The present invention therefore implements cellular large protein reconstitution by providing to a target cell two or more fragments of said large protein fused to split inteins to promote intein-mediated trans-splicing and reconstitute the functional protein.

SUMMARY

The present disclosure provides gene therapy with AAV vectors for diseases due to mutations of genes, in particular of genes with coding regions exceeding 5 kb.

Based on the findings that protein trans-splicing mediated by split-inteins is used by single cell organisms to reconstitute proteins, the inventors have constructed multiple AAV vectors each encoding one of the fragments of either reporter or large therapeutic proteins flanked by short split-inteins, resulting in protein trans-splicing and full-length protein reconstitution in vitro and in vivo.

Advantageously, the AAV-based protein trans-splicing-mediated reconstitution of disease proteins achieved by the present disclosure afforded expression of larger amounts of target proteins than AAV-based methods for large proteins known in the art. This is probably due to the overcoming of various limiting steps required for efficient transduction of dual vector-based systems including: proper DNA concatemer formation, stability of the heterogeneous mRNA and splicing efficiency across the junctions of the vectors.

In one aspect, the present disclosure provides a vector system to express a coding sequence in a cell, said coding sequence consisting of a first portion (CDS1), a second portion (CDS2) and optionally a third portion (CDS3), said vector system comprising:

a) a first vector comprising:

- said first portion of said coding sequence (CDS1),

-a first intein nucleotide sequence coding for a N-Intein, said sequence being located at the 3’ end of CDS1; and

b) a second vector comprising:

- said second portion of said coding sequence (CDS2),

-a second intein nucleotide sequence coding for a C-Intein, said sequence being located at the 5’ end of CDS2;

wherein when the first vector and the second vector are inserted in a cell, the protein product of the coding sequence is produced by protein splicing;

or optionally said vector system comprising:

a’) a first vector comprising:

- said first portion of said coding sequence (CDS1),

-a first intein nucleotide sequence coding for a first N-Intein, said sequence being located at the 3’ end of CDS1; and

b’) a second vector comprising:

- said second portion of said coding sequence (CDS2),

-a second intein nucleotide sequence coding for a first C-Intein, said sequence being located at the 5’ end of CDS2;

-a third intein nucleotide sequence coding for a second N-Intein, said sequence being located at the 3’ end of CDS2; and

c') a third vector comprising:

-said third portion of said coding sequence (CDS3)

-a fourth intein nucleotide sequence coding for a second C-Intein, said sequence being located at the 5’ end of CDS3

wherein the first intein nucleotide sequence is different from the third intein nucleotide sequence when present and the second intein sequence is different from the fourth intein nucleotide sequence when present, wherein when the first vector, the second vector, and the third vector when present are inserted in a cell, the protein product of the coding sequence is produced by protein trans-splicing, and wherein the coding sequence encodes a protein able to correct a retinal degeneration disorder.

Preferably in the vector system the first intein, the second intein, the third intein and the fourth intein encodes for a split intein, preferably said split intein has a maximum length of 150 amino acids, more preferably said split intein is a DnaE or DnaB intein.

According to the present disclosure, an intein is a segment of a protein that is able to excise itself and join the remaining portions (the exteins) with a peptide bond in a process known as protein splicing. The segments are called "intein" for internal protein sequence, and "extein" for external protein sequence, with upstream exteins termed "N-exteins" and downstream exteins called "C-exteins", the upstream intein called "N-Intein" and the downstream intein called "C-Intein"."

Therefore, in the context of the present disclosure, an N-Intein is an intein fragment located at the N-terminus of (and fused with) the first polypeptide and a C-Intein is an intein fragment located at the C-terminus of (and fused with) the second polypeptide, wherein upon expression of the two polypeptides, the two intein fragments undergo protein trans-splicing and are joined to form a full intein, and the two polypeptides are joined, wherein when the two polypeptides form a full length protein, said full length protein is reconstituted.

