

77 bp sequence from the F1 phage genome (AK) has been found to be highly recombinogenic in vitro and in vivo experiments.

Although studies have highlighted the potential of dual vector systems, such as AAV vector systems, for delivery and reconstitution of large transgenes in a tissue of interest, their translation to clinical use remains a significant unmet need.

SUMMARY

During development of a dual vector system for the delivery of a large transgene (e.g. Myosin7A, MYO7A), the inventors unexpectedly discovered a consistent contaminant in their preparation of the vector containing the 5' end portion of the transgene CDS. The inventors analysed the preparations with Southern blots and identified a band corresponding to the expected vector and surprisingly also discovered a smaller size band of about 1.3 kb corresponding to the contaminant.

The inventors then studied the vectors further and identified region of homology between the chimeric promoter intron and the splicing donor (SD) site used in the vector. Further sequencing analysis of purified viral DNA confirmed that a homologous recombination event takes place due to the presence of these regions of homology within the construct, which leads to the deletion of the remaining portion of the intron, the 5' end portion of the transgene CDS and the SD site while retaining AAV inverted terminal repeats (ITRs), thus supporting vector production.

In one example the disclosure provides a vector system for expressing a transgene in a cell, the vector system comprising a first vector and a second vector, wherein:

- (a) the first vector comprises in a 5' to 3' direction: an intron; a 5' end portion of the transgene coding sequence (CDS); and a splice donor sequence;
- (b) the second vector comprises in a 5' to 3' direction: a splice acceptor sequence; and a 3' end portion of the transgene CDS;

wherein the 5' end portion and the 3' end portion together constitute the transgene CDS, and wherein the intron is not capable of homologous recombination with the splice donor sequence to excise the 5' end portion of the transgene CDS.

In another example the disclosure provides a vector system for expressing a transgene in a cell, the vector system comprising a first vector and a second vector, wherein:

- (a) the first vector comprises in a 5' to 3' direction: a promoter; an intron; a 5' end portion of the transgene coding sequence (CDS); and a splice donor sequence;
- (b) the second vector comprises in a 5' to 3' direction: a splice acceptor sequence; and a 3' end portion of the transgene CDS;

wherein the 5' end portion and the 3' end portion together constitute the transgene CDS, and wherein the intron is not capable of homologous recombination with the splice donor sequence to excise the 5' end portion of the transgene CDS.

In one aspect the disclosure provides a vector system for expressing a transgene in a cell, the vector system comprising a first vector and a second vector, wherein:

- (a) the first vector comprises in a 5' to 3' direction: a promoter; an intron; a 5' end portion of the transgene coding sequence (CDS); a splice donor sequence; and a first recombinogenic region;
- (b) the second vector comprises in a 5' to 3' direction: a second recombinogenic region; a splice acceptor sequence; and a 3' end portion of the transgene CDS;

wherein the 5' end portion and the 3' end portion together constitute the transgene CDS, and wherein the intron is not capable of homologous recombination with the splice donor sequence to excise the 5' end portion of the transgene CDS.

In preferred examples, the intron is a simian virus 40 (SV40) intron. The SV40 intron may be a modified SV40 intron.

In some examples, the intron is a minute virus mice (MVM) intron.

In some examples, the intron comprises a nucleotide sequence with at least 95% sequence identity (e.g. at least 96%, 97%, 98% or 99% sequence identity, or 100% sequence identity) to SEQ ID NO: 3 or 4. In preferred examples, the intron comprises a nucleotide sequence with at least 95% sequence identity (e.g. at least 96%, 97%, 98% or 99% sequence identity, or 100% sequence identity) to SEQ ID NO: 3.

In some examples, the splice donor sequence comprises a nucleotide sequence with at least 95% sequence identity to SEQ ID NO: 5.

In preferred examples, the first recombinogenic region and the second recombinogenic region are the same.

In some examples, the first recombinogenic region and the second recombinogenic region are both F1 phage recombinogenic regions or fragments thereof.

