



(51) International Patent Classification:

A61K 48/00 (2006.01)

C12N 15/90 (2006.01)

C12N 15/11 (2006.01)

(21) International Application Number:

PCT/EP2023/061582

(22) International Filing Date:

02 May 2023 (02.05.2023)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

22171218.5

02 May 2022 (02.05.2022)

EP

(71) Applicant: **FONDAZIONE TELETHON ETS** [IT/IT];
Via Varese 16B, 00185 Rome (IT).

(72) Inventors: **AURICCHIO, Alberto**; Fondazione Telethon ETS, Via Varese 16B, 00185 Rome (IT). **DELL'AQUILA, Fabio**; Fondazione Telethon ETS, Via Varese 16B, 00185 Rome (IT). **ESPOSITO, Federica**; Fondazione Telethon ETS, Via Varese 16B, 00185 Rome (IT). **FERLA, Rita**; Fondazione Telethon ETS, Via Varese 16B, 00185 Rome (IT). **LLADO SANTAEULARIA, Manel**; Fondazione Telethon ETS, Via Varese 16B, 00185 Rome (IT).

(74) Agent: **TURRI, Elisa** et al.; De Simone and Partners S.r.l.,
Via Vincenzo Bellini 20, 00198 Rome (IT).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ,

RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))
- with sequence listing part of description (Rule 5.2(a))
- in black and white; the international application as filed contained color or greyscale and is available for download from PATENTSCOPE

(54) Title: HOMOLOGY INDEPENDENT TARGETED INTEGRATION FOR GENE EDITING

(57) Abstract: The present invention relates to a method, preferably a homology independent targeted integration (HITI), of integrating an exogenous DNA sequence into a genome of a cell comprising contacting the cell with a donor nucleic acid comprising said exogenous DNA sequence, optionally one or more albumin exons, wherein said donor nucleic acid is flanked at 5' and 3' by inverted targeting sequences; a complementary strand oligonucleotide homologous to the targeting sequence and a nuclease that recognizes the targeting sequence, wherein said targeting sequence is located at the 3' end of the albumin gene in a region selected from intron 12, intron 13 and intron 14 of said albumin gene. The invention also relates to systems, vectors and pharmaceutical compositions comprising said donor nucleic acid and/or complementary strand oligonucleotide homologous to the targeting sequence and/or nuclease and to medical uses thereof.



HOMOLOGY INDEPENDENT TARGETED INTEGRATION FOR GENE EDITING

TECHNICAL FIELD

The present invention relates to a method of integrating an exogenous DNA sequence into a genome of a cell comprising contacting the cell with a donor nucleic acid, a complementary strand oligonucleotide homologous to the targeting sequence and a nuclease that recognizes the targeting sequence. The invention also relates to constructs, vectors, systems and pharmaceutical compositions comprising said donor nucleic acid and/or complementary strand oligonucleotide homologous to the targeting sequence and/or nuclease and to medical uses thereof.

BACKGROUND OF THE INVENTION

Genome Editing has emerged in the last years as a viable option for the treatment of inherited diseases. Genome editing uses an endonuclease, usually CRISPR/Cas9 [1, 2]. CRISPR-Cas9 is a ribonucleoprotein that binds a sequence called guide RNA (gRNA) and uses it to recognize the target DNA sequence by Watson-Crick base complementarity. This target DNA sequence must be adjacent to a protospacer-adjacent motif (PAM) sequence, which allows Cas9 to bind to the DNA and cleave the target sequence [3]. The RNA-based targeting of Cas9 facilitates its design for targeting different loci and even allows the targeting of 2 different sequences by delivering Cas9 and 2 different gRNAs to the same cell. After Cas9 is targeted to a particular location in the genome, it generates a double-strand break (DSB), which will be repaired by one of two repair mechanisms:

Non homologous end joining (NHEJ) is the most dominant mechanism in most cell types since it is active in all phases of the cell cycle and consists of the insertion or deletion of random bases in the site of the DSB in order to repair it. This random insertion or deletion (INDEL) often causes a change in the reading frame, thus knocking out the expression of the targeted gene [3].

Homology-Directed Repair (HDR) is a process that occurs mainly in the G and S2 phases of the cell cycle, and uses a homologous template, which can be provided by an external donor DNA or by the other allele, for precise correction of the DSB.[3]. Gene correction by HDR has been successfully used *in vitro* [4] and *in vivo* [5-8] , even in the absence of Cas9 [6]. However its efficiency *in vivo* is limited by the low activity of the homologous recombination pathway in differentiated cells [9]. Thus, there is a need in the field for alternative therapeutic gene replacement strategies which enable gene correction in tissues not undergoing active regeneration and in differentiated cells.

Homology-Independent Targeted Integration (HITI) has recently been developed [10, 11] to overcome the limitations of both allele-specific knock-out and gene correction by HDR. HITI uses a

donor DNA that is flanked by the same gRNA target sequences within the gene of interest. After Cas9 cleaves both the gene and the donor DNA, the NHEJ machinery of the cell can include the donor DNA in the repairing of the cleavage, with a surprisingly high (60-80%) rate of integration in the absence of INDELS. The possible inverted integration of the donor DNA is avoided by inverting its gRNA target sequences, so that Cas9 can recognize and cut again the target sequence if inverted integration occurs. Because HITI uses NHEJ, it is effective in terminally differentiated cells like neurons or tissues like liver independently of its regeneration potential (for instance both in adult and children tissues) [11]. In addition, HITI-mediated insertion of a wild-type copy of the therapeutic gene has the potential of being therapeutic independently of the specific disease-causing mutation and of the potential proliferative status of the target cells [11].

The present inventors have previously found that HITI can be used to convert the liver in a factory for systemic release of high levels of a therapeutic protein, which is desirable for therapy of many inherited and common conditions caused by loss-of-function or conditions where the factor to be replaced is secreted from the liver and/or has to reach other target organs through the blood to perform its function, like in hemophilia, LSDs, or diabetes, overcoming limitations of current available therapies as the low efficient enzyme replacement therapy, traditional gene therapy and gene editing.

Vectors based on adeno-associated viruses (AAVs) are the most frequently used for *in vivo* applications of gene therapy, because of their safety profile, wide tropism and ability to provide long-term transgene expression [12]. However, given the episomal status of AAV genomes, hepatic transgene expression from AAV can be lost over time in a developing liver or if there is hepatic damage [13] with limited success for instance in pediatric patients.

Hence there is a need for more stable and efficient hepatic transgene expression.

The HITI developed by the present inventors overcomes said limitations by inserting the coding sequence of a secreted protein of interest in the highly-transcribed Albumin locus [5-8], providing long-term expression of high levels of proteins secreted systemically while preserving the endogenous expression of the Albumin protein.

Mucopolysaccharidosis type VI (MPS VI) is a rare lysosomal storage disorder (LSD) that is caused by arylsulfatase B (ARSB) deficiency, resulting in widespread accumulation and urinary excretion of toxic glycosaminoglycans (GAGs). Clinically, the MPS VI phenotype is characterized by growth retardation, coarse facial features, skeletal deformities, joint stiffness, corneal clouding, cardiac

valve thickening, and organomegaly, with absence of primary cognitive impairment [14]. Therapies for MPS rely on normal lysosomal hydrolases being secreted and then up taken by most cells via the mannose-6-phosphate receptor pathway.

The present inventors previously demonstrated that a single systemic administration of a recombinant AAV vector serotype 8 (AAV2/8), which encodes ARSB under the transcriptional control of the liver-specific thyroxine-binding globulin (TBG) promoter (AAV2/8.TBG.hARSB), results in sustained liver transduction and phenotypic improvement in MPS VI animal models [15-21]. The present inventors also showed that this is at least as effective in MPS VI mice as weekly administrations of enzyme replacement therapy (ERT), which is the current standard of care for this condition [22-24]. The present inventors have recently initiated a phase I/II clinical trial (ClinicalTrials.gov Identifier: NCT03173521) to test both the safety and efficacy of this approach in MPS VI patients.

Haemophilia A (HemA) is a severe bleeding disorder caused by a deficiency or complete absence of the activity of coagulation factor VIII or 8 (FVIII or F8). It is the most common hereditary X-linked recessive coagulation disorder with an incidence of approximately 1 in 5,000 male live births worldwide.

About 50% of the cases of HemA are severe, i.e. have circulating FVIII levels less than 1%. [25]. The severe form of HemA is characterized clinically by spontaneous musculoskeletal and soft tissue bleeding as well as the inability to achieve hemostasis after trauma unless concentrates of clotting factor are infused. The current treatment is prophylactic administration of recombinant or plasma-derived clotting FVIII. Infusions are required frequently (two or three times per week) and can be burdensome. Moreover, they cannot prevent spontaneous bleeding and there is always a high risk of neutralizing anti-FVIII antibody (inhibitor) development (in about 25-30% of patients). Thus, gene therapy as an alternative holds great promise for a single-administration life-long cure.

The use of gene therapy for HemA has been under extensive investigation in the last 20 years after it was observed that even modest improvements in FVIII levels (by 1-2%) can produce a significant reduction in the risk of spontaneous bleeding events and reduce the need for FVIII replacement infusions. Additionally, gene therapy has a wide therapeutic range wherein gene expression does not need to be strictly regulated and has an easy quantifiable therapeutic endpoint (FVIII plasma levels). Several gene transfer strategies for FVIII replacement have been evaluated, and adeno-

associated viral (AAV) vectors are emerging as the most promising one because of the vectors' excellent safety profile and ability to direct long-term transgene expression from post-mitotic tissues such as the liver. However, HemA poses a great challenge to AAV gene therapy because of the size of the F8 gene coding sequence to be transferred (7 kb) that exceeds the canonical AAV cargo capacity of 4.7 kb. Previously a 5 kb expression cassette including a B-domain deleted (BDD) F8 and both short liver-specific promoter and a polyA signal has been packaged into AAV5 and shown to result in therapeutic levels of FVIII in mice and cynomolgus monkeys [26] as well as in HemA patients [27]. However, the genome of this vector is slightly oversized and is packaged into AAV capsids as a library of heterogeneous truncated genomes, which upon reconstitution in target cells result in ineffective transduction. The efficiency of oversize AAV vectors is lower compared to normal size and the quality of such a product with heterogeneous truncated genomes may preclude its further development. All of the AAV-based products under clinical investigation consist of B-domain deleted (BDD) versions of the F8 transgene which are ~4.4 kb in size [28]. Still such large transgene leaves limited space in the vector for the needed regulatory elements, thus restricting the choice of promoters and polyA signals. Moreover, all these vector genomes are on the verge of AAV's normal cargo capacity and at risk of being improperly packaged as a library of heterogeneous truncated genomes. In spite of the ability of such oversize vectors to successfully express large proteins, their long-term efficiency and safety are still to be confirmed [28-32].

HITI in hepatocytes at the highly transcribed Albumin locus has the potential to overcome several limitations of the otherwise safe and effective liver gene therapy with AAV, including: i. levels of transgene expression, which are particularly high from the Albumin locus; ii. Stability of transgene expression guaranteed by the insertion of the therapeutic coding sequence at a genomic locus, which would be replicated should hepatocyte cell loss occur and in developing liver, making it possible to make the therapy available to pediatric patients.

A previous approach for integrating an exogenous DNA sequence into a genome of a cell based on HITI is disclosed in WO2020079033, wherein in particular the second exon of the albumin gene was targeted. In that approach, the targeted albumin locus was disrupted by insertion of the donor DNA carrying the transgene of interest.

There is still the need for gene therapy strategies for diseases requiring stable systemic expression of therapeutic proteins.

SUMMARY OF THE INVENTION

The inventors have now found a new approach for integrating an exogenous DNA sequence into a genome of a cell taking advantage of the HITI system targeting the Albumin gene without disrupting its expression. Proof of efficacy has been provided with ARSB which encodes for arylsulfatase B, the lysosomal enzyme deficient in mucopolysaccharidosis VI (MPS VI), in MPS VI mice, and with the F8 CodopV3 transgene coding for factor VIII, which is missing in Haemophilia A, in hemophilic mice.

The inventors now found that integration of a transgene, such as ARSB or F8, at the 3' end of mouse Albumin (mAlb) through a novel HITI system resulted in an increase in the levels and/or activity of the missing enzyme. In particular, integration of ARSB resulted in circulating supraphysiological levels of enzyme with two tested doses while one dose induces lower levels of circulating enzymes, and phenotypic improvement up to 36 weeks after neonatal delivery in MPS VI mice while F8 activity levels in haemophilic mice treated with the system of the invention was increased of 20% compared to unaffected controls. The inventors also demonstrated integration of a reporter gene at the 3' end of human Albumin (hALB) *in vitro*.

Overall, the novel HITI results in stable, high expression levels of therapeutic transgenes from liver with hepatocyte proliferation.

The present invention relies on insertion of the sequence of interest within the locus of a gene expressed at high levels in the liver, i.e. albumin. The gene of interest is expressed under the Albumin promoter resulting in high levels of expression of the gene of interest, albeit from a relatively small number of cells within the liver parenchima, therefore expression of the gene of interest is sufficiently high to achieve a therapeutic effect. Furthermore, since the gene of interest is stably integrated in the liver genome, upon tissue regeneration (in children or upon liver damage), expression of the gene of interest is not lost. The albumin locus is targeted with a strategy that allows preservation of expression of the Albumin gene.

It is an object of the invention a method of integrating an exogenous DNA sequence into a genome of a cell comprising contacting the cell with:

a) a donor nucleic acid comprising:

- said exogenous DNA sequence
- optionally one or more albumin exons

wherein said donor nucleic acid is flanked at 5' and 3' by inverted targeting sequences;

- b) a complementary strand oligonucleotide homologous to a targeting sequence and
- c) a nuclease that recognizes said targeting sequence,

wherein said targeting sequence is located at the 3' end of the albumin gene in a region selected from intron 9, intron 11, intron 12, intron 13, and intron 14 of said albumin gene.

In the present invention, the donor nucleic acid preferably comprises one or more albumin exons and said exon is exon 10 and/or exon 11 and/or exon 12 and/or exon 13 and/or exon 14 or fragments thereof.

It is also an object of the invention a method of integrating an exogenous DNA sequence into a genome of a cell comprising contacting the cell with:

- a) a donor nucleic acid comprising:
 - said exogenous DNA sequence
 - optionally one or more albumin exons

wherein said donor nucleic acid is flanked at 5' and 3' by inverted targeting sequences;

- b) a complementary strand oligonucleotide homologous to a targeting sequence and
- c) a nuclease that recognizes said targeting sequence,

wherein said targeting sequence is located at the 3' end of the albumin gene in a region selected from intron 12, intron 13, and intron 14 of said albumin gene.

In the present invention, the donor nucleic acid preferably comprises one or more albumin exons and said exon is exon 13 and/or exon 14 or fragments thereof.

In a preferred embodiment, said albumin exon(s) is (are) present and it is (they are) exon 10 and/or exon 11 and/or exon 12 and/or exon 13 and/or exon 14 or fragments thereof. Preferably, it is located at the 5' end of the exogenous DNA sequence from which it can be separated by a ribosomal skipping sequence.

In a preferred embodiment, said albumin exon is present and it is exon 13 and/or exon 14 or fragments thereof. Preferably, it is located at the 5' end of the exogenous DNA sequence from which it can be separated by a ribosomal skipping sequence. In the present invention, the donor nucleic acid preferably comprises one or more albumin exons and said exon is exon 10 and/or exon 11 and/or exon 12 and/or exon 13 and/or exon 14 or fragments thereof.

In the present invention, the donor nucleic acid preferably comprises one or more albumin exons and said exon is exon 13 and/or exon 14 or fragments thereof.

Said albumin exons can be from albumin genes of any origin, preferably they are from human or murine albumin gene.

Preferably, the complementary strand oligonucleotide homologous to a targeting sequence is a guide RNA that hybridizes to a targeting sequence, or to its complementary strand, located within intron 9, intron 11, intron 12, intron 13 or intron 14 of the albumin gene, preferably said guide RNA being adjacent to a protospacer-adjacent motif (PAM) sequence.

Preferably, the complementary strand oligonucleotide homologous to a targeting sequence is a guide RNA that hybridizes to a targeting sequence, or to its complementary strand, located within intron 12, intron 13 or intron 14 of the albumin gene, preferably said guide RNA being adjacent to a protospacer-adjacent motif (PAM) sequence.

Preferably, the targeting sequence is a guide RNA (gRNA) target site and said complementary strand oligonucleotide homologous to the targeting sequence is a guide RNA that hybridizes to a targeting sequence, or to its complementary strand, located within intron 12, intron 13 or intron 14. Said oligonucleotide thus guides the nuclease to cut within the intron 12, 13 or 14 of the Albumin gene. Preferably, said guide RNA is adjacent to a protospacer-adjacent motif (PAM) sequence.

