An optimized acceptor splice site module for biological and biotechnological applications

TECHNICAL FIELD

[1] The present invention relates to a novel acceptor splice region, as well as uses and applications thereof.

BACKGROUND

[2] Most eukaryotic genes contain non-coding introns, which must be removed from the precursor messenger RNA (pre-mRNA) to generate a translatable mature messenger RNA (mRNA) in a process called "splicing". Spicing is mediated via a large ribonucleoprotein complex, the spliceosome, which consists of five conserved small nuclear ribonucleoproteins (snRNPs), i.e., U1, U2, U4, U5 and U6 snRNP, and more than 300 proteins. The spliceosome assembles and disassembles for each intron in a highly dynamic process. For this purpose, specific splicing sequences need to be recognized and hence exon and intron boundaries are defined by distinct sequence motifs, which serve as binding positions for the ribonucleoproteins involved in the regulation of mRNA splicing.

[3] The most relevant splice motifs are the canonical donor and acceptor splice sites, which define the exon-intron boundaries. The 5' donor splice site (DSS) is located at the 3'-end of the exon and at the 5'-end of the downstream intron and the 3' acceptor splice site (ASS) is located at the 3'-end of the intron and the 5'-end of the downstream exon. A functional acceptor splice site (ASS) requires three distinct elements, which are usually located with a range of approximately 50 bp, a branch point (BP), a poly-pyrimidine tract (PPT) and the canonical ASS sequence.

[4] The so-called GT-AG splicing is by far the most common type of mRNA splicing in mammals and defines the first two and the last two nucleotides of the intron at the DNA level. Thus, a part of the canonical DSS sequence represents the first two bases of the 5' intron sequence, which is a guanine followed by a thymine (GT) in the DNA sequence (corresponding to GU in the RNA sequence), and the last two bases of the intron (representing the most conserved part of the consensus ASS) are an adenine followed by a guanine (AG). The GT-AG nucleotides are essential for an effective splice reaction and their disruption or substitution leads to the loss of a functional splice site.

[5] Each splicing cycle consists of two transesterification reactions. For the first reaction, known as branching, the branch point nucleoside (typically adenosine) attacks the phosphate bond at the 5' exon-intron junction. This results in the formation of an intronic

lariat-3' exon intermediate (comprising the exon downstream of the splice site and the intron) and a free 5' exon end (comprising the exon upstream of the splice site). In the second reaction, called exon ligation, the free end of the 5' exon attacks the phosphate at the intron-3' exon junction causing a ligation of the two exons and a release of the intron lariat structure. The splicing reaction is regulated by auxiliary *cis*-acting splicing regulatory elements of the pre-mRNA composed of up to ten nucleotides. According to their function and position, they can be categorized into exonic splice enhancers (ESEs), exonic splice enhancers (ESSs), intronic splice enhancers (ISEs) and intronic splice silencers (ISSs). These elements are able to recruit *trans*-acting proteins for promotion or prevention of exon inclusion and are therefore proposed to pay an important role in alternative splicing.

[6] The splicing efficiency for given exons is expected to depend on the strength of the donor and acceptor splice sites. This DSS or ASS strength in turn depends on the intronic or exonic sequence elements upstream or downstream of the GT or AG. The strength of a functional DSS can be reliably predicted using standard *in silico* prediction software (e.g. NNSplice: http://www.fruitfly.org/seq_tools/splice.html or Human Splice Finder: http://www.umd.be/HSF3/). By contrast, due to its complexity, *in silico* prediction of the ASS strength leads to less reliable results (e.g. Koller et al., (2011) "A novel screening system improves genetic correction by internal exon replacement." Nucleic acids research. 39:e108; Lorain et al., (2013) "Dystrophin rescue by trans-splicing: a strategy for DMD genotypes not eligible for exon skipping approaches." Nucleic acids research. 41:8391-8402). Therefore, the factual ASS strength needs to be validated experimentally. Many biological, biotechnological and therapeutical applications rely on the efficiency of the classical GT-AG mRNA splicing, and, hence, on the usage of strong splice sites.

[7] Apart from the regular *cis*-splicing events to generate a translatable mature mRNA from a pre-mRNA by removing the non-coding introns, splicing can also occur in *trans*, thereby combining two separate pre-mRNA molecules to create a non-co-linear chimeric RNA (Lei et al., (2016) "Evolutionary insights into RNA trans-splicing in vertebrates", Genome Biol. Evol. 8(3):562-577). This process has first been discovered in trypanosomes. Since then, *trans*-splicing events have also been identified in many more species including mice (Hirano M and Noda T., (2004) "Genomic organization of the mouse Msh4 gene producing bicistronic chimeric and antisense mRNA", Gene 342: 165-177) as well as in human cells (Chuang et al., (2018) "Integrative transcriptome sequencing reveals extensive alternative trans-splicing and cis-back splicing in human cells". Nucleic Acids Research, 46(7): 3671-3691), although *trans*-splicing seems to occur less frequently in higher vertebrates only.

[8] Attempts have been made to exploit *trans*-splicing for gene therapy. Viral vectors are an attractive vehicle for gene therapy, however, they often have a restricted loading capacity

and hence there is a limitation as to which genes can be replaced. For example adeno-associated viral vectors have a packaging capacity of up to about 5.0 kb, which allows a transgene of about 4 kb or less. *Trans*-splicing, i.e., joining together two physically separate pre-mRNAs to form a mature mRNA may be one way to overcome these limitations. For example spliceosome-mediated RNA *trans*-splicing (SMaRT) may be used as a tool for gene therapy, comprising an exogenous pre-mRNA *trans*-splicing molecule that is introduced into a target cell to replace only a part of a mutated endogenous pre-mRNA. This may allow for the delivery of shorter coding sequences in the viral vectors.

[9] An alternative way to address the size limitation of viral vectors, particularly for adeno-associated virus (AAV) based vectors, is the use of recombinant AAV (rAAV) dual vector technology. AAV is a single-stranded DNA virus. For rAAV dual vector technologies the coding sequence of the transgene (gene of interest) is split into at least two parts and packaged into two or more separate rAAV vectors. After co-transduction of the target cell by the split genome vectors, the full-length coding sequence is reconstituted. Since a number of cells, such as the photoreceptor cells of the retina show a high co-transduction efficiency of > 90%, the efficient delivery of both rAAVs into the same target cell does not seem to be limiting. An efficient reconstitution of the two transgene halves, however, remains challenging. For several years post the development of rAAV dual vector systems, reconstitution has only been addressed at the DNA level and several strategies have been explored to improve this approach (McClements and MacLaren, (2017) "Adeno-associated virus (AAV) dual vector strategies for gene therapy encoding large transgenes, Yale Journal of Biology and Medicine 90: 611-623).

[10] Often these rAAV dual vector systems are misleadingly referred to as "*trans*-splicing dual vectors", although mRNA splicing in these approaches never actually takes place in *trans*. Rather, reconstitution in previously described rAAV dual vector systems, relies on the concatemerization of the ITR structures and/or homologous recombination of overlapping sequences, resulting in a single pre-mRNA that may be spliced in *cis* to remove the concatemerized ITR elements and/or artificial recombinogenic elements. Such rAAV dual vector systems are, e.g., disclosed by Trapani et al., ("Effective delivery of large genes to the retina by dual AAV vectors", (2014) EMBO Molecular Medicine, 6(2): 194-211). Reconstitution of rAAV dual vector systems at the DNA level can be recognized by the presence of a promoter driving the expression of the 5' part of the coding sequence in the first AAV vector, and by the absence of a promoter driving the expression of the 3' part of the coding sequence in the second AAV vector. The reported *in vivo* efficiency for reconstitution of such rAAV dual vector systems is relatively low, namely below 10% (Carvalho et al., "Evaluating efficiencies of dual AAV approaches for retinal targeting", (2017) Frontiers in Neuroscience, 11(503): 1-8). In contrast, for pre-mRNA splicing to take place in a *trans*

configuration, both vectors of the rAAV dual vector system require a promoter to generate two separate pre-mRNA molecules.

[11] For an optimal performance of mRNA-splicing and particularly mRNA *trans*-splicing, there is an unmet need for identification of optimized and experimentally validated strong ASS. In particular, there is a need for strong ASS regions, which can inter alia be used in the development of further AAV vectors, such as next-generation rAAV dual vector systems or AAVs delivering highly specific and efficient pre-mRNA *trans*-splicing molecules, which may be used in gene therapy.

SUMMARY OF THE INVENTION

[12] The present invention complies with this need as described herein and as described in the Examples, Figures and Claims.

[13] Provided herein is a pre-mRNA *trans*-splicing molecule comprising, (i) an acceptor splice region, comprising (ia) a pyrimidine tract, wherein the pyrimidine tract comprises (iaa) 5 to 25 nucleotides; (iab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U); (ib) an acceptor splice site, (iba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and (ibb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G; (ii) a nucleotide sequence of interest or a portion thereof; wherein the acceptor splice region is localized 3' or 5' to the nucleotide sequence of interest or the portion thereof; (iii) a binding domain targeting pre-mRNA, which is localized 3' or 5' to the nucleic acid sequence of interest or the portion thereof; and (iv) optionally a spacer sequence, wherein the spacer sequence is localized between the binding domain and the acceptor splice region. In one embodiment the pre-mRNA *trans*-splicing molecule of claim 1 comprises (ii) the nucleotide sequence of interest or the portion thereof, wherein the acceptor splice region is localized 5' to the nucleotide sequence of interest or the portion thereof; (iii) a donor splice site, wherein the donor splice site is localized 3' to the nucleotide sequence of interest or the portion thereof; (iv) a first binding domain targeting pre-mRNA located 5' to the nucleotide sequence of interest or the portion thereof; (v) a second binding domain targeting pre-mRNA located 3' to the nucleotide sequence of interest or the portion thereof; (vi) optionally a first spacer sequence, wherein the first spacer is localized between the first binding domain and the acceptor splice region; and (vii) optionally a second spacer sequence, wherein the second spacer is localized between the second binding domain and the donor splice site.

[14] Preferably the acceptor splice region is localized 5' to the nucleotide sequence of interest or the portion thereof and the binding domain is localized 5' to acceptor splice region. The pre-mRNA *trans*-splicing molecule may further comprise a termination

sequence, preferably a polyA sequence. Preferably the pre-mRNA *trans*-splicing molecule according to the invention comprises a pyrimidine tract, wherein the 5 to 25 nucleotides of the pyrimidine tract comprise a sequence encoded by the sequence of TTTTTT or TCTTTT. Additionally or alternatively the sequence between the last pyrimidine of the pyrimidine tract and the acceptor splice site has less than 10, preferably less than 5, more preferably less than 3 bases. Additionally or alternatively the acceptor splice site has a sequence of CAGG.

[15] The acceptor splice region may comprise (a) 7 nucleotides 5' to the pyrimidine tract having at least 4 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotide is C; (b) 7 nucleotides 5' to the pyrimidine tract having at least 5 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotide is C; (c) 7 nucleotides 5' to the pyrimidine tract having at least 4 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAA; or (d) 7 nucleotides 5' to the pyrimidine tract having at least 5 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAAC. In one embodiment the acceptor splice region is encoded by a nucleotide sequence which has the sequence of SEQ ID NO: 3 or 4 or which is at least 85 %, 90 %, 95 %, 97 %, 98 %, 99 % or 100 % identical to the sequence of SEQ ID NO: 3 or 4. Also provided is a DNA molecule, comprising a promoter and a sequence encoding the pre-mRNA *trans*-splicing molecule according to the invention, wherein the DNA molecule is preferably a vector or a plasmid.

[16] In another aspect the invention relates to a method for producing a nucleic acid sequence, the method comprising (A) providing a first nucleic acid sequence comprising one or more donor splice site sequences; (B) providing a second nucleic acid sequence, wherein the second nucleic acid sequence comprises (i) an acceptor splice region comprising (ia) a pyrimidine tract, wherein the pyrimidine tract comprises (iaa) 5 to 25 nucleotides; (iab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U); (ib) an acceptor splice site, (iba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and (ibb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G; (C) cleaving the first nucleic acid sequence at the one or more donor splice site sequence(s) and cleaving the second nucleic acid sequence in the acceptor splice site, (D) ligating the first cleaved nucleic acid sequence to the second cleaved nucleic acid sequence, thereby obtaining a nucleic acid sequence. The first nucleic acid sequence further comprises a nucleotide sequence of interest or a portion thereof, wherein at least a portion of the nucleotide sequence of interest is located 5' to the donor splice site and wherein the second nucleic acid sequence further comprises a nucleotide sequence of interest or a portion thereof, wherein at least a portion of the nucleotide sequence of interest is located 3' to the acceptor splice region. Preferably the first and the second nucleic acid sequence are introduced into a host cell, preferably wherein the first and the second nucleic acid sequence

are a recombinant nucleic acid sequence.

[17] In another aspect of an embodiment of the method of the invention, the method comprises step (A) introducing into a host cell a first nucleic acid sequence comprising a pre-mRNA *trans*-splicing molecule sequence or a nucleic acid sequence encoding said pre-mRNA *trans*-splicing molecule, wherein the first pre-mRNA *trans*-splicing molecule comprises from 5' to 3', (a) a 5' portion of a nucleotide acid sequence of interest; (b) a donor splice site; (c) optionally a spacer sequence; (d) a first binding domain; and (e) optionally a termination sequence, preferably a polyA sequence; step (B) introducing into the host cell a second nucleic acid sequence comprising a pre-mRNA *trans*-splicing molecule sequence or a nucleic acid sequence encoding said pre-mRNA *trans*-splicing molecule, wherein the second pre-mRNA *trans*-splicing molecule comprises from 5' to 3'; (i) a second binding domain, which is complementary to the first target domain of the first nucleic acid sequence; (ii) an acceptor splice region sequence comprising; (iia) a pyrimidine tract, wherein the pyrimidine tract comprises; (iiaa) 5 to 25 nucleotides; (iiab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U); (iib) an acceptor splice site, (iiba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and (iibb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G; (iii) a 3' portion of the nucleotide sequence of interest, and (iv) a termination sequence, preferably a polyA sequence; step (C) cleaving the first nucleic acid sequence at the donor splice site sequence and cleaving the second nucleic acid sequence in the acceptor splice site; and step (D) ligating the first cleaved nucleic acid sequence comprising the 5' portion of the nucleotide sequence of interest to the second cleaved nucleic acid sequence comprising the 3' portion of the nucleotide sequence of interest, thereby obtaining the nucleic acid sequence of interest.

[18] In yet another aspect, the invention relates to an adeno-associated virus (AAV) vector comprising at least two inverted terminal repeats comprising a nucleic acid sequence between these two inverted terminal repeats, wherein said nucleic acid sequence comprises from 5' to 3': (i) a promoter; (ii) a binding domain; (iii) optionally a spacer sequence; (iv) an acceptor splice region sequence comprising (a) a pyrimidine tract, wherein the pyrimidine tract comprises (aa) 5 to 25 nucleotides; (ab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U); (b) a acceptor splice site, wherein the acceptor splice site is located 3' to the pyrimidine tract; and wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G; (v) a nucleotide sequence of interest or a portion thereof; and (vi) optionally a polyA sequence. The AAV vector may also be part of an AAV vector system comprising (I) a first AAV vector comprising at least two inverted terminal

repeats comprising a nucleic acid sequence between these two inverted terminal repeats, wherein said nucleic acid sequence between these two inverted terminal repeats comprises from 5' to 3': (a) a promoter, (b) a nucleotide sequence encoding an N-terminal portion of a polypeptide of interest; (c) a donor splice site; (d) optionally a spacer sequence; (e) a first binding domain; and (f) optionally a termination sequence, preferably a polyA sequence; (II) a second AAV vector comprising a nucleic acid sequence comprising at least two inverted terminal repeats comprising a nucleic acid sequence between these two inverted terminal repeats, wherein said nucleic acid sequence between these two inverted terminal repeats comprises from 5' to 3' (i) a promoter; (ii) a second binding domain, which is complementary to the first binding domain of the first AAV vector; (ii) optionally a spacer sequence (iii) an acceptor splice region sequence comprising (a) a pyrimidine tract, wherein the pyrimidine tract comprises (aa) 5 to 25 nucleotides; (ab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U); (b) an acceptor splice site, wherein the acceptor splice site is located 3' to the pyrimidine tract; and wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G; (iv) a nucleotide sequence encoding a C-terminal portion of the polypeptide of interest; (iva) wherein the C-terminal portion of the polypeptide of interest and the N-terminal portion of the polypeptide of interest reconstitute the polypeptide of interest; and (v) a termination sequence, preferably a polyA sequence. In one embodiment the polypeptide is a full-length polypeptide and the first AAV vector comprises a N-terminal portion of a full-length polypeptide of interest and the second AAV vector comprises a C-terminal portion of the full-length polypeptide of interest.

[19] Also provided is a nucleic acid sequence comprising an acceptor splice region sequence comprising (ia) a pyrimidine tract, wherein the pyrimidine tract comprises (aa) 5 to 25 nucleotides; (ab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U); (ib) a acceptor splice site, ba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and bb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G; and (ii) a nucleotide sequence of interest, wherein the nucleotide sequence of interest (iia) is located 3' or 5' to the acceptor splice region.

[20] In certain embodiments according to the invention the acceptor splice region described herein comprises a pyrimidine tract, wherein the 5 to 25 nucleotides of the pyrimidine tract comprise a sequence encoded by the sequence of TTTTTT or TCTTTT. The sequence between the last pyrimidine of the pyrimidine tract and the acceptor splice site preferably has less than 10, more preferably less than 5, even more preferably less than 3 bases. Moreover, the acceptor splice site preferably has the sequence CAGG. The acceptor splice region may further comprise 7 nucleotides 5' to the pyrimidine tract having at least 4

nucleotides of the sequence CAACGAG, wherein the first 5' nucleotide is C. In a preferred embodiment the splice acceptor region comprises or consists of the sequence of SEQ ID NO: 3 or 4 or is encoded by a nucleic acid sequence comprising or consisting of the sequence of SEQ ID NO: 3 or 4.

BRIEF DESCRIPTION OF THE FIGURES

[21] **Fig. 1** mRNA splicing of *RHO* minigenes in HEK293 cells and in transduced photoreceptors.

A, Schematic true to scale view of the *RHO* gene. The boxes represent the exons and the thick lines in between represent the introns, the respective start and stop codon is indicated. The asterisk symbolizes the c.620T>G mutation in exon 3. **B** and **C**, Rhodopsin minigenes encompassing the coding parts of the exons and the flanking introns were driven either by a CMV promoter for expression in HEK293 cells (B), or by a human rhodopsin (hRHO) promoter for expression in photoreceptors (C). To enable packaging of the *RHO* minigene into AAV vectors, intron 1 was shortened as indicated. For protein visualization and detection, *RHO* was C-terminally fused either to citrine (in B), or to a myc-tag (in C). The primers shown as arrows in B and C were used for specific detection of the wild type (WT) or mutant splicing products derived from HEK293 cells (D) or from transduced photoreceptors (E). **D**, RT-PCR analysis from HEK293 cells transiently transfected with WT and mutant *RHO* minigenes. **E**, Left, schematic representation of the subretinal *RHO* minigene delivery to murine retinas at postnatal day 14 (P14). Right, RT-PCR from injected murine retinas containing the respective WT or mutant *RHO* minigenes. RT-PCR was performed four weeks post injection. All experiments were repeated once.

[22] **Fig. 2** Identification of the most efficient ASS_620 sequence.

A, Composition of the single elements of a functional ASS. The branch point (BP) is underlined representing the consensus sequence for human branch points, with the branch point nucleoside adenosine highlighted in bold. The polypyrimidine tract (Poly-C/T) is represented as an empty box and the ASS spanning the intron-exon boundary is represented by a grey filled box. **B**, Exon-intron organization of the human *RHO* gene and zoomed view in the DNA sequence of exon 3 of the human *RHO* gene for wild type (WT, upper panel) and the c.620T>G mutant (lower panel). Intronic sequences are shown in lower case, exonic sequences are shown in upper case (bold) letters. The single ASS elements are highlighted according to the scheme shown in A, the BP sequence is underlined (thick line), the Poly-C/T sequence is marked by an empty box, the ASS sequence(s) is/are marked by a filled grey box. The ATG sequence in the WT converted to the AGG sequence in the c.620T>G mutant is marked by a thin underline. Note that the c.620T>G mutation generates a novel canonical ASS sequence. The other two elements required for a

functional ASS (BP and Poly-C/T) are already present in WT *RHO* exon 3. **C**, Structure of the human ribosomal protein 27 (*RPS27)* minigene used for the splicing experiments. Primers used for the RT-PCR shown in (D) are symbolized by arrows. The individual sequences containing potential elements of ASS_620 termed RHO_E3a-g as shown below were inserted into exon 3 of the *RPS27* gene as indicated. *RHO*_E3a is the WT *RHO* sequence lacking a functional ASS. Potential BP, Poly-C/T or ASS sequences are indicated as above. **D**, RT-PCR from HEK293 cells transfected with the single chimeric *RPS27* minigenes as indicated. CS, correctly spliced *RPS27* minigene (i.e. using the native exon 3 ASS). AS, aberrantly spliced *RPS27* minigene (i.e. using ASS_620).

[23] **Fig. 3** ASS_620 is a strong acceptor splice site.

A, *In silico* acceptor splice site (ASS) strength prediction using two commonly available prediction tools, NNSplice (0.9 version; January 1997) (http://www.fruitfly.org/seq_tools/splice.html) or human splicing finder (version 3.1) (HSF, http://www.umd.be/HSF3/). **B**, Composition and sequence of the single elements of the 26 bp sequence referred to as vgASS_620 comprising the ASS, the PolyC/T and seven additional nucleotides 5' of the Poly C/T. **C**, Schematic representation of the different minigenes used for determining the strength of vgASS_620. The position of the vgASS_620 insertion within the single minigenes is symbolized by an asterisk and a dashed line. The binding position of the primers used for the RT-PCR shown in D are indicated by arrows. **D**, RT-PCR from HEK293 cells transfected with the respective minigenes containing only the native (nat) acceptor splice site or both, the native acceptor splice site and vgASS_620 (620). All bands were checked by sequencing. **E**, Length of the RT-PCR products originating from the utilization of the native acceptor splice site or vgASS_620.

[24] **Fig. 4** Cerulean reconstitution assay testing different binding domains.

A, Schematic overview of *RHO* intron 2 sequences used as binding domains for the reconstitution of cerulean via mRNA *trans*-splicing. The entire *RHO* intron 2 sequence (a) as well as different 5' (b, d, f, h) and 3' parts (c, e, g, i), or a small 5' part fused to a small 3'part (h+i) of intron 2 were tested providing the results shown in B-G. **B** and **C**, Control qRT-PCR (n=3) from transfected HEK293 cells to compare the mRNA levels of the single constructs comprising the different binding domains shown in A. Shown are the delta CT (ΔCT) values in relation to the housekeeper aminolevulinic acid synthase (ALAS). Binding positions of primers (p1+p2 in B and p3+p4 in C) are displayed in D. Statistical analysis (n=3 for each transfection) was done by one-way ANOVA followed by the Tukey's test for multiple comparisons. **D**, Principle of the cerulean reconstitution assay using the example of the h+i binding domain, a donor splice site (DSS) and the acceptor splice site ASS_620. Cerulean was split at the nucleotide position 154 downstream of the start codon of the full-length cerulean sequence (c1) as indicated by the dashed line. In the control construct (c2) an

artificial intron was introduced into the cerulean coding sequence at the indicated position 154 for creating two artificial cerulean exons. The antibody (α-cerNT) used for Western blotting binds to the N-terminal cerulean half. The control construct (c2) served as reference for quantification in confocal imaging and Western blotting experiments. The different sequences indicated in (A) were tested as binding domains using the dual vector approach. The scheme of the dual vector approach for cerulean reconstitution is exemplarily shown for the h+i binding domain in combination with vgASS_620 (c3) or the native *RHO* exon 3 (*RHO*_E3) acceptor splice site (c4). **E**, Cerulean reconstitution efficiencies (CRE) of the different binding domains were calculated from Western blot band intensities resulting from three independent transfections. For quantification, the band intensities were first normalized to internal tubulin controls. The normalized values are given as percentages of the reference construct intensity (c2). **F**, Representative confocal imaging of living HEK293 cells transfected with c1, c2, c3 or c4 constructs driven by a CMV promoter. For imaging, cerulean-specific laser and filter settings were used. Scale bar, 50 µm. **G**, Representative Western blots from protein lysates of transfected HEK293 cells (Ø, non-transfected cells; IB, immunoblotting; α-Tub, beta-tubulin-specific antibody).

[25] **Fig. 5**: Sequence of the first dual AAV comprising the 5' coding sequence of ABCA4 (SEQ ID NO: 19): ITR sequences at the 5' end and the 3' end are highlighted in grey without underlining (NNN). The spacer sequences are depicted in small letters. The CMV promoter is highlighted in grey and is underlined using a solid line (NNN). The 5' coding sequence of the ABCA4 protein is underlined with a dotted line (NNN). The DSS is depicted in italic letters and is underlined (*Nnn*, with the capital letters representing the coding sequence and the small letter representing the non-coding sequence). The binding domain is highlighted in grey and is underlined with an interrupted line (NNN).

[26] **Fig. 6**: Sequence of the second dual AAV comprising the 3' coding sequence of ABCA4 (SEQ ID NO: 20): ITR sequences at the 5' end and the 3' end are highlighted in grey without underlining (NNN). The spacer sequences are depicted in small letters. The CMV promoter is highlighted in grey and is underlined using a solid line (NNN). The binding domain is highlighted in grey and is underlined with interrupted lines (NNN). The acceptor splice site is depicted in italic letters and is underlined (*NNN*). The 3' coding sequence of the ABCA4 protein is underlined using a dotted line (NNN). The polyA sequence is highlighted in dark grey and underlined with waves (NNN).

[27] **Fig. 7**: Sequence of the first dual AAV comprising the 5' coding sequence of ABCA4 and the ABCA4 promoter (SEQ ID NO: 21): ITR sequences at the 5' end and the 3' end are highlighted in grey without underlining (NNN). The spacer sequences are depicted in small letters. The ABCA4 promoter is highlighted in grey and is underlined using a solid line (NNN). The 5' coding sequence of the ABCA4 protein is underlined using a dotted line

(NNN). The DSS is depicted in italic letters and is underlined (*Nnn*, with the capital letters representing the coding sequence and the small letters representing the non-coding sequence). The binding domain is underlined using an interrupted line (NNN).

[28] **Fig. 8**: Sequence of the second dual AAV comprising the 3' coding sequence of ABCA4 and the ABCA4 promoter (SEQ ID NO: 22): ITR sequences at the 5' end and the 3' end are highlighted in grey without underlining (NNN). The spacer sequences are depicted in small letters. The ABCA4 promoter is highlighted in grey and underlined using a solid line (NNN). The binding domain is underlined using an interrupted line (NNN). The ASS is depicted in italic letters and is underlined (*NNN*). The 3' coding sequence of the ABCA4 protein is highlighted in grey and is underlined using a dotted line (NNN). The polyA sequence is highlighted in dark grey and underlined with waves (NNN).

[29] **Fig. 9:** Impact of acceptor splice site *vs.* binding domain on cerulean reconstitution efficiency. **A**, Binding domains (BDs) and acceptor splice sites (ASSs) tested in this experiment. All BD sequences originate from the human *RHO* gene. BDs and ASSs expected to yield high efficiencies are marked in bold and italic letters. **B**, Confocal live images of HEK293 cells transfected with constructs containing different combinations of the three BDs and the three ASSs shown in A. The strength of the respective BD and ASS is indicated. Scale bar, 50 µm. **C**, Upper panel, RT-PCR of the different BD and ASS combinations. Lower panel, Western blot of the different BD and ASS combinations. *GAPDH* and beta tubulin served as loading controls. **D**, Quantification of the reconstitution efficiency by the radiometric analysis of the cerulean protein bands in relation to the *cis*-ctrl (n = 3 – 8). All protein bands were normalized to beta tubulin prior to quantification.

