

## VECTORS

### FIELD OF THE INVENTION

The present invention relates to dual adeno-associated virus (AAV) vector systems for expressing a COL4A3, COL4A4 or COL4A5 polypeptide. The present invention also relates to cells and pharmaceutical compositions comprising said dual AAV vector systems and uses thereof, for example for preventing and/or treating Alport Syndrome.

### BACKGROUND TO THE INVENTION

Alport syndrome is a genetic condition affecting approximately 1 in 5,000-10,000 of all individuals in continental Europe and the USA. Alport syndrome usually presents during childhood and is associated with a spectrum of phenotypes that include a progressive loss of kidney function, and which can also include hearing loss and eye abnormalities (see e.g. Savage, J., et al., 2021. European Journal of Human Genetics, 29(8), pp.1186-1197).

Alport syndrome is caused by pathogenic variants in the COL4A3, COL4A4 and COL4A5 genes, which result in abnormalities of the collagen IV  $\alpha$ 345 network of basement membranes. The condition can be transmitted in an X-linked, autosomal dominant, or autosomal recessive pattern, with X-linked being the common while autosomal recessive and autosomal dominant account for around 15% and 20% of cases respectively (see e.g. Warady, B.A., et al., 2020. Kidney medicine, 2(5), pp.639-649).

In the absence of treatment, renal disease progresses from microhematuria to proteinuria, progressive renal insufficiency and end-stage renal disease in all males with the X-linked form, and in all males and females with the autosomal recessive form. Alport syndrome can be diagnosed by genetic testing and current treatments include angiotensin converting enzyme (ACE) inhibitors or angiotensin receptor blockers (ARB) to delay onset of end-stage kidney disease. However, at present there is no way to prevent end-stage renal failure, with a renal transplant being the only option (see e.g. Kashtan, C. E et al., 2018. Kidney international, 93(5), pp.1045-1051).

There are significant challenges to overcome in developing a successful gene therapy for Alport syndrome. For example, whilst adeno-associated virus (AAV) vectors are a leading platform for gene delivery, the COL4A5, COL4A3 and COL4A4 polypeptides are each up to around 1600-1700 amino acids in length, rendering them challenging for delivery by an AAV vector, due to limited AAV cargo capacity.

Thus, there is a demand for new gene therapies that can efficiently express COL4A3, COL4A4 or COL4A5 polypeptides in host cells and thereby provide a therapy for example, for the prevention and/or treatment of Alport syndrome and related conditions (e.g. any condition in a subject who has a pathogenic variant of a COL4A3, COL4A4 or COL4A5 gene).

## 5 SUMMARY OF THE INVENTION

The present invention is based on the inventors surprising provision of a dual adeno-associated virus (AAV) vector system for expressing a COL4A3, COL4A4 or COL4A5 polypeptide. The present inventors have shown for the first time that a dual AAV vector system can be used to efficiently express a COL4A3, COL4A4 or COL4A5 polypeptide in a host cell.

10 A dual AAV vector system encoding a COL4A3, COL4A4 or COL4A5 polypeptide may provide an improved therapy for the prevention and/or treatment of Alport syndrome and related conditions. The dual AAV vector system encoding a COL4A3, COL4A4 or COL4A5 polypeptide may provide a therapy for a subject with a pathogenic variant of a COL4A3, COL4A4 or COL4A5 gene.

15 In one aspect, the present invention provides a dual adeno-associated virus (AAV) vector system for expressing a COL4A3, COL4A4 or COL4A5 polypeptide comprising: (a) a first AAV vector comprising a 5' coding sequence (CDS), wherein the 5' CDS encodes an N-terminal part of the COL4A3, COL4A4 or COL4A5 polypeptide; and (b) a second AAV vector comprising a 3' CDS, wherein the 3' CDS encodes a C-terminal part of the COL4A3, COL4A4 or COL4A5 polypeptide.

The COL4A3, COL4A4 or COL4A5 polypeptide is not particularly limited. Any suitable COL4A3, COL4A4 or COL4A5 polypeptide may be used. Suitably, the COL4A3, COL4A4 or COL4A5 polypeptide is a human COL4A3, COL4A4 or COL4A5 polypeptide. In some embodiments the COL4A3, COL4A4 or COL4A5 polypeptide comprises or consists of: (a) an amino acid sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 1 to 7; (b) an amino acid sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 8 to 20; or (c) an amino acid sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 21 to 26. In some embodiments the COL4A3, COL4A4 or COL4A5 polypeptide comprises or consists of:



(a) an amino acid sequence having at least 70% identity to SEQ ID NO: 1; (b) an amino acid sequence having at least 70% identity to SEQ ID NO: 8; or (c) an amino acid sequence having at least 70% identity to SEQ ID NO: 21. In some embodiments, the COL4A3, COL4A4 or COL4A5 polypeptide comprises or consists of: (a) the amino acid sequence of SEQ ID NO: 1;  
 5 (b) the amino acid sequence of SEQ ID NO: 8; or (c) the amino acid sequence of SEQ ID NO: 21.

A CDS encoding the COL4A3, COL4A4 or COL4A5 polypeptide may be formed upon delivery of the first AAV vector and the second AAV vector into a host cell. The CDS encoding the COL4A3, COL4A4 or COL4A5 polypeptide is not particularly limited. Any suitable CDS  
 10 encoding a COL4A3, COL4A4 or COL4A5 polypeptide may be used. In some embodiments, the CDS encoding the COL4A3, COL4A4 or COL4A5 polypeptide comprises or consists of: (a) a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 27; (b) a nucleotide  
 15 sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 81; or (c) a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%,  
 20 at least 99%, or 100% identity to SEQ ID NO: 129. In some embodiments, the CDS encoding the COL4A3, COL4A4 or COL4A5 polypeptide comprises or consists of: (a) the nucleotide sequence of SEQ ID NO: 27; (b) the nucleotide sequence of SEQ ID NO: 81; or (c) the nucleotide sequence of SEQ ID NO: 129.

The 5' CDS may encode any suitable N-terminal part of a COL4A3, COL4A4 or COL4A5  
 25 polypeptide. Suitably, the 5' CDS encodes an N-terminal part of the COL4A3, COL4A4 or COL4A5 polypeptide having a length of from 400 amino acids to 1300 amino acids, from 500 amino acids to 1200 amino acids, from 600 amino acids to 1100 amino acids, or from 700 amino acids to 1000 amino acids. Suitably, the 5' CDS comprises exons 1-10 or more, exons 1-15 or more, exons 1-20 or more, or exons 1-25 or more of a CDS encoding a COL4A3,  
 30 COL4A4 or COL4A5 polypeptide. Suitably, the 5' CDS comprises exons 1-45 or less, exons 1-40 or less, or exons 1-35 or less of a CDS encoding a COL4A3, COL4A4 or COL4A5 polypeptide. Suitably, the 5' CDS comprises or consists of: (a) at least exons 1-15 to at most exons 1-43 of a CDS encoding a COL4A3 polypeptide; (b) at least exons 1-14 to at most exons 1-39 of a CDS encoding a COL4A4 polypeptide; or (c) at least exons 1-15 to at most  
 35 exons 1-41 of a CDS encoding a COL4A5 polypeptide. In some embodiments, the 5' CDS

comprises or consists of: (a) exons 1-28, exons 1-29, exons 1-30, exons 1-31, exons 1-32, exons 1-33, or exons 1-34 of a CDS encoding a COL4A3 polypeptide; (b) exons 1-25, exons 1-26, exons 1-27, exons 1-28, exons 1-29, exons 1-30, or exons 1-31 of a CDS encoding a COL4A4 polypeptide; or (c) exons 1-27, exons 1-28, exons 1-29, exons 1-30, exons 1-31, exons 1-32, or exons 1-33 of a CDS encoding a COL4A5 polypeptide, or exons 1-40 of a CDS encoding a COL4A5 polypeptide. In some embodiments, the 5' CDS comprises or consists of: (a) a nucleotide sequence having at least 70% identity to any of SEQ ID NOs: 181 to 187; (b) a nucleotide sequence having at least 70% identity to any of SEQ ID NOs: 188 to 194; or (c) a nucleotide sequence having at least 70% identity to any of SEQ ID NOs: 195 to 201 or SEQ ID NO: 287. In some embodiments, the 5' CDS comprises or consists of: (a) a nucleotide sequence selected from any of SEQ ID NOs: 181 to 187; (b) a nucleotide sequence selected from any of SEQ ID NOs: 188 to 194; or (c) a nucleotide sequence selected from any of SEQ ID NOs: 195 to 201 or SEQ ID NO: 287.

The 3' CDS may encode any suitable C-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide. Suitably, the 3' CDS encodes an C-terminal part of the COL4A3, COL4A4 or COL4A5 polypeptide having a length of from 500 amino acids to 1200 amino acids, from 600 amino acids to 1100 amino acids, or from 700 amino acids to 1000 amino acids. Suitably, the 3' CDS comprises: (a) exons 48-52 or more, exons 43-52 or more, exons 38-52 or more, or exons 33-52 or more of a CDS encoding a COL4A3 polypeptide; (b) exons 43-47 or more, exons 38-47 or more, exons 33-47 or more, or exons 28-47 or more of a CDS encoding a COL4A4 polypeptide; or (c) exons 47-51 or more, exons 42-51 or more, exons 37-51 or more, or exons 32-51 or more of a CDS encoding a COL4A5 polypeptide. Suitably, the 3' CDS comprises: (a) exons 18-52 or less, exons 23-52 or less, or exons 28-52 or less of a CDS encoding a COL4A3 polypeptide; (b) exons 13-47 or less, exons 18-47 or less, or exons 23-47 or less of a CDS encoding a COL4A4 polypeptide; or (c) exons 12-51 or less, exons 17-51 or less, or exons 22-51 or less of a CDS encoding a COL4A5 polypeptide. Suitably, the 3' CDS comprises or consists of: (a) at least exons 44-52 to at most exons 16-52 of a CDS encoding a COL4A3 polypeptide; (b) at least exons 40-47 to at most exons 15-47 of a CDS encoding a COL4A4 polypeptide; or (c) at least exons 42-51 to at most exons 16-51 of a CDS encoding a COL4A5 polypeptide. In some embodiments, the 3' CDS comprises or consists of: (a) exons 35-52, exons 34-52, exons 33-52, exons 32-52, exons 31-52, exons 30-52, or exons 29-52 of a CDS encoding a COL4A3 polypeptide; (b) exons 32-47, exons 31-47, exons 30-47, exons 29-47, exons 28-47, exons 27-47, or exons 26-47, of a CDS encoding a COL4A4 polypeptide; or (c) exons 36-51, exons 35-51, exons 34-51, exons 33-51, exons 32-51, exons 31-51, exons 30-51, exons 29-51, exons 28-51 or exons 27-51, of a CDS encoding a COL4A5 polypeptide. In some embodiments, the 3' CDS comprises or consists of: (a) a nucleotide

sequence having at least 70% identity to any of SEQ ID NOs: 202 to 208; (b) a nucleotide sequence having at least 70% identity to any of SEQ ID NOs: 209 to 215; or (c) a nucleotide sequence having at least 70% identity to any of SEQ ID NOs: 295 to 297, SEQ ID NOs: 216 to 222 or SEQ ID NOs: 308 to 310. In some embodiments, the 3' CDS comprises or consists of: (a) a nucleotide sequence selected from any of SEQ ID NOs: 202 to 208; (b) a nucleotide sequence selected from any of SEQ ID NOs: 209 to 215; or (c) a nucleotide sequence selected from any of SEQ ID NOs: 295 to 297, SEQ ID NOs: 216 to 222 or SEQ ID NOs: 308 to 310.

The 5' CDS and the 3' CDS may together be used to reconstitute the full-length CDS upon co-delivery of the first AAV vector and the second AAV vector. In some embodiments, the 5' CDS and the 3' CDS are overlapping. In other embodiments, the 5' CDS and the 3' CDS do not overlap. In some embodiments i) the 5' CDS consists of exons 1-24 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 25-51 of a CDS encoding a COL4A5 polypeptide; or ii) the 5' CDS consists of exons 1-31 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 32-51 of a CDS encoding a COL4A5 polypeptide; or iii) the 5' CDS consists of exons 1-32 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 33-51 of a CDS encoding a COL4A5 polypeptide; or iv) the 5' CDS consists of exons 1-33 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 34-51 of a CDS encoding a COL4A5 polypeptide; or v) the 5' CDS consists of exons 1-38 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 39-51 of a CDS encoding a COL4A5 polypeptide. In some embodiments: i) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 278 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 293; or ii) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 199 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 220; or iii) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 186 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 222; or iv) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 285 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 300.

In another aspect, the present invention provides dual adeno-associated virus (AAV) vector system for expressing a COL4A5 polypeptide comprising: (a) a first AAV vector comprising a 5' coding sequence (CDS), wherein the 5' CDS encodes an N-terminal part of the COL4A5 polypeptide; and (b) a second AAV vector comprising a 3' CDS, wherein the 3' CDS encodes a C-terminal part of the COL4A5 polypeptide.

The COL4A5 polypeptide is not particularly limited. Any suitable COL4A5 polypeptide may be used. Suitably, the COL4A5 polypeptide is a human COL4A5 polypeptide. In some embodiments, the COL4A5 polypeptide comprises or consists of an amino acid sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at

least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 21 to 26. In some embodiments, the COL4A5 polypeptide comprises or consists of an amino acid sequence having at least 70%,  
 %, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least  
 5 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 21. In some embodiments, the COL4A5 polypeptide comprises or consists of the amino acid sequence of SEQ ID NO: 21.

A CDS encoding the COL4A5 polypeptide may be formed upon delivery of the first AAV vector and the second AAV vector into a host cell. The CDS encoding the COL4A5 polypeptide is  
 10 not particularly limited. Any suitable CDS encoding a COL4A5 polypeptide may be used. In some embodiments, the CDS encoding the COL4A5 polypeptide comprises or consists of the nucleotide sequence of SEQ ID NO: 129.

The 5' CDS may encode any suitable N-terminal part of a COL4A5 polypeptide. Suitably, the 5' CDS comprises or consists of exons 1-15, exons 1-16, exons 1-17, exons 1-18, exons 1-19, exons 1-20, exons 1-21, exons 1-22, exons 1-23, exons 1-24, exons 1-25, exons 1-26,  
 15 exons 1-27, exons 1-28, exons 1-29, exons 1-30, exons 1-31, exons 1-32, exons 1-33, exons 1-34, exons 1-35, exons 1-36, exons 1-37, exons 1-38, exons 1-39, exons 1-40, or exons 1-41 of a CDS encoding a COL4A5 polypeptide. In some embodiments, the 5' CDS comprises or consists of exons 1-27, exons 1-28, exons 1-29, exons 1-30, exons 1-31, exons 1-32, or  
 20 exons 1-33 of a CDS encoding a COL4A5 polypeptide, optionally wherein the 5' CDS comprises or consists of exons 1-29, exons 1-31, exons 1-32, or exons 1-33 of a CDS encoding a COL4A5 polypeptide. In some embodiments, the 5' CDS comprises or consists of a nucleotide sequence having at least 70% identity to any of SEQ ID NOs: 195 to 201. . In some embodiments, the 5' CDS comprises or consists of the nucleotide sequence selected  
 25 from any of SEQ ID NOs: 195 to 201. In some embodiments, the 5' CDS comprises or consists of the nucleotide sequence selected from any of SEQ ID NOs: 197, 199, 200, or 201. In some embodiments, the 5' CDS comprises or consists of exons 1-40 of a CDS encoding a COL4A5 polypeptide. In some embodiments, the 5' CDS comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least  
 30 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 287. In some embodiments, the 5' CDS comprises or consists of SEQ ID NO: 287.

The 3' CDS may encode any suitable C-terminal part of a COL4A5 polypeptide. Suitably, the 3' CDS comprises or consists of exons 42-51, exons 41-51, exons 40-51, exons 39-51, exons  
 35 38-51, exons 37-51, exons 36-51, exons 35-51, exons 34-51, exons 33-51, exons 32-51,

exons 31-51, exons 30-51, exons 29-51, exons 28-51, exons 27-51, exons 26-51, exons 25-51, exons 24-51, exons 23-51, exons 22-51, exons 21-51, exons 20-51, exons 19-51, exons 18-51, exons 17-51, or exons 16-51 of a CDS encoding a COL4A5 polypeptide. In some embodiments, the 3' CDS comprises or consists of exons 36-41, exons 35-51, exons 34-51, 5 exons 33-51, exons 32-51, exons 31-51, exons 30-51, exons 29-51, exons 28-51 or exons 27-51, of a CDS encoding a COL4A5 polypeptide. In some embodiments, the 3' CDS comprises or consists of a nucleotide sequence selected from any of SEQ ID NOs: 295 to 297, SEQ ID NOs: 216 to 222 or SEQ ID NOs: 308 to 310. In some embodiments, the 3' CDS comprises or consists of exons 34-51, exons 33-51, exons 32-51, exons 31-51, exons 30-51, 10 exons 29-51, or exons 28-51, of a CDS encoding a COL4A5 polypeptide. In some embodiments, the 3' CDS comprises or consists of exons 34-51, exons 33-51, exons 32-51, or exons 30-51 of a CDS encoding a COL4A5 polypeptide. In some embodiments, the 3' CDS comprises or consists of a nucleotide sequence selected from any of SEQ ID NOs: 216 to 222. In some embodiments, the 3' CDS comprises or consists of a nucleotide sequence selected from any of SEQ ID NOs: 218, 220, 221, or 222. In some embodiments, the 3' CDS comprises or consists of a nucleotide sequence selected from any of SEQ ID NOs: 217, 219 or 297. 15

The 5' CDS and the 3' CDS may together be used to reconstitute the full-length CDS upon co-delivery of the first AAV vector and the second AAV vector. In some embodiments: i) the 5' CDS consists of exons 1-24 of a CDS encoding a COL4A5 polypeptide and the 3' CDS 20 consists of exons 25-51 of a CDS encoding a COL4A5 polypeptide; or ii) the 5' CDS consists of exons 1-31 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 32-51 of a CDS encoding a COL4A5 polypeptide; or iii) the 5' CDS consists of exons 1-32 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 33-51 of a CDS encoding a COL4A5 polypeptide; or iv) the 5' CDS consists of exons 1-33 of a CDS encoding 25 a COL4A5 polypeptide and the 3' CDS consists of exons 34-51 of a CDS encoding a COL4A5 polypeptide; or vv) the 5' CDS consists of exons 1-38 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 39-51 of a CDS encoding a COL4A5 polypeptide. In some embodiments; i) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 278 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 293; or ii) the 30 5' CDS consists of the nucleotide sequence of SEQ ID NO: 199 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 220; or iii) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 186 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 222; or iv) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 285 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 300..

In some embodiments the 5' CDS consists of exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 36-41, exons 35-51, exons 34-51, exons 33-51, exons 32-51, exons 31-51, exons 30-51, exons 29-51, exons 28-51 or exons 27-51 of a CDS encoding a COL4A5 polypeptide. In some embodiments: (i) the 5' CDS consists of exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 36-41 of a CDS encoding a COL4A5 polypeptide; (ii) the 5' CDS consists of exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 35-41 of a CDS encoding a COL4A5 polypeptide; (iii) the 5' CDS consists of exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 34-41 of a CDS encoding a COL4A5 polypeptide; (iv) the 5' CDS consists of exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 33-41 of a CDS encoding a COL4A5 polypeptide; (v) the 5' CDS consists of exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 32-41 of a CDS encoding a COL4A5 polypeptide; (vi) the 5' CDS consists of exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 31-41 of a CDS encoding a COL4A5 polypeptide; (vii) the 5' CDS consists of exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 30-41 of a CDS encoding a COL4A5 polypeptide; (viii) the 5' CDS consists of exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 29-41 of a CDS encoding a COL4A5 polypeptide; (ix) the 5' CDS consists of exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 28-41 of a CDS encoding a COL4A5 polypeptide; or (x) the 5' CDS consists of exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 27-41 of a CDS encoding a COL4A5 polypeptide. In some embodiments: (i) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 287 (or at least 90% or at least 95% identity) and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 297 (or at least 90% or at least 95% identity); (ii) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 287 (or at least 90% or at least 95% identity) and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 310 (or at least 90% or at least 95% identity); (iii) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 287 (or at least 90% or at least 95% identity) and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 309 (or at least 90% or at least 95% identity); (iv) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 287 (or at least 90% or at least 95% identity) and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 308 (or at least 90% or at least 95% identity); (v) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 287 (or at least 90% or at least 95% identity) and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 219 (or at least 90% or at least 95% identity); (vi) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 287 (or at least 90% or at least 95% identity) and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 217 (or at least 90% or at least 95% identity); or (vii) the 5' CDS consists of the nucleotide sequence of

SEQ ID NO: 287 (or at least 90% or at least 95% identity) and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 295 (or at least 90% or at least 95% identity).

In another aspect, the present invention provides a dual adeno-associated virus (AAV) vector system for expressing a COL4A3, COL4A4 or COL4A5 polypeptide comprising: (a) a first AAV vector comprising a 5' coding sequence (CDS) wherein the 5' CDS encodes an N-terminal part of the COL4A3, COL4A4 or COL4A5 polypeptide; and (b) a second AAV vector comprising a 3' CDS, wherein the 3' CDS encodes a C-terminal part of the COL4A3, COL4A4 or COL4A5 polypeptide, wherein the 5' CDS and the 3' CDS share an overlapping sequence, and wherein the overlapping sequence has a length of at least 200 nucleotides.

10 In another aspect, the present invention provides a dual adeno-associated virus (AAV) vector system for expressing a COL4A5 polypeptide comprising: (a) a first AAV vector comprising a 5' coding sequence (CDS) wherein the 5' CDS encodes an N-terminal part of the COL4A5 polypeptide; and (b) a second AAV vector comprising a 3' CDS, wherein the 3' CDS encodes a C-terminal part of the COL4A5 polypeptide, wherein the 5' CDS and the 3' CDS share an overlapping sequence, and wherein the overlapping sequence has a length of at least 200 nucleotides.

In some embodiments, the overlapping sequence has a length of at least 250 nucleotides, at least 300 nucleotides, at least 350 nucleotides, at least 400 nucleotides, at least 450 nucleotides, at least 500 nucleotides, at least 550 nucleotides, at least 600 nucleotides, at least 700 nucleotides, at least 800 nucleotides, at least 900 nucleotides, or at least 1000 nucleotides. In some embodiments, the overlapping sequence has a length of 1600 nucleotides or fewer. In some embodiments, the overlapping sequence has a length of from 200 nucleotides to 2000 nucleotides, 450 nucleotides to 1600 nucleotides, 500 nucleotides to 1600 nucleotides, from 600 nucleotides to 1600 nucleotides, from 700 nucleotides to 1600 nucleotides, from 800 nucleotides to 1400 nucleotides, from 900 nucleotides to 1300 nucleotides, or from 1000 nucleotides to 1200 nucleotides.

In some embodiments wherein the 5' CDS and the 3' CDS share an overlapping sequence, the 5' CDS is longer than the 3' CDS. In some embodiments wherein the 5' CDS and the 3' CDS share an overlapping sequence, the 3' CDS is longer than the 5' CDS. For both of these approaches, the length of the overlapping sequence suitably ranges from 200-2000 nucleotides, preferably from 450-1600 nucleotides.

In another aspect, the present invention provides a dual adeno-associated virus (AAV) vector system for expressing a COL4A5 polypeptide comprising: (a) a first AAV vector comprising a 5' coding sequence (CDS) wherein the 5' CDS encodes an N-terminal part of the COL4A5

polypeptide; and (b) a second AAV vector comprising a 3' CDS, wherein the 3' CDS encodes a C-terminal part of the COL4A5 polypeptide, wherein the 5' CDS and the 3' CDS do not share an overlapping sequence, and wherein the 5' CDS and the 3' CDS have approximately equal length.

- 5 In another aspect, the present invention provides a dual adeno-associated virus (AAV) vector system for expressing a COL4A5 polypeptide comprising: (a) a first AAV vector comprising a 5' coding sequence (CDS) wherein the 5' CDS encodes an N-terminal part of the COL4A5 polypeptide; and (b) a second AAV vector comprising a 3' CDS, wherein the 3' CDS encodes a C-terminal part of the COL4A5 polypeptide, wherein the 5' CDS and the 3' CDS do not share  
10 an overlapping sequence, and wherein the 3' CDS is longer than the 5' CDS.

In some embodiments, the 3' CDS is at least about 1000 nucleotides longer than the 5' CDS, at least about 1100 nucleotides longer than the 5' CDS, at least about 1200 nucleotides longer than the 5' CDS, at least about 1300 nucleotides longer than the 5' CDS, at least about 1400 nucleotides longer than the 5' CDS, or at least about 1500 nucleotides longer than the 5' CDS.

- 15 In another aspect, the present invention provides a dual adeno-associated virus (AAV) vector system for expressing a COL4A5 polypeptide comprising: (a) a first AAV vector comprising a 5' coding sequence (CDS) wherein the 5' CDS encodes an N-terminal part of the COL4A5 polypeptide; and (b) a second AAV vector comprising a 3' CDS, wherein the 3' CDS encodes a C-terminal part of the COL4A5 polypeptide, wherein the 5' CDS and the 3' CDS do not share  
20 an overlapping sequence, and wherein the 5' CDS is longer than the 3' CDS.

- In some embodiments, the 5' CDS is at least about 1500 nucleotides longer than the 3' CDS, at least about 1600 nucleotides longer than the 3' CDS, at least about 1700 nucleotides longer than the 3' CDS, at least about 1800 nucleotides longer than the 3' CDS, at least about 1900 nucleotides longer than the 3' CDS, at least about 2000 nucleotides longer than the 3' CDS,  
25 at least about 2100 nucleotides longer than the 3' CDS, at least about 2200 nucleotides longer than the 3' CDS, at least about 2300 nucleotides longer than the 3' CDS, at least about 2400 nucleotides longer than the 3' CDS, or at least about 2500 nucleotides longer than the 3' CDS.

- The first AAV vector may comprise a promoter upstream of the 5' CDS. Any suitable promoter may be used. Suitably, the promoter upstream of the 5' CDS is selected from: a CMV promoter, a CBA promoter, a CAG promoter, a CB7 promoter, an EF1 $\alpha$  promoter (optionally  
30 according to either SEQ ID NO: 318 or 319 or at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 318 or 319), a CMV/EF1 $\alpha$  hybrid promoter, a NF-kB promoter, a pSE-7 promoter, a mPGK promoter,



a mU1a promoter, a U6 promoter, a U7 promoter, a MNDU3 promoter, a HLP promoter, an AAT promoter, an ALB promoter, a ApoE/AAT promoter, a EalbAAT promoter, a LP1 promoter, a TBG promoter, a TTR promoter, a SYN1 promoter, a NSE promoter, a tMCK promoter, a CK8 promoter, a MHCK7 promoter, a SMN promoter, a DES promoter, a RK promoter, a hRHO promoter, a GRK1 promoter, a hCAR promoter, a hRPE65p promoter, a P546 promoter, a PR1.7 promoter, a hRS1 promoter, a VMD2 promoter, an  $\alpha$ -MHC promoter, a FRE1 promoter, a NPHS1 promoter, and a NPHS2 promoter. In some embodiments, the promoter upstream of the 5' CDS is a ubiquitous promoter. In some embodiments, the promoter upstream of the 5' CDS is a CMV promoter, a CBA promoter, or a CAG promoter. In some embodiments, the promoter upstream of the 5' CDS comprises or consists of a nucleotide sequence having at least 70% identity to SEQ ID NO: 223. In other embodiments, the promoter upstream of the 5' CDS is a kidney-specific promoter. In some embodiments, the promoter upstream of the 5' CDS is a podocyte-specific promoter. In some embodiments, the promoter upstream of the 5' CDS is a NPHS1 promoter or a NPHS2 promoter. In some embodiments, the promoter upstream of the 5' CDS is a minimal NPHS1 promoter or a minimal NPHS2 promoter. In some embodiments, the promoter upstream of the 5' CDS comprises or consists of a nucleotide sequence having at least 70% identity to any of SEQ ID NOs: 226 to 231.

The second AAV vector may comprise a polyadenylation sequence downstream from the 3' CDS. Any suitable polyadenylation sequence may be used. In some embodiments, the polyadenylation sequence downstream from the 3' CDS is a bovine growth hormone polyadenylation sequence (bGH), a soluble neuropilin-1 polyadenylation sequence, an early SV40 polyadenylation sequence (SV40pA), or a chicken beta-globin polyadenylation sequence. In some embodiments, the polyadenylation sequence downstream from the 3' CDS comprises or consists of a nucleotide sequence having at least 70% identity to any of SEQ ID NOs: 236 to 239.

In some embodiments, the dual AAV vector system is suitable for reconstitution of the CDS at the AAV genome level. The dual AAV vector system may comprise any elements required for reconstitution of the CDS at the AAV genome level.

Suitably, the first AAV vector comprises a splice donor sequence downstream from the 5' CDS and the second AAV vector comprises a splice acceptor sequence upstream from the 3' CDS. Any suitable splice donor sequence and splice acceptor sequence may be used. In some embodiments, the splice donor sequence comprises or consists of a nucleotide sequence having at least 70% identity to SEQ ID NO: 240. In some embodiments, the splice acceptor

sequence comprises or consists of a nucleotide sequence having at least 70% identity to SEQ ID NO: 242.

Suitably, the first AAV vector and the second AAV vector each comprise an identical recombinogenic sequence, wherein the first AAV vector comprises the recombinogenic sequence downstream from a splice donor sequence, and wherein the second AAV vector comprises the recombinogenic sequence upstream from a splice acceptor sequence. Any recombinogenic sequence may be used. Suitably, the recombinogenic sequence is a recombinogenic region from the F1 phage sequence or a recombinogenic region from the alkaline phosphate gene. In some embodiments, the recombinogenic sequence comprises or consists of a nucleotide sequence having at least 70% identity to any of SEQ ID NOs: 244-247. In some embodiments, the recombinogenic sequence is a recombinogenic region from the F1 phage sequence. In some embodiments, the recombinogenic sequence comprises or consists of a nucleotide sequence having at least 70% identity to SEQ ID NO: 244. Any suitable combination of splice donor/acceptor sequence and recombinogenic sequences may be used. In some embodiments: (a) the first AAV vector comprises a nucleotide sequence having at least 70% identity to SEQ ID NO: 248 downstream from the 5' CDS; and (b) the second AAV vector comprises a nucleotide sequence having at least 70% identity to SEQ ID NO: 250 upstream from the 3' CDS. In some embodiments: (a) the first AAV vector comprises a nucleotide sequence having at least 70% identity to SEQ ID NO: 249 downstream from the 5' CDS; and (b) the second AAV vector comprises a nucleotide sequence having at least 70% identity to SEQ ID NO: 251 upstream from the 3' CDS.

The first AAV vector and/or the second AAV vector may comprise any other suitable elements. In some embodiments, the first AAV vector and/or the second AAV vector comprises one or more regulatory sequences. In some embodiments, the first AAV vector comprises one or more regulatory sequences upstream from the 5' CDS. In some embodiments, the first AAV vector comprises a spliceosomal intron or a fragment thereof comprising a branch point sequence, a polypyrimidine tract and a 3' splice acceptor site, upstream from the 5' CDS. In some embodiments, the spliceosomal intron or a fragment thereof comprises or consists of the nucleotide sequence  $(N)_x Y T N A Y (N)_{n1} (Y)_{n2} (N)_{n3} Y A G G$ , wherein  $x = 10$  to  $100$ ,  $n1 = 2$  to  $22$ , wherein  $n2 = 10$  to  $20$ , and wherein  $n1 + n2 + n3 = 15$  to  $40$ . In some embodiments, the spliceosomal intron or a fragment thereof comprises or consists of a nucleotide sequence having 80% or more identity to SEQ ID NO: 256 or 257. In some embodiments, the second AAV vector comprises one or more regulatory sequences downstream from the 3' CDS. In some embodiments, the second AAV vector comprises a Woodchuck hepatitis post-transcriptional regulatory element (WPRE), downstream from the 3' CDS. In some

embodiments, the WPRE comprises or consists of a nucleotide sequence having 80% or more identity to SEQ ID NO: 263.

In other embodiments, the dual AAV vector system is suitable for reconstitution of the CDS at the RNA level. The dual AAV vector system may comprise any elements required for  
5 reconstitution of the CDS at the RNA level.

Suitably, the second AAV vector further comprises a promoter upstream from the 3' CDS. Any suitable promoter may be used. Suitably, the promoter upstream of the 3' CDS is selected from: a CMV promoter, a CBA promoter, a CAG promoter, a CB7 promoter, an EF1 $\alpha$  promoter (optionally according to either SEQ ID NO: 318 or 319 or at least 70%, at least 75%, at least  
10 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 318 or 319), a CMV/EF1 $\alpha$  hybrid promoter, a NF-kB promoter, a pSE-7 promoter, a mPGK promoter, a mU1a promoter, a U6 promoter, a U7 promoter, a MNDU3 promoter, a HLP promoter, an AAT promoter, an ALB promoter, a ApoE/AAT promoter, a EalbAAT promoter,  
15 a LP1 promoter, a TBG promoter, a TTR promoter, a SYN1 promoter, a NSE promoter, a tMCK promoter, a CK8 promoter, a MHCK7 promoter, a SMN promoter, a DES promoter, a RK promoter, a hRHO promoter, a GRK1 promoter, a hCAR promoter, a hRPE65p promoter, a P546 promoter, a PR1.7 promoter, a hRS1 promoter, a VMD2 promoter, an  $\alpha$ -MHC promoter, a FRE1 promoter, a NPHS1 promoter, and a NPHS2 promoter. In some  
20 embodiments, the promoter upstream of the 3' CDS is a ubiquitous promoter. In some embodiments, the promoter upstream of the 3' CDS is a CMV promoter, a CBA promoter, or a CAG promoter. In some embodiments, the promoter upstream of the 3' CDS is a CMV promoter. In some embodiments, the promoter upstream of the 3' CDS comprises or consists of a nucleotide sequence having at least 70% identity to SEQ ID NO: 223. In other  
25 embodiments, the promoter upstream of the 3' CDS is a kidney-specific promoter. In some embodiments, the promoter upstream of the 3' CDS is a podocyte-specific promoter. In some embodiments, the promoter upstream of the 3' CDS is a NPHS1 promoter or a NPHS2 promoter. In some embodiments, the promoter upstream of the 3' CDS is a minimal NPHS1 promoter or a minimal NPHS2 promoter. In some embodiments, the promoter upstream of the  
30 3' CDS comprises or consists of a nucleotide sequence having at least 70% identity to any of SEQ ID NOs: 226 to 231. In some embodiments, the promoter upstream of the 5' CDS and the promoter upstream of the 3' CDS are identical.

Suitably, the first AAV vector further comprises a polyadenylation sequence downstream from the 5' CDS. Any suitable polyadenylation sequence may be used. In some embodiments, the  
35 polyadenylation sequence downstream from the 5' CDS is a bovine growth hormone

polyadenylation sequence (bGH), a soluble neuropilin-1 polyadenylation sequence, an early SV40 polyadenylation sequence (SV40pA), or a chicken beta-globin polyadenylation sequence. In some embodiments, the polyadenylation sequence downstream from the 5' CDS comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 236 to 239.

Suitably, (a) the first AAV vector further comprises a splice donor sequence and a hybridisation domain downstream from the 5' CDS and upstream from the polyadenylation sequence; and (b) the second AAV vector further comprises a complementary hybridisation domain and a splice acceptor sequence upstream from the 3' CDS and downstream from the promoter. In some embodiments, the first AAV vector comprises one or more regulatory sequences upstream from the 5' CDS and/or the second AAV vector comprises one or more regulatory sequences upstream from the 3' CDS. In some embodiments, the first AAV vector comprises a spliceosomal intron or a fragment thereof comprising a branch point sequence, a polypyrimidine tract and a 3' splice acceptor site, upstream from the 5' CDS; and/or the second AAV vector comprises a spliceosomal intron or a fragment thereof comprising a branch point sequence, a polypyrimidine tract and a 3' splice acceptor site, downstream from the 5' CDS. In some embodiments, the spliceosomal intron or a fragment thereof comprises or consists of the nucleotide sequence  $(N)_x Y T N A Y (N)_{n1} (Y)_{n2} (N)_{n3} Y A G G$ , wherein  $x = 10$  to  $100$ ,  $n1 = 2$  to  $22$ , wherein  $n2 = 10$  to  $20$ , and wherein  $n1 + n2 + n3 = 15$  to  $40$ . In some embodiments, the spliceosomal intron or a fragment thereof comprises or consists of a nucleotide sequence having 80% or more identity to SEQ ID NO: 256 or 257. In some embodiments, the first AAV vector comprises one or more regulatory sequences downstream from the 5' CDS, and/or the second AAV vector comprises one or more regulatory sequences downstream from the 3' CDS. In some embodiments, the first AAV vector comprises a Woodchuck hepatitis post-transcriptional regulatory element (WPRE), downstream from the 5' CDS, and/or the second AAV vector comprises a Woodchuck hepatitis post-transcriptional regulatory element (WPRE), downstream from the 3' CDS. In some embodiments, the WPRE comprises or consists of a nucleotide sequence having 80% or more identity to SEQ ID NO: 263.

The first AAV vector and/or the second AAV vector may be provided in the form of an AAV vector particle. In preferred embodiments, the first AAV vector and the second AAV vector are each in the form of AAV vector particles. The AAV vector particles may be encapsidated by any suitable capsid proteins. In some embodiments, the first AAV vector and/or the second AAV vector is in the form of an AAV vector particle encapsidated by LK03, AAV3B, AAV9,

ShH10, AAV-DJ, AAV2, AAV6.2, AAV5 or KP1 capsid (NCBI accession number MN428626) proteins. In some embodiments, the first AAV vector and/or the second AAV vector is in the form of an AAV vector particle encapsidated by capsid proteins comprising or consisting of an amino acid sequence having at least 90% identity to any of SEQ ID NOs: 264 to 271. In some  
5    embodiments, the first AAV vector and the second AAV vector are encapsidated by the same capsid proteins.

In another aspect, the present invention provides an isolated cell comprising the dual AAV vector system according to the present invention.

In another aspect, the present invention provides a pharmaceutical composition comprising the dual  
10    AAV vector system according to the present invention.

In another aspect, the present invention provides a dual AAV vector system, a cell according to or a pharmaceutical composition according to the present invention, for use as a medicament.

In another aspect, the present invention provides use of a dual AAV vector system, a cell or a  
15    pharmaceutical composition according to the present invention, for the manufacture of a medicament.

In another aspect, the present invention provides a product comprising: (a) a first AAV vector comprising a 5' coding sequence (CDS), wherein the 5' CDS encodes an N-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide; and (b) a second AAV vector comprising a 3' CDS,  
20    wherein the 3' CDS encodes a C-terminal part of the COL4A3, COL4A4 or COL4A5 polypeptide, as a combined preparation for simultaneous, separate or sequential use in therapy, wherein a CDS encoding the COL4A3, COL4A4 or COL4A5 polypeptide is formed and the COL4A3, COL4A4 or COL4A5 polypeptide is expressed upon administration of the first AAV vector and the second AAV vector to a subject. The first AAV vector and the second  
25    AAV vector may be any described herein.

In another aspect, the present invention provides a dual AAV vector system, a cell or a pharmaceutical composition according to the present invention, for use in preventing and/or treating Alport Syndrome or any condition in a subject who has a pathogenic variant of COL4A3, COL4A4 or COL4A5.

In another aspect, the present invention provides use of a dual AAV vector system, a cell or a  
30    pharmaceutical composition, for the manufacture of a medicament for preventing and/or treating Alport Syndrome or any condition in a subject who has a pathogenic variant of COL4A3, COL4A4 or COL4A5.

In another aspect, the present invention provides a product comprising: (a) a first AAV vector comprising a 5' coding sequence (CDS), wherein the 5' CDS encodes an N-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide; and (b) a second AAV vector comprising a 3' CDS, wherein the 3' CDS encodes a C-terminal part of the COL4A3, COL4A4 or COL4A5 polypeptide, as a combined preparation for simultaneous, separate or sequential use in preventing and/or treating Alport Syndrome in a subject (or any condition in a subject who has a pathogenic variant of COL4A3, COL4A4 or COL4A5), wherein a CDS encoding the COL4A3, COL4A4 or COL4A5 polypeptide is formed and the COL4A3, COL4A4 or COL4A5 polypeptide is expressed upon administration of the first AAV vector and the second AAV vector to the subject. The first AAV vector and the second AAV vector may be any described herein.

In another aspect, the present invention provides a method of preventing and/or treating Alport Syndrome (any condition in a subject who has a pathogenic variant of COL4A3, COL4A4 or COL4A5) comprising administering a dual AAV vector system, a cell or a pharmaceutical composition according to the present invention, to a subject in need thereof.

In another aspect, the present invention provides a method of preventing and/or treating Alport Syndrome, or any condition in a subject having a pathogenic variant of COL4A3, COL4A4 or COL4A5, comprising administering to a subject in need thereof: (a) a first AAV vector comprising a 5' coding sequence (CDS), wherein the 5' CDS encodes an N-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide; and (b) a second AAV vector comprising a 3' CDS, wherein the 3' CDS encodes a C-terminal part of the COL4A3, COL4A4 or COL4A5 polypeptide, wherein the first AAV vector and the second AAV vector are administered simultaneously, separately or sequentially, and wherein a CDS encoding the COL4A3, COL4A4 or COL4A5 polypeptide is formed and the COL4A3, COL4A4 or COL4A5 polypeptide is expressed in the subject upon administration of the first AAV vector and the second AAV vector to the subject. The first AAV vector and the second AAV vector may be any described herein.

The dual AAV vector system, cell or pharmaceutical composition may be administered by any suitable route and in any suitable dosage. In some embodiments, the first AAV vector and second AAV vector are administered simultaneously. In some embodiments, the first AAV vector, second AAV vector, dual AAV vector system, cell or pharmaceutical composition is administered by renal artery injection, by intraparenchymal injection, by transparenchymal injection, by renal vein injection, by ureteral injection, by intravenous administration, or by subcapsular injection. In some embodiments, the first AAV vector, second AAV vector, dual AAV vector system, cell or pharmaceutical composition is administered by renal artery injection. In some embodiments, the first AAV vector, second AAV vector, dual AAV vector

system, cell or pharmaceutical composition is administered in a dose of from about  $1 \times 10^6$  vg/kg to about  $1 \times 10^{14}$  vg/kg.

In another aspect, the present invention provides a kit comprising (a) a first AAV vector and (b) a second AAV vector. In some embodiments, the kit is for use in preventing and/or treating Alport Syndrome or for use in preventing and/or treating a condition in a subject who has a pathogenic variant of the COL4A3, COL4A4 or COL4A5 gene, optionally wherein the first AAV vector and the second AAV vector are administered simultaneously, separately or sequentially.

In another aspect, the present invention provides a kit or system for production of AAV vector particles comprising: (a) a first transfer vector encoding a first AAV vector; and (b) a second transfer vector encoding a second AAV vector, wherein the first AAV vector and the second AAV vector are any described herein. In some embodiments, the kit further comprises one or more helper vectors encoding AAV replication and capsid proteins.

In another aspect, the present invention provides an isolated cell comprising the kit or system according to the present invention. In some embodiments, the cell is a packaging cell.

## DESCRIPTION OF DRAWINGS

### Figure 1 – Schematic of non-overlapping dual AAV vector systems encoding COL4A5

(A) Schematics showing example 5' AAV vectors encoding an N-terminal part of a COL4A5 polypeptide. (B) Schematics showing corresponding example 3' AAV vectors encoding a C-terminal part of a COL4A5 polypeptide. SD, splice donor; SA, splice acceptor; 77bp F1, recombinogenic region from the F1 phage sequence; and Chr. Int., endogenous COL4A5 intron.

### Figure 2 – Expression of COL4A5 in AD293 cells

A dual AAV vector system comprising (a) a first AAV vector comprising a CMV promoter and a 5' CDS encoding exons 1-31 from COL4A5; and (b) a second AAV vector comprising a 3' CDS encoding exons 32-51 from COL4A5 was delivered to AD293 cells. Western blot of (A) cell lysate and (B) supernatant, 72 hours post-transduction.

### Figure 3 – Expression of COL4A5 in podocytes

A dual AAV vector system comprising (a) a first AAV vector comprising a CMV promoter and a 5' CDS encoding exons 1-31 from COL4A5; and (b) a second AAV vector comprising a 3' CDS encoding exons 32-51 from COL4A5 was delivered to podocytes. Western blot of (A)

cell lysate and **(B)** supernatant, 72 hours post-transduction. **(C)** Fold change in WPRE expression in podocytes determined by qPCR. **(D)** Three dual AAV vector systems: mid-split 31/32 (a first AAV vector comprising a CMV promoter and a 5' CDS encoding exons 1-31 from COL4A5 and a second AAV vector comprising a 3' CDS encoding exons 32-51 from COL4A5),

5 mid-split 32/33 (a first AAV vector comprising a CMV promoter and a 5' CDS encoding exons 1-32 from COL4A5 (Fig 1A4) and a second AAV vector comprising a 3' CDS encoding exons 33-51 from COL4A5 (Fig1B4)) and mid-split 33/34 (a first AAV vector comprising a CMV promoter and a 5' CDS encoding exons 1-33 from COL4A5 and a second AAV vector comprising a 3' CDS encoding exons 34-51 from COL4A5) were delivered to podocytes.

10 Western blot of supernatant, 72 hours post-transduction.

**Figure 4 – Non-overlapping CDS dual AAV vector systems encoding COL4A5 (with hybrid asymmetric split sites, trans-splicing and no intron constructs) evaluated**

Hybrid designs with asymmetric split sites were evaluated (24/25 and 38/39). Trans-splicing and no intron constructs were also evaluated (both 31/32).

15 **(A)** Shows a schematic of the evaluated AAV vector designs. **(B)** Shows the results. Non-transduced cells, and cells transduced with the 3' vectors alone demonstrate no detectable COL4A5 expression. Cells transduced with the min/max 24/25 hybrid design demonstrated expression of a single, full-length (~200 kD) species of COL4A5, with no other truncated COL4A5 species. Cells transduced with the min/max 38/39 hybrid design or the trans-splicing

20 design demonstrated expression of a full-length (~200 kD) species of COL4A5 together with truncated protein species.

**Figure 5 – Schematic of overlapping dual AAV vector systems encoding COL4A5 with a longer 5' CDS**

25 Schematic of an example 5' AAV vector and variable length 3' AAV vectors encoding overlapping parts of a COL4A5 polypeptide. ITR, inverted terminal repeat; CMV, CMV promoter; V5, V5-tag; WPRE, Woodchuck hepatitis post-transcriptional regulatory element; bGH, bovine growth hormone polyadenylation sequence.

**Figure 6 – Expression of COL4A5 in podocytes**

30 Conditionally immortalised podocytes were transduced with the dual AAV vector systems shown in Figure 5 and expression of COL4A5 determined by Western Blot using **(A,B)** an anti-V5 antibody or **(C)** an anti-COL4A5 antibody (H53). **(A)** lane 4 = ladder, lane 5 = non-transduced, lane 6 = 5' AAV only (exons 1-40), lane 7 = 3' AAV only (exons 29-51), lanes 8-9 = 5'+3' AAV vectors (overlap of 1360 bp), lane 10 = empty, lane 11 = ladder, lane 12 = non-



transduced, lane 13 = 5' AAV only (exons 1-40), lane 14 = 3' AAV only (exons 36-51), lanes 15-16 = 5'+3' AAV vectors (overlap of 498 bp). **(B)** lane 1 = ladder, lane 2= positive control, lane 3 = empty, lane 4 = ladder, lane 5 = non-transduced, lane 6 = 5' AAV only (exons 1-40), lane 7 = 3' AAV only (exons 38-51), lanes 8-9 = 5'+3' AAV vectors (overlap of 230 bp), lane 10 = empty, lane 11 = ladder, lane 12 = non-transduced, lane 13 = 5' AAV only (exons 1-40), lane 14 = 3' AAV only (exons 31-51), lanes 15-16 = 5'+3' AAV vectors (overlap of 1095 bp). **(C)** pos. ctrl = positive control, NT = non-transduced, 5' = 5' AAV only. **(D)** Conditionally immortalised podocytes were transduced with a dual AAV vector system with the 5' construct as shown in Figure 5 (that was modified to replace the CMV promoter with the 265bp NPHS1 promoter) and 3' construct number 3 shown in Figure 5. A full-length COL4A5 species was detected. **(E)** Conditionally immortalised podocytes were transduced with the dual AAV vector systems shown in Figure 5 (the 5' construct and the 3' constructs 2, 3 and 4) expressing murine COL4A5.

#### Figure 7 - Schematic of overlapping dual AAV vector systems encoding COL4A5 with a longer 3' CDS

Schematics of example variable length 5' AAV vectors and a 3' AAV vector encoding overlapping parts of a COL4A5 polypeptide. ITR, inverted terminal repeat; CMV, CMV promoter V5, V5-tag; WPRE, Woodchuck hepatitis post-transcriptional regulatory element; bGH, bovine growth hormone polyadenylation sequence.

#### Figure 8 - Expression of COL4A5 polypeptide in podocytes

Expression of COL4A5 in conditionally immortalised podocytes transduced with the dual AAV vector systems shown in Figure 7 determined by **(A)** Western Blot using an anti-V5 antibody, and **(B)** using a COL4A5-specific enzyme-linked immunosorbent assay (ELISA).

#### DETAILED DESCRIPTION

Various preferred features and embodiments of the present invention will now be described by way of non-limiting examples. This disclosure is not limited by the exemplary methods and materials disclosed herein, and any methods and materials similar or equivalent to those described herein can be used in the practice or testing of embodiments of this disclosure. The skilled person will understand that they can combine all features of the invention disclosed herein without departing from the scope of the invention as disclosed.

It must be noted that as used herein and in the appended claims, the singular forms "a", "an", and "the" include plural referents unless the context clearly dictates otherwise.

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

Numeric ranges are inclusive of the numbers defining the range. As used herein the term "about" means approximately, in the region of, roughly, or around. When the term "about" is used in conjunction with a numerical value or range, it modifies that value or range by  
10 extending the boundaries above and below the numerical value(s) set forth. In general, the terms "about" and "approximately" may be used herein to modify a numerical value(s) above and below the stated value(s) by 10%.

Unless otherwise indicated, any nucleic acid sequences are written left to right in 5' to 3' orientation; amino acid sequences are written left to right in amino to carboxy orientation,  
15 respectively.

The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that such publications constitute prior art to the claims appended hereto.

All publications mentioned in the specification are herein incorporated by reference.

## 20 **Dual AAV vector systems**

The present invention provides a dual adeno-associated virus (AAV) vector system for expressing a COL4A3, COL4A4 or COL4A5 polypeptide.

As used herein, a "dual AAV vector system" may refer to a combination of two separate AAV vectors which each comprise part of a full-length coding sequence (CDS), wherein upon  
25 delivery into a host cell the full-length CDS is formed. To overcome packaging capacity limitation, a full-length CDS may split into two parts and packaged into two separate AAV vectors, typically referred to as dual AAVs. Depending on the approach, the split parts of the CDS can subsequently be reconstituted at the AAV genome, mRNA, or protein level upon co-delivery of the dual AAVs to the target tissue (see e.g. McClements, M.E. and MacLaren, R.E.,  
30 2017. The Yale journal of biology and medicine, 90(4), p.611; and Riedmayr, L.M., et al., 2023. Nature Communications, 14(1), p.6578).

The dual AAV vector system of the present invention comprises: (a) a first AAV vector comprising a 5' coding sequence (CDS), wherein the 5' CDS encodes an N-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide; and (b) a second AAV vector comprising a 3' CDS, wherein the 3' CDS encodes a C-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide.

- 5 A CDS encoding a COL4A3, COL4A4 or COL4A5 polypeptide may be formed upon delivery of the first AAV vector and the second AAV vector into a host cell.

As used herein, a "CDS" (also known as a "coding sequence" or "coding region") may refer to a portion of DNA or RNA that encodes for a protein. In DNA, a CDS is typically flanked by a promoter sequence at the 5' end and a termination sequence at the 3' end. In mRNA, a CDS  
10 is typically flanked by a 5' untranslated region (5'-UTR) at the 5' and a 3' untranslated region (3'-UTR), 5' cap, and poly-A tail at the 3' end.

As used herein, a "5' coding sequence (CDS)" may refer to a coding sequence encoding an N-terminal part of a protein and a "3' coding sequence (CDS)" may refer to a coding sequence encoding a C-terminal part of a protein. A 5' CDS and a 3' CDS may together encode a full-  
15 length protein. In some embodiments, there is no overlap between the 5' CDS and the 3' CDS (e.g. in trans-splicing strategies). In other embodiments, there is some overlap between 5' CDS and the 3' CDS (e.g. in overlapping strategies).

#### COL4A3, COL4A4 and COL4A5 polypeptides

The COL4A3, COL4A4 or COL4A5 polypeptides expressed by the dual AAV vector system of  
20 the present invention are not particularly limited and may be any naturally-occurring or variant COL4A3, COL4A4 or COL4A5 polypeptide known in the art. Suitable COL4A3, COL4A4 or COL4A5 polypeptides are described below.

Type IV collagen forms a complex branch network that is a major component of basement membranes and comprises a family of triple helical isoforms. Each collagen molecule is  
25 composed of three  $\alpha$  chains that share a very similar primary structure: an approximately 25-residue "7S" domain at the amino terminus; a collagenous domain of approximately 1400 Gly-X-Y repeats, which is interrupted by short, non-collagenous regions, and forms the triple helix together with two other  $\alpha$  chains; and an approximately 230-residue non-collagenous (NC1) domain at the carboxyl terminus, which is folded into a globular structure. The  $\alpha$ 3,  $\alpha$ 4, and  $\alpha$ 5  
30 chains are strongly expressed in the highly specialized glomerular BM (GBM) in the kidney where they form a distinct network characterized by loops and supercoiled triple helices that are stabilized by disulfide bonds (see e.g. Heidet, L., et al., 2001. Journal of the American Society of Nephrology, 12(1), pp.97-106).

The COL4A3, COL4A4 or COL4A5 polypeptides may be full-length COL4A3, COL4A4 or COL4A5 polypeptides. As used herein, a “full-length COL4A3, COL4A4 or COL4A5 polypeptide” may refer to a polypeptide that is capable of forming a collagen IV  $\alpha$ 345 network. Full-length COL4A3, COL4A4 or COL4A5 polypeptides may include any naturally-occurring  
5 isoforms and/or variants thereof, provided that they are capable of forming a collagen IV  $\alpha$ 345 network. Suitably, a full-length COL4A3, COL4A4 or COL4A5 polypeptide may comprise a 7S domain, a collagenous domain, and a NC1 domain.

In preferred embodiments, the COL4A3, COL4A4 or COL4A5 polypeptides are human COL4A3, COL4A4 or COL4A5 polypeptides. In other preferred embodiments, the COL4A3,  
10 COL4A4 or COL4A5 polypeptides are full-length human COL4A3, COL4A4 or COL4A5 polypeptides.

The COL4A3, COL4A4 or COL4A5 polypeptides may comprise a signal sequence. As used herein, a “signal sequence” (also known as a signal peptide, targeting signal, localisation sequence, or leader sequence) may refer a short peptide (usually 16-30 amino acids long),  
15 typically present at the N-terminus of newly synthesized proteins that are destined toward the secretory pathway. In some embodiments, the COL4A3, COL4A4 or COL4A5 polypeptides comprise an endogenous signal sequence. In some embodiments, the COL4A3, COL4A4 or COL4A5 polypeptides comprise an alternative signal sequence. Any suitable signal sequence known in the art may be used (see e.g. O’Neill, P., et al., 2023. ACS Synthetic Biology, 12(8),  
20 pp.2339-2352). In some embodiments, the COL4A3, COL4A4 or COL4A5 polypeptides do not comprise a signal sequence.

#### *COL4A3 polypeptides*

In one embodiment, the present invention provides a dual AAV vector system for expressing a COL4A3 polypeptide.

25 The COL4A3 polypeptide may be any suitable COL4A3 polypeptide known in the art. Example human COL4A3 polypeptides include COL4A3 polypeptides having the UniProtKB accession number Q01955, for example isoforms Q01955-1, Q01955-2, Q01955-3, Q01955-4, or Q01955-5. Other example human COL4A3 polypeptides include COL4A3 polypeptides from NCBI Gene ID: 1285, for example isoforms having NCBI accession numbers XP\_005246334,  
30 XP\_011508857, or XP\_047299180.

The COL4A3 polypeptide may be a variant of any COL4A3 polypeptide known in the art. A person skilled in the art would be able to generate variants of known COL4A3 polypeptides using conservative substitutions, based on the known structural and functional features of

COL4A3 (see e.g. Casino, P., et al., 2018. IUCrJ, 5(6), pp.765-779; and Boudko, S.P., et al., 2021. Journal of Biological Chemistry, 296), and/or based on known variants (see e.g. NCBI Gene ID: 1285). Suitably, a variant COL4A3 polypeptide comprises a 7S domain, a collagenous domain and a NC1 domain and retains the ability to form a collagen IV  $\alpha$ 345 network.

The COL4A3 polypeptide may comprise a signal sequence. For example, amino acids 1-28 of each of SEQ ID NOs: 1 to 7 may correspond to a COL4A3 signal sequence. In some embodiments, the COL4A3 polypeptide signal sequence is absent or substituted by an alternative signal sequence. In some embodiments, the COL4A3 polypeptide lacks a signal sequence. In some embodiments, the COL4A3 polypeptide has an alternative signal sequence.

In some embodiments, the COL4A3 polypeptide has a length of 1400 amino acids or more, 1450 amino acids or more, 1500 amino acids or more, 1550 amino acids or more, 1600 amino acids or more, 1650 amino acids or more, 1655 amino acids or more, 1660 amino acids or more, 1665 amino acids or more, or 1670 amino acids or more. In some embodiments, the COL4A3 polypeptide has a length of 1700 amino acids or less, 1690 amino acids or less, 1680 amino acids or less, 1675 amino acids or less, or 1670 amino acids or less. In some embodiments, the COL4A3 polypeptide has a length of from 1640 amino acids to 1670 amino acids. In some embodiments, the COL4A3 polypeptide has a length of 1670 amino acids.

Suitably, a COL4A3 polypeptide may comprise or consist of an amino acid sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 1 to 7 or fragments thereof lacking amino acids 1-28.

Suitably, a COL4A3 polypeptide may comprise or consist of the amino acid sequence of any of SEQ ID NOs: 1 to 7 or fragments thereof lacking amino acids 1-28. Suitably, a COL4A3 polypeptide may consist of the amino acid sequence of any of SEQ ID NOs: 1 to 7 or fragments thereof lacking amino acids 1-28.

Suitably, a COL4A3 polypeptide may comprise or consist of the amino acid sequence of any of SEQ ID NOs: 1 to 7. Suitably, a COL4A3 polypeptide may consist of the amino acid sequence of any of SEQ ID NOs: 1 to 7.

In some embodiments, a COL4A3 polypeptide comprises or consists of the amino acid sequence of SEQ ID NO: 1 or a fragment thereof lacking amino acids 1-28. In some

embodiments, a COL4A3 polypeptide consists of the amino acid sequence of SEQ ID NO: 1 or a fragment thereof lacking amino acids 1-28.

In some embodiments, a COL4A3 polypeptide comprises or consists of the amino acid sequence of SEQ ID NO: 1. In some embodiments, a COL4A3 polypeptide consists of the amino acid sequence of SEQ ID NO: 1.

#### *COL4A4 polypeptides*

In one embodiment, the present invention provides a dual AAV vector system for expressing a COL4A4 polypeptide.

The COL4A4 polypeptide may be any suitable COL4A4 polypeptide known in the art. Example human COL4A4 polypeptides include COL4A4 polypeptides having the UniProtKB accession number P53420. Other example human COL4A4 polypeptides include COL4A4 polypeptides from NCBI Gene ID: 1286, for example isoforms having NCBI accession numbers XP\_011508859, XP\_011508860, XP\_016858786, XP\_006712309, XP\_047299197, XP\_011508863, XP\_005246339, XP\_011508867, XP\_011508869, XP\_011508870, or XP\_011508872.

The COL4A4 polypeptide may be a variant of any COL4A4 polypeptide known in the art. A person skilled in the art would be able to generate variants of known COL4A4 polypeptides using conservative substitutions, based on the known structural and functional features of COL4A4 (see e.g. Casino, P., et al., 2018. IUCrJ, 5(6), pp.765-779; and Boudko, S.P., et al., 2021. Journal of Biological Chemistry, 296), and/or based on known variants (see e.g. NCBI Gene ID: 1286). Suitably, a variant COL4A4 polypeptide comprises a 7S domain, a collagenous domain and a NC1 domain and retains the ability to form a collagen IV  $\alpha$ 345 network.

The COL4A4 polypeptide may comprise a signal sequence. For example, amino acids 1-38 of each of SEQ ID NOs: 8 to 20 may correspond to a COL4A4 signal sequence. In some embodiments, the COL4A4 polypeptide signal sequence is absent or substituted by an alternative signal sequence. In some embodiments, the COL4A4 polypeptide lacks a signal sequence. In some embodiments, the COL4A4 polypeptide has an alternative signal sequence.

In some embodiments, the COL4A4 polypeptide has a length of 1400 amino acids or more, 1450 amino acids or more, 1500 amino acids or more, 1550 amino acids or more, 1600 amino acids or more, 1650 amino acids or more, 1660 amino acids or more, 1670 amino acids or more, 1675 amino acids or more, 1680 amino acids or more, 1685 amino acids or more, or

1690 amino acids or more. In some embodiments, the COL4A4 polypeptide has a length of 1720 amino acids or less, 1710 amino acids or less, 1700 amino acids or less, 1695 amino acids or less, or 1690 amino acids or less. In some embodiments, the COL4A4 polypeptide has a length of from 1650 amino acids to 1690 amino acids. In some embodiments, the

5 COL4A4 polypeptide has a length of 1690 amino acids.

Suitably, a COL4A4 polypeptide may comprise or consist of an amino acid sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 8 to 20 or fragments thereof lacking amino

10 acids 1-38.

