAACGAAAGTCAA
ATTATCGGTACCTTTGAGCTGCAGGGTCCCTCGCCTGACGAAAAGTCCGCGGCTCCGGGGTTGAAACTCA
CTCCGGGGCTGTGGACGTCGGCTTACCTTCGCAAATTTGTACCTGAGGACTACCACGCCCACGAGATTAG
GTTCTACGAAGACCAATCCCGCCCGCCTAATGCGGAGCTTACCGCCTGCGTCATTACCCAGGGCCACATT
CTTGGCCAATTGCAAGCCATCAACAAAGCCCGCCAAGAGTTTCTGCTACGAAAGGGACGGGGGGTTTACT
TGGACCCCCAGTCCGGCGAGGAGCTCAACCCAATCCCCCCGCCGCCGCAGCCCTATCAGCAGCAGCCGCG
GGCCCTTGCTTCCCAGGATGGCACCCAAAAAGAAGCTGCAGCTGCCGCCGCCACCCACGGACGAGGAGGA
ATACTGGGACAGTCAGGCAGAGGAGGTTTTGGACGAGGAGGAGGAGGACATGATGGAAGACTGGGAGAGC
CTAGACGAGGAAGCTTCCGAGGTCGAAGAGGTGTCAGACGAAACACCGTCACCCTCGGTCGCATTCCCCT
CGCCGGCGCCCCAGAAATCGGCAACCGGTTCCAGCATGGCTACAACCTCCGCTCCTCAGGCGCCGCCGGC
ACTGCCCGTTCGCCGACCCAACCGTAGATGGGACACCACTGGAACCAGGGCCGGTAAGTCCAAGCAGCCG
CCGCCGTTAGCCCAAGAGCAACAACAGCGCCAAGGCTACCGCTCATGGCGCGGGCACAAGAACGCCATAG
TTGCTTGCTTGCAAGACTGTGGGGGCAACATCTCCTTCGCCCGCCGCTTTCTTCTCTACCATCACGGCGT
GGCCTTCCCCCGTAACATCCTGCATTACTACCGTCATCTCTACAGCCCATACTGCACCGGCGGCAGCGGC
AGCAACAGCAGCGGCCACACAGAAGCAAAGGCGACCGGATAGCAAGACTCTGACAAAGCCCAAGAAATCC

ACAGCGGCGGCAGCAGCAGGAGGAGGAGCGCTGCGTCTGGCGCCCAACGAACCCGTATCGACCCGCGAGC
TTAGAAACAGGATTTTTCCCACTCTGTATGCTATATTTCAACAGAGCAGGGGCCAAGAACAAGAGCTGAA
AATAAAAAACAGGTCTCTGCGATCCCTCACCCGCAGCTGCCTGTATCACAAAAGCGAAGATCAGCTTCGG
CGCACGCTGGAAGACGCGGAGGCTCTCTTCAGTAAATACTGCGCGCTGACTCTTAAGGACTAGTTTCGCG
CCCTTTCTCAAATTTAAGCGCGAAAACTACGTCATCTCCAGCGGCCACACCCGGCGCCAGCACCTGTTGT
CAGCGCCATTATGAGCAAGGAAATTCCCACGCCCTACATGTGGAGTTACCAGCCACAAATGGGACTTGCG
GCTGGAGCTGCCCAAGACTACTCAACCCGAATAAACTACATGAGCGCGGGACCCCACATGATATCCCGGG
TCAACGGAATACGCGCCCACCGAAACCGAATTCCCTTGGAACAGGCGGCTATTACCACCACACCTCGTAA
TAACCTTAATCCCCGTAGTTGGCCCGCTGCCCTGGTGTACCAGGAAAGTCCCGCTCCCACCACTGTGGTA
CTTCCCAGAGACGCCCAGGCCGAAGTTCAGATGACTAACTCAGGGGCGCAGCTTGCGGGCGGCTTTCGTC
ACAGGGTGCGGTCGCCCGGGC

SEQ ID NO: 18 - CMV promoter WT

AGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAA
TGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTA
ACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTAC
ATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTA
TGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACC
ATGCTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTC
TCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAA
CAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGT
TTAGTGAACCGTC

SEQ ID NO: 19 - CMV promoter inducible (p565-2xTetO)

CGTGAGGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGG
GGTCGGCAATTGAACCGGTGCCTAGAGAAGGTGGCGCGGGGTAAACTGGGAAAGTGATGTCGTGTACTGG
CTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCACTAGTCGCCGTGAACGTCAATGGAA
AGTCCCTATTGGCGTTACTATGGGAACATACGTCATTATTGACGTCAATGACGGTAAATGGCCCGCCTGG
CATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTA
TTACCATGCTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCC
AAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGT
CGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAG
CTGtccctatcagtgatagagatgtccctatcagtgatagagatcgtcgagcagctcGTTTAGTGAACCG
TCAGATC