[30] **Fig. 10:** mRNA *trans*-splicing rAAV dual vector approach *in vivo*. **A**, 5' vector- and 3' vector-constructs used for *in vivo* expression. As an expression control, a citrine and an mCherry sequence was fused 5' of the cerulean 5' coding sequence (CDS) and 3' of the cerulean 3' CDS, respectively. **B**, Representative confocal images of retinal sections two weeks post injection expressing constructs containing BD_h+i. Fluorophore expression is present in the retinal pigment epithelium (RPE). ONL, outer nuclear layer. Scale bar, 20 µm. **C**, Confocal images of RPE cells before (upper panel) and after (lower panel) selective photobleaching of citrine and mCherry fluorophores using a 514 nm laser. Scale bar, 2 µm.

[31] **Fig. 11:** Identification of suitable BDs originating from the *lacZ* gene. **A**, Binding domains (BDs) taken from the bacterial *lacZ* gene and modified to possess no detectable homology to the human genome. **B**, Confocal live images of HEK293 cells transiently co-transfected with constructs containing the BDs shown in A. Scale bar, 50 µm. **C**, Western blot obtained from transfected HEK293 cell lysates. **D**, Quantification of cerulean reconstitution efficiency based on the ratiometric analysis of the western blot band intensity (n = 3 – 8). BD_g efficiency served as a measure for the highest reconstitution obtained so

far (cf. Figure 9).

[32] **Fig. 12:** Deletion of regulatory elements and its impact on mRNA splicing efficiency. **A**, Confocal live images of HEK293 cells transiently co-transfected with cerulean constructs as indicated containing BD_k. 5'^{polyAdel}, 5' vector without polyA signal. 3'^{promdel}, 3' vector without promoter sequence. Scale bar, 50 µm. **B**, Western blot obtained from transfected HEK293 cell shown in A. Beta tubulin served as loading control.

[33] **Fig. 13:** Reconstitution of *Sp*Cas9-VPR. **A**, Plasmids used for mRNA *trans*-splicing of *Sp*Cas9-VPR in HEK293 cells (5' and 3' vector). As positive control served a plasmid containing full-length (FL) *Sp*Cas9-VPR CDS (FL vector). A junction-spanning primer pair was used for RT-PCR (black arrows). **B**, RT-PCR from HEK293 cells co-transfected with the 5' and 3' vector containing BD_k (n = 3) or the FL vector (n = 3). **C**, Representative sequencing result of the reconstituted *Sp*Cas9-VPR product. **D**, Western blot of protein lysates from the respective transfections. Ø, untransfected cells. Beta tubulin served as a loading control.

[34] **Fig. 14:** Reconstitution of ABCA4. **A**, Plasmids used for mRNA *trans*-splicing of ABCA4 in HEK293 cells (5' and 3' vector). Additionally, six short introns of six different genes were incorporated within the CDS of ABCA4, i.e. three into the 5' CDS (5' vector with introns, 5'wi) and three into the 3' CDS (3' vector with introns, 3'wi), respectively. A junction-spanning primer pair was used for RT-PCR (black arrows). myc, myc-tag. **B**, RT-PCRs run with different numbers of cycles from HEK293 cells co-transfected with the respective constructs as indicated. GAPDH served as loading control. Ø, untransfected cells. **C**, Representative sequencing result of reconstituted ABCA4.

[35] **Fig. 15:** rAAV dual vector mRNA trans-splicing of ABCA4 *in vivo*. **A**, 5' vector- and 3' vector-constructs containing BD_k used for *in vivo* expression. hRho, human rhodopsin promoter. **B**, RT-PCRs run with different numbers of cycles from retinal lysates of C54Bl/6J wild type mice co-transduced with the respective constructs as indicated. As negative control served non-injected wild type (WT) mice. NN/GL, capsid variants of AAV2. -RT, negative control without reverse transcriptase. GAPDH was used as loading control. **C**, Representative sequencing result of reconstituted ABCA4. **D**, Preliminary results of qRT-PCR performed from the same retinas shown in B. ABCA4 expression obtained upon co-transduction was normalized to non-injected WT retinas. **E**, Western blot from retinal lysates of the same mice shown in B-D. Human ABCA4 protein was detected by an anti-myc antibody and is indicated via an arrow head. Beta tubulin served as loading control.

[36] **Fig. 16:** Sequence of the minigene encoding the 5' coding sequence of cerulean and the binding domain BD_g (SEQ ID NO: 35). The CMV promoter is highlighted in grey. The 5' coding sequence of the cerulean protein (bold letters) is highlighted in grey and underlined using a thick solid line (NNN). The DSS is underlined using a wavy line (*NNN*). The binding

domain is marked with italic letters and is underlined using a dotted line (*NNN*) and the polyadenylation signal is underlined using a double line (NNN).

[37] Fig. 17: Sequence of the minigene encoding the 3' coding sequence of cerulean and the binding domain BD_g (SEQ ID NO: 36). The CMV promoter is highlighted in grey. The binding domain is marked with italic letters and is underlined using a dotted line (NNN). The ASS is underlined using a wavy line (NNN) and bold letters. The 3' coding sequence of the cerulean protein is highlighted in grey and underlined using a thick solid line (NNN) and the polyadenylation signal is underlined using a double line (NNN).

[38] Fig. 18: Sequence of the minigene encoding the 5' coding sequence of *Sp*Cas9-VPR and the binding domain BD_k (SEQ ID NO: 37). The CMV promoter is highlighted in grey. The 5' coding sequence of the *Sp*Cas9-VPR protein is highlighted in bold letters and underlined using a thick solid line (NNN). The DSS is underlined using a wavy line (*NNN*). The binding domain marked with italic letters and is underlined using a dotted line (NNN), the polyadenylation signal is underlined using a double line (NNN).

[39] Fig. 19: Sequence of the minigene encoding the 3' coding sequence of *Sp*Cas9-VPR and the binding domain BD_k (SEQ ID NO: 38). The CMV promoter is highlighted in grey. The binding domain marked with italic letters and is underlined using a dotted line (NNN). The ASS is underlined using a wavy line and bold letters (*NNN*). The 3' coding sequence of the *Sp*Cas9-VPR protein is highlighted in grey and underlined using a thick solid line (NNN) and the polyadenylation signal is underlined using a double line (NNN).

[40] Fig. 20: Sequence of the first dual AAV comprising the 5' coding sequence of ABCA4 with introns (SEQ ID NO: 39): ITR sequences at the 5' end and the 3' end are highlighted in grey without underlining (NNN). The spacer sequences are depicted in small letters. The CMV promoter is highlighted in grey and is underlined using a solid line (NNN). The 5' coding sequence of the ABCA4 protein is underlined with a dotted line interrupted by introns indicated by a wavy underline (NNNNNNNNN). The DSS is depicted in italic letters and is underlined (*Nnn*, with the capital letters representing the coding sequence and the small letter representing the non-coding sequence). The binding domain is highlighted in grey and is underlined with an interrupted line (NNN). The polyA sequence is highlighted in dark grey and underlined with waves (NNN).

[41] Fig. 21: Sequence of the second dual AAV comprising the 3' coding sequence of ABCA4 (SEQ ID NO: 40): ITR sequences at the 5' end and the 3' end are highlighted in grey without underlining (NNN). The spacer sequences are depicted in small letters. The CMV promoter is highlighted in grey and is underlined using a solid line (NNN). The binding domain is highlighted in grey and is underlined with interrupted lines (NNN). The acceptor splice site is depicted in italic letters and is underlined (*NNN*). The 3' coding sequence of the ABCA4 protein is underlined using a dotted line interrupted by introns indicated by a wavy

underline (NNNNNNNNN). The polyA sequence is highlighted in dark grey and underlined with waves (NNN).

DETAILED DESCRIPTION

[42] The present inventors have found a novel acceptor splice region as described in the Examples. This acceptor splice region is very efficient and can be used in various applications, in which splicing is utilized. Thus, the acceptor splice region of the present invention can inter alia be used in *trans*-splicing, *cis*-splicing, SMaRT technology (Spliceosome-mediated RNA trans-splicing) using pre-mRNA *trans*-splicing molecules, for the reconstitution of split AAV vectors, AAV vector systems comprising 2 AAV vectors and/or trap vectors. However, the acceptor splice region as described herein can also be used for any further application in which splicing is of interest.

[43] The present invention relates to a nucleic acid sequence comprising

(i) an acceptor splice region sequence comprising

(ia) a pyrimidine tract, wherein the pyrimidine tract comprises

(aa) 5 to 25 nucleotides;

(ab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U);

(ib) an acceptor splice site,

ba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and

bb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G; and

(ii) a nucleotide sequence of interest, wherein the nucleotide sequence of interest

(iia) is located 3' or 5' to the acceptor splice region, preferably 3' to the acceptor splice region;

wherein the nucleic acid sequence according to the invention is preferably cleaved in the acceptor splice region thereby separating the nucleotide sequence of interest from the acceptor splice region sequence.

The invention further relates to the use of said nucleic acid sequence for cleaving in the acceptor splice region thereby separating the nucleotide sequence of interest from the acceptor splice region sequence.

[44] The terms "nucleic acid molecule", "nucleic acid sequence" or "nucleotide sequence" are used synonymously herein and encompass any nucleic acid molecule having a nucleotide sequence comprising purine- and pyrimidine bases which are comprised by said nucleic acid molecule/sequence, whereby said bases represent the primary structure of a

nucleic acid molecule. Nucleic acid sequences can include DNA, cDNA, genomic DNA, RNA, both sense and antisense strands. The RNA may for example be pre-mRNA, mRNA, tRNA or rRNA. The polynucleotide of the present invention can be composed of any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. The person skilled in the art will understand that a thymine (T) in a polydeoxynucleotide transcribes into an uracil (U) in a polyribonucleotide. The sequences referred to herein are typically provided as DNA sequences that may be transcribed into the corresponding RNA sequence (prior to any splicing event). Thus, a nucleic acid sequence comprising a T also discloses the corresponding (transcribed) RNA sequence, preferably pre-mRNA sequence, comprising a U.

[45] A variety of modifications can be made to DNA and RNA; thus, the term "nucleic acid molecules" or “nucleotide” can embrace chemically, enzymatically, or metabolically modified forms. "Modified" bases/nucleotides include, for example, tritylated bases and unusual bases such as inosine.

[46] The acceptor splice region sequence of the present invention comprises two features, namely an acceptor splice site and a pyrimidine tract. The pyrimidine tract comprises 5 to 25 nucleotides and wherein at least 60 % of the nucleotides within these (in total) 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U). It is also envisioned that the pyrimidine tract comprises 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 nucleotides, preferably 10 to 18, more preferably 12 to 16 nucleotides. Additionally or alternatively at least 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 97 %, 98 %, 99 % or 100 % of the nucleotides within these (in total) 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U). For example the pyrimidine tract can comprise 6 nucleotides. It is further envisioned that the pyrimidine tract comprises or consists of the sequence TTTTTT. It is further contemplated that the pyrimidine tract comprises or consists of the sequence TCTTTT. It is further contemplated that the pyrimidine tract comprises or consists of the sequence TTTTTTGTCATTT (SEQ ID NO. 11). It is further contemplated that the pyrimidine tract comprises or consists of the sequence TCTTTTGTCATCTA (SEQ ID NO. 12). It is further contemplated that the pyrimidine tract is preceded by 7 nucleotides, together comprising or consisting of the sequence CAACGAGTCTTTTGTCATCTA (SEQ ID NO. 13). The pyrimidine tract may also comprise or consist a sequence of TCTTTTGTCATCT (SEQ ID NO. 1). It is also contemplated that the pyrimidine tract as described herein has a nucleotide sequence which is at least 70%, 75 %, 80 %, 85 %, 90 %, 95 %, 97 %, 98 %, 99 % or 100 % identical to a sequence of SEQ ID NO. 1, 11, 12 or 13 or a sequence of TTTTTT or TCTTTT. In one embodiment the 5 to 25 nucleotides of the pyrimidine tract comprise a sequence of TTTTTT or TCTTTT. Preferably, the sequence between the last pyrimidine of

the pyrimidine tract and the acceptor splice site has less than 10, preferably less than 5, more preferably less than 3 bases. The term "pyrimidine tract" is used interchangeably herein with the term "poly-pyrimidine tract" and may be abbreviated as PPT or PolyC/T. The "last pyrimidine of the pyrimidine tract" refers to the most 3' pyrimidine of the 5 to 25 nucleotides of the pyrimidine tract.

[47] In accordance with the present invention, the term "identical" or "percent identity" in the context of two or more nucleic acid molecules, refers to two or more sequences or subsequences that are the same, or that have a specified percentage of nucleotides that are the same (e.g., at least 95 %, 96 %, 97 %, 98 % or 99 % identity), when compared and aligned for maximum correspondence over a window of comparison, or over a designated region as measured using a sequence comparison algorithm as known in the art, or by manual alignment and visual inspection. Sequences having, for example, 80 % to 95 % or greater sequence identity are considered to be substantially identical. Such a definition also applies to the complement of a test sequence. Those having skill in the art will know how to determine percent identity between/among sequences using, for example, algorithms such as those based on CLUSTALW computer program (Thompson Nucl. Acids Res. 2 (1994), 4673-4680) or FASTDB (Brutlag Comp. App. Biosci. 6 (1990), 237-245), as known in the art.

[48] Also available to those having skill in this art are the BLAST and BLAST 2.0 algorithms (Altschul Nucl. Acids Res. 25 (1977), 3389-3402). The BLASTN program for nucleic acid sequences uses as defaults a word size (W) of 28, an expectation threshold (E) of 10, match/mismatch score 1, -2, gap costs linear and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a word size (W) of 6, and an expectation threshold (E) of 10, gap costs are Existence: 11 and Extension: 1. Furthermore, the BLOSUM62 scoring matrix (Henikoff Proc. Natl. Acad. Sci., USA, 89, (1989), 10915) can be used.

[49] For example, BLAST2.0, which stands for Basic Local Alignment Search Tool (Altschul, Nucl. Acids Res. 25 (1997), 3389-3402; Altschul, J. Mol. Evol. 36 (1993), 290-300; Altschul, J. Mol. Biol. 215 (1990), 403-410), can be used to search for local sequence alignments.

[50] As outlined herein the acceptor splice region sequence of the present invention in addition to the pyrimidine tract comprises an acceptor splice site. As used herein an "acceptor splice site" has the meaning as known to the skilled person and as described inter alia in Alberts B, Johnson A, Lewis J, et al. (2002) "Molecular Biology of the Cell. 4th edition." New York: Garland Science under the headline "from DNA to RNA". The acceptor splice site of the present invention comprises or consists of the nucleotides NAGG, preferably CAGG or CAGGT (or CAGG or CAGGU in RNA). The acceptor splice site is usually located within a sequence defined as "acceptor splice region". The term "acceptor

splice site" (abbreviated to ASS) as used herein refers to the consensus acceptor splice sequence and may also be referred to as "splice acceptor site" (abbreviated to SAS). In the context of this invention the term "acceptor splice region" and "acceptor splice site" are used as distinct terms in order to distinguish between the region and the consensus acceptor splice site, wherein the acceptor splice region comprises the acceptor splice site. In the literature, the acceptor splice region as well as the acceptor splice site may be referred to as ASS or SAS.

[51] The acceptor splice site is called acceptor splice site because at this site the nucleic acid is cut by the so called spliceosome or an artificial variant thereof. The spliceosome and how it works are also known to the skilled person. The structure of the human spliceosome is for example described in Zhang et al. (2017) "An atomic structure of the human spliceosome" Cell 169, 918-926. Artificial spliceosomes include e.g. enzymes that mediate the same cut to a splice site as a spliceosome. Artificial spliceosomes can comprise a DNA enzyme as described in Coppins and Silvermann (2005) "Mimicking the First Step of RNA Splicing: An Artificial DNA Enzyme Can Synthesize Branched RNA Using an Oligonucleotide Leaving Group as a 5'-Exon Analogue" Biochemistry, 44 (41), pp 13439–13446 and Müller (2017) "Design and Experimental Evolution of trans-Splicing Group I Intron Ribozymes." Molecules. 22(1). pii: E75. doi: 10.3390/molecules22010075. The acceptor splice site as described herein has a sequence of NAGG, wherein N can be any nucleotide. For example, N is a nucleotide selected from A, C, T, G or U. The acceptor splice site as described herein can also have a sequence of CAGG. The acceptor splice site as described herein can also have a sequence of CAGGT (or CAGGU for RNA).

[52] The cut at the acceptor splice site of the present invention leads to two fragments, one comprising NAG of the NAGG and/or NAGGT splice site or CAG of the CAGG splice site and/or CAGGT splice site. The other fragment comprising the last G of the NAGG or CAGG splice site or the last GT of the NAGGT or the CAGGT splice site. Thus, it is envisioned that in the uses, nucleic acid sequences, AAV vectors, AAV vector systems and methods as described herein the finally obtained (after cleaving/splicing) nucleotide sequence (of interest) comprises the last G of the NAGG or CAGG sequence or the GT of the NAGGT or the CAGGT sequence. The person skilled in the art will understand that the splicing referred to herein takes place at the RNA level and hence the finally obtained (after cleaving/splicing) RNA nucleotide sequence (of interest) comprises the last G of the NAGG or CAGG sequence or the GU of the CAGGU sequence of the pre-mRNA nucleotide sequence. It is however also contemplated that in the methods and uses the final nucleic acid sequence (of interest) comprises the NAG of the NAGG splice site sequence, and/or the CAG of the CAGG or the CAGGT splice site sequence. It is thus envisioned that cleaving in the acceptor splice site comprises a cleaving between the G and the G of the acceptor

splice site sequence of NAGG, CAGG and/or CAGGT (or CAGGU) thereby separating the nucleotide sequence (of interest) from the intronic acceptor splice region sequence.

[53] The acceptor splice region of the present invention can inter alia be used in applications including *trans*- but also *cis*-splicing or any other splicing. Occurrence of splicing can be measured e.g. as described in Berger et al. (2016) "mRNA trans-splicing in gene therapy for genetic diseases" WIREs RNA 7: 487-498. For example, spliced nucleotide sequences can be quantified by end-point quantitative RT-PCR with specific primers and probes to discriminate the spliced product of interest. Further measuring of splicing can be performed as described in the Examples herein. For example, splicing can be measured by using marker genes as described in Orengo et al. (2006) "A bichromatic fluorescent reporter for cell-based screens of alternative splicing." Nucleic Acids Research 34(22):e148. Further splicing may be detected using restriction enzyme analysis as described in Berger et al., (2016) "Repair of rhodopsin mRNA by spliceosome-mediated RNA trans-splicing: a new approach for autosomal dominant retinitis pigmentosa.", Mol Ther.; 23(5):918-930.

[54] An acceptor splice site when present in a pre-mRNA typically corresponds to the 3'-end of the intron and the 5' end of the next exon. When artificially introducing an acceptor splice site, preferably an acceptor splice region, it is not absolutely required that the acceptor splice site is located at an intron-exon boundary, but may also be located within an open reading frame, within an intron or at the 5' end of a complete or a partial open reading frame. It is also envisioned that the acceptor splice site is located at the 5'-end of an intron. It is further envisioned that the acceptor splice site is not located in a pre-mRNA molecule but within an artificial molecule, such as any nucleic acid molecule.

[55] The acceptor splice site of the present invention is comprised in an acceptor splice region. Notably, the acceptor splice site of the present invention is located 3' to the pyrimidine tract.

[56] The acceptor splice region (or acceptor splice site module) can further comprise a branch point and/or a branch point sequence. In principle any suitable branch point or branch point sequence is contemplated by the present invention. Exemplary branch point(s) or branch point sequence(s) are inter alia described in Gao et al. (2008) "Human branch point consensus sequence is yUnAy" Nucleic Acid Research, vol. 36, no. 7, pp. 2257-2267; Mercer et al. (2016) "Genome-wide discovery of human splicing branch points" Genome Research 25: 290-303.

[57] For example, the branch point sequence may comprise or consist of the sequence UACUA**A**C. Additionally or alternatively the branch point sequence may comprise or consist of the sequence YNYUR**A**Y wherein Y is U or C and wherein R is A or G. Additionally or alternatively the branch point sequence may comprise or consist of the sequence

YNCUR**A**C wherein Y is U or C and wherein R is A or G. Additionally or alternatively the branch point sequence may comprise or consist of the sequence YNCUR**A**Y wherein Y is U or C and wherein R is A or G. Additionally or alternatively the branch point sequence may comprise or consist of the sequence CUR**A**Y wherein Y is U or C and wherein R is A or G. Additionally or alternatively the branch point sequence may comprise or consist of the sequence YUV**A**Y wherein Y is U or C, wherein R is A or G and wherein V is A, C or G. Additionally or alternatively the branch point sequence may comprise or consist of the sequence CUS**A**Y wherein Y is U or C, wherein S is G or C. Additionally or alternatively the branch point sequence may comprise or consist of the sequence CUG**A**C. Additionally or alternatively the branch point sequence may comprise or consist of the sequence CUA**A**C. The branch point may be present within a branch point sequence of C/U-U-N-**A**-C/U, CAACGA or CUC**A**A or GUC**A**A (wherein the branch point sequence is encoded by the respective DNA sequence of C/T-T-N-**A**-C/T, CAACGA or CTC**A**A or GTC**A**A). The underlined adenosine in all respective sequences represents the branch point nucleotide.

[58] The acceptor splice region of the present invention can further comprise a branch point nucleotide sequence (c), wherein the branch point nucleotide sequence

(ca) comprises 1-15 nucleotides;

(cb) comprises a branch point nucleotide, preferably adenosine (A); and

(cc) is located 5' to the pyrimidine tract and the acceptor splice site.

[59] It is also contemplated that the branch point sequence comprises about 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 or more nucleotides. It is further contemplated that the branch point sequence comprises 6 to 8 nucleotides. It is also envisioned that the branch point sequence comprises 8 nucleotides. Notably, the nucleotides of the branch point sequence can however, be present in series, which is why they can also comprise more than 15 such as e.g. 30 (15 + 15), 40, 50, 100, 200 or more nucleotides in total.

[60] The branch point nucleotide can be any nucleotide. Thus, the branch point nucleotide can be any of A (adenosine), T (thymine), G (guanine), C (cytosine), and U (uracil). Preferably, the branch point nucleotide is A (adenosine). The branch point nucleotide is located within the branch point sequence.

[61] The acceptor splice region can additionally or alternatively comprise an intronic splice enhancer. Such splice enhancers are known to the skilled person and are inter alia described in Wang et al. (2012) "Intronic splicing enhancers, cognate splicing factors and context-dependent regulation rules" Nature Structural & Molecular Biology, vol. 19, no. 10, pp. 1044-1053. Two examples are the splice enhancer sequences "AACG" (group F) and "CGAG" (group D). However, any suitable intronic splice region is embraced by this term.

[62] An exemplary intronic splice enhancer has a sequence of TGGGGGGAGG (SEQ ID NO: 2). Another exemplary intronic splice enhancer has a sequence of GTAACGGC. It is further contemplated that the intronic splice enhancer has a sequence of AACG. It is also contemplated that the acceptor splice regions as described herein has a nucleotide sequence which is at least 70%, 75 %, 80 %, 85 %, 90 %, 95 %, 97 %, 98 %, 99 % or 100 % identical to a sequence of SEQ ID NO: 2, GTAACGGC or AACG.

[63] The acceptor splice region of the present invention preferably further comprises about 7 nucleotides (such as 5-12 nucleotides, preferably 6-10, more preferably 6-8) 5' to the polypyrimidine tract. In a preferred embodiment the acceptor splice region further comprises (a) 7 nucleotides 5' to the pyrimidine tract having at least 4 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotide is C; (b) 7 nucleotides 5' to the pyrimidine tract having at least 5 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotide is C; (c) 7 nucleotides 5' to the pyrimidine tract having at least 6 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotide is C; (d) 7 nucleotides 5' to the pyrimidine tract having at least 4 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAA; (e) 7 nucleotides 5' to the pyrimidine tract having at least 4 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAAC; (f) 7 nucleotides 5' to the pyrimidine tract having at least 5 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAA; (g) 7 nucleotides 5' to the pyrimidine tract having at least 5 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAAC; (h) 7 nucleotides 5' to the pyrimidine tract having at least 6 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAA; (i) 7 nucleotides 5' to the pyrimidine tract having at least 6 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAAC or (j) 7 nucleotides 5' to the pyrimidine tract having the sequence CAACGAG. Without being bound by theory, this sequence may act as a splice enhancer.

[64] It is further envisioned that the acceptor splice region comprises 1, 2, 3, 4, 5, or more different/identical branch point sequences. It is further envisioned that the acceptor splice region comprises 1, 2, 3, 4, 5, or more different or identical intronic splice enhancer (sequences).

[65] It is envisioned that the pyrimidine tract, the acceptor splice site and optionally further the branch point sequence and/or the intronic splice enhancer, thus the whole acceptor splice region includes in total about 1000, 500, 250, 100, 50, 45, 40, 35, 30, 25, 20, 15 or less nucleotides, preferably 26 nucleotides.

[66] In some embodiments the branch point sequence or branch point nucleotide used by a specific acceptor splice site may be flexible. This means different branch points may be used, such as 3 or more, 5 or more, 7 or more, 9 or more, or 11 or more. In one embodiment the frequency of usage of each branch point is < 30% or < 20%. In cases where the choice

of the branch point sequence or branch point nucleotide used by a specific acceptor splice site is flexible, the strength of an acceptor splice region does not originate and is not dependent from the presence of a specific contained branch point sequence or branch point nucleotide. The flexibility to use any branch point sequence or branch point nucleotide may render the acceptor splice region highly efficient and versatile and particularly suitable for use in different nucleic acid sequences to induce efficient splicing in a sequence independent manner.

[67] The acceptor splice region as described herein and as present in the nucleic acid, the pre-mRNA *trans*-splicing molecule, the AVV vector, the AVV vector system and as used in the method according to the invention comprises in addition to the acceptor splice site and the polypyrimidine tract further nucleotides 5' of the polypyrimidine tract. In a preferred embodiment the acceptor splice region comprises about 7 nucleotides (such as 5-12 nucleotides, preferably 6-10, more preferably 6-8) 5' to the polypyrimidine tract. In a particularly preferred embodiment the acceptor splice region comprises the sequence CAACGAG 5' to the polypyrimidine tract. The acceptor splice region comprising the acceptor splice site, the polypyrimidine tract and the further nucleotides 5' of the polypyrimidine tract, preferably the further about 7 nucleotides 5' of the polypyrimidine tract, acts as a minimal acceptor splice region and may be inserted into any nucleic acid sequence to introduce a very potent functional acceptor splice region.

[68] It is further envisioned that the acceptor splice region of the present invention comprises or consists of a sequence of CAACGAG*TCTTTTGTCATCTA***CAGG**T (SEQ ID NO: 3). It is further contemplated that the acceptor splice site of the present invention comprises or consists of a sequence of CAACGAG*TTTTTTGTCATCTA***CAGG**T (SEQ ID NO: 4). It is also envisioned that the acceptor splice region comprises or consists of any of CTACACGCTC**A**AGCCGGAGGTC**A**ACAACGAG*TCTTTTGTCATCTA***CAGG**T (SEQ ID NO: 5), GCCGGAGGTC**A**ACAACGAG*TCTTTTGTCATCTA***CAGG**T (SEQ ID NO: 6), GTCTTTTGTCATCTA**CAGG**T (SEQ ID NO: 7), GTCTTTTGTCATCTA**CAGG**TGTTCGTG (SEQ ID NO: 26) or GTCTTTTGTCATCTA**CAGG**TGTTCGTGGTTCGTGGTCCA (SEQ ID NO: 8). Preferably the acceptor splice region of the present invention comprises or consists (or is encoded by a nucleic acid sequence comprising or consisting) of a sequence of CAACGAG*TCTTTTGTCATCTA***CAGG**T (SEQ ID NO: 3) or CAACGAG*TTTTTTGTCATCTA***CAGG**T (SEQ ID NO: 4).

