RepCap plasmid

The present invention relates to recombinant nucleic acid molecules, and particularly, although not exclusively, to vectors (e.g. RepCap plasmids) and host cells comprising the recombinant nucleic acid molecules. The invention also relates to methods for producing recombinant adeno-associated viruses (rAAVs) using the recombinant nucleic acid molecules or vectors according to the invention.

Adeno-associated virus (AAV) is a small, non-enveloped, single-stranded DNA (ssDNA) virus. The efficient replication of AAV requires coinfection with a helper virus, such as adenovirus or herpesvirus. AAV is often used as a vector for gene therapy due to its ability to transduce in a number of cell types. Several medicines of AAV vector have already been launched and will continue to be launched in the future. However, AAV manufacturing efficiency is lower than for other medicines, and so productivity needs to be improved to ensure stable supply and cost reduction for AAV therapies.

Recombinant AAV (rAAV) for medicines are usually produced by cells (e.g. HEK293, Hela, sf9, etc.), in which genes for AAV production are transfected. The most commonly used production system is the "triple-plasmid transfection" of the HEK293 cell, with (i) a Helper plasmid (pHelper), (ii) a plasmid containing the gene of interest (pGOI), and (iii) a RepCap plasmid (pRepCap), as illustrated in Figure 1. Alternatively, pHelper and pRepCap may be combined together for a "two-plasmid transfection" method with the separate pGOI. The pHelper plasmid provides the minimal gene products required for AAV replication, such as E2A, E4 and VA. The pGOI plasmid is an inverted terminal repeat (ITR)-containing plasmid, with the *rep* and *cap* genes replaced by the promoter and gene of interest (GOI). The RepCap plasmid contains *rep* and *cap* genes of AAV, which are responsible for genome replication, capsid expression and encapsidation. As illustrated in Figure 2, the *rep* gene encodes four proteins required for viral replication: Rep78, Rep68, Rep52 and Rep40. The *cap* gene encodes the three capsid subunits, VP1, VP2 and VP3, through alternative splicing and translation from different start codons.

Current rAAV production methods result in low yields, such that the levels needed for human clinical trials remains costly and labour-intensive. As such, there is room for improvement in order to optimise the production system using the HEK293 cell, and in particular, allow the production of increased yields of rAAV.

Splicing of pre-mRNA molecules occurs in several steps that are catalysed by small nuclear ribonucleoproteins (snRNPs). Splicing is initiated by the binding of the U1

snRNP to the 5' splice site. The 5' splice sites of many human and non-human primate AAVs are identical (CAG | GTACCA (SEQ ID No: 11)) and differ from the U1 snRNP consensus binding site (CAG | GTA*AGT* (SEQ ID No: 12)), in the last three nucleotides. As such, the splicing efficiency of *rep* and *cap* is low, which limits AAV production. Therefore, it was hypothesised by the inventors that the enhancement of splicing of *rep* and *cap* gene pre-mRNAs by improving the splice site could result in increased rAAV yields.

However, further improvements to the RepCap plasmid are required in order to significantly improve rAAV production yields. For example, stability of mRNA is also considered important for increasing spliced transcripts. This is because when mRNA stability is poor, spliced transcripts are rapidly degraded and reduced, whereas when mRNA stability is good, spliced transcripts remain longer and are expressed at higher levels.

There is, therefore, the need for an improved RepCap plasmid, which can result in increased AAV productivity.

Accordingly, in a first aspect of the invention, there is provided a recombinant nucleic acid molecule comprising an AAV capsid (*Cap*) gene and an AAV replication (*Rep*) gene, wherein the recombinant nucleic acid molecule comprises a splice site sequence which improves gene splicing efficiency, and a polyA signal sequence.

The inventors set out to improve the RepCap plasmid for the purpose of improving rAAV productivity. They hypothesised that splicing efficiency of mRNA of the *cap* gene is the key factor for AAV production because the splicing is needed for capsid protein expression. Additionally, it was thought that the stability of mRNA is considered important for increasing spliced transcripts. Based on these factors, the inventors constructed modified pRepCap, as follows: i) several bases of the 5' splice site of the intron were changed to strengthen the binding to U1 snRNP; and ii) a polyA signal sequence was inserted after the *cap* gene sequence to increase mRNA stability. As demonstrated in the Examples, by using the modified pRepCap according to the invention, AAV production was surprisingly increased.

The process of gene splicing involves many components, including U1 small nuclear ribonucleoprotein (U1 snRNP). It is well-known to the skilled person that U1 snRNP interacts to the 5' splice site of an intron and is essential to initiate splicing. As shown in Figure 2A, *rep* and *cap* transcripts share the same intron.

The original 5′ splice site sequence of the intron of the RepCap plasmid is provided herein as SEQ ID No: 1, as follows:

GTACCA

[SEQ ID No: 1]

The consensus sequence of the 5′ splice site of the intron for U1 snRNP (i.e. U1 snRNP's consensus binding site) is provided herein as SEQ ID No: 2, as follows:

GTAAGT

[SEQ ID No: 2]

Accordingly, it will be appreciated that the 5' splice site of the intron of the original RepCap plasmids differs from the consensus sequence in the last three bases. Thus, it was hypothesised that splicing efficiency would be improved by modifying the 5′ splice site sequence of the original RepCap plasmid such that it is more similar to the consensus sequence.

Accordingly, in an embodiment, the splice site sequence which improves gene splicing efficiency is the 5′ splice site of the *rep* and *cap* genes. In some embodiments, the splice site sequence which improves gene splicing efficiency is the 5′ splice site of the intron of the *rep* and *cap* genes.

In an embodiment, the splice site sequence which improves gene splicing efficiency comprises a sequence as set out in SEQ ID No: 1, comprising at least one base substitution.

In an embodiment, the splice site comprises a sequence as set out in SEQ ID No: 1, comprising at least two base substitutions. In another embodiment, the splice site comprises a sequence as set out in SEQ ID No: 1, comprising at least three base substitutions. In another embodiment, the splice site comprises a sequence as set out in SEQ ID No: 1, comprising at least four base substitutions.

In one embodiment, the at least one base substitution occurs within the last three bases of SEQ ID No: 1. In another embodiment, the at least two base substitutions occur within the last three bases of SEQ ID No: 1. In another embodiment, the at least three base substitutions occur within the last three bases of SEQ ID No: 1.

Alternatively, the at least one, two or three base substitutions may occur within the first three bases of SEQ ID No: 1.

In another embodiment, the splice site comprises a sequence as set out in SEQ ID No: 2, comprising at least one base substitution. The at least one base substitution may occur at the second base within SEQ ID No: 2. Alternatively, the at least one base substitution occurs at the third base within SEQ ID No: 2.

In another embodiment, the splice site comprises or consists of a sequence having at least 60% sequence identity to SEQ ID No: 2. In some embodiments, the splice site comprises or consists of a sequence having at least 65% sequence identity to SEQ ID No: 2. The splice site may comprise or consist of a sequence having at least 70% sequence identity to SEQ ID No: 2. In some embodiments, the splice site comprises or consists of a sequence having at least 75% sequence identity to SEQ ID No: 2. In some embodiments, the splice site comprises or consists of a sequence having at least 80% sequence identity to SEQ ID No: 2. In some embodiments, the splice site comprises or consists of a sequence having at least 85% sequence identity to SEQ ID No: 2. In some embodiments, the splice site comprises or consists of a sequence having at least 90% sequence identity to SEQ ID No: 2.

In some embodiments, the splice site comprises or consists of a sequence having between 60% and 90% sequence identity to SEQ ID No: 2. In some embodiments, the splice site comprises or consists of a sequence having between 65% and 90% sequence identity to SEQ ID No: 2. In some embodiments, the splice site comprises or consists of a sequence having between 70% and 90% sequence identity to SEQ ID No: 2. In some embodiments, the splice site comprises or consists of a sequence having between 75% and 90% sequence identity to SEQ ID No: 2. In some embodiments, the splice site comprises or consists of a sequence having between 80% and 90% sequence identity to SEQ ID No: 2. In some embodiments, the splice site comprises or consists of a sequence having between 85% and 90% sequence identity to SEQ ID No: 2.

The consensus sequence of the 5′ splice site of the intron for U1 snRNP also serves as the coding sequence (CDS) of the *rep* gene. If the fourth base of the original RepCap sequence (i.e. a cytosine (C)), is changed to an adenine (A) to match the consensus sequence, a stop codon would enter the CDS and translation of Rep78 and Rep52 would stop. Therefore, in an embodiment, the splice site sequence does not comprise a stop codon.

Accordingly, in some embodiments, the splice site does not comprise or consist of a sequence having 100% sequence identity to SEQ ID No: 2.