Suitably, a COL4A4 polypeptide may comprise or consist of the amino acid sequence of any of SEQ ID NOs: 8 to 20 or fragments thereof lacking amino acids 1-38. Suitably, a COL4A4 polypeptide may consist of the amino acid sequence of any of SEQ ID NOs: 8 to 20 or fragments thereof lacking amino acids 1-38.

15 Suitably, a COL4A4 polypeptide may comprise or consist of the amino acid sequence of any of SEQ ID NOs: 8 to 20. Suitably, a COL4A4 polypeptide may consist of the amino acid sequence of any of SEQ ID NOs: 8 to 20.

In some embodiments, a COL4A4 polypeptide comprises or consists of the amino acid sequence of SEQ ID NO: 8 or a fragment thereof lacking amino acids 1-38. In some

20 embodiments, a COL4A4 polypeptide consists of the amino acid sequence of SEQ ID NO: 8 or a fragment thereof lacking amino acids 1-38.

In some embodiments, a COL4A4 polypeptide comprises or consists of the amino acid sequence of SEQ ID NO: 8. In some embodiments, a COL4A4 polypeptide consists of the amino acid sequence of SEQ ID NO: 8.

25 *COL4A5 polypeptides*

In one embodiment, the present invention provides a dual AAV vector system for expressing a COL4A5 polypeptide.

The COL4A5 polypeptide may be any suitable COL4A5 polypeptide known in the art. Example human COL4A5 polypeptides include COL4A5 polypeptides having the UniProtKB accession

30 number P29400, for example isoforms P29400-1 or P29400-2. Other example human COL4A5 polypeptides include COL4A5 polypeptides from NCBI Gene ID: 1287, for example

isoforms having NCBI accession numbers XP\_011529151, XP\_016884748, XP\_016884749, or XP\_047297766.

The COL4A5 polypeptide may be a variant of any COL4A5 polypeptide known in the art. A person skilled in the art would be able to generate variants of known COL4A5 polypeptides using conservative substitutions, based on the known structural and functional features of COL4A5 (see e.g. Casino, P., et al., 2018. IUCrJ, 5(6), pp.765-779; and Boudko, S.P., et al., 2021. Journal of Biological Chemistry, 296), and/or based on known variants (see e.g. NCBI Gene ID: 1287). Suitably, a variant COL4A5 polypeptide comprises a 7S domain, a collagenous domain and a NC1 domain and retains the ability to form a collagen IV  $\alpha$ 345 network.

The COL4A5 polypeptide may comprise a signal sequence. For example, amino acids 1-26 of each of SEQ ID NOs: 21 and 24 may correspond to a COL4A5 signal sequence. In some embodiments, the COL4A5 polypeptide signal sequence is absent or substituted by an alternative signal sequence. In some embodiments, the COL4A5 polypeptide lacks a signal sequence. In some embodiments, the COL4A5 polypeptide has an alternative signal sequence.

In some embodiments, the COL4A5 polypeptide has a length of 1400 amino acids or more, 1450 amino acids or more, 1500 amino acids or more, 1550 amino acids or more, 1600 amino acids or more, 1650 amino acids or more, 1655 amino acids or more, 1660 amino acids or more, 1665 amino acids or more, 1670 amino acids or more, 1675 amino acids or more, 1680 amino acids or more, or 1685 amino acids or more. In some embodiments, the COL4A5 polypeptide has a length of 1720 amino acids or less, 1710 amino acids or less, 1700 amino acids or less, 1695 amino acids or less, 1690 amino acids or less, or 1685 amino acids or less. In some embodiments, the COL4A5 polypeptide has a length of from 1650 amino acids to 1685 amino acids. In some embodiments, the COL4A5 polypeptide has a length of 1685 amino acids.

Suitably, a COL4A5 polypeptide may comprise or consist of an amino acid sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 21 to 26 or fragments thereof lacking a signal sequence.

Suitably, a COL4A5 polypeptide may comprise or consist of the amino acid sequence of any of SEQ ID NOs: 21 to 26 or fragments thereof lacking a signal sequence. Suitably, a COL4A5



polypeptide may consist of the amino acid sequence of any of SEQ ID NOs: 21 to 26 or fragments thereof lacking a signal sequence.

Suitably, a COL4A5 polypeptide may comprise or consist of the amino acid sequence of any of SEQ ID NOs: 21 to 26. Suitably, a COL4A5 polypeptide may consist of the amino acid  
5 sequence of any of SEQ ID NOs: 21 to 26.

In some embodiments, a COL4A5 polypeptide comprises or consists of the amino acid sequence of SEQ ID NO: 21 or a fragment thereof lacking amino acids 1-26. In some embodiments, a COL4A5 polypeptide consists of the amino acid sequence of SEQ ID NO: 21 or a fragment thereof lacking amino acids 1-26.

10 In some embodiments, a COL4A5 polypeptide comprises or consists of the amino acid sequence of SEQ ID NO: 21. In some embodiments, a COL4A5 polypeptide consists of the amino acid sequence of SEQ ID NO: 21.

#### COL4A3, COL4A4 and COL4A5 coding sequences

The CDS formed by the dual AAV vector system of the present invention is not particularly  
15 limited and may be any naturally-occurring or modified CDS encoding a COL4A3, COL4A4 or COL4A5 polypeptide known in the art. Suitable coding sequences are known in the art and are described below.

It will be understood by a skilled person that numerous different coding sequences can encode the same polypeptide as a result of the degeneracy of the genetic code. In addition, it is to be  
20 understood that skilled persons may, using routine techniques, make nucleotide substitutions that do not affect the polypeptide sequence encoded by a CDS to reflect the codon usage of any particular host organism in which the polypeptides are to be expressed.

The CDS may be codon-optimised, e.g. codon-optimised for humans. Different cells differ in their usage of particular codons. This codon bias corresponds to a bias in the relative  
25 abundance of particular tRNAs in the cell type. By altering the codons in the CDS so that they are tailored to match with the relative abundance of corresponding tRNAs, it is possible to increase expression. By the same token, it is possible to decrease expression by deliberately choosing codons for which the corresponding tRNAs are known to be rare in the particular cell type. Thus, an additional degree of translational control is available. Codon usage tables are  
30 known in the art for mammalian cells (e.g. humans), as well as for a variety of other organisms.

The CDS may comprise one or more exons. An exon is a part of a gene that will form a part of the final mature mRNA produced by that gene after introns have been removed. In protein-

coding genes, exons can include both the protein-coding sequences and the 5'- and 3'- untranslated regions (UTRs) (see e.g. Aspden, J.L., et al., 2023. Cell Genomics, 3(4)).

The CDS may exclusively comprise coding exonic regions. As used herein, "coding exonic regions" may refer to exons (or fragments thereof) which include protein-coding sequences.

5 For example, exon 1 and nucleotides 1-101 of exon 2 of NM\_000092.5 (encoding a COL4A4 polypeptide) do not include any protein-coding sequences and may therefore be referred to as "non-coding exonic regions". In the context of the present invention, the term "exon" may be used interchangeably with the term "coding exonic region".

10 The exons (or coding exonic regions) of any CDS may be readily identified from the corresponding mRNA precursor (see e.g. NM\_000091.5, NM\_000092.5, and NM\_000495) and examples are provided herein for coding sequences which encode a COL4A3, COL4A4 or COL4A5 polypeptide.

15 The coding sequences disclosed herein may comprise or lack stop codons at their 3' end. The present disclosure encompasses the SEQ ID NOs disclosed herein with the stop codons present or absent.

The CDS may encode a full-length COL4A3, COL4A4 or COL4A5 polypeptide. In preferred embodiments, the CDS encodes a human COL4A3, COL4A4 or COL4A5 polypeptide. In other preferred embodiments, the CDS encodes a full-length human COL4A3, COL4A4 or COL4A5 polypeptide.

#### 20 *COL4A3 coding sequences*

In one embodiment, a CDS encoding a COL4A3 polypeptide is formed upon delivery of the first AAV vector and the second AAV vector into a host cell.

25 The COL4A3 coding sequence may be any suitable COL4A3 coding sequence known in the art. Example human COL4A3 coding sequences include COL4A3 coding sequences from NCBI Gene ID: 1285, for example coding sequences from mRNA having NCBI accession numbers NM\_000091, XM\_005246277, XM\_011510555, or XM\_047443224.

30 The COL4A3 coding sequence may be a variant of any COL4A3 coding sequence known in the art. A person skilled in the art would be able to generate variants of known COL4A3 coding sequences using the degeneracy of the genetic code, based on codon-optimisation, and/or based on known variants.

The COL4A3 coding sequence may comprise a signal sequence. For example, nucleotide 1-84 of SEQ ID NO: 27 may encode a COL4A3 signal sequence. In some embodiments, the COL4A3 signal sequence is absent or substituted by an alternative signal sequence. In some embodiments, the COL4A3 coding sequence lacks a signal sequence. In some embodiments, the COL4A3 coding sequence has an alternative signal sequence.

In some embodiments, a COL4A3 coding sequence comprises or consists of a nucleotide sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 27 or a fragment thereof lacking nucleotides 1-84.

In some embodiments, a COL4A3 coding sequence comprises or consists of the nucleotide sequence of SEQ ID NO: 27 or a fragment thereof lacking nucleotides 1-84. In some embodiments, a COL4A3 coding sequence consists of the nucleotide sequence of SEQ ID NO: 27 or a fragment thereof lacking nucleotides 1-84.

In some embodiments, a COL4A3 coding sequence comprises or consists of the nucleotide sequence of SEQ ID NO: 27. In some embodiments, a COL4A3 coding sequence consists of the nucleotide sequence of SEQ ID NO: 27.

SEQ ID NO: 27 may correspond to the following coding exonic regions (also referred to herein as "exons"):

| Exon | Sequence                                                                                                        |
|------|-----------------------------------------------------------------------------------------------------------------|
| 1    | atgagcgcggcggaccgccccagggcggcaggtgctcctgctgcccgtcctgctggtgctcctggcgcg<br>gcgcccgcagccagcaag (SEQ ID NO: 28)     |
| 2    | ggttggtgtctgtataagacaaaggccagtgcttctgtgacggggccaaaggggagaag (SEQ ID NO: 29)                                     |
| 3    | ggggagaagggttcttctggacccccgggttctcctggccagaaaggattcacagggtcctgaaggcttg<br>cctggaccgcagggacccaag (SEQ ID NO: 30) |
| 4    | ggctttccaggacttcaggactcacgggttccaaagggtgtaagg (SEQ ID NO: 32)                                                   |
| 5    | ggaataagtggttgccaggattttctggttctcctggacttcca (SEQ ID NO: 33)                                                    |
| 6    | ggcaccgccaggcaataaccgggccttacggacttgctcgggtgtaccaggatgcagtggttctaag (SEQ ID NO: 34)                             |
| 7    | ggtgagcaggggttccaggactcccagggacactgggctaccagggatcccg (SEQ ID NO: 35)                                            |
| 8    | ggtgctgctgggttgaaaggacaaaag (SEQ ID NO: 36)                                                                     |
| 9    | ggtgctcctgctaaagaagaagatatagaacttgatgcaaaaggcgacccccgggttgccaggggctcca<br>ggacccccag (SEQ ID NO: 37)            |
| 10   | ggtttgccaggccctccagggttttctgggcctgttggccacctggctcctccgggattcttt (SEQ ID NO: 38)                                 |
| 11   | ggctttccaggagccatgggacctagaggacctaag (SEQ ID NO: 39)                                                            |
| 12   | ggtcacatgggtgaaagagtgataggacataaaggagagcgg (SEQ ID NO: 40)                                                      |
| 13   | ggtgtgaaaggggttaacaggacccccgggaccaccaggaacagttattgtgacctaaactggccagat<br>aacagaacg (SEQ ID NO: 41)              |
| 14   | gacctcaagggggaaaaaggagacaaggagcaatgggcgagcctggacctcctggaccctca (SEQ ID NO: 42)                                  |

|    |                                                                                                                                                                                                        |
|----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 15 | ggactgcttgagaaatcatatggatctgaaaaggggtgctcctggagaccctggcctgcag (SEQ ID NO: 43)                                                                                                                          |
| 16 | ggaaaaccggaaaagatgggtgttctctggcttccctggaagtgcag (SEQ ID NO: 44)                                                                                                                                        |
| 17 | ggagtcaagggaacacaggggtttccctgggttaatgggtgaagatggcattaag (SEQ ID NO: 45)                                                                                                                                |
| 18 | ggacagaaaggggacattggccctccaggatttctgtgtcca (SEQ ID NO: 46)                                                                                                                                             |
| 19 | acagaatattatgacacataccaggaaaaggagatgaaggcactccaggcccaccaggggccagagga gctcgtggcccaag (SEQ ID NO: 47)                                                                                                    |
| 20 | gtcccagtggtccccccggagttcctggaagtctctg (SEQ ID NO: 48)                                                                                                                                                  |
| 21 | gatcatcaaggcctggcctcagaggagccccctggatggccaggcctgaaaggaagtaaagggaacgag gccgcccaggaaaaggatgccatggggactcctgggtccccagggtgtgctggttcaccagggtcttccag gatcacccgggacctccaggaccgccag (SEQ ID NO: 49)             |
| 22 | gtgacatcggtttttcgcaaggggtccacctggagatcacggactgccaggctatctagggtctccaggaa tcccaggagttgatgggccccaaag (SEQ ID NO: 50)                                                                                      |
| 23 | gagaaccaggcctcctgtgtacacagtgccttatatcccagggcctcccgggtctcccaggattgccag ggttacatggtgtaaaaggaatcccag (SEQ ID NO: 51)                                                                                      |
| 24 | gaagacaaggcgagctggcttgaaaggaagcccagggtccccaggaaaatacagggtcttccaggatttc ca (SEQ ID NO: 52)                                                                                                              |
| 25 | ggtttcccagggtgccagggtgaccaggacttaaggagaaaaaggtgaaacacttcagcctgagggg caagtgggtgtcccagggtgaccgggggtcagaggccaacctgggagaaaagggttgatggaattcct ggaactccgggagtgaaaggattaccaggacctaaggcggaactg (SEQ ID NO: 53) |
| 26 | gctctgagtggtgagaaaggggaccaaggtcctccaggggatcctgggtccccctgggtccccaggacct gcaggaccagctggaccacctggctacggaccccaaggagaacctgggtctccagggcacgcaaggagtt cctggagcccccgaccaccgggagaaagccg (SEQ ID NO: 54)          |
| 27 | gccctaggggagagctcagtggtttcaacaccagttccaggcccaccaggacctccaggggccccctggcc atcctggcccccaaggtccacctg (SEQ ID NO: 55)                                                                                       |
| 28 | gtatccctggatccctggggaatgtggagatcctgggtcttccagggcctgatggtgaaccagggaattc caggaattggatttctctgggcctcctggacctaagg (SEQ ID NO: 56)                                                                           |
| 29 | gagaccaaggttttccagggtacaaaaggatcactgggttgctcctggaaaaatgggagagcctgggttac ctggaagccaggcctcccaggagccaag (SEQ ID NO: 57)                                                                                   |
| 30 | ggagaaccagcagtagccatgcctggaggaccaggaacaccagggttttccaggagaaaagaggcaattct ggggaacatggagaaattggactccctggacttccagggtctcctggaaactccaggaaatgaagggtt gatggaccacgag (SEQ ID NO: 58)                            |
| 31 | gagatccagggcagcctggaccacctggagaacaaggacccccaggaggtgcatagagggtcccaggg gagcccaaggacttccagggttaaatggattgaaagggaacaag (SEQ ID NO: 59)                                                                      |
| 32 | gcagaagaggtaaaaacggggccaaaggagaccagggaattccaggccttgatagatcaggatttctctg gagaaactggatcaccagggaattccagggtcatcaaggtgaaatgggaccactgggtcaaagaggatata caggaaatccgggaatttttagggccaccag (SEQ ID NO: 60)         |
| 33 | gtgaagatggagtgattgggatgatgggttttctctggagccattggcctccaggggccccctgggaacc caggcacaccagggcagaggg (SEQ ID NO: 61)                                                                                           |
| 34 | ggagccctggaattccaggagtaaaaggccagagaggaacccccaggagccaagggggaacaaggagata aaggaaatccccggccttcagagatatcccagctaaggggacaaaggagaaccagggtctcaaag (SEQ ID NO: 62)                                               |
| 35 | gattcgcaggaaatccagggtgagaaaggaaacagaggcgttccagggtgaccagggtttaaaggccctca aaggactaccggaccagcaggaccaccag (SEQ ID NO: 63)                                                                                  |
| 36 | gccccagaggagatttgggcagcactgggaatcctggagaaccaggactgcgtggtataaccagggaagca tggggaacatgggcattgccag (SEQ ID NO: 64)                                                                                         |
| 37 | gttctaaaggaaaaagggaactttgggatttccagggtcagcaggaagaccaggcctcccagggtattc atggtctccagggaataaggagagccagggttatcagaaggtacaaggccaggaccaccgggaccaa cg (SEQ ID NO: 65)                                           |
| 38 | ggggatccaggactgccgggtgatatgggaaagaaaggagaaaatggggcaacctggcccacctggacat ttggggcctgctggacctgagggagccccctggaagtcctggaagtcctggcctcccag (SEQ ID NO: 66)                                                     |
| 39 | gaaagccagggtcctcatggtgatttgggttttaaaaggaatcaaaggcctcctgggcccctccaggaaatca gaggccctccag (SEQ ID NO: 67)                                                                                                 |
| 40 | gtcttccaggatttccaggatctcctggaccaatgggtataagaggtgaccaaggacgtgatggaattc ctggtccagccggagaaaaggagaaacgg (SEQ ID NO: 68)                                                                                    |
| 41 | gtttattgagggccccctccaggcccaaggagggaacctggtgctcaag (SEQ ID NO: 69)                                                                                                                                      |

|    |                                                                                                                                                                                                                                                |
|----|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 42 | gagccaaaggagacagggagccccaggttttctggcctcccgggcagaaaaaggggcatgggagatg<br>ctggacctcgaggacccacagggcatagaaggattcccagggccaccaggtctgcccgggtgcaattatcc<br>ctggccagacaggaatcgtggtccaccaggtcaagaggaagcccag (SEQ ID NO: 70)                               |
| 43 | gtgcgcctgggtccccctggacctccagggagtcagtgaataggcataaaaggagacaaaggtctatgg<br>gccacctggcccaaaaggtccacctggaactgcaggagacatgggaccaccaggtcgtctg (SEQ<br>ID NO: 71)                                                                                      |
| 44 | ggagcaccaggtactccaggtcttccaggacccagaggtgatcctggattccaggggtttccaggcgtg<br>aaag (SEQ ID NO: 72)                                                                                                                                                  |
| 45 | gagaaaagggtaatcctggatttctaggatccattggacctccaggaccaattgggccaagggaccac<br>ctg (SEQ ID NO: 73)                                                                                                                                                    |
| 46 | gtgtacgtggagacctggcacacttaagattatctcccttccaggaagcccagggccacctggcacac<br>ctggagaaccagggatgcagggagaacctgggcccaccagggccacctggaaacctag (SEQ ID NO:<br>74)                                                                                          |
| 47 | gacctgtgggccaagaggttaagccaggcaaggatggaaaaccaggaactcctggaccagctggagaaa<br>aaggcaacaaaggttctaaaggagagccag (SEQ ID NO: 75)                                                                                                                        |
| 48 | gaccagctggatcagatggattgccaggtttgaaaggaaaacgtggagacagtggatcacctgcaacct<br>ggacaacgagaggtttgtcttcaccgacacagtcacaaaccacagcaattccttcattgtccagagggga<br>cagtggcactctacagtgggttttcttttcttttgtacaaggaaatcaacgagcccacggacaagacc<br>ttg (SEQ ID NO: 76) |
| 49 | gaactcttggcagctgcctgcagcgatttaccacaatgccattcttattctgcaatgtcaatgatgtat<br>gtaattttgcatctcgaaatgattattcactactggctgtcaacaccagctctgatgccaatgaacatgg<br>ctcccattactggcagagcccttgagccttatataagcag (SEQ ID NO: 77)                                    |
| 50 | atgcactgtttgtgaaggtcctgcagtcgccatagccgttcacagccaaaccactgacattcctccatg<br>tcctcacggctggatttctctctggaaggattttcattcatcatg (SEQ ID NO: 78)                                                                                                         |
| 51 | ttcacaagtgcaggttctgagggcaccgggcaagcactggcctcccctggctcctgcctggaagaattc<br>cgagccagcccatttctagaatgtcatggaagaggaaacgtgcaactactattcaaatcctacagtttc<br>tggctggcttcattaaaccagaaaagaatgttcag (SEQ ID NO: 79)                                          |
| 52 | aaagcctattccatcaactgtgaaagctggggaattagaaaaataataagtcgctgtcaggtgtgcat<br>gaagaaaagacactga (SEQ ID NO: 80)                                                                                                                                       |

#### COL4A4 coding sequences

In one embodiment, a CDS encoding a COL4A4 polypeptide is formed upon delivery of the first AAV vector and the second AAV vector into a host cell.

- 5 The COL4A4 coding sequence may be any suitable COL4A4 coding sequence known in the art. Example human COL4A4 coding sequences include COL4A4 coding sequences from NCBI Gene ID: 1286, for example coding sequences from mRNA having NCBI accession numbers NM\_000092, XM\_011510557, XM\_011510558, XM\_017003297, XM\_006712246, XM\_047443241, XM\_011510560, XM\_047443242, XM\_011510559, XM\_011510561,
- 10 XM\_011510562, XM\_047443245, XM\_005246282, XM\_011510565, XM\_011510567, XM\_047443249, XM\_011510566, XM\_047443248, XM\_011510568, XM\_011510570, XM\_047443250, or XM\_011510569.

- The COL4A4 coding sequence may be a variant of any COL4A4 coding sequence known in the art. A person skilled in the art would be able to generate variants of known COL4A4 coding sequences using the degeneracy of the genetic code, based on codon-optimisation, and/or
- 15 based on known variants.

The COL4A4 coding sequence may comprise a signal sequence. For example, nucleotide 1-114 of SEQ ID NO: 81 may encode a COL4A4 signal sequence. In some embodiments, the COL4A4 signal sequence is absent or substituted by an alternative signal sequence. In some embodiments, the COL4A4 coding sequence lacks a signal sequence. In some embodiments, the COL4A4 coding sequence has an alternative signal sequence.

In some embodiments, a COL4A4 coding sequence comprises or consists of a nucleotide sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 81 or a fragment thereof lacking nucleotides 1-114.

In some embodiments, a COL4A4 coding sequence comprises or consists of the nucleotide sequence of SEQ ID NO: 81 or a fragment thereof lacking nucleotides 1-114. In some embodiments, a COL4A4 coding sequence consists of the nucleotide sequence of SEQ ID NO: 81 or a fragment thereof lacking nucleotides 1-114.

In some embodiments, a COL4A4 coding sequence comprises or consists of the nucleotide sequence of SEQ ID NO: 81. In some embodiments, a COL4A4 coding sequence consists of the nucleotide sequence of SEQ ID NO: 81.

SEQ ID NO: 81 may correspond to the following coding exonic regions (also referred to herein as "exons"):

| Exon | Sequence                                                                                                                                              |
|------|-------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1    | atgtggtctctgcacatagtactaatgaggtgctccttcagattgaccaagtccttggccacaggtccctg (SEQ ID NO: 82)                                                               |
| 2    | gtcacttatactcattctcttttctgtacaatatgtatatggg (SEQ ID NO: 83)                                                                                           |
| 3    | agtggaaaagaatacattggctcttgtggaggaagagattgctctgtttgccactgtgttcctgaaaaggggtctcgg (SEQ ID NO: 84)                                                        |
| 4    | ggtcaccaggaccaccaggggccacagggtccaattggacccttggagccccaggaccattgggctttcaggagagaaaaggaatgagaggggaccgcgccctcctggagcagcaggggacaaaggagataag (SEQ ID NO: 85) |
| 5    | ggtcacttggtgttccttggaattccagggttagatggcatacct (SEQ ID NO: 86)                                                                                         |
| 6    | gggcacccagggcctcctggacccagaggcaaacctggatagtggtggccacaatggctcaagaggtgaccagggtttccaggaggaagaggagctcttggcccaggaggcccccta (SEQ ID NO: 87)                 |
| 7    | ggccatcctggggaaaaggagaaaaaggaaattcagtgttcattttaggtgccgttaaagggtattcag (SEQ ID NO: 88)                                                                 |
| 8    | ggagacagaggggacccaggactgcctggccttacca (SEQ ID NO: 89)                                                                                                 |
| 9    | ggatcttggggtgcagaggaccggcagggtccacaggatatcctggagagccagggttagtg (SEQ ID NO: 90)                                                                        |
| 10   | ggacctccgggccaaccaggggcgtccagggttgaag (SEQ ID NO: 91)                                                                                                 |
| 11   | ggaaatcccgggtgtggagtaaaggggcaaatgggagaccg (SEQ ID NO: 92)                                                                                             |
| 12   | ggtgaggttggtcagcaaggttctcctggacccaccctgttggtagagccactgacttttgtctctataaaggagaaaag (SEQ ID NO: 93)                                                      |
| 13   | ggtataaaaggaattcctggaatggttgactgccaggaccaccaggacgcaag (SEQ ID NO: 94)                                                                                 |

|    |                                                                                                                                                                                                                                                          |
|----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 14 | ggagaatctggtattggggcaaaaaggagaaaaaggtattcctggatttccagggcctcgg (SEQ ID NO: 95)                                                                                                                                                                            |
| 15 | ggggatcctgggttcctatggatctccagggttttccaggattaaag (SEQ ID NO: 96)                                                                                                                                                                                          |
| 16 | ggagaactaggactggttgagatcctgggctatttggattaattggcccaaag (SEQ ID NO: 97)                                                                                                                                                                                    |
| 17 | ggggatcctggaaatcgagggcaccaggaccaccagggtgttttggtgactccacctcttccactcaaa<br>g (SEQ ID NO: 98)                                                                                                                                                               |
| 18 | gcccaccaggggaccagggttccctggccgctatggagaaacaggggatgttgaccacctgggtcccc<br>cagggtctcttgggcagaccagggggaagcctgtgcag (SEQ ID NO: 99)                                                                                                                           |
| 19 | gcatgataggacccccctgggcccacaaggttttctggtggtcttccctgggcttccaggagaagctggtattc<br>ctgggagacctgattctgctccaggaaaaccagggaagccaggatcacctggcttgctggagcaccag<br>gcctgcagggcctcccaggatcaagtg (SEQ ID NO: 100)                                                       |
| 20 | tgatatactgtagtgttgggaaccccgaccacaaggaataaaaaggcaaagtgtgtccccaggaggaa<br>gaggcccaaaaggagaaaaag (SEQ ID NO: 101)                                                                                                                                           |
| 21 | gaaatgaaggactctgtgcctgtgagcctggaccatgggccccctggccctccaggacttctctggga<br>ggcaggggagtaaggagacttgggctccctggctggcttggaacaaaaggtagaccaggacctcctg<br>gtgctgaaggacctccagggtacca (SEQ ID NO: 102)                                                                |
| 22 | ggaaagcatggtgcctctggaccacctggcaacaaaggggcgaagggtgacatggttgatcaagagtt<br>aaag (SEQ ID NO: 103)                                                                                                                                                            |
| 23 | ggcacaaggagaaaagaggtcctgatgggccccaggatttccagggcagccaggatcacatggtcggg<br>atggacatgctggagaaaaaggggatccaggacctcca (SEQ ID NO: 104)                                                                                                                          |
| 24 | ggggatcatgaagatgcgacccaggtggttaaaggatttccctggacctctgggccccaggcaagca<br>ggacctgtggggccccaggactgggtttcctggtccaccaggagagcgaggccaccaggagttcca<br>ggccaccagggtgtgaggggccccctgatggcttgaagggtcagaaag (SEQ ID NO: 105)                                           |
| 25 | gtgacacaatttcttgcaacgtaacctaccctgggaggcatggccctccagggttttgatggacctccag<br>(SEQ ID NO: 106)                                                                                                                                                               |
| 26 | gtccgaagggtatttccagggtccccaaaggtgccccctgggctgagtgttcagatgggcataaaaggcagac<br>ctggcacaccaggaacagcggaataaccagggtccacctg (SEQ ID NO: 107)                                                                                                                   |
| 27 | gttttcgtggtgacatgggagatccgggttttggaggtgaaaagggtcctccccctgttgggccccag<br>gccctccccggtccaccaggagtgaatggtcagaaaaggaatccccgggagacctgcatttggctacctgg<br>gacccccgggaaagaggggtctttcaggagtgcaggggataaaaaggaccagaggtgatccgggatgtc<br>caggggtgaag (SEQ ID NO: 108) |
| 28 | ggccagctggcattcctggattcctaggtctcaaagggtcccaaaaggcagagagggacatgctgggtttc<br>cagggtgtcccagggtccacctggccattcctgtgaaagaggtgctccagggataccaggggcaaccgggac<br>tcctgggtatccaggtagccag (SEQ ID NO: 109)                                                           |
| 29 | gtgctccagggtgggaaaggacagccgggagatgtggggcctccccgggcccagctggaatgaaaggcctcc<br>ccggactcccaggacggcctggggcacatggtcccccaggcctcccagggaatcccagggtccctttggag<br>atgatgggctacctggtcctccagggtccaaagg (SEQ ID NO: 110)                                               |
| 30 | gacccccgggggtgcctggttttccagggttttccccggagaaaggagaaagcctggtgcagagggatgtc<br>ctggcgcaaaagggagaacctggagagaagggcatgtctggccttccctggagaccggggactgagagggg<br>ccaaag (SEQ ID NO: 111)                                                                            |
| 31 | gagccataggacctccccggagatgaaggagaaatggctatcatttcacaaaagggaacacctggggaac<br>ctggacctcctggagatgatggattcccaggagaaagag (SEQ ID NO: 112)                                                                                                                       |
| 32 | gtgataaaggaaactccccgggatgcaagggagaagaggagagccgggaagatacggaccacctggatttc<br>acagaggggaacctggtgagaaaggtcagccagggcctcctggacccccaggccctccagggtcaactg<br>gtctaagaggggttcattggttttccaggacttccagggtgaccag (SEQ ID NO: 113)                                      |
| 33 | ggtgagccagggttctccagggtccccctggattttcaggaattgatggagcaagaggacctaaag (SEQ<br>ID NO: 114)                                                                                                                                                                   |
| 34 | gaacaaaagggtgacctgcccagtcacttttggtccacctggtccaaagggtgagccaggtagccctggat<br>gtccag (SEQ ID NO: 115)                                                                                                                                                       |
| 35 | ggcatttttgagcatccggagagcagggttgcctggtattcaagggccagaggatcacctggaaggc<br>cagggccacctggctcctctggaccaccagggtgccag (SEQ ID NO: 116)                                                                                                                           |
| 36 | gtgatcacgggatgcctgggctgaggggacagccaggagaaaatgggagacctgggccaagaggcctcc<br>aggggatccagggataccagggtcctccgggaataaaaag (SEQ ID NO: 117)                                                                                                                       |
| 37 | gtccctccggatcacctggcctgaacggccttgcatggattgaaagggtcagaaaaggaactaaagggtgctt<br>cag (SEQ ID NO: 118)                                                                                                                                                        |
| 38 | gtttgcatgatgtggggccacctggtccagtgggaataacctgggctaaaaggggagagaggagacctg<br>ggagcccaggaatctctcctccagggtcctcgtggaagaaagggtcccccaggacccccag (SEQ ID<br>NO: 119)                                                                                               |
| 39 | ggagttcaggaccacctggtcctgcagggtgccacaggaagagctcctaaggacattcctgaccgggtc<br>cacctggagatcagggaacctcctggtcctgatggcccaagag (SEQ ID NO: 120)                                                                                                                    |

|    |                                                                                                                                                                                                                                                                                                                                  |
|----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 40 | gagcacctgggacctccaggcctccctgggagtgttgaccttctgagaggggagccaggtgactgtggtc<br>taccagggccaccaggtccccctggcccaccaggccctccaggatacaaaggctttccaggatgtgatg<br>gaaaagatggccagaaaag (SEQ ID NO: 121)                                                                                                                                          |
| 41 | gaccagtgggattccccgggaccgcaggaccacatggatttccctgggcccactggagagaagggtttac<br>ctggacctccaggagaaaaaggggccactggtcttccgg (SEQ ID NO: 122)                                                                                                                                                                                               |
| 42 | gtcccagag                                                                                                                                                                                                                                                                                                                        |
| 43 | gtgaaccggggccacctgcagatgtggatgactgtccccgaatcccaggccttccctggggcgccaggca<br>tgagaggaccagaaggagccatggggctccctggaatgagaggccccctcaggaccag (SEQ ID NO:<br>124)                                                                                                                                                                         |
| 44 | ggtgcaaaggagagcctgggctggatggcaggaggggtgtggatggcgctccctgggtctcctgggcctc<br>ccggacgtaaagggtgacacaggagaagacggctaccctggaggaccag (SEQ ID NO: 125)                                                                                                                                                                                     |
| 45 | ggcctcctggtcccattggggatcctgggcccagggtttggccctggatacctcggtggcttccctcc<br>tggttctccacagtcagacggaccaggagcccactgccccctgggcatgccagggtcttgactgggt<br>atagtctgttatacctggaagggaagagaaaagctcacaatcaagaccttg (SEQ ID NO: 126)                                                                                                              |
| 46 | gtctggcagggtcttgccttcccgtatttagcacgctgccctttgctactgcaacatccaccaggtgt<br>gccactatgccagagaaacgacagatcctactggctggccagcgctgcgccccctccccatgatgccac<br>tctctgaagaggcgatccgccccctatgtcagccgctgtgcggtatgagaggccccggccaggcggtgg<br>cgggtgcacagccaggaccagtcctatcccccatgtccgcagacctggaggagcctctggatcggttatt<br>cattcctgatg (SEQ ID NO: 127) |
| 47 | cacacaggagctggggaccaaggaggagggcaggcccttatgtcacctggcagctgcctggaagatttc<br>agagcagaccattccttgaatgccaggcgccagggaacttgccactttttcgaaaataagtatagc<br>ttctggctcacaacggtgaaagcagacttgagtttctctgtctccagcaccagacaccttaaaagaa<br>agccaggcccaacgccagaaaatcagccggtgccaggtctgcgtgaagtatagctag (SEQ ID NO:<br>128)                              |

#### COL4A5 coding sequences

In one embodiment, a CDS encoding a COL4A5 polypeptide is formed upon delivery of the first AAV vector and the second AAV vector into a host cell.

- 5 The COL4A5 coding sequence may be any suitable COL4A5 coding sequence known in the art. Example human COL4A5 coding sequences include COL4A5 coding sequences from NCBI Gene ID: 1287, for example coding sequences from mRNA having NCBI accession numbers NM\_000495, NM\_033380, XM\_011530849, XM\_017029259, XM\_017029260, or XM\_047441810.
- 10 The COL4A5 coding sequence may be a variant of any COL4A5 coding sequence known in the art. A person skilled in the art would be able to generate variants of known COL4A5 coding sequences using the degeneracy of the genetic code, based on codon-optimisation, and/or based on known variants.

- 15 The COL4A5 coding sequence may comprise a signal sequence. For example, nucleotide 1-78 of SEQ ID NO: 129 may encode a COL4A5 signal sequence. In some embodiments, the COL4A5 signal sequence is absent or substituted by an alternative signal sequence. In some embodiments, the COL4A5 coding sequence lacks a signal sequence. In some embodiments, the COL4A5 coding sequence has an alternative signal sequence.



In some embodiments, a COL4A5 coding sequence comprises or consists of a nucleotide sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 129 or a fragment thereof lacking nucleotides 1-78.

In some embodiments, a COL4A5 coding sequence comprises or consists of the nucleotide sequence of SEQ ID NO: 129 or a fragment thereof lacking nucleotides 1-78. In some embodiments, a COL4A5 coding sequence consists of the nucleotide sequence of SEQ ID NO: 129 or a fragment thereof lacking nucleotides 1-78.

- 10 In some embodiments, a COL4A5 coding sequence comprises or consists of the nucleotide sequence of SEQ ID NO: 129. In some embodiments, a COL4A5 coding sequence consists of the nucleotide sequence of SEQ ID NO: 129.

SEQ ID NO: 129 may correspond to the following coding exonic regions (also referred to herein as "exons"):

| Exon | Sequence                                                                                                           |
|------|--------------------------------------------------------------------------------------------------------------------|
| 1    | atgaaactgcgtggagtcagcctggctgccggcttggtcttactggccctgagtccttgggggcagcctgcagaggctgcg (SEQ ID NO: 130)                 |
| 2    | gcttgctatgggtgttctccaggatcaaagtgtgactgcagtgccataaaaggggaaaag (SEQ ID NO: 131)                                      |
| 3    | ggagagagaggggtttccagggtttggaaggacaccaggattgcctggatttccagggtccagaagggcctccggggcctcggggacaaaag (SEQ ID NO: 132)      |
| 4    | ggtgatgatggaattccaggggccaccaggacccaaaaggaatcaga (SEQ ID NO: 133)                                                   |
| 5    | ggctctctggacttctctggatttccaggggacaccagggtcttctct (SEQ ID NO: 134)                                                  |
| 6    | ggaatgccaggccacgatggggccccaggacctcaaggtattcccgatgcaatggaaccaag (SEQ ID NO: 135)                                    |
| 7    | ggagaacgtggatttccaggcagtcctcggttttcttggtttacagggtcctcca (SEQ ID NO: 136)                                           |
| 8    | ggacccccctgggatccagggtatgaag (SEQ ID NO: 137)                                                                      |
| 9    | ggtgaaccaggtagtataattatgtcatcactgccaggacccaaagggaatccaggatatccaggctcctcctggaatacaa (SEQ ID NO: 138)                |
| 10   | ggcctacctggtccactggtataccaggggccaattggtccccaggaccaccagggtttgatg (SEQ ID NO: 139)                                   |
| 11   | ggcctctctggtccaccaggacttccaggacctaag (SEQ ID NO: 140)                                                              |
| 12   | gggaatatgggcttaaatttccaggggacccaaagggtgaaaaa (SEQ ID NO: 141)                                                      |
| 13   | ggtgagcaaggctcttcaggggccacctggggccacctgggcagatcagtgaaacagaaaagaccaattgatgtagagtttcagaaaaggagatcag (SEQ ID NO: 142) |
| 14   | ggacttctggtgaccgagggcctcctggacctccagggatacgtggtcctcca (SEQ ID NO: 143)                                             |
| 15   | ggtccccagggtggtgagaaaagggtgagaagggtgagcaaggagagccagggcaaaaga (SEQ ID NO: 144)                                      |
| 16   | ggtaaaccaggcaaaagatggagaaaaatggccaaccagggaattcct (SEQ ID NO: 145)                                                  |
| 17   | ggtttgcttggtgatcctggttacctggtgaaccgggaagggtggtgaaaag (SEQ ID NO: 146)                                              |
| 18   | ggccaaaaagggtgacactggcccacctggacctcctggactt (SEQ ID NO: 147)                                                       |

|    |                                                                                                                                                                                                                           |
|----|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 19 | gtaattcctagacctgggactggtataactatagagaaaaagaaacattgggttgctgggttgctggagaaaaaggagagcgaggatttctggaatacagggtccacctggccttctggacctccag (SEQ ID NO: 148)                                                                          |
| 20 | gggctgcagttatgggtcctcctggcctcctggatttctggagaaaggggtcagaaaggatgaag gaccacctggaatttccattcctggacctcctggacttgacggacagcctggggctcctgggttccag ggctcctggcctgctggcctcacattcctcta (SEQ ID NO: 149)                                  |
| 21 | gtgatgagatatgtgaaccaggccctccaggccccaggatctccagggtgataaaggactccaaggag aacaaggagtgaag (SEQ ID NO: 150)                                                                                                                      |
| 22 | gtgacaaaaggtagaccttgcttcaactgcattggaactgggtatttcagggcctccagggtcaacctgggt tgccagggtctccagggtcctccag (SEQ ID NO: 151)                                                                                                       |
| 23 | gatctcttggtttccctggacagaaaggggaaaaaggacaagctggtgcaactggtcccaaaggattac ca (SEQ ID NO: 152)                                                                                                                                 |
| 24 | ggcattccaggagctccagggtgctccagggttctcctggatctaaagggtgaacctgggtgatatacctcact tttccaggaatgaagggtgacaaaggagagttgggttcccctggagctccagggttccctgggttacct ggactcctggacaggatggattgccagggttctcctggccccgaaaggagagcct (SEQ ID NO: 153) |
| 25 | ggtggaattacttttaagggtgaaagggtccccctgggaaccagggtttaccaggcctccagggaat ataggggcctatgggtccccctgggttccggcctccaggcccagtaggtgaaaaaggcatacaagggtgtg gcaggaaatccaggccagccaggaataccag (SEQ ID NO: 154)                              |
| 26 | gtcctaaagggtgatccagggtcagactataaccagccggggaagcctggcttgctggtaaccaggga gagatggtgatgtaggtcttccag (SEQ ID NO: 155)                                                                                                            |
| 27 | gtgaccttgacttccagggaaccagggttgccagggtacactggttagcaaaaggagaaccagggtatcc ctggaattgggttctcctggacctggtcccaaag (SEQ ID NO: 156)                                                                                                |
| 28 | gctttcctggaattccaggacctccaggagcacctgggacacctggaagaattggtctagaaggccctc ctggggccaccggccttccaggaccaaag (SEQ ID NO: 157)                                                                                                      |
| 29 | ggtgaaccaggatttgcattacctgggccacctgggccaccaggacttccagggttcaaaggagcactt ggtccaaaagggtgatcgtggttccaggacctccgggtcctccaggagcactgggttagatgggctc cctggaccaaag (SEQ ID NO: 158)                                                   |
| 30 | gtgatgttgaccaaattggacaacctggaccaatgggacctcctgggtgccaggaaataggtgttcagg gaccaccaggaccaccagggtattcctgggccaataggtaacctg (SEQ ID NO: 159)                                                                                      |
| 31 | gtttacatggaataccaggagagaaggggatccaggacctcctggacttgatgttccaggaccctccag gtgaaaggagcagtcaggtccagggtatccccggagcacctggtcctataggacctccaggatcaccagggttc caggaaaagcaggtgcctctggtattccag (SEQ ID NO: 160)                          |
| 32 | gtaccaaagggtgaaatgggtatgatgggacctccaggcccaccaggacctttgggaattcctggcagga gtggtgtacctggtcttaag (SEQ ID NO: 161)                                                                                                              |
| 33 | gtgatgatggcttgagggtcagccaggacttccctggccctacaggagaaaaaggtagtaaaaggagagc ctggccttccaggccctcctggaccaatggatccaaatcttctgggtcaaaaaggagagaagggggaac ctggcttaccag (SEQ ID NO: 162)                                                |
| 34 | gtatacctggagtttccaggccaaaagggttatcagggttgcctggagaccagggaacctggactga gtggacaacctggattaccaggaccaccag (SEQ ID NO: 163)                                                                                                       |
| 35 | gtcccaaaggtaacctggtctcctggacagccagggtcttataggacctcctggacttaaaaggaaacca tcggtgatatgggttttccag (SEQ ID NO: 164)                                                                                                             |
| 36 | ggcctcagggtgtggaaggcctcctggaccttctggagttcctggacaacctgggtccccaggattac ctggacagaaaggcgacaaagggtgatcctggtatttcaagcattggtcttccagggtcttctggtccaa ag (SEQ ID NO: 165)                                                           |
| 37 | ggtgagcctggtctgctggataccagggaacctggtatcaaagggttctgtgggagatcctgggttg cccggattaccagggaacctcctggagcaaaaggacaaccaggccttctggtattccag (SEQ ID NO: 166)                                                                          |
| 38 | gaaccccaggccctcctggaccaaagggtattagtgccctcctgggaaccccgcccttccaggagaac ctggtcctgtag (SEQ ID NO: 167)                                                                                                                        |
| 39 | gtggtggaggtcatcctgggcaaccagggtcctccaggcgaaaaaggcaaacccgggtcaagatggtattc ctggaccagctggacagaagggtgaaccag (SEQ ID NO: 168)                                                                                                   |
| 40 | gtcaaccagggtttggaaaccaggacccccctggacttccaggactttctg (SEQ ID NO: 169)                                                                                                                                                      |
| 41 | gcaaaaagggtgatggaggattacctgggattccaggaaatcctggccttccagggtccaaaggcgaaac caggctttacgggttctcctggtgtgcagggtccccaggccctcctggttctccgggtccagctctgg aaggacctaaaggcaacctgggccccaaagggtcctcctgggagaccag (SEQ ID NO: 170)            |
| 42 | gtctaccagggtccagaagggtcctccagggtctcctggaaatggagggtattaaaggagagaagggaaatc caggccaacctgggtacctgggttgctggtttgaaaggagatcaaggaccaccaggactccag (SEQ ID NO: 171)                                                                 |

|    |                                                                                                                                                                                                                                                       |
|----|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 43 | ggtaatcctggccggccgggtctcaatggaatgaaaggagatcctgggtctccctgggtgttccaggattccag (SEQ ID NO: 172)                                                                                                                                                           |
| 44 | gcatgaaaggacccagtgagtagctggatcagctggccctgaggggaaccgggacttattggtcctccag (SEQ ID NO: 173)                                                                                                                                                               |
| 45 | gtcctcctggattacctggtccttcaggacagagtatcataattaaaggagatgctggctcctccaggaa<br>tccctggccagcctgggctaaagggtctaccaggaccccaaggacctcaaggcttaccag (SEQ ID<br>NO: 174)                                                                                            |
| 46 | gtccaactggccctccaggagatcctggacgcaatggactccctggccttgatgggtgcaggagggcgca<br>aaggagaccaggtctgccaggacagccag (SEQ ID NO: 175)                                                                                                                              |
| 47 | gtacccgtggtttggatgggtcccccctgggtccagatggattgcaagggtcccccagggtcccccctggaacct<br>cctctgttgacatggatttcttattacagccacagccagacaacggatgcaccacaatgccacag<br>gaacacttcagggtctatgaaggcttttctctctgtatgtacaaggaaataaaagagcccacgggtcaag<br>acttgg (SEQ ID NO: 176) |
| 48 | ggacggctggcagctgccttcgtcgtctttagtaccatgcctttcatgttctgcaacatcaataatgttt<br>gcaactttgcttcaagaaatgactattcttactggctctctaccccagagcccatgccaatgagcatgc<br>aaccctaaaggccagagcatccagccattcattagtcg (SEQ ID NO: 177)                                            |
| 49 | atgtgcagtatgtgaagctccagctgtgggtgatcgcagttcacagtcagacgatccagatteccattg<br>tccctcagggatgggattctctgtggattgggttattccttcatgatg (SEQ ID NO: 178)                                                                                                            |
| 50 | catacaagtgcaggggcagaaggctcaggtcaagccctagcctccctgggtcctgcttggagagttt<br>cgttcagctcccttcacgaatgtcatgggaggggtacctgtaactactatgccaaactcctacagcttt<br>tggctggcaactgtagatgtgtcagacatgttcag (SEQ ID NO: 179)                                                  |
| 51 | taaacctcagtcagaaacgctgaaagcaggagacttgaggacacgaattagccgatgtcaagtgtgcat<br>gaagaggacataa (SEQ ID NO: 180)                                                                                                                                               |

### 5' AAV vector

The present invention provides an AAV vector comprising a 5' coding sequence (CDS) encoding an N-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide.

- 5 As used herein, a "5' AAV vector" may refer to an AAV vector forming part of a dual AAV vector system which comprises a 5' coding sequence (CDS). The 5' CDS may encode the N-terminal part of a full-length protein.

Suitably, the 5' CDS encodes an N-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide having a length of at least 250 amino acids, at least 300 amino acids, at least 350 amino acids, at least 400 amino acids, at least 450 amino acids, at least 500 amino acids, at least 550 amino acids, at least 600 amino acids, at least 650 amino acids, at least 700 amino acids, at least 750 amino acids, or at least 800 amino acids.

### *COL4A3 5' coding sequences*

- 15 In one embodiment, the 5' AAV vector comprises a 5' CDS encoding an N-terminal part of a COL4A3 polypeptide.

The N-terminal part of a COL4A3 polypeptide may be from any COL4A3 polypeptide known in the art or described herein.

Suitably, the 5' CDS comprises exons 1-10 or more, exons 1-11 or more, exons 1-12 or more, exons 1-13 or more, exons 1-14 or more, exons 1-15 or more, exons 1-16 or more, exons 1-17 or more, exons 1-18 or more, exons 1-19 or more, exons 1-20 or more, exons 1-21 or more, exons 1-22 or more, exons 1-23 or more, exons 1-24 or more, or exons 1-25 or more of a CDS encoding a COL4A3 polypeptide.

Suitably, the 5' CDS comprises exons 1-45 or less, exons 1-44 or less, exons 1-43 or less, exons 1-42 or less, exons 1-41 or less, exons 1-40 or less, exons 1-39 or less, exons 1-38 or less, exons 1-37 or less, exons 1-36 or less, or exons 1-35 or less of a CDS encoding a COL4A3 polypeptide.

Suitably, the 5' CDS comprises at least exons 1-15 to at most exons 1-45, at least exons 1-16 to at most exons 1-44, at least exons 1-17 to at most exons 1-43, at least exons 1-18 to at most exons 1-41, at least exons 1-19 to at most exons 1-41, at least exons 1-20 to at most exons 1-40, at least exons 1-21 to at most exons 1-39, at least exons 1-22 to at most exons 1-38, at least exons 1-23 to at most exons 1-37, at least exons 1-24 to at most exons 1-36, or at least exons 1-25 to at most exons 1-35 of a CDS encoding a COL4A3 polypeptide.

Suitably, the 5' CDS comprises or consists of exons 1-15, exons 1-16, exons 1-17, exons 1-18, exons 1-19, exons 1-20, exons 1-21, exons 1-22, exons 1-23, exons 1-24, exons 1-25, exons 1-26, exons 1-27, exons 1-28, exons 1-29, exons 1-30, exons 1-31, exons 1-32, exons 1-33, exons 1-34, exons 1-35, exons 1-36, exons 1-37, exons 1-38, exons 1-39, exons 1-40, or exons 1-41, exons 1-42, or exons 1-43 of a CDS encoding a COL4A3 polypeptide.

In some embodiments, the N-terminal part of a COL4A3 polypeptide is an N-terminal fragment from an amino acid sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 1 to 7 or fragments thereof lacking amino acids 1-28.

In some embodiments, the 5' CDS is a 5' fragment of a nucleotide sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 27 or a fragment thereof lacking nucleotides 1-84.

Suitably, the 5' CDS comprises or consists of at least nucleotides 1-855 and at most nucleotides 1-3893 of SEQ ID NO: 27. In some embodiments, the 5' CDS comprises or consists of at least nucleotides 1-1505 and at most nucleotides 1-3505, at least nucleotides

1-1605 and at most nucleotides 1-3405, at least nucleotides 1-1705 and at most nucleotides 1-3305, at least nucleotides 1-1805 and at most nucleotides 1-3205, at least nucleotides 1-1905 and at most nucleotides 1-3105, at least nucleotides 1-2005 and at most nucleotides 1-3005, at least nucleotides 1-2105 and at most nucleotides 1-2905, or at least nucleotides 1-2205 and at most nucleotides 1-2805 of SEQ ID NO: 27.

Examples of a 5' CDS encoding an N-terminal part of a COL4A3 polypeptide are shown in SEQ ID NOs: 181 to 187, corresponding to exons 1-28 (SEQ ID NO: 181), exons 1-29 (SEQ ID NO: 182), exons 1-30 (SEQ ID NO: 183), exons 1-31 (SEQ ID NO: 184), exons 1-32 (SEQ ID NO: 185), exons 1-33 (SEQ ID NO: 186), or exons 1-34 of SEQ ID NO: 27 (SEQ ID NO: 187).

In some embodiments, the 5' CDS comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 181 to 187.

In some embodiments, the 5' CDS comprises or consists of a nucleotide sequence selected from any of SEQ ID NOs: 181 to 187.

#### *COL4A4 5' coding sequences*

In one embodiment, the 5' AAV vector comprises a 5' CDS encoding an N-terminal part of a COL4A4 polypeptide.

The N-terminal part of a COL4A4 polypeptide may be from any COL4A4 polypeptide known in the art or described herein.

Suitably, the 5' CDS comprises exons 1-10 or more, exons 1-11 or more, exons 1-12 or more, exons 1-13 or more, exons 1-14 or more, exons 1-15 or more, exons 1-16 or more, exons 1-17 or more, exons 1-18 or more, exons 1-19 or more, exons 1-20 or more, exons 1-21 or more, exons 1-22 or more, exons 1-23 or more, exons 1-24 or more, or exons 1-25 or more of a CDS encoding a COL4A4 polypeptide.

Suitably, the 5' CDS comprises exons 1-45 or less, exons 1-44 or less, exons 1-43 or less, exons 1-42 or less, exons 1-41 or less, exons 1-40 or less, exons 1-39 or less, exons 1-38 or less, exons 1-37 or less, exons 1-36 or less, or exons 1-35 or less of a CDS encoding a COL4A4 polypeptide.

Suitably, the 5' CDS comprises at least exons 1-15 to at most exons 1-45, at least exons 1-16 to at most exons 1-44, at least exons 1-17 to at most exons 1-43, at least exons 1-18 to at most exons 1-41, at least exons 1-19 to at most exons 1-41, at least exons 1-20 to at most exons 1-40, at least exons 1-21 to at most exons 1-39, at least exons 1-22 to at most exons 1-38, at least exons 1-23 to at most exons 1-37, at least exons 1-24 to at most exons 1-36, or at least exons 1-25 to at most exons 1-35 of a CDS encoding a COL4A4 polypeptide.

Suitably, the 5' CDS comprises or consists of exons 1-14, exons 1-15, exons 1-16, exons 1-17, exons 1-18, exons 1-19, exons 1-20, exons 1-21, exons 1-22, exons 1-23, exons 1-24, exons 1-25, exons 1-26, exons 1-27, exons 1-28, exons 1-29, exons 1-30, exons 1-31, exons 1-32, exons 1-33, exons 1-34, exons 1-35, exons 1-36, exons 1-37, or exons 1-38, or exons 1-39 of a CDS encoding a COL4A4 polypeptide.

In some embodiments, the N-terminal part of a COL4A4 polypeptide is an N-terminal fragment from an amino acid sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 8 to 20 or fragments thereof lacking amino acids 1-38.

Suitably, the N-terminal part of a COL4A4 polypeptide comprises or consists of at least amino acids 1-304 and at most amino acids 1-1297 of SEQ ID NO: 8.. In some embodiments, the N-terminal part of a COL4A4 polypeptide comprises or consists of at least amino acids 1-445 and at most amino acids 1-1245, at least amino acids 1-495 and at most amino acids 1-1195, at least amino acids 1-545 and at most amino acids 1-1145, at least amino acids 1-595 and at most amino acids 1-1095, at least amino acids 1-645 and at most amino acids 1-1045, at least amino acids 1-695 and at most amino acids 1-995, or at least amino acids 1-745 and at most amino acids 1-945 of SEQ ID NO: 8.

In some embodiments, the 5' CDS is a 5' fragment of a nucleotide sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 81 or a fragment thereof lacking nucleotides 1-114.

Suitably, the 5' CDS comprises or consists of at least nucleotides 1-915 and at most nucleotides 1-3893 of SEQ ID NO: 81. In some embodiments, the 5' CDS comprises or consists of at least nucleotides 1-1535 and at most nucleotides 1-3535, at least nucleotides 1-1635 and at most nucleotides 1-3435, at least nucleotides 1-1735 and at most nucleotides 1-3335, at least nucleotides 1-1835 and at most nucleotides 1-3235, at least nucleotides 1-

1935 and at most nucleotides 1-3135, at least nucleotides 1-2035 and at most nucleotides 1-3035, at least nucleotides 1-2135 and at most nucleotides 1-2935, or at least nucleotides 1-2235 and at most nucleotides 1-2835 of SEQ ID NO: 81.

5 Examples of a 5' CDS encoding an N-terminal part of a COL4A4 polypeptide are shown in SEQ ID NOs: 188 to 194, corresponding to exons 1-25 (SEQ ID NO: 188), exons 1-26 (SEQ ID NO: 189), exons 1-27 (SEQ ID NO: 190), exons 1-28 (SEQ ID NO: 191), exons 1-29 (SEQ ID NO: 192), exons 1-30 (SEQ ID NO: 193), or exons 1-31 (SEQ ID NO: 194) of SEQ ID NO: 81.

10 In some embodiments, the 5' CDS comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 188 to 194.

In some embodiments, the 5' CDS comprises or consists of a nucleotide sequence selected from any of SEQ ID NOs: 188 to 194.

15 COL4A5 5' coding sequences

In one embodiment, the 5' AAV vector comprises a 5' CDS encoding an N-terminal part of a COL4A5 polypeptide.

The N-terminal part of a COL4A5 polypeptide may be from any COL4A5 polypeptide known in the art or described herein.

20 Suitably, the 5' CDS comprises exons 1-10 or more, exons 1-11 or more, exons 1-12 or more, exons 1-13 or more, exons 1-14 or more, exons 1-15 or more, exons 1-16 or more, exons 1-17 or more, exons 1-18 or more, exons 1-19 or more, exons 1-20 or more, exons 1-21 or more, exons 1-22 or more, exons 1-23 or more, exons 1-24 or more, or exons 1-25 or more of a CDS encoding a COL4A5 polypeptide.

25 Suitably, the 5' CDS comprises exons 1-45 or less, exons 1-44 or less, exons 1-43 or less, exons 1-42 or less, exons 1-41 or less, exons 1-40 or less, exons 1-39 or less, exons 1-38 or less, exons 1-37 or less, exons 1-36 or less, or exons 1-35 or less of a CDS encoding a COL4A5 polypeptide.

30 Suitably, the 5' CDS comprises at least exons 1-15 to at most exons 1-45, at least exons 1-16 to at most exons 1-44, at least exons 1-17 to at most exons 1-43, at least exons 1-18 to at most exons 1-41, at least exons 1-19 to at most exons 1-41, at least exons 1-20 to at most

exons 1-40, at least exons 1-21 to at most exons 1-39, at least exons 1-22 to at most exons 1-38, at least exons 1-23 to at most exons 1-37, at least exons 1-24 to at most exons 1-36, or at least exons 1-25 to at most exons 1-35 of a CDS encoding a COL4A5 polypeptide.

Suitably, the 5' CDS comprises or consists of exons 1-15, exons 1-16, exons 1-17, exons 1-18, exons 1-19, exons 1-20, exons 1-21, exons 1-22, exons 1-23, exons 1-24, exons 1-25, exons 1-26, exons 1-27, exons 1-28, exons 1-29, exons 1-30, exons 1-31, exons 1-32, exons 1-33, exons 1-34, exons 1-35, exons 1-36, exons 1-37, exons 1-38, exons 1-39, or exons 1-40, or exons 1-41 of a CDS encoding a COL4A5 polypeptide.

In some embodiments, the N-terminal part of a COL4A5 polypeptide is an N-terminal fragment from an amino acid sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 21 to 26 or fragments thereof lacking a signal sequence.

Suitably, the N-terminal part of a COL4A5 polypeptide comprises or consists of at least amino acids 1-299 and at most amino acids 1-1297 of SEQ ID NO: 21.

In some embodiments, the 5' CDS is a 5' fragment of a nucleotide sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 129 or a fragment thereof lacking nucleotides 1-78.

Suitably, the 5' CDS comprises or consists of at least nucleotides 1-900 and at most nucleotides 1-3893 of SEQ ID NO: 129.

Examples of a 5' CDS encoding an N-terminal part of a COL4A5 polypeptide are shown in SEQ ID NOs: 274 to 280, 195 to 201 and 281 to 288, corresponding to exons 1-20 (SEQ ID NO: 274), exons 1-21 (SEQ ID NO: 275), exons 1-22 (SEQ ID NO: 276), exons 1-23 (SEQ ID NO: 277), exons 1-24 (SEQ ID NO: 278), exons 1-25 (SEQ ID NO: 279), exons 1-26 (SEQ ID NO: 280), exons 1-27 (SEQ ID NO: 195), exons 1-28 (SEQ ID NO: 196), exons 1-29 (SEQ ID NO: 197), exons 1-30 (SEQ ID NO: 198), exons 1-31 (SEQ ID NO: 199), exons 1-32 (SEQ ID NO: 200), exons 1-33 (SEQ ID NO: 201), exons 1-34 (SEQ ID NO: 281), exons 1-35 (SEQ ID NO: 282), exons 1-36 (SEQ ID NO: 283), exons 1-37 (SEQ ID NO: 284), exons 1-38 (SEQ ID NO: 285), exons 1-39 (SEQ ID NO: 286), exons 1-40 (SEQ ID NO: 287), and exons 1-41 (SEQ ID NO: 288) of SEQ ID NO: 129.



In some embodiments, the 5' CDS comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 274 to 280, 195 to 201 or 281 to 288.

- 5 In some embodiments, the 5' CDS comprises or consists of a nucleotide sequence selected from any of SEQ ID NOs: 274 to 280, 195 to 201 or 281 to 288.

In some embodiments, the 5' CDS consists of a nucleotide sequence selected from any of SEQ ID NOs: 274 to 280, 195 to 201 or 281 to 288.