[69] It is also contemplated that the acceptor splice region as described herein has a nucleotide sequence which is at least 70%, 75 %, 80 %, 85 %, 90 %, 95 %, 97 %, 98 %, 99 % or 100 % identical to any sequence of any one of SEQ ID NO: 3, 4, 5, 6, 7, 26 or 8. Preferably the acceptor splice region comprises a nucleotide sequence which is at least 85

%, 90 %, 95 %, 97 %, 98 %, 99 % or 100 % identical to any sequence of any one of SEQ ID NO: 3, 4, 5, 6, 7, 26 or 8, more preferably at least 85 %, 90 %, 95 %, 97 %, 98 %, 99 % or 100 % identical to any sequence of SEQ ID NO: 3 or 4, even more preferably at least 90 %, 95 %, 97 %, 98 %, 99 % or 100 % identical to any sequence of SEQ ID NO: 3 or 4.

[70] The nucleic acid sequence according to the invention may further comprise in addition to the acceptor splice site a nucleotide sequence of interest. The nucleotide sequence of interest can be any suitable nucleotide sequence. The term "nucleotide sequence of interest" and "nucleic acid sequence of interest" are used interchangeably herein. Typically it is referred to as "nucleotide sequence of interest" to better distinguish the term from the nucleic acid sequence or nucleic acid molecule according to the invention. For example, the nucleotide sequence of interest can be an intronic or exonic nucleotide sequence or comprise both intronic and exonic (i.e. non-coding and coding) sequences. It is further contemplated that the nucleotide sequence of interest is a cDNA, mRNA, rRNA or tRNA. The nucleotide sequence of interest or a portion thereof can be located 3' or 5', preferably 3', to the acceptor splice region. In certain embodiments the nucleotide sequence of interest or a portion thereof as comprised in any of the nucleic acid sequences, pre-mRNA *trans*-splicing molecules, AAV vectors, AAV vectors or as used in any of the methods according to the invention may be a transgene or a portion thereof. It has been found that the splicing efficiency, particularly the *trans*-splicing efficiency, can be improved by the presence of introns in the nucleotide sequence of interest or a portion thereof. Particularly, for a nucleotide sequence of interest or a portion thereof with more than 1000 bp it may be advantageous to introduce or maintain introns in the sequence. Preferably introns are present about every 200 to 1000 bp of the nucleotide sequence of interest or a portion thereof, but may also occur in lower or higher sequence intervals. The introns may originate from the nucleotide sequence of interest and/or from a different gene and/or artificial.

[71] The nucleotide sequence of interest (also referred to as nucleic acid sequence of interest) as used herein can also be a coding sequence, such as a sequence encoding for a polypeptide of interest or portions thereof. In a preferred embodiment the nucleotide sequence of interest is a coding sequence, more preferably a transgene, a sequence encoding for a polypeptide of interest or a portion thereof.

[72] It is further envisioned by the present invention that the nucleotide sequence of interest encodes a therapeutic polypeptide, a therapeutic nucleic acid (such as a therapeutic RNA) or a portion thereof.

[73] The therapeutic nucleic acid or therapeutic polypeptide can be used to treat an ocular disorder such as autosomal recessive severe early-onset retinal degeneration (Leber' s Congenital Amaurosis), congenital achromatopsia, Stargardt's disease, Best disease (vitelliform macular degeneration), Doyne's disease, retinitis pigmentosa (particularly

autosomal dominant, autosomal-recessive, X-linked, di- or polygenic retinitis pigmentosa), (X-linked) retinoschisis, macular degeneration (AMD), age-related macular degeneration, atrophic age-related macular degeneration, neovascular AMD, diabetic maculopathy, proliferative diabetic retinopathy (PDR), cystoid macular oedema, central serous retinopathy, retinal detachment, intra-ocular inflammation, glaucoma, posterior uveitis, congenital stationary night blindness, choroideremia, early-onset retinal dystrophy, cone, rod-cone or cone-rod dystrophy, pattern dystrophies, Usher syndrome and other syndromic ciliopathies, such as Bardet-Biedl syndrome, Joubert syndrome, Senior-Løken syndrome or Alström syndrome.

[74] The nucleic acid, therapeutic nucleic acid, therapeutic protein/polypeptide or therapeutic molecule can be used to treat a disorder affecting the photoreceptor cells (photoreceptor cell disease) such as the rods and/or cones. Non-limiting examples of a photoreceptor cell disease include achromatopsia, age-related macular degeneration, retinal degeneration, retinal dystrophy, retinitis pigmentosa, cone dystrophy, rod-cone dystrophy, color blindness, macular degeneration, night blindness, retinoschisis, choroideremia, diabetic retinopathy, hereditary optic neuropathy, Oguchi disease type I, retinitis punctata albescens (RPA), progressive retinal atrophy (PRA), fundus albipunctatus (FA) or congenital stationary night blindness (CSNB).

[75] The therapeutic nucleic acid or therapeutic polypeptide can be used to treat an inherited retinal disease (IRD). IRDs are a genetically and phenotypically heterogeneous group of disorders. They are the leading cause of blindness in people aged 15 to 45 and have an estimated prevalence of 1 in 1,500 – 1 in 3,000. IRDs can be grouped according to the primarily effected retinal cell type, i.e., rod or cone photoreceptors, and according to the state of disease. Non-limiting examples of IRDs are achromatopsia, age-related macular degeneration, retinal degeneration, retinal dystrophy, retinitis pigmentosa, cone dystrophy, rod-cone dystrophy, color blindness, macular degeneration, retinoschisis, choroideremia, diabetic retinopathy, hereditary optic neuropathy, Oguchi disease type I, retinitis punctata albescens (RPA), progressive retinal atrophy (PRA), fundus albipunctatus (FA) or congenital stationary night blindness (CSNB).

[76] The most common progressive IRD predominantly affecting cones is Stargardt macular dystrophy. It is inherited in an autosomal recessive manner and it is mostly caused by mutations in the *ABCA4* gene encoding the retinal transporter ABCR. In one embodiment the therapeutic nucleic acid encodes *ABCA4* or a portion thereof, wherein the therapeutic nucleic acid is used to treat Stargardt macular dystrophy.

[77] The most common IRD primarily affecting rod photoreceptors is retinitis pigmentosa (RP). RP is a progressive disease leading to night blindness and tunnel vision. In later stages, secondary cone photoreceptor cell death is induced eventually resulting in complete

vision loss. Some forms of retinitis pigmentosa have been associated with mutations in the rhodopsin encoding gene. In one embodiment the therapeutic nucleic acid encodes rhodopsin or a portion thereof, wherein the therapeutic nucleic acid is used to treat retinitis pigmentosa.

[78] The nucleic acid sequence according to the invention may also comprise a donor splice site, wherein the donor splice site is localized 3' to the nucleotide sequence of interest or a portion thereof. The term "donor splice region" is used herein interchangeably with "donor splice site" (abbreviated to DSS) or "splice donor site" (abbreviated as SDS). Thus, the nucleic acid sequence of the invention may also comprise a donor splice site (DSS). Donor splice sites are known to the skilled person and inter alia described in Buckley et al. (2009) "A method for identifying alternative or cryptic donor splice sites within gene and mRNA sequences. Comparisons among sequences from vertebrates, echinoderms and other groups" BMC Genomics 10: 318 and Qu et al. (2017) "A Bioinformatics-Based Alternative mRNA Splicing Code that May Explain Some Disease Mutations Is Conserved in Animals" Front. Genet., vol. 8 Art. 38. Here, any suitable donor splice site is contemplated by this term. A donor splice site is typically located at the 3'-end of the exon and at the 5'-end of the flanking intron. However, in the context of the present invention the donor splice site can also be located somewhere else within the nucleotide sequence of interest. Analogous to the acceptor splice site the donor splice site permits the spliceosome to cut at its site.

[79] In principle, the donor splice site as described herein can be located at any position of a nucleic acid molecule. It is also envisioned that the donor splice site is located at the 3' end of the (first) nucleic acid sequence. For example, the donor splice site sequence may comprise or consist of a sequence of AAGGTAAGT, AAGGTGAGT, CAGGTAAGT, CAGGTGAGT, AAGGTAAG, AAGGTGAG, CAGGTAAG or CAGGTGAG. It is also envisioned that the donor splice site sequence has a sequences which has a sequence identity of at least 70 %, 80 % 85 %, 90 %, 95 %, 98 %, 99% or 100 % to the sequence of AAGGTAAGT, AAGGTGAGT, CAGGTAAGT, CAGGTGAGT, AAGGTAAG, AAGGTGAG, CAGGTAAG or CAGGTGAG.

[80] Similarly, the acceptor splice region as described herein can be located at any position of a nucleic acid molecule. It is also envisioned that the acceptor splice site is located at the 5' end of the (second) nucleic acid sequence.

[81] The nucleotide sequence of interest may additionally or alternatively comprise a polyadenylation signal and/or a promoter as described herein.

[82] The present invention also relates to a method for producing a nucleic acid sequence, the method comprising

(A) providing a first nucleic acid sequence comprising one or more donor splice site sequences;

(B) providing a second nucleic acid sequence, wherein the second nucleic acid sequence comprises

(i) an acceptor splice region sequence comprising

(ia) a pyrimidine tract, wherein the pyrimidine tract comprises

(iaa) 5 to 25 nucleotides;

(iab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U);

(ib) an acceptor splice site,

(iba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and

(ibb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G;

thereby obtaining a nucleic acid sequence. The method is preferably an *in vitro* method. The method may further comprise a step of splicing the first nucleic acid sequence at the one or more donor splice sequence(s) and the second nucleic acid sequence at the acceptor splice site to obtain a spliced nucleic acid sequence. The method may also comprise a step of (C) cleaving the first nucleic acid sequence at the one or more donor splice site sequence(s) and cleaving the second nucleic acid sequence in the acceptor splice site, and (D) ligating the first cleaved nucleic acid sequence to the second cleaved nucleic acid sequence.

[83] This method reflects the so-called *trans*-splicing. In particular, *trans*-splicing is a type of splicing in which exons from two separate nucleic acid molecules (preferably pre-mRNA molecules) are joined together to form one nucleic acid molecule. For example, as is the case when pre-RNA is spliced to a final mRNA, tRNA or rRNA. This process is also known to the skilled person and is inter alia described in Alberts B, Johnson A, Lewis J, et al. (2002) "Molecular Biology of the Cell. 4th edition. New York: Garland Science under the headline "from DNA to RNA". Therefore, the spliced nucleic acid sequences are preferably mRNA, rRNA or tRNA sequences. The nucleotide sequence of interest can thus be a pre-mRNA.

[84] The first and the second nucleic acid sequence may be a DNA or an RNA sequence. In case the first and the second nucleic acid sequence is a DNA sequence, both, the first and the second nucleic acid sequence further comprise a promoter and at least the second, preferably both also comprise a transcription termination sequence, such as a polyA sequence. In case the first and the second nucleic acid sequence is an RNA sequence, at least the second, preferably both nucleic acid sequences comprise a termination sequence, such as a polyA sequence. Preferably the nucleic acid sequence is a DNA sequence, more preferably a vector or a plasmid comprising the DNA sequence.

[85] In one embodiment the first and the second nucleic acid sequence further comprise a

binding domain. In yet another embodiment the first and the second nucleic acid sequence is a pre-mRNA *trans*-splicing molecule or a DNA sequence encoding a pre-mRNA *trans*-splicing molecule as described herein.

[86] Without being bound by theory, the cleaving of the first nucleic acid sequence at the one or more donor splice site sequence(s) and cleaving the second nucleic acid sequence in the acceptor splice region sequence and ligating the first cleaved nucleic acid sequence to the second cleaved nucleic acid sequence is effected using two transesterification reactions in *cis*-splicing as well as in *trans*-splicing and is referred to as "splicing" herein. The term "*trans*-splicing" as used herein refers to RNA splicing, particularly pre-RNA splicing of two separate RNA or pre-RNA molecules to form the mature RNA, such as the final mRNA, tRNA or rRNA.

[87] The term "*in vitro* method" refers to a method outside the body, i.e., human or animal body, but includes cell lines or primary cells outside the body.

[88] The methods and uses as described herein can be performed in any suitable cell-free system such as a cell lysate or in any suitable host cell.

[89] Such suitable host cells are known to the skilled person. For example, suitable host cells comprise the splicing machinery, i.e. the spliceosome. A person of ordinary skill in the art can select regulatory elements for use in appropriate host cells, for example, mammalian or human host cells. Regulatory elements include, for example, promoters, termination sequences (such as transcription termination sequences, translation termination sequences), enhancers, and polyadenylation elements. Moreover, the expression levels of the transgene (nucleic acid sequence of interest) can be enhanced by using regulatory elements like enhancer sequences or the woodchuck hepatitis virus post-transcriptional response element (WPRE). However, these elements further limit the packaging capacity of the rAAVs. Therefore, in some cases, particularly where the size limitation is critical, it is recommended to omit such optional elements and preferably use small mandatory regulatory elements such as short polyadenylation signals as termination sequence for the AAV vector and particularly the dual rAAV vector system according to the invention. The host cell may be a eukaryotic cell or a mammalian cell, such as a human or rodent cell line. The host cell may be a HEK cell such as HEK293(T) cell, a Cos7 cell, a CHO cell, a fibroblast e.g. of human or murine origin, a retinoblastoma cell, a 661W cell, an induced pluripotent stem cell (iPSC) such as a human iPSC, a photoreceptor cell e.g. obtained from a vertebrate, a neuronal cell, or a glial cell.

[90] In one embodiment the method is performed in a host cell *in vitro* (i.e., outside the body) comprising introducing said first nucleic acid sequence and introducing said second nucleic acid sequence. Said first nucleic acid sequence and said second nucleic acid sequence may be introduced by transfection or transduction, more specifically co-

transfection or co-transduction.

[91] Therefore, in one embodiment the method comprising

(A) providing a first nucleic acid sequence comprising one or more donor splice site sequences;

(B) providing a second nucleic acid sequence, wherein the second nucleic acid sequence comprises

(i) an acceptor splice region sequence comprising

(ia) a pyrimidine tract, wherein the pyrimidine tract comprises

(iaa) 5 to 25 nucleotides;

(iab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U);

(ib) an acceptor splice site,

(iba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and

(ibb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G;

(ii) a nucleotide sequence of interest, wherein the nucleotide sequence of interest

(iia) is located 3' to the acceptor splice region;

thereby obtaining a nucleic acid sequence. The method may further comprise a step of splicing the first nucleic acid sequence at the one or more donor splice site sequence(s) and the second nucleic acid sequence at the acceptor splice site to obtain a spliced nucleic acid sequence. The method may also comprise a step of (C) cleaving the first nucleic acid sequence at the one or more donor splice site sequence(s) and cleaving the second nucleic acid sequence in the acceptor splice site, and (D) ligating the first cleaved nucleic acid sequence to the second cleaved nucleic acid sequence thereby obtaining a nucleic acid sequence.

[92] In cases where both nucleic acid sequences comprise a nucleotide sequence of interest the method additionally comprises the feature that the first nucleotide sequence comprises (ii) a nucleotide sequence of interest, wherein the nucleotide sequence of interest is located 5' to the donor splice site.

[93] Both the second and the first nucleic acid sequence can comprise a nucleotide sequence of interest as described herein. Such a nucleotide sequence of interest can be located 5' of the donor splice site. A nucleotide sequence of interest can additionally or alternatively be located 3' to the acceptor splice region. In another embodiment of the invention the polynucleotide comprises the following arrangement order: Preferably, 5'-(sequence of interest) – (donor splice site) – 3' combined with 5' – (acceptor splice region) –

(sequence of interest) -3'. Alternatively, 5'- (sequence of interest) – (acceptor splice region) – 3' combined with 5' – (donor splice site) – (sequence of interest) -3'.

[94] In a preferred embodiment the nucleotide sequence of interest is a sequence encoding for a polypeptide and the first nucleic acid sequence comprises the 5' portion of the sequence coding for a polypeptide and the second nucleic acid sequence comprises the 3' portion of the sequence coding for a polypeptide and *trans*-splicing reconstitutes the 5' portion and the 3' portion of the sequence coding for the polypeptide.

[95] The ligating of the cleaved nucleic acid sequences is a process known to the skilled person. For example, this ligation process can be performed by a ligating molecule. For example, the ligating molecule can be a RNA ligase or a protein with a RNA ligase function. Such RNA ligases are known to the skilled person and inter alia described e.g. in Chambers and Patrick (2015) "Archaeal Nucleic Acid Ligases and Their Potential in Biotechnology" Hindawi Publishing Corporation Archaea Volume 2015, Article ID 170571, 10 pages. Some (t)RNA ligases are for example described in Popow et al. (2012) "Diversity and roles of (t)RNA ligases" Cell Mol Life Sci. 69(16):2657-70. Therefore, the methods which include a ligation step can further comprise the step of the addition of a RNA ligase. The RNA ligase can for example be a mRNA ligase, a tRNA ligase or a rRNA ligase. Methods describing how e.g. ligated mRNA molecules can be analyzed are shown in the Examples.

[96] In a preferred embodiment the method is performed in a host cell and more preferably *in vitro* (i.e., outside the body). Thus, the first and the second nucleic acid sequence are introduced into a host cell, wherein the first and the second nucleic acid sequence are a recombinant nucleic acid sequence and/or heterologous to the host cell. Introducing the first and the second nucleic acid into the host cell comprises transfection or transduction. Transfection may be DNA or RNA transfection and methods for transfecting DNA or RNA are known to the person skilled in the art. Preferably the nucleic acid sequence is a DNA sequence and the DNA sequence is transfected in form of a plasmid or vector comprising said DNA sequence. Transduction as referred to herein means introducing a nucleic acid using a viral vector, wherein the viral vector may comprise DNA or RNA. In a preferred embodiment the viral vector is a DNA viral vector and may comprise single-stranded DNA (such as AAV) or double stranded DNA.

[97] In one embodiment the method according to the invention comprises (A) introducing into a host cell a first nucleic acid sequence comprising a pre-mRNA *trans*-splicing molecule sequence or a nucleic acid sequence encoding said pre-mRNA *trans*-splicing molecule, wherein the first pre-mRNA *trans*-splicing molecule comprises from 5' to 3'(a) a 5' portion of a nucleotide acid sequence of interest; (b) a donor splice site; (c) optionally a spacer sequence; (d) a first binding domain; and (e) optionally a termination sequence, preferably a polyA sequence and (B) introducing into a host cell a second nucleic acid sequence

comprising the pre-mRNA *trans*-splicing molecule sequence or a nucleic acid sequence encoding said pre-mRNA *trans*-splicing molecule, wherein the second pre-mRNA *trans*-splicing molecule comprises from 5' to 3'; (i) a second binding domain, which is complementary to the first binding domain of the first nucleic acid sequence; (ii) an acceptor splice region sequence comprising; (iia) a pyrimidine tract, wherein the pyrimidine tract comprises (iiaa) 5 to 25 nucleotides; (iiab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U); (iib) an acceptor splice site, (iiba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and (iibb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G; and (iii) a 3' portion of the nucleotide sequence of interest. The method may further comprise (C) cleaving the first nucleic acid sequence at the donor splice site sequence and cleaving the second nucleic acid sequence in the acceptor splice site; (D) ligating the first cleaved nucleic acid sequence comprising the 5' portion of the nucleotide sequence of interest to the second cleaved nucleic acid sequence comprising the 3' portion of the nucleotide sequence of interest, thereby obtaining the nucleic acid sequence of interest. In a preferred embodiment the first and the second nucleic acid sequence are a DNA sequence encoding said pre-mRNA *trans*-splicing molecule and the first nucleic acid sequence further comprises a promoter 5' of the 5' portion of the nucleic acid sequence of interest and the second nucleic acid sequence further comprises a promoter 5' of the second binding domain.

[98] The present invention also relates to a nucleic acid sequence, comprising

(i) an acceptor splice region sequence comprising

- (ia) a pyrimidine tract, wherein the pyrimidine tract comprises
  - (iaa) 5 to 25 nucleotides;
  - (iab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U);
- (ib) a acceptor splice site,
  - (iba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and
  - (ibb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G;

(ii) a nucleotide sequence of interest, wherein the nucleotide sequence of interest

- (iia) is located 3' or 5' to the acceptor splice region;

wherein the nucleotide sequence of interest is not exon 3 of rhodopsin gene of SEQ ID NO: 9 or wherein the nucleotide sequence of interest does not comprise a sequence as defined

in SEQ ID No: 9 and/or 10, or wherein the nucleic acid sequence of interest does not comprise exon 3 of rhodopsin gene of SEQ ID NO: 9.

[99] The present invention also relates to a nucleic acid sequence, comprising

(i) an acceptor splice region sequence comprising

(ia) a pyrimidine tract, wherein the pyrimidine tract comprises

(iaa) 5 to 25 nucleotides;

(iab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U);

(ib) a acceptor splice site,

(iba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and

(ibb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G;

(ii) a nucleotide sequence of interest, wherein the nucleotide sequence of interest

(iia) is located 3' or 5' to the acceptor splice region;

wherein the nucleotide sequence of interest is not the rhodopsin mRNA of SEQ ID NO. 10.

[100] It is envisioned that the nucleic acid sequence of the present invention has a length of at most 5800 nucleotides. It is further envisioned that the nucleic acid sequence of the present invention has a length of at most 5500, 5000, 4500, 4000, 3500, 3000, 2500, 2000, 1500, 1000, 500, 400, 300, 200 or at most 150 nucleotides.

[101] A further application in which the acceptor splice region of the present invention can be used in a nucleic acid molecule is a so called exogenous pre-mRNA *trans*-splicing molecule for the Smart technology. Such pre-mRNA *trans*-splicing molecules in which the acceptor splice region of the present invention can be introduced are inter alia described in WO 2011/042556, WO2013/025461 as well as by Berger et al. (2016) "mRNA trans splicing in gene therapy for genetic diseases" WIREs RNA, 7:487-498, Puttaraju et al. (1999) "Spliceosome-mediated RNA trans-splicing as a tool for gene therapy" Nature Biotechnology, vol. 17, pp. 246-252; and Mansfield et al. (2003) "5' Exon replacement and repair by spliceosome-mediated RNA trans-splicing" RNA, vol. 9: 1290-1297. From these references the skilled person also knows how to construct such pre-mRNA *trans*-splicing molecules. Some exemplary pre-mRNA *trans*-splicing molecule nucleic acid molecules are also described herein.

[102] The present invention also concerns a pre-mRNA *trans*-splicing molecule comprising,

(i) an acceptor splice region, comprising

(ia) a pyrimidine tract, wherein the pyrimidine tract comprises

(iaa) 5 to 25 nucleotides;

(iab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C)or uracil (U);

(ib) a acceptor splice site,

(iba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and

(ibb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, U or G;

(ii) a nucleotide sequence of interest; wherein the acceptor splice region is localized 5' to the nucleotide sequence of interest;

(iii) a binding domain targeting pre-mRNA, which is localized 5' to the nucleic acid sequence of interest; and

(iv) optionally a spacer sequence, wherein the spacer sequence is localized between the binding domain and the acceptor splice region.

[103] In a preferred embodiment the acceptor splice region is localized 5' to the nucleotide sequence of interest and the binding domain is localized 5' to acceptor splice region. Preferably the nucleic acid sequence or the pre-mRNA *trans*-splicing molecule further comprise a termination sequence, preferably a polyA sequence.

[104] The present invention further relates to a pre-mRNA *trans*-splicing molecule comprising,

(i) an acceptor splice region comprising

(ia) a pyrimidine tract, wherein the pyrimidine tract comprises

(iaa) 5 to 25 nucleotides;

(iab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymidine (T) and/or uracil (U);

(ib) an acceptor splice site,

(iba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and

(ibb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, U or G;

(ii) a nucleotide sequence of interest, wherein the acceptor splice region is localized 5' to the nucleotide sequence of interest;

(iii) a donor splice site, wherein the donor splice site is localized 3' to the nucleotide sequence of interest;

(iv) a first binding domain targeting pre-mRNA located 5' to the nucleotide sequence of interest;

(v) a second binding domain targeting pre-mRNA located 3' to the nucleotide sequence of interest;

(vi) optionally a first spacer sequence, wherein the first spacer is localized between the first binding domain and the acceptor splice region;

(vii) optionally a second spacer sequence, wherein the second spacer is localized between the second binding domain and the donor splice site.

[105] In one embodiment the first binding domain is localized 5' to acceptor splice region and the second binding domain is localized 3' to the donor splice site. The person skilled in the art will understand that the DNA sequence encoding the pre-mRNA *trans*-splicing molecule comprises a promoter in order to transcribe the pre-mRNA *trans*-splicing molecule, which is a RNA molecule. The term "binding domain targeting pre-mRNA" may also be referred to as "pre-mRNA targeting binding domain", which may be located 5' or 3' of the nucleotide sequence of interest.

[106] The nucleic acid sequences as described herein can reflect a pre-mRNA *trans*-splicing molecule. The person skilled in the art will understand that the nucleic acid sequences and the pre-mRNA *trans*-splicing molecules as described herein are recombinant sequences or molecules. These pre-mRNA *trans*-splicing molecules are known in the art and inter alia described in WO 2011/042556, WO2013/025461 as well as by Berger et al. (2016) "mRNA trans splicing in gene therapy for genetic diseases" WIREs RNA, 7:487-498, Puttaraju et al. (1999) "Spliceosome-mediated RNA trans-splicing as a tool for gene therapy" Nature Biotechnology, vol. 17, pp. 246-252, Mansfield et al. (2003) "5' Exon replacement and repair by spliceosome-mediated RNA trans-splicing" RNA, vol. 9: 1290-1297 and Berger et al. (2016) "mRNA trans-splicing in gene therapy for genetic diseases" WIREs RNA 7: 487-498. Thus, the skilled person knows how to constructs those pre-mRNA *trans*-splicing molecules. It is envisioned that the pre-mRNA *trans*-splicing molecule binds to target pre-mRNA, wherein the pre-mRNA may be a natural pre-mRNA, particularly a pre-mRNA endogenous to the host cell, or a recombinant pre-mRNA, particularly another pre-mRNA *trans*-splicing molecule. It is also contemplated that the pre-mRNA *trans*-splicing molecule preferentially induces the *trans*-splicing reaction more efficiently than the *cis*-splicing reaction.

[107] The term "binding domain targeting pre-mRNA" as used herein can be any suitable "binding domain targeting pre-mRNA", which means that is complementary to a sequence of the targeted pre-mRNA located 5' or 3' to the nucleotide sequence of interest. The "binding domain targeting pre-mRNA" as used herein may also be referred to as "target binding domain", "binding domain" (abbreviated to BD) or "binding sequence". For example, the binding domain may recognize the target pre-mRNA or mRNA by base-pairing. For example, the target on the mRNA can be an intron. The target binding domain of the pre-mRNA *trans*-

splicing molecules described herein may contain one or two binding domains of at least 15 to 30 nucleotides, preferably 80-120 nucleotides, more preferably about 100 nucleotides; or having long target binding domains as described in US Patent Publication No. US 2006-0194317 A1, of up to several hundred nucleotides which are complementary to and in anti-sense orientation to the targeted region of the selected (e.g. endogenous) pre-mRNA. This confers specificity of binding and anchors the endogenous pre-mRNA closely in space so that the spliceosome e.g. of the nucleus of a host cell can *trans*-splice a portion of the pre-mRNA *trans*-splicing molecule to a portion of the (e.g. endogenous) pre-mRNA. Alternatively the binding domain may be reverse complementary to the binding domain of another recombinant pre-mRNA, such as another pre-mRNA *trans*-splicing molecule. This confers specificity of binding and anchors the two pre-mRNA *trans*-splicing molecules closely in space so that the spliceosome e.g. of the nucleus of a host cell can *trans*-splice a portion of the one pre-mRNA *trans*-splicing molecule to a portion of the other pre-mRNA *trans*-splicing molecule. Thus, it is envisioned that the binding domain comprises between about 15 to 250 nucleotides, between about 15 to 200 nucleotides, between about 100 to 200 nucleotides or less than 500, 400, 300 or 200 nucleotides. In one embodiment the binding domain comprises 50-150 nucleotides, preferably 80-120 nucleotides, even more preferably 90 to 110 nucleotides or about 100 nucleotides. Alternatively or more preferably furthermore, an efficient binding domain has a GC content of 45-65% and/or is derived from intronic eukaryotic sequences or bacterial sequences and/or comprises at least one branch point in the last 30 bp (as e.g., evaluated using the human splicing finder 3.1 (http://www.umd.be/HSF/index.html)).