In some embodiments, the splice site sequence which improves gene splicing efficiency is provided herein as SEQ ID No: 3, as follows:

GTACGT

[SEQ ID No: 3]

Alternatively, in an embodiment, the splice site sequence which improves gene splicing efficiency is provided herein as SEQ ID No: 4, as follows:

GAAAGT

[SEQ ID No: 4]

Alternatively, in an embodiment, the splice site sequence which improves gene splicing efficiency is provided herein as SEQ ID No: 5, as follows:

GCAAGT

[SEQ ID No: 5]

Alternatively, in an embodiment, the splice site sequence which improves gene splicing efficiency is provided herein as SEQ ID No: 6, as follows:

GGAAGT

[SEQ ID No: 6]

Alternatively, in an embodiment, the splice site sequence which improves gene splicing efficiency is provided herein as SEQ ID No: 7, as follows:

GTGAGT

[SEQ ID No: 7]

Alternatively, in an embodiment, the splice site sequence which improves gene splicing efficiency is provided herein as SEQ ID No: 8, as follows:

GTCAGT

[SEQ ID No: 8]

Alternatively, in an embodiment, the splice site sequence which improves gene splicing efficiency is provided herein as SEQ ID No: 9, as follows:

GTTAGT

[SEQ ID No: 9]

Accordingly, in some embodiments, the splice site sequence comprises or consists of a nucleotide sequence substantially as set out in any one of SEQ ID Nos: 3, 4, 5, 6, 7, 8 or 9, or a fragment or variant thereof.

In an embodiment, the splice site sequence comprises or consists of a nucleotide sequence substantially as set out in SEQ ID No: 3, or a fragment or variant thereof.

The inventors surprisingly discovered that the inclusion of a polyA signal sequence in the RepCap plasmid further improved the production yield of rAAV significantly. It is thought that the polyA signal sequence exists in the wild-type AAV genome (original RepCap DNA). However, it has been reported that there are mRNAs without the polyA tail in wild-type AAV (Wang, Lina et al. *The Journal of general virology* vol. 96,9 (2015): 2780-2787), and the efficacy of this polyA signal sequence of wild-type AAV is expected to be weak. Thus, the inventors inserted a polyA signal sequence after the *cap* gene, as illustrated in Figure 3.

The polyA signal sequence may be selected from a group consisting of: SV40 polyA signal sequence; human growth hormone polyA signal sequence; and beta-globin polyA signal sequence. In an embodiment, the polyA signal sequence is an SV40 polyA signal sequence.

In one embodiment, the polyA signal sequence is at least 10, 20, 30, 40, 50 or 60 nucleotides in length. In another embodiment, the polyA signal sequence is at least 70, 80, 90, 100, 100 or 120 nucleotides in length.

In one embodiment, the polyA signal sequence is less than 200, 190, 180 or 170 nucleotides in length. In another embodiment, the polyA signal sequence is less than 160, 150, 140 or 130 nucleotides in length.

In one embodiment, the polyA signal sequence is between 10 and 220, 20 and 210, 30 and 200, 40 and 190, or 50 and 180 nucleotides in length. In another embodiment,

the polyA signal sequence is between 60 and 170, 70 and 160, 80 and 150, 90 and 140, or 100 and 130 nucleotides in length.

In one embodiment, at least 10% of the nucleotides in the polyA signal sequence are adenine. In another embodiment, at least 20% of the nucleotides in the polyA signal sequence are adenine. In another embodiment, at least 30% of the nucleotides in the polyA signal sequence are adenine. In another embodiment, at least 35% of the nucleotides in the polyA signal sequence are adenine.

One embodiment of the SV40 polyA signal sequence, is provided herein as SEQ ID No: 10, as follows:

TAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAAT
TTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTT

[SEQ ID No: 10]

Accordingly, in an embodiment, the polyA signal sequence comprises or consists of a nucleotide sequence substantially as set out in SEQ ID No: 10, or a fragment or variant thereof.

As illustrated in Figure 3, the polyA signal sequence is inserted after the *cap* gene. Accordingly, in an embodiment, the polyA signal sequence is disposed 3′ of the *cap* gene.

In one embodiment, the AAV capsid (*Cap*) gene encodes three viral capsid proteins, which together form the outer capsid shell that protects the viral genome. In some embodiments, the AAV capsid (*Cap*) gene encodes VP1, VP2 and VP3.

In one embodiment, the AAV replication (*Rep*) gene encodes four proteins that are required for viral genome replication and packaging. Hence, in some embodiments, the AAV replication (*Rep*) gene encodes Rep78, Rep68, Rep52 and Rep40.

In one embodiment, the AAV capsid (*Cap*) gene encodes a capsid from an AAV of serotype 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, rh.10, 2.7m8, or a variant thereof (e.g., AAV8, AAV9, AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV10, AAV11, AAV12, AAVrh.10, or AAV2.7m8). In an embodiment, the AAV capsid is an AAV serotype 2 (AAV2) capsid, an AAV serotype 8 (AAV8) capsid, an AAV serotype 9 (AAV9) capsid, an AAV2.7m8 serotype capsid, or a modified capsid thereof. The terms

"variant thereof" and "modified capsid thereof", may be understood as a capsid that comprises or is encoded by a sequence comprising at least 70% sequence identity with the wildtype capsid sequence. In some embodiments, the "variant thereof" and "modified capsid thereof", may be a capsid that comprises or is encoded by a sequence comprising at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, or at least 99% sequence identity with the wildtype capsid sequence.

In another embodiment, the AAV replication (*Rep*) gene encodes a Rep protein from an AAV of serotype 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, rh.10, or a variant thereof (e.g., AAV8, AAV9, AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV10, AAV11, AAV12, or AAVrh.10). In an embodiment, the AAV Rep protein is an AAV serotype 2 (AAV2), an AAV serotype 8 (AAV8), or an AAV serotype 9 (AAV9). In an embodiment, the AAV Rep protein is an AAV serotype 2 (AAV2), or a modified capsid thereof. The terms "variant thereof" and "modified capsid thereof", may be understood as a capsid that comprises or is encoded by a sequence comprising at least 70% sequence identity with the wildtype capsid sequence. In some embodiments, the "variant thereof" and "modified capsid thereof", may be a capsid that comprises or is encoded by a sequence comprising at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, or at least 99% sequence identity with the wildtype capsid sequence.

As illustrated in the Examples and Figure 8, the AAV productivity for the AAV9 serotype was particularly high when using "Splicing activation" and "Splicing activation + SV40 polyA signal insertion" RepCap plasmids. For example, "Splicing activation" RepCap plasmids surprisingly increased AAV9 productivity 5.97-fold, and "Splicing activation + SV40 polyA signal insertion" RepCap plasmids increased AAV9 productivity 6.59-fold, which was even more unexpected.

Hence, in an embodiment, the AAV capsid (*Cap*) gene encodes a capsid from an AAV of serotype 9, and the AAV replication (*Rep*) gene encodes a Rep protein from an AAV of serotype 2. Alternatively, the AAV capsid (*Cap*) gene encodes a capsid from an AAV of serotype 9, and the AAV (*Rep*) gene encodes a Rep protein from an AAV of serotype 9.

Accordingly, in a second aspect of the invention, there is provided a recombinant nucleic acid molecule comprising an AAV capsid (*Cap*) gene and an AAV replication (*Rep*) gene, wherein the AAV capsid (*Cap*) gene encodes a capsid from an AAV

of serotype 9, and the AAV replication (*Rep*) gene encodes a Rep protein from an AAV of serotype 2 or 9, and wherein the recombinant nucleic acid molecule:

(i) comprises a splice site sequence which improves gene splicing efficiency, and/or

(ii) comprises a polyA signal sequence.

In an embodiment, the AAV replication (*Rep*) gene encodes a Rep protein from an AAV of serotype 2. In another embodiment, the AAV replication (*Rep*) gene encodes a Rep protein from an AAV of serotype 9.

As illustrated in the Examples and Figure 9, the AAV productivity for the AAV2.7m8 serotype was high when using "Splicing activation + SV40 polyA signal insertion", which indicates that this plasmid is effective not only for the wild-type AAV capsids but also for the modified AAV capsids. For example, "Splicing activation + SV40 polyA signal insertion" RepCap plasmids increased AAV2.7m8 productivity 1.56-fold.

Hence, in an embodiment, the AAV capsid (*Cap*) gene encodes a capsid from an AAV of serotype 2.7m8, and the AAV replication (*Rep*) gene encodes a Rep protein from an AAV of serotype 2.

Accordingly, in another aspect of the invention, there is provided a recombinant nucleic acid molecule comprising an AAV capsid (*Cap*) gene and an AAV replication (*Rep*) gene, wherein the AAV capsid (*Cap*) gene encodes a capsid from an AAV of serotype 2.7m8, and the AAV replication (*Rep*) gene encodes a Rep protein from an AAV of serotype 2, and wherein the recombinant nucleic acid molecule:

(i) comprises a splice site sequence which improves gene splicing efficiency, and/or

(ii) comprises a polyA signal sequence.

In some embodiments, the splice site sequence which improves gene splicing efficiency and the polyA signal sequence are as described above for the first aspect.

In one embodiment, the AAV capsid (*Cap*) gene encodes three viral capsid proteins, which together form the outer capsid shell that protects the viral genome. In some embodiments, the AAV capsid (*Cap*) gene encodes VP1, VP2 and VP3.

In one embodiment, the AAV replication (*Rep*) gene encodes four proteins that are required for viral genome replication and packaging. Hence, in some embodiments, the AAV replication (*Rep*) gene encodes Rep78, Rep68, Rep52 and Rep40.

In one embodiment, the recombinant nucleic acid molecule according to the first or second aspect, may further comprise one or more promoters. Preferably, the one or more promoter(s) is an AAV promoter, e.g., a P5, P19, and/or P40 promoter.