Preferably, the targeting sequence is a guide RNA (gRNA) target site and said complementary strand oligonucleotide homologous to the targeting sequence is a guide RNA that hybridizes to a targeting sequence, or to its complementary strand, located within intron 9, intron 11, intron 12, intron 13 or intron 14. Said oligonucleotide thus guides the nuclease to cut within the intron 9, intron 11, intron 12, intron 13 or intron 14 of the Albumin gene. Preferably, said guide RNA is adjacent to a protospacer-adjacent motif (PAM) sequence.

In the context of the present invention, the albumin gene is preferably a human or murine gene.

Preferably, said targeting sequence comprises or has essentially a sequence having at least 95% of identity to any one of SEQ ID Nos. 1-2, 9-18, 54, 92-98, or fragments thereof, and/or the complementary strand oligonucleotide homologous to the targeting sequence comprises or has essentially a sequence having at least 95% of identity to any one of SEQ ID Nos. 1-2, 9-18, 54, 92-98, or fragments thereof.

In this embodiment, the guide RNA comprising or having at least 95% of identity to SEQ ID NO: 2, SEQ ID N. 10, SEQ ID N.12, SEQ ID N. 14, SEQ ID N.16, SEQ ID NO: 17, SEQ ID NO:18, any one of SEQ ID NO:54, 92-98, or fragments thereof, can bind to a target sequence comprising or having at least 95% of identity to SEQ ID N. 1, SEQ ID N. 9, SEQ ID N. 11, SEQ ID N. 13 , SEQ ID N. 15, SEQ ID N.16, SEQ ID NO: 17 or SEQ ID NO:18, or fragments thereof, or to its complementary strand.

In this embodiment, the guide RNA comprising or having at least 95% of identity to SEQ ID N. 2, SEQ ID N. 10, SEQ ID N. 12, SEQ ID N. 14, SEQ ID N. 16 , SEQ ID N. 17 or 18 or fragments thereof, can bind to a target sequence comprising or having at least 95% of identity to SEQ ID NO: 1, SEQ ID N. 9, SEQ ID N.11, SEQ ID N.13 or SEQ ID N.15, SEQ ID N. 16 , SEQ ID N. 17 or 18 or fragments thereof, or to its complementary strand.

Preferably, said fragments are at least 15 nucleotides long.

In an embodiment, the targeting sequences and the respective gRNAs are as described in Tables 1 and 3.

The present invention also includes embodiments wherein the sequences mentioned above, i.e. SEQ ID N.1, 2, 9-18, 54, 92-98, have a reverse orientation, i.e. from 3' to 5'.

The present invention also includes embodiments wherein the sequences mentioned above, i.e. SEQ ID N.1, 2, 9-18, 54, 92-98, are of RNA.

Said gRNAs are also objects of the invention.

Further objects of the invention are isolated guide ribonucleic acid (gRNA) comprising or consisting of a sequence that is substantially complementary or perfectly annealing to a sequence selected from SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17 or SEQ ID NO:18, 54, 92-98 and to portions thereof at least 15 nucleotides long.

Preferably, the albumin gene is from human or mouse.

Said albumin gene introns can be in an albumin gene of any origin, preferably they are in human or murine albumin gene.

In an embodiment, the complementary strand oligonucleotide homologous to the targeting sequence is under the control of a promoter, preferably the U6 promoter.

In the present invention, guide RNA or gRNA can be used as a synonym of complementary strand oligonucleotide homologous to the targeting sequence.

In a preferred embodiment, said exogenous DNA sequence is a coding sequence of the Arylsulfatase B (ARSB) gene. Preferably, said ARSB coding sequence is human. Preferably it comprises or has essentially a sequence having at least 95% of identity to SEQ ID NO 33. The coding sequence can codify for a variant of Arylsulfatase B (ARSB), for example it can comprise additions, deletions or substitutions with respect to the coding sequence of the wild type Arylsulfatase B (ARSB) gene as long as these protein variants retain substantially the same relevant functional activity as the original ARSB. The coding sequence can also codify for a fragment of Arylsulfatase B (ARSB) as long as this fragment retains substantially the same relevant functional activity as the original ARSB.

Suitably, the coding sequence may be codon optimized for expression in human.

In a further preferred embodiment, said exogenous DNA sequence is a coding sequence of the Factor 8 (F8) gene or of the B domain deleted (BDD) F8 gene. Preferably, said BDD F8 coding sequence is human, more preferably it comprises or has essentially a sequence having at least 95% of identity to SEQ ID NO:36 or 55. Preferably the F8 coding sequence comprises or has essentially a sequence having at least 95% of identity to SEQ ID NO 36 or 55. The coding sequence can codify for a variant of BDD F8 or of F8, for example it can comprise additions, deletions or substitutions with respect to the coding sequence of the wild type BDD F8 gene or of the wild type F8 gene as long as these protein variants retain substantially the same relevant functional activity as the original BDD F8 or F8, respectively. The coding sequence can also codify for a fragment of BDD F8 or of F8 as long as this fragment retains substantially the same relevant functional activity as the original BDD F8 or F8, respectively.

In a further preferred embodiment, said exogenous DNA sequence is a coding sequence of a gene which, in a recessive inherited disease patient, is interested by a mutation which causes loss-of function. The liver may thus be used as a therapeutical target.

In a further preferred embodiment, said exogenous DNA sequence is a coding sequence of F9 gene or genes which are mutated in α 1-anti-trypsin (AAT) deficiency, Wilson disease, OAT deficiency, MPSVII.

The inverted targeting sequences in the context of the present invention are positioned one upstream and one downstream of the donor nucleic acid, which is the DNA construct that is cut and

then integrated in the targeted locus. An inverted targeting sequence is the same exact sequence that recognizes the guide RNA in the target genomic locus, i.e. the targeting sequence, but it is inverted or reversed with respect to the genomic sequence. Inverted or reverse means that if the targeting sequence has a specific 5'-3' sequence, the inverted targeting sequence has the same sequence but with orientation 3'-5'. Therefore, an inverted targeting sequence is complementary to the guide RNA but inverted. This allows to obtain a mono-directional integration, as known in the HITI method. In short, in the case that the donor DNA is integrated in the opposite direction the nuclease, such as Cas9, would be able to recognize again its target site and cleave it. Upon integration in the correct orientation, the nuclease would no longer be able to cleave the target site. Said inverted targeting sequence is preferably an inverted sequence with respect to a target sequence located at the 3' end of the albumin gene in a region selected from intron 9, intron 11, intron 12, intron 13 and intron 14. Preferably, each of said inverted targeting sequence is linked at its 3' to a protospacer-adjacent motif (PAM) sequence.

Said exogenous DNA sequence may also comprise a reporter gene, preferably said reporter gene is selected from at least one of Discosoma Red, Green Fluorescent Protein (GFP), a Red Fluorescent protein (RFP), a luciferase, a β -galactosidase and a β -glucuronidase.

In an embodiment, said donor nucleic acid further comprises one or more of:

- a post-transcriptional regulatory element, preferably localized at the 3' end of the exogenous DNA sequence;
- a transcription termination sequence preferably localized at the 3' end of the post-transcriptional regulatory element or at the 3' end of the exogenous DNA sequence;
- a splice acceptor sequence, preferably localized at the 3' end of the donor nucleic acid, for example linked to an albumin exon, if present;
- a ribosomal skipping sequence, preferably localized between the exogenous DNA sequence and the albumin exon(s).

In an embodiment, said donor nucleic acid further comprises one or more of:

- a splice acceptor sequence, preferably localized at the 5' end of the donor nucleic acid, for example it is linked to an albumin exon, if present;
- a ribosomal skipping sequence, preferably localized between the albumin exon(s) and the exogenous DNA sequence;

- a post-transcriptional regulatory element, preferably localized at the 3' end of the exogenous DNA sequence;
- a transcription termination sequence preferably localized at the 3' end of the post-transcriptional regulatory element or at the 3' end of the exogenous DNA sequence.

Preferably, the ribosomal-skipping sequence is a T2A, P2A, E2A, F2A, preferably T2A sequence. This sequence when expressed in a cell allows to separate the protein of interest from the albumin.

Preferably said post-transcriptional regulatory element is the Woodchuck hepatitis virus post-transcriptional regulatory element (WPRE).

Preferably said transcription termination sequence is a poly-adenylation signal sequence, preferably the bovine growth hormone polyA (BGH polyA), most preferably a short synthetic polyA.

As mentioned above, the donor DNA sequence is flanked at 5' and 3' by the same gRNA target site that the gRNA recognizes, but inverted (e.g. an inverted target site).

Preferably, the targeting sequence comprises or has essentially a sequence having at least 95% of identity to one of the sequences herein mentioned or functional fragments thereof and/or the complementary strand oligonucleotide homologous to the targeting sequence comprises or has essentially a sequence having at least 95% of identity to one of the sequences herein mentioned or functional fragments thereof.

Preferably, the inverted targeting sequence comprises or has essentially a sequence having at least 95% of identity to SEQ ID NO: 20, 54 or 77 or to SEQ ID NO:2 or 1 or to any one of SEQ ID NO:9-18, 92-98 or 54.

Preferably said donor nucleic acid comprises:

- an inverted targeting sequence with its protospacer-adjacent motif (PAM) sequence;
- a splice acceptor sequence
- one or more albumin exons, preferably one or more exons selected from exon 13 and 14;
- a ribosomal skipping sequence, preferably T2A;
- the exogenous DNA sequence, preferably the coding sequence of the marker dsRed, the human ARSB gene or the BDD F8 gene;
- a transcription termination sequence; and
- a further inverted targeting sequence with its protospacer-adjacent motif (PAM) sequence.

Preferably said donor nucleic acid comprises:

- an inverted targeting sequence with its protospacer-adjacent motif (PAM) sequence;
- a splice acceptor sequence
- one or more albumin exons, preferably one or more exons selected from exon 10, exon 11, exon 12, exon 13, and exon 14;
- a ribosomal skipping sequence, preferably T2A;
- the exogenous DNA sequence, preferably the coding sequence of the marker dsRed, the human ARSB gene or the BDD F8 gene;
- a transcription termination sequence; and
- a further inverted targeting sequence with its protospacer-adjacent motif (PAM) sequence.

Preferably, said elements are in the 5'-3' order as listed but other orders may be equally suitable.

Preferably, said transcription termination sequence comprises or has essentially a sequence having at least 95% of identity to SEQ ID NO 26 or SEQ ID N.37 or SEQ ID NO:48 or SEQ ID NO:65.

Preferably, said ribosomal skipping sequence comprises or has essentially a sequence having at least 95% of identity to SEQ ID NO 23 or 63.

Preferably, said albumin exon comprises or has essentially a sequence having at least 95% of identity to SEQ ID NO 22 and/or 78 and/or 79.

Preferably, said splice acceptor sequence comprises or has essentially a sequence having at least 95% of identity to SEQ ID NO 21.

Preferably, said inverted targeting sequence comprises or has essentially a sequence having at least 95% of identity to SEQ ID NO 1 or 2 or 20 or 54 or 77, preferably without the PAM sequence.

Said nuclease can be provided as a protein or as a nucleic acid coding for said nuclease. Said nucleic acid can be DNA or RNA, for example it can be the mRNA of a nuclease or it can be a cDNA or the DNA coding sequence of a nuclease or a DNA construct coding for the nuclease.

Preferably, said nucleic acid coding for a nuclease is a DNA construct comprising a nucleic acid coding for Cas9 or spCas9 preferably under the control of a tissue specific promoter, e.g. a liver specific promoter like a liver hybrid liver promoter (HLP). Said construct may further comprise a poly A, conveniently a short synthetic polyA (synt. polyA). All such elements are well known in the art and may have conventional nucleotide sequences. An exemplary DNA construct coding for a nuclease comprises or has essentially a sequence having at least 95% of identity to SEQ ID NO 43, 47 or 52.

Said nuclease is preferably selected from: a CRISPR nuclease, a TALEN, a DNA-guided nuclease, a meganuclease, and a Zinc Finger Nuclease, preferably said nuclease is a CRISPR nuclease selected from the group consisting of: Cas9, Cpf1, Cas12b (C2c1), Cas13a (C2c2), Cas3, Csf1, Cas13b (C2c6), and C2c3 or variants thereof such as SaCas9, VQR-Cas9-HF1 or dCas9.

Preferably, the cell is contacted with a nucleic acid encoding for said nuclease, preferably said nucleic acid coding for said nuclease is under the control of a tissue specific promoter, e.g. a liver specific hybrid liver promoter (HLP).

Preferably, the nucleic acid coding for said nuclease is under the control of a tissue specific promoter, most preferably a liver specific promoter, for instance a hepatocyte specific promoter, e.g. a liver specific hybrid liver promoter (HLP).

Suitably, said donor nucleic acid, said complementary strand oligonucleotide homologous to a targeting sequence and said nucleic acid coding for said nuclease are comprised in DNA constructs. Preferably, a first DNA construct comprises the donor nucleic acid and the complementary strand oligonucleotide homologous to a targeting sequence and a second DNA construct comprises the nucleic acid coding for the nuclease that recognizes said targeting sequence. Alternatively, a first DNA construct comprises the donor nucleic acid and a second DNA construct comprises the complementary strand oligonucleotide homologous to a targeting sequence and the nucleic acid coding for the nuclease that recognizes said targeting sequence. As a further alternative, three constructs are provided: a first construct comprising the donor nucleic acid, a second construct comprising the complementary strand oligonucleotide homologous to a targeting sequence and a third construct comprising the nucleic acid coding for the nuclease that recognizes said targeting sequence.

Said constructs are also objects of the invention.

Preferably, one or more of said DNA constructs are comprised in a vector, preferably a viral vector, still preferably a lentiviral vector or an adeno-associated vector. Alternatively, all or some of said DNA constructs may be inserted into a non-viral vector, wherein said non-viral vector is selected from a polymer-based, particle-based, lipid-based, peptide-based delivery vehicle or combinations thereof, such as cationic polymers, micelles, liposomes, exosomes, microparticles and nanoparticles including lipid nanoparticles (LNP).

Said vectors are also object of the invention.

Said complementary strand oligonucleotide, said donor nucleic acid and said nucleic acid encoding the nuclease can be comprised in one or more viral or non-viral vectors, preferably said viral vector being selected from: an adeno-associated virus, a lentivirus, a retrovirus and an adenovirus. This means that they can be in the same or in different vectors. Preferably the cell is selected from the group consisting of: liver cells, one or more of lymphocytes, monocytes, neutrophils, eosinophils, basophils, endothelial cells, epithelial cells, hepatocytes, osteocytes, platelets, adipocytes, cardiomyocytes, neurons, retinal cells, smooth muscle cells, skeletal muscle cells, spermatocytes, oocytes, and pancreas cells, induced pluripotent stem cells (iPSCs), stem cells, hematopoietic stem cells, hematopoietic progenitor stem cells, preferably the cell is an hepatocyte of a subject. Another object of the invention is a cell obtainable by the above defined method.

Said cell can be for use as a medicament or for use in treating a genetic disease or for use in treating recessive inherited and common diseases, preferably said diseases comprising diabetes, Lysosomal storage diseases comprising mucopolysaccharidoses (MPSI, MPSII, MPSIIIA, MPSIIIB, MPSIIIC, MPSIVA, MPSIVB, MPSVI, MPSVII), sphingolipidoses (Fabry's Disease, Gaucher Disease, Nieman-Pick Disease, GM1 Gangliosidosis), lipofuscinoses (Batten's Disease and others) and mucolipidoses; gyrate atrophy of the choroid and retina, adenylosuccinate deficiency, hemophilia A and B, ALA dehydratase deficiency, adrenoleukodystrophy.

The cell of the invention can be for use in treating a diseases wherein both the mutant and wildtype alleles are replaced with a correct copy of the gene provided by the donor DNA or for use for the treatment of a recessive inherited and common disease due to loss-of-function, preferably said disease being selected from haemophilia, diabetes, Lysosomal storage diseases comprising mucopolysaccharidoses, such as MPSI, MPSII, MPSIIIA, MPSIIIB, MPSIIIC, MPSIVA, MPSIVB, MPSVI and MPSVII, sphingolipidoses, such as Fabry's Disease, Gaucher Disease, Nieman-Pick Disease and GM1 Gangliosidosis, lipofuscinoses, such as Batten's Disease, and mucolipidoses; gyrate atrophy of the choroid and retina, adenylosuccinate deficiency, hemophilia A and B, ALA dehydratase deficiency, adrenoleukodystrophy.