According to the present disclosure, the first intein sequence is an N-intein sequence and the second intein sequence is a C-Intein sequence, wherein said N-Intein and said C- Intein are preferably derived from the same intein or split intein gene. Alternatively, said N-Intein and said C-Intein derive from two different intein genes which are able to undergo the trans-splicing reaction naturally or are modified to do so. Accordingly, the same gene may be the from the same organism or from different organisms. For instance, widely used split inteins derive from the DnaE gene from different organisms. According to the present disclosure, when the coding sequence of the protein of interest is split into two portions, the N-intein coding sequence is fused in frame with the sequence coding for the N-terminal portion of the protein of interest; the C-Intein coding sequence is fused in frame with the sequence coding for the C-terminal portion of the sequence of interest. Upon expression of the two precursor fusion proteins, the inteins undergo autocatalytic excision and form a ligated extein, eg the reconstituted protein of interest.

According to the present disclosure, the coding sequence of the protein of interest may be split into three portions. Accordingly, the first intein sequence is an N-intein sequence and the second intein sequence is a C-Intein sequence, wherein the first intein coding sequence is fused in frame at the C-terminus to the sequence coding for the N-portion of the protein of interest, and the second intein coding sequence is fused in frame at the N-terminus of the sequence coding for the middle portion of the protein of interest. Accordingly, said N-Intein and said C- Intein are preferably derived from the same intein or split intein gene. Alternatively, said N-Intein and said C-Intein derive from two different intein genes which are able to undergo the trans-splicing reaction naturally or are modified to do so. Accordingly, the same gene may be the from the same organism or from different organisms. Within the present configuration, the third intein is an N-Intein coding sequence fused in frame to the sequence coding for the C-terminus of the middle portion of the protein of interest, and the fourth intein is a C-Intein coding sequence fused in frame to the sequence coding for the N-terminus of the C-portion of the protein of interest. Accordingly, said third and fourth inteins are preferably derived from the same intein or split intein gene. Alternatively, said N-Intein and said C-Intein derive from two different intein genes which are able to undergo the trans-splicing reaction naturally or are modified to do so. Accordingly, the same gene may be the from the same organism or from different organisms. Within the scope of the present disclosure, said first and second inteins and said third and fourth inteins derive from different intein genes and the first intein binds selectively the second intein, while the third intein binds selectively the fourth intein.

In the present disclosure when the first vector, the second vector and optionally the third vector are inserted in a cell, a least two fusion proteins or three fusion proteins are formed and when contacting said two fusion proteins or three fusion proteins, the protein product of the coding sequence is produced. The step of contacting is performed under conditions that permit binding of the N-intein to the C-intein.

In the present disclosure when the first vector, the second vector and the third vector are inserted in a cell, three independent polypeptides are produced, and full-length protein is produced via trans-splicing. Pivotal to the development of the three AAV intein vectors has been the use of different inteins, i.e. DnaE and DnaB, which do not cross-react thus

preventing improper trans-splicing between the polypeptides produced by the first and the third vector.

According to preferred examples of the present disclosure, a vector system to express the coding sequence of a gene of interest in a cell comprise two vectors, each vector comprising a portion of said coding sequence flanked by an intein sequence, wherein the 5'end of said coding sequence is flanked at the 3' terminus by the sequence of an N-intein, and the 3' end of the coding sequence of the gene of interest is flanked by the sequence of a C-Intein, such that when both vectors are expressed in a cell, two fusion proteins are produced and the full length protein of interest is generated as a result of a spontaneous trans-splicing reaction.