In some embodiments, the recombinant nucleic acid molecule comprises a P5 promoter, to produce Rep78 and Rep68. In some embodiments, the recombinant nucleic acid molecule comprises a P19 promoter, to produce Rep52 and Rep40. In some embodiments, the recombinant nucleic acid molecule comprises a P40 promoter to drive the *Cap* gene.

In one embodiment, the recombinant nucleic acid molecule may further comprise a coding sequence for a selectable marker. In one embodiment, the selectable marker is a protein conferring antibiotic resistance, for instance, a protein for kanamycin resistance (e.g., KanR), a protein for ampicillin resistance (e.g., AmpR), or a protein for puromycin resistance (e.g., Pac). In an embodiment, the recombinant nucleic acid molecule comprises a coding sequence for KanR.

Accordingly, in one embodiment, the nucleic acid molecule may further comprise a promoter driving expression of the selectable marker. In some embodiments, the promoter is the AmpR promoter.

In one embodiment, the recombinant nucleic acid molecule may further comprise at least one origin of replication (ori). In an embodiment, the recombinant nucleic acid molecule further comprises a plasmid origin of replication (ori).

The AAV capsid (*Cap*) and AAV replication (*Rep*) genes may be disposed on the recombinant nucleic acid molecule in any order from the 5′ to the 3′. For example, in one embodiment, the *Rep* gene is disposed 5′ of the *Cap* gene. Alternatively, in another embodiment, the *Cap* gene is disposed 5′ of the *Rep* gene. In some embodiments, the *Rep* gene is disposed 5′ of the *Cap* gene.

In an embodiment, the polyA signal sequence is disposed 3′ of the *Cap* gene. Accordingly, in an embodiment, the recombinant nucleic acid molecule may comprise, in this specified order, a *Rep* gene, a *Cap* gene, and a polyA signal sequence. The use

of 5′ and 3′ indicates that the features are either upstream or downstream, and is not intended to indicate that the features are necessarily terminal features. Alternatively, the recombinant nucleic acid molecule may comprise, in this specified order, a *Cap* gene, a *Rep* gene, and a polyA signal sequence.

It will be appreciated by the skilled person that the "Helper" plasmid may be combined with the "RepCap" plasmid, for a "double transfection" method. Accordingly, in some embodiments, the recombinant nucleic acid molecule according to the first and/or second aspect may comprise one or more adenoviral helper genes. The adenoviral helper genes may be selected from a group consisting of: E1A, E1B, E2A, E4, and VA RNA. In an embodiment, the recombinant nucleic acid molecule comprises an E2A gene. In an embodiment, the recombinant nucleic acid molecule comprises an E4 gene. In an embodiment, the recombinant nucleic acid molecule comprises a VA gene. In an embodiment, the recombinant nucleic acid molecule comprises at least two adenoviral helper genes selected from a group consisting of: E2A, E4 and VA. In an embodiment, the recombinant nucleic acid molecule comprises E2A, E4 and VA adenoviral helper genes.

Alternatively, it will be appreciated by the skilled person that the "Transgene" or "GOI" plasmid may be combined with the "RepCap" plasmid. Accordingly, in some embodiments, the recombinant nucleic acid molecule may comprise a transgene coding sequence. The recombinant nucleic acid molecule may further comprise a 5′-inverted terminal repeat (5′-ITR) sequence, a promoter, and a 3′-inverted terminal repeat (3′-ITR) sequence.

Alternatively, the "Transgene" or "GOI" plasmid, the "Helper" plasmid, and the "RepCap" plasmid, may be combined to form one plasmid. Accordingly, in some embodiments, the recombinant nucleic acid molecule according to the first and/or second aspect may comprise one or more adenoviral helper genes and a transgene coding sequence. In an embodiment, the recombinant nucleic acid molecule comprises one or more adenoviral helper genes and a 5′-inverted terminal repeat (5′-ITR) sequence, a promoter, a transgene coding sequence, and a 3′-inverted terminal repeat (3′-ITR) sequence. In an embodiment, the recombinant nucleic acid molecule comprises E2A, E4 and VA adenoviral helper genes, and a 5′-inverted terminal repeat (5′-ITR) sequence, a promoter, a transgene coding sequence, and a 3′-inverted terminal repeat (3′-ITR) sequence.

It will be appreciated that the recombinant nucleic acid molecule according to the first or second aspect may be contained within a vector, e.g. a RepCap plasmid.

Accordingly, in a third aspect of the invention, there is provided a vector comprising the recombinant nucleic acid molecule according to the first or second aspect.

The vector may be selected from a group consisting of: a plasmid, a cosmid, an episome, and a virus. In one embodiment, however, the vector is a plasmid.

In an embodiment, the vector is a RepCap plasmid, i.e. a plasmid comprising AAV Rep and Cap genes and from which AAV Rep and Cap proteins are expressed. In an embodiment, the vector has utility in a triple plasmid transfection for the production of rAAV.

The original RepCap plasmid is commercially available, and thus will be familiar to the skilled person. For example, the RepCap plasmid is available from Takara, as follows:

RepCap plasmid name	Takara Code
pRC1 Vector	6672
pRC2-mi342 Vector	6234
pRC5 Vector	6664
pRC6 Vector	6665
pRC8 Vector	6681

Additionally, the wild-type sequences of the Rep and Cap genes are disclosed in several databases, for example, in GenBank or NCBI Reference Sequence Database, with accession numbers as follows:

Gene	Accession No.
Adeno-associated virus - 2, complete genome	NC_001401.2
Adeno-associated virus - 8, complete genome	NC_006261.1
AAV9 VP1 (*cap*) gene	AY530579.1

As illustrated in the Examples, the RepCap plasmid according to the invention results in increased splicing efficiency and AAV productivity.

Accordingly, in an embodiment, the recombinant nucleic acid molecule according to the first or second aspect, or the vector according to the third aspect, increases AAV productivity. Increased AAV productivity may comprise increased vector genome (vg) titer. In an embodiment, the vector genome (vg) titer may increase by at least 1.2-fold, 1.3-fold, 1.4-fold, or 1.5-fold, compared to the original RepCap plasmid (e.g.

RepCap plasmids that do not comprise the splice site modification and polyA signal sequence). In an embodiment, the vector genome (vg) titer may increase by at least 1.6-fold, 1.7-fold, 1.8-fold, or 1.9-fold, compared to the original RepCap plasmid. In an embodiment, the vector genome (vg) titer may increase by at least 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, or 7-fold, compared to the original RepCap plasmid.

In another embodiment, the recombinant nucleic acid molecule according to the first or second aspect, or the vector according to the third aspect, increases splicing efficiency. As shown in Fig. 2A, *rep* and *cap* transcripts share the same intron, so splicing efficiency can be checked by the ratio of spliced and unspliced Rep protein. In an embodiment, splicing efficiency may be detected by Western blotting of the Rep protein. In an embodiment, low splicing efficiency occurs when the band of spliced small Rep protein (Rep40) is not detectable, or is detected in low levels compared to that of unspliced small Rep protein (Rep52). In an embodiment, high splicing efficiency occurs when the band of spliced small Rep protein (Rep40) is detectable, or is detected at high levels compared to that of unspliced small Rep protein (Rep52).

It will be appreciated that the recombinant nucleic acid molecule according to the first or second aspect, or the vector according to the third aspect, may be delivered, e.g. transfected, into a host cell.

Accordingly, in a fourth aspect of the invention, there is provided a host cell comprising the recombinant nucleic acid molecule according to the first or second aspect, or the vector according to the third aspect.

Any cell that is known in the art and allows production of rAAV through introduction of a construct can be selected, without limitation, as the host cell for use in the present invention. The host cell may be a bacteria, yeast, insect or mammalian cells. Examples of the host cell include animal cells (e.g., CHO cells, HEK293 cells, HeLa cells), insect cells (e.g., Sf9 cells), bacteria (such as Escherichia coli), and yeasts (Saccharomyces spp., Pichia spp.).

In one embodiment, the host cell may be a cell (or a cell line) appropriate for the production of rAAV. The host cell may be selected from a group consisting of: a HEK293, HeLa, Cos-7, A549, BHK, Vero, RD, ARPE-19, and MRC-5 cell. In an embodiment, the host cell is a HEK293 cell or a cell derived therefrom (e.g., Expi293F, VPC2.0, a HEK293T cell).

The host cell may further comprise a plasmid comprising one or more adenoviral helper genes (e.g., E1A, E1B, E2A, E4, and/or VA RNA, etc.) (referred to herein as a "helper" plasmid). Accordingly, in an embodiment, the host cell further comprises a "helper" plasmid. In one embodiment, the "Helper" plasmid comprises E2A, E4 and VA RNA genes.

Additionally, the host cell may further comprise a plasmid comprising a transgene coding sequence (i.e. a gene of interest or GOI). In an embodiment, the plasmid may comprise a 5′-inverted terminal repeat (5′-ITR) sequence, a promoter, a transgene coding sequence, and a 3′-inverted terminal repeat (3′-ITR) sequence (referred to herein as a "GOI" or "Transgene" plasmid). Accordingly, in an embodiment, the host cell further comprises a "Transgene" plasmid. In some embodiments, the "Transgene" plasmid comprises a 5′-inverted terminal repeat (5′-ITR) sequence, a promoter, a transgene coding sequence, and a 3′-inverted terminal repeat (3′-ITR) sequence.