Therefore the cell of the invention may be used for treating haemophilia, diabetes, Lysosomal storage diseases comprising mucopolysaccharidoses, such as MPSI, MPSII, MPSIIIA, MPSIIIB, MPSIIIC, MPSIVA, MPSIVB, MPSVI and MPSVII, sphingolipidoses, such as Fabry's Disease, Gaucher Disease, Nieman-Pick Disease and GM1 Gangliosidosis, lipofuscinoses, such as Batten's Disease, and

mucopolysaccharidoses; gyrate atrophy of the choroid and retina, adenylosuccinate deficiency, hemophilia A and B, ALA dehydratase deficiency or adrenoleukodystrophy.

The cell obtainable according to the invention expresses the exogenous DNA sequence. Suitably, the cell obtainable according to the invention expresses also the full-length Albumin coding sequence.

When the donor nucleic acid contacts the cell it is typically inserted into the target gene via non-homologous end joining.

A further object of the invention is a system comprising:

a) a donor nucleic acid comprising:

- an exogenous DNA sequence
- optionally one or more albumin exons

wherein said donor nucleic acid is flanked at 5' and 3' by inverted targeting sequences;

b) a complementary strand oligonucleotide homologous to the targeting sequence and

c) a nuclease that recognizes said targeting sequence,

wherein said targeting sequence is located at the 3' end of the albumin gene in a region selected from intron 9, intron 11, intron 12, intron 13 and intron 14.

A further object of the invention is a system comprising:

a) a donor nucleic acid comprising:

- an exogenous DNA sequence
- optionally one or more albumin exons

wherein said donor nucleic acid is flanked at 5' and 3' by inverted targeting sequences;

b) a complementary strand oligonucleotide homologous to the targeting sequence and

c) a nuclease that recognizes said targeting sequence,

wherein said targeting sequence is located at the 3' end of the albumin gene in a region selected from intron 12, intron 13 and intron 14.

Said nuclease can be provided as a protein or as a nucleic acid coding for said nuclease. Said nucleic acid can be DNA or RNA, for example it can be the mRNA of a nuclease or it can be a cDNA or the DNA coding sequence of a nuclease or a DNA construct coding for the nuclease.

In an embodiment, said nuclease is codified by a nucleic acid and said donor nucleic acid, said complementary strand oligonucleotide and said nucleic acid codifying for the nuclease are located on DNA constructs, preferably said donor nucleic acid and said complementary strand oligonucleotide are located on the same DNA construct while said nucleic acid codifying for the nuclease is located on a separate DNA construct. Preferably, said construct comprising said donor nucleic acid and said complementary strand oligonucleotide comprises or has essentially a sequence having at least 95% of identity to a sequence comprising the following sequences: SEQ ID N. 20, SEQ ID N.21, SEQ ID N.22, SEQ ID N.23, SEQ ID N.33, SEQ ID N.26, SEQ ID N.20, SEQ ID N.27, SEQ ID N.1 and SEQ ID N.28.

In another embodiment, said construct comprising said donor nucleic acid comprises or has essentially a sequence having at least 95% of identity to a sequence comprising the following sequences: SEQ ID N.20, SEQ ID N.21, SEQ ID N.22, SEQ ID N.23, SEQ ID N.36, SEQ ID N.37, SEQ ID N.20.

A construct comprising said donor nucleic acid and said complementary strand oligonucleotide can comprise or have essentially a sequence having at least 95% of identity to SEQ ID N. 34.

Said construct comprising said donor nucleic acid can comprise or have essentially a sequence having at least 95% of identity to SEQ ID N.38.

In a preferred embodiment the donor nucleic acid and/or the exogenous DNA sequence and/or the albumin exons and/or the inverted targeting sequence and/or the targeting sequences and/or the complementary strand oligonucleotide and/or the nuclease and/or wherein the intron are as defined above or herein.

Another object of the invention is a process for preparing a viral vector particle comprising introducing such DNA constructs into a host cell, and obtaining the viral vector particle is also an object of the invention.

In a preferred embodiment the donor nucleic acid and/or the exogenous DNA sequence and/or the targeting sequences and/or the complementary strand oligonucleotide and/or the nuclease are as defined above.

In a preferred embodiment the donor nucleic acid and/or the exogenous DNA sequence and/or the albumin exons and/or the inverted targeting sequence and/or the targeting sequences and/or the complementary strand oligonucleotide and/or the nuclease and/or wherein the intron are as defined above or herein.

In the context of the present invention, the donor nucleic acid and/or the complementary strand oligonucleotide and/or the nucleic acid encoding the nuclease are comprised in one or more viral or non-viral vector, preferably said viral vector being selected from: an adeno-associated virus, a retrovirus, an adenovirus and a lentivirus; said non-viral vector being preferably selected from non-viral vector is selected from a polymer-based, particle-based, lipid-based, peptide-based delivery vehicle or combinations thereof, such as cationic polymers, micelles, liposomes, exosomes, microparticles and nanoparticles including lipid nanoparticles (LNP).

Preferably, a first vector comprises the donor nucleic acid and the complementary strand oligonucleotide homologous to a targeting sequence and a second vector comprises the nucleic acid coding for the nuclease that recognizes said targeting sequence. Alternatively, a first vector comprises the donor nucleic acid and a second vector comprises the complementary strand oligonucleotide homologous to a targeting sequence and the nucleic acid coding for the nuclease that recognizes said targeting sequence. As a further alternative, three vectors are provided: a first vector comprising the donor nucleic acid, a second vector comprising the complementary strand oligonucleotide homologous to a targeting sequence and a third vector comprising the nucleic acid coding for the nuclease that recognizes said targeting sequence.

Preferably, the system comprises a first vector comprising a nucleic acid expressing a nuclease and a second vector comprising the donor nucleic acid and the complementary strand oligonucleotide homologous to the targeting sequence, wherein such elements are as defined above or herein.

Preferably the system comprises a first vector comprising the donor nucleic acid and a second vector comprising the complementary strand oligonucleotide homologous to a targeting sequence and the nucleic acid coding for the nuclease, wherein such elements are as defined above or herein. The system according to the invention is preferably for medical use, preferably for use in treating a genetic disease or for use in treating inherited and common diseases due to loss-of-function, preferably said diseases comprising diabetes, Lysosomal storage diseases comprising mucopolysaccharidoses (MPSI, MPSII, MPSIIIA, MPSIIIB, MPSIIIC, MPSIVA, MPSIVB, MPSVI, MPSVII), sphingolipidoses (Fabry's Disease, Gaucher Disease, Nieman-Pick Disease, GM1 Gangliosidosis),

lipofuscinoses (Batten's Disease and others) and mucopolysaccharidoses; gyrate atrophy of the choroid and retina, adenylosuccinate deficiency, hemophilia A and B, ALA dehydratase deficiency, adrenoleukodystrophy.

The system of the invention can be for use in the treatment of a diseases wherein both the mutant and wildtype alleles are replaced with a correct copy of the gene provided by the donor DNA or for use for the treatment of a recessive inherited and common disease due to loss-of-function, preferably said disease being selected from haemophilia, diabetes, Lysosomal storage diseases comprising mucopolysaccharidoses, such as MPSI, MPSII, MPSIIIA, MPSIIIB, MPSIIIC, MPSIVA, MPSIVB, MPSVI and MPSVII, sphingolipidoses, such as Fabry's Disease, Gaucher Disease, Nieman-Pick Disease and GM1 Gangliosidosis, lipofuscinoses, such as Batten's Disease, and mucopolysaccharidoses; gyrate atrophy of the choroid and retina, adenylosuccinate deficiency, hemophilia A and B, ALA dehydratase deficiency, adrenoleukodystrophy.

Therefore the system of the invention may be used for treating haemophilia, diabetes, Lysosomal storage diseases comprising mucopolysaccharidoses, such as MPSI, MPSII, MPSIIIA, MPSIIIB, MPSIIIC, MPSIVA, MPSIVB, MPSVI and MPSVII, sphingolipidoses, such as Fabry's Disease, Gaucher Disease, Nieman-Pick Disease and GM1 Gangliosidosis, lipofuscinoses, such as Batten's Disease, and mucopolysaccharidoses; gyrate atrophy of the choroid and retina, adenylosuccinate deficiency, hemophilia A and B, ALA dehydratase deficiency or adrenoleukodystrophy.

A further object of the invention is a vector that comprises the donor nucleic acid and/or the complementary strand oligonucleotide homologous to the targeting sequence and/or a nucleic acid coding for a nuclease that recognizes the targeting sequence as defined above or herein.

In the context of the present invention, preferably the donor nucleic acid and/or the exogenous DNA sequence and/or the albumin exons and/or the inverted targeting sequence and/or the targeting sequences and/or the complementary strand oligonucleotide and/or the nuclease and/or wherein the intron are as defined above or herein.

Preferably, the vector is a viral vector, preferably a lentiviral vector or an adeno-associated vector, or a non-viral vector, preferably selected from a polymer-based, particle-based, lipid-based, peptide-based delivery vehicle or combinations thereof, such as cationic polymers, micelles, liposomes, exosomes, microparticles and nanoparticles including lipid nanoparticles (LNP). Preferably AAV2/8 vectors are used.

In a preferred embodiment of the invention, one vector comprising a nucleic acid expressing Cas9 is used together with a second vector comprising the donor DNA and the complementary strand oligonucleotide homologous to the targeting sequence, i.e. the gRNA, as defined above.

In an alternative preferred embodiment of the invention, one vector comprising a nucleic acid expressing Cas9 and the complementary strand oligonucleotide homologous to the targeting sequence, i.e. the gRNA is used together with a second vector comprising the donor DNA, as defined above.

Preferably, a first vector comprises a nucleic acid coding for Cas9 or spCas9 preferably under the control of a tissue specific promoter, e.g. a liver specific promoter like a liver hybrid liver promoter (HLP). Said vector may further comprise a poly A, conveniently a short synthetic polyA (synt.polyA). Preferably, a second vector comprises the gRNA expression cassette and the donor DNA as defined above. Preferably, the gRNA expression cassette comprises the gRNA as defined above under the U6 promoter. Preferably the donor DNA is flanked at 3' and 5' by the inverted targeting sequences, preferably linked to the respective PAM.

Preferably, the vector is a viral vector, preferably the viral vector is a lentiviral vector, an adeno-associated virus vector, an adenoviral vector, a retroviral vector, a polio viral vector, a murine Maloney-based viral vector, an alpha viral vector, a pox viral vector, a herpes viral vector, a vaccinia viral vector, a baculoviral vector, or a parvoviral vector, preferably the adeno-associated virus is AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8 AAV9, AAV 10, AAVSH19, AAVPHP.BAAV2, AAV9, AAV1, AAVSH19, AAVPHP.B, AAV8, AAV6.

Preferably the viral vector or vector further comprises 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, preferably the ITRs derive from the same virus serotype or from different virus serotypes, preferably the virus is an AAV, preferably of serotype 2.

In an embodiment, said viral vector comprising the gRNA expression cassette and the donor DNA further comprises a 5' inverted terminal repeat (ITR) sequence, preferably of AAV, preferably localized at the 5' end of the construct comprising the gRNA expression cassette and the donor DNA and a 3' inverted terminal repeat (ITR) sequence, preferably of AAV, preferably localized at the 3' end of the construct comprising the gRNA expression cassette and the donor DNA.

Preferably, said ITR comprises or has a sequence having at least 95% of identity to SEQ ID NO 110, SEQ ID NO 29 or 66.

Preferably said viral vector comprising the gRNA expression cassette and the donor DNA comprises:

- an AAV 5'-inverted terminal repeat (5'-ITR) sequence;
- an inverted targeting sequence linked to its protospacer-adjacent motif (PAM) sequence;
- a splice acceptor sequence;
- one or more albumin exons, preferably one or more exons selected from exon 10, exon 11, exon 12, exon 13, and exon 14;
- a ribosomal skipping sequence, preferably, T2A,
- the exogenous DNA sequence, preferably the coding sequence of human ARSB gene;
- a transcription termination sequence;
- a further inverted targeting sequence linked to its protospacer-adjacent motif (PAM) sequence;
- a complementary strand oligonucleotide homologous to the targeting sequence, under the control of a promoter, preferably the U6 promoter;
- a chimeric gRNA scaffold, and
- an AAV 3'-inverted terminal repeat (3'-ITR) sequence.

Preferably said viral vector comprising the gRNA expression cassette and the donor DNA comprises:

- an AAV 5'-inverted terminal repeat (5'-ITR) sequence;
- an inverted targeting sequence linked to its protospacer-adjacent motif (PAM) sequence;
- a splice acceptor sequence;
- one or more albumin exons, preferably one or more exons selected from exon 13 and 14;
- a ribosomal skipping sequence, preferably, T2A,
- the exogenous DNA sequence, preferably the coding sequence of human ARSB gene;
- a transcription termination sequence;
- a further inverted targeting sequence linked to its protospacer-adjacent motif (PAM) sequence;

- a complementary strand oligonucleotide homologous to the targeting sequence, under the control of a promoter, preferably the U6 promoter;
- a chimeric gRNA scaffold, and
- an AAV 3'-inverted terminal repeat (3'-ITR) sequence.

Alternatively, said viral vector comprising the donor DNA comprises:

- an AAV 5'-inverted terminal repeat (5'-ITR) sequence;
 - an inverted targeting sequence linked to its protospacer-adjacent motif (PAM) sequence;
 - a splice acceptor sequence;
 - one or more albumin exons, preferably one or more exons selected from exon 10, exon 11, exon 12, exon 13, and exon 14;
 - a ribosomal skipping sequence, preferably, T2A,
 - the exogenous DNA sequence, preferably the coding sequence of human BDD F8 gene;
 - a transcription termination sequence;
 - a further inverted targeting sequence linked to its protospacer-adjacent motif (PAM) sequence; and
 - an AAV 3'-inverted terminal repeat (3'-ITR) sequence
- or

said viral vector comprising the donor DNA comprises:

- an AAV 5'-inverted terminal repeat (5'-ITR) sequence;
- an inverted targeting sequence linked to its protospacer-adjacent motif (PAM) sequence;
- a splice acceptor sequence;
- one or more albumin exons, preferably one or more exons selected from exon 13 and 14;
- a ribosomal skipping sequence, preferably, T2A,
- the exogenous DNA sequence, preferably the coding sequence of human BDD F8 gene;
- a transcription termination sequence;
- a further inverted targeting sequence linked to its protospacer-adjacent motif (PAM) sequence; and
- an AAV 3'-inverted terminal repeat (3'-ITR) sequence.

Suitably, when said vector does not comprise the gRNA expression cassette, said cassette is comprised in the vector comprising the nucleic acid expressing the nuclease.

The above-mentioned elements may be in the order above defined, from 5' to 3', however other orders are equally suitable, as the skilled person can appreciate.

The vector may further comprise additional viral sequences, such as additional AAV sequences.

Preferably, said vector comprising said donor nucleic acid and said complementary strand oligonucleotide comprises or has essentially a sequence having at least 95% of identity to a sequence comprising the following sequences: SEQ ID N.29, SEQ ID N. 20, SEQ ID N.21, SEQ ID N.22, SEQ ID N.23, SEQ ID N.33, SEQ ID N.26, SEQ ID N.20, SEQ ID N.27, SEQ ID N.2, SEQ ID N.28 and SEQ ID N.29.

In another embodiment, said vector comprising said donor nucleic acid comprises or has essentially a sequence having at least 95% of identity to a sequence comprising the following sequences: SEQ ID N.110, SEQ ID N.20, SEQ ID N.21, SEQ ID N.22, SEQ ID N.23, SEQ ID N.36, SEQ ID N.37, SEQ ID N.20 and SEQ ID N.29.

Another object of the invention is a host cell comprising the constructs, vectors or vector system or system as above defined.

Another object of the invention is a viral particle that comprises the construct, vector, vector system or system as above defined.

Suitably, a viral vector as defined herein encompasses a viral vector particle.

The term "virus particle" or "viral particle" is intended to mean the extracellular form of a non-pathogenic virus, in particular a viral vector, composed of genetic material made from either DNA or RNA surrounded by a protein coat, called capsid, and in some cases an envelope derived from portions of host cell membranes and including viral glycoproteins.

As used herein, a viral vector refers also to a viral vector particle.

Viral vectors encompassed by the present invention are suitable for gene therapy.

Preferably the viral particle comprises capsid proteins of an AAV.

Preferably the viral particle comprises capsid proteins of an AAV of a serotype selected from one or more of the group consisting of AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8 AAV9, AAV 10, AAVSH19, AAVPHP.B ; preferably from the AAV2 or AAV8 serotype.

Another object of the invention is a pharmaceutical composition that comprises one of the following: a system; one or more vectors; a host cell or a viral particle as above defined and a pharmaceutically acceptable carrier.