According to a further preferred example of the disclosure, the vector system to express the coding sequence of a gene of interest in a cell comprises three vectors, each vector comprising a portion of said coding sequence flanked by an intein sequence, wherein the coding sequence is divided in three portions such that the 5'end of said coding sequence is flanked at the 3' terminus by the sequence of a first N-intein; the middle portion of said coding sequence is flanked at the 5' terminus by a first C-Intein, and at the 3' terminus with a second N-Intein; the 3' portion of said coding sequence is flanked at the 5' terminus by a second C-Intein, such that when all three vectors are expressed in a cell, three fusion proteins are produced, and the full length protein of interest is generated as a result of a spontaneous trans-splicing reaction wherein the first N-Intein reacts with the first C-Intein and the second N-Intein reacts with the second C-Intein.

Split inteins of the disclosure may be encoded by one gene which is then engineered to encode two separate intein fragments, eg split inteins; alternatively, naturally occurring split inteins are encoded by two separate genes; for instance in cyanobacteria, DnaE, the catalytic subunit α of DNA polymerase III, is encoded by two separate genes, *dnaE-n* and *dnaE-c*. Preferred inteins within the present disclosure are inteins which derive from intein proteins (eg mini inteins) or split inteins which form intein proteins via trans-splicing reaction, which are 150 aa long or less.

Split inteins of the disclosure may be 100% identical, 98%, 80%, 75%, 70%, 65%, 60%, 55%, 50% identical to naturally occurring inteins or to SEQ ID No. 1 to 14 (homologs), wherein said inteins retain the ability to undergo trans-splicing reactions. Within the scope of the present disclosure are fragments or variants of naturally occurring or modified inteins which retain trans-splicing activity.

Conveniently, split inteins of the disclosure may be derived from the same gene isolated from different organisms. Preferred intein genes are *Dna B* and *Dna E* .

In a preferred example, the intein of the disclosure is a split intein derived from the *DnaE* gene (eg DNA polymerase III subunit alpha) from cyanobacteria including *Nostoc punctiforme (Npu) Synechocystis sp. PCC6803 (Ssp), Fischerella sp. PCC 9605, Scytonema tolypothrichoides, Cyanobacteria bacterium SW_9_47_5, Nodularia spumigena, Nostoc flagelliforme, Crocosphaera watsonii WH 8502, Chroococcidiopsis cubana CCALA 043, Trichodesmium erythraeum*; preferably, the intein of the disclosure is derived from Dna E gene isolated from *nostoc puntiforme or Synechocystis sp. PCC6803*.

In a further preferred example, the intein of the disclosure is a split intein derived from the *DnaB* gene from cyanobacteria including *R. marinus* (Rma), *Synechocystis* sp. PC6803 (Ssp), *Porphyra purpurea* chloroplast (Ppu) which are described for instance in (59).

Preferably,

-the first intein nucleotide sequence encodes for an intein selected from the group consisting of: SEQ ID No 1, 3, 5, 7, 9, 11, 13 or a variant thereof or a fragment thereof or an homolog thereof;

-the second intein nucleotide sequence encodes for an intein selected from the group consisting of: SEQ ID No 2, 4, 6, 8, 10, 12, 14 or a variant thereof or a fragment thereof or an homolog thereof;

-the third intein nucleotide sequence encodes for an intein selected from the group consisting of: SEQ ID No1, 3, 5, 7, 9, 11, 13 or a variant thereof or a fragment thereof or an homolog thereof;

-the fourth intein nucleotide sequence encodes for an intein selected from the group consisting of: SEQ ID No2, 4, 6, 8, 10, 12, 14 or a variant thereof or a fragment thereof or an homolog thereof.

Preferably, wherein when the first or third intein is SEQ ID 1, the second or fourth is SEQ ID 2; or when the first or third intein is SEQ ID 3, the second or fourth intein is SEQ ID 4; or when the first or third intein is SEQ ID 5, the second or fourth is SEQ ID 6; or when the first or third intein is SEQ ID 7, the second or fourth is SEQ ID 8; or when the first or third intein is SEQ ID 9, the second or fourth is SEQ ID 10; or when the first or third intein is SEQ ID 11, the second or fourth is SEQ ID 12.