[108] Within a host cell, the binding domain may bind to endogenous pre-mRNA or to a second heterologous or recombinant pre-mRNA, (provided or introduced to the host cell) as long as the endogenous pre-mRNA or the second heterologous or recombinant pre-mRNA comprises a sequence reverse complement to the target binding domain. Thus for a second heterologous or recombinant pre-mRNA, it comprises a binding domain reverse complement to the binding domain of the first heterologous pre-mRNA. The terms “endogenous” and heterologous” are used herein relative to the host cell. Targeting endogenous pre-mRNA is also referred to as spliceosome-mediated RNA *trans*-splicing and is typically used to replace a part of the endogenous pre-mRNA. Targeting a second heterologous or recombinant pre-mRNA is also referred to as *trans*-splicing and generates a new recombinant mRNA ligating the first nucleotide sequence or a portion thereof (5’) from one pre-mRNA *trans*-splicing molecule to the second nucleotide sequence or a portion thereof (3’) of another pre-mRNA *trans*-splicing molecule. The term “recombinant” as used herein refers to a DNA molecule or sequence formed by laboratory methods of genetic recombination, such as molecular cloning and to a RNA or polypeptide molecule/sequence encoded by said recombinant DNA

molecule or sequence.

[109] The term "pre-mRNA" or "pre-RNA" refers to the RNA prior to *trans*-splicing and typically describes RNA molecules prior to *cis*-splicing (e.g., comprising introns and exons), but may also refer to RNA molecules post *cis*-splicing (e.g., comprising exons only), such as in case where the binding domain target a sequence spanning an exon-exon boundary.

[110] The term "heterologous" as used herein refers to a protein or a nucleic acid molecule or sequence that is experimentally transferred or introduced into a cell and is therefore not endogenous to this cell. Wherein the heterologous protein or nucleic acid molecule or sequence may be from a different cell type or a different species than the recipient, it may as well be from the same cell type or species as the recipient, as long as it is introduced into the recipient host cell. As referred to herein, the terms "heterologous" and "recombinant" are used interchangeably.

[111] An exemplary binding domain suitable for use in the present invention, without being limited thereto, is a binding domain having an RNA sequence encoded by the sequence of SEQ ID NO: 27, 28, 32, 33 or 18. Further target binding domains may be identified using a reporter reconstitution assay, such as the cerulean reconstitution assay for determining reconstitution efficiencies as described and used in Examples 2 and 5 or as described in Riedmayr LM. (2020, SMaRT for Therapeutic Purposes. Methods Mol Biol 2079:219-32. doi: 10.1007/978-1-4939-9904-0_17) or Dallinger G. et al. (2003, Development of spliceosome-mediated RNA *trans*-splicing (SMaRT) for the correction of inherited skin diseases. Exp Dermatol 12:37-46).

[112] A second binding region may be placed at the 3' end of the molecule and can be incorporated into the pre-mRNA *trans*-splicing molecule of the invention. Absolute complementarity, although preferred, is not required. A sequence "complementary" to a portion of the endogenous pre-mRNA, as referred to herein, means a sequence having sufficient complementarity to be able to hybridize with the endogenous pre-mRNA, forming a stable duplex. The ability to hybridize will depend on both the degree of complementarity and the length of the nucleic acid (See, for example, Sambrook et ah, 1989, Molecular Cloning, A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.).

[113] Complementarity as used with regard to binding domains targeting pre-mRNA thus means that the binding domain comprises a sequences, which is at least 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 % 99 % or 100 % complementary to the target sequence on the pre-mRNA, preferably at least 90 %, 95 %, 98 % 99 % or 100 % complementary to the target sequence on the pre-mRNA.

[114] A spacer region to separate the splice site(s) from the binding domain is preferably also included in the pre-mRNA *trans*-splicing molecule. The spacer sequence is a region of the pre-mRNA *trans*-splicing molecule that can cover elements of the 3' and/or 5' splice site

of the pre-mRNA *trans*-splicing molecule by relatively weak complementarity thereby preventing non-specific trans-splicing.

[115] The spacer separating the 5' donor splice site and the 3'-end binding domain can comprise between about 10 and 100 nucleotides, preferably between about 10 to 70, between about 20 and 70 nucleotides, between about 20 to 50 and more preferably between about 30 and 50 nucleotides. This spacer may comprise a downstream intronic splice enhancer (DISE). The spacer sequence that separates the 3' acceptor splice region from the 5'-end binding domain can comprise between about 2 and 100 nucleotides or between about 10 and 100 nucleotides, preferably between about 2 and 50, more preferably between about 5 and 20 nucleotides.

[116] The spacers can be non-coding sequences but may be designed to include features such as stop codons which would block any translation of a spliced pre-mRNA *trans*-splicing molecule. Additional features can be added to the pre-mRNA *trans*-splicing molecule are known to the skilled person as described above. It is further envisioned that the nucleic acid sequences or pre-mRNA *trans*-splicing molecules as described herein are comprised in a (recombinant) vector. Such vectors are known to the skilled person as well. Vectors comprising the nucleotide sequences of the present invention such as the pre-mRNA *trans*-splicing molecules of interest can be plasmid, viral, or others known in the art, used for replication and expression in mammalian cells.

[117] Expression of the nucleotide sequence of interest or a nucleotide sequence as described herein such as pre-mRNA *trans*-splicing molecule can be regulated by any promoter/enhancer sequences known in the art to act in mammalian, preferably human cells. Such promoters/enhancers can be inducible or constitutive. Such promoters are described also elsewhere herein. One exemplary promoter is the human or murine Rhodopsin or short or medium or long wavelength sensitive opsin promoter, etc. Any type of plasmid, cosmid, YAC or viral vector can be used to prepare the recombinant DNA construct which can be introduced directly into the tissue site. Alternatively, viral vectors can be used which selectively infect the desired target cell. Vectors for use in the practice of the invention include any eukaryotic expression vectors, including but not limited to viral expression vectors such as those derived from the class of retroviruses, adenoviruses or adeno-associated viruses. In a preferred embodiment, the recombinant vector of the invention is a eukaryotic expression vector.

[118] In another specific embodiment, the present invention comprises delivering the nucleic acid sequence such as the pre-mRNA *trans*-splicing molecule of the invention or a nucleic acid sequence encoding the pre-mRNA *trans*-splicing molecule of the invention to a target cell. Various delivery systems are known and can be used to transfer the compositions of the invention into cells, e. g. encapsulation in liposomes, microparticles,

microcapsules, recombinant cells capable of expressing the composition, receptor-mediated endocytosis, construction of a nucleic acid as part of a retroviral, adenoviral, adeno-associated viral or other vector, injection of DNA, electroporation, calcium phosphate mediated transfection, etc. The present invention also concerns a cell comprising the nucleic acid sequence such as the pre-mRNA *trans*-splicing molecule of the invention, a nucleic acid sequence encoding the pre-mRNA *trans*-splicing molecule or a recombinant vector comprising the nucleic acid sequence encoding the pre-mRNA *trans*-splicing molecule of the invention.

[119] The nucleic acid sequence of interest or nucleic acid sequences as disclosed herein can be used to provide a gene encoding a functional biologically active molecule to cells of an individual with an inherited genetic disorder where expression of the missing or mutant gene product produces a normal phenotype. This can inter alia be achieved by an adeno-associated virus vector as also disclosed herein.

[120] The invention also relates to a deoxyribonucleic acid (DNA) molecule comprising a promoter and a sequence encoding the pre-mRNA *trans*-splicing molecule as described herein. Preferably the DNA molecule is a vector or a plasmid, wherein the vector may be a viral vector such as an AAV, an adenovirus, or a lentivirus vector or a plasmid. The pre-mRNA *trans*-splicing molecule or the DNA molecule such as the viral vector may be used in therapy, particularly gene therapy, more particularly in gene therapy treating an ocular disorder.

[121] A further application in which the acceptor splice site of the present invention can be used is in any AAV vector or AAV vector system. Such vector systems and how they can be constructed are known to the skilled person and *inter alia* described in Carvalho et al. (2017) "Evaluating efficiencies of dual AAV approaches for retinal targeting" Frontiers in Neuroscience, vol. 11, Article 503, US 2014/0256802 or Trapani et al. (2013) "Effective delivery of large genes to the retina by dual AAV vectors" EMBO Molecular Medicine, vol. 6, no. 2, pp. 194-211. Some exemplary AAV vectors and AAV vectors systems are also described herein.

[122] The present invention also relates to an adeno-associated virus (AAV) vector comprising at least two inverted terminal repeats comprising a nucleic acid sequence between these two inverted terminal repeats, wherein said nucleic acid sequence comprises from 5' to 3 '

(i) a promoter;

(ii) a binding domain;

(iii) optionally a spacer sequence;

(iv) an acceptor splice region sequence comprising

(a) a pyrimidine tract, wherein the pyrimidine tract comprises

(aa) 5 to 25 nucleotides;

(ab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U);

(b) a acceptor splice site,

ba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and

bb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G (preferably A, C, T or G);

(v) a nucleotide sequence of interest

(vi) optionally a termination sequence, such as a poly A sequence.

[123] Structurally, AAVs are small (25 nm), single-DNA stranded non-enveloped viruses with an icosahedral capsid. Naturally occurring or engineered AAV variants (also AAV serotypes) that differ in the composition and structure of their capsid (cap) protein have varying tropism, i.e. ability to transduce different (retinal) cell types (Boye et al. (2013) "A comprehensive review of retinal gene therapy" Molecular therapy 21 509-519). When combined with ubiquitously active promoters this tropism defines the site of gene expression. Whereas in combination with cell-type specific promoters the degree of site-specificity (i.e. transgene expression only in rod or cone photoreceptors) is defined by the combination of both, the tropism of the AAV serotype and the specificity of the promoter (Schön et al. (2015) "Retinal gene delivery by adeno-associated virus (AAV) vectors: Strategies and applications" Eur J Pharm Biopharm. 95(Pt B):343-52).

[124] The terms "adeno-associated virus vector" or "recombinant AAV" or "rAAV" all used interchangeably are meant to include any AAV that comprises a heterologous polynucleotide sequence (also referred to a nucleotide sequence of interest herein) in its viral genome. In general, the heterologous polynucleotide/nucleotide sequence of interest is flanked by at least one, and generally by two naturally occurring or variant AAV inverted terminal repeat sequences (ITRs). The term rAAV vector encompasses both rAAV vector particles and rAAV vector plasmids. Thus, for example, a rAAV that comprises a heterologous polynucleotide sequence would be a rAAV that includes a nucleic acid sequence not normally included in a naturally-occurring, wild-type AAV, for example, a transgene (e.g. a non-AAV RNA-coding polynucleotide sequence, non-AAV protein-coding polynucleotide sequence), a non-AAV promoter sequence, a non-AAV poly-adenylation sequence, etc.

[125] Such recombinant AAV vectors are common general knowledge in the art and so the skilled person also knows how to construct such recombinant AAVs.

[126] A "rAAV vector genome" or "rAAV genome" is an AAV genome (i.e. vDNA) that comprises one or more heterologous nucleic acid sequences. rAAV vectors generally require

only the terminal repeat(s) (TR) in as to generate virus. All other viral sequences are considered dispensable and may be supplied in *trans* (Muzyczka, (1992) Curr Topics Microbiol. Immunol. 158:97). Typically, the rAAV vector genome will only retain the one or more TR sequence so as to maximize the size of the transgene/nucleotide sequence of interest/heterologous nucleic acid sequence that can be efficiently packaged by the vector/capsid. The structural and non-structural protein coding sequences may be provided in *trans* (e.g., from a vector, such as a plasmid, or by stably integrating the sequences into a packaging cell). In embodiments of the invention the rAAV vector genome comprises at least one TR sequence (e g., AAV TR sequence), optionally two TRs (e.g., two AAV TRs), which typically will be at the 5' and 3' ends of the vector genome and flank the heterologous nucleic acid, but need not be contiguous thereto. The TRs can be the same or different from each other.

[127] The term “inverted terminal repeat”, "terminal repeat" or "TR" all used interchangeable includes any viral terminal repeat or synthetic sequence that forms a hairpin structure and functions as an inverted terminal repeat (i.e. if it mediates the desired functions such as replication, virus packaging, integration and/or provirus rescue, and the like). The TR can be an AAV TR or a non-AAV TR. For example, a non-AAV TR sequence such as those of other parvoviruses (e.g. canine parvovirus (CPV), mouse parvovirus (MVM), human parvovirus B-19) or any other suitable virus sequence (e.g., the SV40 hairpin that serves as the origin of SV40 replication) can be used as a TR, which can further be modified by truncation, substitution, deletion, insertion and/or addition. Further, the TR can be partially or completely synthetic, such as the "double-D sequence" as described in United States Patent No. 5,478,745. The terminal repeat can have a length of about 50, 100, 150, 200, 250, 300 or more nucleotide. For example, the terminal repeat has about 145 nucleotides. For example the inverted terminal repeat can have a sequence of SEQ ID NO. 14 and/ or 15. It is also envisioned that the inverted terminal repeat can have a sequence having at least 60 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 % or 100 % sequence identity to a sequence of SEQ ID 14 and/or 15. Preferably, the AAV vector comprises both of SEQ ID NO. 14 and 15 or a sequence having at least 60 % sequence identity to SEQ ID 14 and 15.

[128] An 'AAV terminal repeat" or "AAV TR" may be from any AAV, including but not limited to serotypes 1, 2, 3, 3B, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 or any other AAV now known or later discovered. An AAV terminal repeat need not have the native terminal repeat sequence (e.g., a native AAV TR sequence may be altered by insertion, deletion, truncation and/or missense mutations), as long as the terminal repeat mediates the desired functions, e.g., replication, virus packaging, integration, and/or provirus rescue, and the like. The virus vectors of the invention can further be "targeted" virus vectors (e.g., having a directed tropism) and/or a "hybrid" parvovirus (i.e, in which the viral TRs and viral capsid are from

different parvoviruses) as described in international patent publication WO 00/28004.

[129] It is also envisioned that the recombinant AAV or AAV vector comprises or has a sequence having a sequence identity of at least 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 97 %, 99 % or 100 % with a naturally occurring AAV type 1 (AAV-1), AAV type 2 (AAV-2), AAV type 3 (AAV-3), AAV3B, AAV type 4 (AAV-4), AAV type 5 (AAV-5), AAV type 6 (AAV-6), AAV type 7 (AAV-7), AAV type 8 (AAV-8), AAV9, AAV10, AAV11, AAV12, rh10, avian AAV, bovine AAV, canine AAV, equine AAV, primate AAV, non-primate AAV, and ovine AAV.

[130] According to the invention the nucleic acid sequence between the two inverted terminal repeats further comprises a promotor as described herein. The selection of the promoter to be employed in the AAV may be made from among a wide number of constitutive or inducible promoters that can express the selected transgene/nucleotide sequence of interest in the desired target cell. The target cell can be a photoreceptor cell. The promoter may be derived from any species, including human.

[131] A promoter as described herein can be "cell specific". The term "cell-specific" means that the particular promoter selected for the recombinant vector can direct expression of the selected transgene/nucleotide sequence of interest in a particular cell or ocular cell type. Useful promoters include, without limitation, the rod opsin promoter, the red-green opsin promoter, the blue opsin promoter, the cGMP-P-phosphodiesterase promoter, the mouse opsin promoter (Beltran et al. (2010) “rAAV2/5 gene-targeting to rods:dose-dependent efficiency and complications associated with different promoters.” Gene Ther. 17(9):1162-74), the rhodopsin promoter (Mussolino et al, Gene Ther, July 2011, 18(7):637-45); the alpha-subunit of cone transducin (Morrissey et al, BMC Dev, Biol, Jan 2011, 11 :3); beta phosphodiesterase (PDE) promoter; the retinitis pigmentosa (RP1) promoter (Nicord et al, J. Gene Med, Dec 2007, 9(12): 1015-23); the NXNL2/NXNL 1 promoter (Lambard et al, PLoS One, Oct. 2010, 5(10):el3025), the PE65 promoter; the retinal degeneration slow/peripherin 2 (Rds/perph2) promoter (Cai et al, Exp Eye Res. 2010 Aug;91(2): 186-94); and the VMD2 promoter (Achi et al, Human Gene Therapy, (2009) 20:31- 9).

[132] Useful promoters to be used in the present invention also include, without limitation, the rod opsin promoter (RHO), the red-green opsin promoter, the blue opsin promoter, the cGMP-Phosphodiesterase promoter, the SWS promoter (blue short wavelength-sensitive (SWS) opsin promoter), the mouse opsin promoter (Beltran et al 2010 cited above), the rhodopsin promoter (Mussolino et al, Gene Ther, July 2011, 18(7):637-45); the alpha-subunit of cone transducin (Morrissey et al, BMC Dev, Biol, Jan 2011, 11 :3); the cone arrestin (ARR3) promoter (Kahle NA et al., Hum Gene Ther Clin Dev, September 2018, 29(3):121-131), the beta phosphodiesterase (PDE) promoter; the retinitis pigmentosa (RP1) promoter (Nicord et al, J. Gene Med, Dec 2007, 9(12): 1015-23); the NXNL2/NXNL 1 promoter (Lambard et al, PLoS One, Oct. 2010, 5(10):el3025), the RPE65 promoter; the retinal

degeneration slow/peripherin 2 (Rds/perph2) promoter (Cai et al, Exp Eye Res. 2010 Aug;91(2): 186-94); the VMD2 promoter (Achi et al, Human Gene Therapy, 2009 (20:31- 9)), and the ABCA4 promoter or any hybrid promoter consisting of at least two different promoters.

[133] It is also contemplated that the nucleic acid sequence between the two inverted terminal repeats further comprises a termination signal, such as a polyadenylation signal, located 3' to the nucleic acid sequence of interest. The termination signal, such as the polyadenylation signal, located 3' to the nucleic acid sequence of interest is optional, because the pre-RNA may be *trans*-spliced to an endogenous pre-RNA or a further recombinant pre-RNA to form the 5' end of the mature RNA, such as the mature mRNA. However, it has been shown that *trans*-splicing efficacy is higher in the presence of a polyadenylation signal.

[134] A promotor is required in order to transcribe the pre-RNA, however, the nucleic acid sequence does not necessarily need an ATG (translation start) and/or a Kozak consensus sequence. In case the pre-RNA may be *trans*-spliced to an endogenous pre-RNA or a further recombinant pre-RNA to form the 3' end of the mature RNA, such as the mature mRNA, the lack of an ATG and/or a Kozak consensus sequence may be advantageous to avoid undesired translation from the pre-RNA prior to *trans*-splicing.

[135] Selection of these and other common vector and regulatory elements are conventional and many such sequences are available. See e.g. Ausubel et al., Current Protocols in Molecular Biology, John Wiley & Sons, New York, 1989). Of course, not all vectors and expression control sequences will function equally well to express all of the transgenes as described herein. However, one of skill in the art may make a selection among these, and other, expression control sequences without departing from the scope of this invention.

[136] The present invention also relates to an adeno-associated virus (AAV) vector system comprising

(I) a first AAV vector comprising a nucleic acid sequence at least two inverted terminal repeats comprising a nucleic acid sequence between these two inverted terminal repeats, wherein said nucleic acid sequence between these two inverted terminal repeats comprises from 5' to 3'
  (a) a promoter,
  (b) a nucleotide sequence encoding an N-terminal portion of a polypeptide of interest;
  (c) a donor splice site;
  (d) optionally a spacer sequence;
  (e) a first binding domain; and

(f) optionally a termination signal, preferably a polyA signal:

(II) a second AAV vector comprising a nucleic acid sequence comprising at least two inverted terminal repeats comprising a nucleic acid sequence between these two inverted terminal repeats, wherein said nucleic acid sequence between these two inverted terminal repeats comprises from 5' to 3'

(i) a promoter;

(ii) a second binding domain, which is complementary to the first binding domain of the first AAV vector;

(iii) optionally a spacer sequence;

(iv) an acceptor splice region sequence comprising

(a) a pyrimidine tract, wherein the pyrimidine tract comprises

(aa) 5 to 25 nucleotides;

(ab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U);

(b) an acceptor splice site,

(ba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and

(bb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G;

(v) a nucleotide sequence encoding a C-terminal portion of the polypeptide of interest;

(vb) wherein the C-terminal portion of the polypeptide of interest and the N-terminal portion of the polypeptide of interest reconstitute the polypeptide of interest; and

(vi) a termination signal, preferably a polyA signal.

[137] Thus, the C-terminal portion of the polypeptide of interest corresponds to the portion of the polypeptide of interest that is missing from the N-terminal portion of the polypeptide comprised in the first AAV vector. In one embodiment the polypeptide is a full-length polypeptide and the first AAV vector comprises an N-terminal portion of a full-length polypeptide of interest and the second AAV vector comprises a C-terminal portion of the full-length polypeptide of interest and the C-terminal portion of the full-length polypeptide of interest and the N-terminal portion of the full-length polypeptide of interest reconstitute the polypeptide of interest. Thus, the C-terminal portion of the full-length polypeptide of interest corresponds to the portion of the full-length polypeptide of interest that is missing from the N-terminal portion of the full-length polypeptide comprised in the first AAV vector. More specifically following *trans*-splicing the mRNA comprises a sequence encoding the N-terminal portion of the polypeptide of interest and the C-terminal portion of the polypeptide of

interest (in frame), wherein the mRNA encodes the (full length) polypeptide of interest and hence the polypeptide of interest is reconstituted.

[138] Notably, there exists a size limitation of rAAV or AAV vector genome packaging capacity of about ~ 5 kb. Since cDNAs of many therapeutic proteins are large, devising strategies to deliver large transgenes using rAAV vectors can significantly expand the clinical application of rAAV-mediated gene therapy. Thus, in one embodiment the polypeptide of interest is a polypeptide encoded by a transgene. To deliver transgenes that already exceed the AAV packaging capacity, the *trans*-splicing approach was developed. Briefly, two separate rAAV vectors deliver two parts of a transgene into a target cell, one part containing a splicing donor signal at the 3' end of the 5' part of the transgene and the other part harbouring a splicing acceptor signal at the 5' end of the 3' part of the transgene. While in prior art systems the intermolecular recombination between the two vector genomes generates an intervening ITR junction or an intervening sequence containing a recombinogenic sequence, which is in a next step excised by the cellular *cis*-splicing mechanism from a single pre-mRNA molecule to form a complete full-length transgene cassette, in the AAV vector system according to the invention the two vector genomes are transcribed separately, the pre-mRNAs interact via their complementary binding domains and the two pre-mRNAs are spliced in *trans* to form a complete (full-length) transgene transcript.

[139] The efficiency of the vector systems can be assessed by measuring the expression of the (full-length) protein of interest or mRNA encoding the protein of interest that has been provided by the dual AAV vector approach, e.g., by measures and techniques as described herein or known to the skilled person.

[140] The second binding domain is complementary to the first binding domain of the first AAV vector. Absolute complementarity, although preferred, is not required. A sequence "complementary" to a portion of the first binding domain, as referred to herein, means a sequence having sufficient complementarity to be able to hybridize with the first binding domain, forming a stable duplex. The ability to hybridize will depend on both the degree of complementarity and the length of the nucleic acid (See, for example, Sambrook et ah, 1989, Molecular Cloning, A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.). Complementarity as used with regard to second binding domain thus means that the second binding domain comprises a sequences, which is at least 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 % 99 % or 100 % complementary to the first binding domain comprised in the first AAV, preferably at least 90 %, 95 %, 98 % 99 % or 100 % complementary to the to the first binding domain comprised in the first AAV.

[141] The first binding domain can have or comprise a sequence as depicted in SEQ ID

NO: 17, or a sequence having 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % or 100 % sequence identity to a sequence of SEQ ID 17. The second binding domain can have or comprise a sequence as depicted in SEQ ID NO: 18 or a sequence having 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % or 100 % sequence identity to a sequence of SEQ ID No. 18. The person skilled in the art will understand that the first and the second binding domain may be exchanged, as long as the second binding domain is complementary to the first binding domain. Thus, the first binding domain can have or comprise a sequence as depicted in SEQ ID NO. 18, or a sequence having 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % or 100 % sequence identity to a sequence of SEQ ID 17. The second binding domain can have or comprise a sequence as depicted in SEQ ID NO. 17 or a sequence having 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % or 100 % sequence identity to a sequence of SEQ ID No. 18.

[142] Alternatively, the second binding domain, has or comprises a sequence of any one of SEQ ID NOs: 18, 27, 28, 32 and 33, or a sequence having 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % or 100 % sequence identity to a sequence of any one of SEQ ID NOs: 18, 27, 28, 32 and 33 and the first binding domain has or comprises a sequence complementary to the second binding domain, preferably at least 80 %, 85 %, 90 %, 95 %, 98 % 99 % or 100 % complementary to the second binding domain, more preferably at least 90 %, 95 %, 98 % 99 % or 100 % complementary to the second binding domain, even more preferably 98 % 99 % or 100 % complementary to the second binding domain. The person skilled in the art will understand that the first and the second binding domain may be exchanged, as long as the binding domains are complementary to each other. Thus, the first binding domain can have or comprise a sequence of any one of SEQ ID NOs: 18, 27, 28, 32 and 33, or a sequence having 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % or 100 % sequence identity to a sequence of any one of SEQ ID NOs: 18, 27, 28, 32 and 33 and the second binding domain has or comprises a sequence complementary to the first binding domain, preferably at least 80 %, 85 %, 90 %, 95 %, 98 % 99 % or 100 % complementary to the first binding domain, more preferably at least 90 %, 95 %, 98 % 99 % or 100 % complementary to the first binding domain, even more preferably 98 % 99 % or 100 % complementary to the first binding domain.

[143] It has further been shown that nucleotides 50 to 100 of SEQ ID NO: 27 or 28 are effective as a binding domain. Thus, in one embodiment the second or the first binding domain comprises a sequence of nucleotides 50 to 100 of SEQ ID NO: 27 or 28, or a sequence having 80 %, 85 %, 90 %, 95 %, 98 %, 99 % or 100 % sequence identity to a sequence of nucleotides 50 to 100 of SEQ ID NO: 27 or 28. Preferably the binding domain has at least 80 nucleotides comprising a sequence of nucleotides 50 to 100 of SEQ ID NO: 27 or 28, or a sequence having 80 %, 85 %, 90 %, 95 %, 98 %, 99 % or 100 % sequence

identity to a sequence of nucleotides 50 to 100 of SEQ ID NO: 27 or 28. The person skilled in the art will understand that the other binding domain has or comprises a sequence complementary to the first or the second binding domain, respectively, preferably at least 80 %, 85 %, 90 %, 95 %, 98 % 99 % or 100 % complementary to the first or the second binding domain, respectively, more preferably at least 90 %, 95 %, 98 % 99 % or 100 % complementary to the first or the second binding domain, respectively, even more preferably 98 % 99 % or 100 % complementary to the first or the second binding domain, respectively.