Another object of the invention is a kit comprising : a DNA construct, a system or one or more vectors or a host cell or a viral particle as above defined or a pharmaceutical composition as above defined in one or more containers, optionally further comprising instructions or packaging materials that describe how to administer the nucleic acid construct, vector, host cell, viral particle or pharmaceutical composition to a patient.

The system or one or more vectors or a host cell or a viral particle as above defined or a pharmaceutical composition as above defined are preferably for use as a medicament, preferably for use in the treatment of a diseases herein mentioned, preferably of hepatic diseases, Lysosomal storage diseases comprising mucopolysaccharidoses (such as MPSI, MPSII, MPSIIIA, MPSIIIB, MPSIIIC, MPSIVA, MPSIVB, MPSVI, MPSVII), sphingolipidoses (Fabry's Disease, Gaucher Disease, Nieman-Pick Disease, GM1 Gangliosidosis), lipofuscinoses (Batten's Disease and others) and mucopolipidoses; other diseases where the liver can be used as a factory for production and secretion of therapeutic proteins, like diabetes, gyrate atrophy of the choroid and retina, adenylosuccinate deficiency, hemophilia A and B, ALA dehydratase deficiency, adrenoleukodystrophy.

The system or one or more vectors or a host cell or a viral particle as above defined or a pharmaceutical composition as above defined can be for use in the treatment of a diseases wherein both the mutant and wildtype alleles are replaced with a correct copy of the gene provided by the donor DNA or for use for the treatment of a recessive inherited and common disease due to loss-of-function, preferably said disease being selected from haemophilia, diabetes, Lysosomal storage diseases comprising mucopolysaccharidoses, such as MPSI, MPSII, MPSIIIA, MPSIIIB, MPSIIIC, MPSIVA, MPSIVB, MPSVI and MPSVII, sphingolipidoses, such as Fabry's Disease, Gaucher Disease, Nieman-Pick Disease and GM1 Gangliosidosis, lipofuscinoses, such as Batten's Disease, and mucopolipidoses; gyrate atrophy of the choroid and retina, adenylosuccinate deficiency, hemophilia A and B, ALA dehydratase deficiency, adrenoleukodystrophy.

Therefore the system or one or more vectors or a host cell or a viral particle as above defined or a pharmaceutical composition as above defined may be used for treating haemophilia, diabetes, Lysosomal storage diseases comprising mucopolysaccharidoses, such as MPSI, MPSII, MPSIIIA, MPSIIIB, MPSIIIC, MPSIVA, MPSIVB, MPSVI and MPSVII, sphingolipidoses, such as Fabry's Disease,

Gaucher Disease, Nieman-Pick Disease and GM1 Gangliosidosis, lipofuscinoses, such as Batten's Disease, and mucopolysaccharidoses; gyrate atrophy of the choroid and retina, adenylosuccinate deficiency, hemophilia A and B, ALA dehydratase deficiency or adrenoleukodystrophy.

A further object of the invention is a construct as above defined for the production of viral particles.

It is also an object of the invention a method for treating a subject affected by disease herein mentioned, preferably an inherited disease due to gene loss-of-function comprising administering to the subject an effective amount of the vector system or the vector or the host cell or the viral particle or the pharmaceutical composition as above defined. Preferably said disease is a lysosomal storage disease, such as mucopolysaccharidoses (MPSI, MPSII, MPSIIIA, MPSIIIB, MPSIIIC, MPSIVA, MPSIVB, MPSVI, MPSVII) or haemophilia A or B.

Preferably, object of the invention are the sequences herein mentioned.

Preferably, the donor DNA cassette elements and/or the gRNA expression cassette elements and/or the promoter sequences and/or U6 promoter for gRNA expression and/or the gRNA and/or the gRNA target site and/or the inverted targeting sequences and/or the Cas9 and /or the exogenous DNA sequence and/or the post-transcriptional regulatory element and/or the transcription termination sequence and/or the splice acceptor sequence and/or the ribosomal skipping sequence are the sequences depicted in the following sequences SEQ ID NOs 1-109.

Another object of the invention is a DNA construct comprising the donor nucleic acid and/or the complementary strand oligonucleotide homologous to a targeting sequence and/or a nucleic acid coding for a nuclease that recognizes said targeting sequence, as defined above or herein.

In an embodiment, the methods of the invention are ex-vivo or in vitro.

In an embodiment, in the methods of the invention the cell is an isolated cell from a subject or a patient.

The sequence of albumin is preferably described with the following Accession n. AC140220.4 (GeneBank, NCBI, database; last version) or with the following Accession n. NC_000004.12 .

The invention also provides a pharmaceutical composition comprising the nucleic acids as defined above or the nucleotide sequences as defined above or the vectors as defined above and pharmaceutically acceptable diluents and/or excipients and/or carriers.

Preferably the composition further comprises a therapeutic agent, preferably the therapeutic agent is selected from the group consisting of: enzyme replacement therapy and small molecule therapy.

Preferably the pharmaceutical composition is administered through a route selected from the group consisting of: parenteral, intravenous (for instance through the temporal vein), intraperitoneal, intratumoral, intrahepatic, or any combination thereof. Preferably the vector of the invention is administered through intravenous or parenteral route.

The present invention also provides the vector as defined above for medical use, wherein said vector is administered through a route selected from the group consisting of: parenteral, intravenous (for instance through the temporal vein), intraperitoneal, intratumoral, intrahepatic or any combination thereof. Preferably the vector of the invention is administered through intravenous or parenteral route.

In preferred embodiments of the invention:

- If the targeting sequence is located in intron 9 of albumin gene, albumin exons 10, 11, 12, 13 and 14 or fragments thereof are present;
- If the targeting sequence is located in intron 11 of albumin gene, albumin exons 12, 13 and 14 or fragments thereof are present;
- If the targeting sequence is located in intron 12 of albumin gene, albumin exons 13 and 14 or fragments thereof are present;
- If the targeting sequence is located in intron 13 of albumin gene, albumin exon 14 or fragments thereof is present; or
- If the targeting sequence is located in intron 14 of albumin gene, no albumin exons or fragments thereof are present.

In the present invention a 3XFLAG sequence may be present, preferably it comprises or has essentially a sequence having at least 95% of identity SEQ ID NO:56 or 62.

In the present invention "at least 80 % identity" means that the identity may be at least 80%, or 85 % or 90% or 95% or 100% sequence identity to referred sequences. This applies to all the mentioned % of identity. In the present invention "at least 95 % identity" means that the identity may be at least 95%, 96%, 97%, 98%, 99% or 100% sequence identity to referred sequences. This applies to all the mentioned % of identity. In the present invention "at least 98 % identity" means that the identity may be at least 98%, 99% or 100% sequence identity to referred sequences. This

applies to all the mentioned % of identity. Preferably, the % of identity relates to the full length of the referred sequence.

Included in the present invention are also nucleic acid sequences derived from the nucleotide sequences herein mentioned, e.g. functional fragments, mutants, variants, derivatives, analogues, and sequences having a % of identity of at least 80% with the sequences herein mentioned.

DETAILED DESCRIPTION OF THE INVENTION

Definitions

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

In the present invention “at least 80 % identity” means that the identity may be at least 80%, or 85 % or 90% or 95% or 100% sequence identity to referred sequences. This applies to all the mentioned % of identity. In the present invention “at least 95 % identity” means that the identity may be at least 95%, 96%, 97%, 98%, 99% or 100% sequence identity to referred sequences. This applies to all the mentioned % of identity. In the present invention “at least 98 % identity” means that the identity may be at least 98%, 99% or 100% sequence identity to referred sequences. This applies to all the mentioned % of identity. Preferably, the % of identity relates to the full length of the referred sequence.

Included in the present invention are also nucleic acid sequences derived from the nucleotide sequences herein mentioned, e.g. functional fragments, mutants, variants, derivatives, analogues, and sequences having a % of identity of at least 80% with the sequences herein mentioned, as far as such fragments, mutants, variants, derivatives and analogues maintain the function of the sequence from which they derive.

Figures

Figure 1. *In vivo* integration and expression of *DsRed* transgene into the 3' *mAlb* locus. Wild type mice received a mixture of AAV8-SpCas9 and AAV8-donor-gRNA (or -scRNA as negative control) via temporal vein at 1 days old (p1). **(A)** Schematic of the HITI construct and integration in the 3' mouse albumin locus. **(B)** Representative indel analysis by T7 endonuclease (T7) cleavage assay. Expected band sizes and average indel frequency are depicted (n=5). **(C)** PCR of DNA extracted from liver

samples showing precise 5' (left panel) and 3' (right panel) junction products at the 3' *mAlb* locus using specific primer combinations black arrows indicate expected band size. **(D)** Representative fluorescence microscopy imaging of liver cryo-sections of n=5 gRNA- and n=5 scRNA-treated animals at 20x magnification. gRNA = HITI donor-gRNA + SpCas9; scRNA = HITI donor-scRNA + SpCas9. In the right panel the percentage of Ds-Red positive hepatocytes is reported.

Figure 2. Integration in the 3' Albumin following neonatal administration of AAV-HITI improves the phenotype of a mouse model of MPSVI. **(A)** Schematic of AAV-gRNA-HITI donor and AAV-Cas9 constructs. SAS: synthetic splicing acceptor signal; Exon 14: exon 14 of murine Albumin; T2A: Thosea asigna virus 2A skipping peptide; spA: synthetic bovine growth hormone poliA **(B)** Serum arylsulfatase B (ARSB) activity measured in normal (NR), not treated MPS VI mice (AF NT) and gRNA-treated MPS VI mice (AF gRNA) is reported. Values are reported in logarithmic scale. **(C)** Urinary glycosaminoglycans (GAGs) were measured in normal (NR), not treated MPS VI mice (AF NT) and gRNA-treated MPS VI mice (AF gRNA). Values are reported as percentage of age-matched scramble controls.

Figure 3. HITI mediated F8 codopV3 integration in newborn hemophilic mice at the mouse 3' Alb (*mAlb*) locus. **(A)** Schematic of constructs. U6 = U6 promoter; gRNA = gRNA expression cassette; HLP = Hybrid liver promoter; Cas9 = Sp Cas9 coding sequence; pA = polyadenylation signal; SAS = synthetic splicing acceptor; Ex 14 = mouse albumin exon 14; T2A = Thosea asigna virus 2A skipping peptide; F8= coding sequence of CodopV3; pA = polyadenylation signal. **(B)** Chromogenic assay performed on plasma samples of AF = affected untreated hemophilic mice; NR= unaffected controls; HITI gRNA = affected animals treated with Cas9+U6 gRNA expression cassette and the HITI donor; HITI scRNA = affected animals treated with Cas9+U6 scRNA expression cassette and the HITI donor. Each dot corresponds to a different animal. F8 activity reported in international units per deciliter(IU/dl).

Figure 4: Serum albumin levels. Serum albumin levels measured in all animals treated and not treated with AAV-HITI at p360 after treatment. AFgRNA= affected animals treated with the AAV-HITI guide (gRNA) vector; AF scRNA= affected animals treated with the AAV-HITI scramble RNA (scRNA) vector; NR= unaffected untreated animals. Each bar corresponds to the albumin levels of a single animal. Serum albumin is expressed as mg of albumin/ml of serum.

Figure 5: Mouse Alfa fetoprotein levels (AFP). AFP levels measured in serum samples collected at p360 after treatment. AFgRNA= affected animals treated with the AAV-HITI guide (gRNA)

vector; AFscRNA= affected animals treated with the AAV-HITI scramble RNA (scRNA) vector; AF= affected untreated animals; NR= unaffected untreated animals. Each bar corresponds to the AFP levels of a single animal. AFP levels are expressed as ng of AFP/ml of serum.

Figure 6: CAST-seq analysis on AAV-HITI samples.

Visual representation of the CAST-seq analysis performed on genomic DNA extracted from liver samples of three different AAV-HITI gRNA treated MPSVI mice.

Figure 7: Dose-response of AAV-HITI to treat MPSVI mice.

Serum active ARSB levels are shown. Treatment, genotype and timepoint are reported below the graph. Each dot represents one mouse, mean levels are reported inside each bar (above the bar for the LD treatment). Normal levels of unaffected mice expressing ARSB are indicated by the dashed line and are reported as mean $\pm$ standard error of mean (from Alliegro & Ferla *et al.*, 2016). NT = not treated; HD = high dose; MD = medium dose; LD = low dose; AF = affected or *Arsb*^{-/-} mice; p30 = age 30 days.

Figure 8: Evaluation of INDELS at the 3' *ALB* locus.

Representative indel analysis by T7 endonuclease (T7) cleavage assay. Expected band sizes are indicated by black arrows, % of indels is reported below each gRNA lane and is shown as mean $\pm$ standard error of mean (n = 3 independent experiments). Molecular weight marker is the 100 bp marker. scRNA = scramble RNA; + = sample treated with T7 enzyme; - = sample not treated with T7 enzyme; SEM = standard error of mean

Figure 9: HITI-mediated integration at the 3' *Alb* or the 3' *ALB* locus *in vitro*.

Quantification of DsRed positive (DsRed+) cells upon integration of the donor DNA carrying the promoter less DsRed coding sequence, at the 3' *Alb* or the 3' *ALB* locus. The number of DsRed+ cells as result of the integration induced by the gRNA was normalized to samples receiving the scramble RNA (scRNA) and is reported as % of cells positive for EGFP linked to Cas9. Cell line, gRNA ID and targeted intron of *Alb* or *ALB* are reported below the graph. Each dot represents a biological replicate of transfected cells. scRNA = scramble RNA; HEPA 1-6 = mouse hepatoma cell line 1-6; HUH7 = human hepatoma cell line 7; *Alb* = mouse *Albumin*; intr = intron; *ALB* = human *Albumin*.

Figure 10: AAV-HITI molecular characterization at the on- and off- target sites in mice. A)

Indel percentage (%) at the on-target site, obtained with Illumina-seq NGS analysis. AAV-HITI gRNA treated mice show 29 % of indel at the on-target site while in AAV-HITI scRNA treated mice the % of indel is close to zero. **B)** Reads generated with Illumina-seq were aligned at the on-target

site to detect possible AAV genome integration. Reads containing the ITRs sequences were enriched when either the AAV-HITI or the AAV-Cas9 genome were used as reference. C) NGS Off-target analysis performed at the top 10 predicted off-target sites in DNA samples derived from AAV-HITI gRNA and AAV-HITI scRNA mice showed that inventors' selected gRNA is specific for the on-target site (mouse albumin intron 13).

Albumin gene

The albumin gene is the target genomic locus recognized by gRNAs of the invention in order to insert the exogenous DNA sequences to be expressed under the Albumin promoter.

The sequence of albumin is preferably described with the following Accession n. AC140220.4 or with the following Accession n. NC_000004.12.

The albumin gene (ENSMUSG00000029368) is located on chromosome 5 and has three alternative transcript variants, only one (ENSMUST00000031314.10, containing 15 exons) encodes for the Albumin protein (P07724, 608 aa).

The Albumin protein is abundant in plasma and it is essential for maintaining oncotic pressure that functions as a carrier protein for various molecules such as steroids and fatty acids in blood. This gene is primarily expressed in liver where the encoded protein undergoes proteolytic processing before secretion into the plasma. [provided by RefSeq, Oct 2015]

Therapeutic genes and proteins

Therapeutic genes of the invention are genes responsible for one or more genetic disease, e.g. lysosomal storage diseases comprising mucopolysaccharidoses (MPSI, MPSII, MPSIIIA, MPSIIIB, MPSIIIC, MPSIVA, MPSIVB, MPSVII), sphingolipidoses (Fabry's Disease, Gaucher Disease, Nieman-Pick Disease, GM1 Gangliosidosis), lipofuscinoses (Batten's Disease and others) and mucopolipidoses, gyrate atrophy of the choroid and retina diabetes, adenylosuccinate deficiency, hemophilia A and B, ALA dehydratase deficiency, adrenoleukodystrophy.

Particularly preferred therapeutic genes of the invention are those genes that may be expressed by liver cells to correct a defect in the same tissue or other tissues.

Suitably, according to the present invention the liver can be used as a factory for production and secretion of therapeutic proteins to correct genetic defects within the liver or affecting different tissues.

Therapeutic genes of the invention are also genes which in recessive diseases (autosomal or sex-linked) present loss of function.