Preferably when the first intein is SEQ ID 1 and the second intein is SEQ ID 2, the third intein is not SEQ ID 1 and the fourth intein is not SEQ ID 2; preferably when the first intein is SEQ ID 3 and the second intein is SEQ ID 4, the third intein is not SEQ ID 3 and the fourt intein is not SEQ ID 4; preferably when the first intein is SEQ ID 5 and the second intein is SEQ ID 6, the third intein is not SEQ ID 5 and the fourth intein is not SEQ ID 6; preferably when the first intein is SEQ ID 7 and the second intein is SEQ ID 8, the third intein is not SEQ ID 7 and the fourth intein is not SEQ ID 8; preferably when the first intein is SEQ ID 9 and the second intein is SEQ ID 10, the third intein is not SEQ ID 9 and the fourth intein is not SEQ ID 10; preferably when the first intein is SEQ ID 11 and the second intein is SEQ ID 12, the third intein is not SEQ ID 11 and the fourth intein is not SEQ ID 12.

In a particular example, the first intein is SEQ ID 1, the second intein is SEQ ID 2, the third intein is SEQ ID 3, the fourth Intein is SEQ ID 4; or, the first intein is SEQ ID 5, the second intein is SEQ ID 6, the third intein is SEQ ID 3 and the fourth Intein is SEQ ID 4.

In a preferred example the first vector, the second vector and the third vector further comprise a promoter sequence operably linked to the 5'end portion of said first portion of the coding sequence (CDS1) or of said second portion of the coding sequence (CDS2) or of said third portion of the coding sequence (CDS3).

Preferred promoters are ubiquitous, artificial, or tissue specific promoters, including fragments and variants thereof retaining a transcription promoter activity. Particularly preferred promoters are photoreceptor-specific promoters including photoreceptor-specific

human G protein-coupled receptor kinase 1 (GRK1), Interphotoreceptor retinoid binding protein promoter (IRBP), Rhodopsin promoter (RHO), vitelliform macular dystrophy 2 promoter (VMD2), Rhodopsin kinase promoter (RK); Further particularly preferred promoters are muscle-specific promoters including MCK, MYODI; liver-specific promoters including thyroxine binding globulin (TBG), hybrid liver-specific promoter (HLP) (67); neuron-specific promoters including hSYN1, CaMKIIa; kidney-specific promoters including Ksp-cadherin16, NKCC2. Ubiquitous promoters according to the present disclosure are for instance the ubiquitous cytomegalovirus (CMV)(32) and short CMV (33) promoters More preferred promoters within the scope of the present disclosure are GRK1, TBG, CaMKIIa, Ksp-cadherin16.

In a still preferred example the first vector, the second vector and the third vector further comprise a 5'-terminal repeat (5'-TR) nucleotide sequence and a 3'-terminal repeat (3'-TR) nucleotide sequence, preferably the 5'-TR is a 5'-inverted terminal repeat (5'-ITR) nucleotide sequence and the 3'-TR is a 3'-inverted terminal repeat (3'-ITR) nucleotide sequence.

In a still preferred example the first vector, the second vector and the third vector further comprise a poly-adenylation signal nucleotide sequence.

In a still preferred example the coding sequence is split into the first portion, the second portion and optionally the third portion, at a position consisting of a nucleophile amino acid which does not fall within a structural domain or a functional domain of the encoded protein product, wherein the nucleophile amino acid is selected from serine, threonine, or cysteine.

Preferably at least one of the first vector, the second vector and the third vector further comprises at least one enhancer or regulatory nucleotide sequence, operably linked to the coding sequence.

Preferred enhancer or regulatory nucleotide sequence are the □-globin IgG chimeric intron, the Woodchuck hepatitis virus Post-transcriptional Regulatory Element.

Phenylketonuria, dysferlinopathies, Rett's syndrome, Polycystic kidney disease, Niemann-Pick type C, Huntington's disease.