- 5 In some examples, the first recombinogenic region and the second recombinogenic region both comprise a nucleotide sequence with at least 95% sequence identity to SEQ ID NO: 7 or a fragment thereof.

In some examples, the first vector and the second vector are viral vectors. The viral vectors may be adeno-associated viral (AAV) vectors, adenoviral vectors, retroviral vectors, lentiviral
10 vectors, herpes simplex viral vectors, picornaviral vectors or alphaviral vectors. In some examples, the first vector and the second vector are plasmids. The first and/or second plasmid may, for example, be used to produce the first and/or second viral vector particles (e.g. separately or together in a composition).

In preferred examples, the first vector and the second vector are AAV vectors.

- 15 In some examples, the AAV vectors are of the same serotype (e.g. comprise capsids of the same serotype). In some examples, the AAV vectors are of different serotypes (e.g. comprise capsids of different serotypes).

In some examples, the first vector and the second vector are selected from the group consisting of AAV2, AAV8, AAV5, AAV7, AAV9, AAV-PhP.B and AAV-PhP.eB. In some
20 examples, the first vector and the second vector are selected from the group consisting of hu68 (see, for example, WO 2018/160582), Anc libraries (see, for example, WO 2015/054653 and WO 2017/019994) and AAV2-TT (see, for example, WO 2015/121501).

In some examples, the first vector and the second vector are AAV2 vectors. In some examples, the first vector and the second vector are AAV8 vectors.

- 25 In some examples, the first vector and the second vector comprise capsids selected from the group consisting of AAV2, AAV8, AAV5, AAV7, AAV9, AAV-PhP.B and AAV-PhP.eB. In some examples, the first vector and the second vector comprise capsids selected from the group consisting of hu68 (see, for example, WO 2018/160582), Anc libraries (see, for example, WO 2015/054653 and WO 2017/019994) and AAV2-TT (see, for example, WO 2015/121501).

In some examples, the first vector and the second vector comprise AAV2 capsids. In some examples, the first vector and the second vector comprise AAV8 capsids.

5 In some examples, the first vector further comprises a 5' ITR and a 3' ITR. In some examples, the second vector further comprises a 5' ITR and a 3' ITR. In preferred examples, the first vector further comprises a 5' ITR and a 3' ITR, and the second vector further comprises a 5' ITR and a 3' ITR.

In preferred examples, the ITRs are AAV ITRs, preferably AAV2 ITRs. In some examples, the ITRs are AAV8 ITRs.

In preferred examples, the first vector and the second vector are AAV2/8 vectors.

10 In some examples, the ITRs are from the same AAV serotype. In some examples, the ITRs are from different AAV serotypes.

In preferred examples, the 3' ITR of the first vector and the 5' ITR of the second vector are from the same AAV serotype.

15 In preferred examples, the 5' ITR of the first vector and the 5' ITR of the second vector are from the same AAV serotype. In preferred examples, the 3' ITR of the first vector and the 3' ITR of the second vector are from the same AAV serotype.

In preferred examples, the 5' ITR of the first vector and the 5' ITR of the second vector are AAV2 5' ITRs, and the 3' ITR of the first vector and the 3' ITR of the second vector are AAV2 3' ITRs.

20 In some examples, the 5' ITR of the first vector and the 5' ITR of the second vector are AAV8 5' ITRs, and the 3' ITR of the first vector and the 3' ITR of the second vector are AAV8 3' ITRs.

25 In some examples, the 5' ITR of the first vector and the 5' ITR of the second vector are from different AAV serotypes. In some examples, the 3' ITR of the first vector and the 3' ITR of the second vector are from different AAV serotypes. In some examples, the 5' ITR of the first vector and the 5' ITR of the second vector are from different AAV serotypes, and the 3' ITR of the first vector and the 3' ITR of the second vector are from different AAV serotypes.

In some examples, the 5' ITR of the first vector and the 3' ITR of the second vector are from different AAV serotypes.

In some examples, the first vector and the second vector are viral vector particles.