Factor VIII

Factor VIII gene (ENSG00000185010, Gene Synonyms: FVIII or F8 or DXS1253E or F8C or HEMA) is located on the X chromosome (Xq28) and it encodes for coagulation factor VIII, which participates in the intrinsic pathway of blood coagulation; factor VIII is a cofactor for factor IXa which, in the presence of Ca^{+2} and phospholipids, converts factor X to the activated form Xa. This gene produces two alternatively spliced transcripts. Transcript variant 1 (ENST00000360256.9, 26 exons) encodes a large glycoprotein, isoform a, which circulates in plasma and associates with von Willebrand factor in a noncovalent complex. This protein undergoes multiple cleavage events. Transcript variant 2 (ENST00000330287.10, 5 exons) encodes a putative small protein, isoform b, which consists primarily of the phospholipid binding domain of factor VIIIc. This binding domain is essential for coagulant activity. At least 7 alternative transcripts are annotated (Ensembl.org) Defects in this gene results in hemophilia A, a common recessive X-linked coagulation disorder. [provided by Ref Seq, Jul 2008]

The sequence of Factor VIII is preferably described with the following Accession NM_000132.4 Several modifications of Factor VIII have been engineered to improve its stability and activity as described for instance in in Miao, H. Z. et al. Bioengineering of coagulation factor VIII for improved secretion. Blood (2004). In addition to deletion of the B domain wherein amino acids from 740 to 1649 (B domain) of the WT F8 protein are deleted, linker have been engineered to further improve F VIII secretion by mimicking some of the post-translational modifications that normally occur, for instance the N6 linker as described in Miao, H. Z. et al. Bioengineering of coagulation factor VIII for improved secretion. Blood (2004) and Ward et al.(Ward, N. J. et al. Codon optimization of human factor VIII cDNAs leads to high-level expression. Blood (2011)).

Suitably, a fragment of the Factor VIII coding sequence is within the scope of the present invention. A modified Factor VIII is also within the scope of the present invention.

Suitably, a codon optimized version of the Factor VIII coding sequence or a fragment thereof, for instance a BDD Factor VIII coding sequence, is within the scope of the present invention.

Arylsulfatase B (ARSB)

The gene encoding for Arylsulfatase B (ARSB) (ENSG00000113273) is located on chromosome 5 and at least 7 alternative transcripts are annotated (ensembl.org). The isoform 1 (ENST00000264914.10, 8 exons, corresponding to RefSeq NM_000046.5) encodes for a 533 aa protein (P15848-1). Arylsulfatase B encoded by this gene belongs to the sulfatase family. The arylsulfatase B homodimer hydrolyzes sulfate groups of N-Acetyl-D-galactosamine, chondriotin sulfate, and dermatan sulfate.

The protein is targeted to the lysosome. Mucopolysaccharidosis type VI is an autosomal recessive lysosomal storage disorder resulting from a deficiency of arylsulfatase B. (Provided by RefSeq, Dec 2016).

The sequence of Arylsulfatase B (ARSB) is preferably described with the following Accession n. NM_000046.5.

DNA CONSTRUCTS

Exogenous DNA Sequences

Exogenous DNA sequences mentioned above comprise a fragment of DNA to be incorporated into genomic DNA of a target genome. In some embodiments, the exogenous DNA comprises at least a portion of a gene. The exogenous DNA may comprise a coding sequence e.g. a cDNA related to a wild type gene or to a "codon optimized" sequence for the factor that has to be expressed. In some embodiments, the exogenous DNA comprises at least an exon of a gene. In some embodiments, the exogenous DNA comprises an enhancer element of a gene. In some embodiments, the exogenous DNA comprises a discontinuous sequence of a gene comprising a 5' portion of the gene fused to the 3' portion of the gene. In some embodiments, the exogenous DNA comprises a wild type gene sequence. In some embodiments, the exogenous DNA comprises a mutated gene sequence. In some embodiments, the exogenous DNA comprises a wild type gene sequence. In some embodiments, the exogenous DNA sequence comprises a reporter gene. In some embodiments, the reporter gene is selected from at least one of Discosoma Red (Dsred), a Green Fluorescent Protein (GFP), a Red Fluorescent Protein (RFP), a luciferase, a β -galactosidase, and a β -glucuronidase. In some embodiments, the exogenous DNA sequence comprises a gene transcription regulatory element which may e.g. comprise an enhancer sequence. In some embodiments, the exogenous DNA sequence comprises one or more exons or fragments thereof. In some embodiments, the exogenous DNA sequence comprises one or more introns or fragments thereof. In some embodiments, the exogenous DNA sequence comprises at least a portion of a 3' untranslated region or a 5' untranslated region. In some embodiments, the exogenous DNA sequence comprises an artificial DNA sequence. In some embodiments, the exogenous DNA sequence comprises a nuclear localization sequence and/or a nuclear export sequence. In some embodiments, the exogenous DNA sequence comprises a signal peptide sequence. An exogenous DNA sequence, in some embodiments, comprises a segment of nucleic acid to be integrated at a target genomic locus. The exogenous DNA sequence, in some embodiments, comprises one or more polynucleotides of interest. The exogenous DNA sequence in some embodiments comprises one or more expression

cassettes. Such an expression cassette, in some embodiments, comprises an exogenous DNA sequence of interest, a polynucleotide encoding a selection marker and/or a reporter gene, and regulatory components that influence expression. The exogenous DNA sequence, in some embodiments, comprises a genomic nucleic acid. The genomic nucleic acid is derived from an animal, a mouse, a human, a non-human, a rodent, a non-human, a rat, a hamster, a rabbit, a pig, a bovine, a deer, a sheep, a goat, a chicken, a cat, a dog, a ferret, a primate (e.g., marmoset, rhesus monkey), domesticated mammal or an agricultural mammal, an avian, a bacterium, an archaeon, a virus, or any other organism of interest or a combination thereof. Exogenous DNA sequences of any suitable size are integrated into a target genome. In some embodiments, the exogenous DNA sequence integrated into a genome is less than 0.5, about 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5 or 10, kilobases (kb) in length. In some embodiments, the exogenous DNA sequence integrated into a genome is at least about 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5 kb in length.

Genome insertion sites

The site of the double-strand break (DSB) can be introduced specifically by any suitable technique, for example using a CRISPR/Cas9 system and the guide RNAs disclosed herein. In the present invention, the DSB is introduced into intron 9, intron 11, intron 12, intron 13 or intron 14 of the albumin gene. Exemplary genome insertion sites are in position 733 of intron 9, in position 152 of intron 11, in position 538 of intron 12, in position 927 of intron 12, in position 173 of intron 13, in position 456 of intron 13 or in position 123 of intron 14 of the human albumin gene, wherein position is referred to the first nucleotide of each intron.

The nuclease is directed to said insertion sites preferably by gRNAs comprising or consisting of a sequence selected from SEQ ID N. 1-2, SEQ ID NO 9-18.

Ribosomal skipping sequences: 2A self-cleaving peptides

Ribosomal skipping sequence is a herein used as a synonym of 2A self-cleaving peptide, or 2A peptide.

These are 18–22 aa-long peptides which can induce the cleaving of recombinant proteins in the cell. 2A peptides are derived from the 2A region in the genome of virus.

Four members of 2A peptides family are frequently used in life science research. They are P2A, E2A, F2A and T2A. F2A is derived from foot-and-mouth disease virus 18; E2A is derived from equine rhinitis A virus; P2A is derived from porcine teschovirus-1 2A; T2A is derived from Thosea asigna virus 2.

Name	Sequence
T2A	GSG E G R G S L L T C G D V E E N P G P (SEQ ID NO: 39) E G R G S L L T C G D V E E N P G P (SEQ ID NO: 32)
P2A	GSG A T N F S L L K Q A G D V E E N P G P (SEQ ID NO: 40) A T N F S L L K Q A G D V E E N P G P (SEQ ID NO:89)
E2A	GSG Q C T N Y A L L K L A G D V E S N P G P (SEQ ID NO: 41) Q C T N Y A L L K L A G D V E S N P G P (SEQ ID NO:90)
F2A	GSG V K Q T L N F D L L K L A G D V E S N P G P (SEQ ID NO: 42) V K Q T L N F D L L K L A G D V E S N P G P (SEQ ID NO:91)

Any ribosomal skipping sequence may be utilized within the meaning of the present invention. A preferred one is T2A. Ribosomal skipping peptides, for example 2A peptides, are preferably localized between the albumin exon(s) and the exogenous DNA sequence.

Splice acceptor sequences

RNA splicing is a form of RNA processing in which a newly made precursor messenger RNA (pre-mRNA) transcript is transformed into a mature messenger RNA (mRNA). During splicing, introns (non-coding regions) are removed and exons (coding regions) are joined together.

Within introns, a donor site (5' end of the intron), a branch site (near the 3' end of the intron) and an acceptor site (3' end of the intron) are required for splicing. The splice donor site includes an almost invariant sequence GU at the 5' end of the intron, within a larger, less highly conserved region. The splice acceptor site at the 3' end of the intron terminates the intron with an almost invariant AG sequence. Upstream (5'-ward) from the AG there is a region high in pyrimidines (C and U), or polypyrimidine tract. Further upstream from the polypyrimidine tract is the branchpoint.

A “splice acceptor sequence” is a nucleotide sequence which can function as an acceptor site at the 3' end of the intron. Consensus sequences and frequencies of human splice site regions are described in Ma, S.L., et al., 2015. PLoS One, 10(6), p.e0130729.

Suitably, the splice acceptor sequence may comprise the nucleotide sequence (Y)_nNYAG, where n is 10-20, or a variant with at least 90% or at least 95% sequence identity. Suitably, the splice acceptor sequence may comprise the sequence (Y)_nNCAG, where n is 10-20, or a variant with at least 90% or at least 95% sequence identity.

Regulatory elements

The construct of the invention may comprise one or more regulatory elements which may act pre- or post-transcriptionally. The one or more regulatory elements may facilitate expression in the cells of the invention.

A “regulatory element” is any nucleotide sequence which facilitates expression of a polypeptide, e.g. acts to increase expression of a transcript or to enhance mRNA stability. Suitable regulatory elements include for example promoters, enhancer elements, post-transcriptional regulatory elements and polyadenylation sites.

The subject invention also concerns constructs that can include regulatory elements that are functional in the intended host cell in which the vector comprising the construct is to be expressed. 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, transcription termination sequences, translation termination sequences, enhancers, signal peptides, degradation signals and polyadenylation elements.

A construct of the invention may optionally contain a transcription termination sequence, a translation termination sequence, signal peptide sequence, internal ribosome entry sites (IRES), enhancer elements, and/or post-transcriptional regulatory elements such as the Woodchuck hepatitis virus (WHV) posttranscriptional regulatory element (WPRE). 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. In the system of the invention a transcription termination site is typically included.

Promoters

The nucleic acid construct of the invention can comprise a promoter sequence operably linked to a nucleotide sequence encoding the desired polypeptide. The term “operably linked”, means that the parts (e.g. transgene and promoter) are linked together in a manner which enables both to carry out their function substantially unhindered.

A promoter within the meaning of the present invention may be a ubiquitous promoter, meaning that it drives expression of the gene in a wide range of cells and tissues. A further promoter within the present invention is a tissue- specific promoter that shows selective activity in one or a group of tissues but is less active or not active in other tissue. The promoter may show inducible expression in response to presence of another factor, for example a factor present in a host cell.

Where the vector comprising the construct is administered for therapy, it is preferred that the promoter is functional in the target cell (e.g. liver cell).

In some embodiments, the promoter is a ubiquitous promoter or a liver specific promoter, preferably a hepatocyte specific promoter. Promoters contemplated for use in the subject invention include, but are not limited to, native gene promoters or fragments thereof such as cytomegalovirus (CMV) promoter (KF853603.1, bp 149-735), the U6 promoter [37,38], thyroxine binding globulin (TBG) promoter, hybrid liver specific promoter (HLP). However any suitable promoter known in the art may be used. In a preferred embodiment, the promoter is a CMV, HLP or U6 promoter.

In preferred embodiments, the promoter is a U6 promoter for example a promoter of SEQ ID N.27 or a fragment thereof.

Preferably, the promoter nucleic acid sequence that has at least 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% nucleotide identity to SEQ ID N.27, 46, 59 or 61 or a fragment thereof, preferably wherein the promoter substantially retains the natural function of the promoter of SEQ ID N.27, 46, 59 or 61.

Promoters can be incorporated into a construct using standard techniques known in the art. Multiple copies of promoters or multiple promoters can be used in a construct of the invention. In one embodiment, the promoter can be positioned about the same distance from the transcription start site as it is from the transcription start site in its natural genetic environment. Some variation in this distance is permitted without substantial decrease in promoter activity.

Polyadenylation sequence

The nucleic acid construct of the present invention may comprise a polyadenylation sequence. Suitably, the transgene is operably linked to a polyadenylation sequence. A polyadenylation sequence may be inserted downstream of the transgene to improve transgene expression.

A polyadenylation sequence typically comprises a polyadenylation signal, a polyadenylation site and a downstream element: the polyadenylation signal comprises the sequence motif recognised by the RNA cleavage complex; the polyadenylation site is the site of cleavage at which a poly-A tails is added to the mRNA; the downstream element is a GT-rich region which usually lies just downstream of the polyadenylation site, which is important for efficient processing.

In some embodiments, the polyadenylation sequence is a bovine growth hormone (bGH) polyadenylation sequence or an SV40 polyadenylation sequence; or a fragment thereof that retains the natural function of the polyadenylation sequence.

In preferred embodiments, the polyadenylation sequence is a bovine growth hormone (bGH) polyadenylation sequence, most preferably a short synthetic polyA.

A preferred polyadenylation sequence of the invention is SEQ ID N.26 or SEQ ID N.37 or SEQ ID NO:48 or SEQ ID NO:65.

In some embodiments, the polyadenylation sequence comprises or consists of a nucleic acid sequence that has at least 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% nucleotide identity to SEQ ID N. 26 or SEQ ID N.37 or SEQ ID NO:48 or SEQ ID NO:65, preferably wherein the polyadenylation sequence substantially retains the natural function of the polyadenylation sequence of SEQ ID N. 26 or SEQ ID N.37 or SEQ ID NO:48 or SEQ ID NO:65.

POST-TRANSCRIPTIONAL REGULATORY ELEMENTS

The nucleic acid constructs of the present invention may comprise post-transcriptional regulatory elements. Suitably, the protein-coding sequence is operably linked to one or more further post-transcriptional regulatory elements that may improve gene expression.

The construct of the present invention may comprise a Woodchuck Hepatitis Virus Post-transcriptional Regulatory Element (WPRE).

Suitable WPRE sequences will be well known to those of skill in the art (see, for example, Zufferey et al. (1999) Journal of Virology 73: 2886-2892; Zanta-Boussif et al. (2009) Gene Therapy 16: 605-619). Suitably, the WPRE is a wild-type WPRE or is a mutant WPRE. For example, the WPRE may be mutated to abrogate translation of the woodchuck hepatitis virus X protein (WHX), for example by mutating the WHX ORF translation start codon.

Preferably WPRE comprises or has essentially a sequence having at least 95% of identity to SEQ ID NO: 25.

KOZAK SEQUENCE

The nucleic acid construct of the present invention may comprise a Kozak sequence. is operably linked to a Kozak sequence. A Kozak sequence may be inserted before the start codon to improve the initiation of translation.

Suitable Kozak sequences will be well known to the skilled person (see, for example, Kozak (1987) Nucleic Acids Research 15: 8125-8148).

Guide RNAs

A "guide RNA" (gRNA) confers target sequence specificity to an RNA-guided nuclease. Guide RNAs are non-coding short RNA sequences which bind to the complementary target DNA sequences. For

example, in the CRISPR/Cas9 system, guide RNA first binds to the Cas9 enzyme and the gRNA sequence guides the resulting complex via base-pairing to a specific location on the DNA, where Cas9 performs its nuclease activity by cutting the target DNA strand.

The term “guide RNA” encompasses any suitable gRNA that can be used with any RNA-guided nuclease, and not only those gRNAs that are compatible with a particular nuclease such as Cas9.

The guide RNA may comprise a trans-activating CRISPR RNA (tracrRNA) that provides the stem loop structure and a target-specific CRISPR RNA (crRNA) designed to cleave the gene target site of interest. The tracrRNA and crRNA may be annealed, for example by heating them at 95°C for 5 minutes and letting them slowly cool down to room temperature for 10 minutes. Alternatively, the guide RNA may be a single guide RNA (sgRNA) that consists of both the crRNA and tracrRNA as a single construct.

The guide RNA may comprise of a 3'-end, which forms a scaffold for nuclease binding, and a 5'-end which is programmable to target different DNA sites. For example, the targeting specificity of CRISPR-Cas9 may be determined by the 15-25 bp sequence at the 5' end of the guide RNA. The desired target sequence typically precedes a protospacer adjacent motif (PAM) which is a short DNA sequence usually 2-6 bp in length that follows the DNA region targeted for cleavage by the CRISPR system, such as CRISPR-Cas9. The PAM is required for a Cas nuclease to cut and is typically found 3-4 bp downstream from the cut site. After base pairing of the guide RNA to the target, Cas9 mediates a double strand break about 3-nt upstream of PAM.