More preferably the coding sequence is the coding sequence of a gene selected from the group consisting of: ABCA4, MYO7A, CEP290, CDH23, EYS, PCDH15, CACNA1, SNRNP200, RP1, PRPF8, RP1L1, ALMS1, USH2A, GPR98, HMCN1 or a fragment thereof or an ortholog thereof or a minigene thereof with a coding sequence exceeding 5kb in length, i.e, . a minimal gene fragment that includes one or more and the control regions necessary for the gene to express itself in the same way as a wild type gene fragment.

Still preferably the coding sequence is the coding sequence of a gene selected from the group consisting of: DMD, CFTR, F8, ATP7B, PAH, DYSF, MECP2, PKD, NPC1, HTT or a fragment thereof or an ortholog thereof or a minigene thereof thereof with a coding sequence exceeding 5kb in lentgh, i.e, . a minimal gene fragment that includes one or more and the regulatory elements necessary for the gene to express itself in the same way as a wild type gene fragment.

In a particularly preferred example of the disclosure, the coding sequence encodes the ABCA4 gene. Preferably, said coding sequence is split at a nucleotide corresponding to aa Cys1150, Ser1168, Ser 1090 of said ABCA4 protein, and a split intein is inserted at the split point.

In a further preferred example, the coding sequence encodes the CEP290 gene. Preferably, said coding sequence is split at a nucleotide corresponding to aa Cys1076; Ser1275. More preferably, said coding sequence is split at a nucleotide sequence corresponding to aa Cys 929 and 1474; Ser 453 and Cys 1474 of said CEP290 protein, and two split inteins are inserted at the split points.

EGFP SEQ ID No. 15

The first amino acid of the c-extein is highlighted whitin the sequence.Split Cys.71 **(bold)**

MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQ**C**FSRYPDHMKQHD
FFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKI
RHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKDYKDHDGDYKD
HDIDYKDDDDK*

WKVQHHMAPTKDEFDCKAWAYFSDVDLEKDVHSGLIGPLLVCHTNTLNPAHGRQVTVQEFALFFTIFDE
TKSWYFTENMERNCRAPCNIQMEDPTFKENYRFHAINGYIMDTLPGLVMAQDQRIRWYLLSMGSNENI
HSIHFSGHVFTVRKKEEYKMALYNLYPGVFETVEMLPSKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQT
PLGMASGHIRDFQITASGQYGQWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIIHGIKTQGARQKF
SSLYISQFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIKHNIFNPPIIARYIRLHPTHYSIRSTLRMEL
MGCDLNSCSMPLGMESKAISDAQITASSYFTNMFATWSPSKARLHLQGRSNAWRPQVNNPKEWLQV
DFQKTMKVTGVTTQGVKSLLTSMYVKEFLISSSQDGHQWTLFFQNGKVKVFQGNQDSFTPVVNSLDPP
LLTRYLRIHPQSWVHQIALRMEVLGCEAQDLY*

ecDHFR SEQ ID No. 19

MISLIAALAVDYVIGMENAMPWNLPADLAWFKRNTLNKPVIMGRHTWESIGRPLPGRKNI
ILSSQPSTDDRVTWVKSVDEAIAACGDVPEIMVIGGGRVIEQFLPKAQKLYLTHIDAEVE
GDTHFPDYEPDDWESVFSEFHDADAQNSHSYCFEILERR*

mini ecDHFR SEQ ID No. 20

MISLIAALAVDYVIGMENAMPWNLPADLAWFKRNTLNKPVIMGRHTWESIGRPLPGRKNI
ILSSQPSTDDRVTWVKSVDEAIAACGDVPEIMVIGGGRVIEQFLP*

In a preferred example, the vector system of the disclosure comprises:

a) a first vector comprising in a 5'-3' direction:

- a 5'-inverted terminal repeat (5'-ITR) sequence;

- a promoter sequence;

- a 5' end portion of a coding sequence (CDS1), said 5'end portion being operably linked to and under control of said promoter;

- a first intein nucleotide sequence coding for a N-Intein; and

- a 3'-inverted terminal repeat (3'-ITR) sequence; and

b) a second vector comprising in a 5'-3' direction:

- a 5'-inverted terminal repeat (5'-ITR) sequence;

- a promoter sequence;

- a fourth intein nucleotide sequence coding for a second C-Intein;
- the third portion of the coding sequence (CDS3'); and
- a 3'-inverted terminal repeat (3'-ITR) sequence.