In some examples, the promoter is a CBA promoter or a fragment thereof.

In some examples, the first vector further comprises an enhancer sequence. In preferred examples, the enhancer is a CMV enhancer.

- 5 In some examples, the second vector further comprises a polyadenylation sequence downstream of the 3' end portion of the transgene CDS. In preferred examples, the polyadenylation sequence is a bovine growth hormone (bGH) polyadenylation sequence.

In some examples, the transgene is selected from the group consisting of: Myosin 7A (MYO7A), ABCA4, CEP290, CDH23, EYS, USH2a, GPR98 and ALMS1.

- 10 In preferred examples, the transgene is a Myosin 7A (MYO7A) transgene. In some examples, the transgene is an ABCA4 transgene. In some examples, the transgene CDS is a wild type sequence. In some examples, the transgene CDS is codon optimised (e.g. codon optimised for expression in humans).

In preferred examples:

- 15 (a) the first vector comprises a nucleotide sequence with at least 95% sequence identity to SEQ ID NO: 14; and/or
- (b) the second vector comprises a nucleotide sequence with at least 95% sequence identity to SEQ ID NO: 15.

In particularly preferred examples:

- 20 (a) the first vector comprises the nucleotide sequence of SEQ ID NO: 14; and/or
- (b) the second vector comprises the nucleotide sequence of SEQ ID NO: 15.

In some examples, the first vector and second vector are in a 1:1 genome copy ratio.

- In another example the disclosure provides a method for expressing a transgene in a cell, comprising transducing or transfecting the cell with the first vector and the second vector as disclosed herein, such that the transgene is expressed in the cell.
- 25

In another example the disclosure provides a cell comprising the first vector and the second vector as disclosed herein.

In another example the disclosure provides a cell transduced or transfected with the first vector and the second vector as disclosed herein.

In some examples, the cell is a mammalian cell, a human cell, a retinal cell or a non-embryonic stem cell.

- 5 In another example the disclosure provides a vector, wherein the vector is the first vector as disclosed herein.

In another example the disclosure provides a vector, wherein the vector is the second vector as disclosed herein.

- 10 In another example the disclosure provides a vector comprising in a 5' to 3' direction: an intron; a 5' end portion of a transgene coding sequence (CDS); and a splice donor sequence, wherein the intron is not capable of homologous recombination with the splice donor sequence to excise the 5' end portion of the transgene CDS.

- 15 In another example the disclosure provides a vector comprising in a 5' to 3' direction: a promoter; an intron; a 5' end portion of a transgene coding sequence (CDS); and a splice donor sequence, wherein the intron is not capable of homologous recombination with the splice donor sequence to excise the 5' end portion of the transgene CDS.

- 20 In another aspect the disclosure provides a vector comprising in a 5' to 3' direction: a promoter; an intron; a 5' end portion of a transgene coding sequence (CDS); a splice donor sequence; and a recombinogenic region, wherein the intron is not capable of homologous recombination with the splice donor sequence to excise the 5' end portion of the transgene CDS.

In some examples, the vector further comprises a 5' ITR and a 3' ITR. In preferred examples, the ITRs are AAV ITRs, preferably AAV2 ITRs.

In some examples, the ITRs are from the same AAV serotype. In some examples, the ITRs are from different AAV serotypes.

- 25 In another example the disclosure provides a vector comprising in a 5' to 3' direction: a splice acceptor sequence; and a 3' end portion of a transgene coding sequence (CDS).

In another example the disclosure provides a vector comprising in a 5' to 3' direction: a recombinogenic region; a splice acceptor sequence; and a 3' end portion of a transgene coding sequence (CDS).

In some examples, the vector comprises a nucleotide sequence with at least 95% sequence identity to SEQ ID NO: 14.

In preferred examples, the vector comprises the nucleotide sequence of SEQ ID NO: 14.

5 In another example the disclosure provides a kit comprising the first vector as disclosed herein and the second vector as disclosed herein.