- 10 Other examples of a 5' CDS encoding an N-terminal part of a COL4A5 polypeptide are shown in SEQ ID NOs: 304-306.

In some embodiments, the 5' CDS comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 304 to 306.

- 15 In some embodiments, the 5' CDS comprises or consists of a nucleotide sequence selected from any of SEQ ID NOs: 304 to 306.

|                    |               |
|--------------------|---------------|
| Nucleotides 1-3083 | SEQ ID NO:304 |
| Nucleotides 1-2818 | SEQ ID NO:305 |
| Nucleotides 1-2474 | SEQ ID NO:306 |

#### *Example overlapping 5' CDS (long)*

- 20 In some embodiments, the 5' CDS has a length of at least about 3000 nucleotides and comprises an overlapping sequence with the 3' CDS.

In some embodiments, the 5' CDS has a length of at least about 3100 nucleotides, at least about 3200 nucleotides, at least about 3300 nucleotides, at least about 3400 nucleotides, at least about 3500 nucleotides, at least about 3600 nucleotides and comprises an overlapping sequence with the 3' CDS. In some embodiments, the 5' CDS has a length of about 3891

nucleotides or less, about 3800 nucleotides or less, or about 3700 nucleotides or less. In some embodiments, the 5' CDS has a length of from 3000 nucleotides to 3891 nucleotides, from 3100 nucleotides to 3891 nucleotides, from 3200 nucleotides to 3891 nucleotides, from 3300 nucleotides to 3891 nucleotides, from 3400 nucleotides to 3891 nucleotides, from 3500  
 5 nucleotides to 3891 nucleotides, or from 3600 nucleotides to 3891 nucleotides.

In some embodiments, the 5' CDS comprises or consists of exons 1-36, exons 1-37, exons 1-38, exons 1-39, exons 1-40, or exons 1-41 of a CDS encoding a COL4A5 polypeptide.

In some embodiments, the 5' CDS comprises or consists of exons 1-37, exons 1-38, exons 1-39, exons 1-40, or exons 1-41 of a CDS encoding a COL4A5 polypeptide.

10 In some embodiments, the 5' CDS comprises or consists of exons 1-39, exons 1-40, or exons 1-41 of a CDS encoding a COL4A5 polypeptide.

In some embodiments, the 5' CDS comprises or consists of exons 1-40 of a CDS encoding a COL4A5 polypeptide.

In some embodiments, the N-terminal part of a COL4A5 polypeptide comprises or consists of  
 15 at least amino acids 1-1000 and at most amino acids 1-1297, at least amino acids 1-1100 and at most amino acids 1-1297, at least amino acids 1-1100 and at most amino acids 1-1290, at least amino acids 1-1120 and at most amino acids 1-1280, at least amino acids 1-1130 and at most amino acids 1-1270, at least amino acids 1-1140 and at most amino acids 1-1260, at least amino acids 1-1150 and at most amino acids 1-1250, at least amino acids 1-1160 and  
 20 at most amino acids 1-1240, at least amino acids 1-1170 and at most amino acids 1-1230, at least amino acids 1-1180 and at most amino acids 1-1220, at least amino acids 1-1190 and at most amino acids 1-1210 of SEQ ID NO: 21. In some embodiments, the N-terminal part of a COL4A5 polypeptide consists of about amino acids 1-1201 of SEQ ID NO: 21

In some embodiments, the 5' CDS comprises or consists of at least nucleotides 1-3000 and  
 25 at most nucleotides 1-3893, at least nucleotides 1-3100 and at most nucleotides 1-3893, at least nucleotides 1-3200 and at most nucleotides 1-3893, at least nucleotides 1-3250 and at most nucleotides 1-3893, at least nucleotides 1-3300 and at most nucleotides 1-3893, at least nucleotides 1-3350 and at most nucleotides 1-3893, at least nucleotides 1-3400 and at most nucleotides 1-3800, at least nucleotides 1-3450 and at most nucleotides 1-3750, at least  
 30 nucleotides 1-3500 and at most nucleotides 1-3700, or at least nucleotides 1-3550 and at most nucleotides 1-3650 of SEQ ID NO: 129. In some embodiments, the 5' CDS consists of about nucleotides 1-3604 of SEQ ID NO: 129.

In some embodiments, the 5' CDS comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 283 to 288.

- 5 In some embodiments, the 5' CDS comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 287.

- 10 In some embodiments, the 5' CDS comprises or consists of the nucleotide sequence of any of SEQ ID NOs: 283 to 288. In some embodiments, the 5' CDS comprises or consists of the nucleotide sequence of any of SEQ ID NOs: 286 to 288.

In some embodiments, the 5' CDS comprises or consists of the nucleotide sequence of SEQ ID NO: 287.

- 15 In some embodiments, the 5' CDS comprises of a nucleotide sequence selected from any of SEQ ID NOs: 283 to 288.

In some embodiments, the 5' CDS comprises of a nucleotide sequence selected from any of SEQ ID NOs: 286 to 288.

In some embodiments, the 5' CDS comprises of the nucleotide sequence of SEQ ID NO: 287.

*Example overlapping 5' CDS (short)*

- 20 In some embodiments, the 5' CDS has a length of about 3600 nucleotides or less and comprises an overlapping sequence with the 3' CDS.

In some embodiments, the 5' CDS comprises or consists of exons 1-25, exons 1-26, exons 1-27, exons 1-28, exons 1-29, exons 1-30, exons 1-31, exons 1-32, exons 1-33, exons 1-34, exons 1-35, exons 1-36, exons 1-37, or exons 1-38 of a CDS encoding a COL4A5 polypeptide.

- 25 In some embodiments, the 5' CDS comprises or consists of exons 1-28, exons 1-29, exons 1-30, exons 1-31, exons 1-32, exons 1-33, exons 1-34, or exons 1-35 of a CDS encoding a COL4A5 polypeptide.

In some embodiments, the 5' CDS comprises or consists of exons 1-29, exons 1-30, exons 1-31, exons 1-32, exons 1-33, or exons 1-34 of a CDS encoding a COL4A5 polypeptide.

In some embodiments, the N-terminal part of a COL4A5 polypeptide consists of about amino acids 1-1027 of SEQ ID NO: 21.

In some embodiments, the N-terminal part of a COL4A5 polypeptide consists of about amino acids 1-939 of SEQ ID NO: 21.

- 5 In some embodiments, the N-terminal part of a COL4A5 polypeptide consists of about amino acids 1-824 of SEQ ID NO: 21.

In some embodiments, the 5' CDS consists of about nucleotides 1-3083 of SEQ ID NO: 129.

In some embodiments, the 5' CDS consists of about nucleotides 1-2818 of SEQ ID NO: 129.

In some embodiments, the 5' CDS consists of about nucleotides 1-2474 of SEQ ID NO: 129.

- 10 In some embodiments, the 5' CDS comprises or consists of the nucleotide sequence of any of SEQ ID NOs: 196-201, 281-282, or 304-306.

In some embodiments, the 5' CDS comprises or consists of a nucleotide sequence selected from any of SEQ ID NOs: 304-306.

*Example non-overlapping 5' CDS (mid-split)*

- 15 In some embodiments, the 5' CDS has a length of from about 1500 to about 3500 nucleotides and does not comprise an overlapping sequence with the 3' CDS.

In some embodiments, the 5' CDS comprises or consists of exons or exons 1-31 of a CDS encoding a COL4A5 polypeptide.

- 20 In some embodiments, the N-terminal part of a COL4A5 polypeptide consists of about amino acids 1-892 of SEQ ID NO: 21.

In some embodiments, the 5' CDS consists of about nucleotides 1-2677 of SEQ ID NO: 129

In some embodiments, the 5' CDS comprises or consists of a nucleotide sequence selected from any of SEQ ID NOs: 195 to 201.

- 25 In some embodiments, the 5' CDS comprises or consists of the nucleotide sequence of SEQ ID NO: 199.

*Example non-overlapping 5' CDS (short)*

In some embodiments, the 5' CDS has a length of about 3000 nucleotides or less and does not comprise an overlapping sequence with the 3' CDS.

In some embodiments, the 5' CDS comprises or consists of exons 1-24 of a CDS encoding a COL4A5 polypeptide.

- 5 In some embodiments, the N-terminal part of a COL4A5 polypeptide consists of about amino acids 1-593 of SEQ ID NO: 21.

In some embodiments, the 5' CDS consists of about nucleotides 1-1779 of SEQ ID NO: 129.

In some embodiments, the 5' CDS comprises or consists of the nucleotide sequence of SEQ ID NO: 278.

10 *Example non-overlapping 5' CDS (long)*

In some embodiments, the 5' CDS has a length of at least about 3000 nucleotides and does not comprise an overlapping sequence with the 3' CDS.

- 15 In some embodiments, the 5' CDS comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 283 to 288.

In some embodiments, the 5' CDS comprises or consists of a nucleotide sequence selected from any of SEQ ID NOs: 283 to 288.

*Other elements*

- 20 The 5' AAV vector may further comprise any suitable elements for expressing a COL4A3, COL4A4 or COL4A5 polypeptide and/or forming a CDS encoding a COL4A3, COL4A4 or COL4A5 polypeptide.

The 5' AAV vector may further comprise any suitable elements upstream of the 5' CDS.

- 25 The 5' AAV vector may comprise a promoter upstream of the 5' CDS. Typically, the 5' AAV vector comprises a promoter upstream of the 5' CDS, such that transcription of the full-length CDS may be initiated once it is formed. The promoter may be operably linked to the 5' CDS. The promoter may be any described herein. Suitably, the 5' AAV vector comprises or consists of from 5' to 3': a promoter and a 5' CDS encoding an N-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide.

In some embodiments, the 5' AAV vector comprises one or more regulatory elements upstream of the 5' CDS, which act to increase expression of a COL4A3, COL4A4 or COL4A5 polypeptide. The one or more regulatory elements may be operably linked to the 5' CDS. The one or more regulatory elements upstream of the 5' CDS may be any described herein.

- 5 Suitably, the 5' AAV vector comprises or consists of from 5' to 3': a promoter, one or more regulatory elements, and a 5' CDS encoding an N-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide

- 10 In some embodiments, the one or more regulatory elements upstream of the 5' CDS comprise a spliceosomal intron or a fragment thereof. The spliceosomal intron or a fragment thereof may be operably linked to the 5' CDS. The spliceosomal intron or a fragment thereof may be any described herein. Suitably, the 5' AAV vector comprises or consists of from 5' to 3': a promoter, a spliceosomal intron or a fragment thereof, and a 5' CDS encoding an N-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide.

- 15 In some embodiments, the one or more regulatory elements upstream of the 5' CDS comprise a Kozak sequence. The Kozak sequence may be operably linked to the 5' CDS. The Kozak sequence may be any described herein. Suitably, the 5' AAV vector comprises or consists of from 5' to 3': a promoter, a Kozak sequence, and a 5' CDS encoding an N-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide.

- 20 In some embodiments, the one or more regulatory elements upstream of the 5' CDS comprise a spliceosomal intron or a fragment thereof and a Kozak sequence. The spliceosomal intron or a fragment thereof and the Kozak sequence may each be operably linked to the 5' CDS. Suitably, the 5' AAV vector comprises or consists of from 5' to 3': a promoter, a spliceosomal intron or a fragment thereof, a Kozak sequence, and a 5' CDS encoding an N-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide.

- 25 The 5' AAV vector may further comprise any suitable elements downstream of the 5' CDS.

- 30 In some embodiments, the 5' AAV vector comprises a splice donor sequence downstream from the 5' CDS. As described in more detail below, such a splice donor sequence may be used in DNA trans-splicing or DNA hybrid strategies. The splice donor sequence may be any described herein. Suitably, the 5' AAV vector comprises or consists of from 5' to 3': a promoter, optionally one or more regulatory elements (e.g. a spliceosomal intron or a fragment thereof and/or a Kozak sequence), a 5' CDS encoding an N-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide, and a splice donor sequence.

In some embodiments, the 5' AAV vector further comprises a recombinogenic sequence downstream from the splice donor sequence. As described in more detail below, such a recombinogenic sequence may be used in DNA hybrid strategies. The recombinogenic sequence may be any described herein. Suitably, the 5' AAV vector comprises or consists of  
5 from 5' to 3': a promoter, optionally one or more regulatory elements (e.g. a spliceosomal intron or a fragment thereof and/or a Kozak sequence), a 5' CDS encoding an N-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide, a splice donor sequence, and a recombinogenic sequence.

In other embodiments, the 5' AAV vector comprises one or more regulatory elements  
10 downstream from the 5' CDS. As described in more detail below, one or more regulatory elements may be used in mRNA trans-splicing strategies. The one or more regulatory elements may be operably linked to the 5' CDS. The one or more regulatory elements downstream from the 5' CDS may be any described herein. Suitably, the 5' AAV vector comprises or consists of from 5' to 3': a promoter, optionally one or more regulatory elements  
15 (e.g. a spliceosomal intron or a fragment thereof and/or a Kozak sequence), a 5' CDS encoding an N-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide, and one or more regulatory elements.

In some embodiments, the one or more regulatory elements downstream from the 5' CDS  
20 comprise a polyadenylation sequence. The polyadenylation sequence may be operably linked to the 5' CDS. The polyadenylation sequence may be any described herein. Suitably, the 5' AAV vector comprises or consists of from 5' to 3': a promoter, optionally one or more regulatory elements (e.g. a spliceosomal intron or a fragment thereof and/or a Kozak sequence), a 5' CDS encoding an N-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide, and a polyadenylation sequence.

25 In some embodiments, the one or more regulatory elements downstream from the 5' CDS comprise a WPRE. The WPRE may be operably linked to the 5' CDS. The WPRE may be any described herein. Suitably, the 5' AAV vector comprises or consists of from 5' to 3': a promoter, optionally one or more regulatory elements (e.g. a spliceosomal intron or a fragment thereof and/or a Kozak sequence), a 5' CDS encoding an N-terminal part of a COL4A3, COL4A4 or  
30 COL4A5 polypeptide, and a WPRE.

In some embodiments, the one or more regulatory elements downstream from the 5' CDS comprise a polyadenylation sequence and a WPRE. The polyadenylation sequence and WPRE may each be operably linked to the 5' CDS. Suitably, the 5' AAV vector comprises or consists of from 5' to 3': a promoter, optionally one or more regulatory elements (e.g. a

spliceosomal intron or a fragment thereof and/or a Kozak sequence), a 5' CDS encoding an N-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide, a WPRE, and a polyadenylation sequence.

In some embodiments, the 5' AAV vector comprises a splice donor sequence downstream from the 5' CDS. As described in more detail below, a splice donor sequence may be used in mRNA trans-splicing strategies. The splice donor sequence may be any described herein (e.g. any forward splice donor sequence described herein). Suitably, the 5' AAV vector comprises or consists of from 5' to 3': a promoter, optionally one or more regulatory elements (e.g. a spliceosomal intron or a fragment thereof and/or a Kozak sequence), a 5' CDS encoding an N-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide, a splice donor sequence, and optionally one or more regulatory elements (e.g. a WPRE and/or a polyadenylation sequence).

In some embodiments, the 5' AAV vector further comprises a hybridisation domain downstream from the 5' CDS. As described in more detail below, a hybridisation domain may be used in mRNA trans-splicing strategies. The hybridisation domain may be any described herein. Suitably, the 5' AAV vector comprises or consists of from 5' to 3': a promoter, optionally one or more regulatory elements (e.g. a spliceosomal intron or a fragment thereof and/or a Kozak sequence), a 5' CDS encoding an N-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide, a splice donor sequence, a hybridisation domain, and optionally one or more regulatory elements (e.g. a WPRE and/or a polyadenylation sequence).

### 20 3' AAV vector

The present invention provides an AAV vector comprising a 3' coding sequence (CDS) encoding a C-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide.

As used herein, a "3' AAV vector" may refer to an AAV vector forming part of a dual AAV vector system which comprises a 3' coding sequence (CDS). The 3' CDS may encode the C-terminal part of a full-length protein.

Suitably, the 3' CDS encodes a C-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide having a length of at least 250 amino acids, at least 300 amino acids, at least 350 amino acids, at least 400 amino acids, at least 450 amino acids, at least 500 amino acids, at least 550 amino acids, at least 600 amino acids, at least 650 amino acids, at least 700 amino acids, at least 750 amino acids, or at least 800 amino acids.

Suitably, the 3' CDS encodes a C-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide having a length of 1300 amino acids or fewer, 1250 amino acids or fewer, 1200 amino acids or fewer, 1150 amino acids or fewer, 1100 amino acids or fewer, 1050 amino acids or fewer,



1000 amino acids or fewer, 950 amino acids or fewer, 900 amino acids or fewer, or 850 amino acids or fewer.

Suitably, the 3' CDS has a length of 900 nucleotides or more, 1000 nucleotides or more, 1100 nucleotides or more, 1200 nucleotides or more, 1300 nucleotides or more, 1400 nucleotides or more, 1500 nucleotides or more, 1600 nucleotides or more, 1700 nucleotides or more, 1800 nucleotides or more, 1900 nucleotides or more, 2000 nucleotides or more, 2100 nucleotides or more, 2200 nucleotides or more, 2300 nucleotides or more, or 2400 nucleotides or more.

Suitably, the 3' CDS has a length of 4200 nucleotides or fewer, 4100 nucleotides or fewer, 4000 nucleotides or fewer, 3900 nucleotides or fewer, 3800 nucleotides or fewer, 3700 nucleotides or fewer, 3600 nucleotides or fewer, 3500 nucleotides or fewer, 3400 nucleotides or fewer, 3300 nucleotides or fewer, 3200 nucleotides or fewer, 3100 nucleotides or fewer, 3000 nucleotides or fewer, 2900 nucleotides or fewer, 2800 nucleotides or fewer, 2700 nucleotides or fewer, 2600 nucleotides or fewer, or 2550 nucleotides or fewer.

#### *COL4A3 3' coding sequences*

15 In one embodiment, the 3' AAV vector comprises a 3' CDS encoding a C-terminal part of a COL4A3 polypeptide.

The C-terminal part of a COL4A3 polypeptide may be from any COL4A3 polypeptide known in the art or described herein.

Suitably, the 3' CDS comprises exons 48-52 or more, exons 47-52 or more, exons 46-52 or more, exons 45-52 or more, exons 44-52 or more, exons 43-52 or more, exons 42-52 or more, exons 41-52 or more, exons 40-52 or more, exons 39-52 or more, exons 38-52 or more, exons 37-52 or more, exons 36-52 or more, exons 35-52 or more, exons 34-52 or more, or exons 33-52 or more of a CDS encoding a COL4A3 polypeptide.

Suitably, the 3' CDS comprises exons 18-52 or less, exons 19-52 or less, exons 20-52 or less, exons 21-52 or less, exons 22-52 or less, exons 23-52 or less, exons 24-52 or less, exons 25-52 or less, exons 26-52 or less, exons 27-52 or less, or exons 28-52 or less of a CDS encoding a COL4A3 polypeptide.

Suitably, the 3' CDS comprises at least exons 43-52 to at most exons 18-52, at least exons 42-52 to at most exons 19-52, at least exons 41-52 to at most exons 20-52, at least exons 40-52 to at most exons 21-52, at least exons 39-52 to at most exons 22-52, at least exons 38-52 to at most exons 23-52, at least exons 37-52 to at most exons 24-52, at least exons 36-52 to at most exons 25-52, at least exons 35-52 to at most exons 26-52, at least exons 34-52 to at

most exons 27-52, or at least exons 33-52 to at most exons 28-52 of a CDS encoding a COL4A3 polypeptide.

Suitably, the 3' CDS comprises or consists of exons 44-52, exons 43-52, exons 42-52, exons 41-52, exons 40-52, exons 39-52, exons 38-52, exons 37-52, exons 36-52, exons 35-52, 5 exons 34-52, exons 33-52, exons 32-52, exons 31-52, exons 30-52, exons 29-52, exons 28-52, exons 27-52, exons 26-52, exons 25-52, exons 23-52, exons 23-52, or exons 22-52, exons 21-52, exons 20-52, exons 19-52, exons 18-52, exons 17-52, or exons 16-52 of a CDS encoding a COL4A3 polypeptide.

In some embodiments, the C-terminal part of a COL4A3 polypeptide is a C-terminal fragment 10 from an amino acid sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 1 to 7.

Suitably, the C-terminal part of a COL4A3 polypeptide comprises or consists of at least amino acids 1298-1670 and at most amino acids 285-1670 of SEQ ID NO: 1. In some embodiments, 15 the C-terminal part of a COL4A3 polypeptide comprises or consists of at least amino acids 1236-1670 to at most amino acids 436-1670, at least amino acids 1186-1670 to at most amino acids 486-1670, at least amino acids 1136-1670 to at most amino acids 536-1670, at least amino acids 1086-1670 to at most amino acids 586-1670, at least amino acids 1036-1670 to at most amino acids 636-1670, at least amino acids 986-1670 to at most amino acids 686-20 1670, or at least amino acids 936-1670 to at most amino acids 736-1670 of SEQ ID NO: 1.

In some embodiments, the 3' CDS is a 3' fragment of a nucleotide sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 27.

25 Examples of a 3' CDS encoding a C-terminal part of a COL4A3 polypeptide are shown in SEQ ID NOs: 202 to 208, corresponding to exons 35-52 (SEQ ID NO: 208), exons 34-52 (SEQ ID NO: 207), exons 33-52 (SEQ ID NO: 206), exons 32-52 (SEQ ID NO: 205), exons 31-52 (SEQ ID NO: 204), exons 30-52 (SEQ ID NO: 203), or exons 29-52 (SEQ ID NO: 202) of SEQ ID NO: 27.

30 In some embodiments, the 3' CDS comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 202 to 208.

In some embodiments, the 3' CDS comprises or consists of a nucleotide sequence selected from any of SEQ ID NOs: 202 to 208.

*COL4A4 3' coding sequences*

In one embodiment, the 3' AAV vector comprises a 3' CDS encoding a C-terminal part of a  
5 COL4A4 polypeptide.

The C-terminal part of a COL4A4 polypeptide may be from any COL4A4 polypeptide known in the art or described herein.

Suitably, the 3' CDS comprises exons 43-47 or more, exons 42-47 or more, exons 41-47 or more, exons 40-47 or more, exons 39-47 or more, exons 38-47 or more, exons 37-47 or more,  
10 exons 36-47 or more, exons 35-47 or more, exons 34-47 or more, exons 33-47 or more, exons 32-47 or more, exons 31-47 or more, exons 30-47 or more, exons 29-47 or more, or exons 28-47 or more of a CDS encoding a COL4A4 polypeptide.

Suitably, the 3' CDS comprises exons 13-47 or less, exons 14-47 or less, exons 15-47 or less, exons 16-47 or less, exons 17-47 or less, exons 18-47 or less, exons 19-47 or less, exons 20-  
15 47 or less, exons 21-47 or less, exons 22-47 or less, or exons 23-47 or less of a CDS encoding a COL4A4 polypeptide.

Suitably, the 3' CDS comprises at least exons 38-47 to at most exons 13-47, at least exons 37-47 to at most exons 14-47, at least exons 36-47 to at most exons 15-47, at least exons 35-47 to at most exons 16-47, at least exons 34-47 to at most exons 17-47, at least exons 33-47 to at most exons 18-47, at least exons 32-47 to at most exons 19-47, at least exons 31-47 to at most exons 20-47, at least exons 30-47 to at most exons 21-47, at least exons 29-47 to at most exons 22-47, or at least exons 28-47 to at most exons 23-47 of a CDS encoding a COL4A4 polypeptide.

Suitably, the 3' CDS comprises or consists of exons 40-47, exons 39-47, exons 38-47, exons 37-47, exons 36-47, exons 35-47, exons 34-47, exons 33-47, exons 32-47, exons 31-47, exons 30-47, exons 29-47, exons 28-47, exons 27-47, exons 26-47, exons 25-47, exons 24-47, exons 23-47, exons 22-47, exons 21-47, exons 20-47, or exons 19-47, exons 18-47, exons 17-47, exons 16-47, or exons 15-47 of a CDS encoding a COL4A4 polypeptide.

In some embodiments, the C-terminal part of a COL4A4 polypeptide is a C-terminal fragment  
30 from an amino acid sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least

96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 8 to 20.

Suitably, the C-terminal part of a COL4A4 polypeptide comprises or consists of at least amino acids 1298-1690 and at most amino acids 305-1690 of SEQ ID NO: 8. In some embodiments, the C-terminal part of a COL4A4 polypeptide comprises or consists of at least amino acids 1246-1690 to at most amino acids 446-1690, at least amino acids 1196-1690 to at most amino acids 496-1690, at least amino acids 1146-1690 to at most amino acids 546-1690, at least amino acids 1096-1690 to at most amino acids 596-1690, at least amino acids 1046-1690 to at most amino acids 646-1690, at least amino acids 996-1690 to at most amino acids 696-1690, or at least amino acids 946-1690 to at most amino acids 746-1690 of SEQ ID NO: 8.

In some embodiments, the 3' CDS is a 3' fragment of a nucleotide sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 81.

Suitably, the 3' CDS comprises or consists of at most nucleotides 916-5073 of SEQ ID NO: 81, or a nucleotide sequence having 500 or fewer, 400 or fewer, 300 or fewer, 200 or fewer, 100 or fewer, 90 or fewer, 80 or fewer, 70 or fewer, 60 or fewer, 50 or fewer, 40 or fewer, 30 or fewer, 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 nucleotide substitution(s) thereto.

Suitably, the 3' CDS comprises or consists of at least nucleotides 3894-5073 to at most nucleotides 916-5073 of SEQ ID NO: 81. In some embodiments, the 3' CDS comprises or consists of at least nucleotides 3536-5073 to at most nucleotides 1536-5073, at least nucleotides 3436-5073 to at most nucleotides 1636-5073, at least nucleotides 3336-5073 to at most nucleotides 1736-5073, at least nucleotides 3236-5073 to at most nucleotides 1836-5073, at least nucleotides 3136-5073 to at most nucleotides 1936-5073, at least nucleotides 3036-5073 to at most nucleotides 2036-5073, at least nucleotides 2936-5073 to at most nucleotides 2136-5073, or at least nucleotides 2836-5073 to at most nucleotides 2236-5073 of SEQ ID NO: 81.

Examples of a 3' CDS encoding a C-terminal part of a COL4A4 polypeptide are shown in SEQ ID NOs: 209 to 215, corresponding to exons 32-47 (SEQ ID NO: 215), exons 31-47 (SEQ ID NO: 214), exons 30-47 (SEQ ID NO: 213), exons 29-47 (SEQ ID NO: 212), exons 28-47 (SEQ ID NO: 211), exons 27-47 (SEQ ID NO: 210), or exons 26-47 (SEQ ID NO: 209) of SEQ ID NO: 81.

In some embodiments, the 3' CDS comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 209 to 215.

- 5 In some embodiments, the 3' CDS comprises or consists of a nucleotide sequence selected from any of SEQ ID NOs: 209 to 215.

#### COL4A5 3' coding sequences

In one embodiment, the 3' AAV vector comprises a 3' CDS encoding a C-terminal part of a COL4A5 polypeptide.

- 10 The C-terminal part of a COL4A5 polypeptide may be from any COL4A5 polypeptide known in the art or described herein.

Suitably, the 3' CDS comprises exons 47-51 or more, exons 46-51 or more, exons 45-51 or more, exons 44-51 or more, exons 43-51 or more, exons 42-51 or more, exons 41-51 or more, exons 40-51 or more, exons 39-51 or more, exons 38-51 or more, exons 37-51 or more, exons 36-51 or more, exons 35-51 or more, exons 34-51 or more, exons 33-51 or more, or exons 32-51 or more or more of a CDS encoding a COL4A5 polypeptide.

- 15 Suitably, the 3' CDS comprises exons 12-51 or less, exons 13-51 or less, exons 14-51 or less, exons 15-51 or less, exons 16-51 or less, exons 17-51 or less, exons 18-51 or less, exons 19-51 or less, exons 20-51 or less, exons 21-51 or less, or exons 22-51 or less of a CDS encoding a COL4A5 polypeptide.

- 20 Suitably, the 3' CDS comprises at least exons 42-51 to at most exons 12-51, at least exons 41-51 to at most exons 13-51, at least exons 40-51 to at most exons 14-51, at least exons 39-51 to at most exons 15-51, at least exons 38-51 to at most exons 16-51, at least exons 37-51 to at most exons 17-51, at least exons 36-51 to at most exons 18-51, at least exons 35-51 to at most exons 19-51, at least exons 34-51 to at most exons 20-51, at least exons 33-51 to at most exons 21-51, or at least exons 32-51 to at most exons 22-51 of a CDS encoding a COL4A5 polypeptide.

- 25 Suitably, the 3' CDS comprises or consists of exons 42-51, exons 41-51, exons 40-51, exons 39-51, exons 38-51, exons 37-51, exons 36-51, exons 35-51, exons 34-51, exons 33-51, exons 32-51, exons 31-51, exons 30-51, exons 29-51, exons 28-51, exons 27-51, exons 26-51, exons 25-51, exons 24-51, exons 23-51, exons 22-51, or exons 21-51, exons 20-51, exons 19-51, exons 18-51, exons 17-51, or exons 16-51 of a CDS encoding a COL4A5 polypeptide.

In some embodiments, the C-terminal part of a COL4A5 polypeptide is a C-terminal fragment from an amino acid sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 21 to 26.

Suitably, the C-terminal part of a COL4A5 polypeptide comprises or consists of at least amino acids 1298-1685 of SEQ ID NO: 21.

Suitably, the C-terminal part of a COL4A5 polypeptide comprises or consists of at most amino acids 300-1685 of SEQ ID NO: 21.

10 Suitably, the C-terminal part of a COL4A5 polypeptide comprises or consists of at least amino acids 1298-1685 and at most amino acids 300-1685 of SEQ ID NO: 21.

In some embodiments, the 3' CDS is a 3' fragment of a nucleotide sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 129.

Suitably, the 3' CDS comprises or consists of at least nucleotides 3894-5058 to at most nucleotides 901-5058 of SEQ ID NO: 129.

Examples of a 3' CDS encoding a C-terminal part of a COL4A5 polypeptide are shown in SEQ ID NOs: 289 to 295, 216 to 222 and 296 to 303, corresponding to exons 42-51 (SEQ ID NO: 303), exons 41-51 (SEQ ID NO: 302), exons 40-51 (SEQ ID NO: 301), exons 39-51 (SEQ ID NO: 300), exons 38-51 (SEQ ID NO: 299), exons 37-51 (SEQ ID NO: 298), exons 36-51 (SEQ ID NO: 297), exons 35-51 (SEQ ID NO: 296), exons 34-51 (SEQ ID NO: 222), exons 33-51 (SEQ ID NO: 221), exons 32-51 (SEQ ID NO: 220), exons 31-51 (SEQ ID NO: 219), exons 30-51 (SEQ ID NO: 218), exons 29-51 (SEQ ID NO: 217), exons 28-51 (SEQ ID NO: 216), exons 27-51 (SEQ ID NO: 295), exons 26-51 (SEQ ID NO: 294), exons 25-51 (SEQ ID NO: 293), exons 24-51 (SEQ ID NO: 292), exons 23-51 (SEQ ID NO: 291), exons 22-51 (SEQ ID NO: 290), or exons 21-51 (SEQ ID NO: 289) of SEQ ID NO: 129.

In some embodiments, the 3' CDS comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 289 to 295, 216 to 222 or 296 to 303.

In some embodiments, the 3' CDS comprises or consists of a nucleotide sequence selected from any of SEQ ID NOs: 289 to 295, 216 to 222 or 296 to 303.

Another example of a 3' CDS encoding a C-terminal part of a COL4A5 polypeptide is shown in SEQ ID NO: 307. In some embodiments, the 3' CDS comprises or consists of the nucleotide sequence of SEQ ID NO: 307.

5

|                                    |               |
|------------------------------------|---------------|
| Nucleotides<br>1724-5058<br>COL4A5 | SEQ ID NO:307 |
|------------------------------------|---------------|

Others examples of a 3' CDS encoding a C-terminal part of a COL4A5 polypeptide are shown in SEQ ID NOs: 308 to 310. In some embodiments, the 3' CDS comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 308-310.

10

In some embodiments, the 3' CDS comprises or consists of the nucleotide sequence of any of SEQ ID NOs: 308-310.

In some embodiments, the 3' CDS consists of the nucleotide sequence of any of SEQ ID NOs: 308-310.

15

|                                    |               |
|------------------------------------|---------------|
| Nucleotides<br>2705-5058<br>COL4A5 | SEQ ID NO:308 |
| Nucleotides<br>2855-5058<br>COL4A5 | SEQ ID NO:309 |
| Nucleotides<br>2955-5058<br>COL4A5 | SEQ ID NO:310 |

Another example of a 3' CDS encoding a C-terminal part of a COL4A5 polypeptide is shown in SEQ ID NO: 314. In some embodiments, the 3' CDS comprises or consists of the nucleotide sequence of SEQ ID NO: 314.

## 20 *Example overlapping 3' CDS (short)*

In some embodiments, the 3' CDS has a length of about 3600 nucleotides or less and comprises an overlapping sequence with the 5' CDS.

In some embodiments, the 3' CDS comprises or consists of exons 36-51, exons 35-51, exons 34-51, exons 33-51, exons 32-51, exons 31-51, exons 30-51, exons 29-51, exons 28-51, or exons 27-51 of a CDS encoding a COL4A5 polypeptide.

5 In some embodiments, the 3' CDS comprises or consists of exons 34-51, exons 33-51, exons 32-51, exons 31-51, exons 30-51, exons 29-51, exons 28-51 of a CDS encoding a COL4A5 polypeptide.

In some embodiments, the 3' CDS comprises or consists of exons 33-51, exons 32-51, exons 31-51, exons 30-51, or exons 29-51 of a CDS encoding a COL4A5 polypeptide.

10 In some embodiments, the 3' CDS comprises or consists of exons 32-51, exons 31-51, or exons 30-51, of a CDS encoding a COL4A5 polypeptide.

In some embodiments, the 3' CDS comprises or consists of exons 31-51 of a CDS encoding a COL4A5 polypeptide.

15 In some embodiments, the C-terminal part of a COL4A5 polypeptide comprises or consists of at least amino acids at least amino acids 1100-1685 to at most amino acids 650-1685, at least amino acids 1050-1685 to at most amino acids 700-1685, at least amino acids 1000-1685 to at most amino acids 700-1685, at least amino acids 1000-1685 to at most amino acids 750-1685, at least amino acids 1000-1685 to at most amino acids 800-1685, at least amino acids 950-1685 to at most amino acids 750-1685, at least amino acids 925-1685 to at most amino acids 775-1685, or at least amino acids 900-1685 to at most amino acids 800-1685 of SEQ ID NO: 21. In some embodiments, the C-terminal part of a COL4A5 polypeptide consists of about amino acids 901-1685 of SEQ ID NO: 21.

In some embodiments, the C-terminal part of a COL4A5 polypeptide consists of about amino acids 951-1685 of SEQ ID NO: 21.

25 In some embodiments, the C-terminal part of a COL4A5 polypeptide consists of about amino acids 985-1685 of SEQ ID NO: 21.

In some embodiments, the C-terminal part of a COL4A5 polypeptide consists of about amino acids 837-1685 of SEQ ID NO: 21.

30 In some embodiments, the 3' CDS comprises or consists of at least nucleotides 3200-5058 and at most nucleotides 2000-5058, at least nucleotides 3100-5058 and at most nucleotides 2000-5058, at least nucleotides 3000-5058 and at most nucleotides 2000-5058, at least nucleotides 3000-5058 and at most nucleotides 2100-5058, at least nucleotides 3000-5058



and at most nucleotides 2200-5058, at least nucleotides 2900-5058 and at most nucleotides 2200-5058, at least nucleotides 2800-5058 and at most nucleotides 2200-5058, at least nucleotides 2700-5058 and at most nucleotides 2300-5058, or at least nucleotides 2600-5058 and at most nucleotides 2400-5058 of SEQ ID NO: 129. In some embodiments, the 3' CDS consists of about nucleotides 2705-5058 of SEQ ID NO: 129.

In some embodiments, the 3' CDS consists of about nucleotides 2855-5058 of SEQ ID NO: 129.

In some embodiments, the 3' CDS consists of about nucleotides 2955-5058 of SEQ ID NO: 129.

10 In some embodiments, the 3' CDS consists of about nucleotides 2510-5058 of SEQ ID NO: 129.

In some embodiments, the 3' CDS comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 295, 216 to 222 or 296 to 297 or 308 to 310.

In some embodiments, the 3' CDS comprises or consists of a nucleotide sequence selected from any of SEQ ID NOs: 295, 216 to 222 or 296 to 297 or 308 to 310.

In some embodiments, the 3' CDS comprises or consists of a nucleotide sequence selected from any of SEQ ID NOs: 216 to 222 or 308 to 310.

20 In some embodiments, the 3' CDS comprises or consists of a nucleotide sequence selected from any of SEQ ID NOs: 217 to 221 or 308 to 310.

In some embodiments, the 3' CDS comprises or consists of a nucleotide sequence selected from any of SEQ ID NOs: 218 to 220 or 308 to 310.

In some embodiments, the 3' CDS comprises or consists of a nucleotide sequence selected from any of SEQ ID NOs: 219 or 308 to 310.

In some embodiments, the 3' CDS comprises or consists of the nucleotide sequence of SEQ ID NO: 219.

#### *Example overlapping 3' CDS (long)*

In some embodiments, the 3' CDS has a length of at least about 2700 nucleotides and comprises an overlapping sequence with the 5' CDS.

In some embodiments, the 3' CDS comprises or consists of exons 30-51, exons 29-51, exons 28-51, exons 27-51, exons 26-51, exons 25-51, exons 24-51, exons 23-51, exons 22-51, or exons 21-51 of a CDS encoding a COL4A5 polypeptide.

5 In some embodiments, the 3' CDS comprises or consists of exons 28-51, exons 27-51, exons 26-51, exons 25-51, exons 24-51, exons 23-51, or exons 22-51 of a CDS encoding a COL4A5 polypeptide.

In some embodiments, the 3' CDS comprises or consists of exons 27-51, exons 26-51, exons 25-51, exons 24-51, or exons 23-51 of a CDS encoding a COL4A5 polypeptide.

10 In some embodiments, the 3' CDS comprises or consists of exons 26-51, exons 25-51, or exons 24-51 of a CDS encoding a COL4A5 polypeptide.

In some embodiments, the 3' CDS comprises or consists of exons 25-51 of a CDS encoding a COL4A5 polypeptide.

15 In some embodiments, the C-terminal part of a COL4A5 polypeptide comprises or consists of at least amino acids 800-1685 to at most amino acids 400-1685, at least amino acids 750-1685 to at most amino acids 450-1685, at least amino acids 700-1685 to at most amino acids 500-1685, or at least amino acids 650-1685 to at most amino acids 550-1685 of SEQ ID NO: 21. In some embodiments, the C-terminal part of a COL4A5 polypeptide consists of about amino acids 574-1685 of SEQ ID NO: 21.

20 In some embodiments, the 3' CDS comprises or consists of at least nucleotides 2400-5058 and at most nucleotides 1200-5058, at least nucleotides 2300-5058 and at most nucleotides 1300-5058, at least nucleotides 2200-5058 and at most nucleotides 1400-5058, at least nucleotides 2100-5058 and at most nucleotides 1500-5058, at least nucleotides 2000-5058 and at most nucleotides 1600-5058, at least nucleotides 1900-5058 and at most nucleotides 1600-5058, or at least nucleotides 1800-5058 and at most nucleotides 1600-5058 of SEQ ID  
25 NO: 129. In some embodiments, the 3' CDS consists of about nucleotides 1724-5055 of SEQ ID NO: 129.

In some embodiments, the 3' CDS comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least  
30 99%, or 100% identity to any of SEQ ID NOs: 289-295, 216-218 or 307.

In some embodiments, the 3' CDS comprises or consists of a nucleotide sequence selected from any of SEQ ID NOs: 289-295, 216-218 or 307.

In some embodiments, the 3' CDS comprises or consists of a nucleotide sequence selected from any of SEQ ID NOs: 292-295 or 307.

In some embodiments, the 3' CDS comprises or consists of the nucleotide sequence of SEQ ID NO: 307.

5 *Example non-overlapping 3' CDS (mid-split)*

In some embodiments, the 3' CDS has a length of from about 1500 to about 3500 nucleotides and does not comprise an overlapping sequence with the 5' CDS.

10 In some embodiments, the 3' CDS consists of exons 42-51, exons 41-51, exons 40-51, exons 39-51, exons 38-51, exons 37-51, exons 36-51, exons 35-51, exons 34-51, exons 33-51, exons 32-51, exons 31-51, exons 30-51, exons 29-51, exons 28-51, exons 27-51, exons 26-51, exons 25-51, exons 24-51, exons 23-51, exons 22-51, or exons 21-51, exons 20-51, exons 19-51, exons 18-51, exons 17-51, or exons 16-51 of a CDS encoding a COL4A5 polypeptide.

15 In some embodiments, the 3' CDS comprises or consists of exons 34-51, exons 33-51, exons 32-51, exons 31-51, exons 30-51, exons 29-51, or exons 28-51 of a CDS encoding a COL4A5 polypeptide.

In some embodiments, the 3' CDS consists of exons 34-51, exons 33-51, exons 32-51, exons 31-51, exons 30-51, exons 29-51, or exons 28-51 of a CDS encoding a COL4A5 polypeptide.

In some embodiments, the 3' CDS comprises or consists of exons 34-51, exons 33-51, exons 32-51, or exons 30-51 of a CDS encoding a COL4A5 polypeptide.

20 In some embodiments, the 3' CDS consists of exons 34-51, exons 33-51, exons 32-51, or exons 30-51 of a CDS encoding a COL4A5 polypeptide.

In some embodiments, the 3' CDS consists of exons 34-51, exons 33-51, or exons 32-51 of a CDS encoding a COL4A5 polypeptide.

25 In some embodiments, the C-terminal part of a COL4A5 polypeptide comprises or consists of at least amino acids 1243-1685 to at most amino acids 443-1685, at least amino acids 1193-1685 to at most amino acids 493-1685, at least amino acids 1143-1685 to at most amino acids 543-1685, at least amino acids 1093-1685 to at most amino acids 593-1685, at least amino acids 1043-1685 to at most amino acids 643-1685, at least amino acids 993-1685 to at most amino acids 693-1685, or at least amino acids 943-1685 to at most amino acids 743-1685 of  
30 SEQ ID NO: 21. In some embodiments, the C-terminal part of a COL4A5 polypeptide consists of about amino acids 893-1686 of SEQ ID NO: 21.

In some embodiments, the C-terminal part of a COL4A5 polypeptide consists of about amino acids 923-1686 of SEQ ID NO: 21.

In some embodiments, the C-terminal part of a COL4A5 polypeptide consists of about amino acids 973-1686 of SEQ ID NO: 21.

- 5 In some embodiments, the 3' CDS consists of about nucleotides 2678-5058 of SEQ ID NO: 129.

In some embodiments, the 3' CDS consists of about nucleotides 2768-5058 of SEQ ID NO: 129.

- 10 In some embodiments, the 3' CDS consists of about nucleotides 2918-5058 of SEQ ID NO: 129.

In some embodiments, the 3' CDS comprises or consists of a nucleotide sequence selected from any of SEQ ID NOs: 216 to 222.

In some embodiments, the 3' CDS comprises or consists of a nucleotide sequence selected from any of SEQ ID NOs: 218, 220, 221, or 222.

- 15 In some embodiments, the 3' CDS comprises or consists of a nucleotide sequence selected from any of SEQ ID NOs: 220, 221, or 222.

*Example non-overlapping 3' CDS (long)*

In some embodiments, the 3' CDS has a length of at least about 2500 nucleotides and does not comprise an overlapping sequence with the 5' CDS.

- 20 In some embodiments, the 3' CDS consists of exons 29-51, exons 28-51, exons 27-51, exons 26-51, exons 25-51, exons 24-51, exons 23-51, exons 22-51, exons 21-51, exons 20-51 or exons 19-51 of a CDS encoding a COL4A5 polypeptide.

In some embodiments, the 3' CDS consists of exons 28-51, exons 27-51, exons 26-51, exons 25-51, exons 24-51, exons 23-51, or exons 22-51 of a CDS encoding a COL4A5 polypeptide.

- 25 In some embodiments, the 3' CDS consists of exons 27-51, exons 26-51, exons 25-51, exons 24-51, or exons 23-51 of a CDS encoding a COL4A5 polypeptide. In some embodiments, the 3' CDS consists of exons 26-51, exons 25-51, or exons 24-51 of a CDS encoding a COL4A5 polypeptide. In some embodiments, the 3' CDS consists of exons 25-51 of a CDS encoding a COL4A5 polypeptide.

In some embodiments, the 3' CDS consists of exons 21-51, exons 20-51, or exons 19-51 of a CDS encoding a COL4A5 polypeptide. In some embodiments, the 3' CDS consists of exons 20-51 of a CDS encoding a COL4A5 polypeptide.

5 In some embodiments, the C-terminal part of a COL4A5 polypeptide comprises or consists of at least amino acids 850-1685 to at most amino acids 350-1685, at least amino acids 850-1685 to at most amino acids 400-1685, at least amino acids 800-1685 to at most amino acids 400-1685, at least amino acids 750-1685 to at most amino acids 450-1685, at least amino acids 700-1685 to at most amino acids 500-1685, or at least amino acids 650-1685 to at most amino acids 550-1685 of SEQ ID NO: 21. In some embodiments, the C-terminal part of a  
10 COL4A5 polypeptide consists of about amino acids 594-1685 of SEQ ID NO: 21.

In some embodiments, the C-terminal part of a COL4A5 polypeptide consists of about amino acids 389-1685 of SEQ ID NO: 21.

In some embodiments, the 3' CDS comprises or consists of at least nucleotides 2800-5058 to at most nucleotides 1000-5058, at least nucleotides 2700-5058 to at most nucleotides 1100-  
15 5058, at least nucleotides 2600-5058 to at most nucleotides 1200-5058, at least nucleotides 2500-5058 to at most nucleotides 1200-5058, at least nucleotides 2400-5058 to at most nucleotides 1200-5058, at least 2300-5058 to at most nucleotides 1300-5058, at least nucleotides 2200-5058 to at most nucleotides 1400-5058, at least nucleotides 2100-5058 to at most nucleotides 1500-5058, at least nucleotides 2000-5058 to at most nucleotides 1600-  
20 5058, or at least nucleotides 1900-5058 to at most nucleotides 1700-5058 of SEQ ID NO: 129. In some embodiments, the 3' CDS consists of about nucleotides 1780-5058 of SEQ ID NO: 129.

In some embodiments, the 3' CDS consists of about nucleotides 1166-5058 of SEQ ID NO: 129.

25 In some embodiments, the 3' CDS comprises or consists of a nucleotide sequence selected from any of SEQ ID NOs: 289-295 or 216-217.

In some embodiments, the 3' CDS comprises or consists of a nucleotide sequence selected from any of SEQ ID NOs: 291-295.

In some embodiments, the 3' CDS comprises or consists of a nucleotide sequence selected  
30 from any of SEQ ID NOs: 292-294.

In some embodiments, the 3' CDS comprises or consists of the nucleotide sequence of SEQ ID NO: 293.

*Example non-overlapping 3' CDS (short)*

In some embodiments, the 3' CDS has a length of about 2500 nucleotides or less and does not comprise an overlapping sequence with the 5' CDS.

In some embodiments, the 3' CDS comprises or consists of exons 42-51, exons 41-51, exons 40-51, exons 39-51, exons 38-51, or exons 37-51 of a CDS encoding a COL4A5 polypeptide.

In some embodiments, the 3' CDS comprises or consists of exons 42-51, exons 41-51, exons 40-51, exons 39-51, or exons 38-51 of a CDS encoding a COL4A5 polypeptide.

In some embodiments, the 3' CDS comprises or consists of exons 42-51, exons 41-51, or exons 40-51 of a CDS encoding a COL4A5 polypeptide.

10 In some embodiments, the 3' CDS comprises or consists of exons 42-51 of a CDS encoding a COL4A5 polypeptide.

In some embodiments, the C-terminal part of a COL4A5 polypeptide comprises or consists of at least amino acids 1298-1685 to at most amino acids 1000-1685, at least amino acids 1298-1685 to at most amino acids 1050-1685, at least amino acids 1298-1685 to at most amino acids 1100-1685, at least amino acids 1298-1685 to at most amino acids 1150-1685, at least amino acids 1298-1685 to at most amino acids 1200-1685, at least amino acids 1298-1685 to at most amino acids 1250-1685 of SEQ ID NO: 21. In some embodiments, the C-terminal part of a COL4A5 polypeptide consists of about amino acids 1264-1685 of SEQ ID NO: 21.

20 In some embodiments, the 3' CDS comprises or consists of at least nucleotides 3894-5058 to at most nucleotides 3200-5058, at least nucleotides 3894-5058 to at most nucleotides 3300-5058, at least nucleotides 3894-5058 to at most nucleotides 3400-5058, at least nucleotides 3894-5058 to at most nucleotides 3500-5058, at least nucleotides 3894-3600 to at most nucleotides 3600-5058, or at least nucleotides 3894-5058 to at most nucleotides 3700-5058 of SEQ ID NO: 129. In some embodiments, the 3' CDS consists of about nucleotides 3791-5058 of SEQ ID NO: 129.

In some embodiments, the 3' CDS comprises or consists of a nucleotide sequence selected from any of SEQ ID NOs: 298-303.

In some embodiments, the 3' CDS comprises or consists of a nucleotide sequence selected from any of SEQ ID NOs: 301-303.

30 In some embodiments, the 3' CDS comprises or consists of the nucleotide sequence of SEQ ID NO: 303.

*Other elements*

The 3' AAV vector may further comprise any suitable elements for expressing a COL4A3, COL4A4 or COL4A5 polypeptide and/or forming a full-length CDS encoding a COL4A3, COL4A4 or COL4A5 polypeptide.

- 5 The 3' AAV vector may further comprise any suitable elements downstream of the 3' CDS.

The 3' AAV vector may comprise a polyadenylation sequence downstream of the 3' CDS. Typically, the 3' AAV vector comprises a polyadenylation downstream of the 3' CDS, such that a poly(A) tail is added to the end of the full-length CDS once it is formed. The polyadenylation sequence may be operably linked to the 3' CDS. The polyadenylation sequence may be any  
 10 described herein. Suitably, the 3' AAV vector comprises or consists of from 5' to 3': a 3' CDS encoding a C-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide and a polyadenylation sequence.

In some embodiments, the 3' AAV vector comprises one or regulatory elements downstream of the 3' CDS, which act to increase expression of a COL4A3, COL4A4 or COL4A5  
 15 polypeptide. The one or more regulatory elements may be operably linked to the 3' CDS. The one or more regulatory elements downstream of the 3' CDS may be any described herein. Suitably, the 3' AAV vector comprises or consists of from 5' to 3': a 3' CDS encoding a C-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide, one or more regulatory elements, and a polyadenylation sequence.

20 In some embodiments, the one or more regulatory elements downstream from the 3' CDS comprise a WPRE. The WPRE may be operably linked to the 3' CDS. The WPRE may be any described herein. Suitably, the 3' AAV vector comprises or consists of from 5' to 3': a 3' CDS encoding a C-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide, a WPRE, and a polyadenylation sequence.

- 25 The 3' AAV vector may further comprise any suitable elements upstream of the 3' CDS.

In some embodiments, the 3' AAV vector comprises a splice acceptor sequence upstream from the 3' CDS. As described in more detail below, such a splice acceptor sequence may be used in DNA trans-splicing or hybrid strategies. The splice acceptor sequence may be any described herein. Suitably, the 3' AAV vector comprises or consists of from 5' to 3': a splice  
 30 acceptor sequence, a 3' CDS encoding a C-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide, optionally one or more regulatory elements (e.g. a WPRE), and a polyadenylation sequence.

In some embodiments, the 3' AAV vector further comprises a recombinogenic sequence upstream from the splice acceptor sequence. As described in more detail below, such a recombinogenic sequence may be used in DNA hybrid strategies. The recombinogenic sequence may be any described herein. Suitably, the 3' AAV vector comprises or consists of  
5 from 5' to 3': a recombinogenic sequence, a splice acceptor sequence, a 3' CDS encoding a C-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide, optionally one or more regulatory elements (e.g. a WPRE), and a polyadenylation sequence.

In other embodiments, the 3' AAV vector comprises a promoter upstream from the 3' CDS. As described in more detail below, a promoter may be used in mRNA trans-splicing strategies.  
10 The promoter may be operably linked to the 3' CDS. The promoter may be any described herein. Suitably, the 3' AAV vector comprises or consists of from 5' to 3': a promoter, a 3' CDS encoding a C-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide, optionally one or more regulatory elements (e.g. a WPRE), and a polyadenylation sequence.

In some embodiments, the 3' AAV vector comprises one or more regulatory elements upstream from the 3' CDS. The one or more regulatory elements may be operably linked to the 3' CDS. The one or more regulatory elements upstream from the 3' CDS may be any described herein. Suitably, the 3' AAV vector comprises or consists of from 5' to 3': a promoter,  
15 one or more regulatory elements, a 3' CDS encoding a C-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide, optionally one or more regulatory elements (e.g. a WPRE), and a polyadenylation sequence.  
20

In some embodiments, the one or more regulatory elements upstream of the 3' CDS comprise a spliceosomal intron or a fragment thereof. The spliceosomal intron or a fragment thereof may be operably linked to the 3' CDS. The spliceosomal intron or a fragment thereof may be any described herein. Suitably, the 3' AAV vector comprises or consists of from 5' to 3':  
25 a promoter, a spliceosomal intron or a fragment thereof, a 3' CDS encoding a C-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide, optionally one or more regulatory elements (e.g. a WPRE), and a polyadenylation sequence.

In some embodiments, the one or more regulatory elements upstream of the 3' CDS comprise a Kozak sequence. The Kozak sequence may be operably linked to the 3' CDS. The Kozak sequence may be any described herein. Suitably, the 3' AAV vector comprises or consists of  
30 from 5' to 3': a promoter, a Kozak sequence, a 3' CDS encoding a C-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide, optionally one or more regulatory elements (e.g. a WPRE), and a polyadenylation sequence.



In some embodiments, the one or more regulatory elements upstream of the 3' CDS comprise a spliceosomal intron or a fragment thereof and a Kozak sequence. The spliceosomal intron or a fragment thereof and the Kozak sequence may each be operably linked to the 3' CDS. Suitably, the 3' AAV vector comprises or consists of from 5' to 3': a promoter, a spliceosomal intron or a fragment thereof, a Kozak sequence, a 3' CDS encoding a C-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide, optionally one or more regulatory elements (e.g. a WPRE), and a polyadenylation sequence.

In some embodiments, the 3' AAV vector comprises a splice acceptor sequence upstream from the 3' CDS. As described in more detail below, a splice acceptor sequence may be used in mRNA trans-splicing strategies. The splice acceptor sequence may be any described herein (e.g. any reverse splice acceptor sequence described herein). Suitably, the 3' AAV vector comprises or consists of from 5' to 3': a promoter, optionally one or more regulatory elements (e.g. a spliceosomal intron or a fragment thereof and/or a Kozak sequence), a splice acceptor sequence, a 3' CDS encoding a C-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide, optionally one or more regulatory elements (e.g. a WPRE), and a polyadenylation sequence.

In some embodiments, the 3' AAV vector further comprises a complementary hybridisation domain upstream from the 3' CDS. As described in more detail below, a complementary hybridisation domain may be used in mRNA trans-splicing strategies. The complementary hybridisation domain may be any described herein. Suitably, the 3' AAV vector comprises or consists of from 5' to 3': a promoter, optionally one or more regulatory elements (e.g. a spliceosomal intron or a fragment thereof and/or a Kozak sequence), a complementary hybridisation domain, a splice acceptor sequence, a 3' CDS encoding a C-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide, optionally one or more regulatory elements (e.g. a WPRE), and a polyadenylation sequence.

#### Promoters

The 5' AAV vector of the present invention may comprise a promoter. Suitably, the promoter is upstream of the 5' CDS. Suitably, the promoter may be operably linked to the 5' CDS.

In some embodiments (e.g. in mRNA trans-splicing method), the 3' AAV vector of the present invention comprises a promoter. Suitably, the promoter is upstream of the 3' CDS. Suitably, the promoter may be operably linked to the 3' CDS.

A "promoter" is a region of DNA that leads to initiation of transcription of a gene. Promoters are located near the transcription start sites of genes, upstream on the DNA (towards the 5'

region of the sense strand). Any suitable promoter may be used, the selection of which may be readily made by the skilled person. The promoters comprised in the 5' AAV vector and the 3' AAV vector may be identical or different. In some embodiments, the promoters comprised in the 5' AAV vector and the 3' AAV vector are identical.

- 5 Suitably, the promoter is operable in mammalian cells, e.g. human cells. The promoter may be capable of driving expression of a protein-coding sequence in mammalian cells, e.g. human cells. The promoter may be a mammalian promoter, e.g. a human promoter.

- Suitably, the promoter may be operable in kidney cells. The promoter may be capable of driving expression of a CDS in the kidney. Examples of kidney cells include, but are not limited to glomerular cells, and tubular cells. The promoter may be operable in glomerular cells or tubular cells. The promoter may be capable of driving expression of a CDS in the glomerulus. The promoter may be operable in a podocyte cell. The promoter may be capable of driving expression of a CDS in podocytes.
- 10

Tubular cells are also known in the art as tubular epithelial cells.

- 15 The promoter may be a ubiquitous promoter or a tissue-specific promoter. The promoter may be a constitutive promoter, an inducible promoter, or a repressible promoter.

- In some embodiments, the promoter upstream of the 5' CDS is selected from: a CMV promoter, a CBA promoter, a CAG promoter, a CB7 promoter, an EF1 $\alpha$  promoter, a CMV/EF1 $\alpha$  hybrid promoter, a NF-kB promoter, a pSE-7 promoter, a mPGK promoter, a mU1a promoter, a U6 promoter, a U7 promoter, a MNLU3 promoter, a HLP promoter, an AAT promoter, an ALB promoter, a ApoE/AAT promoter, a EalbAAT promoter, a LP1 promoter, a TBG promoter, a TTR promoter, a SYN1 promoter, a NSE promoter, a tMCK promoter, a CK8 promoter, a MHCK7 promoter, a SMN promoter, a DES promoter, a RK promoter, a hRHO promoter, a GRK1 promoter, a hCAR promoter, a hRPE65p promoter, a P546 promoter, a PR1.7 promoter, a hRS1 promoter, a VMD2 promoter, an  $\alpha$ -MHC promoter, a FRE1 promoter, a NPHS1 promoter, and a NPHS2 promoter.
- 20
- 25

- In some embodiments, the promoter upstream of the 3' CDS is selected from: a CMV promoter, a CBA promoter, a CAG promoter, a CB7 promoter, an EF1 $\alpha$  promoter, a CMV/EF1 $\alpha$  hybrid promoter, a NF-kB promoter, a pSE-7 promoter, a mPGK promoter, a mU1a promoter, a U6 promoter, a U7 promoter, a MNLU3 promoter, a HLP promoter, an AAT promoter, an ALB promoter, a ApoE/AAT promoter, a EalbAAT promoter, a LP1 promoter, a TBG promoter, a TTR promoter, a SYN1 promoter, a NSE promoter, a tMCK promoter, a CK8 promoter, a MHCK7 promoter, a SMN promoter, a DES promoter, a RK promoter, a hRHO
- 30

promoter, a GRK1 promoter, a hCAR promoter, a hRPE65p promoter, a P546 promoter, a PR1.7 promoter, a hRS1 promoter, a VMD2 promoter, an  $\alpha$ -MHC promoter, a FRE1 promoter, a NPHS1 promoter, and a NPHS2 promoter.

- In some embodiments, the promoter upstream of the 5' CDS and the promoter upstream of the 3' CDS are selected from: a CMV promoter, a CBA promoter, a CAG promoter, a CB7 promoter, an EF1 $\alpha$  promoter, a CMV/EF1 $\alpha$  hybrid promoter, a NF-kB promoter, a pSE-7 promoter, a mPGK promoter, a mU1a promoter, a U6 promoter, a U7 promoter, a MNDU3 promoter, a HLP promoter, an AAT promoter, an ALB promoter, a ApoE/AAT promoter, a EalbAAT promoter, a LP1 promoter, a TBG promoter, a TTR promoter, a SYN1 promoter, a NSE promoter, a tMCK promoter, a CK8 promoter, a MHCK7 promoter, a SMN promoter, a DES promoter, a RK promoter, a hRHO promoter, a GRK1 promoter, a hCAR promoter, a hRPE65p promoter, a P546 promoter, a PR1.7 promoter, a hRS1 promoter, a VMD2 promoter, an  $\alpha$ -MHC promoter, a FRE1 promoter, a NPHS1 promoter, and a NPHS2 promoter.

#### *Ubiquitous promoters*

- In some embodiments, the promoter is a ubiquitous promoter. In some embodiments, the promoter upstream of the 5' CDS is a ubiquitous promoter. In some embodiments, the promoter upstream of the 3' CDS is a ubiquitous promoter. In some embodiments, the promoter upstream of the 5' CDS and the promoter upstream of the 3' CDS are each a ubiquitous promoter.
- As used herein, a "ubiquitous promoter" is a promoter which is active in a wide range of cells and tissues. Suitable ubiquitous promoters will be known to the skilled person. Example ubiquitous promoters may include a CMV promoter, a CBA promoter, a CAG promoter, a CB7 promoter, a EF1 $\alpha$  promoter, a CMV/EF1 $\alpha$  hybrid promoter, a NF-kB promoter, a pSE-7 promoter, a mPGK promoter, a mU1a promoter, a U6 promoter, a U7 promoter, and a MNDU3 promoter.