Numerous tools exist for designing guide RNAs (e.g. Cui, Y., et al., 2018. *Interdisciplinary Sciences: Computational Life Sciences*, 10(2), pp.455-465). For example, COSMID is a web-based tool for identifying and validating guide RNAs (Cradick TJ, et al. *Mol Ther - Nucleic Acids*. 2014;3(12):e214).

Chimeric RNA scaffold

A chimeric gRNA scaffold is a dual-RNA structure that directs a Cas9 endonuclease to introduce site-specific double-stranded breaks in target DNA and it is supposed to enhance the efficiency of a Cas nuclease (Martin Jinek# et al. 2012 A programmable dual RNA-guided DNA endonuclease in adaptive bacterial immunity). Within the present invention, preferred chimeric RNA scaffolds of SEQ ID N.28 or 60 are used.

RNA-guided gene editing

The vector system of the present invention may be used to deliver an exogenous DNA sequence into a cell. Subsequently, said exogenous DNA sequence can be introduced into the cell's genome at a site of a double strand break (DSB) by non-homologous end joining (NHEJ). The site of the

double-strand break (DSB) can be introduced specifically by any suitable technique, for example by using an RNA-guided gene editing system.

An "RNA-guided gene editing system" can be used to introduce a DSB and typically comprises a guide RNA and an RNA-guided nuclease. A CRISPR/Cas9 system is an example of a commonly used RNA-guided gene editing system, but other RNA-guided gene editing systems may also be used.

Nucleases

Nucleases recognizing a targeting sequence are known by those of skill in the art and include, but are not limited to, zinc finger nucleases (ZFN), transcription activator-like effector nucleases (TALEN), clustered regularly interspaced short palindromic repeats (CRISPR) nucleases, and meganucleases. Nucleases found in compositions and useful in methods disclosed herein are described in more detail below.

Zinc finger nucleases (ZFNs)

"Zinc finger nucleases" or "ZFNs" are a fusion between the cleavage domain of FokI and a DNA recognition domain containing 3 or more zinc finger motifs. The heterodimerization at a particular position in the DNA of two individual ZFNs in precise orientation and spacing leads to a double-strand break in the DNA. In some cases, ZFNs fuse a cleavage domain to the C-terminus of each zinc finger domain. In order to allow the two cleavage domains to dimerize and cleave DNA, the two individual ZFNs bind opposite strands of DNA with their C-termini at a certain distance apart. In some cases, linker sequences between the zinc finger domain and the cleavage domain require the 5' edge of each binding site to be separated by about 5-7 bp. Exemplary ZFNs that are useful in the present invention include, but are not limited to, those described in Urnov et al., *Nature Reviews Genetics*, 2010, 11 :636-646; Gaj et al., *Nat Methods*, 2012, 9(8):805-7; U.S. Patent Nos. 6,534,261; 6,607,882; 6,746,838; 6,794,136; 6,824,978; 6,866,997; 6,933, 113; 6,979,539; 7,013,219; 7,030,215; 7,220,719; 7,241,573; 7,241,574; 7,585,849; 7,595,376; 6,903,185; 6,479,626; and U.S. Application Publication Nos. 2003/0232410 and 2009/0203140. In some embodiments, a ZFN is a zinc finger nickase which, in some embodiments, is an engineered ZFN that induces site-specific single-strand DNA breaks or nicks. Descriptions of zinc finger nickases are found, e.g., in Ramirez et al., *Nucl Acids Res*, 2012, 40(12):5560-8; Kim et al., *Genome Res*, 2012, 22(7): 1327-33.

TALENs

"TALENs" or "TAL-effector nucleases" are engineered transcription activator-like effector nucleases that contain a central domain of DNA-binding tandem repeats, a nuclear localization signal, and a C-terminal transcriptional activation domain. In some instances, a DNA-binding tandem repeat

comprises 33-35 amino acids in length and contains two hypervariable amino acid residues at positions 12 and 13 that recognize one or more specific DNA base pairs. TALENs are produced by fusing a TAL effector DNA binding domain to a DNA cleavage domain. For instance, a TALE protein may be fused to a nuclease such as a wild-type or mutated FokI endonuclease or the catalytic domain of FokI. Several mutations to FokI have been made for its use in TALENs, which, for example, improve cleavage specificity or activity. Such TALENs are engineered to bind any desired DNA sequence. TALENs are often used to generate gene modifications by creating a double-strand break in a target DNA sequence, which in turn, undergoes NHEJ or HDR. In some cases, a single-stranded donor DNA repair template is provided to promote HDR. Detailed descriptions of TALENs and their uses for gene editing are found, e.g., in U.S. Patent Nos. 8,440,431; 8,440,432; 8,450,471; 8,586,363; and 8,697,853; Scharenberg et al., *Curr Gene Ther*, 2013, 13(4):291-303; Gaj et al., *Nat Methods*, 2012, 9(8):805-7; Beurdeley et al., *Nat Commun*, 2013, 4: 1762; and Joung and Sander, *Nat Rev Mol Cell Biol*, 2013, 14(1):49-55. DNA

Guided Nucleases

"DNA guided nucleases" are nucleases that use a single stranded DNA complementary nucleotide to direct the nuclease to the correct place in the genome by hybridizing to another nucleic acid, for example, the target nucleic acid in the genome of a cell. In some embodiments, the DNA guided nuclease comprises an Argonaute nuclease. In some embodiments, the DNA guided nuclease is selected from TtAgo, PfAgo, and NgAgo. In some embodiments, the DNA guided nuclease is NgAgo.

Meganucleases

"Meganucleases" are rare-cutting endonucleases or homing endonucleases that, in certain embodiments, are highly specific, recognizing DNA target sites ranging from at least 12 base pairs in length, e.g., from 12 to 40 base pairs or 12 to 60 base pairs in length.

Any meganuclease is contemplated to be used herein, including, but not limited to, I-SceI, I-SceII, I-SceIII, I-SceIV, I-SceV, I-SceVI, I-SceVII, I-CeuI, I-CeuAII, I-CreI, I-CrepsbIP, I-CrepsbIIP, I-CrepsbIIIP, I-CrepsbIVP, I-TliI, I-PpoI, PI-PsPI, F-SceI, F-SceII, F-SuvI, F-TevI, F-TevII, I-Amal, I-Anil, I-Chul, I-Cmoel, I-Cpal, I-Cpall, I-Csml, I-Cvul, I-CvuAIP, I-DdI, I-DdII, I-Dirl, I-Dmol, I-Hmul, I-Hmull, I-HsNIP, I-Llal, I-Msol, I-Naal, I-NanI, I-NclIP, I-NgrIP, I-NitI, I-Njal, I-Nsp236IP, I-PakI, I-PboIP, I-PcuIP, I-PcuAI, I-PcuVI, I-PgrIP, I-PobIP, I-PorI, I-PorIIP, I-PbpIP, I-SpBetaIP, I-Scal, I-SexIP, I-SneiP, I-SpomI, I-SpomCP, I-SpomIP, I-SpomIIP, I-SquIP, I-Ssp6803I, I-SthPhiIP, I-SthPhiST3P, I-SthPhiSTe3bP, I-TdeIP, I-TevI, I-TevII, I-TevIII, I-UarAP, I-UarHGPAIP, I-UarHGPA13P, I-VinIP, I-ZbiIP, PI-MtuI, PI-MtuHIIP, PI-MtuHIIIP, PI-Pful, PI-Pfull, PI-Pkol, PI-Pkoll, PI-Rma43812IP, PI-SpBetaIP, PI-SceI, PI-Tful, PI-Tfull, PI-Thyl, PI-

TliI, P1-THII, I-CreI meganuclease, I-CeuI meganuclease, I-MsoI meganuclease, I-SceI meganuclease, or any active variants, fragments, mutants or derivatives thereof.

CRISPR

The CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)/Cas (CRISPR-associated protein) nuclease system is an engineered nuclease system based on a bacterial system that is used for genome engineering. It is based in part on the adaptive immune response of many bacteria and archaea. When a virus or plasmid invades a bacterium, segments of the invader's DNA are converted into CRISPR RNAs (crRNA) by the "immune" response. The crRNA then associates, through a region of partial complementarity, with another type of RNA called tracrRNA to guide the Cas (e.g., Cas9) nuclease to a region homologous to the crRNA in the target DNA called a "protospacer." The Cas (e.g., Cas9) nuclease cleaves the DNA to generate blunt ends at the double-strand break at sites specified by a 20-nucleotide complementary strand sequence contained within the crRNA transcript. The Cas (e.g., Cas9) nuclease, in some embodiments, requires both the crRNA and the tracrRNA for site-specific DNA recognition and cleavage. This system has now been engineered such that, in certain embodiments, the crRNA and tracrRNA are combined into one molecule (the "single guide RNA" or "sgRNA"), and the crRNA equivalent portion of the single guide RNA is engineered to guide the Cas (e.g., Cas9) nuclease to target any desired sequence (see, e.g., Jinek et al. (2012) *Science* 337:816-821; Jinek et al. (2013) *eLife* 2:e00471; Segal (2013) *eLife* 2:e00563).

As used herein, tracrRNA is also defined as scaffold gRNA. Thus, the CRISPR/Cas system can be engineered to create a double-strand break at a desired target in a genome of a cell, and harness the cell's endogenous mechanisms to repair the induced break by homology-directed repair (HDR) or nonhomologous end-joining (NHEJ). In some embodiments, the Cas nuclease has DNA cleavage activity. The Cas nuclease, in some embodiments, directs cleavage of one or both strands at a location in a target DNA sequence. For example, in some embodiments, the Cas nuclease is a nickase having one or more inactivated catalytic domains that cleaves a single strand of a target DNA sequence. Non-limiting examples of Cas nucleases include CasI, CasIB, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (also known as CsnI and CsxI2), CasIO, , CpfI, C2c3, C2c2 and C2cICsyl, Csy2, Csy3, CseI, Cse2, CseI, Cse2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, CmrI, Cmr3, Cmr4, Cmr5, Cmr6, CpfI, CsbI, Csb2, Csb3, CsxI7, CsxI4, CsxIO, CsxI6, CsaX, Csx3, CsxI, CsxI5, CsfI, Csf2, Csf3, Csf4, homologs thereof, variants thereof, mutants thereof, and derivatives thereof. There are three main types of Cas nucleases (type I, type II, and type III), and 10 subtypes including 5 type I, 3 type II, and 2 type III proteins (see, e.g., Hochstrasser and Doudna, *Trends Biochem Sci*, 2015:40(1):58-66). Type

II Cas nucleases include, but are not limited to, Cas1, Cas2, Csn2, and Cas9. These Cas nucleases are known to those skilled in the art. For example, the amino acid sequence of the *Streptococcus pyogenes* wild-type Cas9 polypeptide is set forth, e.g., in NCBI Ref. Seq. No. NP_269215, and the amino acid sequence of *Streptococcus thermophilus* wild-type Cas9 polypeptide is set forth, e.g., in NCBI Ref. Seq. No. WP_011681470. Cas nucleases, e.g., Cas9 polypeptides, in some embodiments, are derived from a variety of bacterial species. "Cas9" refers to an RNA-guided double-stranded DNA-binding nuclease protein or nickase protein. Wild-type Cas9 nuclease has two functional domains, e.g., RuvC and HNH, that cut different DNA strands. Cas9 can induce double-strand breaks in genomic DNA (target DNA) when both functional domains are active. The Cas9 enzyme, in some embodiments, comprises one or more catalytic domains of a Cas9 protein derived from bacteria belonging to the group consisting of *Corynebacter*, *Sutterella*, *Legionella*, *Treponema*, *Filif actor*, *Eubacterium*, *Streptococcus*, *Lactobacillus*, *Mycoplasma*, *Bacteroides*, *Flaviivola*, *Flavobacterium*, *Sphaerochaeta*, *Azospirillum*, *Gluconacetobacter*, *Neisseria*, *Roseburia*, *Parvibaculum*, *Staphylococcus*, *Nitratifactor*, and *Campylobacter*. In some embodiments, the Cas9 is a fusion protein, e.g. the two catalytic domains are derived from different bacteria species. Useful variants of the Cas9 nuclease include a single inactive catalytic domain, such as a RuvC⁻ or HNH⁻ enzyme or a nickase. A Cas9 nickase has only one active functional domain and, in some embodiments, cuts only one strand of the target DNA, thereby creating a single strand break or nick. In some embodiments, the mutant Cas9 nuclease having at least a D10A mutation is a Cas9 nickase. In other embodiments, the mutant Cas9 nuclease having at least a H840A mutation is a Cas9 nickase. Other examples of mutations present in a Cas9 nickase include, without limitation, N854A and N863 A. A double-strand break is introduced using a Cas9 nickase if at least two DNA-targeting RNAs that target opposite DNA strands are used. A double-nicked induced double-strand break is repaired by NHEJ or HDR. This gene editing strategy favors HDR and decreases the frequency of indel mutations at off-target DNA sites. The Cas9 nuclease or nickase, in some embodiments, is codon-optimized for the target cell or target organism. In some embodiments, the Cas nuclease is a Cas9 polypeptide that contains two silencing mutations of the RuvC and HNH nuclease domains (D10A and H840A), which is referred to as dCas9. In one embodiment, the dCas9 polypeptide from *Streptococcus pyogenes* comprises at least one mutation at position D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, A987, or any combination thereof. Descriptions of such dCas9 polypeptides and variants thereof are provided in, for example, International Patent Publication No. WO 2013/176772. The dCas9 enzyme in some embodiments, contains a mutation at D10, E762, H983,

or D986, as well as a mutation at H840 or N863. In some instances, the dCas9 enzyme contains a D10A or DION mutation. Also, the dCas9 enzyme alternatively includes a mutation H840A, H840Y, or H840N. In some embodiments, the dCas9 enzyme of the present invention comprises D10A and H840A; D10A and H840Y; D10A and H840N; DION and H840A; DION and H840Y; or DION and H840N substitutions. The substitutions are alternatively conservative or non-conservative substitutions to render the Cas9 polypeptide catalytically inactive and able to bind to target DNA. For genome editing methods, the Cas nuclease in some embodiments comprises a Cas9 fusion protein such as a polypeptide comprising the catalytic domain of the type IIS restriction enzyme, FokI, linked to dCas9. The FokI-dCas9 fusion protein (fCas9) can use two guide RNAs to bind to a single strand of target DNA to generate a double-strand break.