In another example the disclosure provides a composition comprising the first vector as disclosed herein and the second vector as disclosed herein.

In some examples, the first vector and second vector are in a 1:1 genome copy ratio.

10 In preferred examples, the composition is a pharmaceutical composition comprising a pharmaceutically-acceptable carrier, diluent or excipient.

In another example the disclosure provides the vector system, vector, kit or composition of the disclosure for use in therapy.

15 In another example the disclosure provides the vector system, vector, kit or composition of the disclosure for use in treatment of a retinal degeneration. Preferably the retinal degeneration is an inherited retinal degeneration.

In another example the disclosure provides the first vector as disclosed herein for use in therapy, wherein the first vector is administered simultaneously, sequentially or separately in combination with the second vector as disclosed herein.

20 In another example the disclosure provides the first vector as disclosed herein for use in treatment of a retinal degeneration, wherein the first vector is administered simultaneously, sequentially or separately in combination with the second vector as disclosed herein. Preferably the retinal degeneration is an inherited retinal degeneration.

25 In another example the disclosure provides the second vector as disclosed herein for use in therapy, wherein the second vector is administered simultaneously, sequentially or separately in combination with the first vector as disclosed herein.

In another example the disclosure provides the second vector as disclosed herein for use in treatment of a retinal degeneration, wherein the second vector is administered simultaneously, sequentially or separately in combination with the first vector as disclosed herein. Preferably the retinal degeneration is an inherited retinal degeneration.

In some examples, the use is in treatment or prevention of Usher syndrome, retinitis pigmentosa, Leber congenital amaurosis (LCA), Stargardt disease, Alstrom syndrome or ABCA4-associated diseases.

- 5 In another example the disclosure provides the vector system, vector, kit or composition of the disclosure for use in treatment of Usher syndrome.

In another example the disclosure provides a method of treating or preventing a retinal degeneration comprising administering an effective amount of the vector system, vector, kit or composition of the disclosure to a subject in need thereof. Preferably the retinal degeneration is an inherited retinal degeneration.

10 In another example the disclosure provides a method of treating or preventing Usher syndrome comprising administering an effective amount of the vector system, vector, kit or composition of the disclosure to a subject in need thereof. In yet another aspect, the present disclosure refers to a use of the vector system as disclosed herein in the manufacture of a medicament for treating or preventing Usher syndrome, wherein an effective amount of the medicament is to be administered to a subject in need thereof, optionally wherein the Usher syndrome is Usher syndrome Type 1B.

20 **DESCRIPTION OF THE DRAWINGS**

FIGURE 1. Identification of a contaminant vector.

(A) Southern blot image showing genomes corresponding to the AAV-5'hMYO7A (indicated by top arrow) and to the contaminant vector (indicated by bottom arrow). DNase: treatment with DNase for degradation of contaminant external DNA; plasmid: plasmid DNA containing the DNA sequences to generate AAV8-CBA-Chimeric intron-5'hMYO7A; 5'AAV genome-CI: genomic DNA extracted from AAV8-CBA promoter-Chimeric intron-5'hMYO7A; molecular weight marker expressed in kilobases; bp = base pair. CI: chimeric intron. (B) Representation of AAV-5'hMYO7A genome showing the sequence recognised by the southern blot probe. (C) Pairing mechanism between the chimeric promoter's intron and the SD signal (indicated by dotted lines). (D) Representation of the contaminant vector genome showing the sequence recognised by the Southern blot probe. (E) Southern blot analysis of AAV preparations including the following expression cassettes: 1. 5'CMV ABCA4 AK (dual hybrid); 2. 5'CMV ABCA4 TS (dual trans-splicing); 3. 5'CMV NO INTR ABCA4 OV (dual overlapping); 4. 5'CMV NO INTR ABCA4 AK (dual hybrid); 5. 5'VMD2 ABCA4 AK (dual hybrid); 6. 5'RHO ABCA4 AK (dual hybrid); 7. 5'RHO ABCA4 TS (dual trans-splicing). Chimeric intron is present in vectors 1 and 2 and absent in vectors 3 to 7. The dashed box indicates full-length genomes of the expected sizes; the solid box indicates short, truncated genomes.