SEQ ID NO: 20 - 5′ UTR sequence: Cap sequence

Cagttgcgcagccatcgacgtcagacgcggaagcttcgatcaactacgcagacaggtaccaaaacaaatg
ttctcgtcacgtgggcatgaatctgatgctgtttccctgcagacaatgcgagagaatgaatcagaattca
aatatctgcttcactcacggacagaaagactgtttagagtgctttcccgtgtcagaatctcaacccgttt
ctgtcgtcaaaaaggcgtatcagaaactgtgctacattcatcatatcatgggaaaggtgccagacgcttg
cactgcctgcgatctggtcaatgtggatttggatgactgcatctttgaacaataaatgatttaaatcagg
tatggctgccgatggttatcttccagattggctcgaggacactctctctgaaggaataagacagtggtgg
aagctcaaacctggcccaccaccaccaaagcccgcagagcggcataaggacgacagcaggggtcttgtgc
ttcctgggtacaagtacctcggacccttcaacggactcgacaagggagagccggtcaacgaggcagacgc
cgcggccctcgagcacgacaaagcctacgaccggcagctcgacagcggagacaacccgtacctcaagtac
aaccacgccgacgcggagtttcaggagcgccttaaagaagatacgtcttttgggggcaacctcggacgag
cagtcttccaggcgaaaaagagggttcttgaacctctgggcctggttgaggaacctgttaagacggctcc
gggaaaaaagaggccggtagagcactctcctgtggagccagactcctcctcgggaaccggaaaggcgggc
cagcagcctgcaagaaaaagattgaattttggtcagactggagacgcagactcagtacctgacccccagc
ctctcggacagccaccagcagccccctctggtctgggaactaatacgatggctacaggcagtggcgcacc
aatggcagacaataacgagggcgccgacggagtgggtaattcctcgggaaattggcattgcgattccaca
tggatgggcgacagagtcatcaccaccagcacccgaacctgggccctgcccacctacaacaaccacctct
acaaacaaatttccagccaatcaggagcctcgaacgacaatcactactttggctacagcacccttgggg
gtattttgacttcaacagattccactgccacttttcaccacgtgactggcaaagactcatcaacaacaac
tggggattccgacccaagagactcaacttcaagctctttaacattcaagtcaaagaggtcacgcagaatg
acggtacgacgacgattgccaataaccttaccagcacggttcaggtgtttactgactcggagtaccagct
cccgtacgtcctcggctcggcgcatcaaggatgcctcccgccgttcccagcagacgtcttcatggtgcca
cagtatggatacctcaccctgaacaacgggagtcaggcagtaggacgctcttcattttactgcctggagt
actttccttctcagatgctgcgtaccggaaacaactttaccttcagctacacttttgaggacgttccttt
ccacagcagctacgctcacagccagagtctggaccgtctcatgaatcctctcatcgaccagtacctgtat
tacttgagcagaacaaacactccaagtggaaccaccacgcagtcaaggcttcagttttctcaggccggag
cgagtgacattcgggaccagtctaggaactggcttcctggaccctgttaccgccagcagcgagtatcaaa
gacatctgcggataacaacaacagtgaatactcgtggactggagctaccaagtaccacctcaatggcaga
gactctctggtgaatccgggcccggccatggcaagccacaaggacgatgaagaaaagtttttttcctcaga
gcggggttctcatctttgggaagcaaggctcagagaaaacaaatgtggacattgaaaaggtcatgattac
agacgaagaggaaatcaggacaaccaatcccgtggctacggagcagtatggttctgtatctaccaacctc
cagagaggcaacagacaagcagctaccgcagatgtcaacacacaaggcgttcttccaggcatggtctggc
aggacagagatgtgtaccttcaggggcccatctgggcaaagattccacacacggacggacattttcaccc
ctctcccctcatgggtggattcggacttaaacaccctcctccacagattctcatcaagaacaccccggta
cctgcgaatccttcgaccaccttcagtgcggcaaagtttgcttccttcatcacacagtactccacgggac

aggtcagcgtggagatcgagtgggagctgcagaaggaaaacagcaaacgctggaatcccgaaattcagta
cacttccaactacaacaagtctgttaatgtggactttactgtggacactaatggcgtgtattcagagcct
cgccccattggcaccagatacctgactcgtaatctgtaA

SEQ ID NO: 21 - 3′ UTR:Polyadenylation sequence (SV40)

CAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTAT
TTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAAC
AATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTTTTT

CLAIMS

1. A nucleic acid molecule comprising:

(i) a promoter,

(ii) an AAV *cap* gene, and

(iii) an AAV *rep* gene,

in the above 5' → 3' order, wherein the *cap* gene and the *rep* gene are both operably-associated with the promoter, and wherein the *rep* gene is also operably-associated with an IRES.