In some embodiments, the promoter upstream of the 5' CDS is selected from: a CMV promoter, a CBA promoter, a CAG promoter, a CB7 promoter, an EF1 $\alpha$  promoter, a CMV/EF1 $\alpha$  hybrid promoter, a NF-kB promoter, a pSE-7 promoter, a mPGK promoter, a mU1a promoter, a U6 promoter, and a U7 promoter.

- In some embodiments, the promoter upstream of the 5' CDS is a CMV promoter, a CBA promoter, or a CAG promoter. In some embodiments, the promoter upstream of the 5' CDS is a CMV promoter.

In some embodiments, the promoter upstream of the 3' CDS is selected from: a CMV promoter, a CBA promoter, a CAG promoter, a CB7 promoter, an EF1 $\alpha$  promoter, a CMV/EF1 $\alpha$  hybrid promoter, a NF-kB promoter, a pSE-7 promoter, a mPGK promoter, a mU1a promoter, a U6 promoter, and a U7 promoter.

- 5 In some embodiments, the promoter upstream of the 3' CDS is a CMV promoter, a CBA promoter, or a CAG promoter. In some embodiments, the promoter upstream of the 3' CDS is a CMV promoter.

- 10 In some embodiments, the promoter upstream of the 5' CDS and the promoter upstream of the 3' CDS are selected from: a CMV promoter, a CBA promoter, a CAG promoter, a CB7 promoter, an EF1 $\alpha$  promoter, a CMV/EF1 $\alpha$  hybrid promoter, a NF-kB promoter, a pSE-7 promoter, a mPGK promoter, a mU1a promoter, a U6 promoter, and a U7 promoter.

- 15 In some embodiments, the promoter upstream of the 5' CDS and the promoter upstream of the 3' CDS are both a CMV promoter, a CBA promoter, or a CAG promoter. In some embodiments, the promoter upstream of the 5' CDS and the promoter upstream of the 3' CDS are both a CMV promoter.

The cytomegalovirus (CMV) promoter is commonly included in AAV vectors, because it is a strong promoter and drives constitutive expression of genes under its control. An example CMV promoter is shown in SEQ ID NO: 223.

- 20 In some embodiments, the promoter upstream of the 5' CDS comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 223. In some embodiments, the promoter upstream of the 5' CDS comprises or consists of the nucleotide sequence of SEQ ID NO: 223.

- 25 In some embodiments, the promoter upstream of the 3' CDS comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 223. In some embodiments, the promoter upstream of the 3' CDS comprises or consists of the nucleotide sequence of SEQ ID NO: 223.
- 30

In some embodiments, the promoter upstream of the 5' CDS and the promoter upstream of the 3' CDS each comprise or consist of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at

least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 223. In some embodiments, the promoter upstream of the 5' CDS and the promoter upstream of the 3' CDS each comprise or consist of the nucleotide sequence of SEQ ID NO: 223.

- 5 The chicken  $\beta$ -actin (CBA) promoter is also commonly included in AAV vectors, because it is also a strong promoter and drives constitutive expression of genes under its control. An example CBA promoter is shown in SEQ ID NO: 224.

In some embodiments, the promoter upstream of the 5' CDS comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least  
10 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 224. In some embodiments, the promoter upstream of the 5' CDS comprises or consists of the nucleotide sequence of SEQ ID NO: 224.

In some embodiments, the promoter upstream of the 3' CDS comprises or consists of a  
15 nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 224. In some embodiments, the promoter upstream of the 3' CDS comprises or consists of the nucleotide sequence of SEQ ID NO: 224.

20 In some embodiments, the promoter upstream of the 5' CDS and the promoter upstream of the 3' CDS each comprise or consist of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 224. In some embodiments, the promoter upstream of the 5' CDS and  
25 the promoter upstream of the 3' CDS each comprise or consist of the nucleotide sequence of SEQ ID NO: 224.

The CAG promoter is a strong synthetic promoter frequently used to drive high levels of gene expression from AAV vectors. The CAG promoter comprises the CMV early enhancer element combined with the first exon and the first intron of the CBA promoter. An example CAG  
30 promoter is shown in SEQ ID NO: 225.

In some embodiments, the promoter upstream of the 5' CDS comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at

least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 225. In some embodiments, the promoter upstream of the 5' CDS comprises or consists of the nucleotide sequence of SEQ ID NO: 225.

In some embodiments, the promoter upstream of the 3' CDS comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 225. In some embodiments, the promoter upstream of the 3' CDS comprises or consists of the nucleotide sequence of SEQ ID NO: 225.

In some embodiments, the promoter upstream of the 5' CDS and the promoter upstream of the 3' CDS each comprise or consist of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 225. In some embodiments, the promoter upstream of the 5' CDS and the promoter upstream of the 3' CDS each comprise or consist of the nucleotide sequence of SEQ ID NO: 225.

Other ubiquitous promoters which have been used in AAV vectors include a CBA promoter/CMV enhancer (CB7) promoter, an Elongation Factor-1 alpha (EF1α) promoter, a hybrid CMV/EF1α promoter, a Nuclear Factor-κB (NF-κB) promoter, a Proximal Sequence Element 7 (PSE-7) promoter, a Murine Phosphoglycerate Kinase (mPGK) promoter, a Murine small nuclear RNA promoter (mU1a) and a RNA Polymerase III U6 (U6) promoter (see e.g. Au, H.K.E., et al., 2022. *Frontiers in medicine*, 8, p.809118).

#### *Tissue-specific promoters*

In some embodiments, the promoter is a cell-specific or tissue-specific promoter. In some embodiments, the promoter upstream of the 5' CDS is a cell-specific or tissue-specific promoter. In some embodiments, the promoter upstream of the 3' CDS is a cell-specific or tissue-specific promoter. In some embodiments, the promoter upstream of the 5' CDS and the promoter upstream of the 3' CDS are each a cell-specific or tissue-specific promoter.

As used herein, a "cell-specific promoter" or "tissue-specific promoter" is a promoter which preferentially facilitates expression of a protein-coding sequence in a specific type of cells or tissue (see e.g. Zheng, C. and Baum, B.J., 2008. *Gene Therapy Protocols: Design and Characterization of Gene Transfer Vectors*, pp.205-219). Suitably, a cell-specific or tissue-specific promoter may facilitate higher expression of a protein-coding sequence in one cell-

type or tissue as compared to other cell-types or tissues. For example, a cell-specific or tissue-specific promoter may be a promoter which facilitates expression of a protein-coding sequence at least 10% higher, at least 20% higher, at least 30% higher, at least 40% higher, at least 50% higher, at least 100% higher, at least 200% higher, at least 300% higher, at least 400% higher, at least 500% higher, or at least 1000% higher in one cell-type or tissue as compared to expression in other cell-types or tissues.

Suitable tissue-specific promoters will be known to the skilled person (see e.g. Toscano, M.G., et al., 2011. Gene therapy, 18(2), pp.117-127; and Powell, S.K., et al., 2015. Discovery medicine, 19(102), p.49) and can be generated using methods known in the art (see and Shen, S.Q., et al., 2016. Genome research, 26(2), pp.238-255). In some embodiments, the promoter is a kidney-specific promoter, a neuron-specific promoter, an astrocyte-specific promoter, an oligodendrocyte-specific promoter, a retina-specific promoter, a lung-specific promoter, a liver-specific promoter, a pancreas-specific promoter, a cardiac-specific promoter, or a skeletal muscle-specific promoter.

Kidney-specific promoters include a NPHS1 promoter and a NPHS2 promoter and are described in more detail in the section below.

Neuro-specific promoters include the Proximal region of the Synapsin I (a SYN1 promoter) and a Rat Neuron-Specific Enolase (NSE) promoter.

Retina-specific promoters include a Human Rhodopsin Kinase (RK) promoter, a hRHO promoter, a Human Rhodopsin Kinase (GRK1) promoter, a Human Cone Arrestin (hCAR) promoter, a Human retinal pigmented epithelium (hRPE65p) promoter, a Truncated Mecp2-promoter (a P546 promoter), a 1.7-Kb L-Op sin Promoter (a PR1.7 promoter), a hRS1 promoter, and a vitelliform macular dystrophy-2 (VMD2) promoter.

Liver-specific promoters include a Hybrid Human Liver (HLP) promoter, an A-1 Antitrypsin (AAT) promoter, an albumin (ALB) promoter, a ApoE/AAT promoter, a EalbAAT promoter, LP1 promoter, a Thyroxine-Binding Globulin (TBG) promoter, and a Transthyretin (TTR) Promoter.

Cardiac-specific promoters include an  $\alpha$ -myosin heavy chain ( $\alpha$ -MHC) promoter.

Muscle-specific promoters include a tMCK promoter, Mouse Creatine Kinase promoter/enhancer element (a CK8 promoter), a Murine Muscle Creatine Kinase (CK) and  $\alpha$ -myosin heavy-chain genes (MHCK7) promoter, a Survival Motor Neuron (SMN) promoter, and a Human Desmin enhancer/promoter (DES) promoter.

The cell-specific or tissue-specific promoter may be a minimal promoter. As used, herein, a “minimal promoter” means the minimal sequence that can act as a promoter.

*Kidney-specific promoters*

In preferred embodiments, the promoter is a kidney-specific promoter. In some embodiments, the promoter upstream of the 5' CDS is a kidney-specific promoter. In some embodiments, the promoter upstream of the 3' CDS is a kidney-specific promoter. In some embodiments, the promoter upstream of the 5' CDS and the promoter upstream of the 3' CDS are each a kidney-specific promoter. Suitable kidney-specific promoters will be known to the skilled person.

In some embodiments, the promoter is a tubular cell specific promoter. Tubular cells are also known in the art as tubular epithelial cells. In preferred embodiments, the promoter is a glomerular-specific promoter. In some embodiments, the promoter upstream of the 5' CDS is a glomerular-specific promoter. In some embodiments, the promoter upstream of the 3' CDS is a glomerular-specific promoter. In some embodiments, the promoter upstream of the 5' CDS and the promoter upstream of the 3' CDS are each a glomerular-specific promoter. Suitable glomerular-specific promoters will be known to the skilled person.

In preferred embodiments, the promoter is a podocyte-specific promoter. In some embodiments, the promoter upstream of the 5' CDS is a podocyte-specific promoter. In some embodiments, the promoter upstream of the 3' CDS is a podocyte-specific promoter. In some embodiments, the promoter upstream of the 5' CDS and the promoter upstream of the 3' CDS are each a podocyte-specific promoter. Suitable podocyte-specific promoters will be known to the skilled person.

Suitably, a podocyte-specific promoter may be or may be derived from a promoter associated with a gene with selective expression in human podocytes. Genes selectively expressed in podocytes will be known to those of skill in the art and selective gene expression in podocytes can be readily determined by methods known to those of skill in the art, for instance with microarrays. Genes selectively expressed in podocytes include *NPHS1*, *NPHS2*, *WT1*, *FOXC2*, *ABCA9*, *ACPP*, *ACTN4*, *ADM*, *ANGPTL2*, *ANXA1*, *ASB15*, *ATP8B1*, *B3GALT2*, *BB014433*, *BMP7*, *C1QTNF1*, *CAR13*, *CD2AP*, *CD55*, *CD59A*, *CD59B*, *CDC14A*, *CDH3*, *CDKN1B*, *CDKN1C*, *CEP85L*, *CLIC3*, *CLIC5*, *COL4A1*, *COL4A2*, *COL4A3*, *COL4A4*, *COL4A5*, *COLEC12*, *CRIM1*, *CST12*, *DEGS1*, *DOCK4*, *DOCK5*, *EGF*, *ENPEP*, *EPHX1*, *FAM81A*, *FAT1*, *FGFBP1*, *FOXD1*, *FRYL*, *GABRB1*, *GALC*, *GM10554*, *H2-D1*, *H2-Q7*, *H2BC4*, *H3C15*, *HS3ST3A1*, *HTRA1*, *IFNGR1*, *IL18*, *ILDR2*, *ITGB5*, *ITGB8*, *KIRREL*, *LAMA1*, *LAMA5*, *LAMB1*, *LAMB2*, *LMX1B*, *MAFB*, *MAGI2*, *MELA*, *MERTK*, *MGAT4A*, *MYO1D*,



*MYO1E, MYOM2, MYZAP, NEBL, NES, NOD1, NPR3, NR2F2, NUPR1, OPTN, P3H2, PAK1, PARD3B, PDPN, PLAT, PLCE1, PLSCR2, PODXL, PROS1, PTPRO, RAB3B, RDH1, RDH9, SDC4, SEMA3E, SERPINB6B, SH3BGRL2, SLC41A2, SLCO2A1, ST3GAL6, SYNPO, TDRD5, THSD7A, TIMP3, TJP1, TLR7, TM4SF1, TMEM108, TMEM54, TMTC1, TOP1MT,*  
 5 *TRAV10, TRAV10N, TRAV5-4, TSHB, UACA, UBA1Y, UPRT, VEGFA, VTCN1, ZBTB20, and 5730407107RIK.*

Methods to identify the promoter regions associated with genes will be well known to those of skill in the art. The promoter is usually located just proximal to or overlapping the transcription initiation site and contains several sequence motifs with which transcription factors (TFs)  
 10 interact in a sequence-specific manner.

Suitably, a podocyte-specific promoter is selected from a NPHS1 promoter, a NPHS2 promoter, a WT1 promoter, a FOXC2 promoter, a ABCA9 promoter, a ACPD promoter, a ACTN4 promoter, a ADM promoter, a ANGPTL2 promoter, a ANXA1 promoter, a ASB15 promoter, a ATP8B1 promoter, a B3GALT2 promoter, a BB014433 promoter, a BMP7  
 15 promoter, a C1QTNF1 promoter, a CAR13 promoter, a CD2AP promoter, a CD55 promoter, a CD59A promoter, a CD59B promoter, a CDC14A promoter, a CDH3 promoter, a CDKN1B promoter, a CDKN1C promoter, a CEP85L promoter, a CLIC3 promoter, a CLIC5 promoter, a COL4A1 promoter, a COL4A2 promoter, a COL4A3 promoter, a COL4A4 promoter, a COL4A5 promoter, a COLEC12 promoter, a CRIM1 promoter, a CST12 promoter, a DEGS1 promoter,  
 20 a DOCK4 promoter, a DOCK5 promoter, a EGF promoter, a ENPEP promoter, a EPHX1 promoter, a FAM81A promoter, a FAT1 promoter, a FGFBP1 promoter, a FOXD1 promoter, a FRYL promoter, a GABRB1 promoter, a GALC promoter, a GM10554 promoter, a H2-D1 promoter, a H2-Q7 promoter, a H2BC4 promoter, a H3C15 promoter, a HS3ST3A1 promoter, a HTRA1 promoter, a IFNGR1 promoter, a IL18 promoter, a ILDR2 promoter, a ITGB5  
 25 promoter, a ITGB8 promoter, a KIRREL promoter, a LAMA1 promoter, a LAMA5 promoter, a LAMB1 promoter, a LAMB2 promoter, a LMX1B promoter, a MAFB promoter, a MAGI2 promoter, a MELO promoter, a MERTK promoter, a MGAT4A promoter, a MYO1D promoter, a MYO1E promoter, a MYOM2 promoter, a MYZAP promoter, a NEBL promoter, a NES promoter, a NOD1 promoter, a NPR3 promoter, a NR2F2 promoter, a NUPR1 promoter, a  
 30 OPTN promoter, a P3H2 promoter, a PAK1 promoter, a PARD3B promoter, a PDPN promoter, a PLAT promoter, a PLCE1 promoter, a PLSCR2 promoter, a PODXL promoter, a PROS1 promoter, a PTPRO promoter, a RAB3B promoter, a RDH1 promoter, a RDH9 promoter, a SDC4 promoter, a SEMA3E promoter, a SERPINB6B promoter, a SH3BGRL2 promoter, a SLC41A2 promoter, a SLCO2A1 promoter, a ST3GAL6 promoter, a SYNPO promoter, a  
 35 TDRD5 promoter, a THSD7A promoter, a TIMP3 promoter, a TJP1 promoter, a TLR7

promoter, a TM4SF1 promoter, a TMEM108 promoter, a TMEM54 promoter, a TMTC1 promoter, a TOP1MT promoter, a TRAV10 promoter, a TRAV10N promoter, a TRAV5-4 promoter, a TSHB promoter, a UACA promoter, a UBA1Y promoter, a UPRT promoter, a VEGFA promoter, a VTCN1 promoter, a ZBTB20 promoter, and a 5730407107RIK promoter,  
 5 or a variant thereof.

Suitably, a podocyte-specific promoter is selected from a NPHS1 promoter, a NPHS2 promoter, a WT1 promoter, a FOXC2 promoter, a ACTN4 promoter, a BMP7 promoter, a CD2AP promoter, a CDH3 promoter, a CDKN1B promoter, a CDKN1C promoter, a COL4A1 promoter, a COL4A2 promoter, a COL4A3 promoter, a COL4A4 promoter, a COL4A5  
 10 promoter, a CRIM1 promoter, a FAT1 promoter, a FOXD1 promoter, a KIRREL promoter, a LAMA1 promoter, a LAMA5 promoter, a LAMB1 promoter, a LAMB2 promoter, a LMX1B promoter, a MAFB promoter, a NES promoter, a NR2F2 promoter, a PODXL promoter, a PTPRO promoter, a SYNPO promoter, a TJP1 promoter, and a VEGFA promoter, or a variant thereof.

15 Suitably, a podocyte-specific promoter is a NPHS1 promoter, a NPHS2 promoter, a WT1 promoter, or a FOXC2 promoter, or a variant thereof.

Preferably, a podocyte-specific promoter is a NPHS1 or a NPHS2 promoter, or a variant thereof. More preferably, a podocyte-specific promoter is a NPHS1 promoter, or a variant thereof.

20 In some embodiments, the promoter upstream of the 5' CDS is a minimal kidney-specific promoter. In some embodiments, the promoter upstream of the 3' CDS is a minimal kidney-specific promoter. In some embodiments, the promoter upstream of the 5' CDS and the promoter upstream of the 3' CDS are each a minimal kidney-specific promoter. As used, herein, a "minimal kidney-specific promoter" means the minimal sequence that can act as a  
 25 kidney-specific promoter.

In some embodiments, the promoter upstream of the 5' CDS is a minimal glomerular-specific promoter. In some embodiments, the promoter upstream of the 3' CDS is a minimal glomerular-specific promoter. In some embodiments, the promoter upstream of the 5' CDS and the promoter upstream of the 3' CDS are each a minimal glomerular-specific promoter.

30 In some embodiments, the promoter upstream of the 5' CDS is a minimal podocyte-specific promoter. In some embodiments, the promoter upstream of the 3' CDS is a minimal podocyte-specific promoter. In some embodiments, the promoter upstream of the 5' CDS and the promoter upstream of the 3' CDS are each a minimal podocyte-specific promoter.

In some embodiments, the promoter upstream of the 5' CDS is a minimal NPHS1 or a minimal NPHS2 promoter, or a variant thereof. In some embodiments, the promoter upstream of the 3' CDS is a minimal NPHS1 or a minimal NPHS2 promoter, or a variant thereof. In some embodiments, the promoter upstream of the 5' CDS and the promoter upstream of the 3' CDS are each a minimal NPHS1 or a minimal NPHS2 promoter, or a variant thereof.

In some embodiments, the promoter upstream of the 5' CDS is a minimal NPHS1 promoter, or a variant thereof. In some embodiments, the promoter upstream of the 3' CDS is a minimal NPHS1 promoter, or a variant thereof. In some embodiments, the promoter upstream of the 5' CDS and the promoter upstream of the 3' CDS are each a minimal NPHS1 promoter, or a variant thereof.

In some embodiments, the promoter upstream of the 5' CDS is a minimal human NPHS1 or a minimal human NPHS2 promoter. In some embodiments, the promoter upstream of the 3' CDS is a minimal human NPHS1 or a minimal human NPHS2 promoter. In some embodiments, the promoter upstream of the 5' CDS and the promoter upstream of the 3' CDS are each a minimal human NPHS1 or a minimal human NPHS2 promoter.

In some embodiments, the promoter upstream of the 5' CDS is a minimal human NPHS1 promoter. In some embodiments, the promoter upstream of the 3' CDS is a minimal human NPHS1 promoter. In some embodiments, the promoter upstream of the 5' CDS and the promoter upstream of the 3' CDS are each a minimal human NPHS1 promoter.

#### 20 *NPHS1 promoter*

In some embodiments, the promoter is a NPHS1 promoter, or a variant thereof. In some embodiments, the promoter upstream of the 5' CDS is a NPHS1 promoter, or a variant thereof. In some embodiments, the promoter upstream of the 3' CDS is a NPHS1 promoter, or a variant thereof. In some embodiments, the promoter upstream of the 5' CDS and the promoter upstream of the 3' CDS are each a NPHS1 promoter, or a variant thereof.

The NPHS1 gene encodes nephrin, which is selectively expressed in podocytes. A human NPHS1 promoter has been described in Moeller et al. 2002 J Am Soc Nephrol, 13(6):1561–7 and Wong MA et al. 2000 Am J Physiol Renal Physiol, 279(6):F1027-32. This NPHS1 promoter is a 1.2kb fragment and appears to be podocyte-specific. The 1.2kb promoter region lacks a TATA box, but has recognition motifs for other transcription factors e.g. PAX-2 binding element, E-box and GATA consensus sequences.

Suitably, the NPHS1 promoter may comprise or consist of the nucleotide sequence shown as SEQ ID NO: 226 (example NPHS1 promoter), or a variant which is at least 70% identical to SEQ ID NO: 226. Suitably, the variant may be at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%,  
 5 at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 226.

Suitably, the NPHS1 promoter may comprise or consist of the nucleotide sequence shown as SEQ ID NO: 311 (example NPHS1 promoter), or a variant which is at least 70% identical to SEQ ID NO: 311. Suitably, the variant may be at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%,  
 10 at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 311.

The NPHS1 promoter may be a minimal NPHS1 promoter. Suitably, the NPHS1 promoter has a length of about 1.1 kb or less, about 1.0 kb or less, about 0.9 kb or less, about 0.8 kb or less, about 0.7 kb or less, about 0.6 kb or less, about 0.5 kb or less, about 0.4 kb or less, or about 0.3 kb or less. Suitably, the NPHS1 promoter has a length of about 265 bp or more. In some  
 15 embodiments, the NPHS1 promoter has a length of about 265-1100 bp, 265-1000 bp, 265-900 bp, 265-800 bp, 265-700 bp, 265-600 bp, 265-500 bp, 265-400 bp, or 265-300 bp.

In some embodiments, the NPHS1 promoter may comprise or consist of the nucleotide sequence shown as SEQ ID NO: 227 (example minimal NPHS1 promoter 1), or a variant which is at least 70% identical to SEQ ID NO: 227.  
 20 Suitably, the variant may be at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 227.

The NPHS1 promoter may comprise or consist of a variant of SEQ ID NO: 227 shown as SEQ ID NO: 228 or SEQ ID NO: 229.

25 In some embodiments, the NPHS1 promoter may comprise or consist of the nucleotide sequence shown as SEQ ID NO: 228 or 229, or a variant which is at least 70% identical to SEQ ID NO: 228 or 229. Suitably, the variant may be at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 228 or 229.

30 In some embodiments, the NPHS1 promoter is the NPHS1 (nephrin) 265bp promoter:

GGCCCTGGGGTCACGGAGGCTGGGGAGGCACCGAGGAACGCGCCTGGCATGTGCTG  
 ACAGGGGATTTTATGCTCCAGGAGCAAGACAGAGAGATACTCACAGGGAAGAGGG

GAAGAGGAAAACGAGAAAGGGAGGAGAGTAACGGAAAGAGATAAAAAAGAAAAGCAG  
 GTGGCAGAGACACACAGAGAGGGACCCAGAGAAAGCCAGACAGACGCAGGTGGCTGG  
 CAGCGGGCGCTGTGGGGGTCACAGTAGGGGGACCTGTG (SEQ ID NO: 316)

## 5 *NPHS2 promoter*

In some embodiments, the promoter is a NPHS2 promoter, or a variant thereof. In some embodiments, the promoter upstream of the 5' CDS is a NPHS2 promoter, or a variant thereof. In some embodiments, the promoter upstream of the 3' CDS is a NPHS2 promoter, or a variant thereof. In some embodiments, the promoter upstream of the 5' CDS and the promoter upstream of the 3' CDS are each a NPHS2 promoter, or a variant thereof.

The NPHS2 gene encodes podocin, which is selectively expressed in podocytes. A human NPHS2 promoter has been described in Oleggini R, et al., 2006. Gene Expr. 13(1):59–66.

Suitably, the NPHS2 promoter may comprise or consist of the nucleotide sequence shown as SEQ ID NO: 230 (example NPHS2 promoter), or a variant which is at least 70% identical to SEQ ID NO: 230. Suitably, the variant may be at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 230.

The NPHS2 promoter may be a minimal NPHS2 promoter. Suitably, the NPHS2 promoter has a length of about 2.0 kb or less, about 1.8 kb or less, about 1.6 kb or less, about 1.4 kb or less, about 1.2 kb or less, about 1.0 kb or less, about 0.9 kb or less, about 0.8 kb or less, or about 0.7 kb or less. Suitably, the NPHS2 promoter has a length of about 628 bp or more. In some embodiments, the NPHS2 promoter has a length of about 628-2000 bp, 628-1800 bp, 628-1600 bp, 628-1400 bp, 628-1200 bp, 628-1000 bp, 628-900 bp, 628-800 bp, or 628-700 bp.

In some embodiments, the NPHS2 promoter may comprise or consist of the nucleotide sequence shown as SEQ ID NO: 231 (example minimal NPHS2 promoter), or a variant which is at least 70% identical to SEQ ID NO: 231.

Suitably, the variant may be at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 231.

## 30 *Hybrid enhancer-promoters*

In some embodiments, the promoter is a hybrid enhancer-promoter. In some embodiments, the promoter upstream of the 5' CDS is a hybrid enhancer-promoter. In some embodiments, the promoter upstream of the 3' CDS is a hybrid enhancer-promoter. In some embodiments, the promoter upstream of the 5' CDS and the promoter upstream of the 3' CDS are each a hybrid enhancer-promoter.

In some embodiments, the hybrid enhancer-promoter is a hybrid kidney-specific promoter comprising a kidney-specific enhancer and a ubiquitous or core promoter. In some embodiments, the hybrid enhancer-promoter is a hybrid glomerular-specific promoter comprising a glomerular-specific enhancer and a ubiquitous or core promoter. In some embodiments, the hybrid enhancer-promoter is a hybrid podocyte-specific promoter comprising a podocyte-specific enhancer and a ubiquitous or core promoter.

Suitably, the enhancer may be or may be derived from an enhancer associated with a gene with selective expression in human podocytes. Methods to identify the enhancer regions associated with genes will be well known to those of skill in the art. Preferably, the enhancer is a NPHS1 or a NPHS2 enhancer, or a fragment and/or variant thereof. Preferably, the enhancer is a human enhancer, e.g. a human NPHS1 enhancer or human NPHS2 enhancer.

A NPHS1 enhancer has been described in Guo, G., et al., 2004. Journal of the American Society of Nephrology, 15(11), pp.2851-2856. A 186-bp fragment from the human NPHS1 promoter was capable of directing podocyte-specific expression of a  $\beta$ -galactosidase transgene when placed in front of a heterologous minimal promoter in transgenic mice.

Suitably, a NPHS1 enhancer may comprise or consist of the nucleotide sequence shown as SEQ ID NO: 232 (example NPHS1 enhancer), or a variant which is at least 70% identical to SEQ ID NO: 232.

Suitably, the variant may be at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 96%, at least 98%, or at least 99% identical to SEQ ID NO: 232.

A NPHS2 enhancer has been described in WO 2023/213738. The NPHS2 enhancer comprises NPHS2 motif Lmx1b-FoxC2 and allowed the selective expression of genes of interest in podocytes and other kidney cell lines.

Suitably, a NPHS2 enhancer may comprise or consist of the nucleotide sequence shown as SEQ ID NO: 233 (example NPHS2 enhancer), or a variant which is at least 70% identical to SEQ ID NO: 233.

Suitably, the variant may be at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 96%, at least 98%, or at least 99% identical to SEQ ID NO: 233.

5 The enhancer may be upstream of a ubiquitous promoter or core promoter. The enhancer and the ubiquitous promoter or core promoter may be operably linked. In some embodiments, the enhancer is upstream of a core promoter. In some embodiments, the enhancer is upstream of a super core promoter.

10 The "core promoter" is typically 80 nucleotides long, encompassing from -40 to +40 relative to the transcription start site and consists of several functional sub-regions, termed core elements or motifs.

15 A "super core promoter" may refer to a synthetic core promoter which contains a combination of core promoter elements that drive high levels of transcription. For example, a super core promoter may contain a TATA box, an initiator motif (Inr), a motif ten element (MTE) and a downstream promoter element (DPE). Example super core promoters have been described in Juven-Gershon, T., et al., 2006. Nature methods, 3(11), pp.917-922 and Even, D.Y., et al., 2016. PloS one, 11(2), p.e0148918.

20 A super core promoter may comprise from 5' to 3': a TATA box, an initiator motif (Inr), a motif ten element (MTE), and a downstream promoter element (DPE). The super core promoter may comprise one or more further core promoter elements, such as transcription binding sites (e.g. TFIIB recognition element).

Suitably, the super core promoter is selected from super core promoter 1 (SCP1), super core promoter 2 (SCP2), or super core promoter 3 (SCP3). In some embodiments, the super core promoter is SCP1.

25 In some embodiments, the super core promoter comprises or consists of a nucleotide sequence which is at least 70% identical to SEQ ID NO: 234. Suitably, the super core promoter comprises or consists of a nucleotide sequence which is at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 234. In some embodiments, the super core promoter comprises or consists of the nucleotide sequence SEQ ID NO: 234 (example SCP1).

30

In some embodiments, the hybrid enhancer-promoter comprises or consists of a nucleotide sequence which is at least 70% identical to SEQ ID NO: 235. Suitably, the hybrid enhancer-promoter comprises or consists of a nucleotide sequence which is at least 75%, at least 80%,

at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 235. In some embodiments, the hybrid enhancer-promoter comprises or consists of the nucleotide sequence SEQ ID NO: 235 (example kidney-specific enhancer-promoter).

5 *Constitutive, inducible, and repressible promoters*

The promoter may be a constitutive promoter, an inducible promoter, or a repressible promoter.

10 In some embodiments, the promoter is a constitutive promoter. In some embodiments, the promoter upstream of the 5' CDS is a constitutive promoter and/or the promoter upstream of the 3' CDS is a constitutive promoter. As used herein, a "constitutive promoter" is a promoter which is always active. Suitable constitutive promoters will be known to the skilled person. Example constitutive promoters include the CMV promoter, the EF1a promoter, the CAG promoter, the PGK promoter, the U6 promoter, the T7 promoter, the SV40 promoter, and the Sp6 promoter.

15 In other embodiments, the promoter is an inducible promoter. In some embodiments, the promoter upstream of the 5' CDS is an inducible promoter and/or the promoter upstream of the 3' CDS is an inducible promoter. An "inducible promoter" may refer to a promoter which is activated in response to specific stimuli (e.g. in response to chemicals, temperature, or light). Suitable inducible promoters will be known to the skilled person.

20 In other embodiments, the promoter is a repressible promoter. In some embodiments, the promoter upstream of the 5' CDS is a repressible promoter and/or the promoter upstream of the 3' CDS is a repressible promoter. A "repressible promoter" may refer to a promoter which is de-activated in response to specific stimuli. Suitable repressible promoters will be known to the skilled person.

25 Polyadenylation sequences

The 3' AAV vector of the present invention may comprise a polyadenylation sequence. Suitably, the polyadenylation sequence is downstream of the 3' CDS. Suitably, the polyadenylation sequence may be operably linked to the 3' CDS.

30 In some embodiments (e.g. in mRNA trans-splicing method), the 5' AAV vector of the present invention comprises a polyadenylation sequence. Suitably, the polyadenylation sequence is downstream of the 5' CDS. Suitably, the polyadenylation sequence may be operably linked to the 5' CDS.



A “polyadenylation” is a region of DNA or mRNA that comprises the elements required for polyadenylation of the mRNA. The sequence elements for polyadenylation typically include a polyadenylation signal (with consensus sequence AATAAA) and a polyadenylation site (consensus sequence CA) and may also include a GT-rich downstream element.

- 5 Suitable polyadenylation sequences are well-known in the art and may include a bovine growth hormone polyadenylation sequence (bGH), a soluble neuropilin-1 polyadenylation sequence, an early SV40 polyadenylation sequence (SV40pA), and a chicken beta-globin polyadenylation sequence.

- 10 Suitably, the polyadenylation sequence may comprise or consist of the nucleotide sequence shown as SEQ ID NO: 236 (an example bGH poly(A) sequence), or a variant which is at least 70% identical to SEQ ID NO: 236. Suitably, the variant may be at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 96%, at least 98%, or at least 99% identical to SEQ ID NO: 236.

- 15 Suitably, the polyadenylation sequence may comprise or consist of the nucleotide sequence shown as SEQ ID NO: 237 (an example soluble neuropilin-1 poly (A) sequence), or a variant which is at least 70% identical to SEQ ID NO: 237. Suitably, the variant may be at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 96%, at least 98%, or at least 99% identical to SEQ ID NO: 237.

- 20 Suitably, the polyadenylation sequence may comprise or consist of the nucleotide sequence shown as SEQ ID NO: 238 (an example early SV40 poly (A) sequence), or a variant which is at least 70% identical to SEQ ID NO: 238. Suitably, the variant may be at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 96%, at least 98%, or at least 99% identical to SEQ ID NO: 238.

- 25 Suitably, the polyadenylation sequence may comprise or consist of the nucleotide sequence shown as SEQ ID NO: 239 (an example chicken beta-globin poly (A) sequence), or a variant which is at least 70% identical to SEQ ID NO: 239. Suitably, the variant may be at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 96%, at least 98%, or at least 99% identical to SEQ ID NO: 239.

#### Splice donor and acceptor sequences

##### *Splice donor sequences*

In some embodiments (e.g. in trans-splicing methods), the 5' AAV vector of the present invention comprises a splice donor sequence. Suitably, the splice donor sequence is downstream of the 5' CDS. Suitably, the splice donor sequence may be immediately downstream of the 5' CDS, i.e. without any gap between the 5' CDS and splice donor sequence.

Any suitable splice donor sequence may be used. Suitably, the splice donor sequence comprises a splice donor site at its 5' end. A splice donor site may include an almost invariant sequence GT at the 5' end. Suitably, the splice donor sequence comprises or consists of the nucleotide sequence GTNNNN(N)<sub>x</sub>, wherein x is from about 10 to about 1000 (e.g. from about 10 to about 500, from about 10 to about 200, or from about 10 to about 100).

Suitably, a consensus splice donor site comprises or consists of the nucleotide sequence GTRAGT (see e.g. Gao, K., et al., 2008. Nucleic acids research, 36(7), pp.2257-2267). Suitably, the splice donor sequence comprises or consists of the nucleotide sequence GTRAGT(N)<sub>x</sub>, wherein x is from about 10 to about 1000 (e.g. from about 10 to about 500, from about 10 to about 200, or from about 10 to about 100).

The splice donor sequence may be a synthetic or exogenous splice donor sequence. An example synthetic or exogenous splice donor sequence is provided in SEQ ID NO: 240.

In some embodiments, the splice donor sequence comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 240. In some embodiments, the splice donor sequence comprises or consists of the nucleotide sequence of SEQ ID NO: 240.

The splice donor sequence may be an endogenous splice donor sequence, for example from a COL4A3, COL4A4, or COL4A5 gene.

In some embodiments, the endogenous splice donor sequence corresponds to a 5' fragment intronic region between the exons present in the 5' CDS and the 3' CDS. For example, if the 5' CDS comprises exons 1-32 and the 3' CDS comprises exons 33-51 of a CDS encoding a COL4A5 polypeptide, then the splice donor sequence may comprise a 5' fragment of intron 32 of the CDS encoding a COL4A5 polypeptide. An example endogenous splice donor sequence is provided in SEQ ID NO: 241.

In some embodiments, the splice donor sequence comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at

least 98%, at least 99%, or 100% identity to SEQ ID NO: 241. In some embodiments, the splice donor sequence comprises or consists of the nucleotide sequence of SEQ ID NO: 241.

In some embodiments (e.g. in mRNA trans-splicing methods), the 5' AAV vector of the present invention comprises a forward splice donor sequence. Suitably, the forward splice donor sequence is downstream of the 5' CDS. Suitably, the forward splice donor sequence may be immediately downstream of the 5' CDS, i.e. without any gap between the 5' CDS and forward splice donor sequence.

As used herein, a "forward" splice donor sequence may refer to a splice donor sequence which functions as a splice donor site when transcribed from the sense strand of an AAV vector. As used herein, a "reverse" splice donor sequence may refer to a splice donor sequence which functions as a splice donor site when transcribed from the anti-sense strand of an AAV vector.

Any suitable forward splice donor sequence may be used, for example SEQ ID NO: 240 or a variant thereof, or SEQ ID NO: 241 or a variant thereof.

#### *Splice acceptor sequences*

In some embodiments (e.g. in trans-splicing methods), the 3' AAV vector of the present invention comprises a splice acceptor sequence. Suitably, the splice acceptor sequence is upstream of the 3' CDS. Suitably, the splice acceptor sequence may be immediately upstream of the 3' CDS, i.e. without any gap between the 3' CDS and splice acceptor sequence.

Any suitable splice acceptor sequence may be used. Suitably, the splice acceptor sequence comprises a splice acceptor site at its 3' end. A splice acceptor site may include an almost invariant sequence AG at the 3' end. Suitably, the splice acceptor sequence comprises or consists of the nucleotide sequence (N)<sub>x</sub>NAG, wherein x is from about 10 to about 1000 (e.g. from about 10 to about 500, from about 10 to about 200, or from about 10 to about 100).

Suitably, a consensus splice acceptor site comprises or consists of the nucleotide sequence YAG. Suitably, the splice acceptor sequence comprises or consists of the nucleotide sequence (N)<sub>x</sub>YAG, wherein x is from about 10 to about 1000 (e.g. from about 10 to about 500, from about 10 to about 200, or from about 10 to about 100).

The splice acceptor sequence may be a synthetic or exogenous splice donor sequence. An example synthetic or exogenous splice acceptor sequence is provided in SEQ ID NO: 242.

In some embodiments, the splice acceptor sequence comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least

91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 242. In some embodiments, the splice acceptor sequence comprises or consists of the nucleotide sequence of SEQ ID NO: 242.

- 5 The splice acceptor sequence may be an endogenous splice acceptor sequence, for example from a COL4A3, COL4A4, or COL4A5 gene.

In some embodiments, the endogenous splice acceptor sequence corresponds to a 3' fragment intronic region between the exons present in the 5' CDS and the 3' CDS. For example, if the 5' CDS comprises exons 1-32 and the 3' CDS comprises exons 33-51 of a  
10 CDS encoding a COL4A5 polypeptide, then the splice acceptor sequence may comprise a 3' fragment of intron 32 of the CDS encoding a COL4A5 polypeptide. An example endogenous splice acceptor sequence is provided in SEQ ID NO: 243.

In some embodiments, the splice acceptor sequence comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least  
15 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 243. In some embodiments, the splice acceptor sequence comprises or consists of the nucleotide sequence of SEQ ID NO: 243.

In some embodiments (e.g. in mRNA trans-splicing methods), the 3' AAV vector of the present  
20 invention comprises a reverse splice acceptor sequence. Suitably, the reverse splice acceptor sequence is upstream of the 3' CDS. Suitably, the reverse splice acceptor sequence may be immediately upstream of the 3' CDS, i.e. without any gap between the 3' CDS and reverse splice acceptor sequence.

As used herein, a "forward" splice acceptor sequence may refer to a splice acceptor sequence  
25 which functions as a splice acceptor site when transcribed from the sense strand of an AAV vector. As used herein, a "reverse" splice acceptor sequence may refer to a splice acceptor sequence which functions as a splice acceptor site when transcribed from the anti-sense strand of an AAV vector.

Any suitable reverse splice acceptor sequence may be used. An example reverse synthetic  
30 or exogenous splice acceptor sequence is provided in SEQ ID NO: 272, which is based on SEQ ID NO: 242.

In some embodiments, the reverse splice acceptor sequence comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least

90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 272. In some embodiments, the reverse splice acceptor sequence comprises or consists of the nucleotide sequence of SEQ ID NO: 272.

- 5 An example reverse endogenous splice acceptor sequence is provided in SEQ ID NO: 273, which is based on SEQ ID NO: 243.

In some embodiments, the reverse splice acceptor sequence comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 273. In some  
10 embodiments, the reverse splice acceptor sequence comprises or consists of the nucleotide sequence of SEQ ID NO: 273.

*Example combinations of splice donor and acceptor sequences*

In some embodiments, the 5' AAV vector comprises a splice donor sequence downstream  
15 from the 5' CDS having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 240; and the 3' AAV vector comprises a splice acceptor sequence upstream from the 3' CDS having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at  
20 least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 242.

In some embodiments, the 5' AAV vector comprises a splice donor sequence downstream from the 5' CDS having at least 70% identity to SEQ ID NO: 240 and the 3' AAV vector comprises a splice acceptor sequence upstream from the 3' CDS having at least 70% identity  
25 to SEQ ID NO: 242.

In some embodiments, the 5' AAV vector comprises a splice donor sequence downstream from the 5' CDS having at least 80% identity to SEQ ID NO: 240 and the 3' AAV vector comprises a splice acceptor sequence upstream from the 3' CDS having at least 80% identity to SEQ ID NO: 242.

30 In some embodiments, the 5' AAV vector comprises a splice donor sequence downstream from the 5' CDS having at least 85% identity to SEQ ID NO: 240 and the 3' AAV vector comprises a splice acceptor sequence upstream from the 3' CDS having at least 85% identity to SEQ ID NO: 242.

In some embodiments, the 5' AAV vector comprises a splice donor sequence downstream from the 5' CDS having at least 90% identity to SEQ ID NO: 240 and the 3' AAV vector comprises a splice acceptor sequence upstream from the 3' CDS having at least 90% identity to SEQ ID NO: 242.

- 5 In some embodiments, the 5' AAV vector comprises a splice donor sequence downstream from the 5' CDS having at least 95% identity to SEQ ID NO: 240 and the 3' AAV vector comprises a splice acceptor sequence upstream from the 3' CDS having at least 95% identity to SEQ ID NO: 242.

- 10 In some embodiments, the 5' AAV vector comprises a splice donor sequence downstream from the 5' CDS having at least 96% identity to SEQ ID NO: 240 and the 3' AAV vector comprises a splice acceptor sequence upstream from the 3' CDS having at least 96% identity to SEQ ID NO: 242.

- 15 In some embodiments, the 5' AAV vector comprises a splice donor sequence downstream from the 5' CDS having at least 98% identity to SEQ ID NO: 240 and the 3' AAV vector comprises a splice acceptor sequence upstream from the 3' CDS having at least 98% identity to SEQ ID NO: 242.

- 20 In some embodiments, the 5' AAV vector comprises a splice donor sequence downstream from the 5' CDS comprising or consisting of SEQ ID NO: 240 and the 3' AAV vector comprises a splice acceptor sequence upstream from the 3' CDS comprising or consisting of SEQ ID NO: 242.

- 25 In some embodiments, the 5' AAV vector comprises a splice donor sequence downstream from the 5' CDS having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 241; and the 3' AAV vector comprises a splice acceptor sequence upstream from the 3' CDS having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 243.

- 30 In some embodiments, the 5' AAV vector comprises a splice donor sequence downstream from the 5' CDS having at least 70% identity to SEQ ID NO: 241 and the 3' AAV vector comprises a splice acceptor sequence upstream from the 3' CDS having at least 70% identity to SEQ ID NO: 243.

In some embodiments, the 5' AAV vector comprises a splice donor sequence downstream from the 5' CDS having at least 80% identity to SEQ ID NO: 241 and the 3' AAV vector comprises a splice acceptor sequence upstream from the 3' CDS having at least 80% identity to SEQ ID NO: 243.

- 5 In some embodiments, the 5' AAV vector comprises a splice donor sequence downstream from the 5' CDS having at least 85% identity to SEQ ID NO: 241 and the 3' AAV vector comprises a splice acceptor sequence upstream from the 3' CDS having at least 85% identity to SEQ ID NO: 243.

- 10 In some embodiments, the 5' AAV vector comprises a splice donor sequence downstream from the 5' CDS having at least 90% identity to SEQ ID NO: 241 and the 3' AAV vector comprises a splice acceptor sequence upstream from the 3' CDS having at least 90% identity to SEQ ID NO: 243.

- 15 In some embodiments, the 5' AAV vector comprises a splice donor sequence downstream from the 5' CDS having at least 95% identity to SEQ ID NO: 241 and the 3' AAV vector comprises a splice acceptor sequence upstream from the 3' CDS having at least 95% identity to SEQ ID NO: 243.

- 20 In some embodiments, the 5' AAV vector comprises a splice donor sequence downstream from the 5' CDS having at least 96% identity to SEQ ID NO: 241 and the 3' AAV vector comprises a splice acceptor sequence upstream from the 3' CDS having at least 96% identity to SEQ ID NO: 243.

- In some embodiments, the 5' AAV vector comprises a splice donor sequence downstream from the 5' CDS having at least 97% identity to SEQ ID NO: 241 and the 3' AAV vector comprises a splice acceptor sequence upstream from the 3' CDS having at least 97% identity to SEQ ID NO: 243.

- 25 In some embodiments, the 5' AAV vector comprises a splice donor sequence downstream from the 5' CDS having at least 98% identity to SEQ ID NO: 241 and the 3' AAV vector comprises a splice acceptor sequence upstream from the 3' CDS having at least 98% identity to SEQ ID NO: 243.

- 30 In some embodiments, the 5' AAV vector comprises a splice donor sequence downstream from the 5' CDS having at least 99% identity to SEQ ID NO: 241 and the 3' AAV vector comprises a splice acceptor sequence upstream from the 3' CDS having at least 99% identity to SEQ ID NO: 243.

In some embodiments, the 5' AAV vector comprises a splice donor sequence downstream from the 5' CDS comprising or consisting of SEQ ID NO: 241 and the 3' AAV vector comprises a splice acceptor sequence upstream from the 3' CDS comprising or consisting of SEQ ID NO: 243.

5 Recombinogenic sequences

In some embodiments (e.g. in DNA hybrid methods), the 5' AAV vector of the present invention and the 3' AAV vector of the present invention comprises a recombinogenic sequence.

As used herein, a "recombinogenic sequence" may refer to a sequence which can mediate target gene-independent homologous recombination between the 5' AAV vector and the 3' AAV vector. Suitable recombinogenic sequences are known in the art (see e.g. Trapani, I., 2018. Retinal Gene Therapy: Methods and Protocols, pp.153-175; Trapani, I., et al., 2014. EMBO molecular medicine, 6(2), pp.194-21; Lostal, W., et al., 2014. Human gene therapy, 25(6), pp.552-562; and Ghosh, A., et al., 2011. Human gene therapy, 22(1), pp.77-83) and include, for example, recombinogenic regions from the F1 phage sequence or the alkaline phosphate gene.

Suitably, in the 5' AAV vector, the recombinogenic sequence is downstream of a splice donor sequence. Suitably, the recombinogenic sequence may be immediately downstream of the splice donor sequence, i.e. without any gap between the splice donor sequence and recombinogenic sequence.

20 Suitably, in the 3' AAV vector, the recombinogenic sequence is upstream of a splice acceptor sequence. Suitably, the recombinogenic sequence may be immediately upstream of the splice acceptor sequence, i.e. without any gap between the splice acceptor sequence and recombinogenic sequence.

25 The recombinogenic sequences comprised in the 5' AAV vector and the 3' AAV vector may have a degree of identity such that they can mediate target gene-independent homologous recombination between the 5' AAV vector and the 3' AAV vector. Suitably, the recombinogenic sequences comprised in the 5' AAV vector and the 3' AAV vector are at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical. In some embodiments, the recombinogenic sequences comprised in the 5' AAV vector and the 3' AAV vector are identical.

*Example recombinogenic sequences*



In some embodiments, the recombinogenic sequence is a recombinogenic region from the F1 phage sequence or a recombinogenic region from the alkaline phosphate gene.

In some embodiments, the recombinogenic sequence is a recombinogenic region from the F1 phage sequence (see e.g. Trapani, I., 2018. Retinal Gene Therapy: Methods and Protocols, pp.153-175). An example recombinogenic region from the F1 phage sequence is provided in SEQ ID NO: 244.

In some embodiments, the recombinogenic sequence comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 244. In some embodiments, the recombinogenic sequence comprises or consists of the nucleotide sequence of SEQ ID NO: 244.

In some embodiments, the recombinogenic sequence is a recombinogenic region from the alkaline phosphate gene (see e.g. Ghosh, A., et al., 2011. Human gene therapy, 22(1), pp.77-83). Example recombinogenic regions from the alkaline phosphate gene are provided in SEQ ID NOs: 245 to 247.

In some embodiments, the recombinogenic sequence comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 245. In some embodiments, the recombinogenic sequence comprises or consists of the nucleotide sequence of SEQ ID NO: 245.

In some embodiments, the recombinogenic sequence comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 246. In some embodiments, the recombinogenic sequence comprises or consists of the nucleotide sequence of SEQ ID NO: 246.

In some embodiments, the recombinogenic sequence comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 247. In some embodiments, the

recombinogenic sequence comprises or consists of the nucleotide sequence of SEQ ID NO: 247.

*Example 5' intron fragments*

5 The splice donor sequence and recombinogenic sequence may together be referred to as a "5' intron fragment". In some embodiments, the 5' AAV vector comprises a 5' intron fragment downstream of the 5' CDS.

A 5' intron fragment may comprise any suitable splice donor sequence and any suitable recombinogenic sequence described herein. Suitably, a 5' intron fragment comprises from 5' to 3': a splice donor sequence and a recombinogenic sequence.

10 In some embodiments, a 5' intron fragment comprises SEQ ID NO: 240, or a variant thereof, and any of SEQ ID NOs: 244 to 247, or variants thereof. In some embodiments, a 5' intron fragment comprises SEQ ID NO: 240, or a variant thereof, and SEQ ID NO: 244, or a variant thereof. An example 5' intron fragment is provided by SEQ ID NO: 248.

15 In some embodiments, a 5' intron fragment comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 248. In some embodiments, a 5' intron fragment comprises or consists of the nucleotide sequence of SEQ ID NO: 248.

20 In some embodiments, a 5' intron fragment comprises SEQ ID NO: 241, or a variant thereof, and any of SEQ ID NOs: 244 to 247, or variants thereof. In some embodiments, a 5' intron fragment comprises SEQ ID NO: 241, or a variant thereof, and SEQ ID NO: 244, or a variant thereof. An example 5' intron fragment is provided by SEQ ID NO: 249.

25 In some embodiments, a 5' intron fragment comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 249. In some embodiments, a 5' intron fragment comprises or consists of the nucleotide sequence of SEQ ID NO: 249.

*Example 3' intron fragments*

30 The splice acceptor sequence and recombinogenic sequence may together be referred to as a "3' intron fragment". In some embodiments, the 3' AAV vector comprises a 3' intron fragment upstream of the 3' CDS.

A 3' intron fragment may comprise any suitable splice acceptor sequence and any suitable recombinogenic sequence described herein. Suitably, a 3' intron fragment comprises from 5' to 3': a recombinogenic sequence and a splice acceptor sequence.

5 In some embodiments, a 3' intron fragment comprises SEQ ID NO: 242, or a variant thereof, and any of SEQ ID NOs: 244 to 247, or variants thereof. In some embodiments, a 3' intron fragment comprises SEQ ID NO: 242, or a variant thereof, and SEQ ID NO: 244, or a variant thereof. An example 3' intron fragment is provided by SEQ ID NO: 250.

10 In some embodiments, a 3' intron fragment comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 250. In some embodiments, a 3' intron fragment comprises or consists of the nucleotide sequence of SEQ ID NO: 250.

15 In some embodiments, a 3' intron fragment comprises SEQ ID NO: 243, or a variant thereof, and any of SEQ ID NOs: 244 to 247, or variants thereof. In some embodiments, a 3' intron fragment comprises SEQ ID NO: 243, or a variant thereof, and SEQ ID NO: 244, or a variant thereof. An example 3' intron fragment is provided by SEQ ID NO: 251.

20 In some embodiments, a 3' intron fragment comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 251. In some embodiments, a 3' intron fragment comprises or consists of the nucleotide sequence of SEQ ID NO: 251.

#### *Example combinations of intron fragments*

25 In some embodiments, the first AAV vector comprises a 5' intron fragment downstream from the 5' CDS having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 248; and the second AAV vector comprises a 3' intron fragment upstream from the 3' CDS having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 250.

In some embodiments, the 5' AAV vector comprises a 5' intron fragment downstream from the 5' CDS having at least 70% identity to SEQ ID NO: 248 and the 3' AAV vector comprises a 3' intron fragment upstream from the 3' CDS having at least 70% identity to SEQ ID NO: 250.

In some embodiments, the 5' AAV vector comprises a 5' intron fragment downstream from the 5' CDS having at least 80% identity to SEQ ID NO: 248 and the 3' AAV vector comprises a 3' intron fragment upstream from the 3' CDS having at least 80% identity to SEQ ID NO: 250.

- 5 In some embodiments, the 5' AAV vector comprises a 5' intron fragment downstream from the 5' CDS having at least 85% identity to SEQ ID NO: 248 and the 3' AAV vector comprises a 3' intron fragment upstream from the 3' CDS having at least 85% identity to SEQ ID NO: 250.

In some embodiments, the 5' AAV vector comprises a 5' intron fragment downstream from the 5' CDS having at least 90% identity to SEQ ID NO: 248 and the 3' AAV vector comprises a 3' intron fragment upstream from the 3' CDS having at least 90% identity to SEQ ID NO: 250.

- 10 In some embodiments, the 5' AAV vector comprises a 5' intron fragment downstream from the 5' CDS having at least 95% identity to SEQ ID NO: 248 and the 3' AAV vector comprises a 3' intron fragment upstream from the 3' CDS having at least 95% identity to SEQ ID NO: 250.

- 15 In some embodiments, the 5' AAV vector comprises a 5' intron fragment downstream from the 5' CDS having at least 96% identity to SEQ ID NO: 248 and the 3' AAV vector comprises a 3' intron fragment upstream from the 3' CDS having at least 96% identity to SEQ ID NO: 250.

In some embodiments, the 5' AAV vector comprises a 5' intron fragment downstream from the 5' CDS having at least 97% identity to SEQ ID NO: 248 and the 3' AAV vector comprises a 3' intron fragment upstream from the 3' CDS having at least 97% identity to SEQ ID NO: 250.

- 20 In some embodiments, the 5' AAV vector comprises a 5' intron fragment downstream from the 5' CDS having at least 98% identity to SEQ ID NO: 248 and the 3' AAV vector comprises a 3' intron fragment upstream from the 3' CDS having at least 98% identity to SEQ ID NO: 250.

In some embodiments, the 5' AAV vector comprises a 5' intron fragment downstream from the 5' CDS having at least 99% identity to SEQ ID NO: 248 and the 3' AAV vector comprises a 3' intron fragment upstream from the 3' CDS having at least 99% identity to SEQ ID NO: 250.

- 25 In some embodiments, the 5' AAV vector comprises a 5' intron fragment downstream from the 5' CDS comprising or consisting of SEQ ID NO: 248 and the 3' AAV vector comprises a 3' intron fragment upstream from the 3' CDS comprising or consisting of SEQ ID NO: 250.

- 30 In some embodiments, the first AAV vector comprises a 5' intron fragment downstream from the 5' CDS having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 249; and the second AAV vector

comprises a 3' intron fragment upstream from the 3' CDS having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 251.

- 5 In some embodiments, the 5' AAV vector comprises a 5' intron fragment downstream from the 5' CDS having at least 70% identity to SEQ ID NO: 249 and the 3' AAV vector comprises a 3' intron fragment upstream from the 3' CDS having at least 70% identity to SEQ ID NO: 251.

- In some embodiments, the 5' AAV vector comprises a 5' intron fragment downstream from the 5' CDS having at least 80% identity to SEQ ID NO: 249 and the 3' AAV vector comprises a 3' intron fragment upstream from the 3' CDS having at least 80% identity to SEQ ID NO: 251.
- 10

In some embodiments, the 5' AAV vector comprises a 5' intron fragment downstream from the 5' CDS having at least 85% identity to SEQ ID NO: 249 and the 3' AAV vector comprises a 3' intron fragment upstream from the 3' CDS having at least 85% identity to SEQ ID NO: 251.

- In some embodiments, the 5' AAV vector comprises a 5' intron fragment downstream from the 5' CDS having at least 90% identity to SEQ ID NO: 249 and the 3' AAV vector comprises a 3' intron fragment upstream from the 3' CDS having at least 90% identity to SEQ ID NO: 251.
- 15

In some embodiments, the 5' AAV vector comprises a 5' intron fragment downstream from the 5' CDS having at least 95% identity to SEQ ID NO: 249 and the 3' AAV vector comprises a 3' intron fragment upstream from the 3' CDS having at least 95% identity to SEQ ID NO: 251.

- In some embodiments, the 5' AAV vector comprises a 5' intron fragment downstream from the 5' CDS having at least 96% identity to SEQ ID NO: 249 and the 3' AAV vector comprises a 3' intron fragment upstream from the 3' CDS having at least 96% identity to SEQ ID NO: 251.
- 20

- In some embodiments, the 5' AAV vector comprises a 5' intron fragment downstream from the 5' CDS having at least 97% identity to SEQ ID NO: 249 and the 3' AAV vector comprises a 3' intron fragment upstream from the 3' CDS having at least 97% identity to SEQ ID NO: 251.
- 25

In some embodiments, the 5' AAV vector comprises a 5' intron fragment downstream from the 5' CDS having at least 98% identity to SEQ ID NO: 249 and the 3' AAV vector comprises a 3' intron fragment upstream from the 3' CDS having at least 98% identity to SEQ ID NO: 251.

- In some embodiments, the 5' AAV vector comprises a 5' intron fragment downstream from the 5' CDS having at least 99% identity to SEQ ID NO: 249 and the 3' AAV vector comprises a 3' intron fragment upstream from the 3' CDS having at least 99% identity to SEQ ID NO: 251.
- 30

In some embodiments, the 5' AAV vector comprises a 5' intron fragment downstream from the 5' CDS comprising or consisting of SEQ ID NO: 249 and the 3' AAV vector comprises a 3' intron fragment upstream from the 3' CDS comprising or consisting of SEQ ID NO: 251.

#### Hybridisation sequences

- 5 In some embodiments (e.g. in mRNA trans-splicing methods), the 5' AAV vector of the present invention and the 3' AAV vector of the present invention comprise complementary hybridisation sequences.

As used herein, a "hybridisation sequence" may refer to a sequence which can mediate mRNA-trans-splicing between transcripts from the 5' AAV vector and the 3' AAV vector.

- 10 Suitable hybridisation sequences are known in the art (see e.g. Riedmayr, L.M., et al., 2023. Nature Communications, 14(1), p.6578; and Song, Y., et al., 2009. Human gene therapy, 20(3), pp.267-281) and include, for example, hybridisation sequences derived from the intronic sequence of the human rhodopsin (RHO) and hybridisation sequences derived from the bacterial lacZ gene. In some embodiments, the hybridisation sequence is not found in the
- 15 human genome.

Suitably, in the 5' AAV vector, the hybridisation sequence is downstream of a splice donor sequence (e.g. a forward splice donor sequence). Suitably, the hybridisation sequence may be immediately downstream of the splice donor sequence, i.e. without any gap between the splice donor sequence and hybridisation sequence.

- 20 Suitably, in the 3' AAV vector, the complementary hybridisation sequence is upstream of a splice acceptor sequence (e.g. a reverse splice acceptor sequence). Suitably, the complementary hybridisation sequence may be immediately upstream of the splice acceptor sequence, i.e. without any gap between the splice acceptor sequence and complementary hybridisation sequence.

- 25 The hybridisation sequences comprised in the 5' AAV vector and the 3' AAV vector may have a degree of identity such that they can mediate mRNA-transplicing between transcripts from the 5' AAV vector and the 3' AAV vector. Suitably, the hybridisation sequence comprised in the 5' AAV vector and the complementary hybridisation sequence comprised in the 3' AAV vector are identical reverse complements.

- 30 Other regulatory sequences

The 5' AAV vector and the 3' AAV vector may comprise any other suitable regulatory sequences. The one or more regulatory sequences may be operably linked to the 5' CDS or

3' CDS. As used herein, the term "operably linked" may mean that the elements described are in a relationship permitting them to function in their intended manner.

A "regulatory element" may refer to any sequence which increases expression of a protein-coding sequence, e.g. which acts to increase expression of a transcript or to enhance mRNA stability. Suitably, the regulatory element may act post-transcriptionally (i.e. is a post-transcriptional regulatory element). Suitably, the regulatory element may enhance mRNA stability.

In some embodiments, the 5' AAV vector further comprises one or more regulatory sequences downstream of a promoter and upstream of the 5' CDS. In some embodiments, the 3' AAV vector further comprises one or more regulatory sequence downstream of the 3' CDS and upstream of a polyadenylation sequence. In some embodiments, the 5' AAV vector further comprises one or more regulatory sequence downstream of a promoter and upstream of the 5' CDS; and the 3' AAV vector further comprises one or more regulatory sequence downstream of the 3' CDS and upstream of a polyadenylation sequence.