In one embodiment, the "GOI" or "Transgene" plasmid may further comprise one or more sequence elements selected from an enhancer, an intron, and a polyadenylation signal.

As described in the Examples, the inventors have demonstrated that the RepCap plasmid according to the invention can be used in a method of producing high yields of rAAV.

Hence, according to a fifth aspect of the invention, there is provided use of the recombinant nucleic acid molecule according to the first or second aspect, or the vector according to the third aspect, to produce recombinant AAV (rAAV).

According to a sixth aspect of the invention, there is provided a method of producing recombinant AAV (rAAV), the method comprising:

(i) delivering the recombinant nucleic acid molecule according to the first or second aspect, or the vector according to the third aspect, to a host cell; and

(ii) culturing the host cell under suitable conditions to produce rAAV.

Any cell that is known in the art and allows production of rAAV through introduction of a construct can be selected, without limitation, as the host cell for use in the present invention. The host cell may be a bacterium, yeast, insect or mammalian cell. Examples of the host cell include animal cells (e.g., CHO cells, HEK293 cells, HeLa

cells), insect cells (e.g., Sf9 cells), bacteria (such as Escherichia coli), and yeasts (Saccharomyces spp., Pichia spp.).

In one embodiment, the host cell may be a cell (or a cell line) appropriate for the production of rAAV. The host cell may be selected from a group consisting of: a HEK293, HeLa, Cos-7, A549, BHK, Vero, RD, ARPE-19, and MRC-5 cell. In an embodiment, the host cell is a HEK293 cell or a cell derived therefrom (e.g., Expi293F, VPC2.0, a HEK293T cell).

The vector may be delivered to the host cell by any suitable method, including transfection, electroporation, liposome-based delivery, and membrane fusion techniques. In one embodiment, the vector is delivered to the host cell via transfection.

Step (i) of the method may further comprise delivering (e.g. transfecting) a "Helper" plasmid to the host cell. In some embodiments, the "Helper" plasmid comprises E2A, E4 and VA RNA.

Step (i) of the method may further comprise delivering (e.g. transfecting) a "GOI" or "Transgene" plasmid to the host cell. In some embodiments, the "Transgene" plasmid comprises a transgene coding sequence, and, for example may also comprise, a 5′-inverted terminal repeat (5′-ITR) sequence, a promoter, a transgene coding sequence, and a 3′-inverted terminal repeat (3′-ITR) sequence.

In one embodiment, the "GOI" or "Transgene" plasmid may further comprise one or more sequence elements selected from an enhancer, an intron, and a polyadenylation signal.

Step (i) of the method may comprise delivering (e.g. transfecting): (a) the recombinant nucleic acid molecule according to the first or second aspect, or the vector according to the third aspect; (b) a "Helper" plasmid; and (c) a "GOI" or "Transgene" plasmid, to the host cell. It will be appreciated by the skilled person that this is a "triple transfection" method.

Alternatively, it will be appreciated by the skilled person that the "Helper" plasmid may be combined with the "RepCap" plasmid, for a "double transfection" method. Accordingly, step (i) of the method may comprise delivering (e.g. transfecting): (a) the recombinant nucleic acid molecule according to the first or second aspect, or the

vector according to the third aspect (e.g. the combined "Helper" and "RepCap" plasmid); and (b) a "GOI" or "Transgene" plasmid.

Accordingly, in this embodiment, the recombinant nucleic acid molecule according to the first and/or second aspect, or the vector according to the third aspect, may comprise one or more adenoviral helper genes. The adenoviral helper genes may be selected from a group consisting of: E1A, E1B, E2A, E4, and VA RNA. In an embodiment, the recombinant nucleic acid molecule or vector comprises an E2A gene. In an embodiment, the recombinant nucleic acid molecule or vector comprises an E4 gene. In an embodiment, the recombinant nucleic acid molecule or vector comprises a VA gene. In an embodiment, the recombinant nucleic acid molecule or vector comprises at least two adenoviral helper genes selected from a group consisting of: E2A, E4 and VA. In an embodiment, the recombinant nucleic acid molecule or vector comprises E2A, E4 and VA adenoviral helper genes.

Alternatively, it will be appreciated by the skilled person that the "Transgene" or "GOI" plasmid may be combined with the "RepCap" plasmid. Accordingly, step (i) of the method may comprise delivering (e.g. transfecting): (a) the recombinant nucleic acid molecule according to the first or second aspect, or the vector according to the third aspect (e.g. the combined Transgene" or "GOI" plasmid and "RepCap" plasmid); and (b) a "Helper" plasmid.

Accordingly, in this embodiment, the recombinant nucleic acid molecule or vector may comprise a transgene coding sequence. The recombinant nucleic acid molecule or vector may further comprise a 5′-inverted terminal repeat (5′-ITR) sequence, a promoter, and a 3′-inverted terminal repeat (3′-ITR) sequence.

Alternatively, the "Transgene" or "GOI" plasmid, the "Helper" plasmid, and the "RepCap" plasmid, may be combined to form one plasmid, used in a single transfection method. Accordingly, in this embodiment, the recombinant nucleic acid molecule or vector may comprise one or more adenoviral helper genes and a transgene coding sequence. In an embodiment, the recombinant nucleic acid molecule or vector comprises one or more adenoviral helper genes and a 5′-inverted terminal repeat (5′-ITR) sequence, a promoter, a transgene coding sequence, and a 3′-inverted terminal repeat (3′-ITR) sequence. In an embodiment, the recombinant nucleic acid molecule or vector comprises E2A, E4 and VA adenoviral helper genes, and a 5′-inverted terminal repeat (5′-ITR) sequence, a promoter, a transgene coding sequence, and a 3′-inverted terminal repeat (3′-ITR) sequence.

In some embodiments, the method further comprises a step (iii) of: collecting a culture solution from the host cell and/or a lysate of the host cell. The lysate can be obtained, for example, by treating the host cell with a surfactant or an ultrasonic wave.

In some embodiments, the method further comprises a step (iv) of: purifying the rAAV from the culture solution and/or lysate. To purify the rAAV vector from the lysate, for example, ion-exchange chromatography and/or hydrophobic interaction chromatography, caesium chloride density-gradient centrifugation, sucrose gradient centrifugation, iodixanol density-gradient centrifugation, ultrafiltration, diafiltration, affinity chromatography, polyethylene glycol precipitation, and ammonium sulfate precipitation may be used.

It will be appreciated that the invention extends to any nucleic acid or peptide or variant, derivative or analogue thereof, which comprises substantially the amino acid or nucleic acid sequences of any of the sequences referred to herein, including variants or fragments thereof. The terms "substantially the amino acid/nucleotide/peptide sequence", "variant" and "fragment", can be a sequence that has at least 40% sequence identity with the amino acid/nucleotide/peptide sequences of any one of the sequences referred to herein, for example 40% identity with the sequence identified as SEQ ID No:1-12, and so on.

Amino acid/polynucleotide/polypeptide sequences with a sequence identity which is greater than 65%, in some embodiments, greater than 70%, in some embodiments, greater than 75%, and in some embodiments, greater than 80% sequence identity to any of the sequences referred to are also envisaged. In some embodiments, the amino acid/polynucleotide/polypeptide sequence has at least 85% identity with any of the sequences referred to, in some embodiments at least 90% identity, in some embodiments at least 92% identity, in some embodiments at least 95% identity, in some embodiments at least 97% identity, in some embodiments at least 98% identity and, in some embodiments at least 99% identity with any of the sequences referred to herein.

The skilled technician will appreciate how to calculate the percentage identity between two amino acid/polynucleotide/polypeptide sequences. In order to calculate the percentage identity between two amino acid/polynucleotide/polypeptide sequences, an alignment of the two sequences must first be prepared, followed by calculation of the

sequence identity value. The percentage identity for two sequences may take different values depending on:- (i) the method used to align the sequences, for example, ClustalW, BLAST, FASTA, Smith-Waterman (implemented in different programs), or structural alignment from 3D comparison; and (ii) the parameters used by the alignment method, for example, local vs global alignment, the pair-score matrix used (e.g. BLOSUM62, PAM250, Gonnet etc.), and gap-penalty, e.g., functional form and constants.

Having made the alignment, there are many different ways of calculating percentage identity between the two sequences. For example, one may divide the number of identities by: (i) the length of shortest sequence; (ii) the length of alignment; (iii) the mean length of sequence; (iv) the number of non-gap positions; or (iv) the number of equivalenced positions excluding overhangs. Furthermore, it will be appreciated that percentage identity is also strongly length dependent. Therefore, the shorter a pair of sequences is, the higher the sequence identity one may expect to occur by chance.