40 **FIGURE 2. In vitro comparison of Chimeric intron, SV40 intron, MVM intron and no intron by EGFP fluorescence.**

CLAIMS

1. A vector system for expressing a transgene in a cell, the vector system comprising a first vector and a second vector, wherein:
 - (a) the first vector comprises in a 5' to 3' direction: a promoter; an intron; a 5' end portion of the transgene coding sequence (CDS); a splice donor sequence; and a first recombinogenic region;
 - (b) the second vector comprises in a 5' to 3' direction: a second recombinogenic region; a splice acceptor sequence; and a 3' end portion of the transgene CDS;wherein the 5' end portion and the 3' end portion together constitute the transgene CDS, and wherein the intron is not capable of homologous recombination with the splice donor sequence to excise the 5' end portion of the transgene CDS.
2. The vector system of claim 1, wherein the intron does not comprise a region of at least 20, 30, 40, 50, 60, 70, 80, 90 or 100 contiguous nucleotides having at least 95%, 96%, 97%, 98%, 99% or 100% (preferably 100%) sequence identity to a region of the splice donor sequence.
3. The vector system of claim 1 or 2, wherein the intron:
 - (a) is a simian virus 40 (SV40) intron or a minute virus mice (MVM) intron; and/or
 - (b) comprises a nucleotide sequence with at least 95% sequence identity to SEQ ID NO: 3 or 4.
4. The vector system of any preceding claim, wherein the splice donor sequence comprises a nucleotide sequence with at least 95% sequence identity to SEQ ID NO: 5.
5. The vector system of any preceding claim, wherein the first recombinogenic region and the second recombinogenic region:
 - (a) are both F1 phage recombinogenic regions or fragments thereof; and/or
 - (b) both comprise a nucleotide sequence with at least 95% sequence identity to SEQ ID NO: 7 or a fragment thereof.

6. The vector system of any preceding claim, wherein the first vector and the second vector are viral vectors.
7. The vector system of any preceding claim, wherein the first vector and the second vector are AAV vectors, optionally wherein the first vector further comprises a 5' ITR and a 3' ITR, and the second vector further comprises a 5' ITR and a 3' ITR.
8. The vector system of any preceding claim, wherein the promoter is a CBA promoter or a fragment thereof.
9. The vector system of any preceding claim, wherein the second vector further comprises a polyadenylation sequence downstream of the 3' end portion of the transgene CDS.
10. The vector system of any preceding claim, wherein the transgene is a Myosin 7A (MYO7A) transgene.
11. The vector system of any preceding claim, wherein:
 - (a) the first vector comprises a nucleotide sequence with at least 95% sequence identity to SEQ ID NO: 14; and/or
 - (b) the second vector comprises a nucleotide sequence with at least 95% sequence identity to SEQ ID NO: 15.
12. A method for expressing a transgene in a cell, comprising transducing or transfecting the cell with the first vector and the second vector as defined in any preceding claim, such that the transgene is expressed in the cell.
13. A vector comprising in a 5' to 3' direction: a promoter; an intron; a 5' end portion of a transgene coding sequence (CDS); a splice donor sequence; and a recombinogenic region, wherein the intron is not capable of homologous recombination with the splice donor sequence to excise the 5' end portion of the transgene CDS.
14. The vector of claim 13, wherein the vector comprises a nucleotide sequence with at least 95% sequence identity to SEQ ID NO: 14.
15. The vector system of any one of claims 1-11 for use in therapy.
16. The vector system of any one of claims 1-11 for use in treatment of Usher syndrome, optionally Usher syndrome Type 1B.

17. Use of the vector system of any one of claims 1-11 in the manufacture of a medicament for treating or preventing Usher syndrome, wherein an effective amount of the medicament is to be administered to a subject in need thereof, optionally wherein the Usher syndrome is Usher syndrome Type 1B.