[144] In another embodiment the second or the first binding domain comprises a sequence of nucleotides 1 to 50 of SEQ ID NO: 18, 32 or 33, or a sequence having 80 %, 85 %, 90 %, 95 %, 98 %, 99 % or 100 % sequence identity to a sequence of nucleotides 1 to 50 of SEQ ID NO: 18, 32 or 33. Preferably the binding domain has at least 80 nucleotides comprising a sequence of nucleotides 1 to 50 of SEQ ID NO: 18, 32 or 33, or a sequence having 80 %, 85 %, 90 %, 95 %, 98 %, 99 % or 100 % sequence identity to a sequence of nucleotides 1 to 50 of SEQ ID NO: 18, 32 or 33. The person skilled in the art will understand that the other binding domain has or comprises a sequence complementary to the first or the second binding domain, respectively, preferably at least 80 %, 85 %, 90 %, 95 %, 98 % 99 % or 100 % complementary to the first or the second binding domain, respectively, more preferably at least 90 %, 95 %, 98 % 99 % or 100 % complementary to the first or the second binding domain, respectively, even more preferably 98 % 99 % or 100 % complementary to the first or the second binding domain, respectively.

[145] The binding domain may have between about 15 to 250 nucleotides, between about 15 to 200 nucleotides, between about 100 to 200 nucleotides or less than 500, 400, 300 or 200 nucleotides. In one embodiment the binding domain comprises 50-150 nucleotides, preferably 80-120 nucleotides, even more preferably 90 to 110 nucleotides or about 100 nucleotides. The binding domains as described herein may also be used in the pre-mRNA trans-splicing molecules or the methods according to the invention.

[146] The first and the second binding domain as used according to the invention may be derived from human sequences or non-human sequences, such as bacterial sequences. For human use, such as in therapy (particularly gene therapy), non-human sequences are preferred to avoid off-target effects in human cells.

[147] The second AAV vector and optionally the first AAV vector as described herein may comprise a termination signal, such as a polyadenylation (polyA) signal, located 3' to the nucleic acid sequence. PolyA signals/sequences are known to the skilled person and may be derived from many suitable species, including, without limitation SV-40, human and bovine. "PolyA" (A=adenylic acid) refers to a nucleic acid sequence comprising multiple adenosine monophosphates, such as a nucleic acid sequence that comprises the AAUAAA consensus sequence, which enables polyadenylation of a processed transcript. In a gene

disruption or selection cassette (GDSC), the polyA sequence is located downstream to the reporter and/or selectable marker gene and signals the end of the transcript to the RNA-polymerase. It is envisioned that the AAV vector can comprise a polyA sequence of SEQ ID 16 or a sequence having at least 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 % or 100 % sequence identity to a sequence depicted in SEQ ID No. 16.

[148] The AAV vectors or other vectors of the present invention may further optionally include one or more transcription termination sequences, one or more translation termination sequences, one or more signal peptide sequences, one or more internal ribosome entry sites (IRES), and/or one or more enhancer elements, or any combination thereof. Transcription termination regions can typically be obtained from the 3′ untranslated region of a eukaryotic or viral gene sequence. Transcription termination sequences can be positioned downstream of a coding sequence to provide for efficient termination. Signal peptide sequences are amino-terminal peptidic sequences that encode information responsible for the location of an operably-linked polypeptide to one or more post-translational cellular destinations, including, for example, specific organelle compartments, or to the sites of protein synthesis and/or activity, and even to the extracellular environment.

[149] Enhancers—cis-acting regulatory elements that increase gene transcription—may also be included in one of the disclosed AAV vectors or vectors. A variety of enhancer elements are known to those of ordinary skill in the relevant arts, and include, without limitation, a CaMV 35S enhancer element, a cytomegalovirus (CMV) early promoter enhancer element, an SV40 enhancer element, as well as combinations and/or derivatives thereof. One or more nucleic acid sequences that direct or regulate polyadenylation of the mRNA encoded by a structural gene of interest, may also be optionally included in one or more of the vectors of the present invention.

[150] The disclosed dual-vector systems may be introduced into one or more selected mammalian cells using any one or more of the methods that are known to those of ordinary skill in the gene therapy and/or viral arts. Such methods include, without limitation, transfection, microinjection, electroporation, lipofection, cell fusion, and calcium phosphate precipitation, as well as biolistic methods. In one embodiment, the vectors of the invention may be introduced in vivo, including, for example, by lipofection (i.e., DNA transfection via liposomes prepared from one or more cationic lipids). Synthetic cationic lipids (LIPOFECTIN, Invitrogen Corp., La Jolla, Calif., USA) may be used to prepare liposomes that will encapsulate the vectors to facilitate their introduction into one or more selected cells. A vector system of the invention can also be introduced in vivo as “naked” DNA using methods known to those of ordinary skill in the art.

[151] The present invention also relates to a kit comprising the nucleic acid sequence and/or the AAV vector and/or the AAV vector system of the present invention.

[152] The present invention also relates to a method for producing a nucleic acid sequence of interest, the method comprising

(A) contacting a nucleic acid sequence with a host cell, wherein the nucleic acid sequence comprises

(i) one or more donor splice site sequences;

(ii) an acceptor splice region sequence comprising

(a) a pyrimidine tract, wherein the pyrimidine tract comprises

(aa) 5 to 25 nucleotides;

(ab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U);

(b) a acceptor splice site,

ba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and

bb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G;

(iii) a nucleotide sequence of interest or a portion therof, wherein the nucleotide sequence of interest or the portion therof;

(a) is located 3' to the donor splice site and 5' to the acceptor splice region; and

(B) cleaving the nucleic acid sequence at the one or more donor splice site sequence(s) of (i) and in the acceptor splice region sequence of (ii), thereby separating the nucleotide sequence of interest or the portion thereof from the donor splice site and the acceptor splice region.

[153] As it is known to the skilled person splicing can occur in different manners. Known to the skilled person are for example, the *cis* splicing and the *trans*-splicing. *Cis*-splicing is a process in which intron sequences are excised from RNA transcripts in *cis* (mRNA or other RNAs). This process in a host cell takes place in the nucleus. *Cis*-splicing is known to the skilled person and inter alia described in Alberts B, Johnson A, Lewis J, et al. (2002) "Molecular Biology of the Cell. 4th edition." New York: Garland Science under the headline "from DNA to RNA").

[154] Generally therapy or treatment or prevention of a disease comprises administering to a mammalian subject in need thereof, an effective amount of a composition comprising an AAV vector, an AAV vector system, vector or nucleic acid molecule such as pre-mRNA *trans*-splicing molecule described herein, e.g. carrying a nucleic acid sequence encoding transgene/nucleotide sequence of interest, or fragment thereof, under the control of regulatory sequences which express the product of the gene in the subject's target cells

such as ocular cells, and optionally further a pharmaceutically acceptable carrier.

[155] The present invention therefore also relates to a pharmaceutical composition comprising an AAV vector, AAV vector system, vector or nucleic acid sequence (such as a pre-mRNA *trans*-splicing molecule) as described herein. Such pharmaceutical compositions can further comprise a carrier, preferably a pharmaceutically acceptable carrier.

[156] The pharmaceutical composition can be in the form of an injectable solution. The injectable solutions or suspensions may be formulated according to known art, using suitable non-toxic, pharmaceutically acceptable diluents or solvents, such as mannitol, 1,3-butanediol, water, Ringer's solution or isotonic sodium chloride solution, or suitable dispersing or wetting and suspending agents, such as sterile, bland, fixed oils, including synthetic mono- or diglycerides, and fatty acids, including oleic acid.

[157] For injectable formulations, the pharmaceutical compositions can be in lyophilized powder in admixture with suitable excipients in a suitable vial or tube. Before use in the clinic, the drugs may be reconstituted by dissolving the lyophilized powder in a suitable solvent system to form a composition suitable for intravenous or intramuscular injection or for subretinal, intravitreal or subconjunctival injection.

[158] It is also envisaged that the pharmaceutical composition of the present invention is formulated/administered as eye drops.

[159] The AAV vector, AAV vector system, vector, or nucleic acid sequence (such as a pre-mRNA *trans*-splicing molecule) of the present invention or the pharmaceutical composition of the present invention can be administered in a therapeutically effective amount. The “therapeutically effective amount” for the AAV vector, AAV vector system, nucleic acid sequence such as pre-mRNA *trans*-splicing molecule or vector can vary with factors including but not limited to stability of the active compound in the patient's body, the severity of the conditions to be alleviated, the total weight of the patient treated, the route of administration, the ease of absorption, distribution, and excretion of the active compound by the body, the age and sensitivity of the patient to be treated, adverse events, and the like, as will be apparent to a skilled artisan. The amount of administration can be adjusted as the various factors change over time.

[160] In principle the AAV vector, AAV vector system, vector, or nucleic acid sequence such as pre-mRNA *trans*-splicing molecule of the present invention or the pharmaceutical composition of the present invention can be administered in any suitable way. The AAV vector, AAV vector system, vector, nucleic acid sequence or pharmaceutical composition of the present invention e.g. comprising the desired transgene (i.e., nucleotide sequence of interest) for use in targeting photoreceptor cells can be formulated into a pharmaceutical composition intended for subretinal or intravitreal injection. Other forms of administration that may be useful in the methods described herein include, but are not limited to, direct delivery

to a desired organ (e.g., the eye e.g. eye drops), oral, inhalation, intranasal, intratracheal, intravenous, intramuscular, subcutaneous, intradermal, and other parental routes of administration. Routes of administration may be combined, if desired.

[161] Furthermore, it may be desirable to perform non-invasive retinal imaging and functional studies to identify areas of specific ocular cells to be targeted for therapy.

[162] The AAV vector, AAV vector system, vector, or nucleic acid sequence such as pre-mRNA *trans*-splicing molecule as described herein or the pharmaceutical composition of the present invention may be administered in a physiologically acceptable carrier to a subject, as described herein. The concentration of AAV in the pharmaceutical composition or upon administration can be between 10E8 and 10E12 total vector genomes per µl, preferably between 6x10E8 and 6x0E10 per µl. The AAV vectors can also be administered in a concentration of about 10E9 vector genomes per µl.

[163] The AAV vector, AAV vector system, vector, or nucleic acid sequence (such as a pre-mRNA *trans*-splicing molecule) or the pharmaceutical composition of the present invention may be administered alone or in combination with other treatments. Thus, also the pharmaceutical composition may additionally or alternatively comprise one or more further active ingredients.

[164] The pharmaceutical composition, AAV vector, AAV vector system, vector, pre-mRNA *trans*-splicing molecule, or nucleic acid sequence for the use of the invention can be administered to a subject. The AAV vector, AAV vector system, vector, pre-mRNA *trans*-splicing molecule, nucleic acid sequence or pharmaceutical compositions as described herein are applicable for both human therapy and veterinary applications, preferably for treating ocular disorder, particularly for gene therapy of an ocular disorder. Examples for suitable ocular disorders without being limited thereto are autosomal recessive severe early-onset retinal degeneration (Leber' s Congenital Amaurosis), congenital achromatopsia, Stargardt's disease, Best disease (vitelliform macular degeneration), Doyne's disease, retinitis pigmentosa (particularly autosomal dominant, autosomal-recessive, X-linked, di- or polygenic retinitis pigmentosa), (X-linked) retinoschisis, macular degeneration (AMD), age-related macular degeneration, atrophic age-related macular degeneration, neovascular AMD, diabetic maculopathy, proliferative diabetic retinopathy (PDR), cystoid macular oedema, central serous retinopathy, retinal detachment, intra-ocular inflammation, glaucoma, posterior uveitis, congenital stationary night blindness, choroideremia, early-onset retinal dystrophy, cone, rod-cone or cone-rod dystrophy, pattern dystrophies, Usher syndrome and other syndromic ciliopathies, such as Bardet-Biedl syndrome, Joubert syndrome, Senior-Løken syndrome or Alström syndrome.

[165] The subject can be a mammal or any other vertebrate. Examples of suitable mammals include, but are not limited to, a mouse, a rat, a cow, a goat, a sheep, a pig, a dog,

a cat, a horse, a guinea pig, a canine, a hamster, a mink, a seal, a whale, a camel, a chimpanzee, a rhesus monkey and a human, with human being preferred. Examples of other vertebrates include, but are not limited to, a zebra fish, a salamander, a turkey, a chicken, a goose, a duck, a teal, a mallard, a starling, a Northern pintail, a gull, a swan, a Guinea fowl or water fowl to name a few.

[166] The present invention also relates to an AAV vector, AAV vector system, vector, nucleic acid sequence (such as a pre-mRNA *trans*-splicing molecule) or pharmaceutical composition of the present invention for use in treating a photoreceptor cell disease. In these embodiments the AAV vector (e.g. rAAV) AAV vector, AAV vector system, vector, nucleic acid sequence (such as a pre-mRNA *trans*-splicing molecule) or pharmaceutical composition may comprise a nucleotide sequence of interest that is a heterologous nucleic acid encoding a therapeutic polypeptide or a portion thereof, a therapeutic nucleic acid or a portion thereof, a therapeutic protein/polypeptide or a portion thereof or a therapeutic molecule or a portion thereof.

[167] Similar to mRNA splicing, which evolved to remove non-coding RNA sequences, dispensable proteins sequences can be excised during a process known as protein splicing. In this context, protein fragments referred to as exteins (analog to exons) can be spliced together upon removal of the so-called inteins (analog to introns). In contrast to mRNA splicing which requires a complex splicing machinery, protein splicing is an autocatalytic chemical process. Protein splicing can also occur between two different proteins, a process known as protein *trans*-splicing. For this purpose, inteins are split into two parts and each of these parts is tagged to the protein to be fused. This approach leads to a scarless fusion of two proteins or polypeptides or parts thereof, without the need of additional factors. The split-intein technology may further be combined with mRNA trans-splicing to further enhance protein reconstitution efficiency. Thus, in one embodiment the first nucleic acid sequence or the first AAV vector further comprises a nucleic acid sequence that encodes the N-terminal part of an intein between the nucleotide sequence encoding the N-terminal portion of the polypeptide of interest (or the 5' portion of the nucleic acid sequence of interest) and the donor splice site and the second nucleic acid sequence or the second AAV vector further comprises a nucleic acid sequence that encodes the C-terminal part of an intein between the nucleotide sequence encoding the C-terminal portion of the polypeptide of interest (or the 3' portion of the nucleic acid sequence of interest) and the acceptor splice region.

[168] It is clear that all possible embodiments described for the inventive ASS can be used mutatis mutandis in the methods, nucleic acid sequences such as pre-mRNA *trans*-splicing molecules, kits, AAV vectors, AAV vector systems, and uses as described herein.

[169] The present invention is further characterized by the following items:

[170] 1. Use of a nucleic acid sequence comprising

(i) an acceptor splice region sequence comprising

(ia) a pyrimidine tract, wherein the pyrimidine tract comprises

(aa) 5 to 25 nucleotides;

(ab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U);

(ib) an acceptor splice site,

ba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and

bb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G; and

(ii) a nucleotide sequence of interest, wherein the nucleotide sequence of interest

(iia) is located 3' or 5' to the splice region;

for cleaving in the acceptor splice region thereby separating the nucleotide sequence of interest from the acceptor splice region sequence.

[171] 2. A nucleic acid sequence comprising

(i) an acceptor splice region sequence comprising

(ia) a pyrimidine tract, wherein the pyrimidine tract comprises

(aa) 5 to 25 nucleotides;

(ab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U);

(ib) an acceptor splice site,

ba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and

bb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G,

optionally for cleaving in the acceptor splice region thereby separating the nucleotide sequence of interest from the acceptor splice region sequence.

[172] 3. The use of a nucleic acid sequence of item one or the nucleic acid sequence of item 2, wherein the acceptor splice region further comprises a branch point nucleotide, preferably adenosine and/or an intronic splice enhancer.

[173] 4. The use or the nucleic acid sequence of any one of the preceding items, wherein the acceptor splice region further comprises a branch point nucleotide sequence (c), wherein the branch point nucleotide sequence

(ca) comprises 1-15 nucleotides;

(cb) comprises the branch point nucleotide, preferably adenosine (A); and

(cc) is located 5' to the pyrimidine tract and the acceptor splice site.

[174] 5. The use of a nucleic acid sequence or the nucleic acid sequence of any one of the preceding items, wherein the pyrimidine tract, the acceptor splice region and optionally further the branch point sequence and/or the intronic splice enhancer includes in total about 200, 150, 100, 50, 45, 40, 35, 30, 25, 20, 15 or less nucleotides, preferably 26 nucleotides.

[175] 6. The use of a nucleic acid sequence or the nucleic acid sequence of any one of the preceding items, wherein

(a) the 5 to 25 nucleotides of the pyrimidine tract comprise the sequence TTTTTT or TCTTTT;

(b) the sequence between the last pyrimidine of the pyrimidine tract and the acceptor splice site has less than 10, preferably less than 5, more preferably less than 3 bases; and/or

(c) the acceptor splice site has a sequence of CAGG.

[176] 7. The use of a nucleic acid sequence or the nucleic acid sequence of any one of the preceding items, wherein the acceptor splice region further comprises

(a) 7 nucleotides 5' to the pyrimidine tract having at least 4 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotide is C;

(b) 7 nucleotides 5' to the pyrimidine tract having at least 5 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotide is C;

(c) 7 nucleotides 5' to the pyrimidine tract having at least 6 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotide is C;

(d) 7 nucleotides 5' to the pyrimidine tract having at least 4 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAA;

(e) 7 nucleotides 5' to the pyrimidine tract having at least 4 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAAC;

(f) 7 nucleotides 5' to the pyrimidine tract having at least 5 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAA;

(g) 7 nucleotides 5' to the pyrimidine tract having at least 5 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAAC;

(h) 7 nucleotides 5' to the pyrimidine tract having at least 6 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAA; or

(i) 7 nucleotides 5' to the pyrimidine tract having at least 6 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAAC.

[177] 8. The use of a nucleic acid sequence or the nucleic acid sequence of any one of the preceding items, wherein the acceptor splice region has the sequence of SEQ ID NO: 3 or 4.

[178] 8. The use of a nucleic acid sequence or the nucleic acid sequence of any one of the preceding items, wherein the nucleic acid sequence further comprises a donor splice site.

[179] 10. The use of a nucleic acid sequence or the nucleic acid sequence of any one of the preceding items, wherein the nucleic acid sequence further comprises a promoter, preferably wherein the nucleic acid sequence is a DNA sequence and further comprises a promoter.

[180] 11. The use of a nucleic acid sequence or the nucleic acid sequence of any of the preceding items wherein the nucleic acid sequence comprises

(i) a acceptor splice region, comprising

(iia) a pyrimidine tract, wherein the pyrimidine tract comprises

(iiaa) 5 to 25 nucleotides;

(iiab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U);

(iib) a acceptor splice site,

(iiba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and

(iibb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G;

(ii) a nucleotide sequence of interest; wherein the acceptor splice region is localized 3' or 5' to the nucleotide sequence of interest;

(iii) a binding domain targeting pre-mRNA, which is localized 3' or 5' to the nucleotide sequence of interest; and

(iv) optionally a spacer sequence, wherein the spacer sequence is localized between the binding domain and the acceptor splice region.

[181] 12. The use of a nucleic acid sequence or the nucleic acid sequence of any of the preceding items wherein the nucleic acid sequence comprises

(i) an acceptor splice region comprising

(ia) a pyrimidine tract, wherein the pyrimidine tract comprises

(iiiaa) 5 to 25 nucleotides;

(iiiab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U);

(ib) an acceptor splice site,

(iba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and

(ibb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G;

(ii) a nucleotide sequence of interest, wherein the acceptor splice region is localized 5' to the nucleotide sequence of interest;

(iii) a donor splice site, wherein the donor splice site is localized 3' to the nucleotide sequence of interest;

(iv) a first binding domain targeting pre-mRNA located 5' to the nucleotide sequence of interest;

(v) a second binding domain targeting pre-mRNA located 3' to the nucleotide sequence of interest;

(vi) optionally a first spacer sequence, wherein the first spacer is localized between the first binding domain and the acceptor splice region; and

(vii) optionally a second spacer sequence, wherein the second spacer is localized between the second binding domain and the donor splice site.

[182] 13. The use of a nucleic acid sequence or the nucleic acid sequence of any of the preceding items wherein the nucleic acid sequence comprises an adeno-associated virus (AAV) vector, comprising at least two inverted terminal repeats comprising a nucleic acid sequence between these two inverted terminal repeats, wherein said nucleic acid sequence comprises from 5' to 3'

(i) a promoter;

(ii) optionally a binding domain;

(iii) an acceptor splice region sequence comprising

(a) a pyrimidine tract, wherein the pyrimidine tract comprises

(aa) 5 to 25 nucleotides;

(ab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U);

(b) a acceptor splice site,

ba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and

bb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G;

(iv) a nucleotide sequence of interest

(v) optionally a termination sequence, preferably a polyA sequence.

[183] 14. The use of a nucleic acid sequence or the nucleic acid sequence of any of the preceding items wherein the acceptor splice region is localized within a vector, within an AAV vector and or within a pre-mRNA *trans*-splicing molecule.

[184] 15. The use of a nucleic acid sequence or the nucleic acid sequence of any of the preceding items, wherein the acceptor splice region is used for trans-splicing.

[185] 16. A method for producing a nucleic acid sequence, the method comprising

(A) providing a first nucleic acid sequence comprising one or more donor splice site sequences;

(B) providing a second nucleic acid sequence, wherein the second nucleic acid sequence comprises

(i) an acceptor splice region sequence comprising

(ia) a pyrimidine tract, wherein the pyrimidine tract comprises

(iaa) 5 to 25 nucleotides;

(iab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U);

(ib) a acceptor splice site,

(iba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and

(ibb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G;

(ii) a nucleotide sequence of interest, wherein the nucleotide sequence of interest

(iia) is located 3' to the acceptor splice region; and

(C) obtaining a nucleic acid sequence.

[186] 17. The method of item 16, wherein the method further comprises the step of

(D) cleaving the first nucleic acid sequence at the one or more donor splice site sequence(s) and cleaving the second nucleic acid sequence in the acceptor splice site,

(E) ligating the first cleaved nucleic acid sequence to the second cleaved nucleic acid sequence,

thereby obtaining a nucleic acid sequence.

[187] 18. The method of item 16 or 17, wherein the first nucleic acid sequence further comprises a nucleotide sequence of interest or a portion thereof, wherein at least a portion of the nucleotide sequence of interest is located 5' to the donor splice site and wherein the second nucleic acid sequence further comprises a nucleotide sequence of interest or a portion thereof, wherein at least a portion of the nucleotide sequence of interest is located 3' to the splice acceptor splice region.

[188] 19. The method of any one of items 16 to 18, wherein the first and the second nucleic acid sequence are introduced into a host cell, preferably wherein the first and the second nucleic acid sequence are a recombinant nucleic acid sequence.

[189] 20. A method for producing a nucleic acid sequence, the method comprising

(A) providing a first nucleic acid sequence comprising one or more donor splice site sequences;

(B) providing a second nucleic acid sequence, wherein the second nucleic acid sequence comprises

(i) an acceptor splice region sequence comprising

(ia) a pyrimidine tract, wherein the pyrimidine tract comprises

(iaa) 5 to 25 nucleotides;

(iab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U);

(ib) a acceptor splice site,

(iba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and

(ibb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G; and

(C) cleaving the first nucleic acid sequence at the one or more donor splice site sequence(s) and cleaving the second nucleic acid sequence in the acceptor splice site,

(D) ligating the first cleaved nucleic acid sequence to the second cleaved nucleic acid sequence,

thereby obtaining a nucleic acid sequence.

[190] 21. The method of any one of items 16 to 20, wherein the first nucleic acid sequence further comprises a nucleotide sequence of interest or a portion thereof, wherein at least a portion of the nucleotide sequence of interest is located 5' to the donor splice site and wherein the second nucleic acid sequence further comprises a nucleotide sequence of interest or a portion thereof, wherein at least a portion of the nucleotide sequence of interest is located 3' to the splice acceptor splice region.

[191] 22. A method for producing a nucleic acid or the method of any one of items 16 to 21, comprising

(A) introducing into a host cell a first nucleic acid sequence comprising a pre-mRNA trans-splicing molecule sequence or a nucleic acid sequence encoding said pre-mRNA *trans*-splicing molecule, wherein the first pre-mRNA *trans*-splicing molecule comprises from 5' to 3'

(a) a 5' portion of a nucleotide acid sequence of interest;

(b) a donor splice site;

(c) optionally a spacer sequence;

(d) a first binding domain; and

(e) optionally a termination sequence, preferably a poly A sequence and

(B) introducing into a host cell a second nucleic acid sequence comprising the pre-mRNA *trans*-splicing molecule sequence or a nucleic acid sequence encoding said pre-mRNA *trans*-splicing molecule, wherein the second pre-mRNA *trans*-splicing molecule comprises from 5' to 3';

(i) a second binding domain, which is complementary to the first target domain of the first nucleic acid sequence;

(ii) an acceptor splice region sequence comprising;

(iia) a pyrimidine tract, wherein the pyrimidine tract comprises

(iiaa) 5 to 25 nucleotides;

(iiab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U);

(iib) an acceptor splice site,

(iiba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and

(iibb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G; and

(iii) a 3' portion of the nucleotide sequence of interest and

thereby obtaining the nucleic acid sequence of interest, optionally further comprising

(C) cleaving the first nucleic acid sequence at the donor splice site sequence and cleaving the second nucleic acid sequence in the acceptor splice site; and

(D) ligating the first cleaved nucleic acid sequence comprising the 5 portion of the nucleotide sequence of interest to the second cleaved nucleic acid sequence comprising the 3' portion of the nucleotide sequence of interest, thereby obtaining the nucleic acid sequence of interest.

[192] 23. The method of any one of items 16 to 22, wherein

(a) the 5 to 25 nucleotides of the pyrimidine tract comprise the sequence TTTTTT or TCTTTT;

(b) the sequence between the last pyrimidine of the pyrimidine tract and the acceptor splice site has less than 10, preferably less than 5, more preferably less than 3 bases; and/or

(c) the acceptor splice site has a sequence of CAGG.

[193] 24. The method of any one of items 16 to 23, wherein the acceptor splice region further comprises

(a) 7 nucleotides 5' to the pyrimidine tract having at least 4 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotide is C;

(b) 7 nucleotides 5' to the pyrimidine tract having at least 5 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotide is C;

(c) 7 nucleotides 5' to the pyrimidine tract having at least 6 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotide is C;

(d) 7 nucleotides 5' to the pyrimidine tract having at least 4 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAA;

(e) 7 nucleotides 5' to the pyrimidine tract having at least 4 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAAC;

(f) 7 nucleotides 5' to the pyrimidine tract having at least 5 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAA;

(g) 7 nucleotides 5' to the pyrimidine tract having at least 5 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAAC;

(h) 7 nucleotides 5' to the pyrimidine tract having at least 6 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAA; or

(i) 7 nucleotides 5' to the pyrimidine tract having at least 6 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAAC.

[194] 25. A nucleic acid sequence, comprising

(i) an acceptor splice region sequence comprising

(ia) a pyrimidine tract, wherein the pyrimidine tract comprises

(iaa) 5 to 25 nucleotides;

(iab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U);

(ib) a acceptor splice site,

(iba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and

(ibb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G;

(ii) a nucleotide sequence of interest, wherein the nucleotide sequence of interest

(iia) is located 3' or 5' to the acceptor splice region.

[195] 26. A nucleic acid sequence, comprising

(i) an acceptor splice region sequence comprising

(ia) a pyrimidine tract, wherein the pyrimidine tract comprises

(iaa) 5 to 25 nucleotides;

(iab) wherein at least 60 % of the nucleotides within these 5 to 25

nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U);

(ib) a acceptor splice site,

(iba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and

(ibb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G;

(ii) a nucleotide sequence of interest, wherein the nucleotide sequence of interest

(iia) is located 3' or 5' to the acceptor splice region;

and optionally wherein the nucleotide sequence of interest is not the rhodopsin gene of SEQ ID NO: 10 and/or not exon 3 of the rhodopsin gene of SEQ ID NO: 9 and/or is not exon 3 of the rhodopsin gene of SEQ ID NO: 9 and/or wherein the nucleotide sequence of interest does not comprise a sequence as shown in SEQ ID No: 9 and/or 10.

[196] 27. The nucleic acid sequence of item 25 or 26, wherein the nucleic acid sequence has a length of at most 150 nucleotides.