In some embodiments (e.g. in DNA overlapping, trans-splicing or hybrid methods), the 3' AAV vector does not comprise one or more regulatory sequences upstream of the 3' CDS (e.g. other than a splice acceptor sequence). In some embodiments (e.g. in DNA overlapping, trans-splicing or hybrid methods), the 5' AAV vector does not comprise one or more regulatory sequence downstream of the 5' CDS (e.g. other than a splice donor sequence). In some embodiments (e.g. in DNA overlapping, trans-splicing or hybrid methods), the 3' AAV vector does not comprise one or more regulatory sequences upstream of the 3' CDS (e.g. other than a splice acceptor sequence); and the 5' AAV vector does not comprise one or more regulatory sequence downstream of the 5' CDS (e.g. other than a splice donor sequence).

In other embodiments (e.g. in mRNA trans-splicing methods), the 3' AAV vector further comprises one or more regulatory sequence downstream of a promoter and upstream of the 3' CDS. In some embodiments (e.g. in mRNA trans-splicing methods), the 5' AAV vector further comprises one or more regulatory sequence downstream of the 5' CDS and upstream of a polyadenylation sequence. In some embodiments (e.g. in mRNA trans-splicing methods), the 3' AAV vector further comprises one or more regulatory sequence downstream of a promoter and upstream of the 3' CDS; and the 5' AAV vector further comprises one or more regulatory sequence downstream of the 5' CDS and upstream of a polyadenylation sequence.

*Spliceosomal introns*

In some embodiments, the 5' AAV vector comprises a spliceosomal intron or fragment thereof. The spliceosomal intron or fragment thereof may be downstream of a promoter and upstream of the 5' CDS.

In some embodiments (e.g. mRNA trans-splicing methods), the 3' AAV vector comprises a spliceosomal intron or fragment thereof. The spliceosomal intron or fragment thereof may be downstream of a promoter and upstream of the 3' CDS.

A "spliceosomal intron" may refer to a non-coding sequence excised from pre-mRNAs by the spliceosome during mRNA splicing. Spliceosomal introns have been found in most eukaryotic genes and their lengths vary between species, from just tens of bases in some protists to hundreds of kilobases in some mammals (see e.g. Roy, S. W. and Gilbert, W., 2006. *Nature Reviews Genetics*, 7(3), pp.211-221).

The spliceosomal intron may be a naturally-occurring spliceosomal intron, a chimeric spliceosomal intron (e.g. with elements derived from two or more naturally-occurring spliceosomal introns), or a variant spliceosomal intron.

Examples of suitable naturally-occurring spliceosomal introns include the rabbit beta globin intron I, the rabbit beta globin intron II, the minute virus of mice (MVM) intron, and the F.IX truncated intron 1. Examples of suitable chimeric spliceosomal introns include  $\beta$ -globin SD / immunoglobulin heavy chain SA, Adenovirus SD / immunoglobulin SA, SV40 late SD / SA (19S/16S), and Hybrid adenovirus SD / IgG SA (see e.g. Powell, S.K., et al., 2015. *Discovery medicine*, 19(102), pp.49-57). Examples of suitable variant spliceosomal introns are described below.

A spliceosomal intron may be capable of increasing expression of a protein-coding sequence, in particular a protein-coding sequence downstream thereof. The spliceosomal intron may be operable in mammalian cells, e.g. human cells. The spliceosomal intron may be capable of increasing expression of a protein-coding sequence in mammalian cells, e.g. human cells.

The spliceosomal intron may be operable in kidney cells. The spliceosomal intron may be capable of increasing expression of a protein-coding sequence in the kidney. Examples of kidney cells include, but are not limited, to glomerular cells. The spliceosomal intron may be operable in glomerular cells. The spliceosomal intron may be capable of increasing expression of a protein-coding sequence in the glomerulus. The mature glomerulus contains four cell types: Parietal epithelial cells that form Bowman's capsule, podocytes that cover the outermost layer of the glomerular filtration barrier, glycocalyx-coated fenestrated endothelial cells that are in direct contact with blood, and mesangial cells that sit between the capillary loops (see



e.g. Vaughan, M.R. and Quaggin, S.E., 2008. Journal of the American Society of Nephrology, 19(1), pp.24-33). The spliceosomal intron may be operable in a podocyte cell. The spliceosomal intron may be capable of increasing expression of a protein-coding sequence in podocytes.

- 5 Expression of a protein-coding sequence may be measured by any suitable method known in the art. For example, by measuring the expression of a reporter transgene, e.g. Luciferase, placed downstream of the spliceosomal intron, wherein expression of the reporter transgene correlates with the ability of the spliceosomal intron to increase expression of a protein-coding sequence. Expression of the reporter transgene, e.g. Luciferase may be determined by any
- 10 suitable method. Suitable cell lines will be well known to those of skill in the art. For example, a suitable podocyte cell line is CIHP-1 and methods to generate immortalized podocytes will be well known to those of skill in the art (see e.g. Ni, L., et al., 2012. Nephrology, 17(6), pp.525-531).

- A spliceosomal intron typically contains the following conserved elements from 5' to 3': a 5'
- 15 splice donor site, also known as a 5' splice site; a branch point sequence, comprising the branch point; a polypyrimidine tract; and a 3' splice acceptor site, also known as a 3' splice site (see e.g. Padgett, R.A., et al., 1986. Annual review of biochemistry, 55(1), pp.1119-1150).

- In human spliceosomal introns, the human branch point consensus sequence is yUnAy, where the branch point 'A' is underlined. Branch points are typically located 21–34 nucleotides
- 20 upstream of the 3' end of an intron. A polypyrimidine stretch typically spans 4–24 nucleotides downstream of the branch point (see e.g. Gao, K., et al., 2008. Nucleic acids research, 36(7), pp.2257-2267).

- The spliceosomal intron may comprise each of the 5' splice donor site, the branch point sequence, the polypyrimidine tract, and the 3' splice acceptor site in consensus positions.
- 25 Suitably, the 5' splice donor site is from about 10 bp to about 1000 bp upstream of the branch point or from about 50 bp to about 100 bp upstream of the branch point. Suitably, the polypyrimidine tract is from about 4 bp to about 24 bp downstream of the branch point. Suitably, the 3' splice point is from about 21 bp to about 34 bp downstream of the branch point.

- The spliceosomal intron may comprise a consensus 5' splice donor site, a consensus branch
- 30 point sequence, a consensus polypyrimidine tract, and/or a consensus 3' splice acceptor site. In some embodiments, the spliceosomal intron comprises at least a consensus branch point sequence and a consensus 3' splice acceptor site. Suitably, a consensus 5' splice donor site comprises or consists of the nucleotide sequence GTRAGT. Suitably, a consensus branch point sequence comprises or consists of the nucleotide sequence YTNAY. In some

embodiments, the branch point sequence comprises or consists of the nucleotide sequence TGCTGAC. Suitably, a consensus polypyrimidine tract comprises or consists of (Y)<sub>n2</sub>, where n2 is from about 10 to about 20. Suitably, a consensus 3' splice acceptor site comprises or consists of the nucleotide sequence YAGG.

- 5 Suitably, the spliceosomal intron has a length of about 1000 bp or less, about 900 bp or less, about 800 bp or less, about 700 bp or less, about 600 bp or less, about 500 bp or less, about 400 bp or less, about 300 bp or less, about 200 bp or less, about 150 bp or less, about 140 bp or less, or about 130 bp or less.

- Suitably, the spliceosomal intron has a length of about 40 bp or more, about 50 bp or more, about 60 bp or more, about 70 bp or more, about 80 bp or more, about 90 bp or more, about 100 bp or more, about 100 bp or more, about 105 bp or more, about 110 bp or more, about 115 bp or more, or about 120 bp or more.
- 10

- Suitably, the spliceosomal intron may comprise or consist of the nucleotide sequence GTRAGT(N)<sub>x</sub>YTNAY(N)<sub>n1</sub>(Y)<sub>n2</sub>(N)<sub>n3</sub>YAGG (SEQ ID NO: 315), wherein x is from about 10 to about 1000 (e.g. from about 10 to about 100), n1 is from about 2 to about 22, n2 is from about 10 to about 20, and n1 + n2 + n3 is from about 15 to about 40, or a nucleotide sequence having five or fewer, four or fewer, three or fewer, two or fewer, or one nucleotide substitution.
- 15

A variant spliceosomal intron may be derived from any naturally-occurring or chimeric spliceosomal intron, referred to herein as the "wild-type spliceosomal intron".

- 20 Suitably, a variant spliceosomal intron may have 70% or more, 75% or more, 80% or more, 81% or more, 82% or more, 83% or more, 84% or more, 85% or more, 86% or more, 87% or more, 88% or more, 89% or more, 90% or more, 91% or more, 92% or more, 93% or more, 94% or more, 95% or more, 96% or more, 97% or more, 98% or more, 99% or more, or 100% identity to a wild-type spliceosomal intron.
- 25 A variant spliceosomal intron may be obtained by substituting, inserting, and/or deleting 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 nucleotide(s) into the wild-type spliceosomal intron. In some embodiments, a variant spliceosomal intron has 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 nucleotide substitution(s) compared to a wild-type spliceosomal intron.
- 30 Suitably, a variant spliceosomal intron retains or improves the wild-type spliceosomal introns ability to increase expression of a protein-coding sequence. Expression of a protein-coding sequence may be measured by any suitable method known in the art. For example by measuring the expression of a reporter transgene, e.g. Luciferase, placed downstream of the

spliceosomal intron, wherein expression of the reporter transgene correlates with the ability of the spliceosomal intron to increase expression of a protein-coding sequence. Suitably, the variant spliceosomal intron increases expression of a protein-coding sequence to at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 100% of the level of the wild-type spliceosomal intron.

Suitably, a wild-type spliceosomal intron may be modified to increase (or decrease) the consensus, to insert further nucleotide sequences, and/or to delete nucleotide sequences which are not required for the spliceosomal intron to increase expression of a protein-coding sequence. In some embodiments, the wild-type spliceosomal intron is modified to increase the consensus.

Suitably, a variant spliceosomal intron may be obtained by modifying the wild-type spliceosomal intron to increase the consensus. Suitably, the 5' splice donor site may be altered to consist of the consensus sequence GTRAGT; the branch point sequence may be altered to consist of the consensus sequence YTNAY; the polypyrimidine tract may be altered to consist of (Y)<sub>n2</sub>, where n2 is from about 10 to about 20; and/or the 3' splice acceptor site may be altered to consist of YAGG. In some embodiments, the branch point sequence is altered to consist of the nucleotide sequence TGCTGAC. In some embodiments, the polypyrimidine tract is altered to consist of the nucleotide sequence (Y)<sub>n2</sub>, where n2 is from about 10 to about 20.

Alternatively (or additionally), a variant spliceosomal intron may be obtained by modifying the wild-type spliceosomal intron to decrease the consensus, e.g. by making five or fewer, four or fewer, three or fewer, two or fewer, or one or fewer nucleotide substitution. In particular, the 5' splice donor site and/or polypyrimidine tract may each be modified by making one or more nucleotide substitution. Such a modification may be made without affecting the ability of the spliceosomal intron to increase expression of a protein-coding sequence.

Suitably, a variant spliceosomal intron may be obtained by modifying the wild-type spliceosomal intron to insert and/or delete nucleotide sequences such that the polypyrimidine tract and/or the 3' splice acceptor site are in consensus positions. Suitably, one or more nucleotides sequences are inserted and/or deleted such that the polypyrimidine tract is about 4 to about 24 bp downstream of the branch point. Suitably, one or more nucleotides sequences are inserted and/or deleted such that the 3' splice point is from about 21 to about 34 bp downstream of the branch point.

Suitably, a variant spliceosomal intron may be obtained by modifying the wild-type spliceosomal intron to insert nucleotide sequences. For example, one or more nucleotide sequences (e.g. one or more restriction site) may be introduced upstream of the branch point

sequence without affecting the ability of the spliceosomal intron to increase expression of a protein-coding sequence. Suitably, the insertion is about 100 bp or less, about 50 bp or less, about 45 bp or less, about 40 bp or less, about 40 bp or less, about 35 bp or less, about 30 bp or less, about 25 bp or less, about 20 bp or less, about 15 bp or less, about 10 bp or less, 5 or about 5 bp or less in length. In some embodiments, the insertion is about 12 bp in length.

Suitably, a variant spliceosomal intron may be obtained by modifying the wild-type spliceosomal intron to delete nucleotide sequences upstream of the branch point sequence. Sequences downstream of the 5' splice donor site and upstream of the branch point sequence may be deleted without affecting the ability of the spliceosomal intron to increase expression 10 of a protein-coding sequence. Moreover, the present inventors have surprisingly found that deletion of the 5' splice donor site may not affect the ability of a spliceosomal intron to increase expression of a protein-coding sequence. In some embodiments, the 5' splice donor site is deleted. Suitably, the deletion is about 80 bp or less, about 75 bp or less, about 70 bp or less, about 65 bp or less, about 60 bp or less, about 55 bp or less, about 50 bp or less, about 45 15 bp or less, about 40 bp or less, about 40 bp or less, about 35 bp or less, about 30 bp or less, or about 25 bp or less in length. In some embodiments, the deletion is about 22 bp in length and includes the 5' splice donor site.

Suitably, a variant spliceosomal intron may have a length of about 1000 bp or less, about 900 bp or less, about 800 bp or less, about 700 bp or less, about 600 bp or less, about 500 bp or 20 less, about 400 bp or less, about 300 bp or less, about 200 bp or less, about 150 bp or less, about 140 bp or less, or about 130 bp or less.

Suitably, a variant spliceosomal intron may have a length of about 40 bp or more, about 50 bp or more, about 60 bp or more, about 70 bp or more, about 80 bp or more, about 90 bp or more, about 100 bp or more, about 100 bp or more, about 105 bp or more, about 110 bp or more, 25 about 115 bp or more, or about 120 bp or more.

Suitably, a variant spliceosomal intron may comprise or consist of the nucleotide sequence  $(N)_x Y T N A Y (N)_{n1} (Y)_{n2} (N)_{n3} Y A G G$  (SEQ ID NO: 312), wherein x is from about 10 to about 1000 (e.g. from about 10 to about 100), n1 is from about 2 to about 22, n2 is from about 10 to about 20, and  $n1 + n2 + n3$  is from about 15 to about 40, or a nucleotide sequence having five or 30 fewer, four or fewer, three or fewer, two or fewer, or one nucleotide substitution.

The spliceosomal intron may be a rabbit beta globin intron or a variant thereof.

The rabbit beta globin (RBG) gene may have GenBank accession number V00882. The RBG intron I may comprise or consist of the nucleotide sequence of SEQ ID NO: 252. The RBG intron II may comprise or consist of the nucleotide sequence of SEQ ID NO: 253.

5 In some embodiments, the spliceosomal intron is a rabbit beta globin intron or a variant thereof having 70% or more, 75% or more, 80% or more, 81% or more, 82% or more, 83% or more, 84% or more, 85% or more, 86% or more, 87% or more, 88% or more, 89% or more, 90% or more, 91% or more, 92% or more, 93% or more, 94% or more, 95% or more, 96% or more, 97% or more, 98% or more, 99% or more, or 100% identity thereto.

10 In some embodiments, the spliceosomal intron is the rabbit beta globin intron I or a variant thereof having 70% or more, 75% or more, 80% or more, 81% or more, 82% or more, 83% or more, 84% or more, 85% or more, 86% or more, 87% or more, 88% or more, 89% or more, 90% or more, 91% or more, 92% or more, 93% or more, 94% or more, 95% or more, 96% or more, 97% or more, 98% or more, 99% or more, or 100% identity thereto.

15 In some embodiments, the spliceosomal intron is the rabbit beta globin intron I or a variant thereof having 80% or more, 81% or more, 82% or more, 83% or more, 84% or more, 85% or more, 86% or more, or 87% or more identity thereto. In some embodiments, the variant has 87% or less identity thereto.

20 An example RBG intron I variant is provided in SEQ ID NO: 254. Compared to the RBG intron I, the polypyrimidine tract sequence has been altered by substituting the native G and A residues in this region with T to increase the polypyrimidine tract consensus.

In some embodiments, the spliceosomal intron does not comprise a HTLV-IR 5' splice donor site. In some embodiments, the spliceosomal intron does not comprise the nucleotide sequence GTAAGT upstream of the branch point sequence.

25 In some embodiments, the spliceosomal intron does not comprise a HTLV-IR 5' intronic splice donor region.

30 In some embodiments, the spliceosomal intron does not comprise a nucleotide sequence having 70% or more, 75% or more, 80% or more, 81% or more, 82% or more, 83% or more, 84% or more, 85% or more, 86% or more, 87% or more, 88% or more, 89% or more, 90% or more, 91% or more, 92% or more, 93% or more, 94% or more, 95% or more, 96% or more, 97% or more, 98% or more, 99% or more, or 100% identity to SEQ ID NO: 255.

In some embodiments, the spliceosomal intron does not comprise the nucleotide sequence of SEQ ID NO: 255 (HTLV-IR 5' intronic splice donor region).

*Spliceosomal intron fragments*

In some embodiments, the 5' AAV vector comprises a regulatory element derived from a spliceosomal intron. The regulatory element may be downstream of a promoter and upstream of the 5' CDS.

- 5 In some embodiments (e.g. mRNA trans-splicing methods), the 3' AAV vector comprises a regulatory element derived from a spliceosomal intron. The regulatory element may be downstream of a promoter and upstream of the 3' CDS.

The regulatory element may, in particular, be capable of increasing expression of a protein-coding sequence downstream thereof. The regulatory element may be a 5'UTR regulatory element. A "5'UTR regulatory element" may refer to a sequence which increases expression of a protein-coding sequence downstream thereof. Suitably, the regulatory element is operable in mammalian cells, e.g. human cells. The regulatory element may be capable of increasing expression of a protein-coding sequence in mammalian cells, e.g. human cells. The regulatory element may be a mammalian regulatory element.

10

- 15 The regulatory element may be operable in kidney cells. The regulatory element may be capable of increasing expression of a protein-coding sequence in the kidney. Examples of kidney cells include, but are not limited, to glomerular cells. The regulatory element may be operable in glomerular cells. The regulatory element may be capable of increasing expression of a protein-coding sequence in the glomerulus. The regulatory element may be operable in a podocyte cell. The regulatory element may be capable of increasing expression of a protein-coding sequence in podocytes.
- 20

The regulatory element may be a fragment of any naturally-occurring, chimeric, or variant spliceosomal intron, comprising or consisting of the following conserved elements from 5' to 3': a branch point sequence; a polypyrimidine tract; and a 3' splice acceptor site.

- 25 The regulatory element preferably does not comprise a 5' splice donor site. In some embodiments, the regulatory element does not comprise the nucleotide sequence GTRAGT upstream of the branch point sequence. In some embodiments, the regulatory element does not comprise the nucleotide sequence GTTGGT upstream of the branch point sequence. In some embodiments, the regulatory element does not comprise a HTLV-IR 5' splice donor site.
- 30 In some embodiments, the regulatory element does not comprise the nucleotide sequence GTAAGT upstream of the branch point sequence.

The regulatory element may comprise each of the branch point sequence, the polypyrimidine tract, and the 3' splice acceptor site in consensus positions. Suitably, the polypyrimidine tract

is from about 4 to about 24 bp downstream of the branch point. Suitably, the 3' splice point is from about 21 to about 34 bp downstream of the branch point.

The regulatory element may comprise a consensus branch point sequence, a consensus polypyrimidine tract, and/or a consensus 3' splice acceptor site. In some embodiments, the regulatory element comprises at least a consensus branch point sequence and a consensus 3' splice acceptor site. In some embodiments, the regulatory element comprises a consensus branch point sequence, a consensus polypyrimidine tract, and a consensus 3' splice acceptor site.

Suitably, the regulatory element has a length of about 150 bp or less, about 145 bp or less, about 140 bp or less, about 135 bp or less, about 130 bp or less, about 125 bp or less, or about 120 bp or less.

Suitably, the regulatory element has a length of about 25 bp or more, about 30 bp or more, about 35 bp or more, about 40 bp or more, about 45 bp or more, about 50 bp or more, about 55 bp or more, about 60 bp or more, about 65 bp or more, about 70 bp or more, about 75 bp or more, about 80 bp or more, about 85 bp or more, about 90 bp or more, about 95 bp or more, about 100 bp or more, about 105 bp or more, about 110 bp or more, or about 115 bp or more.

Suitably, the regulatory element may comprise or consist of the nucleotide sequence  $(N)_x YTNAY(N)_{n1}(Y)_{n2}(N)_{n3}YAGG$  (SEQ ID NO: 312), wherein x is from about 10 to about 1000 (e.g. from about 10 to about 100), n1 is from about 2 to about 22, n2 is from about 10 to about 20, and  $n1 + n2 + n3$  is from about 15 to about 40, or a nucleotide sequence having five or fewer, four or fewer, three or fewer, two or fewer, or one nucleotide substitution.

Suitably, the regulatory element may comprise or consist of the nucleotide sequence  $(N)_x YTNAY(N)_{n1}(Y)_{n2}(N)_{n3}YAGG$  (SEQ ID NO: 312), wherein x is from about 10 to about 1000 (e.g. from about 10 to about 100), n1 is from about 2 to about 22, n2 is from about 10 to about 20, and  $n1 + n2 + n3$  is from about 15 to about 40.

Suitably, the regulatory element may comprise or consist of the nucleotide sequence  $YTNAY(N)_{n1}(Y)_{n2}(N)_{n3}YAGG$  (SEQ ID NO: 313), wherein n1 is from about 2 to about 22, n2 is from about 10 to about 20, and  $n1 + n2 + n3$  is from about 15 to about 40, or a nucleotide sequence having five or fewer, four or fewer, three or fewer, two or fewer, or one nucleotide substitution.

Suitably, the regulatory element may comprise or consist of the nucleotide sequence  $YTNAY(N)_{n1}(Y)_{n2}(N)_{n3}YAGG$  (SEQ ID NO: 313), wherein n1 is from about 2 to about 22, n2 is from about 10 to about 20, and  $n1 + n2 + n3$  is from about 15 to about 40.

Suitably, a regulatory element may have 70% or more, 75% or more, 80% or more, 81% or more, 82% or more, 83% or more, 84% or more, 85% or more, 86% or more, 87% or more, 88% or more, 89% or more, 90% or more, 91% or more, 92% or more, 93% or more, 94% or more, 95% or more, 96% or more, 97% or more, 98% or more, 99% or more, or 100% identity to a naturally-occurring, chimeric, or variant spliceosomal intron.

A regulatory element may be obtained from a naturally-occurring, chimeric, or variant spliceosomal intron by substituting, inserting, and/or deleting 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 nucleotide(s).

- 10 A regulatory element may be derived from a naturally-occurring, chimeric, or variant spliceosomal intron whilst retaining its ability to increase expression of a protein-coding sequence. Expression of a protein-coding sequence may be measured by any suitable method known in the art. For example by measuring the expression of a reporter transgene, e.g. Luciferase, placed downstream of the regulatory element, wherein expression of the reporter
- 15 transgene correlates with the ability of the regulatory element to increase expression of a protein-coding sequence. Suitably, the regulatory element increases expression of a protein-coding sequence to at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 100% of the level of the naturally-occurring, chimeric, or variant spliceosomal intron.

- Suitably, the naturally-occurring, chimeric, or variant spliceosomal intron may be modified to
- 20 increase (or decrease) the consensus, to insert further nucleotide sequences, and/or to delete nucleotide sequences which are not required for the regulatory element to increase expression of a protein-coding sequence.

- Suitably, a regulatory element may be derived from a naturally-occurring, chimeric, or variant spliceosomal intron by modifying the sequence to increase the consensus. Suitably, the
- 25 branch point sequence may be altered to consist of the consensus sequence YTNAY, the polypyrimidine tract may be altered to consist of  $(Y)_{n2}$ , where  $n2$  is from about 10 to about 20, and/or the 3' splice acceptor site may be altered to consist of YAGG. In some embodiments, the branch point sequence is altered to consist of the nucleotide sequence TGCTGAC. In some embodiments, the polypyrimidine tract is altered to consist of the nucleotide sequence
- 30  $(Y)_{n2}$ , where  $n2$  is from about 10 to about 20.

Alternatively (or additionally), a regulatory element may be derived from a naturally-occurring, chimeric, or variant spliceosomal intron by modifying the sequence to decrease the consensus, e.g. by making five or fewer, four or fewer, three or fewer, two or fewer, or one or fewer nucleotide substitution. In particular, the polypyrimidine tract may each be modified by



making one or more nucleotide substitution. Such a modification may be made without affecting the ability of the spliceosomal intron to increase expression of a protein-coding sequence.

5 Suitably, a regulatory element may be derived from a naturally-occurring, chimeric, or variant spliceosomal intron by modifying the sequence to insert and/or delete nucleotide sequences such that the polypyrimidine tract and/or the 3' splice acceptor site are in consensus positions. Suitably, one or more nucleotides sequences are inserted and/or deleted such that the polypyrimidine tract is about 4 to about 24 bp downstream of the branch point. Suitably, one or more nucleotides sequences are inserted and/or deleted such that the 3' splice point is from  
10 about 21 to about 34 bp downstream of the branch point.

Suitably, a regulatory element may be derived from a naturally-occurring, chimeric, or variant spliceosomal intron by modifying the sequence to insert nucleotide sequences. For example, one or more nucleotide sequences (e.g. one or more restriction site) may be introduced upstream of the branch point sequence without affecting the ability of the regulatory element  
15 to increase expression of a protein-coding sequence. Suitably, the insertion is about 100 bp or less, about 50 bp or less, about 45 bp or less, about 40 bp or less, about 40 bp or less, about 35 bp or less, about 30 bp or less, about 25 bp or less, about 20 bp or less, about 15 bp or less, about 10 bp or less, or about 5 bp or less in length. In some embodiments, the insertion is about 12 bp in length.

20 Suitably, a regulatory element may be derived a naturally-occurring, chimeric, or variant spliceosomal intron by modifying the sequence to delete nucleotide sequences upstream of the branch point sequence. Sequences upstream of the branch point sequence may be deleted without affecting the ability of the regulatory element to increase expression of a protein-coding sequence. Suitably, the deletion is about 80 bp or less, about 75 bp or less,  
25 about 70 bp or less, about 65 bp or less, about 60 bp or less, about 55 bp or less, about 50 bp or less, about 45 bp or less, about 40 bp or less, about 40 bp or less, about 35 bp or less, about 30 bp or less, or about 25 bp or less in length. In some embodiments, the deletion is about 22 bp in length and includes the 5' splice donor site.

In some embodiments, the regulatory element is derived from a RBG intron or a variant  
30 thereof. In some embodiments, the regulatory element is derived from the RBG intron I.

An example regulatory element derived from the RBG intron I is provided in SEQ ID NO: 256. Compared to the RBG intron I, the polypyrimidine tract sequence has been altered by substituting the native G and A residues in this region with T to increase the polypyrimidine

tract consensus; and 22 nucleotides at the 5' end have been deleted, including the 5' splice donor site.

In some embodiments, the regulatory element comprises or consists of a nucleotide sequence having 70% or more, 75% or more, 80% or more, 81% or more, 82% or more, 83% or more, 5 84% or more, 85% or more, 86% or more, 87% or more, 88% or more, 89% or more, 90% or more, 91% or more, 92% or more, 93% or more, 94% or more, 95% or more, 96% or more, 97% or more, 98% or more, 99% or more, or 100% identity to SEQ ID NO: 256.

In some embodiments, the regulatory element comprises or consists of a nucleotide sequence having 90% or more identity to SEQ ID NO: 256.

10 In some embodiments, the regulatory element comprises or consists of a nucleotide sequence having 95% or more identity to SEQ ID NO: 256.

In some embodiments, the regulatory element comprises or consists of a nucleotide sequence having 96% or more identity to SEQ ID NO: 256.

In some embodiments, the regulatory element comprises or consists of a nucleotide sequence 15 having 97% or more identity to SEQ ID NO: 256.

In some embodiments, the regulatory element comprises or consists of a nucleotide sequence having 98% or more identity to SEQ ID NO: 256.

In some embodiments, the regulatory element comprises or consists of a nucleotide sequence having 99% or more identity to SEQ ID NO: 256.

20 In some embodiments, the regulatory element comprises or consists of the nucleotide sequence SEQ ID NO: 256, or a variant thereof having 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 nucleotide substitution(s), deletion(s), and/or insertion(s).

In some embodiments, the regulatory element comprises or consists of the nucleotide sequence SEQ ID NO: 256, or a variant thereof having 5 or fewer, 4 or fewer, 3 or fewer, 2 or 25 fewer, or 1 nucleotide substitution(s).

In some embodiments, the regulatory element comprises or consists of the nucleotide sequence SEQ ID NO: 256.

Another example regulatory element derived from the RBG intron I is provided in SEQ ID NO: 257. Compared to SEQ ID NO: 256, 12 nucleotides have been inserted at the 5' end (including 30 a restriction site but not including a 5' splice donor site).

In some embodiments, the regulatory element comprises or consists of a nucleotide sequence having 70% or more, 75% or more, 80% or more, 81% or more, 82% or more, 83% or more, 84% or more, 85% or more, 86% or more, 87% or more, 88% or more, 89% or more, 90% or more, 91% or more, 92% or more, 93% or more, 94% or more, 95% or more, 96% or more, 5 97% or more, 98% or more, 99% or more, or 100% identity to SEQ ID NO: 257.

In some embodiments, the regulatory element comprises or consists of a nucleotide sequence having 90% or more identity to SEQ ID NO: 257.

In some embodiments, the regulatory element comprises or consists of a nucleotide sequence having 95% or more identity to SEQ ID NO: 257.

10 In some embodiments, the regulatory element comprises or consists of a nucleotide sequence having 96% or more identity to SEQ ID NO: 257.

In some embodiments, the regulatory element comprises or consists of a nucleotide sequence having 97% or more identity to SEQ ID NO: 257.

15 In some embodiments, the regulatory element comprises or consists of a nucleotide sequence having 98% or more identity to SEQ ID NO: 257.

In some embodiments, the regulatory element comprises or consists of a nucleotide sequence having 99% or more identity to SEQ ID NO: 257.

20 In some embodiments, the regulatory element comprises or consists of the nucleotide sequence SEQ ID NO: 257, or a variant thereof having 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 nucleotide substitution(s), deletion(s), and/or insertion(s).

In some embodiments, the regulatory element comprises or consists of the nucleotide sequence SEQ ID NO: 257, or a variant thereof having 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 nucleotide substitution(s).

25 In some embodiments, the regulatory element comprises or consists of the nucleotide sequence SEQ ID NO: 257.

In some embodiments, the regulatory element consists of the nucleotide sequence SEQ ID NO: 257, or a variant thereof having 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 nucleotide substitution(s), deletion(s), and/or insertion(s).

In some embodiments, the regulatory element consists of the nucleotide sequence SEQ ID NO: 257, or a variant thereof having 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 nucleotide substitution(s).

5 In some embodiments, the regulatory element consists of the nucleotide sequence SEQ ID NO: 257.

In some embodiments, the regulatory element does not comprise a HTLV-IR 5' intronic splice donor region.

10 In some embodiments, the regulatory element does not comprise a nucleotide sequence having 70% or more, 75% or more, 80% or more, 81% or more, 82% or more, 83% or more, 84% or more, 85% or more, 86% or more, 87% or more, 88% or more, 89% or more, 90% or more, 91% or more, 92% or more, 93% or more, 94% or more, 95% or more, 96% or more, 97% or more, 98% or more, 99% or more, or 100% identity to SEQ ID NO: 255.

In some embodiments, the regulatory element does not comprise the nucleotide sequence of SEQ ID NO: 255.

#### 15 *Enhancers*

In some embodiments, the 5' AAV vector comprises an enhancer. The enhancer may be upstream of a promoter.

In some embodiments (e.g. mRNA trans-splicing methods), the 3' AAV vector comprises an enhancer. The enhancer may be upstream of a promoter.

20 An "enhancer" is a region of DNA that can be bound by proteins (activators) to increase the likelihood that transcription of a particular gene will occur. Enhancers are cis-acting. They can be located up to 1 Mbp (1,000,000 bp) away from the gene, upstream or downstream from the start site. Any suitable enhancer may be used, the selection of which may be readily made by the skilled person.

25 The 5' AAV vector and/or 3' AAV vector may comprise a kidney-specific enhancer. The 5' AAV vector and/or 3' AAV vector may comprise a glomerular-specific enhancer. The 5' AAV vector and/or 3' AAV vector may comprise a podocyte-specific enhancer. Suitable enhancers will be well known to those of skill in the art. Suitably, the enhancer may be or may be derived from an enhancer associated with a gene with selective expression in human podocytes. Methods  
30 to identify the enhancer regions associated with genes will be well known to those of skill in the art.

Preferably, the enhancer is a NPHS1 or a NPHS2 enhancer, or a fragment and/or variant thereof. Preferably, the enhancer is a human enhancer, e.g. a human NPHS1 enhancer or a human NPHS2 enhancer.

Suitably, a NPHS1 enhancer may comprise or consist of the nucleotide sequence shown as  
 5 SEQ ID NO: 232, or a variant which is at least 70% identical to SEQ ID NO: 232. Suitably, the variant may be at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 96%, at least 98%, or at least 99% identical to SEQ ID NO: 232.

Suitably, a NPHS2 enhancer may comprise or consist of the nucleotide sequence shown as  
 10 SEQ ID NO: 233, or a variant which is at least 70% identical to SEQ ID NO: 233. Suitably, the variant may be at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 96%, at least 98%, or at least 99% identical to SEQ ID NO: 233.

The enhancer may be used with the corresponding promoter, for example the NPHS1  
 15 enhancer may be used with the NPHS1 promoter. Alternatively, the enhancer may be used with a different promoter, for example a promoter which is not podocyte-specific e.g. a hsp promoter.

#### *Kozak sequences*

In some embodiments, the 5' AAV vector comprises a Kozak sequence. The Kozak sequence  
 20 may be downstream of a promoter and upstream of the 5' CDS.

In some embodiments (e.g. mRNA trans-splicing methods), the 3' AAV vector comprises a Kozak sequence. The Kozak sequence may be downstream of a promoter and upstream of the 3' CDS.

A Kozak sequence may be inserted around the start codon of a protein-coding sequence to  
 25 improve the initiation of translation. Suitable Kozak sequences will be well known to those of skill in the art (see e.g. Kozak, M., 2002. Gene, 299(1-2), pp.1-34). A consensus Kozak sequence in vertebrates may have the sequence of SEQ ID NO: 258 or SEQ ID NO: 259.

Suitably, the Kozak sequence may comprise or consist of the nucleotide sequence of SEQ ID  
 30 NO: 258 or 259, or variants thereof which have five or fewer deletions, substitutions or insertions. Suitably, the variants may have four or fewer, three or fewer, two or fewer, or one deletion(s), substitution(s) or insertion(s). Suitably, the variants may have three or fewer, two or fewer, or one deletion(s) and/or three or fewer, two or fewer, or one substitution(s). Suitably,

the variants may have three or fewer, two or fewer, or one deletion(s) and/or three or fewer, two or fewer, or one substitution(s). Suitably, the variants may have one deletion and/or one substitution. Suitably, the variants may have one deletion and one substitution.

- 5 Suitably, the Kozak sequence may comprise or consist of the nucleotide sequence of any of SEQ ID NOs: 260 to 262, or variants which have five or fewer deletions, substitutions or insertions. Suitably, the variants may have four or fewer, three or fewer, two or fewer, or one deletion(s), substitution(s) or insertion(s). Suitably, the variants may have three or fewer, two or fewer, or one deletion(s) and/or three or fewer, two or fewer, or one substitution(s). Suitably,
- 10 the variants may have one deletion and/or one substitution. Suitably, the variants may have one deletion and one substitution. Suitably, the Kozak sequence may comprise or consist of the nucleotide sequence of any of SEQ ID NOs: 260 to 262.

- Suitably, the Kozak sequence may comprise or consist of the nucleotide sequence shown as SEQ ID NO: 260, or a variant which is at least 65% identical to SEQ ID NO: 260. Suitably, the
- 15 variant may be at least 75%, at least 85%, or at least 90% identical to SEQ ID NO: 260. Suitably, the Kozak sequence may comprise or consist of the nucleotide sequence shown as SEQ ID NO: 260.

- Suitably, the Kozak sequence may comprise or consist of the nucleotide sequence shown as SEQ ID NO: 261, or a variant which is at least 60% identical to SEQ ID NO: 261. Suitably, the
- 20 variant may be at least 70%, at least 80%, or at least 90% identical to SEQ ID NO: 261. Suitably, the Kozak sequence may comprise or consist of the nucleotide sequence shown as SEQ ID NO: 261.

- Suitably, the Kozak sequence may comprise or consist of the nucleotide sequence shown as SEQ ID NO: 262, or a variant which is at least 65% identical to SEQ ID NO: 262. Suitably, the
- 25 variant may be at least 75%, or at least 85% identical to SEQ ID NO: 262. Suitably, the Kozak sequence may comprise or consist of the nucleotide sequence shown as SEQ ID NO: 262.

tccaccatg  
Example Kozak sequence 3 (SEQ ID NO: 262)

- 30 *Post-transcriptional regulatory elements*

In some embodiments, the 3' AAV vector comprises a post-transcriptional regulatory element. The post-transcriptional regulatory element may be downstream of the 3' CDS and upstream of a polyadenylation sequence.

In some embodiments (e.g. mRNA trans-splicing methods), the 5' AAV vector comprises a post-transcriptional regulatory element. The post-transcriptional regulatory element may be downstream of the 5' CDS and upstream of a polyadenylation sequence.

5 The post-transcriptional regulatory element may be a Woodchuck Hepatitis Virus Post-transcriptional Regulatory Element (WPRE). The WPRE sequence may have mutations within the X-antigen promoter and/or the initiation codon of the X-antigen. This may prevent the production of a functional X-antigen.

Suitably, the WPRE may comprise or consist of the nucleotide sequence shown as SEQ ID NO: 263, or a variant which is at least 70% identical to SEQ ID NO: 263. Suitably, the variant  
10 may be at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 96%, at least 98%, or at least 99% identical to SEQ ID NO: 263.

#### Example dual AAV vector systems

Any suitable dual AAV vector approach may be used to reconstitute the CDS upon co-delivery  
15 of the dual AAV vectors.

For example, the CDS can subsequently be reconstituted at the AAV genome, mRNA, or protein level (see e.g. McClements, M.E. and MacLaren, R.E., 2017. The Yale journal of biology and medicine, 90(4), p.611; and Riedmayr, L.M., et al., 2023. Nature Communications, 14(1), p.6578).

20 In some embodiments, the CDS encoding a COL4A3, COL4A4 or COL4A5 polypeptide is reconstituted at the AAV genome level or mRNA level.

In some embodiments, the CDS encoding a COL4A3, COL4A4 or COL4A5 polypeptide is reconstituted at the AAV genome level.

Gene reconstitution at the AAV genome level (also referred to herein as DNA level) was the  
25 first dual AAV vector approach applied *in vivo*. The different versions of this method rely on homologous recombination and/or concatemerization of the AAV genome (see e.g. McClements, M.E. and MacLaren, R.E., 2017. The Yale journal of biology and medicine, 90(4), p.611). General methods of reconstitution at the DNA level which are described herein include "DNA overlapping methods"; "DNA trans-splicing methods" and "DNA hybrid methods",  
30 although any other suitable method may be used to reconstitute the CDS at the DNA level.

In some embodiments, the dual AAV vector system comprises:

(a) a first AAV vector comprising from 5' to 3': a promoter, optionally one or more regulatory sequences, and a 5' coding sequence (CDS), wherein the 5' CDS encodes an N-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide; and

5 (b) a second AAV vector comprising from 5' to 3': a 3' CDS, and optionally one or more regulatory sequences (e.g. a polyadenylation sequence), wherein the 3' CDS encodes a C-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide,

wherein the 5' CDS and the 3' CDS share an overlapping sequence or do not share an overlapping sequence.

10 In some embodiments, the CDS encoding a COL4A3, COL4A4 or COL4A5 polypeptide is reconstituted at the mRNA level.

Gene reconstitution at the mRNA level (also referred to herein as RNA level) has been used to express the cystic fibrosis transmembrane conductance regulator (CFTR) gene (see e.g. Song, Y., et al., 2009. Human gene therapy, 20(3), pp.267-281). General methods of reconstitution at the RNA level which are described herein include "mRNA trans-splicing methods", although any other suitable method may be used to reconstitute the CDS at the  
15 RNA level.

In some embodiments, the dual AAV vector system comprises:

20 (a) a first AAV vector comprising from 5' to 3': a promoter, optionally one or more regulatory sequences, a 5' coding sequence (CDS), and optionally one or more regulatory sequences (e.g. a polyadenylation sequence), wherein the 5' CDS encodes an N-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide; and

(b) a second AAV vector comprising from 5' to 3': a promoter, optionally one or more regulatory sequences, a 3' CDS, and optionally one or more regulatory sequences (e.g. a polyadenylation sequence), wherein the 3' CDS encodes a C-terminal part of a  
25 COL4A3, COL4A4 or COL4A5 polypeptide,

wherein the 5' CDS and the 3' CDS do not share an overlapping sequence.

#### *DNA overlapping methods*

In some embodiments, the CDS encoding a COL4A3, COL4A4 or COL4A5 polypeptide is reconstituted at the DNA level by an overlapping method.



In DNA overlapping methods, the 5' CDS and the 3' CDS share overlapping sequences. It is generally believed that some form of homologous recombination occurs between the two partial genome components to reconstitute the large gene expression cassette (see e.g. Duan, D., et al., 2001. *Molecular therapy*, 4(4), pp.383-391).

- 5 In some embodiments, the 5' CDS and the 3' CDS share an overlapping sequence.

In some embodiments, the dual AAV vector system comprises:

- (a) a first AAV vector comprising from 5' to 3': a promoter, optionally one or more regulatory sequences, and a 5' coding sequence (CDS), wherein the 5' CDS encodes an N-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide; and
- 10 (b) a second AAV vector comprising from 5' to 3': a 3' CDS, and optionally one or more regulatory sequences (e.g. a polyadenylation sequence), wherein the 3' CDS encodes a C-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide,

wherein the 5' CDS and the 3' CDS share an overlapping sequence.

- 15 In some embodiments, the 5' CDS sequence is longer than the 3' CDS sequence and in some embodiments the 3' CDS sequence is longer than the 5' CDS sequence. For both of these approaches, the overlapping sequence suitably has a length of from 200-2000 nucleotides, preferably from 450-1600 nucleotides.

#### *DNA trans-splicing methods*

- 20 In some embodiments, the CDS encoding a COL4A3, COL4A4 or COL4A5 polypeptide is reconstituted at the DNA level by a trans-splicing method.

In DNA trans-splicing methods, the 5' CDS has a splice donor sequence at the 3' end and the 3' CDS has a splice acceptor sequence at the 5' end. This allows trans-splicing following head-to-tail concatemerization of the two AAV genomes (see e.g. Yan, Z., et al., 2005. *Journal of virology*, 79(1), pp.364-379).

- 25 In some embodiments, the 5' AAV vector comprises a splice donor sequence downstream of the 5' CDS. The splice donor sequence may be immediately downstream of the 5' CDS. The splice donor sequence may be any described herein.

- In some embodiments, the 3' AAV vector comprises a splice acceptor sequence upstream of the 3' CDS. The splice acceptor sequence may be immediately upstream of the 3' CDS. The  
30 splice acceptor sequence may be any described herein.

In some embodiments, the 5' AAV vector comprises a splice donor sequence downstream of the 5' CDS and the 3' AAV vector comprises a splice acceptor sequence upstream of the 3' CDS.

In some embodiments, the dual AAV vector system comprises:

- 5 (a) a first AAV vector comprising from 5' to 3': a promoter, optionally one or more regulatory sequences, a 5' coding sequence (CDS), and a splice donor sequence, wherein the 5' CDS encodes an N-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide; and
- 10 (b) a second AAV vector comprising from 5' to 3': a splice acceptor sequence, a 3' CDS, and optionally one or more regulatory sequences (e.g. a polyadenylation sequence), wherein the 3' CDS encodes a C-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide,

wherein the 5' CDS and the 3' CDS do not share an overlapping sequence.

#### *DNA hybrid methods*

- 15 In some embodiments, the CDS encoding a COL4A3, COL4A4 or COL4A5 polypeptide is reconstituted at the DNA level by a hybrid method.

DNA hybrid method combine the mechanisms of both DNA overlapping and trans-splicing methods. To drive homologous recombination, a highly recombinogenic region serves as the homologous overlapping sequence. The 5' CDS has a splice donor sequence at the 3' end and the 3' CDS has a splice acceptor sequence at the 5' end. Homologous recombination can rejoin the two AAV vectors, and the pre-engineered splicing signal then removes the overlapping region to form the full-length CDS.

- 25 In some embodiments, the 5' AAV vector comprises a splice donor sequence downstream of the 5' CDS and further comprises a recombinogenic sequence downstream of the splice donor sequence. The recombinogenic sequence may be immediately downstream of the splice donor sequence. The recombinogenic sequence may be any described herein.

- 30 In some embodiments, the 3' AAV vector comprises a splice acceptor sequence upstream of the 3' CDS and further comprises a recombinogenic sequence upstream of the splice acceptor sequence. The recombinogenic sequence may be immediately upstream of the splice donor sequence. The recombinogenic sequence may be any described herein.

In some embodiments, the 5' AAV vector comprises a splice donor sequence downstream of the 5' CDS and further comprises a recombinogenic sequence downstream of the splice donor sequence; and the 3' AAV vector comprises a splice acceptor sequence upstream of the 3' CDS and further comprises a recombinogenic sequence upstream of the splice acceptor sequence. The recombinogenic sequences may be identical.

In some embodiments, the dual AAV vector system comprises:

- (a) a first AAV vector comprising from 5' to 3': a promoter, optionally one or more regulatory sequences, a 5' coding sequence (CDS), a splice donor sequence, and a recombinogenic sequence, wherein the 5' CDS encodes an N-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide; and
  - (b) a second AAV vector comprising from 5' to 3': a recombinogenic sequence, a splice acceptor sequence, a 3' CDS, and optionally one or more regulatory sequences (e.g. a polyadenylation sequence), wherein the 3' CDS encodes a C-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide,
- wherein the 5' CDS and the 3' CDS do not share an overlapping sequence, and preferably wherein the recombinogenic sequences are identical.

#### *mRNA trans-splicing methods*

In some embodiments, the CDS encoding a COL4A3, COL4A4 or COL4A5 polypeptide is reconstituted at the RNA level by a trans-splicing method.

- In mRNA trans-splicing methods, the 5' CDS has a splice donor sequence at the 3' end and the 3' CDS has a splice acceptor sequence at the 5' end. This allows trans-splicing following transcription of the mRNA. The AAV vectors may further comprise complementary hybridisation domains to promote trans-splicing (see e.g. Riedmayr, L.M., et al., 2023. Nature Communications, 14(1), p.6578).
- The 5' AAV vector and the 3' AAV vector may each comprise all elements necessary for mRNA transcript expression, such as a promoter and a polyadenylation sequence. In some embodiments, the 5' AAV vector comprises one or more regulatory sequences downstream of the 5' CDS (e.g. a polyadenylation sequence) and the 3' AAV vector comprises a promoter and optionally one or more regulatory sequences upstream of the 3' CDS (e.g. a spliceosomal intron or a fragment thereof and/or a Kozak sequence).

In some embodiments, the 5' AAV vector comprises a splice donor sequence downstream of the 5' CDS. The splice donor sequence may be immediately downstream of the 5' CDS. The splice donor sequence may be any described herein (e.g. any forward splice donor sequence described herein).

- 5 In some embodiments, the 3' AAV vector comprises a splice acceptor sequence upstream of the 3' CDS. The splice acceptor sequence may be immediately upstream of the 3' CDS. The splice acceptor sequence may be any described herein (e.g. any reverse splice acceptor sequence described herein).

- 10 In some embodiments, the 5' AAV vector comprises a splice donor sequence downstream of the 5' CDS and the 3' AAV vector comprises a splice acceptor sequence upstream of the 3' CDS.

- 15 In some embodiments, the 5' AAV vector further comprises a hybridisation domain downstream of the splice donor sequence. The hybridisation domain may be immediately downstream of the splice donor sequence. The hybridisation domain may be any described herein.

In some embodiments, the 3' AAV vector further comprises a hybridisation domain upstream of the splice acceptor sequence. The hybridisation domain may be immediately upstream of the splice donor sequence. The hybridisation domain may be any described herein.

- 20 In some embodiments, the 5' AAV vector comprises a splice donor sequence downstream of the 5' CDS and further comprises a hybridisation domain downstream of the splice donor sequence; and the 3' AAV vector comprises a splice acceptor sequence upstream of the 3' CDS and further comprises a complementary hybridisation domain upstream of the splice acceptor sequence.

In some embodiments, the dual AAV vector system comprises:

- 25 (a) a first AAV vector comprising from 5' to 3': a promoter, optionally one or more regulatory sequences, a 5' coding sequence (CDS), a splice donor sequence, a hybridisation domain, and optionally one or more regulatory sequences (e.g. a polyadenylation sequence), wherein the 5' CDS encodes an N-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide; and
- 30 (b) a second AAV vector comprising from 5' to 3': a promoter, optionally one or more regulatory sequences, a complementary hybridisation domain, a splice acceptor sequence, a 3' CDS, and optionally one or more regulatory sequences (e.g. a

polyadenylation sequence), wherein the 3' CDS encodes a C-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide,

wherein the 5' CDS and the 3' CDS do not share an overlapping sequence.

The regulatory sequences in the 5' AAV vector and the 3' AAV vector may be identical or different. In some embodiments, the promoters are identical. In some embodiments, the upstream regulatory sequences are identical. In some embodiments, the downstream regulatory sequences (e.g. polyadenylation sequences) are identical. In some embodiments, the promoters are identical, the upstream regulatory sequences are identical, and the downstream regulatory sequences (e.g. polyadenylation sequences) are identical.

#### 10 *Overlapping 5' CDS and 3' CDS sequences*

In some embodiments, for example in DNA overlapping methods, the 5' CDS and the 3' CDS share an overlapping sequence. The total length of the 5' CDS and the 3' CDS combined is greater than the length of a CDS encoding a COL4A3, COL4A4, or COL4A5 polypeptide.

In some embodiments, the overlapping sequence has a length of at least one exon, at least two exons, at least three exons, at least four exons, at least five exons, at least six exons, at least seven exons, at least eight exons, at least nine exons, or at least ten exons of a CDS encoding a COL4A3, COL4A4 or COL4A5 polypeptide. In some embodiments, the overlapping sequence has a length of at least five exons of a CDS encoding a COL4A3, COL4A4 or COL4A5 polypeptide.

20 In some embodiments, the overlapping sequence has a length of fifteen exons or less, twelve exons or less, or ten exons or less of a CDS encoding a COL4A3, COL4A4 or COL4A5 polypeptide. In some embodiments, the overlapping sequence has a length of twelve exons or less of a CDS encoding a COL4A3, COL4A4 or COL4A5 polypeptide.

25 In some embodiments, the overlapping sequence has a length of from one exon to fifteen exons, from two exons to twelve exons, or from three exons to ten exons of a CDS encoding a COL4A3, COL4A4 or COL4A5 polypeptide. In some embodiments, the overlapping sequence has a length of from five exons to fifteen exons, from five exons to fourteen exons, from five exons to thirteen exons, or from five exons to twelve exons of a CDS encoding a COL4A3, COL4A4 or COL4A5 polypeptide.

30 In preferred embodiments, the overlapping sequence has a length of from five exons to fifteen exons, from six exons to fourteen exons, from seven exons to thirteen exons, from eight exons to twelve exons, from nine exons to eleven exons, or about ten exons of a CDS encoding a

COL4A3, COL4A4 or COL4A5 polypeptide. In some embodiments, the overlapping sequence has a length of about ten exons of a CDS encoding a COL4A3, COL4A4 or COL4A5 polypeptide.

In some embodiments:

- 5 (a) the 5' CDS comprises or consists of exons 1 to x of a CDS encoding a COL4A3 polypeptide and the 3' CDS comprises or consists of exons (x-y) to 52 of a CDS encoding a COL4A3 polypeptide, wherein x is from 28 to 43, and wherein y is from 0 to 14 (e.g. 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14);
- (b) the 5' CDS comprises or consists of exons 1 to x of a CDS encoding a COL4A4 polypeptide and the 3' CDS comprises or consists of exons (x-y) to 47 of a CDS encoding a COL4A4 polypeptide, wherein x is from 25 to 39, and wherein y is from 0 to 14 (e.g. 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14); or
- 10 (c) the 5' CDS comprises or consists of exons 1 to x of a CDS encoding a COL4A5 polypeptide and the 3' CDS comprises or consists of exons (x-y) to 51 of a CDS encoding a COL4A5 polypeptide, wherein x is from 27 to 41, and wherein y is from 0 to 14 (e.g. 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14).

In some embodiments:

- 20 (a) the 5' CDS comprises or consists of exons 1 to x of a CDS encoding a COL4A3 polypeptide and the 3' CDS comprises or consists of exons (x-y) to 52 of a CDS encoding a COL4A3 polypeptide, wherein x is from 28 to 43, and wherein y is from 3 to 10 (e.g. 3, 4, 5, 6, 7, 8, 9, 10);
- (b) the 5' CDS comprises or consists of exons 1 to x of a CDS encoding a COL4A4 polypeptide and the 3' CDS comprises or consists of exons (x-y) to 47 of a CDS encoding a COL4A4 polypeptide, wherein x is from 25 to 39, and wherein y is from 3 to 10 (e.g. 3, 4, 5, 6, 7, 8, 9, 10); or
- 25 (c) the 5' CDS comprises or consists of exons 1 to x of a CDS encoding a COL4A5 polypeptide and the 3' CDS comprises or consists of exons (x-y) to 51 of a CDS encoding a COL4A5 polypeptide, wherein x is from 27 to 41, and wherein y is from 3 to 10 (e.g. 3, 4, 5, 6, 7, 8, 9, 10).

30 In some embodiments:

- (a) the 5' CDS comprises or consists of exons 1 to x of a CDS encoding a COL4A3 polypeptide and the 3' CDS comprises or consists of exons (x-y) to 52 of a CDS encoding a COL4A3 polypeptide, wherein x is from 28 to 43, and wherein y is from 4 to 14 (e.g. 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14);
- 5 (b) the 5' CDS comprises or consists of exons 1 to x of a CDS encoding a COL4A4 polypeptide and the 3' CDS comprises or consists of exons (x-y) to 47 of a CDS encoding a COL4A4 polypeptide, wherein x is from 25 to 39, and wherein y is from 4 to 14 (e.g. 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14); or
- 10 (c) the 5' CDS comprises or consists of exons 1 to x of a CDS encoding a COL4A5 polypeptide and the 3' CDS comprises or consists of exons (x-y) to 51 of a CDS encoding a COL4A5 polypeptide, wherein x is from 27 to 41, and wherein y is from 4 to 14 (e.g. 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14).

In some embodiments, y is from 4 to 13, from 4 to 12, or from 4 to 11. In some embodiments, y is from 5 to 13, from 6 to 12, from 7 to 11, from 8 to 10, or 9.

- 15 In some embodiments, the overlapping sequence encodes at least 50 amino acids, at least 60 amino acids, at least 70 amino acids, at least 80 amino acids, at least 90 amino acids, at least 100 amino acids, at least 150 amino acids, at least 200 amino acids, at least 250 amino acids, at least 300 amino acids, at least 350 amino acids, or at least 365 amino acids.

- 20 In some embodiments, the overlapping sequence encodes at least 100 amino acids, at least 105 amino acids, at least 110 amino acids, at least 115 amino acids, at least 120 amino acids, at least 125 amino acids, at least 130 amino acids, at least 135 amino acids, at least 140 amino acids, at least 145 amino acids, at least 150 amino acids, at least 155 amino acids, at least 160 amino acids, or at least 165 amino acids. In some embodiments, the overlapping sequence encodes at least 150 amino acids, at least 155 amino acids, at least 160 amino acids, or at least 165 amino acids. In some embodiments, the overlapping sequence encodes at least 166 amino acids.
- 25

- 30 In some embodiments, the overlapping sequence encodes 500 amino acids or less, 400 amino acids or less, 450 amino acids or less, or 350 amino acids or less. In some embodiments, the overlapping sequence encodes 470 amino acids or less, 465 amino acids or less, 460 amino acids or less, or 455 amino acids or less. In some embodiments, the overlapping sequence encodes 453 amino acids or less.

In some embodiments:

5 (a) the 5' CDS encodes amino acids 1 to x and the 3' CDS encodes amino acids (x-y) to 1670 of SEQ ID NO: 1 or a variant thereof having 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 amino acid substitution(s), wherein x is from 735 to 1297 and y is from 150 to 500;

10 (b) the 5' CDS encodes amino acids 1 to x and the 3' CDS encodes amino acids (x-y) to 1690 of SEQ ID NO: 8 or a variant thereof having 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 amino acid substitution(s), wherein x is from 745 to 1297 and y is from 150 to 500; or

15 (c) the 5' CDS encodes amino acids 1 to x and the 3' CDS encodes amino acids (x-y) to 1685 of SEQ ID NO: 21 or a variant thereof having 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 amino acid substitution(s), wherein x is from 742 to 1297 and y is from 150 to 500.

In some embodiments, y is from 200 to 500, from 250 to 500, from 300 to 450, from 350 to 400, or about 364.

In some embodiments:

20 (a) the 5' CDS encodes amino acids 1 to x and the 3' CDS encodes amino acids (x-y) to 1670 of SEQ ID NO: 1 or a variant thereof having 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 amino acid substitution(s), wherein x is from 735 to 1297 and y is from 150 to 350;

25 (b) the 5' CDS encodes amino acids 1 to x and the 3' CDS encodes amino acids (x-y) to 1690 of SEQ ID NO: 8 or a variant thereof having 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 amino acid substitution(s), wherein x is from 745 to 1297 and y is from 150 to 350; or

30 (c) the 5' CDS encodes amino acids 1 to x and the 3' CDS encodes amino acids (x-y) to 1685 of SEQ ID NO: 21 or a variant thereof having 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 amino acid substitution(s), wherein x is from 742 to 1297 and y is from 150 to 350.



In some embodiments, the overlapping sequence has a length of at least 100 nucleotides, at least 150 nucleotides, at least 200 nucleotides, at least 250 nucleotides, at least 300 nucleotides, at least 350 nucleotides, at least 400 nucleotides, at least 450 nucleotides, at least 500 nucleotides, at least 550 nucleotides, at least 600 nucleotides, at least 650  
5 nucleotides, at least 700 nucleotides, at least 750 nucleotides, at least 800 nucleotides, at least 850 nucleotides, at least 900 nucleotides, at least 900 nucleotides, or at least 1000 nucleotides.

In some embodiments, the overlapping sequence has a length of at least 498 nucleotides.

In some embodiments, the overlapping sequence has a length of 2000 nucleotides or fewer,  
10 1500 nucleotides or fewer, 1400 nucleotides or fewer, 1300 nucleotides or fewer, 1200 nucleotides or fewer, 1100 nucleotides or fewer, or 1000 nucleotides or fewer.

In some embodiments, the overlapping sequence has a length of 1600 nucleotides or fewer.

In preferred embodiments, the overlapping sequence has a length of from 400 nucleotides to 1600 nucleotides, from 450 nucleotides to 1600 nucleotides, from 500 nucleotides to 1600  
15 nucleotides, from 550 nucleotides to 1600 nucleotides, from 600 nucleotides to 1600 nucleotides, from 650 nucleotides to 1600 nucleotides, from 700 nucleotides to 1600 nucleotides, from 750 nucleotides to 1450 nucleotides, from 800 nucleotides to 1400 nucleotides, from 850 nucleotides to 1350 nucleotides, from 900 nucleotides to 1300 nucleotides, from 950 nucleotides to 1250 nucleotides, from 1000 nucleotides to 1200  
20 nucleotides, or about 1095 nucleotides.

In some embodiments:

- (a) the 5' CDS consists of nucleotides 1 to x and the 3' CDS consists of nucleotides (x-y) to 5013 of a CDS encoding a COL4A3 polypeptide, wherein x is from 2205 to 3893 and y is from 300 to 1500;
- 25 (b) the 5' CDS consists of nucleotides 1 to x and the 3' CDS consists of nucleotides (x-y) to 5073 of a CDS encoding a COL4A4 polypeptide, wherein x is from 2235 to 3893 and y is from 300 to 1500; or
- (c) the 5' CDS consists of nucleotides 1 to x and the 3' CDS consists of nucleotides (x-y) to 5058 of a CDS encoding a COL4A5 polypeptide, wherein x is from 2225 to 3893  
30 and y is from 300 to 1500.

In some embodiments:

5 (a) the 5' CDS consists of nucleotides 1 to x and the 3' CDS consists of nucleotides (x-y) to 5013 of SEQ ID NO: 27 or a variant thereof having 1000 or fewer, 500 or fewer, 400 or fewer, 300 or fewer, 200 or fewer, 100 or fewer, 90 or fewer, 80 or fewer, 70 or fewer, 60 or fewer, 50 or fewer, 40 or fewer, 30 or fewer, 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 nucleotide substitution(s), wherein x is from 2205 to 3893 and y is from 300 to 1500;

10 (b) the 5' CDS consists of nucleotides 1 to x and the 3' CDS consists of nucleotides (x-y) to 5073 of SEQ ID NO: 81 or a variant thereof having 1000 or fewer, 500 or fewer, 400 or fewer, 300 or fewer, 200 or fewer, 100 or fewer, 90 or fewer, 80 or fewer, 70 or fewer, 60 or fewer, 50 or fewer, 40 or fewer, 30 or fewer, 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 nucleotide substitution(s), wherein x is from 2235 to 3893 and y is from 300 to 1500; or

15 (c) the 5' CDS consists of nucleotides 1 to x and the 3' CDS consists of nucleotides (x-y) to 5058 of SEQ ID NO: 129 or a variant thereof having 1000 or fewer, 500 or fewer, 400 or fewer, 300 or fewer, 200 or fewer, 100 or fewer, 90 or fewer, 80 or fewer, 70 or fewer, 60 or fewer, 50 or fewer, 40 or fewer, 30 or fewer, 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 nucleotide substitution(s), wherein x is from 2225 to 3893 and y is from 300 to 1500.