Hence, it will be appreciated that the accurate alignment of protein or DNA sequences is a complex process. The popular multiple alignment program ClustalW (Thompson *et al.*, 1994, Nucleic Acids Research, 22, 4673-4680; Thompson *et al.*, 1997, Nucleic Acids Research, 24, 4876-4882) is one way for generating multiple alignments of proteins or DNA in accordance with the invention. Suitable parameters for ClustalW may be as follows: For DNA alignments: Gap Open Penalty = 15.0, Gap Extension Penalty = 6.66, and Matrix = Identity. For protein alignments: Gap Open Penalty = 10.0, Gap Extension Penalty = 0.2, and Matrix = Gonnet. For DNA and Protein alignments: ENDGAP = -1, and GAPDIST = 4. Those skilled in the art will be aware that it may be necessary to vary these and other parameters for optimal sequence alignment.

In some embodiments, calculation of percentage identities between two amino acid/polynucleotide/polypeptide sequences may then be calculated from such an alignment as (N/T)*100, where N is the number of positions at which the sequences share an identical residue, and T is the total number of positions compared including gaps but excluding overhangs. In some embodiments, overhangs are included in the calculation. Hence, one method for calculating percentage identity between two sequences comprises (i) preparing a sequence alignment using the ClustalW program using a suitable set of parameters, for example, as set out above; and (ii) inserting the values of N and T into the following formula:- Sequence Identity = (N/T)*100.

Alternative methods for identifying similar sequences will be known to those skilled in the art. For example, a substantially similar nucleotide sequence will be encoded by a sequence which hybridizes to DNA sequences or their complements under stringent conditions. By stringent conditions, we mean the nucleotide hybridizes to filter-bound DNA or RNA in 3x sodium chloride/sodium citrate (SSC) at approximately 45ºC followed by at least one wash in 0.2x SSC/0.1% SDS at approximately 20-65ºC. Alternatively, a substantially similar polypeptide may differ by at least 1, but less than 5, 10, 20, 50 or 100 amino acids from the sequences shown in, for example, SEQ ID Nos: 1-12.

Due to the degeneracy of the genetic code, it is clear that any nucleic acid sequence described herein could be varied or changed without substantially affecting the sequence of the protein encoded thereby, to provide a functional variant thereof. Suitable nucleotide variants are those having a sequence altered by the substitution of different codons that encode the same amino acid within the sequence, thus producing a silent change. Other suitable variants are those having homologous nucleotide sequences but comprising all, or portions of, sequence, which are altered by the substitution of different codons that encode an amino acid with a side chain of similar biophysical properties to the amino acid it substitutes, to produce a conservative change. For example small non-polar, hydrophobic amino acids include glycine, alanine, leucine, isoleucine, valine, proline, and methionine. Large non-polar, hydrophobic amino acids include phenylalanine, tryptophan and tyrosine. The polar neutral amino acids include serine, threonine, cysteine, asparagine and glutamine. The positively charged (basic) amino acids include lysine, arginine and histidine. The negatively charged (acidic) amino acids include aspartic acid and glutamic acid. It will therefore be appreciated which amino acids may be replaced with an amino acid having similar biophysical properties, and the skilled technician will know the nucleotide sequences encoding these amino acids.

All of the features described herein (including any accompanying claims, abstract and drawings), and/or all of the steps of any method or process so disclosed, may be combined with any of the above aspects in any combination, except combinations where at least some of such features and/or steps are mutually exclusive.

For a better understanding of the invention, and to show how embodiments of the same may be carried into effect, reference will now be made, by way of example, to the accompanying Figures, in which:-

Figure 1 shows the triple-plasmid transfection of the HEK293 cell, with pHelper, pGOI and pRepCap, for rAAV production. The pGOI plasmid is an inverted terminal repeat (ITR)-containing plasmid, with the *rep* and *cap* genes replaced by the promoter and gene of interest (GOI). The pHelper plasmid provides the minimal gene products required for AAV replication, e.g. E2A, E4 and VA. The RepCap plasmid contains *rep* and *cap* genes of AAV for encapsidation.

Figure 2 shows (A) an illustration of the *rep* and *cap* genes and their major transcripts. The promoters for expression of each transcript are described on the left (p5, p19 and p40). Fold lines show the intron of each transcript. The 5′ splice site of the intron, where the base substitutions were conducted for producing the "Splicing activation RepCap plasmids" (i.e. the plasmids according to the invention), is shown by a circle. (B) Sequences of the 5′ splice site. "Consensus" shows the consensus sequence of U1 snRNA binding site. "Original pRepCap" shows the sequence of original RepCap plasmids (pRC-AAV8, pRC-AAV2 and pRC-AAV9). Bases that differ from the consensus sequence are shown in bold. "Splicing activation" shows the sequence used in splicing activation plasmids (pRC-AAV8_bDsp, pRC-AAV8_bDsp_SV40pA, pRC-AAV2_bDsp_SV40pA and pRC-AAV9_bDsp_SV40pA), i.e. the plasmids according to the invention. The substituted bases from original RepCap plasmids are described in italic bold.

Figure 3 shows an illustration of the recombinant nucleic acid molecule according to the invention encoding a *rep* and *cap* gene, and comprising a SV40 polyA signal sequence. The SV40 polyA signal sequence is inserted under the coding sequence (CDS) of cap gene.

Figure 4 illustrates the effect of the "Splicing activation RepCap plasmid" (i.e. the plasmid according to the invention) on AAV productivity improvement. The relative vector genome (vg) titer when using FectoVIR-AAV process are described. Plasmid X was used as a transgene plasmid, pRC-AAV8, pRC-AAV8_bDsp and pRC-AAV8_cDsp were used as RepCap plasmids in Original, Splicing activation and Consensus conditions, respectively. Expi293F cells (A) and VPC2.0 (B) were used in this experiment, respectively.

Figure 5 shows Western blot analysis of small Rep (Rep52 and Rep40). Extracts of Expi293F cells were obtained on Day 1 following transfection by FectorVIR-AAV and were analysed by Western blotting using an anti-Rep antibody. Plasmid X was used as

a transgene plasmid, pRC-AAV8 and pRC-AAV8_bDsp were used as RepCap plasmids in Original and Splicing activation conditions, respectively.

Figure 6 illustrates the effect of "Splicing activation", "SV40 polyA signal insertion" and their combination RepCap plasmid on AAV productivity improvement. The relative vector genome (vg) titer when using FectoVIR-AAV process are described. Plasmid X was used as a transgene plasmid, pRC-AAV8, pRC-AAV8_bDsp, pRC-AAV8_SV40pA and pRC-AAV8_bDsp_SV40pA were used as RepCap plasmids in Original, Splicing activation, SV40 polyA signal insertion, and Splicing activation + SV40 polyA signal insertion conditions, respectively. Expi293F cells (A) and VPC2.0 (B) were used in this experiment, respectively.

Figure 7 illustrates the productivity improvement effect of modified RepCap plasmids in AAV-MAX transfection process. The relative vector genome (vg) titer when using AAV-MAX transfection process are described. Plasmid X was used as a transgene plasmid, pRC-AAV8, pRC-AAV8_bDsp and pRC-AAV8_bDsp_SV40pA were used as RepCap plasmids in Original, Splicing activation and Splicing activation + SV40 polyA signal insertion conditions, respectively. VPC2.0 were used as a production cell in this experiment.

Figure 8 shows the productivity improvement effect of modified RepCap plasmids in various serotypes. The relative vector genome (vg) titer when using AAV-MAX transfection process are described. Plasmid Y was used as a transgene plasmid, pRC-AAV2/8/9, pRC-AAV2/8/9_bDsp and pRC-AAV2/8/9_bDsp_SV40pA were used as RepCap plasmids in Original, Splicing activation and Splicing activation + SV40 polyA signal insertion conditions, respectively. VPC2.0 were used as a production cell in this experiment.

Figure 9 shows the productivity improvement effect of modified RepCap plasmids for AAV2.7m8 capsid. The relative vector genome (vg) titer when using AAV-MAX transfection process are described. Plasmid Z was used as a transgene plasmid, pRC-7m8 and pRC-7m8_bDsp_SV40pA were used as RepCap plasmids in Original and Splicing activation + SV40 polyA signal insertion conditions, respectively. VPC2.0 were used as a production cell in this experiment.

Figure 10 shows a plasmid map of pRC-AAV8 comprising the *rep* and *cap* genes with the original splice site sequence. pRC-AAV8 further comprises a protein for kanamycin resistance (KanR), an AmpR promoter, and an ori and f1 ori.

Figure 11 shows a plasmid map of pRC-AAV8_bDsp comprising the *rep* and *cap* genes with the modified splice site sequence (CA to GT). pRC-AAV8_bDsp further comprises a protein for kanamycin resistance (KanR), an AmpR promoter, and an ori and f1 ori.

Figure 12 shows a plasmid map of pRC-AAV8_cDsp comprising the *rep* and *cap* genes with the consensus splice site sequence (CCA to AGT). pRC-AAV8_cDsp further comprises a protein for kanamycin resistance (KanR), an AmpR promoter, and an ori and f1 ori.

Figure 13 shows a plasmid map of pRC-AAV8_SV40pA comprising the *rep* and *cap* genes and an SV40 polyA signal sequence. pRC-AAV8_SV40pA further comprises a protein for kanamycin resistance (KanR), an AmpR promoter, and an ori and f1 ori.

Figure 14 shows a plasmid map of pRC-AAV8_bDsp_SV40pA comprising the *rep* and *cap* genes with the modified splice site sequence (CA to GT) and an SV40 polyA signal sequence. pRC-AAV8_bDsp_SV40pA further comprises a protein for kanamycin resistance (KanR), an AmpR promoter, and an ori and f1 ori.