[197] 28. The nucleic acid sequence of any one of items 25 to 27, wherein the nucleic acid sequence has a length of at most 5500 nucleotides.

[198] 29. The nucleic acid sequence of any one of items 25 to 28, wherein (a) the 5 to 25 nucleotides of the pyrimidine tract comprise the sequence TTTTTT or TCTTTT and/or (b) the sequence between the last pyrimidine of the pyrimidine tract and the acceptor splice site has less than 10, preferably less than 5, more preferably less than 3 bases.

[199] 30. An acceptor splice region sequence comprising

(ia) a pyrimidine tract, wherein the pyrimidine tract comprises

(iaa) 5 to 25 nucleotides;

(iab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U);

(ib) a acceptor splice site,

(iba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and

(ibb) wherein the acceptor splice site comprises from 5' to 3' a sequence of CAGG.

[200] 31. The nucleic acid or acceptor splice region sequence of any one of items 25 to 30, wherein

(a) the 5 to 25 nucleotides of the pyrimidine tract comprise the sequence TTTTTT or TCTTTT;

(b) the sequence between the last pyrimidine of the pyrimidine tract and the acceptor splice site has less than 10, preferably less than 5, more preferably less than 3 bases; and/or

(c) the acceptor splice site has a sequence of CAGG.

[201] 32. The nucleic acid or acceptor splice region sequence of any one of items 25 to 31, wherein the acceptor splice region further comprises

(a) 7 nucleotides 5' to the pyrimidine tract having at least 4 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotide is C;

(b) 7 nucleotides 5' to the pyrimidine tract having at least 5 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotide is C;

(c) 7 nucleotides 5' to the pyrimidine tract having at least 6 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotide is C;

(d) 7 nucleotides 5' to the pyrimidine tract having at least 4 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAA;

(e) 7 nucleotides 5' to the pyrimidine tract having at least 4 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAAC;

(f) 7 nucleotides 5' to the pyrimidine tract having at least 5 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAA;

(g) 7 nucleotides 5' to the pyrimidine tract having at least 5 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAAC;

(h) 7 nucleotides 5' to the pyrimidine tract having at least 6 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAA; or

(i) 7 nucleotides 5' to the pyrimidine tract having at least 6 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAAC.

[202] 33. A pre-mRNA trans-splicing molecule comprising,

(i) a acceptor splice region, comprising

(iia) a pyrimidine tract, wherein the pyrimidine tract comprises

(iiaa) 5 to 25 nucleotides;

(iiab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U);

(iib) a acceptor splice site,

(iiba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and

(iibb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G;

(ii) a nucleotide sequence of interest or a portion thereof; wherein the acceptor splice region is localized 3' or 5' to the nucleotide sequence of interest or the portion thereof;

(iii) a binding domain targeting pre-mRNA, which is localized 5' to the nucleic acid

sequence of interest or the portion thereof; and

(iv) optionally a spacer sequence, wherein the spacer sequence is localized between the binding domain and the acceptor splice region.

[203] 34. The pre-mRNA *trans*-splicing molecule of item 33, further comprising a donor splice site 3' to the nucleic acid molecule and to the acceptor splice region.

[204] 35. The pre-mRNA *trans*-splicing molecule of item 33 or 34 comprising,

(i) an acceptor splice region comprising

(ia) a pyrimidine tract, wherein the pyrimidine tract comprises

(iiiaa) 5 to 25 nucleotides;

(iiiab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U);

(ib) a acceptor splice site,

(iba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and

(ibb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G;

(ii) a nucleotide sequence of interest, wherein the acceptor splice region is localized 5' to the nucleotide sequence of interest;

(iii) a donor splice site, wherein the donor splice site is localized 3' to the nucleotide sequence of interest;

(iv) a first binding domain targeting pre-mRNA located 5' to the nucleotide sequence of interest;

(v) a second binding domain targeting pre-mRNA located 3' to the nucleotide sequence of interest;

(vi) optionally a first spacer sequence, wherein the first spacer is localized between the first binding domain and the acceptor splice region;

(vii) optionally a second spacer sequence, wherein the second spacer is localized between the second binding domain and the donor splice site.

[205] 36. The pre-mRNA *trans*-splicing molecule of any one of items 33 to 35, wherein the acceptor splice region is localized 5' to the nucleotide sequence of interest or the portion thereof and the binding domain is localized 5' to acceptor splice region.

[206] 37. The pre-mRNA *trans*-splicing molecule of any one of items 33 to 36 further comprising a termination sequence, preferably a polyA sequence.

[207] 38. The pre-mRNA *trans*-splicing molecule of any one of items 33 to 37, wherein (a) the 5 to 25 nucleotides of the pyrimidine tract comprise the sequence TTTTTT or TCTTTT;

(b) the sequence between the last pyrimidine of the pyrimidine tract and the acceptor splice site has less than 10, preferably less than 5, more preferably less than 3 bases; and/or

(c) the acceptor splice site has a sequence of CAGG.

[208] 39. The pre-mRNA *trans*-splicing molecule of any one of items 33 to 38, wherein the acceptor splice region further comprises

(a) 7 nucleotides 5' to the pyrimidine tract having at least 4 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotide is C;

(b) 7 nucleotides 5' to the pyrimidine tract having at least 5 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotide is C;

(c) 7 nucleotides 5' to the pyrimidine tract having at least 6 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotide is C;

(d) 7 nucleotides 5' to the pyrimidine tract having at least 4 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAA;

(e) 7 nucleotides 5' to the pyrimidine tract having at least 4 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAAC;

(f) 7 nucleotides 5' to the pyrimidine tract having at least 5 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAA;

(g) 7 nucleotides 5' to the pyrimidine tract having at least 5 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAAC;

(h) 7 nucleotides 5' to the pyrimidine tract having at least 6 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAA; or

(i) 7 nucleotides 5' to the pyrimidine tract having at least 6 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAAC.

[209] 40. A DNA molecule comprising a promoter and a sequence encoding the pre-mRNA *trans*-splicing molecule according to any one of items 33 to 39.

[210] 41. An adeno-associated virus (AAV) vector system comprising

(I) a first AAV vector comprising at least two inverted terminal repeats comprising a nucleic acid sequence between these two inverted terminal repeats, wherein said nucleic acid sequence between these two inverted terminal repeats comprises from 5' to 3'

(a) a promoter,

(b) a nucleotide sequence encoding an N-terminal portion of a polypeptide of interest;

(c) a donor splice site;

(d) optionally a spacer sequence;

(e) a first binding domain; and

(f) optionally a termination sequence, preferably a polyA sequence;

(II) a second AAV vector comprising a nucleic acid sequence comprising at least two inverted terminal repeats comprising a nucleic acid sequence between these two inverted terminal repeats, wherein said nucleic acid sequence between these two inverted terminal repeats comprises from 5' to 3'

(i) a promoter;

(ii) a second binding domain, which is complementary to the first binding domain of the first AAV vector;

(iii) an acceptor splice region sequence comprising

(a) a pyrimidine tract, wherein the pyrimidine tract comprises

(aa) 5 to 25 nucleotides;

(ab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U);

(b) an acceptor acceptor splice site,

(ba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and

(bb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G;

(iv) a nucleotide sequence of interest encoding a C-terminal portion of the polypeptide of interest;

(iva) wherein the C-terminal portion of the polypeptide of interest and the N-terminal portion of the polypeptide of interest reconstitute the polypeptide of interest; and

(v) a termination sequence, preferably poly A sequence.

[211] 42. The AAV vector system of item 41, wherein the polypeptide is a full-length polypeptide and the first AAV vector comprises an N-terminal portion of a full-length polypeptide of interest and the second AAV vector comprises a C-terminal portion of the full-length polypeptide of interest.

[212] 43. The AAV vector system of item 41 or 42, wherein the C-terminal portion of the polypeptide of interest, corresponds to the portion of the polypeptide of interest, optionally the full-length protein of interest that is missing from the N-terminal portion of the polypeptide comprised in the first AAV vector.

[213] 44. The AAV vector system of any one of items 41 to 43, wherein the first and the second AAV vector comprises a termination sequence, preferably a polyA sequence.

[214] 45. An adeno-associated virus (AAV) vector comprising at least two inverted terminal repeats comprising a nucleic acid sequence between these two inverted

terminal repeats, wherein said nucleic acid sequence comprises from 5' to 3'

(i) a promoter;

(ii) optionally a spacer sequence;

(iii) a binding domain;

(iv) an acceptor splice region sequence comprising

(a) a pyrimidine tract, wherein the pyrimidine tract comprises

(aa) 5 to 25 nucleotides;

(ab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U);

(b) a acceptor splice site,

ba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and

bb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G;

(iv) a nucleotide sequence of interest

(v) optionally a termination sequence, preferably a poly A sequence.

[215] 46. The AAV vector or the AAV vector system of any one of items 41 to 45, wherein

(a) the 5 to 25 nucleotides of the pyrimidine tract comprises the sequence TTTTTT or TCTTTT;

(b) the sequence between the last pyrimidine of the pyrimidine tract and the acceptor splice site has less than 10, preferably less than 5, more preferably less than 3 bases; and/or

(c) the acceptor splice site has a sequence of CAGG.

[216] 47. The AAV vector or the AAV vector system of any one of items 41 to 46, wherein the acceptor splice region further comprises

(a) 7 nucleotides 5' to the pyrimidine tract having at least 4 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotide is C;

(b) 7 nucleotides 5' to the pyrimidine tract having at least 5 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotide is C;

(c) 7 nucleotides 5' to the pyrimidine tract having at least 6 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotide is C;

(d) 7 nucleotides 5' to the pyrimidine tract having at least 4 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAA;

(e) 7 nucleotides 5' to the pyrimidine tract having at least 4 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAAC;

(f) 7 nucleotides 5' to the pyrimidine tract having at least 5 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAA;

(g) 7 nucleotides 5' to the pyrimidine tract having at least 5 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAAC;

(h) 7 nucleotides 5' to the pyrimidine tract having at least 6 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAA; or

(i) 7 nucleotides 5' to the pyrimidine tract having at least 6 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAAC.

[217] 48. An acceptor splice region, comprising

(ia) a pyrimidine tract, wherein the pyrimidine tract comprises

(iaa) 5 to 25 nucleotides;

(iab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U);

(ib) a acceptor splice site,

(iba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and

(ibb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G.

[218] 49. The acceptor splice region of any one of the precedings for use in a vector, an AAV vector or within a pre-mRNA *trans*-splicing molecule according to the invention.

[219] 50. The acceptor splice region sequence of any one of items 48 or 49, wherein

(a) the 5 to 25 nucleotides of the pyrimidine tract comprise the sequence TTTTTT or TCTTTT;

(b) the sequence between the past pyrimidine of the pyrimidine tract and the acceptor splice site has less than 10, preferably less than 5, more preferably less than 3 bases; and/or

(c) the acceptor splice site has a sequence of CAGG.

[220] 51. The acceptor splice region sequence of any one of items 48 to 50, wherein the acceptor splice region further comprises

(a) 7 nucleotides 5' to the pyrimidine tract having at least 4 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotide is C;

(b) 7 nucleotides 5' to the pyrimidine tract having at least 5 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotide is C;

(c) 7 nucleotides 5' to the pyrimidine tract having at least 6 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotide is C;

(d) 7 nucleotides 5' to the pyrimidine tract having at least 4 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAA;

(e) 7 nucleotides 5' to the pyrimidine tract having at least 4 nucleotides of the

sequence CAACGAG, wherein the first 5' nucleotides are CAAC;

(f) 7 nucleotides 5' to the pyrimidine tract having at least 5 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAA;

(g) 7 nucleotides 5' to the pyrimidine tract having at least 5 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAAC;

(h) 7 nucleotides 5' to the pyrimidine tract having at least 6 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAA; or

(i) 7 nucleotides 5' to the pyrimidine tract having at least 6 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAAC.

[221] 52. A kit comprising the acceptor splice region, the AAV vector, the AAV vector system, the vector, the DNA, the nucleic acid sequence or the pre-mRNA *trans*-splicing molecule according to the invention.

[222] 53. A method for producing a nucleic acid sequence of interest, the method comprising

(A) providing a nucleic acid sequence to a host cell, wherein the nucleic acid sequence comprises

(i) one or more donor splice site sequences;

(ii) an acceptor splice region sequence comprising

(a) a pyrimidine tract, wherein the pyrimidine tract comprises

(aa) 5 to 25 nucleotides;

(ab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U);

(b) a acceptor splice site,

ba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and

bb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G;

(iii) a nucleotide sequence of interest, wherein the nucleotide sequence of interest;

(a) is located 3' to the donor splice site and 5' to the acceptor splice region; and

(B) cleaving the nucleic acid sequence at the one or more donor splice site sequence(s) of (i) and between the G and G of the splice acceptor sequence (site) of NAGG, thereby separating the nucleotide sequence of interest from the donor splice site and the acceptor splice region.

[223] 54. A method for producing/cleaving a nucleic acid sequence of interest, the method comprising

(a) providing a nucleic acid sequence comprising

(i) an acceptor splice region sequence comprising

(ia) a pyrimidine tract, wherein the pyrimidine tract comprises

(aa) 5 to 25 nucleotides;

(ab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U);

(ib) a acceptor splice site,

ba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and

bb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G; and

(ii) a nucleotide sequence of interest, wherein the nucleotide sequence of interest

(iia) is located 3' or 5' to the splice region;

(b) cleaving the acceptor splice region thereby separating the nucleotide sequence of interest from the acceptor splice region sequence.

[224] 55. Pharmaceutical composition comprising the nucleic acid molecule, the pre-mRNA *trans*-splicing molecule, the AAV vector, the AAV vector system, the vector of any of the preceding items.

[225] 56. The nucleic acid molecule, the pre-mRNA *trans*-splicing molecule, the AAV vector, the AAV vector system, the vector, the DNA molecule or the pharmaceutical composition of any of the preceding items for use in therapy, preferably for use in treating an ocular disorder.

[226] 57. A method of treating a subject for a nucleic acid sequence of interest, the method comprising

(a) administering a (therapeutically effective amount of a) nucleic acid molecule, the pre-mRNA *trans*-splicing molecule, the AAV vector, the AAV vector system, the vector the DNA molecule or the pharmaceutical composition of any of the preceding items to a subject (in need thereof), wherein the subject preferably has an ocular disorder.

[227] 58. A nucleic acid molecule, pre-mRNA *trans*-splicing molecule, AAV vector, AAV vector system, vector, DNA molecule or pharmaceutical composition of any of the preceding items for the manufacture of a medicament.

[228] 59. The nucleic acid molecule, the pre-mRNA *trans*-splicing molecule, the AAV vector, the AAV vector system, the vector, the DNA molecule, the pharmaceutical composition for use or the method of any of the preceding items, wherein the ocular disorder is selected from the group consisting of autosomal recessive severe early-onset retinal degeneration (Leber' s Congenital Amaurosis), congenital achromatopsia, Stargardt's disease, Best disease (vitelliform macular degeneration), Doyne's disease, retinitis

pigmentosa (particularly autosomal dominant, autosomal-recessive, X-linked, di- or polygenic retinitis pigmentosa), (X-linked) retinoschisis, macular degeneration (AMD), age-related macular degeneration, atrophic age-related macular degeneration, neovascular AMD, diabetic maculopathy, proliferative diabetic retinopathy (PDR), cystoid macular oedema, central serous retinopathy, retinal detachment, intra-ocular inflammation, glaucoma, posterior uveitis, congenital stationary night blindness, choroideremia, early-onset retinal dystrophy, cone, rod-cone or cone-rod dystrophy, pattern dystrophies, Usher syndrome and other syndromic ciliopathies, such as Bardet-Biedl syndrome, Joubert syndrome, Senior-Løken syndrome or Alström syndrome.

[229] It must be noted that as used herein, the singular forms “a”, “an”, and “the”, include plural references unless the context clearly indicates otherwise. Thus, for example, reference to “a reagent” includes one or more of such different reagents and reference to “the method” includes reference to equivalent steps and methods known to those of ordinary skill in the art that could be modified or substituted for the methods described herein.

[230] Furthermore, the term "about," as used herein when referring to a measurable value such as an amount or the length of a polynucleotide or polypeptide sequence, dose, time, temperature, and the like, is meant to encompass variations of ± 20%, ± 10%, ± 5%, ± 1 %, ± 0 5%, or ± 0 1 % of the specified amount. Also as used herein, "and/or" refers to and encompasses any and all possible combinations of one or more of the associated listed items, as well as the lack of combinations when interpreted in the alternative (“or").

[231] Unless the context indicates otherwise, it is specifically intended that the various features of the invention described herein can be used in any combination.

[232] Moreover, the present invention also contemplates that in some embodiments of the invention, any feature or combination of features set forth herein can be excluded or omitted.

[233] All publications and patents cited in this disclosure are incorporated by reference in their entirety. To the extent the material incorporated by reference contradicts or is inconsistent with this specification, the specification will supersede any such material.

[234] Unless otherwise indicated, the term "at least" preceding a series of elements is to be understood to refer to every element in the series. Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the present invention.

[235] Throughout this specification and the claims which follow, unless the context requires otherwise, the word “comprise”, and variations such as “comprises” and “comprising”, will be understood to imply the inclusion of a stated integer or step or group of integers or steps but not the exclusion of any other integer or step or group of integer or step. When used herein the term “comprising” can be substituted with the term “containing” or sometimes when used

herein with the term “having”. However, the term “comprising of” or equivalents as used herein encompass “consisting of” or “consisting essentially of” as defined below and hence may be replaced with the term “consisting of” or “consisting essentially of”.

[236] As used herein “consisting of" excludes any element, step, or ingredient not specified in the claim element. As used herein, "consisting essentially of" does not exclude materials or steps that do not materially affect the basic and novel characteristics of the claim.

[237] Sequences as used in the present invention:

[238] The following sequences having more than 10 nucleotides have been referred to herein:

Table 1:

Seq	Name	Sequence
1	Pyrimidine tract	*TCTTTTGTCATCT*
2	Intronic splice enhancer	TGGGGGGAGG
3	vgASS_620 (RHOE3d in Fg.2) (ASS region)	CAACGAG*TCTTTTGTCATCTA***CAGG**T
4	vgASS_620 mod (ASS region)	CAACGAG*TTTTTTGTCATCTA***CAGG**T
5	ASS region (RHOE3b in Fg.2)	CTACACGCTC**A**AGCCGGAGGTC**A**ACAACGAG *TCTTTTGTCATCTA***CAGG**T
6	ASS region (RHOE3c in Fg.2)	GCCGGAGGTC**A**ACAACGAG*TCTTTTGTCATCTA***CAGG**T
7	ASS region (RHOE3e in Fg.2)	G*TCTTTTGTCATCTA***CAGG**T
8	ASS region (RHOE3f in Fg.2)	G*TCTTTTGTCATCTA***CAGG**TGTTCGTGGTTCGTGGTCCA
9	Mutated exon 3 of rhodopsin gene (the c.620T>G is underlined)	GTACATCCCCGAGGGCCTGCAGTGCTCGTGTGGAATCGACTACTACA CGCTCAAGCCGGAGGTCAACAACGAGTCTTTTGTCATCTACAGGTTC GTGGTCCACTTCACCATCCCCATGATTATCATCTTTTTCTGCTATGGG CAGCTCGTCTTCACCGTCAAGGAG
10	Mutated rhodopsin gene GenBank: BC112104.1 Homo sapiens rhodopsin, mRNA (cDNA clone MGC:138309 IMAGE:8327572), complete cds (the stop and start codon as well as the	CCAGCTGGAGCCCTGAGTGGCTGAGCTCAGGCCTTCGCAGCATTCTT GGGTGGGAGCAGCCACGGGTCAGCCACAAGGGCCACAGCCATGAAT GGCACAGAAGGCCCTAACTTCTACGTGCCCTTCTCCAATGCGACGGG TGTGGTACGCAGCCCCTTCGAGTACCCACAGTACTACCTGGCTGAGC CATGGCAGTTCTCCATGCTGGCCGCCTACATGTTTCTGCTGATCGTG CTGGGCTTCCCCATCAACTTCCTCACGCTCTACGTCACCGTCCAGCA CAAGAAGCTGCGCACGCCTCTCAACTACATCCTGCTCAACCTAGCCG TGGCTGACCTCTTCATGGTCCTAGGTGGCTTCACCAGCACCCTCTAC ACCTCTCTGCATGGATACTTCGTCTTCGGGCCCACAGGATGCAATTTG GAGGGCTTCTTTGCCACCCTGGGCGGTGAAATTGCCCTGTGGTCCTT GGTGGTCCTGGCCATCGAGCGGTACGTGGTGGTGTGTAAGCCCATG AGCAACTTCCGCTTCGGGGAGAACCATGCCATCATGGGCGTTGCCTT CACCTGGGTCATGGCGCTGGCCTGCGCCGCACCCCCACTCGCCGGC TGGTCCAGGTACATCCCCGAGGGCCTGCAGTGCTCGTGTGGAATCG ACTACTACACGCTCAAGCCGGAGGTCAACAACGAGTCTTTTGTCATCT

	c.620T>G mutation are underlined)	ACAGGTTCGTGGTCCACTTCACCATCCCCATGATTATCATCTTTTTCT GCTATGGGCAGCTCGTCTTCACCGTCAAGGAGGCCGCTGCCCAGCA GCAGGAGTCAGCCACCACACAGAAGGCAGAGAAGGAGGTCACCCGC ATGGTCATCATCATGGTCATCGCTTTCCTGATCTGCTGGGTGCCCTAC GCCAGCGTGGCATTCTACATCTTCACCCACCAGGGCTCCAACTTCGG TCCCATCTTCATGACCATCCCAGCGTTCTTTGCCAAGAGCGCCGCCA TCTACAACCCTGTCATCTATATCATGATGAACAAGCAGTTCCGGAACT GCATGCTCACCACCATCTGCTGCGGCAAGAACCCACTGGGTGACGAT GAGGCCTCTGCTACCGTGTCCAAGACGGAGACGAGCCAGGTGGCCC CGGCCTAAGACCTGCCTAGGACTCTGTGGCCGACTATAGGCGTCTCC CATCCCCTACACCTTCCCCCAGCCACAGCCATCCCACCAG
11	Pyrimidine tract	*TTTTTTGTCATTT*
12	Pyrimidine tract	*TCTTTTGTCATCTA*
13	Pyrimidine tract + 7nt	CAACGAG*TCTTTTGTCATCTA*
14	Inverted repeat (5‘)	CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGG CGTCGGGCGACCTTTGGTCGCCCGGCCTCAGTGAGCGAGCGAGCGC GCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCCT
15	Inverted repeat (3‘)	AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGC TCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCCGGGCT TTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAG
16	Poly A sequence	GCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTA TTTGTAACCATTATAAGCTGCAATAAACAAGTT
17	First binding domain sequence	AGTGCATCAAGGCGATCACATCAGTGAAAAAAAGCCAGACAGGCGGT TAAACCAACGCAGATTAAACAGCAGGATGCAAAAATTCGCAGGTGGT CAGATG
18	Second binding domain sequence (BD_k)	CATCTGACCACCTGCGAATTTTTGCATCCTGCTGTTTAATCTGCGTTG GTTTAACCGCCTGTCTGGCTTTTTTTCACTGATGTGATCGCCTTGATG CACT
19	Dual AAV 5’ coding sequence of ABCA4 (CMV promoter)	Sequence as shown in Figure 5
20	Dual AAV 3’ coding sequence of ABCA4 (CMV promoter)	Sequence as shown in Figure 6
21	Dual AAV comprising the 5’ coding sequence of ABCA4 (ABCA4 promoter)	Sequence as shown in Figure 7
22	Dual AAV comprising the 3’ coding sequence of ABCA4 (ABCA4 promoter)	Sequence as shown in Figure 8
23	WT ASS sequence in Figure 2	ctcggcagccaccttggctgttcccaagtccctcacaggcagggtctccctacctgcctgtcctcaggta catccccgagggcctgcagtgctcgtgtggaatcgactactacacgctcaagccggaggtcaacaac gagtcttttgtcatctacatgttcgtggtccacttcaccatccccatgattatcatctttttctgctatgggcagc tcgtcttcaccgtcaaggaggtacggg
24	c.620T>G ASS sequence in	ctcggcagccaccttggctgttcccaagtccctcacaggcagggtctccctacctgcctgtcctcaggta catccccgagggcctgcagtgctcgtgtggaatcgactactacacgctcaagccggaggtcaacaac

	Figure 2 (the mutation is underlined)	gagtcttttgtcatctacaggttcgtggtccacttcaccatccccatgattatcatctttttctgctatgggcag ctcgtcttcaccgtcaaggaggtacggg
25	ASS region (RHOE3a in Fg.2)	ctacacgctcaagccggaggtcaacaacgagtcttttgtcatctacatgt
26	ASS region (RHOE3f in Fg.2)	gtcttttgtcatctacaggtgttcgtg
27	BD_g	GTTCCAGGGAGGGAATGTGAAGCCCCAGAAAGGGCCAGCGCCAGGT GGAATTCGCTAGCTCGGCAGCCACCTTGGCTGTTCCCAAGTCCCTCA CAGGCAGGG
28	BD_h+i	GTTAACCCTCCAGTCAGGACTCAAACCCAGTAGTGTCTGGTTCCAGG CACTGACCTGCTAGCTCGGCAGCCACCTTGGCTGTTCCCAAGTCCCT CACAGGCAGGG
29	BD sequence (PTM1, B2) obtained from *RHO* intron 1 (Berger et al., 2015)	CACCATTCATGGTGATAGCCGGGCTGCTGTTTGTGCAGGGCTGGCAC TGAACACTGCCTTGATCTTATTTGGAGCAATATGCGCTTGTCTAATTTC ACAGCAAGAAAACTGAGCTGAGGCTCAAAGAAGTCAAGCGCCCTGCT GGGGCG
30	ITR2 forward primer	GGAACCCCTAGTGATGGAGTT
31	ITR2 reverse primer	CGGCCTCAGTGAGCGA
32	BD_j	CATCTGACCACCAGCGAATTTTTGCATCCAGCTGTTTAATCAGCGTTG GTTTAACCGCCAGTCAGGCTTTCTTTCAAAGATGTGATCGCCTTGCAG CACT
33	BD_l	CCATCCCGCATCTGACCACCAGCGAAATGGATTTTTGCATCCAGCTG GGTAACGTTGGCAATTTAACCGCCAGTCAGGCTTTCTTTCAAAGATGT GGATT
34	BD_m	CCATCCCGCATCTGACCACCTGCGAAATGGATTTTTGCATCCTGCTG GGTAACGTTGGCAATTTAACCGCCTGTCTGGCTTTCTTTCAATGATGT GGATT
35	5' cerulean + BD_g	Sequence as shown in Figure 16
36	3' cerulean + BD_g	Sequence as shown in Figure 17
37	5' SpCas9-VPR + BD_k	Sequence as shown in Figure 18
38	3' SpCas9-VPR + BD_k	Sequence as shown in Figure 19
39	Dual AAV vector-5'ABCA4_w introns	Sequence as shown in Figure 20
40	Dual AAV vector-3'ABCA4_w introns	Sequence as shown in Figure 21

Table 1: Putative branch point sequences are underlined, the branch point is underlined twice and

depicted in bold letters, the pyrimidine tract sequence is listed in italic letters and the acceptor splice site is shown in bold letters. Underlined in SEQ ID NO: 9 and 10 is the mutation from 620T>G and the start and stop codons.

EXAMPLES

[239] The following examples illustrate the invention. These examples should not be construed as to limit the scope of this invention. The examples are included for purposes of illustration and the present invention is limited only by the claims.

Materials and methods

CELL CULTURE AND TRANSFECTION

[240] Human embryonic kidney 293 (HEK293) cells (DMSZ) were maintained in DMEM medium + GlutaMAX + 1 g/l glucose + pyruvate + 10 % FBS (Biochrom) + 1 % penicillin/streptomycin (Biochrom) in a CO_2 incubator (Heraeus, Thermo Fisher Scientific) at 37 °C and 10 % CO2. The HEK293 derived Lenti-X 293T (HEK293T) cells (Clontech, Takara) were cultured in DMEM medium + GlutaMAX + 4.5 g/l glucose + 10 % FBS + 1 % penicillin/streptomycin under the same conditions. Both cell lines were passaged two times per week at a confluence of approximately 90 %.