Preferably said first, second and third vector are independently a viral vector, preferably an adeno viral vector or adeno-associated viral (AAV) vector, preferably said first, second and third adeno-associated viral (AAV) vectors are selected from the same or different AAV serotypes, preferably the serotype is selected from the serotype 2, the serotype 8, the serotype 5, the serotype 7 or the serotype 9, serotype 7m8, serotype sh10; serotype 2(quad Y-F).

In one aspect, the present disclosure also provides a host cell transformed with the vector system as defined above.

Preferably the vector system or the host cell are for medical use, preferably for use in gene therapy, preferably for use in the treatment and/or prevention of a pathology or disease characterized by a retinal degeneration, a metabolic disorder, a blood disorder, a neurodegenerative disorder, hearing loss, channelopathy, lung disease, myopathy, heart disease, muscular dystrophy. In one aspect, the present disclosure provides a use of the vector system or the host cell as disclosed herein for the manufacture of a medicament for gene therapy, preferably for the treatment and/or prevention of a pathology or disease characterized by a retinal degeneration.

Preferably the retinal degeneration is inherited, preferably the pathology or disease is selected from the group consisting of: retinitis pigmentosa (RP), Leber congenital amaurosis (LCA), Stargardt disease (STGD), Usher disease (USH), Alstrom syndrome, congenital stationary night blindness (CSNB), macular dystrophy, occult macular dystrophy, a disease caused by a mutation in the *ABCA4* gene.

Preferably the vector system or the host cell is for use in the prevention and/or treatment of Duchenne muscular dystrophy, cystic fibrosis, hemophilia A, Wilson disease, Phenylketonuria, dysferlinopathies, Rett's syndrome, Polycystic kidney disease, Niemann-Pick type C, Huntington's disease.

In one aspect, the present disclosure also provides a pharmaceutical composition comprising the vector system or the host cell of the disclosure and pharmaceutically acceptable vehicle.

CLAIMS

1-A vector system to express a coding sequence in a cell, said coding sequence consisting of a first portion (CDS1), a second portion (CDS2) and optionally a third portion (CDS3), said vector system comprising:

a) a first vector comprising:

- said first portion of said coding sequence (CDS1),

-a first intein nucleotide sequence coding for a N-Intein, said sequence being located at the 3' end of CDS1; and

b) a second vector comprising:

- said second portion of said coding sequence (CDS2),

-a second intein nucleotide sequence coding for a C-Intein, said sequence being located at the 5' end of CDS2;

wherein when the first vector and the second vector are inserted in a cell, the protein product of the coding sequence is produced by protein splicing;

or optionally said vector system comprising:

a') a first vector comprising:

- said first portion of said coding sequence (CDS1),

-a first intein nucleotide sequence coding for a first N-Intein, said sequence being located at the 3' end of CDS1; and

b') a second vector comprising:

- said second portion of said coding sequence (CDS2),

-a second intein nucleotide sequence coding for a first C-Intein, said sequence being located at the 5' end of CDS2;

-a third intein nucleotide sequence coding for a second N-Intein, said sequence being located at the 3’ end of CDS2; and

c') a third vector comprising:

-said third portion of said coding sequence (CDS3)

-a fourth intein nucleotide sequence coding for a second C-Intein, said sequence being located at the 5’ end of CDS3;

wherein the first intein nucleotide sequence is different from the third intein nucleotide sequence when present and the second intein sequence is different from the fourth intein nucleotide sequence when present, wherein when the first vector, the second vector, and the third vector when present are inserted in a cell, the protein product of the coding sequence is produced by protein splicing, and wherein the coding sequence encodes a protein able to correct a retinal degeneration disorder.