Figure 15 shows a plasmid map of pRC-AAV2 comprising the *rep* and *cap* genes with the original splice site sequence. pRC-AAV2 further comprises a protein for kanamycin resistance (KanR), an AmpR promoter, and an ori and f1 ori.

Figure 16 shows a plasmid map of pRC-AAV2_bDsp comprising the *rep* and *cap* genes with the modified splice site sequence (CA to GT). pRC-AAV2_bDsp further comprises a protein for kanamycin resistance (KanR), an AmpR promoter, and an ori and f1 ori.

Figure 17 shows a plasmid map of pRC-AAV2_bDsp_SV40pA comprising the *rep* and *cap* genes with the modified splice site sequence (CA to GT) and an SV40 polyA signal sequence. pRC-AAV2_bDsp_SV40pA further comprises a protein for kanamycin resistance (KanR), an AmpR promoter, and an ori and f1 ori.

Figure 18 shows a plasmid map of pRC-AAV9 comprising the *rep* and *cap* genes with the original splice site sequence. pRC-AAV9 further comprises a protein for kanamycin resistance (KanR), an AmpR promoter, and an ori and f1 ori.

Figure 19 shows a plasmid map of pRC-AAV9_bDsp comprising the *rep* and *cap* genes with the modified splice site sequence (CA to GT). pRC-AAV9_bDsp further comprises a protein for kanamycin resistance (KanR), an AmpR promoter, and an ori and f1 ori.

Figure 20 shows a plasmid map of pRC-AAV9_bDsp_SV40pA comprising the *rep* and *cap* genes with the modified splice site sequence (CA to GT) and an SV40 polyA signal sequence. pRC-AAV9_bDsp_SV40pA further comprises a protein for kanamycin resistance (KanR), an AmpR promoter, and an ori and f1 ori.

Figure 21 shows a plasmid map of pRC-7m8 comprising the *rep* and *cap* genes with the original splice site sequence. pRC-7m8 further comprises a protein for kanamycin resistance (KanR), an AmpR promoter, and an ori and f1 ori.

Figure 22 shows a plasmid map of pRC-7m8_bDsp_SV40pA comprising the *rep* and *cap* genes with the modified splice site sequence (CA to GT) and an SV40 polyA signal sequence. pRC-AAV9_bDsp_SV40pA further comprises a protein for kanamycin resistance (KanR), an AmpR promoter, and an ori and f1 ori.

Examples

The inventors set out to evaluate two hypotheses: (1) Cap protein expression is the rate-limiting factor of rAAV production in the cell, and (2) the very low efficiency of splicing of *cap* gene transcripts reduces the cap protein expression. Therefore, the inventors hypothesised that improving the amount of spliced transcripts of the *cap* gene would improve the expression of cap protein expression, which would lead to an increase in rAAV production. To confirm the hypothesis and improve AAV productivity, new modified RepCap plasmids were constructed and used for evaluation studies.

Materials and Methods

RepCap plasmid construction

RepCap plasmids used in the examples are listed in Table 1.

Table 1

Plasmid name	Serotype	Description
pRC-AAV8	AAV8	Original
pRC-AAV8_bDsp	AAV8	Splicing activation
pRC-AAV8_cDsp	AAV8	Consensus
pRC-AAV8_SV40pA	AAV8	SV40 polyA signal insertion
pRC-AAV8_bDsp_SV40pA	AAV8	Splicing activation + SV40 polyA signal insertion

pRC-AAV2	AAV2	Original
pRC-AAV2_bDsp	AAV2	Splicing activation
pRC-AAV2_bDsp_SV40pA	AAV2	Splicing activation + SV40 polyA signal insertion
pRC-AAV9	AAV9	Original
pRC-AAV9_bDsp	AAV9	Splicing activation
pRC-AAV9_bDsp_SV40pA	AAV9	Splicing activation + SV40 polyA signal insertion
pRC-7m8	AAV2.7m8	Original
pRC-7m8_bDsp_SV40pA	AAV2.7m8	Splicing activation + SV40 polyA signal insertion

Original RepCap plasmids (pRC-AAV8, pRC-AAV2, pRC-AAV9 and pRC-7m8) have AAV2 rep genes and cap genes for each serotype (AAV8, AAV2, AAV9 and AAV2.7m8, respectively). "Splicing activation" RepCap plasmids (pRC-AAV8_bDsp, pRC-AAV2_bDsp and pRC-AAV9_bDsp) were constructed by making two base substitutions of the 5' splice site of the *rep* and *cap* intron from "GTACCA" (SEQ ID No: 1) to "GTACGT" (SEQ ID No: 3) in original RepCap plasmids. Consensus RepCap plasmid (pRC-AAV8_cDsp) was constructed by making three base substitutions of the 5' splice site of *rep* and *cap* intron from "GTACCA" (SEQ ID No: 1) to "GTAAGT" (SEQ ID No: 2) in original RepCap plasmids. SV40 polyA signal insertion plasmid (pRC-AAV8_SV40pA) was constructed by inserting the SV40 polyA signal sequence (SEQ ID No: 10: 5'-TAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTT-3') downstream to the CDS region of the *cap* gene in original RepCap plasmids. "Splicing activation + SV40 polyA signal insertion" plasmids (pRC-AAV8_bDsp_SV40pA, pRC-AAV2_bDsp_SV40pA, pRC-AAV9_bDsp_SV40pA and pRC-7m8_bDsp_SV40pA) were constructed by incorporating both modifications described above in original RepCap plasmids.

rAAV production by FectoVIR-AAV

Expi293 Cells (Thermo Fisher Scientific, catalog No. A14527) or Viral Production Cells 2.0 (VPC2.0, Thermo Fisher Scientific, catalog No. A49784) in Viral Production Medium (VPM, Thermo Fisher Scientific, catalog No. A4817901) containing 4% GlutaMAX™ Supplement (GultaMAX, Thermo Fisher Scientific, catalog No. 35050061) were diluted to a density of 38 x 10^5 cells/mL using VPM containing 4% GultaMAX. 15-20 mL of the cell suspension was put into the 125mL shaker flask. The transgene plasmid X, the helper plasmid which has E2 and E4 gene and VA1 RNA coding region of Adenovirus V, and the RepCap plasmid were used as transfected DNA for a triple transfection system. The plasmids were added into FreeStyle™ 293 Expression Medium (Thermo Fisher Scientific, catalog No. 12338-018). FevtoVIR-AAV (PolyPlus, catalog No. 120-010) was added to the DNA solution. 60 minutes later, the complex was added into

the flask. 3 days after the transfection, Triton X-100 was added to the culture medium at a concentration of 0.2%, which was left to stand at room temperature for 1 hour to lyse the cells. The resulting solution was centrifuged at 10,000xg for 5 min at room temperature. The supernatant was collected as the rAAV sample.

rAAV production by AAV-MAX transfection system

Viral Production Cells 2.0 (VPC2.0, Thermo Fisher Scientific, catalog No. A49784) in Viral Production Medium (VPM, Thermo Fisher Scientific, catalog No. A4817901) containing 4% GlutaMAX™ Supplement (GultaMAX, Thermo Fisher Scientific, catalog No. 35050061) were diluted to a density of 4 x 10^5 cells/mL using VPM containing 4% GultaMAX and cultivated. When the target density of 3.8±0.3 E+6 cells/mL (3.5 – 4.1E+6 cells/mL range) was reached (approximately 3 days), Gibco AAV-MAX Enhancer, one of the components of AAV-MAX Transfection kit (Thermo Fisher Scientific, catalog No. A50515), was added into the cell culture. The transgene plasmid X or plasmid Y, the helper plasmid which has E2 and E4 genes and VA1 RNA coding region of Adenovirus V, and the RepCap plasmid were used as transfected DNA for a triple transfection system. The remaining two reagents of AAV-MAX transfection kit were added to the buffer in the order described in the kit protocol. 10 minutes later, the complex was added into the flask. 2 or 3 days after the transfection, Triton X-100 was added to the culture medium at a concentration of 0.2%, which was left to stand at room temperature for 1 hour to lyse the cells. In several serotypes, NaCl was also added along with Triton X-100. The resulting solution was centrifuged at 10,000xg for 5 min at room temperature. The supernatant was collected as the rAAV sample.