[241] Transient transfections were performed using the calcium phosphate method. For this purpose, cells were seeded onto 6 cm cell culture plates. When transfecting SpCas9-containing plasmids for western blotting, 10 cm cell culture plates were used. The cells were incubated overnight until they reached the desired confluence of approximately 70 %. The transfection mix components were added to a 15 ml Falcon tube in the indicated order. 2x BBS was added dropwise during vortexing.

[242] The transfection mix was incubated for 3 – 4 minutes at RT and added dropwise to the culture medium. For the initial experiments (Example 2, Figure 4), the cells have been incubated in a 5% CO_2 setting for 24 h and the cells were harvested without replacing the medium. For the optimized protocol (Examples 3, 5-8; Figure 9 and 11-14), the cells were incubated in a 5 % CO_2 setting for 3 – 4 h, the culture medium was replaced and the cells were maintained at 10 % CO_2 for approximately 48 h. When fluorophore-containing plasmids were transfected, successful transfection and expression was evaluated via the EVOS® FL cell imaging system (Life Technologies, Thermo Fisher Scientific).

[243] The 661W cells were kindly provided by Prof. Muayyad Al-Ubaidi (University of Houston). This cell line was cloned from murine retinal tumors and was found to exhibit molecular characteristics of cone photoreceptors (al-Ubaidi et al., 1992, Tan et al., 2004). 661W cells were maintained in DMEM medium + GlutaMAX + 1 g/l glucose + pyruvate + 10 % FBS (Biochrom) + 1 % Antibiotic-Antimycotic in a CO_2 incubator (Heraeus, Thermo Fisher

Scientific) at 37 °C and 10 % CO_2. They were passaged twice per week at a confluence of approximately 90 %. Transient transfections were performed using the calcium phosphate method as described above.

[244] Mouse embryonic fibroblasts (MEFs) were generated as described (Jat et al., 1986, Xu, 2005). The cells were maintained in DMEM medium + GlutaMAX + 1 g/l glucose + pyruvate + 10 % FBS (Biochrom) + 1 % penicillin/streptomycin (Biochrom) in a CO_2 incubator (Heraeus, Thermo Fisher Scientific) at 37 °C and 5 % CO_2. They were passaged once per week at a confluence of approximately 90 %. MEFs were transiently transfected using the TurboFectTM Transfection Reagent (Thermo Fisher Scientific). Cells were seeded onto a 6 cm cell culture plate and incubated until 70 – 90 % confluence was reached. The reaction mix was prepared in the following order.

[245] After addition of each component, the solution was mixed vigorously by vortexing. The transfection mix was incubated for 15 – 20 min at RT and subsequently added dropwise to the culture plate. The medium was exchanged after 3 h and the cells were harvested 48 h post transfection. When fluorophore-containing plasmids were transfected, successful transfection and expression was evaluated via the EVOS® FL cell imaging system (Life Technologies, Thermo Fisher Scientific).

PRODUCTION OF RECOMBINANT ADENO-ASSOCIATED VIRUSES

[246] Recombinant Adeno-associated viruses (rAAVs) were produced by a triple calcium phosphate transfection of a pAAV2.1 plasmid containing the gene of interest, a pAD. Helper plasmid and a plasmid encoding the desired capsid. For subretinal injections into the murine retina, the 2/8Y733F capsid variant was selected due to its high efficiency in transducing photoreceptors and the RPE (Petrs-Silva et al., 2009, Mol. Ther, 17, 463-71). HEK293T cells were seeded onto 15x 15 cm cell culture plates and incubated over night until a confluence of 60 – 80 % was reached. Prior to transfection, the FBS-containing medium was replaced by serum-free medium. The transfection reagents were added to a 50 ml Falcon tube in the indicated order: pAAV2.1 plasmid 270 µg, pAD Helper plasmid X µg, Capsid plasmid Y µg, H_2O ad 11.85 ml, Polybrene (8 mg/ml) 15 µl, Dextran (10 mg/ml) 1.5 ml, $CaCl_2$ (2.5 M) 1.5 ml, 2x BBS 15 ml. The required amounts of pAD Helper and capsid plasmid were calculated using the following formulas: X µg=270 µg ×MM of pAD HelperMM of pAAV2.1 Y µg= 270 µg ×MM of capsid plasmid MM of pAAV2.1. $CaCl_2$ and 2x BBS were added dropwise during vortexing. 2 ml of the transfection mix were added dropwise to each of the 15 culture plates. The plates were rocked gently and subsequently placed in a 5 % CO_2 setting for 24 h. The medium was exchanged afterwards and the plates were placed in a 10 % CO_2 setting for another 48 h.

[247] The virus-containing culture medium was harvested twice. The first harvest took place 72 h post transfection by collecting the entire medium of all plates and adding fresh medium. The second harvest took place after another 72 h incubation period. The culture medium was collected into a 500 ml centrifugation tube. The residual cells were removed from the medium by centrifugation at 4,000 rpm and 4 °C for 15 min (JA-10 rotor, J2-MC High speed centrifuge, Beckman Coulter) and filtration of the supernatant through a 0.45 µm PES filter unit (Nalgene, Thermo Fisher Scientific). A 40 % polyethylene glycol (PEG) solution was added to the flow-through to a final concentration of 8 % and kept at 4 °C overnight to precipitate the viral particles. Subsequently, the solution was centrifuged at 4,000 rpm and 4 °C for 15 min (JA-10 rotor, J2-MC High speed centrifuge, Beckman Coulter). The supernatant was discarded and the virus-containing pellet was stored at -20 °C until further processing.

[248] For Iodixanol Density Gradient Centrifugation, the pellet was re-suspended in 7.5 ml sterilely filtrated PBS and incubated with Benzonase® (VWR) at a final concentration of 50 U/ml in a water bath (Haake) at 37 °C for 30 min to remove residual, non-packaged DNA. Next, the virus suspension was pipetted into a Quick-Seal polypropylene tube (39 ml, Beckman Coulter) and a density gradient was established by adding solutions with different iodixanol concentrations below the virus suspension in the following order: 7 ml of 15 %, 6 ml of 25 %, 5 ml of 40 % and 6 ml of 60 % iodixanol solution. For this purpose, the MINIPULS 3 peristaltic pump (Gilson) and a long glass pipette were used. Subsequently, the tubes were sealed with the Beckman Tube Topper and centrifuged at 70,000 rpm and 18 °C for 1 h 45 min in an Optima L-80K ultracentrifuge (70 Ti rotor, Beckman Coulter). Afterwards, the top of the tube was perforated with a cannula to ensure air flow. The 40 % iodixanol phase enriched with viral particles was collected from the gradient using a 20 G cannula and 20 ml syringe by puncturing the tube laterally at the border between the 40 % and the 60 % phase. The virus-containing solution was stored at -80 °C until further processing.

[249] To further purify the virus, an anion exchange chromatography was performed employing the ÄKTAprime plus chromatography system (GE Healthcare), the 5 ml HiTrapTM Q FF anion exchange chromatography column (GE Healthcare) and the PrimeView 5.31 software (GE Healthcare). Before starting, the column was equilibrated with buffer A (20 mM Tris, 15 mM NaCl, pH 8.5) and the virus-containing solution was diluted with this buffer in a 1:1 ratio. The solution was loaded onto the column via a loop injector (50 ml Superloop, GE Healthcare). The UV light-absorbance and conductance properties of the collected fractions were monitored providing information about the amount of contained virus. Remaining bound molecules were removed from the column using a 2.5 M NaCl solution. All virus-containing fractions were pooled and used for subsequent processing. 3.5.4 Increase of rAAV Concentration

[250] To increase the virus concentration, Amicon® Ultra-4 centrifugation filter units (Merck) with a molecular weight cut-off of 100 kDa were used. The virus-containing solution was loaded on top of the filter unit and centrifuged at 4,000 rpm (JA-10 rotor, J2-MC High speed centrifuge, Beckman Coulter) and 4 °C in 20 min intervals until the volume was reduced to 500 µl. Subsequently, the filter unit was washed with 1 ml 0.014 % Tween/PBS-MK (10xPBS 50ml, 1M MgCl 500 µl, 2.5 M KCl 500 µl, add 500 ml water). The solution was further centrifuged under the same conditions until the volume was reduced to 100 µl of concentrated viral solution. 10 µl aliquots were prepared and stored at -80 °C until use.

[251] To determine the titer of the produced rAAVs, a qPCR was performed using the StepOnePlus Real-Time PCR system (Applied Biosystems, Thermo Fisher Scientific). A standard curve was produced to serve as a reference. For this purpose, a fragment containing a part of the ITR was amplified by PCR using the following primers.

ITR2 forward: 5' GGAACCCCTAGTGATGGAGTT 3' (SEQ ID NO: 30)

ITR2 reverse: 5' CGGCCTCAGTGAGCGA 3' (SEQ ID NO: 31)

[252] Next, the amplicon was purified and the concentration was determined using the NanodropTM 2000c spectrophotometer (Thermo Fisher Scientific). After preparation of a standard solution with concentration c, a dilution series ranging from 10^{10} to 10^{1} copies was made. To attain the standard curve, a qPCR was run with three technical replicates of the standard dilution series. For this purpose, a MicroAmp™ Fast Optical 96-Well Reaction Plate (Applied Biosystems, Thermo Fisher Scientific) and the PowerUpTM SYBRTM Green Master Mix (Thermo Fisher Scientific) were used. The virus solution was diluted 100-fold in H2O and run on the same reaction plate with three technical replicates. The reaction mix was prepared as follows. The obtained data was analyzed using the StepOnePlus Real-Time PCR system software (Applied Biosystems, Thermo Fisher Scientific). Baseline settings and cycling threshold position were adjusted manually if necessary. The standard curve was attained by plotting the resulting cycle threshold (Ct) values against the logarithm of the dilution. The number of viral genomes per µl (vg/µl) of the produced rAAVs could be inferred from the standard curve.

SUBRETINAL INJECTION

[253] For subretinal injections, C57Bl6/J mice at postnatal day 21 (P21) were anesthetized by an intraperitoneal injection of ketamine (40 mg/kg body weight) and xylazine (20 mg/kg body weight). After a complete absence of the paw withdrawal reflex the pupils were dilated by administration of atropine (1 %)- and tropicamide (0.5 %)-containing eye drops (Mydriaticum Stulln, Pharma Stulln GmbH). The fundus of the eye was focused using a surgical microscope (OPMI 1 FR pro, Zeiss). 1 µl containing 1010 rAAV particles was injected subretinally by a single injection with a NANOFIL 10 µl syringe (World Precision

Instruments) and a 34 G beveled needle (World Precision Instruments). The injected eyes were treated with an eye ointment containing 5 mg/g gentamicin and 0.3 mg/g dexamethasone. The mice were kept on a heating plate (Leica HI1120, Leica Biosystems) at 37 °C until full recovery from anesthesia. Two to four weeks post injection, all injected retinas were harvested and processed for RT-PCR analysis or immunohistochemistry.

IMMUNOHISTOCHEMISTRY

[254] For immunohistochemistry subretinally injected mice were euthanized via cervical dislocation. The eyes were removed and placed in 0.1 M phosphate buffer (PB). Subsequently, the eyeball was punctured at the ora serrata with a 21 G cannula and fixed in 4 % paraformaldehyde (PFA, Sigma Aldrich, pH adjusted to 7.4) for 5 min. Next, the eye was positioned below a stereomicroscope (Stemi 2000, Zeiss) on a filter paper drenched with 0.1 M PB. By cutting alongside the ora serrata using surgical scissors (SuperFine Vannas, World Precision Instruments), the cornea, lens and vitreous body were removed. The remaining part of the eyeball containing the retina was fixed in 4 % PFA for 45 min at RT and subsequently washed 3 times for 5 min in 0.1 M PB. For cryopreservation, the eyeball was placed in a 30 % sucrose solution (w/v) overnight at 4 °C.

[255] The next day, the eyeball was embedded in tissue freezing medium (Sakura) and chilled on dry ice until the medium solidified. The retina was sectioned using a cryostat (Leica CM3050 S, Leica Biosystems) into 10 µm thick slices, collected on coated glass object slides (Superfrost Plus microscopic slides, Thermo Fisher Scientific) and stored at -20 °C.

[256] For immunohistochemical staining, the retinal sections were thawed at RT and encircled using a Super PAP Pen Liquid Blocker (Science Services). Subsequently, the sections were rehydrated with 0.1 M PB for 5 min and fixed with 4 % PFA for 10 min. After washing the sections three times for 5 min each with 0.1 M phosphate buffer, pH 7.4 (PB), a solution containing the primary antibodies, 5 % ChemiBLOCKER (Merck) and 0.3 % Triton X-100 in 0.1 M PB was applied. The cryosections were incubated with the primary antibody solution overnight at 4 °C. The next day, the retinas were washed three times with 0.1 M PB for 5 min and incubated with a solution containing the secondary antibodies and 2 % ChemiBLOCKER in 0.1 M PB for 1.5 h at RT. After subsequent washing with 0.1 M PB three times for 5 min, the cell nuclei were stained with 5 µg/ml Hoechst 33342 solution (Invitrogen). Finally, the sections were washed with 0.1 M PB, embedded in Fluoromount-G Mounting Medium (Thermo Fisher Scientific), covered by a cover slip and stored at 4 °C.

CONFOCAL MICROSCOPY

[257] Images of the stained retinas were obtained using a Leica TCS SP8 inverted confocal laser scanning microscope (Leica Microsystems) equipped with a 405 nm diode as well as a 552 nm and a 633 nm optically pumped semiconductor laser appropriate for the excitation of Hoechst 33342, Cy3 and Cy5, respectively. The filter settings were chosen according to the emission spectrum of the respective dye. Images were acquired as z-stacks (1 µm steps) with a HC PL APO 40x/1.30 Oil CS2 objective (Leica Microsystems) and type F immersion liquid (Leica Microsystems) using the LAS X software (Leica Microsystems). With the same software, z-stacks were condensed into 2D images by applying a maximum intensity projection. The images were processed further with the ImageJ 1.48v software (National Institutes of Health).

[258] Images of transiently transfected living cells were obtained using the Leica TCS SP8 spectral confocal laser scanning microscope (Leica Microsystems) equipped with a 448 nm, a 514 nm and a 552 nm optically pumped semiconductor laser appropriate for the excitation of cerulean, citrine and mCherry, respectively. The filter settings were chosen according to the emission spectrum of the respective fluorophore. Images were acquired with a HCX APO 20x/1.00 W objective (Leica Microsystems). All images were processed with the ImageJ 1.48v software.

RNA EXTRACTION

[259] For RNA extraction from injected retinas, the mice were euthanized via cervical dislocation. The retinas were collected by placing blunt forceps below the eye, incising the eye ball using a sterile scalpel (Swann-Morton) and gradually moving the forceps upward. Three retinas were pooled per construct and RNA was extracted using the RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. For disruption, the tissue was homogenized by adding 350 µl of RLT buffer (Qiagen, provided with the kit) + 3.5 µl β-mercaptoethanol (β-ME, Sigma Aldrich) and passing it through a 20 G needle fitted onto a sterile syringe at least five times. The remaining steps were executed according to protocol. The RNA was eluted in 30 µl RNAse-free H2O.

[260] For RNA extraction from transiently transfected cells, the RNeasy Mini Kit Plus (Qiagen) was used. For this purpose, the medium was removed from the 6 cm culture plates and the cells were scrapped off using a 16 cm cell scraper (Sarstedt). The cells were collected in 500 µl medium into a 2 ml safe-lock tube (Eppendorf) and centrifuged at 3,000 x g and 4 °C for 10 min. The supernatant was discarded and the pellet was re-suspended in 600 µl RLT Plus buffer (Qiagen, provided with the kit) + 6 µl β-ME (Sigma Aldrich). A steel ball was placed into each tube and the cells were disrupted using the mixer mill MM400 (Retsch) at 30 Hz for 1 min. Afterwards, the balls were removed and the suspension was

centrifuged at 21,000 x g and RT for 5 min. The remaining steps were executed according to protocol including an optional step removing the genomic DNA via gDNA eliminator spin columns. The RNA was eluted in 30 µl RNAse-free H2O.

[261] The RNA concentrations were measured using the NanodropTM 2000c spectrophotometer (Thermo Fisher Scientific). The RNA was kept on ice until further use or stored at -20 °C for short-term or -80 °C for long-term storage.

cDNA SYNTHESIS

[262] For cDNA synthesis, the Revert Aid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific) was used according to the manufacturer's instructions. Equal amounts of RNA were used per experiment. The cDNA reaction mix was incubated in a Mastercycler® nexus gradient. The cDNA was kept on ice until further use or stored at -20 °C for short-term or -80 °C for long-term storage.

REVERSE TRANSCRIPTION PCR

[263] Reverse transcription PCR (RT-PCR) was performed using the Herculase II fusion DNA polymerase (Agilent Technologies) or the VWR Taq DNA polymerase (VWR).

BRANCH POINT ANALYSIS

[264] For Nested Lariat RT-PCR, RNA was extracted as described above. 10 µg RNA were incubated with RNase R (Lucigen) following the manufacturer's instructions to remove any non-circular RNA. Subsequent cDNA synthesis (Revert Aid First Strand cDNA Sythesis Kit, Life Technologies) was performed as described above. For this reaction, only the random hexamer primer was used. Next, the lariats were amplified via a nested RT-PCR using the Herculase II fusion DNA polymerase (Agilent Technologies). For the first round of amplification, the reaction mix was prepared as described in 3.12. For the second round of amplification, 5 µl of the first PCR were added to the reaction mix instead of the cDNA. Moreover, a second primer pair binding 25 – 30 bp downstream of the first primer pair was used. The applied cycling conditions are listed in Table 11.

[265] For TOPO Cloning and Analysis of Lariats, the products of the nested RT-PCR were subcloned into plasmids. For this purpose, 3'-adenine overhangs were added to the DNA fragments post amplification by incubating one unit of Taq Polymerase (VWR) with the PCR reaction for 10 min at 72 °C. Subsequently, the lariats were subcloned into a TOPO vector (TOPO TA Cloning Kit, Thermo Fisher Scientific) according to the manufacturer's instructions. The plasmids were transformed into bacteria and a small-scale plasmid preparation was performed. The resulting plasmids were sequenced (Eurofins Genomics)

and the obtained lariat sequences were analyzed by alignment with the investigated intron using the DNAMAN software (Lynnon Biosoft).

PROTEIN EXTRACTION

[266] For protein extraction, the medium was removed from the 6 or 10 cm culture plates. The cells were scrapped off using a 16 cm cell scraper (Sarstedt) and collected in 500 µl medium into a 2 ml safe-lock tube (Eppendorf). The cell suspension was centrifuged at 3,000 x g and 4 °C for 10 min, the supernatant was discarded and the pellet re-suspended in 150 µl and 250 µl Triton X-100 (TX) lysis buffer (2.5 ml Triton X-100, 15 ml 5 mM NaCl ml, 0.4 ml 2.5 M $CaCl_2$, add 500 ml water; c0mplete™ ULTRA protease Inhibitor Cocktail tablets (Roche) were added right before use (1 tablet/10 ml)) for 6 cm and 10 cm plates, respectively. A steel ball was added to each safe-lock tube and the cells were disrupted using the mixer mill MM400 (Retsch) at 30 Hz for 1 min. Afterwards, the tubes containing the balls were rotated end-over-end (VWRTM tube rotator) for 20 min at 4°C. Subsequently, the balls were removed and the lysates were centrifuged at 5,000 x g for 10 min at 4 °C. The protein-containing supernatant was transferred into a new reaction tube and stored at -20 °C.

[267] Total protein concentration was determined using the Bradford assay. 5 µl of the protein lysate was mixed with 95 µl of a 0.15 M NaCl solution and transferred into a PMMA standard disposable cuvette (BRAND). Subsequently, 1 ml of Coomassie blue solution was added, thoroughly mixed by pipetting and incubated at RT for 2 min. Using the BioPhotometer (Eppendorf), the absorption of the solution was measured against a blank control containing 5 µl of the TX lysis buffer. The obtained value represents the total amount of protein contained in 5 µl lysate.

Example 1: Identification of an optimized ASS module

[268] The impact of disease-linked rhodopsin mutations on mRNA splicing has been analyzed using human rhodopsin (*RHO*) minigenes in HEK293 cells and in transduced murine photoreceptors. Among those, we found one mutation in exon 3 of the *RHO* gene (c.620T>G) which creates a novel ASS (Fig. 1). The sequence of the ASS and the predicted ASS elements are shown in Figure 2A and B.

[269] *In silico* prediction showed that the acceptor splice site formed by the c.620T>G mutation (hereafter referred to as ASS_620) has similar splice scores as the native rhodopsin acceptor splice site of exon 3 (cf. Fig. 3A). This suggests that both splice sites could be used by the splicing machinery alternatively. However, experimental data showed that ASS_620 was exclusively used in both, HEK293 cells (Fig. 1D) and in murine photoreceptors (Fig. 1E). This suggests that ASS_620 is a strong acceptor splice site.

[270] For potential usage in biotechnological applications, the strength and functionality of an ASS should be independent of the gene environment. To test whether ASS_620 is functional in another non-native environment and to determine the ASS_620 elements required for most efficient splicing, sequences of variable lengths flanking the c.620T>G mutation were introduced into the exon 3 of the *RPS27* gene (Fig. 2C). The single sequences led to variable splicing efficiencies at the ASS_620 position (Fig. 2D). Highest splicing efficiencies (close to 100%) were obtained when using the *RHO*-E3d comprising *RPS27* minigene, i.e., the minigene comprising a 26 bp sequence (CAACGAG*TCTTTTGTCATCTA***CAGG**T; SEQ ID NO: 3) consisting of a 7 bp sequence at the 5'end, a polypyrimidine tract and the canonical ASS. Surprisingly, this sequence did not contain any of the predicted branch points present in *RHO* exon 3. The 7 bp sequence at the 5' end (CAACGAG) could harbor a currently not characterized effective branch point sequence, or, alternatively, an intronic splice enhancer recognition site required for efficient usage of the ASS. The 26 bp sequence (SEQ ID NO: 3) is referred to herein as "vgASS_620".

[271] To further confirm the splicing efficiency of vgASS_620, it was introduced into four additional randomly chosen genes (*HBQ1*, *S100A12*, *CLRN1 and CNGB1*) with variable predicted acceptor splice site strengths (Fig. 3C). Compared to vgASS_620, the predicted strength of the native exon acceptor splice site was higher in the case of *RPS27*, similar in the case of *S100A12*, *CLRN1* and *HBQ1*, and lower in the case of *CNGB1* (Fig. 3A). In subsequent RT-PCR experiments it was shown that vgASS_620 was not only exclusively used in *RPS27*, but also in all additional genes tested herein (Fig. 3D-E). These results demonstrate that vgASS_620 is a very strong acceptor splice site independent of the gene environment. It was therefore hypothesized that vgASS_620 can be utilized to improve the currently rather low trans-splicing efficiencies in SMaRT or dual adeno associated virus (AAV) vector hybrid technologies.

Example 2: Cerulean reporter assay for *in vitro* applications

[272] To test this hypothesis, a splice reporter assay has been generated by splitting the coding sequence of cerulean into two artificial exons interrupted by an artificial intron (Fig. 4D). This intron contained vgASS_620 and a strong donor splice site separated from each other by artificial 145 bp. Confocal imaging of transfected HEK293 cells revealed a strong and robust cerulean fluorescence when both splice sites were provided in *cis* (Figure 4F(c1)). RT-PCR experiments and sequencing of the corresponding bands confirmed that in this configuration both cerulean exons are spliced together efficiently (data not shown). As shown in Figure 4G, Western blotting experiments using a specific antibody against the N-terminal part of cerulean detected a specific immunosignal at the expected size (27 kDa).

Taken together, these findings suggest that the reporter assay is functional and very efficient in the *cis* configuration. In subsequent experiments, this construct was therefore used as a reference to determine the cerulean reconstitution efficiency when the two cerulean splice fragments were provided in *trans* (i.e. on separate plasmids). The DSS used in this experiment has a sequence of AAGGTAAG.

[273] Next, the impact of the binding domain and the acceptor splice site strength for the cerulean reconstitution efficiencies were analyzed using a reporter assay, wherein reconstitution of the reporter requires mRNA *trans*-splicing.

[274] For this purpose, a fluorescent reporter-based assay has been developed for which the coding sequence of the cyan fluorescent protein "cerulean" was again split into two halves at position 154 (Figure 4D). The first DNA construct comprising the 5' part under the control of a CMV promoter was further equipped with a strong DSS (AAGGTAAG) followed by a binding domain and the second DNA construct comprising the 3' part under the control of a CMV promoter was equipped with vgASS_620 as ASS following a binding domain complementary to the binding domain of the first DNA construct, as shown in Figure 4D (c3). After transcription, base-pairing of the BDs brings the splicing elements into close proximity, which facilitates splicing in *trans* and results in a full-length mature cerulean mRNA. Reconstitution of cerulean may be detected optically, e.g., by microscopy (as in Figure 4F) or flow cytometry, on mRNA level (e.g., using RT-PCR) or on protein level (e.g., using Western blot as in Figure 4G).

[275] As an exemplary template for optimization of the binding domain, intron 2 of the human rhodopsin gene was chosen for two reasons: (1) An optimized type of the *RHO* intron 2 binding domain can be used for a SMaRT-based replacement of the c.620T>G splice mutation, as this mutation is localized in exon 3 (Fig. 1). (2) The *RHO* intron 2 sequence is not homologous to any of the sequences in the mouse genome (data not shown) and is therefore not expected to cause off-target effects. Consequently, the optimized *RHO* intron 2 binding domain could also be utilized for reconstitution of other large human genes in mouse retinas using the dual AAV vector approach.

[276] HEK 293 cells were transiently co-transfected with the first and the second DNA construct as described above and shown exemplary for BD_h+i in Figure 4D(c3) and the presence of cerulean fluorescence was evaluated via confocal live cell imaging. A construct containing both cerulean parts in *cis* with an intervening artificial intron comprising the same splicing elements was used as a *cis*-splicing reference control (*cis*-ctrl). When the two halves were transfected separately, no fluorescence could be detected. When co-transfecting both constructs, fluorescent cells could be observed, indicating successful trans-splicing and reconstitution of the cerulean coding sequence.

[277] The *RHO* intron 2 binding domain was optimized by varying its size and position (Fig.

4A). Moreover, artificial BDs were created and tested by fusing sequences originating from the 5' and the 3' end of the intron as for BD_h+i.

[278] Following co-transfection of HEK293 cells mRNA expression of the two constructs was analyzed compared to the housekeeping gene *ALAS* using primers p1 and p2 for the 5' construct and primers p3 and p4 for the 3' construct, demonstrating only small differences in expression on pre-mRNA level (Figure 4B and C). Reconstitution efficiencies were determined by the radiometric analysis of the cerulean protein bands relative to the *cis*-ctrl and varied greatly using different BDs (Figure 4E). All protein bands were normalized to beta tubulin prior to quantification. Two binding domains BD_g (SEQ ID NO: 27) and BD_h+i (SEQ ID NO: 28), both about 100 bp in length resulted in a high cerulean reconstitution efficiency reaching >30%. Importantly, this high efficiency by far outperformed those known from previous studies using an alternative approach based on the reconstitution of a split gene on the genomic level (hybrid, overlapping and genomic "*trans*"-splicing approach) (Carvalho et al., 2017, Frontiers in Neurosciences, 11, Article 503). All three strategies were tested *in vitro* using the reporter gene *lacZ*. The highest reported reconstitution efficiencies in this setting amounted to 17.7 %.