25 In some embodiments, y is from 400 to 1500, from 450 to 1500, from 500 to 1500, from 550 to 1500, from 600 to 1500, from 650 to 1500, from 700 to 1500, from 750 to 1500, from 800 to 1500, from 850 to 1500, from 900 to 1500, from 950 to 1500, or from 1000 to 1500. In some embodiments, y is from 400 to 1500, from 500 to 1500, from 600 to 1500, from 700 to 1500, from 750 to 1450, from 800 to 1400, from 850 to 1350, from 900 to 1300, from 950 to 1250, from 1000 to 1200, or from 1050 to 1150.

In some embodiments, y is about 1094. In some embodiments, the term "about" includes  $\pm 100$ ,  $\pm 50$ ,  $\pm 40$ ,  $\pm 30$ ,  $\pm 20$ ,  $\pm 10$ , or  $\pm 5$ .

30 In some embodiments:

(a) the 5' CDS consists of nucleotides 1 to x and the 3' CDS consists of nucleotides (x-y) to 5013 of SEQ ID NO: 27 or a variant thereof having 1000 or fewer, 500 or fewer, 400 or fewer, 300 or fewer, 200 or fewer, 100 or fewer, 90 or fewer, 80 or fewer, 70 or

fewer, 60 or fewer, 50 or fewer, 40 or fewer, 30 or fewer, 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 nucleotide substitution(s), wherein x is from 2205 to 3893 and y is from 400 to 1000;

- 5 (b) the 5' CDS consists of nucleotides 1 to x and the 3' CDS consists of nucleotides (x-y) to 5073 of SEQ ID NO: 81 or a variant thereof having 1000 or fewer, 500 or fewer, 400 or fewer, 300 or fewer, 200 or fewer, 100 or fewer, 90 or fewer, 80 or fewer, 70 or fewer, 60 or fewer, 50 or fewer, 40 or fewer, 30 or fewer, 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 nucleotide substitution(s), wherein x is from 2235 to 3893 and y is from 400 to 1000; or

- 10 (c) the 5' CDS consists of nucleotides 1 to x and the 3' CDS consists of nucleotides (x-y) to 5058 of SEQ ID NO: 129 or a variant thereof having 1000 or fewer, 500 or fewer, 400 or fewer, 300 or fewer, 200 or fewer, 100 or fewer, 90 or fewer, 80 or fewer, 70 or fewer, 60 or fewer, 50 or fewer, 40 or fewer, 30 or fewer, 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 nucleotide substitution(s), wherein x is from 2225 to 3893 and y is from 400 to 1000.

*Example overlapping 5' CDS and 3' CDS sequences (long/short)*

- 20 In some embodiments, the 5' CDS and the 3' CDS share an overlapping sequence and the 5' CDS is longer than the 3' CDS.

Examples of suitable 5' CDS and 3' CDS sequences for these embodiments are provided in the sections entitled "Example overlapping 5' CDS (long)" and "Example overlapping 3' CDS (short)", respectively.

- 25 In some embodiments, the 5' CDS comprises or consists of exons 1 to x of a CDS encoding a COL4A5 polypeptide and the 3' CDS comprises or consists of exons (x-y) to 51 of a CDS encoding a COL4A5 polypeptide, wherein x is from 36 to 41 (e.g. 36, 37, 38, 39, 40, 41), and wherein y is from 4 to 14 (e.g. 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14).

- 30 In some embodiments, x is 40. In some embodiments, y is 9. In some embodiments, x is 40 and y is 9

In some embodiments, the 5' CDS encodes amino acids 1 to x and the 3' CDS encodes amino acids (x-y) to 1685 of SEQ ID NO: 21 or a variant thereof having 20 or fewer, 15 or fewer, 10

or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 amino acid substitution(s), wherein x is from 1100 to 1297 and y is from 150 to 500.

In some embodiments, x is about 1201. In some embodiments, y is about 364. In some  
5 embodiments, x is about 1201 and y is about 364. In some embodiments, the 5' CDS consists of nucleotides 1 to x and the 3' CDS consists of nucleotides (x-y) to 5058 of a CDS encoding a COL4A5 polypeptide, wherein x is from 3200 to 3800 and y is from 400 to 1500. In some  
10 embodiments, the 5' CDS consists of nucleotides 1 to x and the 3' CDS consists of nucleotides (x-y) to 5058 of SEQ ID NO: 129 or a variant thereof having 1000 or fewer, 500 or fewer, 400 or fewer, 300 or fewer, 200 or fewer, 100 or fewer, 90 or fewer, 80 or fewer, 70 or fewer, 60 or fewer, 50 or fewer, 40 or fewer, 30 or fewer, 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1  
nucleotide substitution(s), wherein x is from 3200 to 3800 and y is from 400 to 1500.

In some embodiments, x is about 3604. In some embodiments, y is about 1094. In some  
15 embodiments, x is about 3604 and y is about 1094

In some embodiments, the 5' CDS and 3' CDS consist of the nucleotide sequence pairs described in table 1 or variants thereof having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity thereto; or variants  
20 thereof having 500 or fewer, 400 or fewer, 300 or fewer, 200 or fewer, 100 or fewer, 90 or fewer, 80 or fewer, 70 or fewer, 60 or fewer, 50 or fewer, 40 or fewer, 30 or fewer, 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 nucleotide substitution(s), deletion(s), and/or addition(s). Table 1:

| 5' CDS | 3' CDS |  | 5' CDS | 3' CDS |  | 5' CDS | 3' CDS |
|--------|--------|--|--------|--------|--|--------|--------|
| 287    | 295    |  | 287    | 220    |  | 287    | 308    |
| 287    | 216    |  | 287    | 221    |  | 287    | 309    |
| 287    | 217    |  | 287    | 222    |  | 287    | 310    |
| 287    | 218    |  | 287    | 296    |  |        |        |
| 287    | 219    |  | 287    | 297    |  |        |        |

*Example overlapping 5' CDS and 3' CDS sequences (short/long)*

In some embodiments, the 5' CDS and the 3' CDS share an overlapping sequence and the 3' CDS is longer than the 5' CDS.

5 Examples of suitable 5' CDS and 3' CDS sequences for these embodiments are provided in the sections entitled "Example overlapping 5' CDS (short)" and "Example overlapping 3' CDS (long)", respectively.

10 In some embodiments, the 5' CDS comprises or consists of exons 1 to (x+y) of a CDS encoding a COL4A5 polypeptide and the 3' CDS comprises or consists of exons x to 51 of a CDS encoding a COL4A5 polypeptide, wherein x is from 21 to 30 (e.g. 21, 22, 23, 24, 25, 26, 27, 28, 29, 30), and wherein y is from 4 to 14 (e.g. 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14).

In some embodiments, x is 25.

In some embodiments, x is 25 and y is 4. In some embodiments, x is 25 and y is 7. In some embodiments, x is 25 and y is 9.

15 In some embodiments, the 5' CDS encodes amino acids 1 to (x+y) and the 3' CDS encodes amino acids x to 1685 of SEQ ID NO: 21 or a variant thereof having 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 amino acid substitution(s), wherein x is from 400 to 800 and y is from 150 to 500.

20 In some embodiments, x is about 574. In some embodiments, y is about 250. In some embodiments, y is about 365. In some embodiments, y is about 450.

In some embodiments, x is about 574 and y is about 250. In some embodiments, x is about 574 and y is about 365. In some embodiments, x is about 574 and y is about 450.

25 In some embodiments, the 5' CDS consists of nucleotides 1 to (x+y-1) and the 3' CDS consists of nucleotides x to 5058 of a CDS encoding a COL4A5 polypeptide, wherein x is from 1200 to 2400 and y is from 400 to 1500. In some embodiments, the 5' CDS consists of nucleotides 1 to (x+y-1) and the 3' CDS consists of nucleotides x to 5058 of SEQ ID NO: 129 or a variant thereof having 1000 or fewer, 500 or fewer, 400 or fewer, 300 or fewer, 200 or fewer, 100 or fewer, 90 or fewer, 80 or fewer, 70 or fewer, 60 or fewer, 50 or fewer, 40 or fewer, 30 or fewer, 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 nucleotide substitution(s), wherein x is from 1200 to 2400 and y is from 400 to 1500.

30

In some embodiments, x is about 1724. In some embodiments, y is about 750. In some embodiments, y is about 1094. In some embodiments, y is about 1359.

In some embodiments, x is about 1724 and y is about 750. In some embodiments, x is about 1724 and y is about 1094. In some embodiments, x is about 1724 and y is about 1359.

- 5 In some embodiments, (i) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 304 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 307; (ii) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 305 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 307; or (iii) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 306 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 307, or variants thereof having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity thereto.

- 15 In some embodiments, (i) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 304 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 307; (ii) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 305 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 307; or (iii) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 306 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 307, or variants thereof having 500 or fewer, 400 or fewer, 300 or fewer, 200 or fewer, 100 or fewer, 90 or fewer, 80 or fewer, 70 or fewer, 60 or fewer, 50 or fewer, 40 or fewer, 30 or fewer, 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 nucleotide substitution(s), deletion(s), and/or addition(s).

#### *Non-overlapping sequences*

- 25 In some embodiments, for example in DNA trans-splicing methods, DNA hybrid methods, and mRNA trans-splicing methods, the 5' CDS and the 3' CDS do not share an overlapping sequence. In these embodiments, the total length of the 5' CDS and the 3' CDS combined is equal to the length of a full-length CDS encoding a COL4A3, COL4A4, or COL4A5 polypeptide.

- 30 In some embodiments, the 5' CDS comprises or consists of exons 1 to x of a CDS encoding a COL4A3 polypeptide and the 3' CDS comprises or consists of exons (x+1) to 52 of a CDS encoding a COL4A3 polypeptide, wherein x is from 15 to 43 (e.g. 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42 or 43). In some embodiments, x is from 28 to 34 (e.g. 28, 29, 30, 31, 32, 33 or 34).

In some embodiments, (i) the 5' CDS comprises or consists of exons 1-28 of a CDS encoding a COL4A3 polypeptide and the 3' CDS comprises or consists of exons 29-52 of a CDS encoding a COL4A3 polypeptide; (ii) the 5' CDS comprises or consists of exons 1-29 of a CDS encoding a COL4A3 polypeptide and the 3' CDS comprises or consists of exons 30-52 of a CDS encoding a COL4A3 polypeptide; (iii) the 5' CDS comprises or consists of exons 1-30 of a CDS encoding a COL4A3 polypeptide and the 3' CDS comprises or consists of exons 31-52 of a CDS encoding a COL4A3 polypeptide; (iv) the 5' CDS comprises or consists of exons 1-31 of a CDS encoding a COL4A3 polypeptide and the 3' CDS comprises or consists of exons 32-52 of a CDS encoding a COL4A3 polypeptide; (v) the 5' CDS comprises or consists of exons 1-32 of a CDS encoding a COL4A3 polypeptide and the 3' CDS comprises or consists of exons 33-52 of a CDS encoding a COL4A3 polypeptide; (vi) the 5' CDS comprises or consists of exons 1-33 of a CDS encoding a COL4A3 polypeptide and the 3' CDS comprises or consists of exons 34-52 of a CDS encoding a COL4A3 polypeptide; or (vii) the 5' CDS comprises or consists of exons 1-34 of a CDS encoding a COL4A3 polypeptide and the 3' CDS comprises or consists of exons 35-52 of a CDS encoding a COL4A3 polypeptide.

In some embodiments, the 5' CDS comprises or consists of exons 1 to x of a CDS encoding a COL4A4 polypeptide and the 3' CDS comprises or consists of exons (x+1) to 47 of a CDS encoding a COL4A4 polypeptide, wherein x is from 14 to 39 (e.g. 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38 or 39). In some embodiments, x is from 25 to 31 (e.g. 25, 26, 27, 28, 29, 30 or 31).

In some embodiments, (i) the 5' CDS comprises or consists of exons 1-25 of a CDS encoding a COL4A4 polypeptide and the 3' CDS comprises or consists of exons 26-47 of a CDS encoding a COL4A4 polypeptide; (ii) the 5' CDS comprises or consists of exons 1-26 of a CDS encoding a COL4A4 polypeptide and the 3' CDS comprises or consists of exons 27-47 of a CDS encoding a COL4A4 polypeptide; (iii) the 5' CDS comprises or consists of exons 1-27 of a CDS encoding a COL4A4 polypeptide and the 3' CDS comprises or consists of exons 28-47 of a CDS encoding a COL4A4 polypeptide; (iv) the 5' CDS comprises or consists of exons 1-28 of a CDS encoding a COL4A4 polypeptide and the 3' CDS comprises or consists of exons 29-47 of a CDS encoding a COL4A4 polypeptide; (v) the 5' CDS comprises or consists of exons 1-29 of a CDS encoding a COL4A4 polypeptide and the 3' CDS comprises or consists of exons 30-47 of a CDS encoding a COL4A4 polypeptide; (vi) the 5' CDS comprises or consists of exons 1-30 of a CDS encoding a COL4A4 polypeptide and the 3' CDS comprises or consists of exons 31-47 of a CDS encoding a COL4A4 polypeptide; or (vii) the 5' CDS comprises or consists of exons 1-31 of a CDS encoding a COL4A4 polypeptide and the 3' CDS comprises or consists of exons 32-47 of a CDS encoding a COL4A4 polypeptide.

In some embodiments, the 5' CDS comprises or consists of exons 1 to x of a CDS encoding a COL4A5 polypeptide and the 3' CDS comprises or consists of exons (x-y) to 51 of a CDS encoding a COL4A5 polypeptide, wherein x is from 15 to 41 (e.g. 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40 or 41).

- 5 In some embodiments, the 5' CDS encodes amino acids 1 to x and the 3' CDS encodes amino acids (x+1) to 1670 of SEQ ID NO: 1 or a variant thereof having 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 amino acid substitution(s), wherein x is from 284 to 1297.

- 10 In some embodiments, the 5' CDS encodes amino acids 1 to x and the 3' CDS encodes amino acids (x+1) to 1690 of SEQ ID NO: 8 or a variant thereof having 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 amino acid substitution(s), wherein x is from 304 to 1297.

- 15 In some embodiments, the 5' CDS encodes amino acids 1 to x and the 3' CDS encodes amino acids (x+1) to 1685 of SEQ ID NO: 21 or a variant thereof having 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 amino acid substitution(s), wherein x is from 299 to 1297.

- 20 In some embodiments, the 5' CDS consists of nucleotides 1 to x and the 3' CDS consists of nucleotides (x+1) to 5013 of a CDS encoding a COL4A3 polypeptide, wherein x is from 855 to 3893. In some embodiments, the 5' CDS consists of nucleotides 1 to x and the 3' CDS consists of nucleotides (x+1) to 5013 of SEQ ID NO: 27 or a variant thereof having 1000 or fewer, 500 or fewer, 400 or fewer, 300 or fewer, 200 or fewer, 100 or fewer, 90 or fewer, 80 or fewer, 70 or fewer, 60 or fewer, 50 or fewer, 40 or fewer, 30 or fewer, 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 nucleotide substitution(s), wherein x is from 855 to 3893.

- 25 In some embodiments, (i) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 181 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 202; (ii) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 182 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 203; (iii) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 183 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 204;
- 30 (iv) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 184 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 205; (v) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 185 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 206; (vi) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 186 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 207; or (vii) the 5' CDS



consists of the nucleotide sequence of SEQ ID NO: 187 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 208, or variants thereof having 1000 or fewer, 500 or fewer, 400 or fewer, 300 or fewer, 200 or fewer, 100 or fewer, 90 or fewer, 80 or fewer, 70 or fewer, 60 or fewer, 50 or fewer, 40 or fewer, 30 or fewer, 20 or fewer, 15 or fewer, 10 or fewer,  
 5 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 nucleotide substitution(s).

In some embodiments, the 5' CDS consists of nucleotides 1 to x and the 3' CDS consists of nucleotides (x+1) to 5073 of a CDS encoding a COL4A44 polypeptide, wherein x is from 915 to 3893. In some embodiments, the 5' CDS consists of nucleotides 1 to x and the 3' CDS  
 10 consists of nucleotides (x+1) to 5073 of SEQ ID NO: 81 or a variant thereof having 1000 or fewer, 500 or fewer, 400 or fewer, 300 or fewer, 200 or fewer, 100 or fewer, 90 or fewer, 80 or fewer, 70 or fewer, 60 or fewer, 50 or fewer, 40 or fewer, 30 or fewer, 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 nucleotide substitution(s), wherein x is from 915 to 3893.

15 In some embodiments, (i) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 188 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 209; (ii) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 189 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 210; (iii) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 190 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 211;  
 20 (iv) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 191 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 212; (v) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 192 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 213; (vi) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 193 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 214; or (vii) the 5' CDS  
 25 consists of the nucleotide sequence of SEQ ID NO: 194 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 215, or variants thereof having 1000 or fewer, 500 or fewer, 400 or fewer, 300 or fewer, 200 or fewer, 100 or fewer, 90 or fewer, 80 or fewer, 70 or fewer, 60 or fewer, 50 or fewer, 40 or fewer, 30 or fewer, 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or  
 30 1 nucleotide substitution(s).

In some embodiments, the 5' CDS consists of nucleotides 1 to x and the 3' CDS consists of nucleotides (x+1) to 5058 of a CDS encoding a COL4A5 polypeptide, wherein x is from 900 to 3893. In some embodiments, the 5' CDS consists of nucleotides 1 to x and the 3' CDS consists of nucleotides (x+1) to 5058 of SEQ ID NO: 129 or a variant thereof having 1000 or  
 35 fewer, 500 or fewer, 400 or fewer, 300 or fewer, 200 or fewer, 100 or fewer, 90 or fewer, 80 or

fewer, 70 or fewer, 60 or fewer, 50 or fewer, 40 or fewer, 30 or fewer, 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 nucleotide substitution(s), wherein x is from 900 to 3893.

*Example non-overlapping 5' CDS and 3' CDS sequences (mid-split)*

- 5 In some embodiments, the 5' CDS and the 3' CDS do not share an overlapping sequence and the 5' CDS and the 3' CDS have approximately equal length.

Examples of suitable 5' CDS and 3' CDS sequences for these embodiments are provided in the sections entitled "Example non-overlapping 5' CDS (mid-split)" and "Example non-overlapping 3' CDS (mid-split)", respectively.

- 10 In some embodiments, the 5' CDS comprises or consists of exons 1 to x of a CDS encoding a COL4A5 polypeptide and the 3' CDS comprises or consists of exons (x+1) to 51 of a CDS encoding a COL4A5 polypeptide, wherein x is from 27 to 33 (e.g. 27, 28, 29, 30, 31, 32 or 33). In some embodiments, x is 29, 31, 32, or 33. In some embodiments, x is 31, 32, or 33.

In some embodiments, the 5' CDS and 3' CDS comprise or consist of the combinations of

- 15 exons given in table 2 encoding a COL4A5 polypeptide. Table 2:

| 5' CDS | 3' CDS |  | 5' CDS | 3' CDS |  | 5' CDS | 3' CDS |
|--------|--------|--|--------|--------|--|--------|--------|
| 1-27   | 28-51  |  | 1-30   | 31-51  |  | 1-32   | 33-51  |
| 1-28   | 29-51  |  | 1-31   | 32-51  |  | 1-33   | 34-51  |
| 1-29   | 30-51  |  |        |        |  |        |        |

- In some embodiments, the 5' CDS encodes amino acids 1 to x and the 3' CDS encodes amino acids (x+1) to 1685 of SEQ ID NO: 21 or a variant thereof having 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 amino acid substitution(s), wherein x is from 442 to 1242.
- 20

In some embodiments, the 5' CDS encodes amino acids 1 to x and the 3' CDS encodes amino acids (x+1) to 1685 of SEQ ID NO: 21 or a variant thereof having 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 amino acid substitution(s), wherein x is 892, 922, or 972.

In some embodiments, the 5' CDS consists of nucleotides 1 to x and the 3' CDS consists of nucleotides (x+1) to 5058 of a CDS encoding a COL4A5 polypeptide, wherein x is from 1525 to 3525, from 1625 to 3425, from 1725 to 3325, from 1825 to 3225, from 1925 to 3125, from 2025 to 3025, from 2125 to 2925, or from 2225 to 2825. In some embodiments, the 5' CDS  
5 consists of nucleotides 1 to x and the 3' CDS consists of nucleotides (x+1) to 5058 of SEQ ID NO: 129 or a variant thereof having 1000 or fewer, 500 or fewer, 400 or fewer, 300 or fewer, 200 or fewer, 100 or fewer, 90 or fewer, 80 or fewer, 70 or fewer, 60 or fewer, 50 or fewer, 40 or fewer, 30 or fewer, 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 nucleotide substitution(s), wherein  
10 x is from 1525 to 3525.

In some embodiments, the 5' CDS consists of nucleotides 1 to x and the 3' CDS consists of nucleotides (x+1) to 5058 of SEQ ID NO: 129 or a variant thereof having 1000 or fewer, 500 or fewer, 400 or fewer, 300 or fewer, 200 or fewer, 100 or fewer, 90 or fewer, 80 or fewer, 70 or fewer, 60 or fewer, 50 or fewer, 40 or fewer, 30 or fewer, 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 nucleotide substitution(s), wherein x is 2677, 2767, or 2917.  
15

In some embodiments, the 5' CDS and 3' CDS comprise or consist of the nucleotide sequence pairs given in table 3 or variants thereof having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%,  
20 at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity thereto; or variants thereof having 1000 or fewer, 500 or fewer, 400 or fewer, 300 or fewer, 200 or fewer, 100 or fewer, 90 or fewer, 80 or fewer, 70 or fewer, 60 or fewer, 50 or fewer, 40 or fewer, 30 or fewer, 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 nucleotide substitution(s): Table 3:

| 5' CDS | 3' CDS |  | 5' CDS | 3' CDS |  | 5' CDS | 3' CDS |
|--------|--------|--|--------|--------|--|--------|--------|
| 195    | 216    |  | 198    | 219    |  | 200    | 221    |
| 196    | 217    |  | 199    | 220    |  | 201    | 222    |
| 197    | 218    |  |        |        |  |        |        |

25

*Example non-overlapping 5' CDS and 3' CDS sequences (short/long)*

In some embodiments, the 5' CDS and the 3' CDS do not share an overlapping sequence and the 3' CDS is longer than the 5' CDS.

In some embodiments, the 3' CDS is at least about 1000 nucleotides longer than the 5' CDS, at least about 1100 nucleotides longer than the 5' CDS, at least about 1200 nucleotides longer than the 5' CDS, at least about 1300 nucleotides longer than the 5' CDS, at least about 1400 nucleotides longer than the 5' CDS, or at least about 1500 nucleotides longer than the 5' CDS. In some embodiments, the 3' CDS is about 3000 nucleotides or less longer than the 5' CDS, about 2900 nucleotides or less longer than the 5' CDS, about 2800 nucleotides or less longer than the 5' CDS, about 2700 nucleotides or less longer than the 5' CDS, about 2600 nucleotides or less longer than the 5' CDS, about 2500 nucleotides or less longer than the 5' CDS, about 2400 nucleotides or less longer than the 5' CDS, about 2300 nucleotides or less longer than the 5' CDS, about 2200 nucleotides or less longer than the 5' CDS, about 2100 nucleotides or less longer than the 5' CDS, about 2000 nucleotides or less longer than the 5' CDS, about 1900 nucleotides or less longer than the 5' CDS, about 1800 nucleotides or less longer than the 5' CDS, about 1700 nucleotides or less longer than the 5' CDS, about 1600 nucleotides or less longer than the 5' CDS, or about 1500 nucleotides or less longer than the 5' CDS.

Examples of suitable 5' CDS and 3' CDS sequences for these embodiments are provided in the sections entitled "Example non-overlapping 5' CDS (short)" and "Example non-overlapping 3' CDS (long)", respectively.

In some embodiments, the 5' CDS comprises or consists of exons 1 to x of a CDS encoding a COL4A5 polypeptide and the 3' CDS comprises or consists of exons (x+1) to 51 of a CDS encoding a COL4A5 polypeptide, wherein x is from 18 to 30 (e.g. 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30). In some embodiments, x is from 20 to 28, from 21 to 27, from 22 to 26, from 23 to 25, or 24. In some embodiments, x is from 18 to 20, or 19.

In some embodiments the 5' CDS and 3' CDS comprise or consist of the combinations of exons given in table 4 encoding a COL4A5 polypeptide. Table 4:

| 5' CDS | 3' CDS | 5' CDS | 3' CDS | 5' CDS | 3' CDS |
|--------|--------|--------|--------|--------|--------|
| 1-18   | 19-51  | 1-22   | 23-51  | 1-26   | 27-51  |
| 1-19   | 20-51  | 1-23   | 24-51  | 1-27   | 28-51  |
| 1-20   | 21-51  | 1-24   | 25-51  | 1-28   | 29-51  |

|      |       |  |      |       |  |  |  |
|------|-------|--|------|-------|--|--|--|
| 1-21 | 22-51 |  | 1-25 | 26-51 |  |  |  |
|------|-------|--|------|-------|--|--|--|

In some embodiments, the 5' CDS encodes amino acids 1 to x and the 3' CDS encodes amino acids (x+1) to 1685 of SEQ ID NO: 21 or a variant thereof having 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 amino acid substitution(s), wherein x is from 350 to 850. In some embodiments, x is 593. In some embodiments, x is 388.

In some embodiments, the 5' CDS consists of nucleotides 1 to x and the 3' CDS consists of nucleotides (x+1) to 5058 of a CDS encoding a COL4A5 polypeptide, wherein x is from 1000 to 2800. In some embodiments, the 5' CDS consists of nucleotides 1 to x and the 3' CDS consists of nucleotides (x+1) to 5058 of SEQ ID NO: 129 or a variant thereof having 1000 or fewer, 500 or fewer, 400 or fewer, 300 or fewer, 200 or fewer, 100 or fewer, 90 or fewer, 80 or fewer, 70 or fewer, 60 or fewer, 50 or fewer, 40 or fewer, 30 or fewer, 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 nucleotide substitution(s), wherein x is from 1000 to 2800. In some embodiments, x is 1779. In some embodiments, x is 1165.

In some embodiments, the 5' CDS and 3' CDS consist of the nucleotide sequence pairs described in table 5; or variants thereof having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity thereto; or variants thereof having 1000 or fewer, 500 or fewer, 400 or fewer, 300 or fewer, 200 or fewer, 100 or fewer, 90 or fewer, 80 or fewer, 70 or fewer, 60 or fewer, 50 or fewer, 40 or fewer, 30 or fewer, 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 nucleotide substitution(s). Table 5:

|        |        |  |        |        |  |        |        |
|--------|--------|--|--------|--------|--|--------|--------|
| 5' CDS | 3' CDS |  | 5' CDS | 3' CDS |  | 5' CDS | 3' CDS |
| 196    | 217    |  | 279    | 294    |  | 276    | 291    |
| 159    | 216    |  | 278    | 293    |  | 275    | 290    |
| 280    | 295    |  | 277    | 292    |  | 274    | 289    |

Example non-overlapping 5' CDS and 3' CDS sequences (long/short)

In some embodiments, the 5' CDS and the 3' CDS do not share an overlapping sequence and the 5' CDS is longer than the 3' CDS.

In some embodiments, the 5' CDS is at least about 1000 nucleotides longer than the 3' CDS, about 1100 nucleotides longer than the 3' CDS, about 1200 nucleotides longer than the 3' CDS, about 1300 nucleotides longer than the 3' CDS, about 1400 nucleotides longer than the 3' CDS, at least about 1500 nucleotides longer than the 3' CDS, about 1600 nucleotides longer than the 3' CDS, about 1700 nucleotides longer than the 3' CDS, about 1800 nucleotides longer than the 3' CDS, about 1900 nucleotides longer than the 3' CDS, at least about 2000 nucleotides longer than the 3' CDS, at least about 2100 nucleotides longer than the 3' CDS, at least about 2200 nucleotides longer than the 3' CDS, at least about 2300 nucleotides longer than the 3' CDS, at least about 2400 nucleotides longer than the 3' CDS, or at least about 2500 nucleotides longer than the 3' CDS. In some embodiments, the 5' CDS about 3000 nucleotides or less longer than the 3' CDS, about 2900 nucleotides or less longer than the 3' CDS, about 2800 nucleotides or less longer than the 3' CDS, about 2700 nucleotides or less longer than the 3' CDS, or about 2600 nucleotides or less longer than the 3' CDS.

Examples of suitable 5' CDS and 3' CDS sequences for these embodiments are provided in the sections entitled "Example non-overlapping 5' CDS (long)" and "Example non-overlapping 3' CDS (short)", respectively.

In some embodiments, the 5' CDS comprises or consists of exons 1 to x of a CDS encoding a COL4A5 polypeptide and the 3' CDS comprises or consists of exons (x+1) to 51 of a CDS encoding a COL4A5 polypeptide, wherein x is from 36 to 41 (e.g. 36, 37, 38, 39, 40, 41). In some embodiments, x is from 37 to 41, from 38 to 41, from 39 to 41, from 40 to 41, or 41.

In some embodiments, (i) the 5' CDS comprises or consists of exons 1-36 of a CDS encoding a COL4A5 polypeptide and the 3' CDS comprises or consists of exons 37-51 of a CDS encoding a COL4A5 polypeptide; (ii) the 5' CDS comprises or consists of exons 1-37 of a CDS encoding a COL4A5 polypeptide and the 3' CDS comprises or consists of exons 38-51 of a CDS encoding a COL4A5 polypeptide; (iii) the 5' CDS comprises or consists of exons 1-38 of a CDS encoding a COL4A5 polypeptide and the 3' CDS comprises or consists of exons 39-51 of a CDS encoding a COL4A5 polypeptide; (iv) the 5' CDS comprises or consists of exons 1-39 of a CDS encoding a COL4A5 polypeptide and the 3' CDS comprises or consists of exons 40-51 of a CDS encoding a COL4A5 polypeptide; (v) the 5' CDS comprises or consists of exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS comprises or consists of exons 41-51 of a CDS encoding a COL4A5 polypeptide; or (vi) the 5' CDS comprises or

consists of exons 1-41 of a CDS encoding a COL4A5 polypeptide and the 3' CDS comprises or consists of exons 42-51 of a CDS encoding a COL4A5 polypeptide.

In some embodiments, the 5' CDS encodes amino acids 1 to x and the 3' CDS encodes amino acids (x+1) to 1685 of SEQ ID NO: 21 or a variant thereof having 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 amino acid substitution(s), wherein x is from 1000 to 1297. In some embodiments, x is 1263.

In some embodiments, the 5' CDS consists of nucleotides 1 to x and the 3' CDS consists of nucleotides (x+1) to 5058 of a CDS encoding a COL4A5 polypeptide, wherein x is from 3000 to 3893. In some embodiments, the 5' CDS consists of nucleotides 1 to x and the 3' CDS consists of nucleotides (x+1) to 5058 of SEQ ID NO: 129 or a variant thereof having 1000 or fewer, 500 or fewer, 400 or fewer, 300 or fewer, 200 or fewer, 100 or fewer, 90 or fewer, 80 or fewer, 70 or fewer, 60 or fewer, 50 or fewer, 40 or fewer, 30 or fewer, 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 nucleotide substitution(s), wherein x is from 3000 to 3893. In some embodiments, x is 3790.

In some embodiments, the 5' CDS and 3' CDS consist of the nucleotide sequence pairs described in table 6; or variants thereof having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity thereto; or variants thereof having 1000 or fewer, 500 or fewer, 400 or fewer, 300 or fewer, 200 or fewer, 100 or fewer, 90 or fewer, 80 or fewer, 70 or fewer, 60 or fewer, 50 or fewer, 40 or fewer, 30 or fewer, 20 or fewer, 15 or fewer, 10 or fewer, 9 or fewer, 8 or fewer, 7 or fewer, 6 or fewer, 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 nucleotide substitution(s). Table 6:

| 5' CDS | 3' CDS |  | 5' CDS | 3' CDS |  | 5' CDS | 3' CDS |
|--------|--------|--|--------|--------|--|--------|--------|
| 283    | 298    |  | 285    | 300    |  | 287    | 302    |
| 284    | 299    |  | 286    | 301    |  | 288    | 303    |

Adeno-associated viral (AAV) vectors

The 5' AAV vector and/or the 3' AAV vector may each be in the form of an AAV vector particle. In some embodiments, the 5' AAV vector and the 3' AAV vector are each in the form of an AAV vector particle.

*AAV genome*

- 5 An AAV vector or AAV vector particle may comprise an AAV genome or a fragment or derivative thereof.

An AAV genome is a polynucleotide sequence, which may encode functions needed for production of an AAV particle. These functions include those operating in the replication and packaging cycle of AAV in a host cell, including encapsidation of the AAV genome into an AAV  
10 particle. Naturally occurring AAVs are replication-deficient and rely on the provision of helper functions in trans for completion of a replication and packaging cycle. Accordingly, the AAV genome of an AAV vector of the invention is typically replication-deficient.

The AAV genome may be in single-stranded form (ssAAV), either positive or negative-sense, or alternatively in double-stranded form (dsAAV). The use of a double-stranded form allows  
15 bypass of the DNA replication step in the target cell and so can accelerate transgene expression. The maximum packaging capacity of the single-stranded form is larger than the double-stranded form. Suitably, the AAV genome is in single-stranded form.

AAVs occurring in nature may be classified according to various biological systems. The AAV genome may be from any naturally derived serotype, isolate or clade of AAV.

- 20 AAV may be referred to in terms of their serotype. A serotype corresponds to a variant subspecies of AAV which, owing to its profile of expression of capsid surface antigens, has a distinctive reactivity which can be used to distinguish it from other variant subspecies. Typically, an AAV vector particle having a particular AAV serotype does not efficiently cross-react with neutralising antibodies specific for any other AAV serotype. AAV serotypes include  
25 AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10 and AAV11, and derivatives thereof.

The 5' AAV vector genome and the 3' AAV vector genome may each be the same or different serotypes. In some embodiments, the 5' AAV vector genome and the 3' AAV vector genome are the same serotype. Suitably, the AAV serotype is selected from AAV1, AAV2, AAV3,  
30 AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10 and AAV11, and derivatives thereof.

AAV may also be referred to in terms of clades or clones. This refers to the phylogenetic relationship of naturally derived AAVs, and typically to a phylogenetic group of AAVs which



can be traced back to a common ancestor, and includes all descendants thereof. Additionally, AAVs may be referred to in terms of a specific isolate, i.e. a genetic isolate of a specific AAV found in nature. The term genetic isolate describes a population of AAVs which has undergone limited genetic mixing with other naturally occurring AAVs, thereby defining a recognisably distinct population at a genetic level.

Typically, the AAV genome of a naturally derived serotype, isolate or clade of AAV comprises at least one inverted terminal repeat sequence (ITR). An ITR sequence acts in cis to provide a functional origin of replication and allows for integration and excision of the vector from the genome of a cell. ITRs may be the only sequences required in cis next to the therapeutic gene.

- 10 An AAV genome may also comprise packaging genes, such as rep and/or cap genes which encode packaging functions for an AAV particle. A promoter may be operably linked to each of the packaging genes. Specific examples of such promoters include the p5, p19 and p40 promoters. For example, the p5 and p19 promoters are generally used to express the rep gene, while the p40 promoter is generally used to express the cap gene. The rep gene  
15 encodes one or more of the proteins Rep78, Rep68, Rep52 and Rep40 or variants thereof. The cap gene encodes one or more capsid proteins such as VP1, VP2 and VP3 or variants thereof. These proteins make up the capsid of an AAV particle, which determines the AAV serotype. VP1, VP2, and VP3 may be produced by alternate mRNA splicing (Trempe, J.P. and Carter, B.J., 1988. Journal of virology, 62(9), pp.3356-3363). Thus, VP1, VP2 and VP3  
20 may have identical sequences, but wherein VP2 is truncated at the N-terminus relative to VP1, and VP3 is truncated at the N-terminus relative to VP2.

- An AAV genome may be the full genome of a naturally occurring AAV. For example, a vector comprising a full AAV genome may be used to prepare an AAV vector or vector particle. Preferably, an AAV genome is derivatised for the purpose of administration to patients. Such  
25 derivatisation is standard in the art and the invention encompasses the use of any known derivative of an AAV genome, and derivatives which could be generated by applying techniques known in the art. An AAV genome may be a derivative of any naturally occurring AAV. Suitably, the AAV genome is a derivative of AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, or AAV11. Suitably, the AAV genome is a derivative of AAV2.

- 30 The 5' AAV vector genome and the 3' AAV vector genome may each be derived from the same or different naturally occurring AAVs. In some embodiments, the 5' AAV vector genome and the 3' AAV vector genome are each derived from the same naturally occurring AAV. Suitably, the AAV genome is a derivative of AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, or AAV11. Suitably, the AAV genome is a derivative of AAV2.

Derivatives of an AAV genome include any truncated or modified forms of an AAV genome which allow for expression of a transgene from an AAV vector of the invention *in vivo*. Typically, it is possible to truncate the AAV genome significantly to include minimal viral sequence yet retain the above function. This is preferred for safety reasons to reduce the risk of recombination of the vector with wild-type virus, and also to avoid triggering a cellular immune response by the presence of viral gene proteins in the target cell.

Typically, a derivative will include at least one inverted terminal repeat sequence (ITR), preferably more than one ITR, such as two ITRs or more. One or more of the ITRs may be derived from AAV genomes having different serotypes, or may be a chimeric or mutant ITR. A preferred mutant ITR is one having a deletion of a trs (terminal resolution site). This deletion allows for continued replication of the genome to generate a single-stranded genome which contains both coding and complementary sequences, i.e. a self-complementary AAV (scAAV) genome. This allows for bypass of DNA replication in the target cell, and so enables accelerated transgene expression. However, the maximum packaging capacity of a scAAV is reduced. Suitably, the AAV genome is not a scAAV genome.

Suitably, the one or more ITRs flank the AAV genome at either end. The inclusion of one or more ITRs is may aid concatamer formation of an AAV vector in the nucleus of a host cell, for example following the conversion of single-stranded vector DNA into double-stranded DNA by the action of host cell DNA polymerases. The formation of such episomal concatamers protects the AAV vector during the life of the host cell, thereby allowing for prolonged expression of the protein-coding sequence *in vivo*. Suitably, an AAV genome may comprise a 5' ITR and a 3' ITR.

The AAV genome may comprise one or more ITR sequences from any naturally derived serotype, isolate or clone of AAV or a variant thereof. The AAV genome may comprise at least one, such as two, AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, or AAV11 ITRs, or variants thereof. Suitably, the AAV genome may comprise at least one, such as two, AAV2 ITRs.

The 5' AAV vector genome and the 3' AAV vector genome may comprise the same or different ITRs. In some embodiments, the 5' AAV vector genome and the 3' AAV vector genome each comprise one or more ITR sequences which are the same. Suitably, the 5' AAV vector genome and the 3' AAV vector genome each comprise at least one, such as two, AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, or AAV11 ITRs, or variants thereof. Suitably, the 5' AAV vector genome and the 3' AAV vector genome each comprise at least one, such as two, AAV2 ITRs.

ITR sequences for use in the invention may be derived from the following accession numbers for AAV whole genome sequences: Adeno-associated virus 1 NC\_002077; Adeno-associated virus 2 NC\_001401; Adeno-associated virus 3 NC\_001729; Adeno-associated virus 4 NC\_001829; Adeno-associated virus 5 NC\_006152; Adeno-associated virus 6 AF028704; 5 Adeno-associated virus 7 NC\_006260; and Adeno-associated virus 8 NC\_006261. Other suitable ITRs are described in Wilmott, P., et al., 2019. Human gene therapy methods, 30(6), pp.206-213.

The inclusion of one or more ITRs is preferred to aid concatamer formation of the AAV vector in the nucleus of a host cell, for example following the conversion of single-stranded vector 10 DNA into double-stranded DNA by the action of host cell DNA polymerases. The formation of such episomal concatamers protects the AAV vector during the life of the host cell, thereby allowing for prolonged expression of the transgene in vivo.

Suitably, ITR elements will be the only sequences retained from the native AAV genome in the derivative. A derivative will preferably not include the rep and/or cap genes of the native 15 genome and any other sequences of the native genome. This is preferred for the reasons described above, and also to reduce the possibility of integration of the vector into the host cell genome. Additionally, reducing the size of the AAV genome allows for increased flexibility in incorporating other sequence elements (such as regulatory elements) within the vector in addition to the protein-coding sequence.

20 The following portions could therefore be removed in a derivative of the invention: one inverted terminal repeat (ITR) sequence, the replication (rep) and capsid (cap) genes. However, derivatives may additionally include one or more rep and/or cap genes or other viral sequences of an AAV genome. Naturally occurring AAV integrates with a high frequency at a specific site on human chromosome 19, and shows a negligible frequency of random 25 integration, such that retention of an integrative capacity in the AAV vector may be tolerated in a therapeutic setting.

Suitably, the 5' AAV vector comprises a 5' ITR and a 3' ITR. Suitably, the 5' AAV vector comprises from 5' to 3': a 5' ITR, a promoter, optionally one or more regulatory sequences, the 5' CDS, and a 3' ITR. Suitably, the 3' AAV vector comprises a 5' ITR and a 3' ITR. Suitably, 30 the 3' AAV vector comprises from 5' to 3': a 3' ITR, the 3' CDS, optionally one or more regulatory sequences, and a 3' ITR.

The invention additionally encompasses the provision of sequences of an AAV genome in a different order and configuration to that of a native AAV genome. The invention also encompasses the replacement of one or more AAV sequences or genes with sequences

from another virus or with chimeric genes composed of sequences from more than one virus. Such chimeric genes may be composed of sequences from two or more related viral proteins of different viral species.

*AAV serotype and capsid proteins*

- 5 An AAV vector particle may be encapsidated by capsid proteins. The serotype may facilitate the transduction of glomerular cells (e.g. podocytes), for example specific transduction of glomerular cells (e.g. podocytes).

- The AAV vector particles may be a kidney-specific vector particle. Preferably, the AAV vector particles are glomerular-specific (e.g. podocyte-specific) vector particle. The AAV vector  
10 particles may be encapsidated by a glomerular-specific (e.g. podocyte-specific) capsid. The AAV vector particles may comprise a glomerular-specific (e.g. podocyte-specific) capsid protein.

- Suitably, the AAV vector particles may be transcapsidated forms wherein an AAV genome or derivative having an ITR of one serotype is packaged in the capsid of a different serotype. The  
15 AAV vector particle also includes mosaic forms wherein a mixture of unmodified capsid proteins from two or more different serotypes makes up the viral capsid. The AAV vector particle also includes chemically modified forms bearing ligands adsorbed to the capsid surface. For example, such ligands may include antibodies for targeting a particular cell surface receptor.

- 20 Where a derivative comprises capsid proteins i.e. VP1, VP2 and/or VP3, the derivative may be a chimeric, shuffled or capsid-modified derivative of one or more naturally occurring AAVs. In particular, the invention encompasses the provision of capsid protein sequences from different serotypes, clades, clones, or isolates of AAV within the same vector (i.e. a pseudotyped vector). An AAV vector may be in the form of a pseudotyped AAV vector particle.

- 25 Chimeric, shuffled or capsid-modified derivatives will be typically selected to provide one or more desired functionalities for the AAV vector. Thus, these derivatives may display increased efficiency of gene delivery, decreased immunogenicity (humoral or cellular), an altered tropism range and/or improved targeting of podocytes compared to an AAV vector comprising a naturally occurring AAV genome. Increased efficiency of gene delivery may be effected by  
30 improved receptor or co-receptor binding at the cell surface, improved internalisation, improved trafficking within the cell and into the nucleus, improved uncoating of the viral particle and improved conversion of a single-stranded genome to double-stranded form. Increased

efficiency may also relate to an altered tropism range or targeting of podocytes, such that the vector dose is not diluted by administration to tissues where it is not needed.

Chimeric capsid proteins include those generated by recombination between two or more capsid coding sequences of naturally occurring AAV serotypes. This may be performed for example by a marker rescue approach in which non-infectious capsid sequences of one serotype are co-transfected with capsid sequences of a different serotype, and directed selection is used to select for capsid sequences having desired properties. The capsid sequences of the different serotypes can be altered by homologous recombination within the cell to produce novel chimeric capsid proteins.

Chimeric capsid proteins also include those generated by engineering of capsid protein sequences to transfer specific capsid protein domains, surface loops or specific amino acid residues between two or more capsid proteins, for example between two or more capsid proteins of different serotypes.

Shuffled or chimeric capsid proteins may also be generated by DNA shuffling or by error-prone PCR. Hybrid AAV capsid genes can be created by randomly fragmenting the sequences of related AAV genes e.g. those encoding capsid proteins of multiple different serotypes and then subsequently reassembling the fragments in a self-priming polymerase reaction, which may also cause crossovers in regions of sequence homology. A library of hybrid AAV genes created in this way by shuffling the capsid genes of several serotypes can be screened to identify viral clones having a desired functionality. Similarly, error prone PCR may be used to randomly mutate AAV capsid genes to create a diverse library of variants which may then be selected for a desired property.

The sequences of the capsid genes may also be genetically modified to introduce specific deletions, substitutions or insertions with respect to the native wild-type sequence. In particular, capsid genes may be modified by the insertion of a sequence of an unrelated protein or peptide within an open reading frame of a capsid coding sequence, or at the N- and/or C-terminus of a capsid coding sequence. The unrelated protein or peptide may advantageously be one which acts as a ligand for a particular cell type, thereby conferring improved binding to a target cell or improving the specificity of targeting of the vector to a particular cell population. The unrelated protein may also be one which assists purification of the viral particle as part of the production process, i.e. an epitope or affinity tag. The site of insertion will typically be selected so as not to interfere with other functions of the viral particle e.g. internalisation, trafficking of the viral particle.

The capsid protein may be an artificial or mutant capsid protein. The term "artificial capsid" as used herein means that the capsid particle comprises an amino acid sequence which does not occur in nature or which comprises an amino acid sequence which has been engineered (e.g. modified) from a naturally occurring capsid amino acid sequence. In other words the artificial capsid protein comprises a mutation or a variation in the amino acid sequence compared to the sequence of the parent capsid from which it is derived where the artificial capsid amino acid sequence and the parent capsid amino acid sequences are aligned.

The capsid protein may comprise a mutation or modification relative to the wild type capsid protein which improves the ability to transduce podocytes relative to an unmodified or wild type viral particle. Improved ability to transduce podocytes may be measured for example by measuring the expression of a reporter transgene, e.g. GFP, carried by the AAV vector particle, wherein expression of the transgene in podocytes correlates with the ability of the AAV vector particle to transduce podocytes.

The 5' AAV vector genome and the 3' AAV vector genome may each comprises the same or different capsid proteins. In some embodiments, the 5' AAV vector and the 3' AAV vector are comprise the same capsid proteins.

The 5' AAV vector and/or the 3' AAV vector may be an AAV3B, LK03, AAV9, ShH10, AAV-DJ, AAV2, AAV6.2, AAV5, or AAV8 vector particle. In some embodiments, the 5' AAV vector and the 3' AAV vector is an AAV3B, LK03, AAV9, ShH10, AAV-DJ, AAV2, AAV6.2, AAV5, or AAV8 vector particle. In some embodiments, the 5' AAV vector and/or the 3' AAV vector is an AAV3B vector particle or an LK03 vector particle. In one embodiment, the 5' AAV vector and/or the 3' AAV vector is an LK03 vector particle. In other embodiments, the 5' AAV vector and/or the 3' AAV vector is a ShH10, AAV-DJ, AAV2, AAV6.2, or AAV5 vector particle. In some embodiments, the 5' AAV vector and the 3' AAV vector is an AAV3B vector particle or an LK03 vector particle. In one embodiment, the 5' AAV vector and the 3' AAV vector is an LK03 vector particle. In other embodiments, the 5' AAV vector and the 3' AAV vector is a ShH10, AAV-DJ, AAV2, AAV6.2, or AAV5 vector particle.

The 5' AAV vector particle and/or the 3' AAV vector particle may comprise an AAV3B, LK03, AAV9, ShH10, AAV-DJ, AAV2, AAV6.2, AAV5, or AAV8 capsid protein. In some embodiments, the 5' AAV vector particle and/or the 3' AAV vector particle each comprise an AAV3B, LK03, AAV9, ShH10, AAV-DJ, AAV2, AAV6.2, AAV5, or AAV8 capsid protein. In some embodiments, the 5' AAV vector particle and/or the 3' AAV vector particle comprises an AAV3B capsid protein or an LK03 capsid protein. In one embodiment, the 5' AAV vector particle and/or the 3' AAV vector particle comprises an LK03 capsid protein. In other

embodiments, the 5' AAV vector particle and/or the 3' AAV vector particle comprises a ShH10, AAV-DJ, AAV2, AAV6.2, or AAV5 capsid protein. In some embodiments, the 5' AAV vector particle and the 3' AAV vector particle comprises an AAV3B capsid protein or an LK03 capsid protein. In one embodiment, the 5' AAV vector particle and the 3' AAV vector particle  
5 comprises an LK03 capsid protein. In other embodiments, the 5' AAV vector particle and the 3' AAV vector particle comprises a ShH10, AAV-DJ, AAV2, AAV6.2, or AAV5 capsid protein.

The 5' AAV vector particle and/or the 3' AAV vector particle may comprise AAV3B, LK03, AAV9, ShH10, AAV-DJ, AAV2, AAV6.2, AAV5, or AAV8 capsid proteins VP1, VP2 and VP3. In some embodiments, the 5' AAV vector particle and the 3' AAV vector particle each comprise  
10 AAV3B, LK03, AAV9, ShH10, AAV-DJ, AAV2, AAV6.2, AAV5, or AAV8 capsid proteins VP1, VP2 and VP3. In some embodiments, the 5' AAV vector particle and/or the 3' AAV vector particle comprises AAV3B or LK03 capsid proteins VP1, VP2 and VP3. In one embodiment, the 5' AAV vector particle and/or the 3' AAV vector particle comprises LK03 capsid proteins VP1, VP2 and VP3. In other embodiments, the 5' AAV vector particle and/or the 3' AAV vector  
15 particle comprises ShH10, AAV-DJ, AAV2, AAV6.2, or AAV5 capsid proteins VP1, VP2 and VP3. In some embodiments, the 5' AAV vector particle and the 3' AAV vector particle comprises AAV3B or LK03 capsid proteins VP1, VP2 and VP3. In one embodiment, the 5' AAV vector particle and the 3' AAV vector particle comprises LK03 capsid proteins VP1, VP2 and VP3. In other embodiments, the 5' AAV vector particle and the 3' AAV vector particle  
20 comprises ShH10, AAV-DJ, AAV2, AAV6.2, or AAV5 capsid proteins VP1, VP2 and VP3.

The 5' AAV vector particle and/or the 3' AAV vector particle may comprise one or more AAV2 ITR sequences and AAV3B capsid proteins, LK03 capsid proteins, AAV9 capsid proteins, ShH10 capsid proteins, AAV-DJ capsid proteins, AAV2 capsid proteins, AAV6.2 capsid proteins, AAV5 capsid proteins, or AAV8 capsid proteins. In some embodiments, the 5' AAV  
25 vector particle and the 3' AAV vector particle each comprise one or more AAV2 ITR sequences and AAV3B capsid proteins, LK03 capsid proteins, AAV9 capsid proteins, ShH10 capsid proteins, AAV-DJ capsid proteins, AAV2 capsid proteins, AAV6.2 capsid proteins, AAV5 capsid proteins, or AAV8 capsid proteins. In some embodiments, the 5' AAV vector particle and/or the 3' AAV vector particle comprises one or more AAV2 ITR sequences and AAV3B or  
30 LK03 capsid proteins. In other embodiments, the 5' AAV vector particle and/or the 3' AAV vector particle comprises one or more AAV2 ITR sequences and ShH10, AAV-DJ, AAV2, AAV6.2, or AAV5 capsid proteins. In some embodiments, the 5' AAV vector particle and the 3' AAV vector particle comprises one or more AAV2 ITR sequences and AAV3B or LK03 capsid proteins. In other embodiments, the 5' AAV vector particle and the 3' AAV vector

particle comprises one or more AAV2 ITR sequences and ShH10, AAV-DJ, AAV2, AAV6.2, or AAV5 capsid proteins.

- The 5' AAV vector particle and/or the 3' AAV vector particle may have an AAV2 genome and AAV3B capsid proteins (AAV2/3B), an AAV2 genome and LK03 capsid proteins, an AAV2 genome and AAV9 capsid proteins (AAV2/9), an AAV2 genome and ShH10 capsid proteins, an AAV2 genome and AAV-DJ capsid proteins, an AAV2 genome and AAV2 capsid proteins, an AAV2 genome and AAV6.2 capsid proteins, an AAV2 genome and AAV5 capsid proteins, or an AAV2 genome and AAV8 capsid proteins (AAV2/8). In some embodiments, the 5' AAV vector particle and the 3' AAV vector particle each have an AAV2 genome and AAV3B capsid proteins (AAV2/3B), an AAV2 genome and LK03 capsid proteins, an AAV2 genome and AAV9 capsid proteins (AAV2/9), an AAV2 genome and ShH10 capsid proteins, an AAV2 genome and AAV-DJ capsid proteins, an AAV2 genome and AAV2 capsid proteins, an AAV2 genome and AAV6.2 capsid proteins, an AAV2 genome and AAV5 capsid proteins, or an AAV2 genome and AAV8 capsid proteins (AAV2/8).
- The nomenclature AAVX/Y may denote a pseudotyped AAV, for example where the ITR sequences are from AAVX and flank a cassette harbouring a payload which is encapsidated into serotype AAVY (i.e. with AAVY capsid proteins).

#### *LK03 serotype*

- The 5' AAV vector particle and/or the 3' AAV vector particle may comprise a LK03 capsid protein. Suitably, the 5' AAV vector particle and/or the 3' AAV vector particle may be encapsidated by LK03 capsid proteins.

In some embodiments, the 5' AAV vector particle and the 3' AAV vector particle each comprise a LK03 capsid protein. In some embodiments, the 5' AAV vector particle and the 3' AAV vector particle are both encapsidated by LK03 capsid proteins.

- The AAV-LK03 cap sequence consists of fragments from seven different wild-type serotypes (AAV1, 2, 3B, 4, 6, 8, 9) and is described in Lisowski, L., et al., 2014. Nature, 506(7488), pp.382-386. The present inventors have demonstrated that AAV-LK03 vectors can achieve high transduction in human podocytes *in vitro*.

- The 5' AAV vector particle and/or the 3' AAV vector particle may comprise an LK03 VP1 capsid protein, an LK03 VP2 capsid protein, and/or an LK03 VP3 capsid protein. Suitably, the 5' AAV vector particle and/or the 3' AAV vector particle may be encapsidated by LK03 VP1 capsid proteins, LK03 VP2 capsid proteins, and/or LK03 VP3 capsid proteins. Suitably, the 5' AAV



vector particle and/or the 3' AAV vector particle may be encapsidated by LK03 VP1, VP2, and VP3 capsid proteins.

Suitably, the LK03 VP1 capsid protein may comprise or consist of the amino acid sequence shown as SEQ ID NO: 264, or a variant which is at least 90% identical to SEQ ID NO: 264.

- 5 Suitably, the variant may be at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 264.

Suitably, the LK03 VP2 and VP3 capsid proteins may be N-terminal truncations of SEQ ID NO: 264, or N-terminal truncations of a variant which is at least 90% identical, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 264.

10 *AAV3B serotype*

The 5' AAV vector particle and/or the 3' AAV vector particle may comprise an AAV3B capsid protein. Suitably, the 5' AAV vector particle and/or the 3' AAV vector particle may be encapsidated by AAV3B capsid proteins.

- 15 In some embodiments, the 5' AAV vector particle and the 3' AAV vector particle each comprise an AAV3B capsid protein. In some embodiments, the 5' AAV vector particle and the 3' AAV vector particle are both encapsidated by AAV3B capsid proteins.

- 20 Two distinct AAV3 isolates (AAV3A and AAV3B) have been cloned. In comparison with vectors based on other AAV serotypes, it is thought that AAV3 vectors inefficiently transduce most cell types. However, AAV3B may efficiently transduce podocytes. AAV3B has been described in Rutledge, E.A., et al., 1998. Journal of virology, 72(1), pp.309-319.

- 25 The AAV vector particle may comprise an AAV3B VP1 capsid protein, an AAV3B VP2 capsid protein, and/or an AAV3B VP3 capsid protein. Suitably, the AAV vector particle may be encapsidated by AAV3B VP1 capsid proteins, AAV3B VP2 capsid proteins, and/or AAV3B VP3 capsid proteins. Suitably, the AAV vector particle may be encapsidated by AAV3B VP1, VP2, and VP3 capsid proteins.

Suitably, the AAV3B VP1 capsid protein may comprise or consist of the amino acid sequence shown as SEQ ID NO: 265, or a variant which is at least 90% identical to SEQ ID NO: 265.

Suitably, the variant may be at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 265.

Suitably, the AAV3B VP2 and VP3 capsid proteins may be N-terminal truncations of SEQ ID NO: 265, or N-terminal truncations of a variant which is at least 90% identical, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 265.

#### *AAV9 serotype*

- 5 The 5' AAV vector particle and/or the 3' AAV vector particle may comprise an AAV9 capsid protein. Suitably, the 5' AAV vector particle and/or the 3' AAV vector particle may be encapsidated by AAV9 capsid proteins.

In some embodiments, the 5' AAV vector particle and the 3' AAV vector particle each comprise an AAV9 capsid protein. In some embodiments, the 5' AAV vector particle and the 3' AAV  
10 vector particle are both encapsidated by AAV9 capsid proteins.

The 5' AAV vector particle and/or the 3' AAV vector particle may comprise an AAV9 VP1 capsid protein, an AAV9 VP2 capsid protein, and/or an AAV9 VP3 capsid protein. Suitably, the 5' AAV vector particle and/or the 3' AAV vector particle may be encapsidated by AAV9 VP1 capsid proteins, AAV9 VP2 capsid proteins, and/or AAV9 VP3 capsid proteins. Suitably,  
15 the 5' AAV vector particle and/or the 3' AAV vector particle may be encapsidated by AAV9 VP1, VP2, and VP3 capsid proteins.

Suitably, the AAV9 VP1 capsid protein may comprise or consist of the amino acid sequence shown as SEQ ID NO: 266, or a variant which is at least 90% identical to SEQ ID NO: 266.

Suitably, the variant may be at least 95%, at least 96%, at least 97%, at least 98% or at least  
20 99% identical to SEQ ID NO: 266.

Suitably, the AAV9 VP2 and VP3 capsid proteins may be N-terminal truncations of SEQ ID NO: 266, or N-terminal truncations of a variant which is at least 90% identical, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 266.

#### *ShH10 serotype*

- 25 The 5' AAV vector particle and/or the 3' AAV vector particle may comprise a ShH10 capsid protein. Suitably, the 5' AAV vector particle and/or the 3' AAV vector particle may be encapsidated by ShH10 capsid proteins.

In some embodiments, the 5' AAV vector particle and the 3' AAV vector particle each comprise a ShH10 capsid protein. In some embodiments, the 5' AAV vector particle and the 3' AAV  
30 vector particle are both encapsidated by ShH10 capsid proteins.

The ShH10 variant is derived from AAV6 and has increased specificity and efficiency for Müller cells (see e.g. Klimczak, R.R., et al., 2009. PloS one, 4(10), p.e7467).

The 5' AAV vector particle and/or the 3' AAV vector particle may comprise a ShH10 VP1 capsid protein, a ShH10 VP2 capsid protein, and/or a ShH10 VP3 capsid protein. Suitably, the 5' AAV vector particle and/or the 3' AAV vector particle may be encapsidated by ShH10 VP1 capsid proteins, ShH10 VP2 capsid proteins, and/or ShH10 VP3 capsid proteins. Suitably, the 5' AAV vector particle and/or the 3' AAV vector particle may be encapsidated by ShH10 VP1, VP2, and VP3 capsid proteins.

Suitably, the ShH10 VP1 capsid protein may comprise or consist of the amino acid sequence shown as SEQ ID NO: 267, or a variant which is at least 90% identical to SEQ ID NO: 267.

Suitably, the variant may be at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 267.

Suitably, the ShH10 VP2 and VP3 capsid proteins may be N-terminal truncations of SEQ ID NO: 267, or N-terminal truncations of a variant which is at least 90% identical, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 267.

#### *AAV-DJ serotype*

The 5' AAV vector particle and/or the 3' AAV vector particle may comprise an AAV-DJ capsid protein. Suitably, the 5' AAV vector particle and/or the 3' AAV vector particle may be encapsidated by AAV-DJ capsid proteins.

In some embodiments, the 5' AAV vector particle and the 3' AAV vector particle each comprise an AAV-DJ capsid protein. In some embodiments, the 5' AAV vector particle and the 3' AAV vector particle are both encapsidated by AAV-DJ capsid proteins.

AAV-DJ is a hybrid vector created from DNA shuffling of eight AAV serotypes, which mediates efficient gene expression both in vitro and in vivo (see e.g. Mao, Y., et al., 2016. BMC biotechnology, 16, pp.1-8).