2. A nucleic acid molecule as claimed in claim 1, wherein:

(a) the *rep* gene encodes Rep78, Rep68, Rep52 and Rep40; or

(b) the *rep* gene only encodes one, two or three of Rep78, Rep68, Rep52 and Rep40.

3. A nucleic acid molecule as claimed in claim 1, wherein the *rep* gene:

(a) encodes Rep78 and Rep52, but does not encode Rep68 or Rep40;

(b) encodes Rep68 and Rep40, but does not encode Rep78 or Rep52;

(c) encodes Rep68 and Rep52, but does not encode Rep78 or Rep40; or

(d) encodes Rep78 and Rep40, but does not encode Rep68 or Rep52.

4. A nucleic acid molecule as claimed in any one of the preceding claims, wherein Rep52 and/or Rep40 are not transcribed from the said promoter.

5. A nucleic acid molecule as claimed in claim 4, wherein Rep52 and/or Rep40 are transcribed from a p19 promoter.

6. A nucleic acid molecule as claimed in any one of the preceding claims, wherein the nucleotide sequences of the *rep* and *cap* genes are AAV serotype 2 *rep* or *cap* genes.

7. A nucleic acid molecule as claimed in any one of the preceding claims, wherein the promoter is a cytomegalovirus immediate early (CMV) promoter.

8. A nucleic acid molecule as claimed in any one of claims 1 to 6, wherein the promoter is an inducible promoter.

9. A nucleic acid molecule as claimed in any one of the preceding claims, wherein the IRES is a Picornavirus IRES (Encephalomyocarditis virus, EMCV IRES) or an Aphthovirus IRES (Foot-and-mouth disease virus, FMDV IRES).

10. A RNA molecule comprising:

(i) an AAV *cap* gene, and

(ii) an AAV *rep* gene,

in the above 5' → 3' order, wherein the *rep* gene is operably-associated with an IRES.

11. A plasmid or vector comprising a nucleic acid molecule as defined in any one of claims 1 to 9.

12. A kit comprising

(a) a plasmid or vector as claimed in claim 11, together with one or both of:

(b) an AAV Transfer Plasmid comprising a transgene flanked by ITRs;

(c) a Helper Plasmid comprising one or more genes selected from E1A, E1B, E2A, E4 and VA.

13. A kit as claimed in claim 12, wherein the Helper Plasmid does not comprise an E2A gene.

14. A kit comprising

(a) a plasmid or vector as defined in claim 11, together with one or both of the following:

(b) an AAV Transfer Plasmid comprising a transgene flanked by ITRs;

(c) a mammalian host cell, preferably HEK293, comprising one or more viral genes selected from E1A, E1B, E2A, E4 and VA expressible from the host cell genome.

15. A kit as claimed in claim 14, wherein the mammalian host cell does not comprise an E2A gene expressible from the host cell genome.

16. A mammalian cell comprising a nucleic acid molecule, plasmid or vector as claimed in any one of claims 1 to 9 or 11, preferably wherein the mammalian cell is a human cell, more preferably a HEK293 cell or a derivative thereof.

17. Use of a mammalian cell as claimed in claim 16 in the production of an AAV particle.

18. A process for producing an AAV packaging cell, the process comprising the steps:

(a) stably integrating a nucleic acid molecule as claimed in any one of claims 1 to 9 or a plasmid or vector as claimed in claim 11 into a mammalian cell,

thereby producing a mammalian cell that expresses AAV *rep* and *cap* genes.

19. A process for producing AAVs, the process comprising the steps:

(a) introducing a Transfer Plasmid comprising 5'- and 3'-AAV ITRs flanking a transgene into an AAV packaging cell, the AAV packaging cell comprising a nucleic acid molecule as claimed in any one of claims 1 to 9 and sufficient helper genes, preferably selected from one or more of E1A, E1B, E2A, E4 and VA, for packaging the Transfer Plasmid, the helper genes either being present in an episomal Helper Plasmid within the cell or being integrated into the packaging cell genome;

(b) culturing the cell under conditions such that AAVs are assembled and secreted by the cell; and

(c) harvesting packaged AAVs from the supernatant, and optionally purifying the harvested AAVs.

20. A process as claimed in claim 19, wherein the helper genes do not include an E2A gene.

21. A process as claimed in claim 19 or claim 20, wherein the transgene encodes a CRISPR enzyme, preferably Cas9 or Cpf1, or a CRISPR sgRNA.