[279] A recent study postulated that the efficiency of *trans*-splicing is not influenced by the splice site strength (Lorain et al., 2013). To test this assumption, the cerulean reconstitution efficiency of vgASS_620 was compared with the native acceptor splice site of *RHO* exon 3 in combination with the most efficient binding domain BD_h+i (SEQ ID NO: 28). As may be taken from Figure 4E, compared to vgASS_620, reconstitution efficiencies originating from the native ASS were strikingly lower (35.7 ± 4.6% *vs* 0.7 ± 0.1%). This finding provides clear evidence that the *trans*-splicing efficiency in the dual vector approach strongly depends on the acceptor splice site strength (Fig. 4D-G).

[280] Reconstitution efficacy could be increased in later experiments (see e.g., Example 3, Figure 9D). First of all the number of independent samples (n) has still been very low in the initial experiment. By increasing the number of independent samples (n) and optimizing the transfection protocol, the reconstitution efficiencies could be quantified more reliably and higher values, such as 60% or even higher reconstitution efficiency (Figure 9), could be reached, as the co-transfection efficiency has been increased.

[281] In addition to HEK293 cells reconstitution efficacy trans-splicing has also been tested in 661W cells and MEF cells. Using the binding domain BD_g reconstitution efficacies of > 40% in 661W cells and of > 50% in MEF cells have been observed (data not shown). No significant differences in *trans*-splicing efficacies could be detected when comparing the two cell lines. This suggests that the mRNA trans-splicing approach is cell-type independent.

[282] Collectively, the *in vitro* results indicate that *trans*-splicing efficiencies can be markedly increased not only by varying the sequence and length of the binding domain but

also by optimizing the acceptor splice site strength. These promising findings open new avenues for further optimization of trans-splicing based technologies.

[283] In combination with vgASS_620, the binding domain "g" shown in Fig. 4 can be used to establish a SMaRT-based gene therapy approach of rhodopsin mutations in or downstream of exon 3. In addition, vgASS_620 can also be combined with a variety of other binding domains to treat other inherited retinal disease (IRD) genes via SMaRT or via the dual AAV vector approach. The disclosure provided herein applies to the usage of any DNA or RNA sequence containing the vgASS_620 splice module apart from its naturally occurring context (i.e. apart from the rhodopsin gene in patients carrying the c.620T>G mutation). vgASS_620 of SEQ ID NO: 3 or a nucleotide acid sequence having at least 70 % sequence identity to SEQ ID 3 could be used for following applications:

1) Design of minigenes (e.g. for the analysis of mutations on mRNA splicing)

2) To target endogenous mRNAs in biological or therapeutic context (e.g. using the SMaRT technology)

3) Reconstitution of two mRNA fragments in *trans*. This is of special importance in the context of overcoming the limited genome packaging capacity of adeno-associated virus (AAV) vectors (approx. 5.0 kb, preferably < 4.7 kb). Figures 5 to 8 depict sequences of AAV vectors which can be used in the dual AAV vectors systems described herein.

4) Design of gene expression or targeting cassettes to selectively increase the presence of certain splice products in alternatively spliced genes containing weak ASS splice sites.

Example 3: Impact of the acceptor splice site versus the binding domain (BD) on reconstitution efficacy.

[284] The BD length and sequence represent key determinants of the reconstitution efficiencies. The BDs most likely influence the tight binding and, potentially, folding of the mRNAs, but they are not expected to directly facilitate the efficiency or precision of the subsequent splicing process. As mentioned above, DSSs are well characterized and the prediction of their strength is well-matched on their experimental performance. Consequently, there is no apparent need for optimization of this splice site in the framework of the split fluorophore reconstitution assay. By contrast, due to their complexity, the strength of the ASS cannot be reliably predicted. The results suggest that vgASS_620 is an exceptionally strong acceptor splice site. Given that the splice site strength can impact on the reconstitution efficiencies of the split fluorophore assay, vgASS_620 should lead to higher values when compared to other acceptor splice sites.

[285] To analyze this, the reconstitution efficiencies in the presence of vgASS_620 or two other ASSs, i.e. the native ASS of *RHO* exon 3 (S3, Figure 9) and a hybrid ASS (S2) which

was created by replacing the poly-pyrimidine tract (PPT) of vgASS_620 with the PPT of the native *RHO* exon 3 ASS, were compared. Moreover, each ASS was combined with three different binding domains: a strong (BD_g, B1; SEQ ID NO: 27) and a weak (BD_f, B3) binding domain originating from this study and a published BD sequence (PTM1, B2) obtained from *RHO* intron 1 (SEQ ID NO: 29). This sequence was shown to yield high efficiencies for the repair of mutated *RHO* transcripts via spliceosome-mediated mRNA *trans*-splicing (Berger et al., "Repair of rhodopsin mRNA by spliceosome-mediated RNA trans-splicing: A new approach for autosomal dominant retinitis pigmentosa", (2015) Mol. Ther. 23(5): 918-930). All combinations were analyzed via confocal live cell imaging, RT-PCR and Western blotting (Figure 9). This experiment resulted in several important findings. First, it revealed that both, the BD and the ASS are key components determining reconstitution efficiencies. Second, the most potent BD identified in this study (BD_g, B1) outperforms the published *RHO* binding domain (B2). Third, the ASS strength appears to influence the reconstitution efficiencies even more potently than the BD, as the weakest BD combined with the strongest ASS (B3+S1) still results in detectable *trans*-splicing, whereas the strongest BD combined with the weakest ASS (B1+S3) does not lead to a detectable reconstitution of the coding sequence.

[286] The vgASS_620 consists of an ASS, a PPT and an additional 7 bp sequence upstream, the deletion of which is impairing ASS_620 recognition. It has therefore been speculated that this 7 bp sequence might contain a very potent branch point potentially explaining the universal and efficient recognition of vgASS_620 independent of the gene environment. However, there were no potent branch points predicted within this sequence. Instead, some other sequences were predicted to serve as branch points positioned up to 40 bp upstream of the c.620T>G ASS. Nevertheless, when simultaneously mutating all predicted branch point nucleotides and all potential branch point adenines contained in the 7 bp sequence upstream of the PPT, splicing could not be altered for the c.620T>G mutant (data not shown). This finding indicates that the branch point(s) are either located elsewhere or that the c.620T>G ASS possesses a high flexibility in branch point choice. To identify the branch point(s) utilized for splicing at the c.620T>G ASS more directly, a nested lariat RT-PCR was performed using HEK293 cells transiently expressing the *RHO* c.620T>G minigene. HEK293 cells transfected with the *RHO* WT minigene served as a reference. When performing the lariat RT-PCR, one band was obtained for the WT and for the mutant minigene, respectively, both differing in their size. Both bands appeared somewhat diffuse suggesting that the corresponding lariats were of different size. To identify the single sequences contained in these diffuse bands, the lariat RT-PCR products were subcloned into a TOPO vector and the obtained clones were analyzed individually via sequencing. When investigating the *RHO* WT lariats, two major branch points (used in 42 % and 33 % of

cases) and three minor branch points (8 % of cases each) were identified (Table 2). The two major branch points closely resembled the consensus sequence and yielded high prediction scores. All *RHO* WT branch points were found 57 – 184 bp upstream of the intron-exon-junction. As more than 90 % of human branch points are found within 50 bp upstream of the ASS sequence, mRNA splicing of *RHO* exon 3 seems unconventional (Corvelo et al., 2010, PLoS Comput. Biol., 6(11): e1001016).

Table 2: Detected branch points and corresponding prediction scores

***RHO* WT minigene**			
Branch point	**Relative position**	**HSF score**	**Frequency of usage**
cactgAt	-184 bp	84.3	41.6 %
tccacAa	-177 bp	76.8	8.3 %
tccccAt	-158 bp	85.6	33.3 %
ttatgAt	-113 bp	60.6	8.3 %
cagccAc	-57 bp	58.4	8.3 %
***RHO* c.620T>G minigene**			
atcgAt	-245 bp	83.6	16.7 %
attccAt	-236 bp	55.6	8.3 %
tgcagAc	-213 bp	73.9	8.3 %
cgtttAt	-208 bp	55.9	8.3 %
atcccCt	-200 bp	not predicted	16.7 %
cagggAg	-189 bp	not predicted	8.3 %
gaagcCc	-174 bp	not predicted	8.3 %
ttcccAa	-132 bp	79.5	8.3 %
caggcAg	-118 bp	not predicted	8.3 %
tccctAc	-107 bp	77.6	8.3 %

[287] Nevertheless, the branch point profile obtained for *RHO* c.620T>G was remarkably different, when compared to the WT minigene. First, various branch points have been identified for the c.620T>G mutant. However, no major branch point(s) could be detected and none of them were identical with the ones obtained for the WT minigene. Second, the branch points were located even further upstream from the used ASS, i.e. 107 – 245 bp corresponding to 21 – 159 bp upstream of the native exon 3 ASS. Third, almost half of the detected branch points hardly resembled the consensus sequence and accordingly were not predicted using the human splice finder (HSF) splice prediction tool. All in all, these data suggest that the strength of vgASS_620 does not originate from the presence of a contained, very potent branch point but might partially be caused by a high flexibility in branch point choice. This could explain its unusually efficient performance and renders it a very attractive tool for biotechnological applications requiring efficient splicing.

Example 4: Investigation of mRNA *trans*-splicing-based rAAV dual vectors *in vivo*

[288] The most potent application of the mRNA *trans*-splicing-based assay evaluated in the previous sections is the reconstitution of large genes in the framework of dual rAAV vectors. Consequently, the mRNA *trans*-splicing approach was tested in the mouse retina using rAAVs. For this purpose, a slightly modified version of the split fluorophore assay was used. To control the rAAV vector-born expression in the cells transduced with the single viruses, both dual rAAV vector cassettes were equipped with a fluorophore sequence, i.e. a citrine at the 5' end of the coding sequence for the 5' vector and a mCherry at the 3' end of the coding sequence for the 3' vector. One of the BDs yielding the highest reconstitution efficiencies *in vitro*, i.e. BD_h+i, was used *in vivo* (Figure 10A). In this experimental setup, cerulean fluorescence should be present in cells expressing citrine as well as mCherry.

[289] The titer-matched viruses were injected subretinally at postnatal day 21 (P21) into WT C57Bl6/J mice. After harvesting the retinas two weeks post injection, a solid fluorophore expression could be detected in the RPE (Figure 10B). Moreover, cerulean fluorescence could be observed in all areas where citrine and mCherry were expressed indicating successful mRNA *trans*-splicing in cells co-transduced with both AAVs. However, citrine and cerulean possess partially overlapping excitation and emission spectra. To exclude the possibility of the cerulean fluorescence being an artefact caused by e.g. bleed-through of citrine or mCherry, both fluorophores were bleached selectively in a small area of the RPE by exciting the fluorophores with a high intensity 514 nm laser. With this procedure, the citrine and mCherry fluorescence could be removed completely (Figure 10C). Nevertheless, the cerulean fluorescence remained unchanged indicating that it originates solely from *trans*-spliced cerulean mRNA. This experiment provides proof of principle for the usability of mRNA *trans*-splicing for reconstitution of genes expressed from AAVs *in vivo.*

Example 5: Identification of potent binding domains suitable for human gene therapy

[290] So far, all binding domain (BD) sequences were obtained from human intronic regions. When used for human gene therapy, they thus could potentially also bind to endogenous mRNAs and induce *trans*-splicing with these transcripts. For application of mRNA *trans*-splicing in human gene therapy, it is therefore necessary to identify BDs which do not contain sequences homologous to the human genome. For this purpose, a random 100 bp sequence was taken from the bacterial *lacZ* gene and modified via random insertions, deletions and substitutions to obtain four sequences without any homology to the human genome (Figure 11A).

[291] When co-transfecting HEK293 cells with the 5' vector- and 3' vector-constructs containing the respective binding domains, very high cerulean reconstitution efficiencies of

78.3 % ± 2.1 % could be observed for one of the BDs (BD_k, Figure 11B – D). BD_k was thus more efficient than the best performing human BDs shown in Figure 4 and was therefore used for preliminary experiments evaluating reconstitution of large genes via mRNA *trans*-splicing.

Example 6: Impact of polyadenylation signal on reconstitution efficiency and involvement of DNA-based reconstitution in *in vitro* assay

[292] As the reconstitution of the two pre-mRNA molecules originating from the 5' and the 3' vector should be taking place in the nucleus, the polyadenylation signal (pA), which is necessary for stabilizing and translating the mature mRNA, could in theory be omitted in the 5' vector. The advantage of this deletion is that any residual non-spliced 5' pre-mRNA should not result in the translation of a truncated protein. Therefore, the impact of a deletion of the pA in the 5' vector on the cerulean reconstitution efficiency was investigated. For this purpose, a 5' vector lacking the pA was co-transfected with the regular 3' vector. Compared to the co-transfected vectors both containing a pA, the reconstitution seems to be slightly reduced (Fig. 12A and B). However, this experiment shows that the pA signal can be omitted if necessary. Moreover, the successful reconstitution observed so far could in theory also be mediated by homologous recombination at the DNA level as in the prior art hybrid dual vector approach, as all necessary components, i.e. a recombinogenic sequence and splicing elements for removal of this sequence via *cis*-splicing, are present in the mRNA *trans*-splicing vectors (Carvalho et al., 2017). To exclude this possibility, the promoter in the 3' vector was deleted to prevent its transcription into a pre-mRNA and hence resembling the 3' vectors used for the hybrid dual vector approach known so far. After co-transfecting this construct with the regular 5' vector, no reconstitution could be observed. This result confirms that all observed reconstitution of cerulean is mediated exclusively via mRNA *trans*-splicing.

Example 7: Proof-of-principle for reconstitution of large genes via mRNA *trans*-splicing

[293] In addition to the AAV cerulean split reporter assay for evaluation of mRNA *trans*-splicing a protein with therapeutic potential has been tested. The transcriptional activator *SpCas*9-VPR, a catalytically inactive nuclease fused to the transcriptional activator domains VP64-p65-Rta (VPR), is a recently developed novel tool for gene therapy. Due to its large size (5.8 kb), it needs to be delivered via dual vectors for *in vivo* applications. Consequently, *SpCas*9-VPR represents a suitable candidate for reconstitution via mRNA *trans*-splicing. The coding sequence of *SpCas*9-VPR was split into two halves at c.2185 and the two halves of the coding sequence were equipped with BD_k or its complementary sequence and the appropriate splicing element, i.e. DSS or vgASS_620. A full length (FL) *SpCas*9-VPR

construct was used as a positive control. The RT-PCR from HEK293 cells co-transfected with the split constructs revealed that *SpCas*9-VPR mRNA was reconstituted successfully and that no unwanted side-products were generated (Figure 13B). Sequencing of the PCR product confirmed an accurate restoration of the reading frame (Figure 13C). Moreover, FL *SpCas*9-VPR protein could be detected via western blotting (Figure 13D). The reconstitution efficiency of *SpCas*9-VPR was 13.2 % ± 0.9 %. This provides a proof of principle for the applicability of the mRNA *trans*-splicing approach for reconstitution of large genes.

Example 8: Reconstitution of ABCA4 (Figure 14)

[294] To finally also investigate the reconstitution of a large gene of human origin, the *ABCA4* gene (6.8 kb) encoding the retinal ATP-binding cassette transporter was selected. This gene is a suitable candidate for gene therapy, as its mutations cause the inherited retinal dystrophy Stargardt macular dystrophy. *ABCA4* was split into two halves at c.3243 and equipped with BD_ k, DSS and vgASS_620 (Figure 14A). Both halves contain a long intronless *ABCA4* coding sequence (CDS) which does not resemble native pre-mRNAs composed of exons and introns. This therefore might impede the recruitment of the splice factors to the pre-mRNA and consequently might also reduce the mRNA *trans*-splicing efficiency. For the split gene to resemble endogenous human pre-mRNA more closely, additional constructs were designed containing three short (80 bp) intervening introns in both CDS halves. HEK293 cells were transiently co-transfected with the intronless or intron-containing constructs to examine the reconstitution-capability *in vitro*.

The RT-PCR revealed that *ABCA4* was reconstituted successfully on mRNA level for the split constructs with and without intervening introns (Figure 14B). Interestingly, the split constructs containing introns seem to be *trans*-spliced more efficiently. Again, no unspecific splice products could be detected and the reading frame was correctly restored (Figure 14C).

Example 9: rAAV dual vector mRNA *trans*-splicing of ABCA4 *in vivo* (Figure 15)

[295] In a final experiment, *ABCA4* reconstitution was additionally tested *in vivo*. For this purpose, the CMV promoter was replaced with a human rhodopsin (hRHO) promoter to ensure a photoreceptor-specific expression (Figure 15A). AAVs were produced using the in-house optimized NN and GL capsid variants derived from the wildtype AAV2 capsid as described in WO 2019/076856. The titer-matched viruses containing the 5' or the 3' CDS of *ABCA4* were co-injected subretinally into 1 month old C57Bl6/J wildtype mice. After 4 weeks, retinas were harvested and successful reconstitution was evaluated on mRNA as well as on protein level. The RT-PCR performed with a junction-spanning primer pair revealed that both capsid variants, i.e. NN and GL, resulted in successful reconstitution with

the NN capsid yielding higher levels of reconstituted mRNA most likely due to higher (co-)transduction efficiencies (Figure 15B). A seamless ligation of the two separate pre-mRNA molecules was confirmed via sequencing (Figure 15C). For a more exact quantification, a qRT-PCR was performed. These preliminary results (n = 1) show that a relative *ABCA4* expression ranging from a 10-fold to a 42-fold increase relative to a non-injected C57Bl6/J wildtype retina was reached, again confirming a successful and efficient reconstitution at the mRNA level (Figure 15D). Finally, to investigate *ABCA4* reconstitution at the protein level, protein lysates of injected retinas were used for western blotting. To ensure a specific detection of the transgene-derived protein, an anti-myc antibody was used. The results show that in both cases the reconstituted mRNA resulted in a successful *in vivo* protein expression (Fig. 15E).

CLAIMS

1. A pre-mRNA *trans*-splicing molecule comprising,
   (i) an acceptor splice region, comprising
      (ia) a pyrimidine tract, wherein the pyrimidine tract comprises
         (iaa) 5 to 25 nucleotides;
         (iab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U);
      (ib) an acceptor splice site,
         (iba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and
         (ibb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G;
   (ii) a nucleotide sequence of interest or a portion thereof; wherein the acceptor splice region is localized 3' or 5' to the nucleotide sequence of interest or the portion thereof;
   (iii) a binding domain targeting pre-mRNA, which is localized 3' or 5' to the nucleic acid sequence of interest or the portion thereof; and
   (iv) optionally a spacer sequence, wherein the spacer sequence is localized between the binding domain and the acceptor splice region.

2. The pre-mRNA *trans*-splicing molecule of claim 1 comprising,
   (ii) the nucleotide sequence of interest or the portion thereof, wherein the acceptor splice region is localized 5' to the nucleotide sequence of interest or the portion thereof;
   (iii) a donor splice site, wherein the donor splice site is localized 3' to the nucleotide sequence of interest or the portion thereof;
   (iv) a first binding domain targeting pre-mRNA located 5' to the nucleotide sequence of interest or the portion thereof;
   (v) a second binding domain targeting pre-mRNA located 3' to the nucleotide sequence of interest or the portion thereof;
   (vi) optionally a first spacer sequence, wherein the first spacer is localized between the first binding domain and the acceptor splice region; and
   (vii) optionally a second spacer sequence, wherein the second spacer is localized between the second binding domain and the donor splice site.

3. The pre-mRNA *trans*-splicing molecule of claim 1 or 2, wherein the acceptor splice

region is localized 5' to the nucleotide sequence of interest or the portion thereof and the binding domain is localized 5' to acceptor splice region.

4. The pre-mRNA *trans*-splicing molecule of any one of claims 1 to 3 further comprising a termination sequence, preferably a polyA sequence.

5. The pre-mRNA trans-splicing molecule of any one of claims 1 to 4, wherein
(a) the 5 to 25 nucleotides of the pyrimidine tract comprise a sequence encoded by the sequence of TTTTTT or TCTTTT;
(b) the sequence between the last pyrimidine of the pyrimidine tract and the acceptor splice site has less than 10, preferably less than 5, more preferably less than 3 bases; and/or
(c) the acceptor splice site has a sequence of CAGG.

6. The pre-mRNA trans-splicing molecule of any one of claims 1 to 5, wherein the acceptor splice region further comprises
(a) 7 nucleotides 5' to the pyrimidine tract having at least 4 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotide is C;
(b) 7 nucleotides 5' to the pyrimidine tract having at least 5 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotide is C;
(c) 7 nucleotides 5' to the pyrimidine tract having at least 4 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAA; or
(d) 7 nucleotides 5' to the pyrimidine tract having at least 5 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotides are CAAC;

7. A DNA molecule comprising a promoter and a sequence encoding the pre-mRNA *trans*-splicing molecule according to any one of claims 1 to 6.

8. A method for producing a nucleic acid sequence, the method comprising
(A) providing a first nucleic acid sequence comprising one or more donor splice site sequences;
(B) providing a second nucleic acid sequence, wherein the second nucleic acid sequence comprises
(i) an acceptor splice region sequence comprising
(ia) a pyrimidine tract, wherein the pyrimidine tract comprises
(iaa) 5 to 25 nucleotides;
(iab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as

cytosine (C), thymine (T) and/or uracil (U);

(ib) an acceptor splice site,

(iba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and

(ibb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G;

(C) cleaving the first nucleic acid sequence at the one or more donor splice site sequence(s) and cleaving the second nucleic acid sequence in the acceptor splice site,

(D) ligating the first cleaved nucleic acid sequence to the second cleaved nucleic acid sequence,

thereby obtaining a nucleic acid sequence.

9. The method of claim 8, wherein the first nucleic acid sequence further comprises a nucleotide sequence of interest or a portion thereof, wherein at least a portion of the nucleotide sequence of interest is located 5' to the donor splice site and wherein the second nucleic acid sequence further comprises a nucleotide sequence of interest or a portion thereof, wherein at least a portion of the nucleotide sequence of interest is located 3' to the splice acceptor splice region.

10. The method of claim 8 or 9, wherein the first and the second nucleic acid sequence are introduced into a host cell, preferably wherein the first and the second nucleic acid sequence are a recombinant nucleic acid sequence.

11. The method of any one of claims 8 to 10, comprising

(A) introducing into a host cell a first nucleic acid sequence comprising a pre-mRNA trans-splicing molecule sequence or a nucleic acid sequence encoding said pre-mRNA *trans*-splicing molecule, wherein the first pre-mRNA *trans*-splicing molecule comprises from 5' to 3'

(a) a 5' portion of a nucleotide acid sequence of interest;

(b) a donor splice site;

(c) optionally a spacer sequence;

(d) a first binding domain; and

(e) optionally a termination sequence, preferably a polyA sequence, and

(B) introducing into the host cell a second nucleic acid sequence comprising a pre-mRNA *trans*-splicing molecule sequence or a nucleic acid sequence encoding said pre-mRNA *trans*-splicing molecule, wherein the second pre-mRNA *trans*-splicing molecule comprises from 5' to 3';

(i) a second binding domain, which is complementary to the first target domain of the first nucleic acid sequence;
(ii) an acceptor splice region sequence comprising;
(iia) a pyrimidine tract, wherein the pyrimidine tract comprises
(iiaa) 5 to 25 nucleotides;
(iiab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U);
(iib) a acceptor splice site,
(iiba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and
(iibb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G;
(iii) a 3' portion of the nucleotide sequence of interest, and
(iv) a termination sequence, preferably a polyA sequence,

(C) cleaving the first nucleic acid sequence at the donor splice site sequence and cleaving the second nucleic acid sequence in the acceptor splice site; and
(D) ligating the first cleaved nucleic acid sequence comprising the 5' portion of the nucleotide sequence of interest to the second cleaved nucleic acid sequence comprising the 3' portion of the nucleotide sequence of interest, thereby obtaining the nucleic acid sequence of interest.

12. An adeno-associated virus (AAV) vector comprising at least two inverted terminal repeats comprising a nucleic acid sequence between these two inverted terminal repeats, wherein said nucleic acid sequence comprises from 5' to 3 '
(i) a promoter;
(ii) a binding domain;
(iii) optionally a spacer sequence;
(iv) an acceptor splice region sequence comprising
(a) a pyrimidine tract, wherein the pyrimidine tract comprises
(aa) 5 to 25 nucleotides;
(ab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U);
(b) a acceptor splice site,
ba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and
bb) wherein the acceptor splice site comprises from 5' to 3' a

sequence of NAGG, wherein N is a A, C, T/U or G;

(v) a nucleotide sequence of interest or a portion thereof; and

(vi) optionally a poly A sequence.

13. An adeno-associated virus (AAV) vector system comprising

(I) a first AAV vector comprising at least two inverted terminal repeats comprising a nucleic acid sequence between these two inverted terminal repeats, wherein said nucleic acid sequence between these two inverted terminal repeats comprises from 5' to 3'

(a) a promoter,

(b) a nucleotide sequence encoding an N-terminal portion of a polypeptide of interest;

(c) a donor splice site;

(d) optionally a spacer sequence;

(e) a first binding domain; and

(f) optionally a termination sequence, preferably a polyA sequence;

(II) a second AAV vector comprising a nucleic acid sequence comprising at least two inverted terminal repeats comprising a nucleic acid sequence between these two inverted terminal repeats, wherein said nucleic acid sequence between these two inverted terminal repeats comprises from 5' to 3'

(i) a promoter;

(ii) a second binding domain, which is complementary to the first binding domain of the first AAV vector;

(ii) optionally a spacer sequence

(iii) an acceptor splice region sequence comprising

(a) a pyrimidine tract, wherein the pyrimidine tract comprises

(aa) 5 to 25 nucleotides;

(ab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U);

(b) an acceptor acceptor splice site,

(ba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and

(bb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G;

(iv) a nucleotide sequence encoding a C-terminal portion of the polypeptide of interest;

(iva) wherein the C-terminal portion of the polypeptide of interest

and the N-terminal portion of the polypeptide of interest reconstitute the polypeptide of interest; and

(v) a termination sequence, preferably a polyA sequence.

14. The AAV vector system of claim 13, wherein the polypeptide is a full-length polypeptide and the first AAV vector comprises an N-terminal portion of a full-length polypeptide of interest and the second AAV vector comprises a C-terminal portion of the full-length polypeptide of interest.

15. A nucleic acid sequence comprising

(i) an acceptor splice region sequence comprising

(ia) a pyrimidine tract, wherein the pyrimidine tract comprises

(aa) 5 to 25 nucleotides;

(ab) wherein at least 60 % of the nucleotides within these 5 to 25 nucleotides are pyrimidine bases such as cytosine (C), thymine (T) and/or uracil (U);

(ib) a acceptor splice site,

ba) wherein the acceptor splice site is located 3' to the pyrimidine tract; and

bb) wherein the acceptor splice site comprises from 5' to 3' a sequence of NAGG, wherein N is a A, C, T/U or G; and

(ii) a nucleotide sequence of interest, wherein the nucleotide sequence of interest

(iia) is located 3' or 5' to the acceptor splice region.

16. The method of any one of claims 8 to 11 the AAV vector of claim 12, the AAV vector system of claim 13 or 14 and the nucleic acid of claim 15, wherein

(a) the 5 to 25 nucleotides of the pyrimidine tract comprise the sequence TTTTTT or TCTTTT;

(b) the sequence between the last pyrimidine of the pyrimidine tract and the acceptor splice site has less than 10, preferably less than 5, more preferably less than 3 bases;

(c) the acceptor splice site has a sequence of CAGG;

(d) the acceptor splice region further comprises 7 nucleotides 5' to the pyrimidine tract having at least 4 nucleotides of the sequence CAACGAG, wherein the first 5' nucleotide is C; and/or

(e) wherein the splice acceptor region comprises the sequence of SEQ ID NO: 3 or 4.

17. The pre-mRNA trans-splicing molecule of any one of claims 1 to 6, the adeno-associated virus vector of claim 12 or the adeno-associated virus vector system of claim 13 or 14 for use in therapy.