2- The vector system according to claim 1, wherein the first intein, the second intein, the third intein and the fourth intein when present encodes for a split intein, preferably said split intein has a maximum length of 150 amino acids, more preferably said split intein is a DnaE or DnaB intein.

3- The vector system according to claim 1 or 2, wherein:

-the first intein nucleotide sequence encodes for an intein selected from the group consisting of: SEQ ID No 1, 3, 5, 7, 9, 11, 13 or a variant thereof or a fragment thereof or an homolog thereof;

-the second intein nucleotide sequence encodes for an intein selected from the group consisting of: SEQ ID No 2, 4, 6, 8, 10, 12, 14 or a variant thereof or a fragment thereof or an homolog thereof;

-the third intein nucleotide sequence when present encodes for an intein selected from the group consisting of: SEQ ID No 1, 3, 5, 7, 9, 11, 13 or a variant thereof or a fragment thereof or an homolog thereof;

-the fourth intein nucleotide sequence when present encodes for an intein selected from the group consisting of: SEQ ID No 2, 4, 6, 8, 10, 12, 14 or a variant thereof or a fragment thereof or an homolog thereof.

4- The vector system according to any one of the previous claims, wherein the first vector, the second vector and the third vector when present further comprise a promoter sequence operably linked to the 5'end portion of said first portion of the coding sequence (CDS1) or of said second portion of the coding sequence (CDS2) or of said third portion of the coding sequence (CDS3).

5- The vector system according to any one of the previous claims, wherein the first vector, the second vector and the third vector further comprise a 5'-terminal repeat (5'-TR) nucleotide sequence and a 3'-terminal repeat (3'-TR) nucleotide sequence, preferably the 5'-TR is a 5'-inverted terminal repeat (5'-ITR) nucleotide sequence and the 3'-TR is a 3'-inverted terminal repeat (3'-ITR) nucleotide sequence.

6- The vector system according to any one of the previous claims, wherein the first vector, the second vector and the third vector further comprise a poly-adenylation signal nucleotide sequence and/or wherein at least one of the first vector or the second vector or the third vector further comprises a nucleotide sequence coding for a degradation signal.

7- The vector system according to claim 6 wherein the degradation signal is selected from the group consisting of CL1, PB29, SMN, CIITA, ODc, ecDHFR or a fragment thereof.

8- The vector system according to any one of the previous claims, wherein the coding sequence is split into the first portion, the second portion and optionally the third portion, at a position consisting of a nucleophile amino acid which does not fall within a structural domain or a functional domain of the encoded protein product, wherein the nucleophile amino acid is selected from serine, threonine, or cysteine.

9- The vector system according to any one of previous claims, wherein at least one of the first vector, the second vector and the third vector when present further comprises at least one enhancer or regulatory nucleotide sequence, operably linked to the coding sequence.

10- The vector system according to any one of the previous claims, wherein the coding sequence encodes a protein able to correct an inherited retinal degeneration disorder, preferably selected from the group consisting of: retinitis pigmentosa (RP), Leber congenital amaurosis (LCA), Stargardt disease (STGD), Usher disease (USH), Alstrom syndrome, congenital stationary night blindness (CSNB), macular dystrophy, occult macular dystrophy, a disease caused by a mutation in the *ABCA4* gene.

11- The vector system according to any one of claims 1 to 10, wherein the coding sequence is the coding sequence of a gene selected from the group consisting of: ABCA4, MYO7A, CEP290, CDH23, EYS, PCDH15, CACNA1, SNRNP200, RP1, PRPF8, RP1L1, ALMS1, USH2A, GPR98, HMCN1.