The 5' AAV vector particle and/or the 3' AAV vector particle may comprise an AAV-DJ VP1 capsid protein, an AAV-DJ VP2 capsid protein, and/or an AAV-DJ VP3 capsid protein. Suitably, the 5' AAV vector particle and/or the 3' AAV vector particle may be encapsidated by AAV-DJ VP1 capsid proteins, AAV-DJ VP2 capsid proteins, and/or AAV-DJ VP3 capsid proteins. Suitably, the 5' AAV vector particle and/or the 3' AAV vector particle may be encapsidated by AAV-DJ VP1, VP2, and VP3 capsid proteins.

Suitably, the AAV-DJ VP1 capsid protein may comprise or consist of the amino acid sequence shown as SEQ ID NO: 268, or a variant which is at least 90% identical to SEQ ID NO: 268.

Suitably, the variant may be at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 268.

- 5 Suitably, the AAV-DJ VP2 and VP3 capsid proteins may be N-terminal truncations of SEQ ID NO: 268, or N-terminal truncations of a variant which is at least 90% identical, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 268.

#### *AAV2 serotype*

- 10 The 5' AAV vector particle and/or the 3' AAV vector particle may comprise an AAV2 capsid protein. Suitably, the 5' AAV vector particle and/or the 3' AAV vector particle may be encapsidated by AAV2 capsid proteins.

In some embodiments, the 5' AAV vector particle and the 3' AAV vector particle each comprise an AAV2 capsid protein. In some embodiments, the 5' AAV vector particle and the 3' AAV vector particle are both encapsidated by AAV2 capsid proteins.

- 15 The 5' AAV vector particle and/or the 3' AAV vector particle may comprise an AAV2 VP1 capsid protein, an AAV2 VP2 capsid protein, and/or an AAV2 VP3 capsid protein. Suitably, the 5' AAV vector particle and/or the 3' AAV vector particle may be encapsidated by AAV2 VP1 capsid proteins, AAV2 VP2 capsid proteins, and/or AAV2 VP3 capsid proteins. Suitably, the 5' AAV vector particle and/or the 3' AAV vector particle may be encapsidated by AAV2
- 20 VP1, VP2, and VP3 capsid proteins.

Suitably, the AAV2 VP1 capsid protein may comprise or consist of the amino acid sequence shown as SEQ ID NO: 269, or a variant which is at least 90% identical to SEQ ID NO: 269.

Suitably, the variant may be at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 269.

- 25 Suitably, the AAV2 VP2 and VP3 capsid proteins may be N-terminal truncations of SEQ ID NO: 269, or N-terminal truncations of a variant which is at least 90% identical, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 269.

#### *AAV6.2 serotype*

The 5' AAV vector particle and/or the 3' AAV vector particle may comprise an AAV6.2 capsid protein. Suitably, the 5' AAV vector particle and/or the 3' AAV vector particle may be encapsidated by AAV6.2 capsid proteins.

- 5 In some embodiments, the 5' AAV vector particle and the 3' AAV vector particle each comprise an AAV6.2 capsid protein. In some embodiments, the 5' AAV vector particle and the 3' AAV vector particle are both encapsidated by AAV6.2 capsid proteins.

The AAV6.2 vector mutant was created by mutating the phenylalanine (F) residue at position 129 in AAV6 to leucine (L) (see e.g. Limberis, M.P., et al., 2009. Molecular Therapy, 17(2), pp.294-301).

- 10 The 5' AAV vector particle and/or the 3' AAV vector particle may comprise an AAV6.2 VP1 capsid protein, an AAV6.2 VP2 capsid protein, and/or an AAV6.2 VP3 capsid protein. Suitably, the 5' AAV vector particle and/or the 3' AAV vector particle may be encapsidated by AAV6.2 VP1 capsid proteins, AAV6.2 VP2 capsid proteins, and/or AAV6.2 VP3 capsid proteins. Suitably, the 5' AAV vector particle and/or the 3' AAV vector particle may be encapsidated by  
15 AAV6.2 VP1, VP2, and VP3 capsid proteins.

Suitably, the AAV6.2 VP1 capsid protein may comprise or consist of the amino acid sequence shown as SEQ ID NO: 270, or a variant which is at least 90% identical to SEQ ID NO: 270.

Suitably, the variant may be at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 270.

- 20 Suitably, the AAV6.2 VP2 and VP3 capsid proteins may be N-terminal truncations of SEQ ID NO: 270, or N-terminal truncations of a variant which is at least 90% identical, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 270.

#### *AAV5 serotype*

- 25 The 5' AAV vector particle and/or the 3' AAV vector particle may comprise an AAV5 capsid protein. Suitably, the 5' AAV vector particle and/or the 3' AAV vector particle may be encapsidated by AAV5 capsid proteins.

In some embodiments, the 5' AAV vector particle and the 3' AAV vector particle each comprise an AAV5 capsid protein. In some embodiments, the 5' AAV vector particle and the 3' AAV vector particle are both encapsidated by AAV5 capsid proteins.

- 30 The 5' AAV vector particle and/or the 3' AAV vector particle may comprise an AAV5 VP1 capsid protein, an AAV5 VP2 capsid protein, and/or an AAV5 VP3 capsid protein. Suitably,

the 5' AAV vector particle and/or the 3' AAV vector particle may be encapsidated by AAV5 VP1 capsid proteins, AAV5 VP2 capsid proteins, and/or AAV5 VP3 capsid proteins. Suitably, the 5' AAV vector particle and/or the 3' AAV vector particle may be encapsidated by AAV5 VP1, VP2, and VP3 capsid proteins.

- 5 Suitably, the AAV5 VP1 capsid protein may comprise or consist of the amino acid sequence shown as SEQ ID NO: 271, or a variant which is at least 90% identical to SEQ ID NO: 271.

Suitably, the variant may be at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 271.

- 10 Suitably, the AAV5 VP2 and VP3 capsid proteins may be N-terminal truncations of SEQ ID NO: 271, or N-terminal truncations of a variant which is at least 90% identical, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 271.

#### **Variants, derivatives, homologues and fragments**

In addition to the specific proteins and nucleotides mentioned herein, the invention also encompasses variants, derivatives, homologues and fragments thereof.

- 15 In the context of the invention, a "variant" of any given sequence is a sequence in which the specific sequence of residues (whether amino acid or nucleic acid residues) has been modified in such a manner that the polypeptide or polynucleotide in question retains at least one or all of its endogenous functions. A variant sequence can be obtained by addition, deletion, substitution, modification, replacement and/or variation of at least one residue present in the  
20 given sequence.

The term "derivative" as used herein in relation to proteins or polypeptides of the invention includes any substitution of, variation of, modification of, replacement of, deletion of and/or addition of one (or more) amino acid residues from or to the sequence, providing that the resultant protein or polypeptide retains at least one or all of its endogenous functions.

- 25 Typically, amino acid substitutions may be made, for example from 1, 2 or 3, to 10 or 20 substitutions, provided that the modified sequence retains the required activity or ability. Amino acid substitutions may include the use of non-naturally occurring analogues.

- Polypeptides used in the invention may also have deletions, insertions or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent  
30 protein. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity and/or the amphipathic nature of the residues

as long as the endogenous function is retained. For example, negatively charged amino acids include aspartic acid and glutamic acid; positively charged amino acids include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values include asparagine, glutamine, serine, threonine and tyrosine.

- 5 Conservative substitutions may be made, for example according to the table below. Amino acids in the same block in the second column and preferably in the same line in the third column may be substituted for each other:

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

- The effect of additions, deletions, substitutions, modifications, replacements and/or variations may be predicted using any suitable prediction tool e.g. SIFT (Vaser, R., et al., 2016. Nature protocols, 11(1), pp.1-9), PolyPhen-2 (Adzhubei, I., et al., 2013. Current protocols in human genetics, 76(1), pp.7-20), CADD (Rentzsch, P., et al., 2021. Genome medicine, 13(1), pp.1-12), REVEL (Ioannidis, N.M., et al., 2016. The American Journal of Human Genetics, 99(4), pp.877-885), MetaLR (Dong, C., et al., 2015. Human molecular genetics, 24(8), pp.2125-2137), and/or MutationAssessor (Reva, B., et al., 2011. Nucleic acids research, 39(17), pp.e118-e118) or based on clinical data e.g. ClinVar (Landrum, M.J., et al., 2016. Nucleic acids research, 44(D1), pp.D862-D868). Suitable additions, deletions, substitutions, modifications, replacements and/or variations may be considered tolerated, benign, and/or likely benign.

- In the present context, a variant sequence is taken to include an amino acid sequence which may be at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85% or at least 90% identical, suitably at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to the subject sequence. Although a variant can also be considered in terms of similarity (i.e. amino acid residues having similar chemical properties/functions), in the context of the present invention it is preferred to express it in terms of sequence identity.

In the present context, a variant sequence is taken to include a nucleotide sequence which may be at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85% or at least 90% identical, suitably at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to the subject sequence. Although a variant can

also be considered in terms of similarity, in the context of the present invention it is preferred to express it in terms of sequence identity.

The term "homologue" as used herein means a variant having a certain similarity with the wild type amino acid sequence or the wild type nucleotide sequence, e.g. having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85% or at least 90% similarity, suitably at least 95%, at least 96%, at least 97%, at least 98% or at least 99% similarity to the subject sequence.

Suitably, reference to a sequence which has a percent identity to any one of the SEQ ID NOs detailed herein refers to a sequence which has the stated percent identity over the entire length of the SEQ ID NO referred to.

Sequence identity comparisons can be conducted by eye, or more usually, with the aid of readily available sequence comparison programs. These commercially available computer programs can calculate percent identity between two or more sequences.

Percent identity may be calculated over contiguous sequences, i.e. one sequence is aligned with the other sequence and each amino acid or nucleotide in one sequence is directly compared with the corresponding amino acid or nucleotide in the other sequence, one residue at a time. This is called an "ungapped" alignment. Typically, such ungapped alignments are performed only over a relatively short number of residues.

Although this is a very simple and consistent method, it fails to take into consideration that, for example, in an otherwise identical pair of sequences, one insertion or deletion in the amino acid or nucleotide sequence may cause the following residues or codons to be put out of alignment, thus potentially resulting in a large reduction in percent identity when a global alignment is performed. Consequently, most sequence comparison methods are designed to produce optimal alignments that take into consideration possible insertions and deletions without penalising unduly the overall identity score. This is achieved by inserting "gaps" in the sequence alignment to try to maximise local identity.

However, these more complex methods assign "gap penalties" to each gap that occurs in the alignment so that, for the same number of identical amino acids or nucleotides, a sequence alignment with as few gaps as possible, reflecting higher relatedness between the two compared sequences, will achieve a higher score than one with many gaps. "Affine gap costs" are typically used that charge a relatively high cost for the existence of a gap and a smaller penalty for each subsequent residue in the gap. This is the most commonly used gap scoring system. High gap penalties will produce optimised alignments with fewer gaps. Most alignment



programs allow the gap penalties to be modified. However, it is preferred to use the default values when using such software for sequence comparisons. For example when using the GCG Wisconsin Bestfit package the default gap penalty for amino acid sequences is -12 for a gap and -4 for each extension.

- 5 Calculation of maximum percent identity therefore firstly requires the production of an optimal alignment, taking into consideration gap penalties. A suitable computer program for carrying out such an alignment is the GCG Wisconsin Bestfit package (see e.g. Devereux, J., et al., 1984. *Nucleic acids research*, 12(1), pp.387-395). Examples of other software that can perform sequence comparisons include, but are not limited to, the BLAST package (see e.g. Altschul, S.F., et al., 1990. *Journal of molecular biology*, 215(3), pp.403-410), BLAST 2 (see e.g. Tatusova, T.A. and Madden, T.L., 1999. *FEMS microbiology letters*, 174(2), pp.247-250), FASTA (see e.g. Pearson, W.R. and Lipman, D.J., 1988. *PNAS*, 85(8), pp.2444-2448.), EMBOSS Needle (Madeira, F., et al., 2019. *Nucleic acids research*, 47(W1), pp.W636-W641) and the GENWORKS suite of comparison tools. For some applications, it is preferred to use
- 10
- 15 EMBOSS Needle.

Although the final percent identity can be measured, the alignment process itself is typically not based on an all-or-nothing pair comparison. Instead, a scaled similarity score matrix is generally used that assigns scores to each pairwise comparison based on chemical similarity or evolutionary distance. An example of such a matrix commonly used is the BLOSUM62

20 matrix.

Once the software has produced an optimal alignment, it is possible to calculate percent sequence identity. The software typically does this as part of the sequence comparison and generates a numerical result. The percent sequence identity may be calculated as the number of identical residues as a percentage of the total residues in the SEQ ID NO referred to.

- 25 The term "fragment" as used herein refers to a variant sequence that is a portion of a full-length polypeptide or polynucleotide. Fragments are typically selected regions of the polypeptide or polynucleotide that is of interest either functionally or, for example, in an assay.

Such variants, derivatives, homologues and fragments may be prepared using standard recombinant DNA techniques such as site-directed mutagenesis. Where insertions are to be made, synthetic DNA encoding the insertion together with 5' and 3' flanking regions corresponding to the naturally-occurring sequence either side of the insertion site may be made. The flanking regions will contain convenient restriction sites corresponding to sites in the naturally-occurring sequence so that the sequence may be cut with the appropriate enzyme(s) and the synthetic DNA ligated into the cut. The DNA is then expressed in

30

accordance with the invention to make the encoded protein. These methods are only illustrative of the numerous standard techniques known in the art for manipulation of DNA sequences and other known techniques may also be used.

#### Vectors, kits, and systems

- 5 The present invention provides a vector encoding the 5' AAV vector, 3' AAV vector, or dual AAV vector system of the invention. The vector may be a transfer vector, as described herein. For example, the vector may be a plasmid and/or the viral genome may be operably linked to a promoter (e.g. a viral promoter, such as a CMV promoter).

- 10 The present invention also provides a kit or system for producing the 5' AAV vector, 3' AAV vector, or dual AAV vector system of the invention. The kit or system may be a virus packaging kit or system or a virus production kit or system. As used herein, a "virus packaging kit or system" may comprise one or more components, and optionally instructions, for packaging the viral vector of the present invention. As used herein, a "virus production kit or system" may comprise one or more components, and optionally instructions, for producing the viral vector  
15 of the present invention.

- The kit or system may comprise a transfer vector encoding the 5' AAV vector, 3' AAV vector, or dual AAV vector of the invention and optionally one or more helper vectors. The kit or system may further comprise host cells (e.g. packaging cells or producer cells) and/or other reagents (e.g. transfection reagent, culture medium, etc.). The kit or system may further  
20 comprise any other suitable components, and optionally instructions for packaging and/or producing the 5' AAV vector, 3' AAV vector, or dual AAV vector of the invention.

- In some embodiments, the kit is for production of AAV vector particles and comprises a transfer vector (e.g. plasmid) encoding the 5' AAV vector of the present invention, and one or more helper plasmids encoding AAV replication and capsid proteins. In some embodiments,  
25 the kit is for production of AAV vector particles and comprises a transfer vector (e.g. plasmid) encoding the 3' AAV vector of the present invention, and one or more helper plasmids encoding AAV replication and capsid proteins. In some embodiments, the kit is for production of AAV vector particles and comprises a transfer vector (e.g. plasmid) encoding the 5' AAV vector of the present invention, a transfer vector (e.g. plasmid) encoding the 3' AAV vector of  
30 the present invention, and one or more helper plasmids encoding AAV replication and capsid proteins.

#### Cells

The present invention provides a cell comprising the 5' AAV vector, 3' AAV vector, or dual AAV vector system of the invention. The present invention also provides a cell comprising the vector, kit or system of the invention.

5 The cell may be any cell type known in the prior art. The cell may be an isolated cell. The cell may be a human cell, suitably an isolated human cell.

Suitably, the cell may be a kidney cell or glomerular cell, for example a podocyte. Suitably, the cell may be an immortalized kidney cell or glomerular cell, for example an immortalized podocyte. Suitable podocyte cell lines will be well known to those of skill in the art, for example CIHP-1. Methods to generate immortalized podocytes will be well known to those of skill in the art. Suitable methods are described in Ni, L., et al., 2012. Nephrology, 17(6), pp.525-531.

10 Suitably, the cell may be a producer cell. The term "producer cell" includes a cell that produces viral particles, after transient transfection, stable transfection or vector transduction of all the elements necessary to produce the viral particles or any cell engineered to stably comprise the elements necessary to produce the viral particles. In some embodiments, the producer cell is an AAV producer cell. Suitable producer cells will be known to those of skill in the art (see e.g. Martin, J., et al. 2013. Human gene therapy methods, 24(4), pp.253-269) and may include HEK293, COS-1, COS-7, CV-1, HeLa, CHO, and A549 cell lines. In some embodiments, the producer cell is a HEK293 cell, or a derivative thereof (e.g. a HEK293T cell).

20 Suitably, the cell may be a packaging cell. The term "packaging cell" includes a cell which contains some or all of the elements necessary for packaging a recombinant virus genome. Typically, such packaging cells contain one or more vectors which are capable of expressing viral structural proteins (e.g. AAV *rep* and *cap* genes) and/or one or more genes encoding the viral structural proteins have been integrated into the genome of the packaging cell. Cells comprising only some of the elements required for the production of enveloped viral particles are useful as intermediate reagents in the generation of viral particle producer cell lines, through subsequent steps of transient transfection, transduction or stable integration of each additional required element. These intermediate reagents are encompassed by the term "packaging cell". In some embodiments, the packaging cell is an AAV packaging cell. Suitable packaging cells will be known to those of skill in the art (see e.g. Martin, J., et al. 2013. Human gene therapy methods, 24(4), pp.253-269).

#### Pharmaceutical compositions

The present invention provides a pharmaceutical composition comprising the 5' AAV vector, 3' AAV vector, dual AAV vector system, or cell of the invention. In preferred embodiments, the pharmaceutical composition comprises the dual AAV vector system of the invention.

5 A pharmaceutical composition is a composition that comprises or consists of a therapeutically effective amount of a pharmaceutically active agent i.e. the dual AAV vector system. It preferably includes a pharmaceutically acceptable carrier, diluent or excipient (including combinations thereof).

By "pharmaceutically acceptable" is included that the formulation is sterile and pyrogen free. The carrier, diluent, and/or excipient must be "acceptable" in the sense of being compatible  
10 with the vector and not deleterious to the recipients thereof. Typically, the carriers, diluents, and excipients will be saline or infusion media which will be sterile and pyrogen free; however, other acceptable carriers, diluents, and excipients may be used.

Acceptable carriers, diluents, and excipients for therapeutic use are well known in the pharmaceutical art. The choice of pharmaceutical carrier, excipient or diluent can be selected  
15 with regard to the intended route of administration and standard pharmaceutical practice. The pharmaceutical compositions may comprise as - or in addition to - the carrier, excipient or diluent any suitable binder(s), lubricant(s), suspending agent(s), coating agent(s) or solubilising agent(s).

Examples of pharmaceutically acceptable carriers include, for example, water, salt solutions,  
20 alcohol, silicone, waxes, petroleum jelly, vegetable oils, polyethylene glycols, propylene glycol, liposomes, sugars, gelatin, lactose, amylose, magnesium stearate, talc, surfactants, silicic acid, viscous paraffin, perfume oil, fatty acid monoglycerides and diglycerides, petroethral fatty acid esters, hydroxymethyl-cellulose, polyvinylpyrrolidone, and the like.

The 5' AAV vector, 3' AAV vector, dual AAV vector system, cell, or pharmaceutical  
25 composition according to the present invention may be administered in a manner appropriate for treating and/or preventing the diseases described herein. The quantity and frequency of administration will be determined by such factors as the condition of the subject, and the type and severity of the subject's disease, although appropriate dosages may be determined by clinical trials. The pharmaceutical composition may be formulated accordingly.

30 The 5' AAV vector, 3' AAV vector, dual AAV vector system, cell or pharmaceutical composition according to the present invention may be administered parenterally, for example, intravenously, or by infusion techniques. The 5' AAV vector, 3' AAV vector, dual AAV vector system, cell or pharmaceutical composition may be administered in the form of a sterile

aqueous solution which may contain other substances, for example, enough salts or glucose to make the solution isotonic with blood. The aqueous solution may be suitably buffered (preferably to a pH of from 3 to 9). The pharmaceutical composition may be formulated accordingly. The preparation of suitable parenteral formulations under sterile conditions is readily accomplished by standard pharmaceutical techniques well-known to those skilled in the art.

The 5' AAV vector, 3' AAV vector, dual AAV vector system, cell or pharmaceutical composition according to the present invention may be administered systemically, for example by intravenous injection.

- 10 The 5' AAV vector, 3' AAV vector, dual AAV vector system, cell or pharmaceutical composition according to the present invention may be administered locally, for example by targeting administration to the kidney. Suitably, the 5' AAV vector, 3' AAV vector, dual AAV vector system, cell or pharmaceutical composition may be administered by injection into the renal artery, by intraparenchymal injection, by transparenchymal injection, by renal vein injection, or by ureteral or subcapsular injection. In some embodiments, the 5' AAV vector, 3' AAV vector, dual AAV vector system, cell or pharmaceutical composition is administered by injection into the renal artery.

The pharmaceutical compositions may comprise the 5' AAV vector, 3' AAV vector, dual AAV vector system, or cell of the invention in infusion media, for example sterile isotonic solution.

- 20 The pharmaceutical composition may be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

The 5' AAV vector, 3' AAV vector, dual AAV vector system, cell or pharmaceutical composition may be administered in a single or in multiple doses. Particularly, the 5' AAV vector, 3' AAV vector, dual AAV vector system, cell or pharmaceutical composition may be administered in a single, one-off dose. The pharmaceutical composition may be formulated accordingly.

The 5' AAV vector, 3' AAV vector, dual AAV vector system, cell or pharmaceutical composition may be administered at varying doses (e.g. measured in vector genomes (vg) per kg). The physician in any event will determine the actual dosage which will be most suitable for any individual subject and it will vary with the age, weight and response of the particular subject.

- 30 Typically, however, for AAV vectors, doses of  $10^{10}$  to  $10^{14}$  vg/kg, or  $10^{11}$  to  $10^{13}$  vg/kg may be administered. Suitably, the 5' AAV vector and the 3' AAV vector are administered at the same dose.

The pharmaceutical composition may further comprise one or more other therapeutic agents, such as an angiotensin converting enzyme (ACE) inhibitor, an aldosterone antagonist (e.g., spironolactone) or an angiotensin receptor blocker (ARB).

- 5 The invention further includes the use of kits comprising the 5' AAV vector, 3' AAV vector, dual AAV vector system, cell and/or pharmaceutical composition of the present invention. Preferably said kits are for use in the methods and used as described herein, e.g., the therapeutic methods as described herein. Preferably, said kits comprise instructions for use of the kit components.

#### **Methods for treating and/or preventing disease**

- 10 The 5' AAV vector, 3' AAV vector, dual AAV vector system, cell or pharmaceutical composition of the invention may be administered to any subject in need thereof. Upon administration of the 5' AAV vector and the 3' AAV vector to a subject, a CDS encoding a COL4A3, COL4A4 or COL4A5 polypeptide is formed and a COL4A3, COL4A4 or COL4A5 polypeptide is expressed.

- 15 In one aspect, the present invention provides the 5' AAV vector, 3' AAV vector, dual AAV vector system, cell or pharmaceutical composition of the invention for use as a medicament.

In another aspect, the present invention provides use of the 5' AAV vector, 3' AAV vector, dual AAV vector system, cell or pharmaceutical composition of the invention in the manufacture of a medicament.

- 20 In another aspect, the present invention provides a method of administering the 5' AAV vector, 3' AAV vector, dual AAV vector system, cell or pharmaceutical composition of the invention to a subject in need thereof.

In another aspect, the present invention provides a product or kit comprising: (a) the 5' AAV vector of the invention; and (b) the 3' AAV vector of the invention, as a combined preparation for simultaneous, separate or sequential use in therapy.

- 25 In another aspect, the present invention provides use of a product or kit comprising: (a) the 5' AAV vector of the invention; and (b) the 3' AAV vector of the invention in the manufacture of a combined preparation for simultaneous, separate or sequential use in therapy.

- 30 In another aspect, the present invention provides a method of administering a therapeutically effective amount of (a) the 5' AAV vector of the invention; and (b) the 3' AAV vector of the invention, wherein the 5' AAV vector and the 3' AAV vector are administered simultaneously, separately or sequentially.

In another aspect, the present invention provides a pharmaceutical composition comprising: (a) the 5' AAV vector of the invention; and (b) the 3' AAV vector of the invention for use as a medicament.

5 In another aspect, the present invention provides use of pharmaceutical composition comprising: (a) the 5' AAV vector of the invention; and (b) the 3' AAV vector of the invention in the manufacture of a medicament.

In another aspect, the present invention provides a method of administering a therapeutically effective amount of a pharmaceutical composition comprising: (a) the 5' AAV vector of the invention; and (b) the 3' AAV vector of the invention.

10 Subject

The 5' AAV vector, 3' AAV vector, dual AAV vector system, cell or pharmaceutical composition of the invention may be administered to any subject in need thereof.

The subject may be a mammal. In preferred embodiments, the subject is a human.

15 The subject may be an adult, an adolescent, a child, a toddler, and infant, or a neonate. In some embodiments, the subject is a paediatric patient. In some embodiments, the subject is a child, an infant, or a neonate. In some embodiments, the subject is an infant or a neonate.

In some embodiments, the subject is aged about 18 years or older. In some embodiments, the subject is aged about 16 years or younger. In some embodiments, the subject is aged about 12 years or younger. In some embodiments, the subject is aged about 6 months or  
20 older. In some embodiments, the subject is aged about 1 years or older.

The subject may be male or female. The subject may be a male with X-chromosome linked Alport syndrome. In some embodiments, the subject is a male child, infant, or neonate with X-chromosome linked Alport syndrome.

25 The subject may have or may be at risk of Alport Syndrome. In some embodiments, the subject has or is at risk of Alport Syndrome. The Alport Syndrome may be any described below in the section entitled "Alport Syndrome".

Alport Syndrome

30 In one aspect, the present invention provides the 5' AAV vector, 3' AAV vector, dual AAV vector system, cell or pharmaceutical composition of the invention for use in preventing and/or treating Alport Syndrome.

In another aspect, the present invention provides use of the 5' AAV vector, 3' AAV vector, dual AAV vector system, cell or pharmaceutical composition of the invention in the manufacture of a medicament for preventing and/or treating Alport Syndrome.

5 In another aspect, the present invention provides a method of preventing and/or treating Alport Syndrome, the method comprising administering the 5' AAV vector, 3' AAV vector, dual AAV vector system, cell or pharmaceutical composition of the invention to a subject in need thereof.

In another aspect, the present invention provides a product or kit comprising: (a) the 5' AAV vector of the invention; and (b) the 3' AAV vector of the invention as a combined preparation for simultaneous, separate or sequential use in preventing and/or treating Alport syndrome.

10 In another aspect, the present invention provides use of a product or kit comprising: (a) the 5' AAV vector of the invention; and (b) the 3' AAV vector of the invention in the manufacture of a combined preparation for simultaneous, separate or sequential use in preventing and/or treating Alport syndrome.

15 In another aspect, the present invention provides a method of preventing and/or treating Alport syndrome, the method comprising administering a therapeutically effective amount of (a) the 5' AAV vector of the invention; and (b) the 3' AAV vector of the invention to a subject in need thereof, wherein the 5' AAV vector and the 3' AAV vector are administered simultaneously, separately or sequentially.

20 In another aspect, the present invention provides a pharmaceutical composition comprising: (a) the 5' AAV vector of the invention; and (b) the 3' AAV vector of the invention for use in preventing and/or treating Alport syndrome.

In another aspect, the present invention provides use of pharmaceutical composition comprising: (a) the 5' AAV vector of the invention; and (b) the 3' AAV vector of the invention in the manufacture of a medicament for preventing and/or treating Alport syndrome.

25 In another aspect, the present invention provides a method of preventing and/or treating Alport syndrome, the method comprising administering a therapeutically effective amount of a pharmaceutical composition comprising: (a) the 5' AAV vector of the invention; and (b) the 3' AAV vector of the invention to a subject in need thereof.

30 Alport syndrome is a genetic condition characterised by kidney disease and extrarenal manifestations, such as loss of hearing and eye abnormalities. Alport syndrome is also known as familial nephritis, hereditary nephritis, thin basement membrane disease and thin basement membrane nephropathy. Alport syndrome is caused by pathogenic variants in the COL4A3,



COL4A4 and COL4A5 genes, which result in abnormalities of the collagen IV  $\alpha$ 345 network of basement membranes. The condition can be transmitted in an X-linked, autosomal dominant, or autosomal recessive pattern, with X-linked being the common while autosomal recessive and autosomal dominant account for around 15% and 20% of cases respectively (see e.g. Warady, B.A., et al., 2020. *Kidney medicine*, 2(5), pp.639-649).

The 5' AAV vector, 3' AAV vector, dual AAV vector system, cell or pharmaceutical composition according to the present invention may be administered to a subject with Alport Syndrome in order to reverse the rejection or slow down progression of Alport Syndrome, or to lessen, reduce, or improve at least one symptom of Alport Syndrome.

Alport syndrome may be diagnosed by any method known in the art, including urinalysis, kidney biopsy and/or by genetic testing. For example, all males with X-linked Alport syndrome, as well as all males and females with autosomal recessive Alport syndrome, have persistent microhematuria (see e.g. Kashtan, C.E., 2021. *American Journal of Kidney Diseases*, 77(2), pp.272-279).

The subject may have haematuria, microalbuminuria, macroalbuminuria or nephrotic range proteinuria. Subjects with Alport syndrome typically present with haematuria, which may progress to proteinuria. Haematuria can be determined by the presence of erythrocytes in urine when viewed microscopically. A basal microalbuminuria level of less than 30 mg/day is usually considered non-pathological. Levels of about 30 mg/day to about 300 mg/day are termed microalbuminuria, which is considered pathologic. Albumin levels of over 300 mg/day are termed macroalbuminuria and levels of proteinuria over 3.5 g/day are considered to be nephrotic range proteinuria. Treating subjects prior to onset of proteinuria may slow or prevent progression of proteinuria and thereby delay or prevent end-stage renal failure. Subjects with nephrotic range proteinuria may also be treated. As the collagen IV  $\alpha$ 345 network of the glomerular basement membrane is changed and normalised or repaired by the transgene, proteinuria levels should be progressively reduced.

The subject may additionally or alternatively test positive for a pathogenic variant of COL4A3, COL4A4 or COL4A5 (see e.g. Savige, J., et al., 2019. *Pediatric Nephrology*, 34, pp.1175-1189). Pathogenic variants of COL4A3 and COL4A4 can be heterozygous (autosomal dominant), or biallelic (autosomal recessive). Pathogenic variants in COL4A5 are hemizygous or heterozygous (X-linked). The subject is preferably treated with a dual AAV vector system encoding a COL4A3, COL4A4, or COL4A5 polypeptide corresponding to the gene(s) for which the subject has a pathogenic variant. For example, a subject having a pathogenic COL4A5 variant can be treated with a dual AAV vector system encoding a COL4A5 polypeptide. The

subject may test positive for two or more pathogenic variants of COL4A3, COL4A4 or COL4A5. Such subjects may be treated with two or more dual AAV vector systems encoding different transgenes, i.e., each dual AAV vector system encoding a COL4A3, COL4A4, or COL4A5 polypeptide corresponding to the genes for which the subject has pathogenic variants.

In some embodiments, the pathogenic variant may be a mutation in the coding sequence or non-coding sequence of the COL4A3, COL4A4 or COL4A5 gene. In some embodiments, the pathogenic variant may be a mutation in the associated promoter or regulatory element(s) of the COL4A3, COL4A4 or COL4A5 gene.

10 In some embodiments, the Alport Syndrome is X-chromosome linked Alport syndrome or autosomal Alport syndrome.

In some embodiments, the Alport Syndrome is X-chromosome linked Alport syndrome. X-chromosome linked Alport syndrome (XLAS) is typically associated with a pathogenic variant of COL4A5 and accounts for about 85% of cases of Alport syndrome (see e.g. Hashimura, Y., et al., 2014. Kidney international, 85(5), pp.1208-1213). Over 300 mutations have been observed in the COL4A5 genes in families with XLAS, including splice-site mutations, missense mutations, and deletions of fewer than ten base pairs.

In some embodiments, the Alport Syndrome is autosomal Alport syndrome. Autosomal Alport syndrome is typically associated with pathogenic variants of COL4A3 or COL4A4. In some embodiments, the Alport Syndrome is autosomal recessive Alport syndrome. Autosomal recessive Alport syndrome (ARAS) accounts for about 10-15% of cases of Alport syndrome. In some embodiments, the Alport Syndrome is autosomal dominant Alport syndrome. Autosomal dominant Alport syndrome (ADAS) is a rare form of Alport syndrome. In patients with Alport syndrome, to date, only six mutations in the COL4A3 gene and twelve mutations in the COL4A4 gene are seen in patients with ARAS. The mutations include frameshift deletions, amino acid substitutions, missense mutations, splicing mutations, and in-frame deletions (see e.g. Watson S, et al. Alport Syndrome. In: StatPearls).

#### Dual AAV vector system administration

In preferred embodiments, the dual AAV vector system is administered to the subject.

30 The 5' AAV vector and the 3' AAV vector may be administered simultaneously, separately or sequentially. In some embodiments, the 5' AAV vector and the 3' AAV vector may be administered simultaneously. In some embodiments, the 5' AAV vector and the 3' AAV vector are administered in the form of pharmaceutical compositions.

The dual AAV vector system may be administered in any suitable dosage (e.g. measured in vector genomes (vg) or vg per kg). The dosage may be determined by such factors as the condition of the subject, age of the subject, weight of the subject, and the type and severity of the subject's disease, and appropriate dosages may be determined by a physician. The dual

5 AAV vector system may be formulated accordingly.

Suitably, each AAV vector comprised in the dual AAV vector system is administered in a dose of about  $1 \times 10^6$  vg/kg or more, about  $1 \times 10^7$  vg/kg or more, about  $1 \times 10^8$  vg/kg or more, about  $1 \times 10^9$  vg/kg or more, about  $1 \times 10^{10}$  vg/kg or more, about  $1 \times 10^{11}$  vg/kg or more, or about  $1 \times 10^{12}$  vg/kg or more. Suitably, each AAV vector comprised in the dual AAV vector system is  
10 administered in a dose of about  $1 \times 10^{14}$  vg/kg or less or about  $1 \times 10^{13}$  vg/kg or less. Suitably, each AAV vector comprised in the dual AAV vector system is administered in a dose of from about  $1 \times 10^6$  vg/kg to about  $1 \times 10^{14}$  vg/kg. Suitably, each AAV vector comprised in the dual AAV vector system is administered in a dose of from about  $1 \times 10^6$  vg/kg to about  $1 \times 10^{13}$  vg/kg. In some embodiments, each AAV vector comprised in the dual AAV vector system is  
15 administered in a dose of from about  $1 \times 10^9$  vg/kg to about  $1 \times 10^{12}$  vg/kg.

Suitably, each AAV vector is administered in a dose of about  $1 \times 10^8$  vg or more, about  $1 \times 10^9$  vg or more, about  $1 \times 10^{10}$  vg or more, about  $1 \times 10^{11}$  vg or more, about  $1 \times 10^{12}$  vg or more, about  $1 \times 10^{13}$  vg or more, or about  $1 \times 10^{14}$  vg or more. Suitably, each AAV vector is administered in a dose of about  $1 \times 10^{15}$  vg or less or about  $1 \times 10^{14}$  vg or less. Suitably, each AAV vector is  
20 administered in a dose of from about  $1 \times 10^8$  vg to about  $1 \times 10^{15}$  vg. Suitably, each AAV vector is administered in a dose of from about  $1 \times 10^8$  vg to about  $5 \times 10^{14}$  vg. In some embodiments, each AAV vector is administered in a dose of from about  $1 \times 10^{11}$  vg to about  $1 \times 10^{14}$  vg.

The dual AAV vector system may be administered in the form of an AAV vector formulation. As used herein, an "AAV vector formulation" may refer to a composition that comprises or  
25 consists of a therapeutically effective amount of one or more AAV vector. It preferably includes a pharmaceutically acceptable carrier, diluent or excipient (including combinations thereof). By "pharmaceutically acceptable" is included that the formulation is sterile and pyrogen free. The carrier, diluent, and/or excipient must be "acceptable" in the sense of being compatible with the AAV vectors and not deleterious to the recipients thereof. Typically, the carriers,  
30 diluents, and excipients will be saline or infusion media which will be sterile and pyrogen free, however, other acceptable carriers, diluents, and excipients may be used.

The dual AAV vector system may be administered in the form of a sterile aqueous solution which may contain other substances, for example, enough salts or glucose to make the solution isotonic with blood. The aqueous solution may be suitably buffered (preferably to a

pH of from 3 to 9). The preparation of suitable parenteral formulations under sterile conditions is readily accomplished by standard pharmaceutical techniques well-known to those skilled in the art. Suitably, the AAV vector formulation comprises an isotonic buffer (e.g. at about pH 7.4).

- 5 The dual AAV vector system may be delivered by any suitable route of administration. In some embodiments, the dual AAV vector system is administered by intravenous injection. In some embodiments, the dual AAV vector system is administered by injection into the renal artery, by intraparenchymal injection, by transparenchymal injection, by renal vein injection, or by ureteral or subcapsular injection. In some embodiments, the dual AAV vector system is administered by renal artery injection.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of chemistry, biochemistry, molecular biology, microbiology and immunology, which are within the capabilities of a person of ordinary skill in the art. Such techniques are explained in the literature. See, for example: Skoog, D.A., et al. (2013) Fundamentals of Analytical Chemistry, 9th edition, Cengage learning; Walker J.M. (2009) The Protein Protocols Handbook, 3rd edition, Springer Nature; Green, M.R. and Sambrook, J. (2012) Molecular Cloning: A Laboratory Manual, 4th Edition, Cold Spring Harbor Laboratory Press; Ausubel, F.M., et al. (2003) Current Protocols in Molecular Biology, John Wiley & Sons; Hill, A. J. (2013) DNA Sequencing Protocols, Humana Press; Nielsen, B.S. and Jones, J. (2021) In Situ Hybridization Protocols, Springer US; Herdewijn, P. (2010) Oligonucleotide Synthesis: Methods and Applications, Humana Press; and Luo, Y. (2019) CRISPR Gene Editing: Methods and Protocols, Springer New York. Each of these general texts is herein incorporated by reference.

## EXAMPLES

- 25 The invention will now be further described by way of Examples, which are meant to serve to assist one of ordinary skill in the art in carrying out the invention and are not intended in any way to limit the scope of the invention. These examples relate to human COL4A5, unless murine COL4A5 is stated.

### Example 1 – Non-overlapping dual AAV vector system encoding COL4A5

- 30 Non-overlapping dual AAV vector systems encoding a COL4A5 polypeptide were designed, generated, and tested.

#### Design and generation of non-overlapping dual AAV vector systems

Schematics of example non-overlapping dual AAV vector systems encoding a COL4A5 polypeptide are shown in **Figure 1**. Each of these non-overlapping dual AAV vector systems are based on the hybrid dual AAV vector strategy using a recombinogenic region from the F1 phage sequence.

- 5       • (1) (a) a first AAV vector encoding exons 1-29 of a CDS encoding a COL4A5 polypeptide and an exogenous splice donor; and (b) a second AAV vector encoding exons 30-51 of a CDS encoding a COL4A5 polypeptide and an exogenous splice acceptor
- 10       • (2) (a) a first AAV vector encoding exons 1-31 of a CDS encoding a COL4A5 polypeptide and an exogenous splice donor; and (b) a second AAV vector encoding exons 32-51 of a CDS encoding a COL4A5 polypeptide and an exogenous splice acceptor
- 15       • (3) (a) a first AAV vector encoding exons 1-32 of a CDS encoding a COL4A5 polypeptide and an exogenous splice donor; and (b) a second AAV vector encoding exons 33-51 of a CDS encoding a COL4A5 polypeptide and an exogenous splice acceptor
- 20       • (4) (a) a first AAV vector encoding exons 1-32 of a CDS encoding a COL4A5 polypeptide and part of the COL4A5 intron between exons 32 and 33 comprising a natural splice donor site; and (b) a second AAV vector encoding exons 33-51 of a CDS encoding a COL4A5 polypeptide and part of the COL4A5 intron between exons 32 and 33 comprising a natural splice acceptor site
- 25       • (5) (a) a first AAV vector encoding exons 1-33 of a CDS encoding a COL4A5 polypeptide and an exogenous splice donor; and (b) a second AAV vector encoding exons 34-51 of a CDS encoding a COL4A5 polypeptide and an exogenous splice acceptor

#### Expression of COL4A5 polypeptide in AD293 cells

AD293 cells were seeded into 6-well plates at a density of 2e5 cells per well the day before transduction. The following day viral particles were added to the wells as necessary at a dose of 1e5 AAV vg per cell. Cells were incubated for a total of 72 hours and cell culture supernatant was collected for western blot analysis.

Non-transduced cells, and cells transduced with either 5' or 3' vectors alone demonstrate no detectable COL4A5 expression. Cells transduced with both 5' and 3' vectors demonstrate expression of COL4A5 when transduced with a 31/32 split hybrid design, including a band at the expected size of ~200kDa (see **Figures 2A and 2B**).

#### Expression of COL4A5 polypeptide in podocytes

Conditionally immortalised podocytes were seeded into 6-well plates at a density of 1e5 cells per well the day before transduction and allowed to adhere overnight. The following day viral particles were added to the wells as necessary at a dose of 1e5 AAV vg per cell. Cells were  
5 incubated for a total of 72 hours and cell culture supernatant was collected for western blot analysis. RNA was extracted and gene expression measured via qPCR using primers targeting the WPRE region.

Non-transduced cells, and cells transduced with either 5' or 3' vectors alone demonstrate no detectable COL4A5 expression. Cells transduced with both 5' and 3' vectors demonstrate  
10 expression of COL4A5 when transduced with a 31/32 split hybrid design, including a band at the expected size of ~200kDa (see **Figures 3A and 3B**).

Non-transduced cells and cells transduced with the 5' vector only demonstrate no WPRE expression. Cells transduced with 3' vectors only demonstrate very low expression of WPRE. Cells transduced with both 5' and 3' vectors demonstrate high levels of WPRE expression  
15 (see **Figure 3C**). In Figure 3C, "Vector A" contains a V5 tag at the 3' end of the COL4A5 sequence. "Vector B" does not contain a tag.

Expression of full-length COL4A5 in podocytes is also achieved with the 33/34 hybrid design shown in Figure 1, albeit that truncated COL4A5 species are also expressed (see **Figure 3D**). Results using the 31/32 hybrid design are also shown alongside. No full-length COL4A5 was  
20 observed with the 32/33 design including the chromosomal intron instead of the splice donor/acceptor sites.

#### Expression of a single full-length species of COL4A5 polypeptide in podocytes

Conditionally immortalised podocytes were seeded into 6-well plates at a density of 1e5 cells per well the day before transduction and allowed to adhere overnight. The following day viral  
25 particles were added to the wells as necessary at a dose of 1e5 AAV vg per cell. Cells were incubated for a total of 72 hours and cell culture supernatant was collected for western blot analysis.

Cells were transduced with variations on the hybrid designs (of Figure 1) evaluated in Figure 3, where either the recombinogenic F1 region alone was removed (Trans-splicing), or both the  
30 F1 and splice donor/acceptor sites were removed (No intron) to minimise off-target splicing. Hybrid designs with asymmetric split sites were also evaluated (24/25 and 38/39) (see **Figures 4A and 4B**).

Non-transduced cells and cells transduced with the 3' vectors alone demonstrate no detectable COL4A5 expression. Cells transduced with both 5' and 3' vectors of the min/max 24/25 hybrid design demonstrated expression of a single, full-length species of COL4A5 (i.e. a band at the expected size of ~200 kD). Cells transduced with both 5' and 3' vectors of the min/max 38/39 hybrid design or the trans-splicing design demonstrated expression of a full-length (~200 kD) species of COL4A5 together with truncated protein species (see **Figure 4B**). The no intron design did not express a full-length COL4A5 polypeptide.

### Conclusions

Western blot analysis in both AD293 cells and podocytes demonstrates successful expression of full-length COL4A5 only in the presence of both 5' and 3' vectors (of the asymmetric min/max 24/25 split, the asymmetric min/max 38/39 split, the trans-splicing design, the 31/32 mid-split design and the 33/34 mid-split design), indicating successful expression is mediated by recombination of the AAVs following transduction. Similarly, expression of high levels of WPRE in cells transduced specifically with both 5' and 3' vectors indicates the successful recombination and transcription of the full cassette.

Together, these results demonstrate the feasibility of expressing full-length COL4A5 using a split DNA dual vector AAV approach. Figures 1-4 demonstrate expression of full-length COL4A5 with the asymmetric min/max 24/25 split, the asymmetric min/max 38/39 split, the trans-splicing design, the 31/32 mid-split design and the 33/34 mid-split design (where the 5' CDS and 3' CDS do not share an overlapping sequence). However, as can be seen from Figures 2, 3 and 4 (e.g. bands at ~160 and ~120 kDa), as well as full-length COL4A5 (i.e. at ~200kDa), there also appears to be truncated protein expressed with the 31/32 mid-split design, the 33/34 mid-split design, the trans-splicing design and the asymmetric min/max 38/39 split design. Therefore, the present inventors went on to design and test alternative dual vector AAV approaches, e.g. overlapping approaches where the 5' CDS and 3' CDS share an overlapping sequence.

### Materials and methods

#### *Cell Culture*

AD293 cells were cultured in DMEM(Sigma, D5796) supplemented with 10% FBS(Gibco, Cat.No. 16629525). Cells were passaged at a ratio of 1:10 once they reached a confluency of ~80%. AD293 cells were maintained at 37°C, 5% CO<sub>2</sub>

Conditionally immortalised Podocytes were cultured in RPMI(Sigma, R8758) supplemented with 10% FBS and 1X insulin, transferrin and selenium (ITS) (Gibco, 12097549). Conditionally

immortalised Podocytes were maintained at 33°C, 5% CO<sub>2</sub>. Cells were passaged at ~80% confluency at a ratio of 1:4.

Viral Production Cells (ThermoFisher, A48400) were cultured in Viral Production Medium (Gibco, A4817901), supplemented with 1% Glutamax (ThermoFisher, 35050061). Cells were  
5 passaged every 3-4 days, seeding at a density of 1e6 cells/ml in shake flasks.

#### *AAV Production & QC*

For AAV production Thermo cells were seeded at a density of 3e6/ml and transfected with Helper, Rep/Cap and Genome plasmids at a 1:1:1 molar ratio using AAV Max transfection kit (Invitrogen, A50515). Cells were lysed using AAV Max Lysis Buffer (Gibco, 17331899), debris  
10 removed by filtration through 0.22µm filtration unit, and concentrated via ultracentrifugation (55,000rpm, 75 minutes) on an iodixanol gradient (Sigma). Concentrated AAV was washed 5 times in with PBS supplemented with 0.001% Pluronic (ThermoFisher, 24040032) in a 15ml amicon filtration unit (Millipore, UFC910024). Finally, the AAV preparation was filter sterilised using 0.22µm syringe filters. AAV particles were quantified using qPCR methods, with primers  
15 targeting either the promoter (CMV) or WPRE regions in the vectors.

#### *Cell transduction and harvest*

For AD293 transduction experiments, cells were seeded into 6-well plates at a density of 2e5 cells per well the day before transduction, and allowed to adhere overnight at 37°C, 5% CO<sub>2</sub>. The following day viral particles were added to the wells as necessary at a dose of 1e5 AAV  
20 vg per cell. Cells were incubated overnight at 37°C, 5% CO<sub>2</sub>. Media was changed 24 hours following transduction, and cells were incubated for a further 48-hours. Cell culture supernatant was collected for western blot analysis.

For Podocyte transduction experiments, cells were seeded into 6-well plates at a density of 7.3e4 cells per well the day before transduction, and allowed to adhere overnight at 33°C, 5%  
25 CO<sub>2</sub>. The following day viral particles were added to the wells as necessary at a dose of 1e5 AAV vg per cell. Cells were incubated overnight at 33°C, 5% CO<sub>2</sub>. Media was changed 24 hours following transduction, and cells were thermoswitched at 37°C. 48 hours later, cell culture supernatant was collected for western blot analysis. Cells were washed with PBS and RNA extracted using RNeasy kit (Qiagen). cDNA was synthesised using SuperScript IL VILO  
30 cDNA synthesis kit (Invitrogen), and gene expression was measured via qPCR using primers targeting the WPRE region.

#### **Example 2 - Overlapping dual AAV vector system encoding COL4A5**



Design and generation of overlapping dual AAV vector systems with a longer 5' CDS

Listed below are 6 example overlapping dual AAV vector systems encoding a COL4A5 polypeptide. Each of these overlapping dual AAV systems comprise the same 5' AAV vector comprising a CMV promoter and exons 1-40 of a CDS encoding a COL4A5 polypeptide. The 3' AAV vector each comprise a V5 tag and additional regulatory elements. Schematics of these example overlapping dual AAV vector systems encoding a COL4A5 polypeptide are shown in **Figure 5**.

- (1) (a) a first AAV vector encoding exons 1-40 of a CDS encoding a COL4A5 polypeptide and a CMV promoter; (b) a second AAV vector encoding exons 27-51 of a CDS encoding a COL4A5 polypeptide, a WPRE and a polyadenylation sequence. The length of overlap is 1566 bp. Labelled as number 1 in Figure 5.
- (2) (a) a first AAV vector encoding exons 1-40 of a CDS encoding a COL4A5 polypeptide and a CMV promoter; (b) a second AAV vector encoding exons 29-51 of a CDS encoding a COL4A5 polypeptide, a WPRE and a polyadenylation sequence. The length of overlap is 1360 bp. Labelled as number 2 in Figure 5.
- (3) (a) a first AAV vector encoding exons 1-40 of a CDS encoding a COL4A5 polypeptide and a CMV promoter; (b) a second AAV vector encoding exons 31-51 of a CDS encoding a COL4A5 polypeptide, a WPRE and a polyadenylation sequence. The length of overlap is 1095 bp. Labelled as number 3 in Figure 5.
- (4) (a) a first AAV vector encoding exons 1-40 of a CDS encoding a COL4A5 polypeptide and a CMV promoter; (b) a second AAV vector encoding nucleotides 2855 - 5058 of a CDS encoding a COL4A5 polypeptide, a WPRE and a polyadenylation sequence. The length of overlap is 750 bp. Labelled as number 4 in Figure 5.
- (5) (a) a first AAV vector encoding exons 1-40 of a CDS encoding a COL4A5 polypeptide and a CMV promoter; (b) a second AAV vector encoding exons 36-51 of a CDS encoding a COL4A5 polypeptide, a WPRE and a polyadenylation sequence. The length of overlap is 498 bp. Labelled as number 5 in Figure 5.
- (6) (a) a first AAV vector encoding exons 1-40 of a CDS encoding a COL4A5 polypeptide and a CMV promoter; (b) a second AAV vector encoding exons 38-51 of a CDS encoding a COL4A5 polypeptide, a WPRE and a polyadenylation sequence. The length of overlap is 230 bp. Labelled as number 6 in Figure 5.

Expression of COL4A5 polypeptide in podocytes

Cell culture, AAV production & QC, and cell transduction and harvest were carried out as described above in Example 1, unless otherwise stated.

Conditionally immortalised podocytes were seeded into 6-well plates at a density of 7.3e4 cells per well the day before transduction, and allowed to adhere overnight. The following day viral particles were added to the wells as necessary at a dose of 1e5 AAV vg per cell. Cells were incubated overnight at 33°C, 5% CO<sub>2</sub>. Media was changed 24 hours following transduction, and cells were thermoswitched at 37°C. 48 hours later, cell culture supernatant was collected for western blot analysis.

Cells transduced with both 5' and 3' vectors demonstrate expression of COL4A5, including a band at the expected size of ~200kDa. Using an anti-V5 antibody to detect expression of the V5-tagged COL4A5, expression of full-length COL4A5 was seen with overlaps of 1566bp, 1360bp, 1095bp, 750bp and 498bp and a smaller amount was observed with an overlap of 230bp. Non-transduced cells, and cells transduced with the 5' vector alone demonstrate no detectable COL4A5 expression (see **Figure 6A and 6B**). Supernatant samples from cells transduced with 1360bp, 1095bp, 498bp and 230bp overlaps were also analysed by western blot using an anti-COL4A5 antibody. Expression of full-length COL4A5 was seen with overlaps of 1360bp, 1095bp and 498bp and a smaller amount was observed with an overlap of 230bp. Non-transduced cells and cells transduced with the 5' vector alone demonstrate no detectable COL4A5 expression (see **Figure 6C**). COL4A5 expression was also detected in cells transduced with the 1095bp overlap design when driven by the nephrin(265) promoter (see **Figure 6D**, western blot analysis using an anti-V5 antibody) and with overlapping designs encoding murine COL4A5 (see **Figure 6E**, western blot analysis using an anti-V5 antibody). This demonstrates the feasibility of expressing full-length COL4A5 using an overlapping dual AAV vector approach. This overlapping approach, with a longer 5' CDS, achieved good full-length COL4A5 expression levels, particularly for an overlapping sequence length of 450-1600 nucleotides.

#### Design and generation of overlapping dual AAV vector systems with a longer 3' CDS.

Schematics of example overlapping dual AAV vector systems encoding a COL4A5 polypeptide are shown in **Figure 7**. Each of these overlapping dual AAV systems comprise the same 3' AAV vector comprising nucleotides 1724-5055 of a CDS encoding a COL4A5 polypeptide, a V5 tag, and additional regulatory elements. The 5' AAV vector each comprise a CMV promoter and a CDS encoding a COL4A5 polypeptide.

- (1) (a) a first AAV vector encoding nucleotides 1 - 2474 of a CDS encoding a COL4A5 polypeptide and a CMV promoter; (b) a second AAV vector encoding nucleotides – 1724-5055 of a CDS encoding a COL4A5 polypeptide, a WPRE and a polyadenylation sequence. The length of overlap is 750 bp. Labelled as number 1 in **Figure 7**.

- (2) (a) a first AAV vector encoding nucleotides 1 - 2818 of a CDS encoding a COL4A5 polypeptide and a CMV promoter; (b) a second AAV vector encoding nucleotides – 1724-5055 of a CDS encoding a COL4A5 polypeptide, a WPRE and a polyadenylation sequence. The length of overlap is 1094 bp. Labelled as number 2 in Figure 7.
- 5     • (3) (a) a first AAV vector encoding nucleotides 1 - 3083 of a CDS encoding a COL4A5 polypeptide and a CMV promoter; (b) a second AAV vector encoding nucleotides – 1724-5055 of a CDS encoding a COL4A5 polypeptide, a WPRE and a polyadenylation sequence. The length of overlap is 1359 bp. Labelled as number 3 in Figure 7.

#### Expression of COL4A5 polypeptide in podocytes

- 10     Cell culture, AAV production & QC, and cell transduction and harvest were carried out as described above for the design and generation of 1<sup>st</sup> generation dual vectors.

Cells transduced with both 5' and 3' vectors demonstrate expression of COL4A5, including a band at the expected size of ~200kDa. Using an anti-V5 antibody to detect expression of the V5-tagged COL4A5, expression of full-length COL4A5 was seen with overlaps of 1359bp,  
 15     1094bp, and 750bp. Non-transduced cells, demonstrate no detectable COL4A5 expression (see **Figures 8A**). Using a COL4A5-specific enzyme-linked immunosorbent assay (ELISA), total COL4A5 protein was quantified in the supernatant of cells transduced with generation 1 and generation 2 overlapping vectors as well as the single 5' and 3' vector controls (see **Figure 8B**).

- 20     This further demonstrates the feasibility of expressing full-length COL4A5 using another overlapping dual AAV vector approach.

**EMBODIMENTS**

Various embodiments of the present invention will now be described with reference to the following numbered paragraphs (paras).

1. A dual adeno-associated virus (AAV) vector system for expressing a COL4A5 polypeptide comprising:

(a) a first AAV vector comprising a 5' coding sequence (CDS), wherein the 5' CDS encodes an N-terminal part of the COL4A5 polypeptide and a promoter upstream of the 5' CDS; and

(b) a second AAV vector comprising a 3' CDS, wherein the 3' CDS encodes a C-terminal part of the COL4A5 polypeptide,

wherein a CDS encoding the COL4A5 polypeptide is formed upon delivery of the first AAV vector and the second AAV vector into a host cell;

optionally wherein the COL4A5 polypeptide is human COL4A5 polypeptide, preferably wherein the COL4A5 polypeptide is full-length human COL4A5 polypeptide.

2. A dual adeno-associated virus (AAV) vector system of paragraph 1, wherein the 5' CDS sequence is longer than the 3' CDS sequence, and wherein the overlapping sequence has a length of from 450 nucleotides to 1600 nucleotides, and wherein the first AAV vector comprises a kidney specific promoter (e.g. a podocyte-specific promoter); and optionally wherein the 5' CDS encodes exons 1-40 of the COL4A5 polypeptide.

3. A dual adeno-associated virus (AAV) vector system of paragraph 2 for expressing a COL4A5 polypeptide, wherein

(a) the first AAV vector encodes exons 1-40 of a CDS encoding a COL4A5 polypeptide and a kidney specific promoter (e.g. a podocyte-specific promoter); and

(b) the second AAV vector encodes

i) exons 27-51 of a CDS encoding a COL4A5 polypeptide, a WPRE and a polyadenylation sequence; or

ii) exons 29-51 of a CDS encoding a COL4A5 polypeptide, a WPRE and a polyadenylation sequence; or

iii) exons 31-51 of a CDS encoding a COL4A5 polypeptide, a WPRE and a polyadenylation sequence; or

iv) nucleotides 2855 - 5058 of a CDS encoding a COL4A5 polypeptide, a WPRE and a polyadenylation sequence; or

v) exons 36-51 of a CDS encoding a COL4A5 polypeptide, a WPRE and a polyadenylation sequence; or

vi) exons 38-51 of a CDS encoding a COL4A5 polypeptide, a WPRE and a polyadenylation sequence.

4. A dual adeno-associated virus (AAV) vector system according to para 3, for expressing a COL4A5 polypeptide wherein

the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 287 and

the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 217, or SEQ ID NO: 219, or SEQ ID NO: 297, SEQ ID NO: 299, SEQ ID NO: 295 or SEQ ID NO: 309.

5. A dual adeno-associated virus (AAV) vector system according to paragraph 1, for expressing a COL4A5 polypeptide wherein

(a) the first AAV vector encodes a 5'CDS encoding a COL4A5 polypeptide and a kidney specific promoter (e.g. a podocyte-specific promoter) wherein the 5' CDS consists of:

i) nucleotides 1-2474 of a CDS encoding a COL4A5 polypeptide; or

ii) nucleotides 1-2818 of a CDS encoding a COL4A5 polypeptide; or

iii) nucleotides 1-3083 of a CDS encoding a COL4A5 polypeptide; and

(b) the second AAV vector encodes a 3'CDS encoding a COL4A5 polypeptide and a WPRE and a polyadenylation sequence, wherein the 3' CDS consists of nucleotides 1724-5055 of a CDS encoding a COL4A5 polypeptide.

6. A dual adeno-associated virus (AAV) vector system according to para 5, for expressing a COL4A5 polypeptide wherein

the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 314 and

the 5' CDS consists of the nucleotide sequence of SEQ ID NO:304 or SEQ ID NO:305 or SEQ ID NO: 306.

To be clear, for embodiments of the invention using an overlapping CDS approach, the location of the CDS split into a 5'CDS and 3'CDS does not have to be precise, all that is required is sufficient overlap for homologous recombination to occur.

7. A dual adeno-associated virus (AAV) vector system according to paragraph 1,

i) wherein the 5' CDS and the 3' CDS do not share an overlapping sequence, and wherein the CDS is formed by DNA recombination, by trans-splicing or hybrid approaches, or by RNA recombination, by mRNA trans-splicing; and wherein the length of the 5' CDS and 3' CDS sequences are approximately equal; optionally wherein the 5' CDS consists of exons **1-31** of a CDS encoding a COL4A5 polypeptide, and the 3' CDS consists of exons **32-51** of a CDS encoding a COL4A5 polypeptide; optionally wherein the 5' CDS consists of exons **1-32** of a CDS encoding a COL4A5 polypeptide, and the 3' CDS consists of exons **33-51** of a CDS encoding a COL4A5 polypeptide; optionally wherein the 5' CDS consists of the nucleotide sequence SEQ ID No: 199, and the 3' CDS consists of the nucleotide sequence SEQ ID NO: 220; or

ii) wherein the 5' CDS and the 3' CDS do not share an overlapping sequence, and wherein the CDS is formed by DNA recombination, by trans-splicing or hybrid approaches; and wherein the length of the 5' CDS sequence is shorter than the 3' CDS sequence; and optionally wherein 5' CDS consists of exons **1-24** of a CDS encoding a COL4A5 polypeptide, and the 3' CDS consists of exons **25-51** of a CDS encoding a COL4A5 polypeptide; optionally wherein the 5' CDS consists of the nucleotide sequence SEQ ID No: 278, and the 3' CDS consists of the nucleotide sequence SEQ ID NO: 293; or

iii) wherein the 5' CDS and the 3' CDS do not share an overlapping sequence, and wherein the CDS is formed by DNA recombination, by trans-splicing or hybrid approaches, or by RNA recombination, by mRNA trans-splicing; and wherein the length of the 5' CDS and 3' CDS sequences are approximately equal; optionally wherein the 5' CDS consists of exons **1-33** of a CDS encoding a COL4A5 polypeptide, and the 3' CDS consists of exons **34-51** of a CDS encoding a COL4A5 polypeptide; optionally wherein the 5' CDS consists of the nucleotide sequence SEQ ID No: 186, and the 3' CDS consists of the nucleotide sequence SEQ ID NO: 222; or

iv) wherein the 5' CDS and the 3' CDS do not share an overlapping sequence, and wherein the CDS is formed by DNA recombination, by trans-splicing or hybrid approaches, or by RNA recombination, by mRNA trans-splicing; and wherein the length of the 5' CDS sequence is longer than the 3' CDS sequence; optionally wherein the 5' CDS consists of exons **1-38** of a CDS encoding a COL4A5 polypeptide, and the 3' CDS consists of exons **39-51** of a CDS

encoding a COL4A5 polypeptide; optionally wherein the 5' CDS consists of the nucleotide sequence SEQ ID No: 285, and the 3' CDS consists of the nucleotide sequence SEQ ID NO: 300.