Targeting Sequences

Targeting sequences herein are nucleic acid sequences recognized and cleaved by a nuclease. In some embodiments, the targeting sequence is about 9 to about 12 nucleotides in length, from about 12 to about 18 nucleotides in length, from about 18 to about 21 nucleotides in length, from about 21 to about 40 nucleotides in length, from about 40 to about 80 nucleotides in length, or any combination of subranges (e.g., 9-18, 9-21, 9-40, and 9-80 nucleotides). In some embodiments, the targeting sequence comprises a nuclease binding site. In some embodiments the targeting sequence comprises a nick/cleavage site. In some embodiments, the targeting sequence comprises a protospacer adjacent motif (PAM) sequence. In some embodiments, the target nucleic acid sequence (e.g., protospacer) is 20 nucleotides. In some embodiments, the target nucleic acid is less than 20 nucleotides. In some embodiments, the target nucleic acid is at least 5, 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30 or more nucleotides. The target nucleic acid, in some embodiments, is at most 5, 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30 or more nucleotides. In some embodiments, the target nucleic acid sequence is 16, 17, 18, 19, 20, 21, 22, or 23 bases immediately 5' of the first nucleotide of the PAM. In some embodiments, the target nucleic acid sequence is 16, 17, 18, 19, 20, 21, 22, or 23 bases immediately 3' of the last nucleotide of the PAM. In some embodiments, the target nucleic acid sequence is 20 bases immediately 5' of the first nucleotide of the PAM. In some embodiments, the target nucleic acid sequence is 20 bases immediately 3' of the last nucleotide of the PAM. In some embodiments, the target nucleic acid sequence is 5' or 3' of the PAM. A targeting sequence, in some embodiments includes nucleic acid sequences present in a target nucleic acid to which a nucleic acid-targeting segment of a complementary strand nucleic acid binds. For example, targeting sequences, in some embodiments, include sequences to which a

complementary strand nucleic acid is designed to have base pairing. Targeting sequences include cleavage sites for nucleases. A targeting sequence, in some embodiments, is adjacent to cleavage sites for nucleases. The nuclease cleaves the nucleic acid, in some embodiments, at a site within or outside of the nucleic acid sequence present in the target nucleic acid to which the nucleic acid-targeting sequence of the complementary strand binds. The cleavage site, in some embodiments, includes the position of a nucleic acid at which a nuclease produces a single-strand break or a double-strand break. For example, formation of a nuclease complex comprising a complementary strand nucleic acid hybridized to a protease recognition sequence and complexed with a protease results in cleavage of one or both strands in or near (e.g., within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 19, 20, 23, 50, or more base pairs from) the nucleic acid sequence present in a target nucleic acid to which a spacer region of a complementary strand nucleic acid binds. The cleavage site, in some embodiments, is on only one strand or on both strands of a nucleic acid. In some embodiments, cleavage sites are at the same position on both strands of the nucleic acid (producing blunt ends) or are at different sites on each strand (producing staggered ends). Site-specific cleavage of a target nucleic acid by a nuclease, in some embodiments, occurs at locations determined by base-pairing complementarity between the complementary strand nucleic acid and the target nucleic acid. Site-specific cleavage of a target nucleic acid by a nuclease protein, in some embodiments, occurs at locations determined by a short motif, called the protospacer adjacent motif (PAM), in the target nucleic acid. For example, the PAM flanks the nuclease recognition sequence at the 3' end of the recognition sequence. In some cases, the cleavage produces blunt ends. In some cases, the cleavage produces staggered or sticky ends with 5' overhangs. In some cases, the cleavage produces staggered or sticky ends with 3' overhangs. Orthologs of various nuclease proteins utilize different PAM sequences. For example different Cas proteins, in some embodiments, recognize different PAM sequences. For example, in *S. pyogenes*, the PAM is a sequence in the target nucleic acid that comprises the sequence 5'-XRR-3', where R is either A or G, where X is any nucleotide and X is immediately 3' of the target nucleic acid sequence targeted by the spacer sequence. The PAM sequence of *S. pyogenes* Cas9 (SpyCas9) is 5'-XGG-3', where X is any DNA nucleotide and is immediately 3' of the nuclease recognition sequence of the non-complementary strand of the target DNA. The PAM of Cpf1 is 5'-TTX-3', where X is any DNA nucleotide and is immediately 5' of the nuclease recognition sequence. Preferably, The Cas9/sgRNA complex introduces DSBs 3 base pairs upstream of the PAM sequence in the genomic target sequence, resulting in two blunt ends. The exact same Cas9/sgRNA target sequence is loaded onto the donor DNA in the reverse direction.

Targeted genomic loci, as well as the donor DNA, are cleaved by Cas9/gRNA and the linearized donor DNAs are integrated into target sites via the NHEJ DSB repair pathway. If donor DNA is integrated in the correct orientation, junction sequences are protected from further cleavage by Cas9/gRNA. If donor DNA integrates in the reverse orientation, Cas9/gRNA will excise the integrated donor DNA due to the presence of intact Cas9/gRNA target sites.

In embodiments of the present invention the PAM has a sequence selected from TGG, AGG, GGG, CGG.

VECTORS

The present invention also relates to a vector comprising the nucleic acid constructs as described herein.

Such vector may therefore contain any of the elements above described in relation to the constructs. In particular, it can comprise, one or more regulatory elements including, for example, promoters, transcription termination sequences, translation termination sequences, enhancers, signal peptides, degradation signals and polyadenylation elements, in particular as above defined.

Vectors suitable for the delivery and expression of nucleic acids into cells for gene therapy are encompassed by the present invention.

Vectors of the invention include viral and non-viral vectors.

Non-viral vectors include non-viral agents commonly used to introduce or maintain nucleic acid into cells. Said agents include in particular polymer-based, particle-based, lipid-based, peptide-based delivery vehicles or combinations thereof, such as cationic polymers, micelles, liposomes, exosomes, microparticles and nanoparticles including lipid nanoparticles (LNP).

Among viral delivery, genetically engineered viruses, including adeno-associated viruses, are currently amongst the most popular tools for gene delivery. The concept of virus-based gene delivery is to engineer the virus so that it can express the gene(s) of interest or regulatory sequences such as promoters and introns. Depending on the specific application and the type of virus, most viral vectors contain mutations that hamper their ability to replicate freely as wild-type viruses in the host. Viruses from several different families have been modified to generate viral vectors for gene delivery. These viruses include retroviruses, lentiviruses, adenoviruses, adeno-associated viruses, herpes viruses, baculoviruses, picornaviruses, and alphaviruses.

Viral vectors of the invention may be derived from non-pathogenic parvovirus such as adeno-associated virus (AAV), retrovirus such as gammaretrovirus, spumavirus and lentivirus, adenovirus, poxvirus and an herpes virus.

Particularly preferred viruses according to the present invention are lentivirus and adeno-associated virus.

Viral vectors are by nature capable of penetrating into cells and delivering nucleic acids of interest into cells, according to a process known as viral transduction.

As used herein, the term “viral vector” refers to a non-replicating, non-pathogenic virus engineered for the delivery of genetic material into cells. Viral genes essential for replication and virulence are replaced with an expression cassette for the transgene of interest. Thus, the viral vector genome comprises the transgene expression cassette flanked by the viral sequences required for viral vector production.

The term “virus particle” or “viral particle” is intended to mean the extracellular form of a non-pathogenic virus, in particular a viral vector, composed of genetic material made from either DNA or RNA surrounded by a protein coat, called capsid, and in some cases an envelope derived from portions of host cell membranes and including viral glycoproteins.

As used herein, a viral vector refers also to a viral vector particle.

Viral vectors encompassed by the present invention are suitable for gene therapy.

Viral particles can be for example obtained using vectors that are capable of accommodating genes of interest and helper cells that can provide the viral structural proteins and enzymes to allow for the generation of vector-containing infectious viral particles.

Adeno-associated virus (AAV)

Adeno-associated virus is a family of viruses that differs in nucleotide and amino acid sequence, genome structure, pathogenicity, and host range. This diversity provides opportunities to use viruses with different biological characteristics to develop different therapeutic applications.

An ideal adeno-associated virus-based vector for gene delivery must be efficient, cell-specific, regulated, and safe. The efficiency of delivery may determine the efficacy of the therapy. Current efforts are aimed at achieving cell-type-specific infection and gene expression with adeno-associated viral vectors. In addition, adeno-associated viral vectors are being developed to regulate

the expression of the gene of interest, since the therapy may require long-lasting or regulated expression.

Adeno-associated virus (AAV) is a small virus which infects humans and some other primate species. AAV is not currently known to cause disease and consequently the virus causes a very mild immune response. Gene therapy vectors using AAV can infect both dividing and quiescent cells and persist in an extrachromosomal state without integrating into the genome of the host cell. These features make AAV a very attractive candidate for creating viral vectors for gene therapy, and for the creation of isogenic human disease models.

Wild-type AAV has attracted considerable interest from gene therapy researchers due to a number of features. Chief amongst these is the virus's apparent lack of pathogenicity. It can also infect non-dividing cells and has the ability to stably integrate into the host cell genome at a specific site (designated AAVS1) in the human chromosome 19. Development of AAVs as gene therapy vectors, however, has eliminated this integrative capacity by removal of the *rep* and *cap* from the DNA of the vector. The desired gene together with a promoter to drive transcription of the gene is inserted between the ITRs that aid in concatemer formation in the nucleus after the single-stranded vector DNA is converted by host cell DNA polymerase complexes into double-stranded DNA. AAV-based gene therapy vectors form episomal concatemers in the host cell nucleus. In non-dividing cells, these concatemers remain intact for the life of the host cell. In dividing cells, AAV DNA is lost through cell division, since the episomal DNA is not replicated along with the host cell DNA. Random integration of AAV DNA into the host genome is detectable but occurs at very low frequency. AAVs also present very low immunogenicity, seemingly restricted to generation of neutralizing antibodies, while they induce no clearly defined cytotoxic response. This feature, along with the ability to infect quiescent cells make AAV particularly suitable for human gene therapy.

The AAV genome is built of single-stranded deoxyribonucleic acid (ssDNA), either positive- or negative-sensed, which is about 4.7 kilobase long. The genome comprises inverted terminal repeats (ITRs) at both ends of the DNA strand, and two open reading frames (ORFs): *rep* and *cap*. The former is composed of four overlapping genes encoding Rep proteins required for the AAV life cycle, and the latter contains overlapping nucleotide sequences of capsid proteins: VP1, VP2 and VP3, which interact together to form a capsid of an icosahedral symmetry.

The Inverted Terminal Repeat (ITR) sequences received their name because of their symmetry, which was shown to be required for efficient multiplication of the AAV genome. Another property

of these sequences is their ability to form a hairpin, which contributes to so-called self-priming that allows primase-independent synthesis of the second DNA strand. The ITRs were also shown to be required for efficient encapsidation of the AAV DNA combined with generation of a fully assembled, deoxyribonuclease-resistant AAV particles.

With regard to gene therapy, ITRs seem to be the only sequences required in cis next to the therapeutic gene: structural (cap) and packaging (rep) genes can be delivered in trans. With this assumption many methods were established for efficient production of recombinant AAV (rAAV) vectors containing a reporter or therapeutic gene.

The AAV vector comprises an AAV capsid able to transduce the target cells of interest. The AAV capsid may be from one or more AAV natural or artificial serotypes.

AAV may be referred to in terms of their serotype. A serotype corresponds to a variant subspecies of AAV which, owing to its profile of expression of capsid surface antigens, has a distinctive reactivity which can be used to distinguish it from other variant subspecies. Typically, an AAV vector particle having a particular AAV serotype does not efficiently cross-react with neutralising antibodies specific for any other AAV serotype.

All of the known serotypes can infect cells from multiple diverse tissue types. Tissue specificity is determined by the capsid serotype and pseudotyping of AAV vectors to alter their tropism range affects their use in therapy.

The inverted terminal repeat (ITR) sequences used in an AAV vector system of the present invention can be any AAV ITR. The ITRs used in an AAV vector can be the same or different. For example, a vector may comprise an ITR of AAV serotype 2 and an ITR of AAV serotype 5. In one embodiment of a vector of the invention, an ITR is from AAV serotype 2, 4, 5, or 8. In the present invention ITRs of AAV serotype 2 are preferred. AAV ITR sequences are well known in the art (for example, see for ITR2, GenBank Accession Nos. AF043303.1 ; NC_001401.2; J01901.1 ; JN898962.1; see for ITR5, GenBank Accession No. NC_006152.1).

Serotype 2 (AAV2) has been the most extensively examined so far. AAV2 presents natural tropism towards skeletal muscles, neurons, vascular smooth muscle cells and hepatocytes.

Three cell receptors have been described for AAV2: heparan sulfate proteoglycan (HSPG), $\alpha\beta 5$ integrin and fibroblast growth factor receptor 1 (FGFR-1). The first functions as a primary receptor, while the latter two have a co-receptor activity and enable AAV to enter the cell by receptor-

mediated endocytosis. HSPG functions as the primary receptor, though its abundance in the extracellular matrix can scavenge AAV particles and impair the infection efficiency.

Although AAV2 is the most popular serotype in various AAV-based research, it has been shown that other serotypes can be effective as gene delivery vectors. For instance AAV6 appears much better in infecting airway epithelial cells, AAV7 presents very high transduction rate of murine skeletal muscle cells (similarly to AAV1 and AAV5), AAV8 is superb in transducing hepatocytes and photoreceptors and AAV1 and 5 were shown to be very efficient in gene delivery to vascular endothelial cells. In the brain, most AAV serotypes show neuronal tropism, while AAV5 also transduces astrocytes. AAV6, a hybrid of AAV1 and AAV2, also shows lower immunogenicity than AAV2.

Serotypes can differ with the respect to the receptors they are bound to. For example AAV4 and AAV5 transduction can be inhibited by soluble sialic acids (of different form for each of these serotypes), and AAV5 was shown to enter cells via the platelet-derived growth factor receptor.

Methods for preparing viruses and virions comprising a heterologous polynucleotide or construct are known in the art. In the case of AAV, cells can be coinfecting or transfected with adenovirus or polynucleotide constructs comprising adenovirus genes suitable for AAV helper function. Examples of materials and methods are described, for example, in U.S. Patent Nos. 8,137,962 and 6,967,018. An AAV virus or AAV vector of the invention can be of any AAV serotype, including, but not limited to, serotype AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, and AAV11, AAV-PhP.B and AAV-PhP.eB.

In a specific embodiment, an AAV2 or an AAV5 or an AAV7 or an AAV8 or an AAV9 serotype is utilized. Preferably, the AAV2-8 is used.

Suitably, the AAV genome is derivatised for the purpose of administration to patients. Such derivatisation is standard in the art and the invention encompasses the use of any known derivative of an AAV genome, and derivatives which could be generated by applying techniques known in the art. The AAV genome may be a derivative of any naturally occurring AAV. Suitably, the AAV genome is a derivative of AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, or AAV11.

Derivatives of an AAV genome include any truncated or modified forms of an AAV genome which allow for expression of a transgene from an AAV vector of the invention in vivo. In one embodiment,

the AAV serotype provides for one or more tyrosine to phenylalanine (Y-F) mutations on the capsid surface.

The DNA constructs described above can be used to generate the AAV vector of the invention. The AAV vector can be for example produced by triple transfection of producer cells, such as HEK293 cells, a method known in the field wherein the plasmid comprising the gene of interest, is transfected along with two additional plasmids into a producer cell wherein the viral particles will then be produced.

Plasmid

It is also within the invention a plasmid for the generation of a viral vector as herein defined.

The plasmid may comprise DNA constructs as above described. The plasmid usually further comprises backbone elements which are typically required for the for the large-scale plasmid production in bacteria, such as bacterial origin of replication, bacterial promoter, antibiotic resistance gene.

It is within the invention the use of said plasmid for the generation of a vector according to the invention.

The vector, for example an AAV vector, can be for example produced by triple transfection of producer cells, such as HEK293 cells, a method known in the field wherein the plasmid comprising the DNA constructs of interest is transfected along with two additional plasmids into a producer cell wherein the viral particles will then be produced.

HITI DNA Genome editing system

As used herein, a “genome-editing system” is a system which comprises all components necessary to edit a genome, preferably using the constructs or the vectors of the invention.

Within the present invention, a genome editing system is a system comprising a donor nucleic acid comprising the exogenous DNA sequence and optionally one or more exons of the Albumin gene, a complementary strand oligonucleotide homologous to a targeting sequence, eg a gRNA homologous to a targeting sequence within the Albumin gene, preferably within intron 12, 13 or 14 of the Albumin gene as defined herein, and a nuclease that recognizes said targeting sequence.

Suitably, the genome editing system of the present invention comprises nucleotide sequences, DNA constructs, vectors, eg non viral or viral vectors, and/or viral particles of the present invention.

HOST CELL

The subject invention also concerns a host cell comprising the viral vector of the invention. The host cell can be a cultured cell or a primary cell, i.e., isolated directly from an organism, e.g., a human. The host cell can be an adherent cell or a suspended cell, i.e., a cell that grows in suspension. Suitable host cells are known in the art and include, for instance, DH5 α , E. coli cells, Chinese hamster ovarian cells, monkey VERO cells, COS cells, HEK293 cells, and the like. The cell can be a human cell or from another animal. In one embodiment, the cell is a retina cell, particularly a photoreceptor cell, an RPE cell or a cone cell. The cell may also be liver cell, particularly a hepatocyte. The selection of an appropriate host is deemed to be within the scope of those skilled in the art from the teachings herein. Preferably, said host cell is an animal cell, and most preferably a human cell. The cell can express a nucleotide sequence provided in the viral vector of the invention.

The man skilled in the art is well aware of the standard methods for incorporation of a polynucleotide or vector into a host cell, for example transfection, lipofection, electroporation, microinjection, viral infection, thermal shock, transformation after chemical permeabilisation of the membrane or cell fusion.

As used herein, the term "host cell or host cell genetically engineered" relates to host cells which have been transduced, transformed or transfected with the viral vector of the invention.

COMPOSITIONS

Pharmaceutical compositions within the meaning of the present invention comprise a system, one or more vectors, a host cell or a viral particle of the invention in combination with a pharmaceutically acceptable carrier, diluent, excipient or adjuvant. The choice of pharmaceutical carrier, excipient or diluent can be selected with regard to the intended route of administration and standard pharmaceutical practice. The pharmaceutical compositions may comprise as - or in addition to - the carrier, excipient or diluent any suitable binder(s), lubricant(s), suspending agent(s), coating agent(s), solubilising agent(s), and other carrier agents that may aid or increase the viral entry into the target site (such as for example a lipid delivery system). The vector can be administered in vivo or ex vivo.