12- The vector system according to any one of the previous claims comprising:

a) a first vector comprising in a 5'-3' direction:

- a 5'-inverted terminal repeat (5'-ITR) sequence;

- a promoter sequence;

- a 5' end portion of a coding sequence (CDS1), said 5'end portion being operably linked to and under control of said promoter;

- a first intein nucleotide sequence coding for a N-Intein; and

- a 3'-inverted terminal repeat (3'-ITR) sequence; and

b) a second vector comprising in a 5'-3' direction:

- a 5'-inverted terminal repeat (5'-ITR) sequence;

- a promoter sequence;

- a second intein nucleotide sequence coding for a C-Intein;

- a 3'end portion of the coding sequence (CDS2); and

- a 3'-inverted terminal repeat (3'-ITR) sequence;

or optionally comprising:

a') a first vector comprising in a 5'-3' direction:

- a 5'-inverted terminal repeat (5'-ITR) sequence;

- a promoter sequence;

- a 5' end portion of a coding sequence (CDS1'), said 5'end portion being operably linked to and under control of said promoter;

- a first intein nucleotide sequence coding for a first N-Intein ; and

- a 3'-inverted terminal repeat (3'-ITR) sequence; and

b') a second vector comprising in a 5'-3' direction:

- a 5'-inverted terminal repeat (5'-ITR) sequence;

- a promoter sequence;

- a second intein nucleotide sequence coding for a first C-Intein;

- the second portion of the coding sequence (CDS2'); and

- a third intein nucleotide sequence coding for a second N-intein;

- a 3'-inverted terminal repeat (3'-ITR) sequence; and

c') a third vector comprising in a 5'-3' direction:

- a 5'-inverted terminal repeat (5'-ITR) sequence;

- a promoter sequence;

- a fourth intein nucleotide sequence coding for a second C-Intein;

- the third portion of the coding sequence (CDS3'); and

- a 3'-inverted terminal repeat (3'-ITR) sequence.

13. The vector system according to any one of the previous claims wherein the coding sequence encodes the ABCA4 gene, preferably, said coding sequence is split at a nucleotide corresponding to aa Cys1150, Ser1168, Ser 1090 of the ABCA4 protein, and a split intein is inserted at the split point or the coding sequence encodes the CEP290 gene, preferably, said coding sequence is split at a nucleotide corresponding to aa Cys1076; Ser1275 of the CEP290 protein, preferably, the coding sequence encoding the CEP290 gene is split at a nucleotide sequence corresponding to aa Cys 929 and 1474; Ser 453 and Cys 1474 of said CEP290 protein, and two split inteins are inserted at the split points.

14- The vector system according to any one of the previous claims wherein said first, second and third vector when present are independently a viral vector, preferably an adeno viral vector or adeno-associated viral (AAV) vector, preferably said first, second and third adeno-associated viral (AAV) vectors are selected from the same or different AAV serotypes, preferably the serotype is selected from the serotype 2, the serotype 8, the serotype 5, the serotype 7 or the serotype 9, serotype 7m8, serotype sh10; serotype 2(quad Y-F).

15- A host cell transformed with the vector system according to any one of the previous claims.

16- The vector system according to any one of claims 1 to 14 or the host cell according to claim 15 for use in therapy.

17- Use of the vector system according to any one of claims 1 to 14 or the host cell according to claim 15 for the manufacture of a medicament for gene therapy, preferably for the treatment and/or prevention of a pathology or disease characterized by a retinal degeneration.

18- The use according to claim 17 wherein the retinal degeneration is inherited, preferably the pathology or disease is selected from the group consisting of: retinitis pigmentosa (RP), Leber congenital amaurosis (LCA), Stargardt disease (STGD), Usher disease (USH), Alstrom syndrome, congenital stationary night blindness (CSNB), macular dystrophy, occult macular dystrophy, a disease caused by a mutation in the *ABCA4* gene.

19- A pharmaceutical composition comprising the vector system according to any one of claims 1 to 14 or the host cell according to claim 15 and pharmaceutically acceptable vehicle.