Western blotting

200 µL of cell culture 1 day after transfection was sampled and cultured cells were collected by centrifuge. After removing the supernatant, collected cells were resuspended by RIPA Lysis and Extraction Buffer (Thermo Fisher Scientific, catalog No. 89901) containing Halt™ Protease and Phosphatase Inhibitor Cocktail, EDTA-free (100X) (Thermo Fisher Scientific, catalog No. 78441) and BENZONASE (Merck, catalog No. 1.01656.0001). The cell suspension was heated at 95°C for 10 min. The heated sample was mixed with NuPAGE LDS Sample Buffer (4×) (Thermo Fisher Scientific, catalog No. NP0007) and NuPAGE Reducing Agent (10X) (Thermo Fisher Scientific, catalog No. NP0004), and were used as a loading sample SDS-PAGE. SDS-PAGE was conducted by using Xcell SureLock (Thermo Fisher Scientific, catalog No. EI0001), Bolt 4-12% Bis-Tris Plus Gels, 17-well (Thermo Fisher Scientific, catalog No. NW04122BOX), and NuPAGE MES SDS Running Buffer (20X) (Thermo Fisher Scientific, catalog No. NP0002) according to manufacturer's protocol. Membrane transfer was

conducted by iBlot Dry Blotting System (Thermo Fisher Scientific, catalog No. IB1001) and iBlot Gel Transfer Stacks PVDF, Mini (Thermo Fisher Scientific, catalog No. IB401002) according to manufacturer's protocol. Membrane blocking, wash and antibody incubation were conducted by iBind Western Device (Thermo Fisher Scientific, catalog No. SLF1000), iBind Solution Kit (Thermo Fisher Scientific, catalog No. SLF1020) and iBind cards (Thermo Fisher Scientific, catalog No. SLF1010) according to manufacturer's protocol. Anti-Adeno-Associated Virus 2 (AAV) (Replicase Rep 78,68,52,40), Mouse-Mono (303.9) (Origene, catalog No. AM09104PU-N) and anti-IgG Mouse, Goat-poly, HRP (RSD, catalog No. HAF007) were used as 1st and 2nd antibody, respectively. The membrane was treated by SuperSignal West Dura Extended Duration Substrate (Thermo Fisher Scientific, catalog No. 34075) and the signal was detected by Amersham Imager 680 (Cytiva).

Vector genome (vg) titer quantification

rAAV samples were diluted by Ringer's lactate solution (Terumo, catalog No. TP-AB05NR) with Pluronic F-68 Non-ionic Surfactant (100X) (Thermo Fisher Scientific, catalog No. 24040032) and treated with DNase I (New England Biolabs, catalog No. M0303S). The solution was further diluted, and digital droplet PCR (QX ONE Droplet Digita PCR system, Bio-Rad Laboratories, Inc.) was then performed using an ddPCR Supermix for Probes (no dUTP) (Bio-Rad Laboratories, catalog No. 1863024), FAM probe and primers designed for transgene or ITR construct. The vg titer concentrations of the supernatant were calculated by multiplying back the dilution rate.

Results and Discussion

Example 1 – Modification of 5′ splice site

Splicing involves many components, and U1 small nuclear RNA (U1 snRNA) is one of them. It is known that U1 snRNA anneals to the 5' splice site of the intron and is essential to initiate splicing. The structure of the *rep* and *cap* genes and their major transcripts, and the 5′ splice site in RepCap plasmids are illustrated in Fig.2A. The consensus sequence for the 5' splice site recognition of the u1 snRNA and the 5′ splice site sequence of original RepCap plasmids are shown in Fig. 2B. As shown in the Figure, the 5' splice site of the original RepCap plasmids differs from the consensus sequence in the last three bases. Thus, it was hypothesised that splicing efficiency would be improved by changing the 5′ splice site sequence so it is closer to the consensus sequence. However, this sequence also serves as the CDS of the *rep* gene,

and if the 4th C is changed to an A, a stop codon will enter the CDS and translation of Rep78 and Rep52 will stop.

Rep proteins are also important for AAV production, and it was thought that the nonsense mutation might reduce productivity. Therefore, two new RepCap plasmids were constructed, one in which two bases were modified from the original to get closer to the consensus sequence as shown in Fig. 2B "Splicing activation" (plasmid name: pRC-AAV8_bDsp), and the other in which the 5' splice site sequence matched exactly to the consensus sequence (plasmid name: pRC-AAV8_cDsp), and they were evaluated for AAV productivity.

The results of the AAV production study are shown in Fig.4. This study was conducted by FectoVIR-AAV process, and Expi293F (A) and VPC2.0 (B) were used as production cells. As shown in Fig.4, using "Splicing activation" RepCap plasmid resulted in higher viral productivity compared to the original RepCap plasmid, with increases of 1.31- and 1.68-fold in Expi293 cells and VPC2.0, respectively.

Splicing efficiency was confirmed by Western blotting of the Rep protein. As shown in Fig. 2A, *rep* and *cap* transcripts share the same intron, so splicing efficiency can be checked by the ratio of spliced and unspliced Rep protein. As illustrated in Fig. 5, in the original RepCap plasmid, the band of spliced small Rep protein (Rep40) was hardly detectable compared to that of unspliced small Rep protein (Rep52), implying low splicing efficiency. On the other hand, "Splicing activation" plasmid exhibited an increase in Rep40 band intensity, indicating a splicing efficiency increase. These results indicate that the 5' splice site sequence modification described in Fig.2B increases splicing efficiency and improves AAV productivity.

On the contrary, using the consensus RepCap plasmid caused AAV productivity to decrease compared to original RepCap plasmid, with 0.61- and 0.39-fold in Expi293 cells and VPC2.0, respectively (see Fig.4). As described above, because the stop codon is inserted into the CDS of Rep78 and Rep52, AAV productivity may be lower than using the original RepCap condition.

Example 2 – Insertion of the SV40 polyA signal sequence (SV40pA)

To further improve AAV productivity, another RepCap plasmid was constructed. Stability of mRNA was also considered important for increasing spliced transcripts. This is because when mRNA stability is poor, spliced transcripts are rapidly degraded and reduced, whereas when mRNA stability is high, spliced transcripts remain longer

and are expressed at higher levels. There are several known factors involved in mRNA stability, and the addition of polyadenylation (polyA) is one of them. It is known that a polyA signal sequence exists under the *cap* gene in original RepCap plasmids. However, the sequence was taken from the wild-type AAV genome, which is not optimised for rAAV production, and there is room for improvement. Thus, RepCap plasmids with inserted polyA signal sequences were constructed and evaluated for AAV production.

The illustration of the construct is shown in Fig. 3. The polyadenylation signal sequence of simian virus 40 (SV40 polyA signal sequence) was used for these newly constructed RepCap plasmids. As shown in Fig. 3, the SV40 polyA signal sequence was inserted under the CDS of the *cap* gene in the original pRepCap (plasmid name: pRC-AAV8) and the "Splicing activation" RepCap plasmid (plasmid name: pRC-AAV8_bDsp). The resultant plasmids were called "SV40 polyA signal insertion" RepCap plasmid (plasmid name: pRC-AAV8_SV40pA) and "Splicing activation + SV40 polyA signal insertion" RepCap plasmid (plasmid name: pRC-AAV8_bDsp_SV40pA), respectively.

The results of the AAV production study are shown in Fig.6. This study was conducted by FectoVIR-AAV process, and Expi293F (A) cells and VPC2.0 (B) were used as production cells. As shown in Fig.6, using these newly constructed RepCap plasmids resulted in higher viral productivity compared to the original RepCap plasmid. By using "SV40 polyA signal insertion" RepCap plasmid, AAV productivity increased 1.34- and 1.58-fold in Expi293 cells and VPC2.0, respectively, compared to original RepCap condition. In addition, the use of "Splicing activation + SV40 polyA signal insertion" RepCap plasmid resulted in higher productivity than any of the individual modifications alone, suggesting synergistic activity, with AAV productivity increasing by 1.57- and 1.73-fold in Expi293 cells and VPC2.0, respectively, compared to original RepCap condition.

To further confirm the efficacy of the modified RepCap plasmids, evaluation studies of AAV production was performed using a different process. The AAV production study using AAV-MAX transfection system was conducted and the results are shown in Fig. 7. As with the FectoVIR process, AAV productivity improved by using "Splicing activation" RepCap plasmid, with a 1.53-fold increase compared to the original RepCap plasmid. Moreover, "Splicing activation + SV40 polyA signal insertion" RepCap plasmid resulted in further AAV productivity improvement, with a 1.93-fold increase compared to the original RepCap plasmid.

From a different perspective, evaluations of modified RepCap plasmids were also conducted for different serotypes. "Splicing activation" and "Splicing activation + SV40 polyA signal insertion" RepCap plasmids for AAV2 and AAV9 were constructed in the same way as RepCap plasmids for AAV8 and used for the evaluation study. Transgene Plasmid Y was used as a transgene plasmid, and the process by AAV-max transfection system was used in the evaluation study. As shown in Fig. 8, the AAV productivity of "Splicing activation" RepCap plasmids was higher than that of the original RepCap plasmids, and the productivity of "Splicing activation + SV40 polyA signal insertion" RepCap plasmids was even higher in all serotypes. Using "Splicing activation + SV40 polyA signal insertion" RepCap plasmids resulted in higher viral productivity compared to original RepCap plasmid, with increases of 2.15-, 2.17-, and 6.59-fold in AAV2, AAV8 and AAV9, respectively.

For further confirmation of efficacy of modified RepCap plasmids, evaluations of modified RepCap plasmids were also conducted for AAV2.7m8 serotype. "Splicing activation + SV40 polyA signal insertion" RepCap plasmids for AAV2.7m8 was constructed in the same way as RepCap plasmids for AAV8 and used for the evaluation study. Transgene Plasmid Z was used as a transgene plasmid, and the process by AAV-max transfection system was used in the evaluation study. As shown in Fig. 9, using "Splicing activation + SV40 polyA signal insertion" RepCap plasmid resulted in higher viral productivity compared to original RepCap plasmid, with an increase of 1.56-fold.