8. A dual adeno-associated virus (AAV) vector system for expressing a full length COL4A5 polypeptide comprising:

(a) a first AAV vector comprising, in the 5' to 3' direction, a kidney specific promoter (e.g. a podocyte specific promoter), a 5' coding sequence (CDS) encoding an N-terminal part of the COL4A5 polypeptide, optionally a splice donor sequence and optionally a recombinogenic sequence; and

(b) a second AAV vector comprising, in the 5 to 3 direction, an optional recombinogenic sequence, an optional splice acceptor sequence and a 3' CDS encoding a C-terminal part of the COL4A5 polypeptide; and wherein

i) the 5' CDS of the first AAV vector comprises or consists of exons 1-31 of a CDS encoding a COL4A5 polypeptide and the 3' CDS of the second AAV vector comprises or consists of exons 32-51 of a CDS encoding a COL4A5 polypeptide; or

ii) the 5' CDS of the first AAV vector comprises or consists of exons 1-24 of a CDS encoding a COL4A5 polypeptide and the 3' CDS of the second AAV vector comprises or consists of exons 25-51 of a CDS encoding a COL4A5 polypeptide;

iii) the 5' CDS of the first AAV vector comprises or consists of exons 1-33 of a CDS encoding a COL4A5 polypeptide and the 3' CDS of the second AAV vector comprises or consists of exons 34-51 of a CDS encoding a COL4A5 polypeptide; or

iiii) the 5' CDS of the first AAV vector comprises or consists of exons 1-38 of a CDS encoding a COL4A5 polypeptide and the 3' CDS of the second AAV vector comprises or consists of exons 39-51 of a CDS encoding a COL4A5 polypeptide;

and wherein a CDS encoding the COL4A5 polypeptide is formed upon delivery of the first AAV vector and the second AAV vector into a host cell.

9. A dual adeno-associated virus (AAV) vector system of para 8, wherein the COL4A5 polypeptide is human COL4A5 polypeptide.

10. A dual AAV vector system according to any preceding para, wherein the first AAV vector and/or the second AAV vector is in the form of an AAV vector particle encapsidated by a LK03

(optionally according to SEQ ID NO: 264), or a KP1 capsid protein (optionally according to NCBI accession number MN428626).

## FURTHER EMBODIMENTS

Various further embodiments of the present invention will now be described with reference to the following numbered items.

1. A dual adeno-associated virus (AAV) vector system for expressing a COL4A3, COL4A4 or COL4A5 polypeptide comprising:

(a) a first AAV vector comprising a 5' coding sequence (CDS), wherein the 5' CDS encodes an N-terminal part of the COL4A3, COL4A4 or COL4A5 polypeptide, and a promoter upstream of the 5' CDS; and

(b) a second AAV vector comprising a 3' CDS, wherein the 3' CDS encodes a C-terminal part of the COL4A3, COL4A4 or COL4A5 polypeptide,

wherein a CDS encoding the COL4A3, COL4A4 or COL4A5 polypeptide is formed upon delivery of the first AAV vector and the second AAV vector into a host cell.

2. The dual AAV vector system according to item 1, wherein the COL4A3, COL4A4 or COL4A5 polypeptide is a human COL4A3, COL4A4 or COL4A5 polypeptide.

3. The dual AAV vector system according to item 1 or 2, wherein the COL4A3, COL4A4 or COL4A5 polypeptide comprises or consists of:

(a) an amino acid sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 1 to 7;

(b) an amino acid sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 8 to 20; or

(c) an amino acid sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 21 to 26.



4. The dual AAV vector system according to any preceding item, wherein the COL4A3, COL4A4 or COL4A5 polypeptide comprises or consists of:

(a) an amino acid sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 1;

(b) an amino acid sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 8; or

(c) an amino acid sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 21.

5. The dual AAV vector system according to any preceding item, wherein the COL4A3, COL4A4 or COL4A5 polypeptide comprises or consists of:

(a) the amino acid sequence of SEQ ID NO: 1;

(b) the amino acid sequence of SEQ ID NO: 8; or

(c) the amino acid sequence of SEQ ID NO: 21.

6. The dual AAV vector system according to any preceding item, wherein the CDS encoding the COL4A3, COL4A4 or COL4A5 polypeptide comprises or consists of:

(a) a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 27;

(b) a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 81; or

(c) a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 129.

7. The dual AAV vector system according to any preceding item, wherein the CDS encoding the COL4A3, COL4A4 or COL4A5 polypeptide comprises or consists of:

(a) the nucleotide sequence of SEQ ID NO: 27;

(b) the nucleotide sequence of SEQ ID NO: 81; or

(c) the nucleotide sequence of SEQ ID NO: 129.

8. The dual AAV vector system according to any preceding item, wherein the 5' CDS encodes an N-terminal part of the COL4A3, COL4A4 or COL4A5 polypeptide having a length of from 400 amino acids to 1300 amino acids, from 500 amino acids to 1200 amino acids, from 600 amino acids to 1100 amino acids, or from 700 amino acids to 1000 amino acids.

9. The dual AAV vector system according to any preceding item, wherein the 5' CDS comprises exons 1-10 or more, exons 1-15 or more, exons 1-20 or more, or exons 1-25 or more of a CDS encoding a COL4A3, COL4A4 or COL4A5 polypeptide.

10. The dual AAV vector system according to any preceding item, wherein the 5' CDS comprises exons 1-45 or less, exons 1-40 or less, or exons 1-35 or less of a CDS encoding a COL4A3, COL4A4 or COL4A5 polypeptide.

11. The dual AAV vector system according to any preceding item, wherein the 5' CDS comprises or consists of:

(a) exons 1-15, exons 1-16, exons 1-17, exons 1-18, exons 1-19, exons 1-20, exons 1-21, exons 1-22, exons 1-23, exons 1-24, exons 1-25, exons 1-26, exons 1-27, exons 1-28, exons 1-29, exons 1-30, exons 1-31, exons 1-32, exons 1-33, exons 1-34, exons 1-35, exons 1-36, exons 1-37, exons 1-38, exons 1-39, exons 1-40, exons 1-41, exons 1-42, or exons 1-43 of a CDS encoding a COL4A3 polypeptide;

(b) exons 1-14, exons 1-15, exons 1-16, exons 1-17, exons 1-18, exons 1-19, exons 1-20, exons 1-21, exons 1-22, exons 1-23, exons 1-24, exons 1-25, exons 1-26, exons 1-27, exons 1-28, exons 1-29, exons 1-30, exons 1-31, exons 1-32, exons 1-33, exons 1-34, exons 1-35, exons 1-36, exons 1-37, exons 1-38, or exons 1-39 of a CDS encoding a COL4A4 polypeptide; or

(c) exons 1-15, exons 1-16, exons 1-17, exons 1-18, exons 1-19, exons 1-20, exons 1-21, exons 1-22, exons 1-23, exons 1-24, exons 1-25, exons 1-26, exons 1-27, exons 1-28, exons 1-29, exons 1-30, exons 1-31, exons 1-32, exons 1-33, exons 1-34, exons 1-35, exons 1-36, exons 1-37, exons 1-38, exons 1-39, exons 1-40, or exons 1-41 of a CDS encoding a COL4A5 polypeptide.

12. The dual AAV vector system according to any preceding item, wherein the 5' CDS comprises or consists of:

- (a) exons 1-28, exons 1-29, exons 1-30, exons 1-31, exons 1-32, exons 1-33, or exons 1-34 of a CDS encoding a COL4A3 polypeptide;
- (b) exons 1-25, exons 1-26, exons 1-27, exons 1-28, exons 1-29, exons 1-30, or exons 1-31 of a CDS encoding a COL4A4 polypeptide; or
- (c) exons 1-27, exons 1-28, exons 1-29, exons 1-30, exons 1-31, exons 1-32, or exons 1-33 of a CDS encoding a COL4A5 polypeptide, or exons 1-40 of a CDS encoding a COL4A5 polypeptide.

13. The dual AAV vector system according to any preceding item, wherein the 5' CDS comprises or consists of:

- (a) a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 181 to 187;
- (b) a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 188 to 194; or
- (c) a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 195 to 201 or SEQ ID NO: 287.

14. The dual AAV vector system according to any preceding item, wherein the 5' CDS comprises or consists of:

- (a) a nucleotide sequence selected from any of SEQ ID NOs: 181 to 187;

(b) a nucleotide sequence selected from any of SEQ ID NOs: 188 to 194; or

(c) a nucleotide sequence selected from any of SEQ ID NOs: 195 to 201 or SEQ ID NO: 287.

15. The dual AAV vector system according to any preceding item, wherein the 3' CDS encodes an C-terminal part of the COL4A3, COL4A4 or COL4A5 polypeptide having a length of from 500 amino acids to 1200 amino acids, from 600 amino acids to 1100 amino acids, or from 700 amino acids to 1000 amino acids.

16. The dual AAV vector system according to any preceding item, wherein the 3' CDS comprises:

(a) exons 48-52 or more, exons 43-52 or more, exons 38-52 or more, or exons 33-52 or more of a CDS encoding a COL4A3 polypeptide;

(b) exons 43-47 or more, exons 38-47 or more, exons 33-47 or more, or exons 28-47 or more of a CDS encoding a COL4A4 polypeptide; or

(c) exons 47-51 or more, exons 42-51 or more, exons 37-51 or more, or exons 32-51 or more of a CDS encoding a COL4A5 polypeptide.

17. The dual AAV vector system according to any preceding item, wherein the 3' CDS comprises:

(a) exons 18-52 or less, exons 23-52 or less, or exons 28-52 or less of a CDS encoding a COL4A3 polypeptide;

(b) exons 13-47 or less, exons 18-47 or less, or exons 23-47 or less of a CDS encoding a COL4A4 polypeptide; or

(c) exons 12-51 or less, exons 17-51 or less, or exons 22-51 or less of a CDS encoding a COL4A5 polypeptide.

18. The dual AAV vector system according to any preceding item, wherein the 3' CDS comprises or consists of:

(a) exons 44-52, exons 43-52, exons 42-52, exons 41-52, exons 40-52, exons 39-52, exons 38-52, exons 37-52, exons 36-52, exons 35-52, exons 34-52, exons 33-52, exons 32-52, exons 31-52, exons 30-52, exons 29-52, exons 28-52, exons 27-52, exons 26-52, exons 25-52, exons 23-52, exons 23-52, exons 22-52, exons 21-52,

exons 20-52, exons 19-52, exons 18-52, exons 17-52, or exons 16-52 of a CDS encoding a COL4A3 polypeptide;

(b) exons 40-47, exons 39-47, exons 38-47, exons 37-47, exons 36-47, exons 35-47, exons 34-47, exons 33-47, exons 32-47, exons 31-47, exons 30-47, exons 29-47, exons 28-47, exons 27-47, exons 26-47, exons 25-47, exons 24-47, exons 23-47, exons 22-47, exons 21-47, exons 20-47, exons 19-47, exons 18-47, exons 17-47, exons 16-47, or exons 15-47 of a CDS encoding a COL4A4 polypeptide; or

(c) exons 42-51, exons 41-51, exons 40-51, exons 39-51, exons 38-51, exons 37-51, exons 36-51, exons 35-51, exons 34-51, exons 33-51, exons 32-51, exons 31-51, exons 30-51, exons 29-51, exons 28-51, exons 27-51, exons 26-51, exons 25-51, exons 24-51, exons 23-51, exons 22-51, exons 21-51, exons 20-51, exons 19-51, exons 18-51, exons 17-51, or exons 16-51 of a CDS encoding a COL4A5 polypeptide.

19. The dual AAV vector system according to any preceding item, wherein the 3' CDS comprises or consists of:

(a) exons 35-52, exons 34-52, exons 33-52, exons 32-52, exons 31-52, exons 30-52, or exons 29-52 of a CDS encoding a COL4A3 polypeptide;

(b) exons 32-47, exons 31-47, exons 30-47, exons 29-47, exons 28-47, exons 27-47, or exons 26-47, of a CDS encoding a COL4A4 polypeptide; or

(c) exons 36-51, exons 35-51, exons 34-51, exons 33-51, exons 32-51, exons 31-51, exons 30-51, exons 29-51, exons 28-51, or exons 27-51, of a CDS encoding a COL4A5 polypeptide.

20. The dual AAV vector system according to any preceding item, wherein the 3' CDS comprises or consists of:

(a) a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 202 to 208;

(b) a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 209 to 215; or

(c) a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 295 to 297, SEQ ID NOs: 216 to 222 or SEQ ID NOs: 308 to 310.

21. The dual AAV vector system according to any preceding item, wherein the 3' CDS comprises or consists of:

- (a) a nucleotide sequence selected from any of SEQ ID NOs: 202 to 208;
- (b) a nucleotide sequence selected from any of SEQ ID NOs: 209 to 215; or
- (c) a nucleotide sequence selected from any of SEQ ID NOs: 295 to 297, SEQ ID NOs: 216 to 222 or SEQ ID NOs: 308 to 310.

22. The dual AAV vector system according to any preceding item, wherein:

- (a)
  - (i) the 5' CDS consists of exons 1-28 of a CDS encoding a COL4A3 polypeptide and the 3' CDS consists of exons 29-52 of a CDS encoding a COL4A3 polypeptide;
  - (ii) the 5' CDS consists of exons 1-29 of a CDS encoding a COL4A3 polypeptide and the 3' CDS consists of exons 30-52 of a CDS encoding a COL4A3 polypeptide;
  - (iii) the 5' CDS consists of exons 1-30 of a CDS encoding a COL4A3 polypeptide and the 3' CDS consists of exons 31-52 of a CDS encoding a COL4A3 polypeptide;
  - (iv) the 5' CDS consists of exons 1-31 of a CDS encoding a COL4A3 polypeptide and the 3' CDS consists of exons 32-52 of a CDS encoding a COL4A3 polypeptide;
  - (v) the 5' CDS consists of exons 1-32 of a CDS encoding a COL4A3 polypeptide and the 3' CDS consists of exons 33-52 of a CDS encoding a COL4A3 polypeptide;
  - (vi) the 5' CDS consists of exons 1-33 of a CDS encoding a COL4A3 polypeptide and the 3' CDS consists of exons 34-52 of a CDS encoding a COL4A3 polypeptide; or
  - (vii) the 5' CDS consists of exons 1-34 of a CDS encoding a COL4A3 polypeptide and the 3' CDS consists of exons 35-52 of a CDS encoding a COL4A3 polypeptide;
- (b)
  - (i) the 5' CDS consists of exons 1-25 of a CDS encoding a COL4A4 polypeptide and the 3' CDS consists of exons 26-47 of a CDS encoding a COL4A4 polypeptide;
  - (ii) the 5' CDS consists of exons 1-26 of a CDS encoding a COL4A4 polypeptide and the 3' CDS consists of exons 27-47 of a CDS encoding a COL4A4 polypeptide;
  - (iii) the 5' CDS consists of exons 1-27 of a CDS encoding a COL4A4 polypeptide and the 3' CDS consists of exons 28-47 of a CDS encoding a COL4A4 polypeptide;
  - (iv) the 5' CDS consists of exons 1-28 of a CDS encoding a COL4A4 polypeptide and the 3' CDS consists of exons 29-47 of a CDS encoding a COL4A4 polypeptide;
  - (v) the 5' CDS

consists of exons 1-29 of a CDS encoding a COL4A4 polypeptide and the 3' CDS consists of exons 30-47 of a CDS encoding a COL4A4 polypeptide; (vi) the 5' CDS consists of exons 1-30 of a CDS encoding a COL4A4 polypeptide and the 3' CDS consists of exons 31-47 of a CDS encoding a COL4A4 polypeptide; or (vii) the 5' CDS consists of exons 1-31 of a CDS encoding a COL4A4 polypeptide and the 3' CDS consists of exons 32-47 of a CDS encoding a COL4A4 polypeptide; or

(c) (i) the 5' CDS consists of exons 1-27 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 28-51 of a CDS encoding a COL4A5 polypeptide; (ii) the 5' CDS consists of exons 1-28 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 29-51 of a CDS encoding a COL4A5 polypeptide; (iii) the 5' CDS consists of exons 1-29 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 30-51 of a CDS encoding a COL4A5 polypeptide; (iv) the 5' CDS consists of exons 1-30 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 31-51 of a CDS encoding a COL4A5 polypeptide; (v) the 5' CDS consists of exons 1-31 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 32-51 of a CDS encoding a COL4A5 polypeptide; (vi) the 5' CDS consists of exons 1-32 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 33-51 of a CDS encoding a COL4A5 polypeptide; or (vii) the 5' CDS consists of exons 1-33 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 34-51 of a CDS encoding a COL4A5 polypeptide.

23. The dual AAV vector system according to any preceding item, wherein:

(a) (i) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 181 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 202; (ii) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 182 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 203; (iii) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 183 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 204; (iv) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 184 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 205; (v) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 185 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 206; (vi) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 186 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 207; or (vii) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 187 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 208;

(b) (i) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 188 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 209; (ii) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 189 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 210; (iii) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 190 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 211; (iv) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 191 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 212; (v) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 192 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 213; (vi) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 193 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 214; or (vii) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 194 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 215; or

(c) (i) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 195 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 216; (ii) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 196 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 217; (iii) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 197 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 218; (iv) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 198 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 219; (v) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 199 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 220; (vi) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 200 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 221; or (vii) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 201 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 222.

24. A dual adeno-associated virus (AAV) vector system for expressing a COL4A5 polypeptide comprising:

(a) a first AAV vector comprising a 5' coding sequence (CDS), wherein the 5' CDS encodes an N-terminal part of the COL4A5 polypeptide, and a promoter upstream of the 5' CDS; and

(b) a second AAV vector comprising a 3' CDS, wherein the 3' CDS encodes a C-terminal part of the COL4A5 polypeptide,



wherein a CDS encoding the COL4A5 polypeptide is formed upon delivery of the first AAV vector and the second AAV vector into a host cell.

25. The dual AAV vector system according to any preceding item, wherein the COL4A5 polypeptide is a human COL4A5 polypeptide.

26. The dual AAV vector system according to item 24 or 25, wherein the COL4A5 polypeptide comprises or consists of an amino acid sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 21 to 26.

27. The dual AAV vector system according to any of items 24-26, wherein the COL4A5 polypeptide comprises or consists of an amino acid sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 21.

28. The dual AAV vector system according to any of items 24-27, wherein the COL4A5 polypeptide comprises or consists of the amino acid sequence of SEQ ID NO: 21.

29. The dual AAV vector system according to any of items 24-28, wherein the CDS encoding the COL4A5 polypeptide comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 129.

30. The dual AAV vector system according to any of items 24-29, wherein the CDS encoding the COL4A5 polypeptide comprises or consists of the nucleotide sequence of SEQ ID NO: 129.

31. The dual AAV vector system according to any of items 24-30, wherein the 5' CDS comprises or consists of exons 1-15, exons 1-16, exons 1-17, exons 1-18, exons 1-19, exons 1-20, exons 1-21, exons 1-22, exons 1-23, exons 1-24, exons 1-25, exons 1-26, exons 1-27, exons 1-28, exons 1-29, exons 1-30, exons 1-31, exons 1-32, exons 1-33, exons 1-34, exons 1-35, exons 1-36, exons 1-37, exons 1-38, exons 1-39, exons 1-40, or exons 1-41 of a CDS encoding a COL4A5 polypeptide.

32. The dual AAV vector system according to any of items 24-31, wherein the 5' CDS comprises or consists of exons 1-27, exons 1-28, exons 1-29, exons 1-30, exons 1-31, exons

1-32, or exons 1-33 of a CDS encoding a COL4A5 polypeptide, optionally wherein the 5' CDS comprises or consists of exons 1-29, exons 1-31, exons 1-32, or exons 1-33 of a CDS encoding a COL4A5 polypeptide; or wherein the 5' CDS comprises or consists of exons 1-40 of a CDS encoding a COL4A5 polypeptide.

33. The dual AAV vector system according to any of items 24-32, wherein the 5' CDS comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 195 to 201, optionally wherein the 5' CDS comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 197, 199, 200, or 201; or wherein the 5' CDS comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 287.

34. The dual AAV vector system according to any of items 24-33, wherein the 5' CDS comprises or consists of the nucleotide sequence selected from any of SEQ ID NOs: 195 to 201, optionally wherein the 5' CDS comprises or consists of the nucleotide sequence selected from any of SEQ ID NOs: 197, 199, 200, or 201; or wherein the 5' CDS comprises or consists of SEQ ID NO: 287.

35. The dual AAV vector system according to any of items 24-34, wherein the 3' CDS comprises or consists of exons 42-51, exons 41-51, exons 40-51, exons 39-51, exons 38-51, exons 37-51, exons 36-51, exons 35-51, exons 34-51, exons 33-51, exons 32-51, exons 31-51, exons 30-51, exons 29-51, exons 28-51, exons 27-51, exons 26-51, exons 25-51, exons 24-51, exons 23-51, exons 22-51, exons 21-51, exons 20-51, exons 19-51, exons 18-51, exons 17-51, or exons 16-51 of a CDS encoding a COL4A5 polypeptide.

36. The dual AAV vector system according to any of items 24-35, wherein the 3' CDS comprises or consists of exons 36-41, exons 35-51, exons 34-51, exons 33-51, exons 32-51, exons 31-51, exons 30-51, exons 29-51, exons 28-51 or exons 27-51, of a CDS encoding a COL4A5 polypeptide, optionally wherein the 3' CDS comprises or consists of exons 34-51, exons 33-51, exons 32-51, or exons 30-51 of a CDS encoding a COL4A5 polypeptide or the 3' CDS comprises or consists of exons 36-41, exons 31-51 or exons 29-31.

37. The dual AAV vector system according to any of items 24-36, wherein the 3' CDS comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 295 to 297, SEQ ID NOs: 216 to 222 or SEQ ID NOs: 308 to 310, optionally wherein the 3' CDS comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 218, 220, 221, or 222 or SEQ ID NOs: 217, 219 or 297.

38. The dual AAV vector system according to any of items 24-37, wherein the 3' CDS comprises or consists of a nucleotide sequence selected from any of SEQ ID NOs: 295 to 297, SEQ ID NOs: 216 to 222 or SEQ ID NOs: 308 to 310, optionally wherein the 3' CDS comprises or consists of a nucleotide sequence selected from any of SEQ ID NOs: 218, 220, 221, or 222 or SEQ ID NOs: 217, 219 or 297.

39. The dual AAV vector system according to any of items 24-38, wherein:

- (a) (i) the 5' CDS consists of exons 1-27 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 28-51 of a CDS encoding a COL4A5 polypeptide;
- (ii) the 5' CDS consists of exons 1-28 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 29-51 of a CDS encoding a COL4A5 polypeptide;
- (iii) the 5' CDS consists of exons 1-29 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 30-51 of a CDS encoding a COL4A5 polypeptide;
- (iv) the 5' CDS consists of exons 1-30 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 31-51 of a CDS encoding a COL4A5 polypeptide;
- (v) the 5' CDS consists of exons 1-31 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 32-51 of a CDS encoding a COL4A5 polypeptide;
- (vi) the 5' CDS consists of exons 1-32 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 33-51 of a CDS encoding a COL4A5 polypeptide; or
- (vii) the 5' CDS consists of exons 1-33 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 34-51 of a CDS encoding a COL4A5 polypeptide; or
- (b) (i) the 5' CDS consists of exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 36-51 of a CDS encoding a COL4A5 polypeptide;

- (ii) the 5' CDS consists of exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 35-51 of a CDS encoding a COL4A5 polypeptide;
- (iii) the 5' CDS consists of exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 34-51 of a CDS encoding a COL4A5 polypeptide;
- (iv) the 5' CDS consists of exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 33-51 of a CDS encoding a COL4A5 polypeptide;
- (v) the 5' CDS consists of exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 32-51 of a CDS encoding a COL4A5 polypeptide;
- (vi) the 5' CDS consists of exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 31-51 of a CDS encoding a COL4A5 polypeptide;
- (vii) the 5' CDS consists of exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 30-51 of a CDS encoding a COL4A5 polypeptide;
- (viii) the 5' CDS consists of exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 29-51 of a CDS encoding a COL4A5 polypeptide;
- (ix) the 5' CDS consists of exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 28-51 of a CDS encoding a COL4A5 polypeptide; or
- (x) the 5' CDS consists of exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS consists of exons 27-51 of a CDS encoding a COL4A5 polypeptide.

40. The dual AAV vector system according to any of items 24-39, wherein:

- (a) (i) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 195 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 216;
- (ii) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 196 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 217;
- (iii) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 197 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 218;
- (iv) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 198 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 219;

- (v) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 199 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 220;
- (vi) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 200 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 221; or
- (vii) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 201 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 222; or
- (b) (i) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 287 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 297;
- (ii) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 287 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 310;
- (iii) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 287 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 309;
- (iv) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 287 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 308;
- (v) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 287 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 219;
- (vi) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 287 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 217; or
- (vii) the 5' CDS consists of the nucleotide sequence of SEQ ID NO: 287 and the 3' CDS consists of the nucleotide sequence of SEQ ID NO: 295.

41. The dual AAV vector system according to any preceding item, wherein the promoter upstream of the 5' CDS is selected from: a CMV promoter, a CBA promoter, a CAG promoter, a CB7 promoter, an EF1 $\alpha$  promoter, a CMV/EF1 $\alpha$  hybrid promoter, a NF-kB promoter, a pSE-7 promoter, a mPGK promoter, a mU1a promoter, a U6 promoter, a U7 promoter, a MNDU3 promoter, a HLP promoter, an AAT promoter, an ALB promoter, a ApoE/AAT promoter, a EalbAAT promoter, a LP1 promoter, a TBG promoter, a TTR promoter, a SYN1 promoter, a NSE promoter, a tMCK promoter, a CK8 promoter, a MHCK7 promoter, a SMN promoter, a DES promoter, a RK promoter, a hRHO promoter, a GRK1 promoter, a hCAR promoter, a hRPE65p promoter, a P546 promoter, a PR1.7 promoter, a hRS1 promoter, a VMD2 promoter, an  $\alpha$ -MHC promoter, a FRE1 promoter, a NPHS1 promoter, and a NPHS2 promoter.

42. The dual AAV vector system according to any preceding item, wherein the promoter upstream of the 5' CDS is a ubiquitous promoter.

43. The dual AAV vector system according to any preceding item, wherein the promoter upstream of the 5' CDS is a CMV promoter, a CBA promoter, or a CAG promoter, optionally wherein the promoter upstream of the 5' CDS is a CMV promoter.

44. The dual AAV vector system according to any preceding item, wherein the promoter upstream of the 5' CDS comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 223.

45. The dual AAV vector system according to any of items 1-41, wherein the promoter upstream of the 5' CDS is a kidney-specific promoter, optionally wherein the promoter upstream of the 5' CDS is a podocyte-specific promoter.

46. The dual AAV vector system according to item 45, wherein the promoter upstream of the 5' CDS is a NPHS1 promoter or a NPHS2 promoter, optionally wherein the promoter upstream of the 5' CDS is a minimal NPHS1 promoter or a minimal NPHS2 promoter.

47. The dual AAV vector system according to item 45 or 46, wherein the promoter upstream of the 5' CDS comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 226 to 231.

48. The dual AAV vector system according to any preceding item, wherein the second AAV vector comprises a polyadenylation sequence downstream from the 3' CDS.

49. The dual AAV vector system according to item 48, wherein the polyadenylation sequence downstream from the 3' CDS is a bovine growth hormone polyadenylation sequence (bGH), a soluble neuropilin-1 polyadenylation sequence, an early SV40 polyadenylation sequence (SV40pA), or a chicken beta-globin polyadenylation sequence; and/or wherein the polyadenylation sequence downstream from the 3' CDS comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 236 to 239.

50. The dual AAV vector system according to any preceding item, wherein the first AAV vector comprises a splice donor sequence downstream from the 5' CDS, and wherein the second AAV vector comprises a splice acceptor sequence upstream from the 3' CDS.

51. The dual AAV vector system according to item 50, wherein the splice donor sequence comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 240.

52. The dual AAV vector system according to item 50 or 51, wherein the splice acceptor sequence comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 242.

53. The dual AAV vector system according to any of items 50-52, wherein the first AAV vector and the second AAV vector each comprise an identical recombinogenic sequence, wherein the first AAV vector further comprises the recombinogenic sequence downstream from the splice donor sequence, and wherein the second AAV vector further comprises the recombinogenic sequence upstream from the splice acceptor sequence.

54. The dual AAV vector system according to item 53, wherein the recombinogenic sequence is a recombinogenic region from the F1 phage sequence or a recombinogenic region from the alkaline phosphatase gene.

55. The dual AAV vector system according to item 53 or 54, wherein the recombinogenic sequence comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 244-247.

56. The dual AAV vector system according to any of items 53-55, wherein the recombinogenic sequence is a recombinogenic region from the F1 phage sequence.

57. The dual AAV vector system according to any of items 53-56, wherein the recombinogenic sequence comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 244.

58. The dual AAV vector system according to any preceding item, wherein:

(a) the first AAV vector comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 248 downstream from the 5' CDS; and

(b) the second AAV vector comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 250 upstream from the 3' CDS.

59. The dual AAV vector system according to any preceding item, wherein:

(a) the first AAV vector comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 249 downstream from the 5' CDS; and

(b) the second AAV vector comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 251 upstream from the 3' CDS.

60. The dual AAV vector system according to any preceding item, wherein the first AAV vector and/or the second AAV vector comprises one or more regulatory sequences.

61. The dual AAV vector system according to any preceding item, wherein the first AAV vector comprises one or more regulatory sequences upstream from the 5' CDS.

62. The dual AAV vector system according to item 61, wherein the first AAV vector comprises a spliceosomal intron or a fragment thereof comprising a branch point sequence, a polypyrimidine tract and a 3' splice acceptor site, upstream from the 5' CDS.

63. The dual AAV vector system according to item 62, wherein the spliceosomal intron or a fragment thereof comprises or consists of the nucleotide sequence  $(N)_x Y T N A Y (N)_{n1} (Y)_{n2} (N)_{n3} Y A G G$ , wherein  $x = 10$  to  $100$ ,  $n1 = 2$  to  $22$ , wherein  $n2 = 10$  to  $20$ , and wherein  $n1 + n2 + n3 = 15$  to  $40$ .

64. The dual AAV vector system according to item 62 or 63, wherein the spliceosomal intron or a fragment thereof comprises or consists of a nucleotide sequence having 80% or more,



85% or more, 90% or more, 91% or more, 92% or more, 93% or more, 94% or more, 95% or more, 96% or more, 97% or more, 98% or more, 99% or more, or 100% identity to SEQ ID NO: 256 or 257.

65. The dual AAV vector system according to any preceding item, wherein the second AAV vector comprises one or more regulatory sequences downstream from the 3' CDS.

66. The dual AAV vector system according to item 65, wherein the second AAV vector comprises a Woodchuck hepatitis post-transcriptional regulatory element (WPRE), downstream from the 3' CDS.

67. The dual AAV vector system according to item 65 or 66, wherein the WPRE comprises or consists of a nucleotide sequence having 80% or more, 85% or more, 90% or more, 91% or more, 92% or more, 93% or more, 94% or more, 95% or more, 96% or more, 97% or more, 98% or more, 99% or more, or 100% identity to SEQ ID NO: 263.

68. The dual AAV vector system according to any of items 1-49, wherein the second AAV vector further comprises a promoter upstream from the 3' CDS.

69. The dual AAV vector system according to item 68, wherein the promoter upstream of the 3' CDS is selected from: a CMV promoter, a CBA promoter, a CAG promoter, a CB7 promoter, an EF1 $\alpha$  promoter, a CMV/EF1 $\alpha$  hybrid promoter, a NF-kB promoter, a pSE-7 promoter, a mPGK promoter, a mU1a promoter, a U6 promoter, a U7 promoter, a MNDU3 promoter, a HLP promoter, an AAT promoter, an ALB promoter, a ApoE/AAT promoter, a EalbAAT promoter, a LP1 promoter, a TBG promoter, a TTR promoter, a SYN1 promoter, a NSE promoter, a tMCK promoter, a CK8 promoter, a MHCK7 promoter, a SMN promoter, a DES promoter, a RK promoter, a hRHO promoter, a GRK1 promoter, a hCAR promoter, a hRPE65p promoter, a P546 promoter, a PR1.7 promoter, a hRS1 promoter, a VMD2 promoter, an  $\alpha$ -MHC promoter, a FRE1 promoter, a NPHS1 promoter, and a NPHS2 promoter.

70. The dual AAV vector system according to item 68 or 69, wherein the promoter upstream of the 3' CDS is a ubiquitous promoter.

71. The dual AAV vector system according to any of items 68-70, wherein the promoter upstream of the 3' CDS is a CMV promoter, a CBA promoter, or a CAG promoter, optionally wherein the promoter upstream of the 3' CDS is a CMV promoter.

72. The dual AAV vector system according to any of items 68-71, wherein the promoter upstream of the 3' CDS comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least

93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 223.

73. The dual AAV vector system according to item 68 or 69, wherein the promoter upstream of the 3' CDS is a kidney-specific promoter, optionally wherein the promoter upstream of the 3' CDS is a podocyte-specific promoter.

74. The dual AAV vector system according to item 73, wherein the promoter upstream of the 3' CDS is a NPHS1 promoter or a NPHS2 promoter, optionally wherein the promoter upstream of the 3' CDS is a minimal NPHS1 promoter or a minimal NPHS2 promoter.

75. The dual AAV vector system according to item 73 or 74, wherein the promoter upstream of the 3' CDS comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 226 to 231.

76. The dual AAV vector system according to any of items 68-75, wherein the promoter upstream of the 5' CDS and the promoter upstream of the 3' CDS are identical.

77. The dual AAV vector system according to any of items 68-76, wherein the first AAV vector further comprises a polyadenylation sequence downstream from the 5' CDS.

78. The dual AAV vector system according to item 77, wherein the polyadenylation sequence downstream from the 5' CDS is a bovine growth hormone polyadenylation sequence (bGH), a soluble neuropilin-1 polyadenylation sequence, an early SV40 polyadenylation sequence (SV40pA), or a chicken beta-globin polyadenylation sequence.

79. The dual AAV vector system according to item 77 or 78, wherein the polyadenylation sequence downstream from the 5' CDS comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 236 to 239.

80. The dual AAV vector system according to any of items 68-79, wherein:

- (a) the first AAV vector further comprises a splice donor sequence and a hybridisation domain downstream from the 5' CDS and upstream from the polyadenylation sequence; and

(b) the second AAV vector further comprises a complementary hybridisation domain and a splice acceptor sequence upstream from the 3' CDS and downstream from the promoter.

81. The dual AAV vector system according to any of items 68-80, wherein the first AAV vector comprises one or more regulatory sequences upstream from the 5' CDS and/or the second AAV vector comprises one or more regulatory sequences upstream from the 3' CDS.

82. The dual AAV vector system according to item 81, wherein the first AAV vector comprises a spliceosomal intron or a fragment thereof comprising a branch point sequence, a polypyrimidine tract and a 3' splice acceptor site, upstream from the 5' CDS; and/or the second AAV vector comprises a spliceosomal intron or a fragment thereof comprising a branch point sequence, a polypyrimidine tract and a 3' splice acceptor site, downstream from the 5' CDS.

83. The dual AAV vector system according to item 82, wherein the spliceosomal intron or a fragment thereof comprises or consists of the nucleotide sequence  $(N)_x YTNAY(N)_{n1}(Y)_{n2}(N)_{n3}YAGG$ , wherein  $x = 10$  to  $100$ ,  $n1 = 2$  to  $22$ , wherein  $n2 = 10$  to  $20$ , and wherein  $n1 + n2 + n3 = 15$  to  $40$ .

84. The dual AAV vector system according to item 82 or 83, wherein the spliceosomal intron or a fragment thereof comprises or consists of a nucleotide sequence having 80% or more, 85% or more, 90% or more, 91% or more, 92% or more, 93% or more, 94% or more, 95% or more, 96% or more, 97% or more, 98% or more, 99% or more, or 100% identity to SEQ ID NO: 256 or 257.

85. The dual AAV vector system according to any of items 68-84, wherein the first AAV vector comprises one or more regulatory sequences downstream from the 5' CDS, and/or the second AAV vector comprises one or more regulatory sequences downstream from the 3' CDS.

86. The dual AAV vector system according to item 85, wherein the first AAV vector comprises a Woodchuck hepatitis post-transcriptional regulatory element (WPRE), downstream from the 5' CDS, and/or the second AAV vector comprises a Woodchuck hepatitis post-transcriptional regulatory element (WPRE), downstream from the 3' CDS.

87. The dual AAV vector system according to item 86, wherein the WPRE comprises or consists of a nucleotide sequence having 80% or more, 85% or more, 90% or more, 91% or more, 92% or more, 93% or more, 94% or more, 95% or more, 96% or more, 97% or more, 98% or more, 99% or more, or 100% identity to SEQ ID NO: 263.

88. The dual AAV vector system according to any preceding item, wherein the first AAV vector and/or the second AAV vector is in the form of an AAV vector particle, preferably wherein the first AAV vector and the second AAV vector are each in the form of AAV vector particles.

89. The dual AAV vector system according to any preceding item, wherein the first AAV vector and/or the second AAV vector is in the form of an AAV vector particle encapsidated by LK03, AAV3B, AAV9, ShH10, AAV-DJ, AAV2, AAV6.2, or AAV5 capsid proteins.

90. The dual AAV vector system according to any preceding item, wherein the first AAV vector and/or the second AAV vector is in the form of an AAV vector particle encapsidated by capsid proteins comprising or consisting of an amino acid sequence having at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 264 to 271.

91. The dual AAV vector system according to item 89 or 90, wherein the first AAV vector and the second AAV vector are encapsidated by the same capsid proteins.

92. An isolated cell comprising the dual AAV vector system according to any preceding item.

93. A pharmaceutical composition comprising the dual AAV vector system according to any of items 1 to 91 or a cell according to item 92.

94. A dual AAV vector system according to any of items 1 to 91, a cell according to item 92 or a pharmaceutical composition according to item 93, for use as a medicament.

95. Use of a dual AAV vector system according to any of items 1 to 91, a cell according to item 92 or a pharmaceutical composition according to item 93, for the manufacture of a medicament.

96. A product comprising:

(a) a first AAV vector comprising a promoter and a 5' coding sequence (CDS), wherein the 5' CDS encodes an N-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide; and

(b) a second AAV vector comprising a 3' CDS, wherein the 3' CDS encodes a C-terminal part of the COL4A3, COL4A4 or COL4A5 polypeptide,

as a combined preparation for simultaneous, separate or sequential use in therapy, wherein a CDS encoding the COL4A3, COL4A4 or COL4A5 polypeptide is formed and the COL4A3,

COL4A4 or COL4A5 polypeptide is expressed upon administration of the first AAV vector and the second AAV vector to a subject.

97. The product for use according to item 96, wherein the first AAV vector and the second AAV vector are defined according to any of items 1 to 91.

98. A dual AAV vector system according to any of items 1 to 91, a cell according to item 92 or a pharmaceutical composition according to item 93, for use in preventing and/or treating Alport Syndrome.

99. Use of a dual AAV vector system according to any of items 1 to 91, a cell according to item 92 or a pharmaceutical composition according to item 93, for the manufacture of a medicament for preventing and/or treating Alport Syndrome.

100. A product comprising:

(a) a first AAV vector comprising a promoter and a 5' coding sequence (CDS), wherein the 5' CDS encodes an N-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide; and

(b) a second AAV vector comprising a 3' CDS, wherein the 3' CDS encodes a C-terminal part of the COL4A3, COL4A4 or COL4A5 polypeptide,

as a combined preparation for simultaneous, separate or sequential use in preventing and/or treating Alport Syndrome in a subject, wherein a CDS encoding the COL4A3, COL4A4 or COL4A5 polypeptide is formed and the COL4A3, COL4A4 or COL4A5 polypeptide is expressed upon administration of the first AAV vector and the second AAV vector to the subject.

101. The product for use according to item 100, wherein the first AAV vector and the second AAV vector are defined according to any of items 1 to 91.

102. A method of preventing and/or treating Alport Syndrome comprising administering a dual AAV vector system according to any of items 1 to 91, a cell according to item 92 or a pharmaceutical composition according to item 93, to a subject in need thereof.

103. A method of preventing and/or treating Alport Syndrome comprising administering to a subject in need thereof:

(a) a first AAV vector comprising a promoter and a 5' coding sequence (CDS), wherein the 5' CDS encodes an N-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide; and

(b) a second AAV vector comprising a 3' CDS, wherein the 3' CDS encodes a C-terminal part of the COL4A3, COL4A4 or COL4A5 polypeptide,

wherein the first AAV vector and the second AAV vector are administered simultaneously, separately or sequentially, and wherein a CDS encoding the COL4A3, COL4A4 or COL4A5 polypeptide is formed and the COL4A3, COL4A4 or COL4A5 polypeptide is expressed in the subject upon administration of the first AAV vector and the second AAV vector to the subject.

104. The method according to item 103, wherein the first AAV vector and the second AAV vector are defined according to any of items 1 to 91.

105. The dual AAV vector system, cell or pharmaceutical composition for use according to item 94 or 98, the use according to item 95 or 99, the product for use according to any of items 96-97 or 100-101, or the method according to any of items 102-104, wherein the first AAV vector and second AAV vector are administered simultaneously.

106. The dual AAV vector system, cell or pharmaceutical composition for use according to any of items 94, 98 or 105, the use according to any of items 95, 99 or 105, the product for use according to any of items 96-97, 100-101 or 105, or the method according to any of items 102-105, wherein the first AAV vector, second AAV vector, cell or pharmaceutical composition is administered by renal artery injection, by intraparenchymal injection, by transparenchymal injection, by renal vein injection, by ureteral injection, by intravenous administration, or by subcapsular injection, preferably wherein the first AAV vector, second AAV vector, cell or pharmaceutical composition is administered by renal artery injection.

107. The dual AAV vector system, cell or pharmaceutical composition for use according to any of items 94, 98 or 105-106, the use according to any of items 95, 99 or 105-106, the product for use according to any of items 96-97, 100-101 or 105-106, or the method according to any of items 102-106, wherein the first AAV vector, second AAV vector, cell or pharmaceutical composition is administered in a dose of from about  $1 \times 10^6$  vg/kg to about  $1 \times 10^{14}$  vg/kg.

108. A kit or system for production of AAV vector particles comprising:

(a) a first transfer vector encoding a first AAV vector; and

(b) a second transfer vector encoding a second AAV vector,

wherein the first AAV vector and the second AAV vector are defined according to any of items 1 to 91.

109. The kit or system according to item 108, wherein the kit further comprises one or more helper vectors encoding AAV replication and capsid proteins.

110. An isolated cell comprising the kit or system according to item 108 or 109.

111. The cell according to item 110, wherein the cell is a packaging cell.

**CLAIMS**

1. A dual adeno-associated virus (AAV) vector system for expressing a COL4A3, COL4A4 or COL4A5 polypeptide in a host cell.

2. A dual adeno-associated virus (AAV) vector system according to claim 1 comprising:

(a) a first AAV vector comprising a 5' coding sequence (CDS), wherein the 5' CDS encodes an N-terminal part of the COL4A3, COL4A4 or COL4A5 polypeptide, and a promoter upstream of the 5' CDS; and

(b) a second AAV vector comprising a 3' CDS, wherein the 3' CDS encodes a C-terminal part of the COL4A3, COL4A4 or COL4A5 polypeptide,

wherein a CDS encoding the COL4A3, COL4A4 or COL4A5 polypeptide is formed upon delivery of the first AAV vector and the second AAV vector into the host cell.

3. A dual adeno-associated virus (AAV) vector system according to claim 2,

wherein the 5' CDS and the 3' CDS share an overlapping sequence.

4. A dual adeno-associated virus (AAV) vector system according to claim 3, wherein the overlapping sequence has a length of at least 200 nucleotides.

5. A dual AAV vector system of claim 4, wherein the overlapping sequence has a length of at least 200 nucleotides, at least 250 nucleotides, at least 300 nucleotides, at least 350 nucleotides, at least 400 nucleotides, at least 450 nucleotides, at least 500 nucleotides, at least 550 nucleotides, at least 600 nucleotides, at least 700 nucleotides, at least 750 nucleotides, at least 800 nucleotides, at least 900 nucleotides, at least 1000 nucleotides, at least 1300 nucleotides, at least 1400 nucleotides, at least 1500 nucleotides, at least 1600 nucleotides, at least 1700 nucleotides, at least 1800 nucleotides or at least 1900 nucleotides.

6. A dual AAV vector system according to any one of claims 3 to 5, wherein, the overlapping sequence has a length of 2000 nucleotides or fewer, preferably wherein the overlapping sequence has a length of from 200 nucleotides to 2000 nucleotides, from 450 nucleotides to 2000 nucleotides, from 450 nucleotides to 1600 nucleotides, from 750 nucleotides to 1600 nucleotides, from 1000 nucleotides to 1600 nucleotides, 1300 nucleotides to 1600 nucleotides, from 750 nucleotides to 1300 nucleotides, from 750 nucleotides to 1400 nucleotides, or from 1000 nucleotides to 1400 nucleotides.



7. A dual AAV vector system of any of claims 3 to 6, wherein the 5' CDS sequence is longer than 3' CDS sequence, and optionally wherein the overlapping sequence has a length of from 200 nucleotides to 2000 nucleotides, optionally wherein the overlapping sequence has a length of from 450 nucleotides to 1600 nucleotides.
8. A dual AAV vector system of claim 7, for expressing a COL4A5 polypeptide.
9. A dual AAV vector system of claim 8, wherein the 5' CDS comprises a minimum of exons 1-36 to a maximum of exons 1-44 of a CDS encoding a COL4A5 polypeptide, and wherein the overlapping sequence has a length of from 450 nucleotides to 1600 nucleotides, optionally wherein the 5'CDS comprises exons 1-40.
10. A dual AAV vector system of claim 8, wherein the 5' CDS comprises a minimum of exons 1-36 to a maximum of exons 1-44 of a CDS encoding a COL4A5 polypeptide, and wherein the 3' CDS comprises a minimum of exons 45-51 to a maximum of exons 37-51 of a CDS encoding a COL4A5 polypeptide.
11. A dual AAV vector system of claim 8, wherein the overlapping sequence has a length selected from: between 130 and 330 nucleotides, optionally 230 nucleotides; between 398 and 598, nucleotides, optionally 498 nucleotides; between 650 to 850 nucleotides, optionally 750 nucleotides; between 995 to 1195 nucleotides, optionally 1095 nucleotides; between 1260 and 1460 nucleotides, optionally 1360 nucleotides; and between 1466 and 1666 nucleotides, optionally 1566 nucleotides.
12. A dual AAV vector system of claim 8, wherein
- i) the 5' CDS comprises exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS comprises exons 38-51 of a CDS encoding a COL4A5 polypeptide, wherein the overlapping sequence has a length of between 130 and 330 nucleotides, optionally 230 nucleotides;
  - ii) the 5' CDS comprises exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS comprises exons 36-51 of a CDS encoding a COL4A5 polypeptide, wherein the overlapping sequence has a length of between 398 and 598, nucleotides, optionally 498 nucleotides;
  - iii) the 5' CDS comprises exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS comprises nucleotides 2855 - 5058 of a CDS encoding a COL4A5 polypeptide, wherein the overlapping sequence has a length of between 650 to 850 nucleotides, optionally 750 nucleotides;

iv) the 5' CDS comprises exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS comprises exons 31-51 of a CDS encoding a COL4A5 polypeptide, wherein the overlapping sequence has a length of between 995 to 1195 nucleotides, optionally 1095 nucleotides;

v) the 5' CDS comprises exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS comprises exons 29-51 of a CDS encoding a COL4A5 polypeptide, wherein the overlapping sequence has a length of between 1260 and 1460 nucleotides, optionally 1360 nucleotides; or

vi) the 5' CDS comprises exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS comprises exons 27-51 of a CDS encoding a COL4A5 polypeptide, wherein the overlapping sequence has a length of between 1466 and 1666 nucleotides, optionally 1566 nucleotides.

13. A dual adeno-associated virus (AAV) vector system according to claim 12, for expressing a COL4A5 polypeptide wherein

(i) the 5' CDS comprises exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS comprises exons 38-51 of a CDS encoding a COL4A5 polypeptide.

(ii) the 5' CDS comprises exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS comprises exons 37-51 of a CDS encoding a COL4A5 polypeptide;

(iii) the 5' CDS comprises exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS comprises exons 36-51 of a CDS encoding a COL4A5 polypeptide;

(iv) the 5' CDS comprises exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS comprises exons 35-51 of a CDS encoding a COL4A5 polypeptide;

(v) the 5' CDS comprises exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS comprises of exons 34-51 of a CDS encoding a COL4A5 polypeptide;

(vi) the 5' CDS comprises exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS comprises exons 33-51 of a CDS encoding a COL4A5 polypeptide;

(vii) the 5' CDS comprises exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS comprises exons 32-51 of a CDS encoding a COL4A5 polypeptide;

(viii) the 5' CDS comprises exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS comprises exons 31-51 of a CDS encoding a COL4A5 polypeptide;

(ix) the 5' CDS comprises exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS comprises exons 30-51 of a CDS encoding a COL4A5 polypeptide;

(x) the 5' CDS comprises exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS comprises exons 29-51 of a CDS encoding a COL4A5 polypeptide;

(xi) the 5' CDS comprises exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS comprises exons 28-51 of a CDS encoding a COL4A5 polypeptide; or

(xii) the 5' CDS comprises exons 1-40 of a CDS encoding a COL4A5 polypeptide and the 3' CDS comprises exons 27-51 of a CDS encoding a COL4A5 polypeptide.

14. A dual adeno-associated virus (AAV) vector system according to claim 13, for expressing a COL4A5 polypeptide wherein

the 5' CDS consists of exons 1-40 of a CDS encoding a COL4A5 polypeptide (optionally according to SEQ ID NO: 287)

and the 3' CDS consists of:

i) exons 27-51 of a CDS encoding a COL4A5 polypeptide (optionally according to SEQ ID NO: 295); or

ii) exons 29-51 of a CDS encoding a COL4A5 polypeptide (optionally according to SEQ ID NO: 217); or

iii) exons 31-51 of a CDS encoding a COL4A5 polypeptide (optionally according to SEQ ID NO: 219); or

iv) nucleotides 2855 - 5058 of a CDS encoding a COL4A5 polypeptide (optionally according to SEQ ID NO: 309); or

v) exons 36-51 of a CDS encoding a COL4A5 polypeptide (optionally according to SEQ ID NO: 297); or

vi) exons 38-51 of a CDS encoding a COL4A5 polypeptide (optionally according to SEQ ID NO: 299).

15. A dual adeno-associated virus (AAV) vector system according to any one of claims 1-6, for expressing a COL4A5 polypeptide, wherein the 3' CDS sequence is longer than 5' CDS sequence.

16. A dual AAV vector system of claim 15, for expressing a COL4A5 polypeptide.
17. A dual AAV vector system of claim 16, wherein the 3' CDS comprises nucleotides 1724-5055 of a CDS encoding a COL4A5 polypeptide, and wherein the overlapping sequence has a length of from 450 nucleotides to 1600 nucleotides, optionally wherein the 3'CDS comprises nucleotides 1024-5058 of a CDS encoding a COL4A5 polypeptide.
18. A dual AAV vector system of claim 16, wherein the 3' CDS comprises nucleotides 1724-5055 of a CDS encoding a COL4A5 polypeptide, and wherein the 5' CDS comprises a minimum of nucleotides 1-2000 to a maximum of nucleotides 1-3500 of a CDS encoding a COL4A5 polypeptide.
19. A dual AAV vector system of claim 16, wherein the overlapping sequence has a length selected from: between 130 and 330 nucleotides, optionally 230 nucleotides; between 398 and 598, nucleotides, optionally 498 nucleotides; between 650 to 850 nucleotides, optionally 750 nucleotides; between 995 to 1195 nucleotides, optionally 1094 nucleotides; between 1260 and 1460 nucleotides, optionally 1359 nucleotides; and between 1466 and 1666 nucleotides, optionally 1566 nucleotides.
20. A dual AAV vector system of claim 16, wherein
- i) the 3' CDS comprises nucleotides 1724-5055 of a CDS encoding a COL4A5 polypeptide and the 5' CDS comprises nucleotides 1-2474 of a CDS encoding a COL4A5 polypeptide, wherein the overlapping sequence has a length of between 650 and 850 nucleotides, optionally 750 nucleotides;
  - ii) the 3' CDS comprises nucleotides 1724-5055 of a CDS encoding a COL4A5 polypeptide and the 5' CDS comprises nucleotides 1-2818 of a CDS encoding a COL4A5 polypeptide, wherein the overlapping sequence has a length of between 995 and 1195, nucleotides, optionally 1094 nucleotides; or
  - iii) the 3' CDS comprises nucleotides 1724-5055 of a CDS encoding a COL4A5 polypeptide and the 5' CDS comprises nucleotides 1-3083 of a CDS encoding a COL4A5 polypeptide, wherein the overlapping sequence has a length of between 1260 to 1460 nucleotides, optionally 1359 nucleotides.
21. A dual adeno-associated virus (AAV) vector system according to claim 20, for expressing a COL4A5 polypeptide wherein

the 3' CDS comprises nucleotides 1724-5055 of a CDS encoding a COL4A5 polypeptide (optionally according to SEQ ID NO: 314)

and the 5' CDS comprises of:

i) nucleotides 1-2474 of a CDS encoding a COL4A5 polypeptide (optionally according to SEQ ID NO: 306; or

ii) nucleotides 1-2818 of a CDS encoding a COL4A5 polypeptide (optionally according to SEQ ID NO: 305); or

iii) nucleotides 1-3083 of a CDS encoding a COL4A5 polypeptide (optionally according to SEQ ID NO: 304).

22. A dual adeno-associated virus (AAV) vector system according to claim 21, for expressing a COL4A5 polypeptide wherein

the 3' CDS consists of nucleotides 1724-5055 of a CDS encoding a COL4A5 polypeptide (optionally according to SEQ ID NO: 314)

and the 5' CDS consists of:

i) nucleotides 1-2474 of a CDS encoding a COL4A5 polypeptide (optionally according to SEQ ID NO: 306; or

ii) nucleotides 1-2818 of a CDS encoding a COL4A5 polypeptide (optionally according to SEQ ID NO: 305); or

iii) nucleotides 1-3083 of a CDS encoding a COL4A5 polypeptide (optionally according to SEQ ID NO: 304).

23. A dual adeno-associated virus (AAV) vector system according to any one of the preceding claims, for expressing a human COL4A5 polypeptide.

24. A dual adeno-associated virus (AAV) vector system according to any one of the preceding claims, for expressing a full-length COL4A5 polypeptide, preferably a full-length human COL4A5 polypeptide.

25. The dual AAV vector system according to any of claims 1 to 24, wherein the promoter upstream of the 5' CDS is selected from: a CMV promoter, a CBA promoter, a CAG promoter, a CB7 promoter, an EF1 $\alpha$  promoter, a CMV/EF1 $\alpha$  hybrid promoter, a NF-kB promoter, a pSE-7 promoter, a mPGK promoter, a mU1a promoter, a U6 promoter, a U7 promoter, a MNDU3

promoter, a HLP promoter, an AAT promoter, an ALB promoter, a ApoE/AAT promoter, a EalbAAT promoter, a LP1 promoter, a TBG promoter, a TTR promoter, a SYN1 promoter, a NSE promoter, a tMCK promoter, a CK8 promoter, a MHCK7 promoter, a SMN promoter, a DES promoter, a RK promoter, a hRHO promoter, a GRK1 promoter, a hCAR promoter, a hRPE65p promoter, a P546 promoter, a PR1.7 promoter, a hRS1 promoter, a VMD2 promoter, an  $\alpha$ -MHC promoter, a FRE1 promoter, a NPHS1 promoter, and a NPHS2 promoter.

26. The dual AAV vector system according to any one of claims 1 to 25, wherein the promoter upstream of the 5' CDS is a kidney-specific promoter, optionally wherein the promoter upstream of the 5' CDS is a podocyte-specific promoter or a tubular cell specific promoter.

27. The dual AAV vector system according to claim 26, wherein the promoter upstream of the 5' CDS is a NPHS1 promoter or a NPHS2 promoter, optionally wherein the promoter upstream of the 5' CDS is a minimal NPHS1 promoter, a minimal NPHS2 promoter or the NPHS1 265bp promoter.

28. The dual AAV vector system according to any one of claims 1 to 27, wherein the second AAV vector comprises one or more regulatory sequences downstream from the 3' CDS.

29. The dual AAV vector system according to claim 28, wherein the second AAV vector comprises a Woodchuck hepatitis post-transcriptional regulatory element (WPRE), downstream from the 3' CDS.

30. The dual AAV vector system according to claim 29, wherein the WPRE comprises or consists of a nucleotide sequence having 80% or more, 85% or more, 90% or more, 91% or more, 92% or more, 93% or more, 94% or more, 95% or more, 96% or more, 97% or more, 98% or more, 99% or more, or 100% identity to SEQ ID NO: 263.

31. The dual AAV vector system according to any one of claims 1 to 30, wherein the second AAV vector comprises a polyadenylation sequence downstream from the 3' CDS.

32. The dual AAV vector system according to claim 31, wherein the polyadenylation sequence downstream from the 3' CDS is a bovine growth hormone polyadenylation sequence (bGH), a soluble neuropilin-1 polyadenylation sequence, an early SV40 polyadenylation sequence (SV40pA), or a chicken beta-globin polyadenylation sequence; and/or wherein the polyadenylation sequence downstream from the 3' CDS comprises or consists of a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 236 to 239.

33. The dual AAV vector system according to any one of claims 1 to 32, wherein the first AAV vector comprises in the 5'-3' direction:

a 5'-inverted terminal repeat (5'-ITR);

a kidney-specific promoter, optionally a podocyte-specific promoter;

the 5' CDS as defined in any of claims 1-27, said 5' CDS being operably linked to and under control of the kidney-specific promoter, optionally the podocyte-specific promoter; and

a 3'-inverted terminal repeat (3'-ITR)

and the second AAV vector comprises in the 5' to 3' direction:

a 5'-ITR;

a 3' CDS as defined in any of claims 1-32;

a WPRE;

a poly-adenylation sequence; and

a 3'-ITR, optionally wherein the first AAV vector comprises one or more regulatory elements.

34. The dual AAV vector system according to any of claims 1 to 33, wherein the first AAV vector and/or the second AAV vector is in the form of an AAV vector particle encapsidated by LK03, AAV3B, AAV9, ShH10, AAV-DJ, AAV2, AAV6.2, AAV5 or KP1 capsid proteins.

35. The dual AAV vector system according to any of claims 1 to 34, wherein the first AAV vector and/or the second AAV vector is in the form of an AAV vector particle encapsidated by the LK03 capsid protein, or the KP1 capsid protein.

36. An isolated cell comprising the dual AAV vector system according to any preceding claim.

37. A pharmaceutical composition comprising the dual AAV vector system according to any of claims 1 to 35 or a cell according to claim 36.

38. A dual AAV vector system according to any of claims 1 to 35, a cell according to claim 36 or a pharmaceutical composition according to claim 37, for use as a medicament.

39. Use of a dual AAV vector system according to any of claims 1 to 35, a cell according to claim 36 or a pharmaceutical composition according to claim 37, for the manufacture of a medicament.

40. A product comprising:

(a) a first AAV vector comprising a promoter and a 5' coding sequence (CDS), wherein the 5' CDS encodes an N-terminal part of a COL4A3, COL4A4 or COL4A5 polypeptide; and

(b) a second AAV vector comprising a 3' CDS, wherein the 3' CDS encodes a C-terminal part of the COL4A3, COL4A4 or COL4A5 polypeptide,

as a combined preparation for simultaneous, separate or sequential use in therapy, wherein a CDS encoding the COL4A3, COL4A4 or COL4A5 polypeptide is formed and the COL4A3, COL4A4 or COL4A5 polypeptide is expressed upon administration of the first AAV vector and the second AAV vector to a subject.

41. The product for use according to claim 40, wherein the first AAV vector and the second AAV vector are defined according to any of claims 1 to 35.

42. A dual AAV vector system according to any of claims 1 to 35, a cell according to claim 36, a pharmaceutical composition according to claim 37, or the product of claim 40 or 41 for use in preventing and/or treating Alport Syndrome or for use in preventing and/or treating a condition in a subject who has a pathogenic variant in the COL4A3, COL4A4 or COL4A5 gene.

43. A kit comprising: (a) the first AAV vector as defined in any of claims 2 to 35; and (b) the second AAV vector as defined in any of claims 2 to 35.

44. The kit according to claim 43, for use in preventing and/or treating Alport Syndrome or for use in preventing and/or treating a condition in a subject who has a pathogenic variant in the COL4A3, COL4A4 or COL4A5 gene, optionally wherein the first AAV vector and the second AAV vector are administered simultaneously, separately or sequentially.

45. A method of transducing a glomerular cell *ex vivo*, wherein the cell is transduced with the dual vector system defined according to any one of claims 1 to 35, optionally wherein the glomerular cell is a podocyte.