Therefore, these results suggest that the RepCap plasmids according to the invention can improve AAV productivity in a wide range of processes and serotypes. In particular, the results demonstrate that the modification of the *rep*/*cap* splice site sequence to improve gene splicing efficiency, and the inclusion of the polyA signal sequence, have a synergistic effect, resulting in significantly increased splicing efficiency and AAV productivity.

Claims

1. A recombinant nucleic acid molecule comprising an AAV capsid (*Cap*) gene and an AAV replication (*Rep*) gene, wherein the recombinant nucleic acid molecule comprises a splice site sequence which improves gene splicing efficiency, and a polyA signal sequence.

2. A recombinant nucleic acid molecule comprising an AAV capsid (*Cap*) gene and an AAV replication (*Rep*) gene, wherein the AAV capsid (*Cap*) gene encodes a capsid from an AAV of serotype 9, and the AAV replication (*Rep*) gene encodes a Rep protein from an AAV of serotype 2 or 9, and wherein the recombinant nucleic acid molecule:

(i) comprises a splice site sequence which improves gene splicing efficiency, and/or

(ii) comprises a polyA signal sequence.

3. A recombinant nucleic acid molecule comprising an AAV capsid (*Cap*) gene and an AAV replication (*Rep*) gene, wherein the AAV capsid (*Cap*) gene encodes a capsid from an AAV of serotype 2.7m8, and the AAV replication (*Rep*) gene encodes a Rep protein from an AAV of serotype 2, and wherein the recombinant nucleic acid molecule:

(i) comprises a splice site sequence which improves gene splicing efficiency, and/or

(ii) comprises a polyA signal sequence.

4. The recombinant nucleic acid molecule according to any preceding claim, wherein the splice site sequence which improves gene splicing efficiency is the 5′ splice site of the *rep* and *cap* genes, optionally wherein the splice site sequence which improves gene splicing efficiency is the 5′ splice site of the intron of the *rep* and *cap* genes.

5. The recombinant nucleic acid molecule according to any preceding claim, wherein the splice site sequence which improves gene splicing efficiency comprises a sequence as set out in SEQ ID No: 1, comprising at least one base substitution.

6. The recombinant nucleic acid molecule according to any preceding claim, wherein the splice site comprises a sequence as set out in SEQ ID No: 1, comprising at least two base substitutions, or wherein the splice site comprises a sequence as set out in SEQ ID No: 1, comprising at least three base substitutions, or wherein the

splice site comprises a sequence as set out in SEQ ID No: 1, comprising at least four base substitutions.

7. The recombinant nucleic acid molecule according to claim 5 or claim 6, wherein the at least one base substitution occurs within the last three bases of SEQ ID No: 1, or wherein the at least two base substitutions occur within the last three bases of SEQ ID No: 1, or wherein the at least three base substitutions occur within the last three bases of SEQ ID No: 1.

8. The recombinant nucleic acid molecule according to any one of claims 5-7, wherein the at least one, two or three base substitutions occur within the first three bases of SEQ ID No: 1.

9. The recombinant nucleic acid molecule according to any preceding claim, wherein the splice site comprises a sequence as set out in SEQ ID No: 2, comprising at least one base substitution.

10. The recombinant nucleic acid molecule according to claim 9, wherein the at least one base substitution occurs at the second base within SEQ ID No: 2, or wherein the at least one base substitution occurs at the third base within SEQ ID No: 2.

11. The recombinant nucleic acid molecule according to any preceding claim, wherein the splice site comprises or consists of a sequence having at least 60% sequence identity to SEQ ID No: 2.

12. The recombinant nucleic acid molecule according to any preceding claim, wherein the splice site comprises or consists of a sequence having at least 65%, 70%, 75%, 80%, 85% or 90% sequence identity to SEQ ID No: 2.

13. The recombinant nucleic acid molecule according to any preceding claim, wherein (i) the splice site does not comprise or consist of a sequence having 100% sequence identity to SEQ ID No: 2; and/or (ii) the splice site sequence does not comprise a stop codon.

14. The recombinant nucleic acid molecule according to any preceding claim, wherein the splice site sequence comprises or consists of a nucleotide sequence substantially as set out in any one of SEQ ID Nos: 3, 4, 5, 6, 7, 8 or 9, or a fragment or variant thereof.

15. The recombinant nucleic acid molecule according to any preceding claim, wherein the splice site sequence comprises or consists of a nucleotide sequence substantially as set out in SEQ ID No: 3, or a fragment or variant thereof.

16. The recombinant nucleic acid molecule according to any preceding claim, wherein the polyA signal sequence is selected from a group consisting of: SV40 polyA signal sequence; human growth hormone polyA signal sequence; and beta-globin polyA signal sequence.

17. The recombinant nucleic acid molecule according to any preceding claim, wherein the polyA signal sequence is an SV40 polyA signal sequence.

18. The recombinant nucleic acid molecule according to any preceding claim, wherein the polyA signal sequence is at least 10, 20, 30, 40, 50 or 60 nucleotides in length, or wherein the polyA signal sequence is at least 70, 80, 90, 100, 100 or 120 nucleotides in length.

19. The recombinant nucleic acid molecule according to any preceding claim, wherein the polyA signal sequence is between 60 and 170, 70 and 160, 80 and 150, 90 and 140, or 100 and 130 nucleotides in length.

20. The recombinant nucleic acid molecule according to any preceding claim, wherein at least 10%, 20% or 30% of the nucleotides in the polyA signal sequence are adenine.

21. The recombinant nucleic acid molecule according to any preceding claim, wherein the polyA signal sequence comprises or consists of a nucleotide sequence substantially as set out in SEQ ID No: 10, or a fragment or variant thereof.

22. The recombinant nucleic acid molecule according to any preceding claim, wherein the polyA signal sequence is disposed 3′ of the *cap* gene.

23. The recombinant nucleic acid molecule according to claim 1, wherein the AAV capsid (*Cap*) gene encodes a capsid from an AAV of serotype 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, rh.10, 2.7m8, or a variant thereof, optionally wherein the AAV capsid is an AAV serotype 2 (AAV2) capsid, an AAV serotype 8 (AAV8) capsid, an AAV serotype 9 (AAV9) capsid, an AAV2.7m8 serotype capsid, or a modified capsid thereof.

24. The recombinant nucleic acid molecule according to claim 1, wherein the AAV replication (*Rep*) gene encodes a Rep protein from an AAV of serotype 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, rh.10, or a variant thereof, optionally wherein the AAV Rep protein is an AAV serotype 2 (AAV2), an AAV serotype 8 (AAV8), or an AAV serotype 9 (AAV9), or a modified capsid thereof.

25. The recombinant nucleic acid molecule according to any preceding claim, wherein the AAV replication (*Rep*) gene encodes a Rep protein from an AAV of serotype 2, or wherein the AAV replication (*Rep*) gene encodes a Rep protein from an AAV of serotype 9.

26. The recombinant nucleic acid molecule according to any preceding claim, wherein the recombinant nucleic acid molecule further comprises one or more promoters, optionally wherein the one or more promoters is an AAV promoter.

27. The recombinant nucleic acid molecule according to claim 26, wherein the one or more promoters is a P5, P19, and/or a P40 promoter.

28. The recombinant nucleic acid molecule according to any preceding claim, wherein the recombinant nucleic acid molecule further comprises one or more adenoviral helper genes, optionally wherein the adenoviral helper genes are selected from a group consisting of: E1A, E1B, E2A, E4, and VA RNA.

29. The recombinant nucleic acid molecule according to any preceding claim, wherein the recombinant nucleic acid molecule further comprises a transgene coding sequence.

30. The recombinant nucleic acid molecule according to any preceding claim, wherein the recombinant nucleic acid molecule further comprises one or more adenoviral helper genes and a transgene coding sequence.

31. A vector comprising the recombinant nucleic acid molecule according to any one of claims 1 to 29.

32. The vector according to claim 31, wherein the vector is selected from a group consisting of: a plasmid, a cosmid, an episome, and a virus, optionally wherein the vector is a plasmid.

33. A host cell comprising the recombinant nucleic acid molecule according to any one of claims 1 to 30, or the vector according to claim 31 or claim 32.

34. The host cell according to claim 33, wherein the host cell is selected from a group consisting of: a HEK293, HeLa, Cos-7, A549, BHK, Vero, RD, ARPE-19, and MRC-5 cell.

35. The host cell according to claim 33 or claim 34, wherein the host cell is a HEK293 cell or a cell derived therefrom, optionally wherein the cell derived from a HEK293 cell is Expi293F, VPC2.0, or a HEK293T cell.

36. The host cell according to any one of claims 33 to 35, wherein the host cell further comprises a plasmid comprising one or more adenoviral helper genes, and/or a plasmid comprising a transgene coding sequence.

37. Use of the recombinant nucleic acid molecule according to any one of claims 1 to 30, or the vector according to claim 31 or claim 32, to produce recombinant AAV (rAAV).

38. A method of producing recombinant AAV (rAAV), the method comprising:

(i) delivering the recombinant nucleic acid molecule according to any one of claims 1 to 30, or the vector according to claim 31 or claim 32, to a host cell; and

(ii) culturing the host cell under suitable conditions to produce rAAV.