Pharmaceutical compositions adapted for parenteral administration, comprising an amount of a compound, constitute a preferred embodiment of the invention. For parenteral administration, the compositions may be best used in the form of a sterile aqueous solution which may contain other

substances, for example enough salts or monosaccharides to make the solution isotonic with blood. In a preferred embodiment the vector or the pharmaceutical composition is systemically delivered, for example by intravenous injection.

The methods of the present invention can be used with humans and other animals. As used herein, the terms "patient" and "subject" are used interchangeably and are intended to include such human and non-human species. Likewise, in vitro methods of the present invention can be carried out on cells of such human and non-human species.

KITS

The subject invention also concerns kits comprising DNA constructs, a system, one or more vectors, a host cell or a viral particle of the invention in one or more containers. Kits of the invention can optionally include pharmaceutically acceptable carriers and/or diluents. In one embodiment, a kit of the invention includes one or more other components, adjuncts, or adjuvants as described herein. In one embodiment, a kit of the invention includes instructions or packaging materials that describe how to administer a vector system of the kit. Containers of the kit can be of any suitable material, e.g., glass, plastic, metal, etc., and of any suitable size, shape, or configuration. In one embodiment, the viral vector or the host cell of the invention is provided in the kit as a solid. In another embodiment, the viral vector or the host cell of the invention is provided in the kit as a liquid or solution. In one embodiment, the kit comprises an ampoule or syringe containing the viral vector or the host cell of the invention in liquid or solution form.

Delivery

The vectors of the present invention may be administered to a patient. Said administration may be an "in vivo" administration or an "ex vivo" administration. A skilled worker would be able to determine appropriate dosage rates. The term "administered" includes delivery by viral or non-viral techniques. Viral delivery mechanisms include but are not limited to adenoviral vectors, adeno-associated viral (AAV) vectors, herpes viral vectors, retroviral vectors, lentiviral vectors, and baculoviral vectors etc as described above. Non-viral delivery systems include polymer-based, particle-based, lipid-based, peptide-based delivery vehicle or combinations thereof, such as cationic polymers, micelles, liposomes, exosomes, microparticles and nanoparticles including lipid nanoparticles (LNP), DNA transfection such as electroporation. . The delivery of one or more therapeutic genes by a vector system according to the present invention may be used alone or in combination with other treatments or components of the treatment.

Any suitable delivery method is contemplated to be used for delivering the compositions of the disclosure. The individual components of the HITI genome editing system (e.g., gRNA, nuclease and/or the exogenous DNA sequence), in some embodiments, are delivered simultaneously or temporally separated. The choice of method of genetic modification is dependent on the type of cell being transformed and/or the circumstances under which the transformation is taking place (e.g., in vitro, ex vivo, or in vivo). A general discussion of these methods is found in Ausubel, et al., Short Protocols in Molecular Biology, 3rd ed., Wiley & Sons, 1995.

The term "contacting the cell" comprises all the delivery method herein disclosed. In some embodiments, a method as disclosed herein involves contacting a target DNA or introducing into a cell (or a population of cells) one or more nucleic acids comprising nucleotide sequences encoding a complementary strand nucleic acid (e.g., gRNA), a site-directed modifying polypeptide (e.g., Cas protein) or a nucleic acid coding thereof, and/or a exogenous DNA sequence.

Methods of Genomic DNA editing

The term "genome editing" refers to a type of genetic engineering in which DNA is inserted, replaced, or removed from a target DNA, e.g. the genome of a cell, using one or more nucleases and/or nickases.

Provided herein are homology-independent targeted integration (HITI) methods and compositions for making changes to nucleic acid, such as genomic DNA, including genomic DNA in dividing or non-dividing or terminally differentiated cells. Methods herein, at least in some embodiments, are homology independent, using non-homologous end-joining to insert exogenous DNA into a target DNA, such as a genomic DNA of a cell, such as a dividing or non-dividing or terminally differentiated cell. In some embodiments, methods herein comprise a method of integrating an exogenous DNA sequence into a genome of a dividing or non-dividing cell comprising contacting the non-dividing cell with a composition comprising one or more targeting constructs comprising the exogenous DNA sequence and a targeting sequence, a complementary strand oligonucleotide homologous to the targeting sequence, and a nuclease, wherein the exogenous DNA sequence comprises at least one nucleotide difference compared to the genome and the targeting sequence is recognized by the nuclease. In some embodiments of HITI methods disclosed herein, exogenous DNA sequences are fragments of DNA containing the desired sequence to be inserted into the genome of the target cell or host cell. At least a portion of the exogenous DNA sequence has a sequence homologous to a portion of the genome of the target cell or host cell and at least a portion of the exogenous DNA sequence has a sequence not homologous to a portion of the genome of the target cell or host cell.

For example, in some embodiments, the exogenous DNA sequence may comprise a portion of a host cell genomic DNA sequence with a mutation therein. Therefore, when the exogenous DNA sequence is integrated into the genome of the host cell or target cell, the mutation found in the exogenous DNA sequence is carried into the host cell or target cell genome. In some embodiments of HITI methods disclosed herein, the exogenous DNA sequence is flanked by at least one targeting sequence. In some embodiments, the exogenous DNA sequence is flanked by two targeting sequences. The targeting sequence comprises a specific DNA sequence that is recognized by at least one nuclease. In some embodiments, the targeting sequence is recognized by the nuclease in the presence of a complementary strand oligonucleotide having a homologous sequence to the targeting sequence. In some embodiments, in HITI methods disclosed herein, a targeting sequence comprises a nucleotide sequence that is recognized and cleaved by a nuclease. Nucleases recognizing a targeting sequence are known by those of skill in the art and include but are not limited to zinc finger nucleases (ZFN), transcription activator-like effector nucleases (TALEN), and clustered regularly interspaced short palindromic repeats (CRISPR) nucleases. ZFNs, in some embodiments, comprise a zinc finger DNA-binding domain and a DNA cleavage domain, fused together to create a sequence specific nuclease. TALENs, in some embodiments, comprise a TAL effector DNA binding domain and a DNA cleavage domain, fused together to create a sequence specific nuclease. CRISPR nucleases, in some embodiments, are naturally occurring nucleases that recognize DNA sequences homologous to clustered regularly interspaced short palindromic repeats, commonly found in prokaryotic DNA. CRISPR nucleases include, but are not limited to, Cas9 Cpf1, C2c3, C2c2, and C2cl. Conveniently, a Cas 9 of the present invention is a variant with reduced off target activity as SpCas9 D10A (Ran, F.A., et al., *Genome engineering using the CRISPR-Cas9 system*. Nat Protoc, 2013. **8**(11): p. 2281-2308. (with Inactivation of RuvC domain cleavage activity), SpCas9 N863A (Ran, F.A., et al., *Genome engineering using the CRISPR-Cas9 system*. Nat Protoc, 2013. **8**(11): p. 2281-2308) (Inactivation of HNH domain cleavage activity), SpCas9-HF1 (Kleinstiver, B.P., et al., *High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects*. Nature, 2016. **529**(7587): p. 490-5) (Reduction of Cas9 binding energy by protein engineering), eSpCas9 (laymaker, I.M., et al., *Rationally engineered Cas9 nucleases with improved specificity*. Science, 2016. **351**(6268): p. 84-8) (Reduction of positive charge of Cas9), EvoCas9 (asini, A., et al., *A highly specific SpCas9 variant is identified by in vivo screening in yeast*. Nat Biotechnol, 2018. **36**(3): p. 265-271) (Mutagenesis of REC3 domain), KamiCas9 (Merienne, N., et al., *The Self-*

Inactivating KamiCas9 System for the Editing of CNS Disease Genes. Cell Rep, 2017. **20**(12): p. 2980-2991) (Knockout of Cas9 after expression).

HITI DNA genomic editing methods disclosed herein, in some embodiments, are capable of introducing exogenous DNA sequences into a host genome or a target genome. In some embodiments, insertions comprise a specific number of nucleotides ranging from 1 to 4,700 base pairs, for example 1-10, 5-20, 15-30, 20-50, 40-80, 50-100, 100-1000, 500-2000, 1000- 4,700 base pairs. In some embodiments, the method comprises eliminating at least one gene, or fragment thereof, eg one or more exons or fragments thereof, from the host genome or target genome. In some embodiments, the method comprises introducing an exogenous gene (herein also defined as Exogenous DNA sequence or gene of interest), or fragment thereof, into the host genome or target genome. HITI genome editing methods disclosed herein have increased capabilities in making changes to genomic DNA in dividing and non-dividing cells. Non-dividing cells include, but are not limited to : cells in the central nervous system including neurons, oligodendrocytes, microglia and ependymal cells; sensory transducer cells; autonomic neuron cells; sense organ and peripheral neuron supporting cells; cells in the retina including photoreceptors, rods and cones; cells in the kidney including parietal cells, glomerulus podocytes, proximal tubule brush border cells, loop of henle thin segment cells, distal tubule cells, collecting duct cells; cells in the hematopoietic lineage including lymphocytes, monocytes, neutrophils, eosinophils, basophils, thrombocytes; preferred non-dividing cells of the invention are liver cells including hepatocytes, stellate cells, the Kupffer cells and the liver endothelial cells, preferably hepatocytes. In some embodiments, HITI genome editing methods disclosed herein provide a method of making changes to genomic DNA in dividing cells, wherein the method has higher efficiency than previous methods disclosed in the art. In some embodiments, the donor nucleic acid, the complementary strand oligonucleotide, and/or the polynucleotide encoding the nuclease for HITI methods described herein are introduced into the target cell or the host cell by a virus. Viruses, in some embodiments, infect the target cell and express the targeting construct, the complementary strand oligonucleotides, and the nuclease, which allows the exogenous DNA of the targeting construct to be integrated into the host genome. In some embodiments, the virus comprises a sendai virus, a retrovirus, a lentivirus, a baculovirus, an adenovirus, or an adeno-associated virus. In some embodiments the virus is a pseudotyped virus. In some embodiments, the donor nucleic acid, the complementary strand oligonucleotide, and/or the polynucleotide encoding the nuclease for HITI methods described herein are introduced into the target cell or the host cell by a non-viral gene delivery method. Non-viral gene delivery methods, in

some embodiments, deliver the genetic materials (including DNA, RNA and protein) into the target cell and express the donor nucleic acid, the complementary strand oligonucleotide, and the nuclease, which allows the exogenous DNA of the donor nucleic acid to be integrated into the host genome. In some embodiments, the non-viral method comprises transfection reagent (including nanoparticles) for DNA mRNA or protein, or electroporation.

Methods of treating disease

Also provided herein are methods and compositions for treating disease, such as genetic diseases. Genetic diseases are those that are caused by mutations in inherited DNA. In some embodiments, genetic diseases are caused by mutations in genomic DNA. Genetic mutations are known by those of skill in the art and include, single base-pair changes or point mutations, insertions, and deletions. In some embodiments, methods provided herein include a method of treating a genetic disease in a subject in need thereof, wherein the genetic disease results from a mutated gene having at least one changed nucleotide compared to a wild-type gene, wherein the method comprises contacting at least one cell of the subject with a composition comprising DNA constructs, vectors, e.g. non viral or viral vectors or a system according to the present invention such that a donor nucleic acid comprising the exogenous DNA sequence and optionally one or more exons of the Albumin gene, a complementary strand oligonucleotide homologous to a targeting sequence, eg a gRNA homologous to a targeting sequence, and a nuclease that recognizes said targeting sequence are introduced into said cell, wherein said targeting sequence is located at the 3' end of the albumin gene in a region selected from intron 9, intron 11, intron 12, intron 13, and intron 14 of said albumin gene. Then, the donor DNA is inserted into the target locus by means of NHEJ, the Albumin gene is reconstituted and the therapeutic gene is expressed in the target cells under the Albumin promoter. Genetic diseases that are treated by methods disclosed herein include but are not limited to Lysosomal storage diseases comprising mucopolysaccharidoses (MPSI, MPSII, MPSIIIA, MPSIIIB, MPSIIIC, MPSIVA, MPSIVB, MPSVII), sphingolipidoses (Fabry's Disease, Gaucher Disease, Nieman-Pick Disease, GM1 Gangliosidosis), lipofuscinoses (Batten's Disease and others) and mucopolipidoses; other diseases where the liver can be used as a factory for production and/or secretion of therapeutic proteins, like diabetes, gyrate atrophy of the choroid and retina, adenylosuccinate deficiency, hemophilia A and B, ALA dehydratase deficiency, adrenoleukodystrophy.

The term "genome editing" refers to a type of genetic engineering in which DNA is inserted, replaced, or removed from a target DNA, e.g. the genome of a cell, using one or more nucleases and/or nickases.

The term "nonhomologous end joining" or "NHEJ" refers to a pathway that repairs double-strand DNA breaks in which the break ends are directly ligated without the need for a homologous template.

The term "polynucleotide," "oligonucleotide", "nucleic acid", "nucleotide" and "nucleic acid molecule" may be used interchangeably refers to deoxyribonucleic acids (DNA), ribonucleic acids (RNA) and polymers thereof in either single, double- or multi- stranded form. The term includes, but is not limited to, single-, double- or multi-stranded DNA or RNA, genomic DNA, cDNA, DNA-RNA hybrids, or a polymer comprising purine and/or pyrimidine bases or other natural, chemically modified, biochemically modified, non-natural, synthetic, or derivatized nucleotide bases. It also includes modifications, such as by methylation and/or by capping, and unmodified forms of the polynucleotide. More particularly, the terms "polynucleotide," "oligonucleotide," "nucleic acid" and "nucleic acid molecule" include polydeoxyribonucleotides (containing 2-deoxy-D-ribose), polyribonucleotides (containing D-ribose), any other type of polynucleotide which is an N- or C-glycoside of a purine or pyrimidine base, and other polymers containing non nucleotidic backbones, for example, polyamide (e.g., peptide nucleic acids (PNAs)) and polymorpholino (commercially available from the Anti-Virals, Inc., Corvallis, Oreg., as Neugene) polymers, and other synthetic sequence-specific nucleic acid polymers providing that the polymers contain nucleobases in a configuration which allows for base pairing and base stacking, such as is found in DNA and RNA. In some embodiments, a nucleic acid can comprise a mixture of DNA, RNA, and analogs thereof. Unless specifically limited, the term encompasses nucleic acids containing known analogs of natural nucleotides that have similar binding properties as the reference nucleic acid and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (e.g., degenerate codon substitutions), alleles, orthologs, single nucleotide polymorphisms (SNPs), and complementary sequences as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues. The term nucleic acid is used interchangeably with gene, cDNA, and mRNA encoded by a gene. The term "gene" or "nucleotide sequence encoding a polypeptide" means the segment of DNA involved in producing a polypeptide chain. The DNA segment may include regions preceding and following the coding region (leader and trailer) involved in the transcription/translation of the gene product and

the regulation of the transcription/translation, as well as intervening sequences (introns) between individual coding segments (exons).

The terms "polypeptide," "peptide," and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers and non-naturally occurring amino acid polymers. As used herein, the terms encompass amino acid chains of any length, including full-length proteins, wherein the amino acid residues are linked by covalent peptide bonds.

A "recombinant expression vector" is a nucleic acid construct, generated recombinantly or synthetically, with a series of specified nucleic acid elements that permit transcription of a particular polynucleotide sequence in a host cell. An expression vector may be part of a plasmid, viral genome, or nucleic acid fragment. Typically, an expression vector includes a polynucleotide to be transcribed, operably linked to a promoter.

As used herein, the term "administering" includes oral administration, topical contact, administration as a suppository, intravenous, intraperitoneal, intramuscular, intralesional, intrathecal, intranasal, or subcutaneous administration to a subject. Administration is by any route, including parenteral and transmucosal (e.g., buccal, sublingual, palatal, gingival, nasal, vaginal, rectal, or transdermal). Parenteral administration includes, e.g., intravenous, intramuscular, intra-arteriole, intradermal, subcutaneous, intraperitoneal, intraventricular, and intracranial. Other modes of delivery include, but are not limited to, the use of liposomal formulations, intravenous infusion, transdermal patches, etc.

The term "treating" refers to an approach for obtaining beneficial or desired results including but not limited to a therapeutic benefit and/or a prophylactic benefit. By therapeutic benefit is meant any therapeutically relevant improvement in or effect on one or more diseases, conditions, or symptoms under treatment. Slowing the progression of a disease is considered a therapeutic improvement within the meaning of the present invention. For prophylactic benefit, the compositions may be administered to a subject at risk of developing a particular disease, condition, or symptom, or to a subject reporting one or more of the physiological symptoms of a disease, even though the disease, condition, or symptom may not have yet been manifested. The term "effective amount" or "sufficient amount" refers to the amount of an agent (e.g., DNA nuclease, etc.) that is sufficient to effect beneficial or desired results. The therapeutically effective amount may vary depending upon one or more of: the subject and disease condition being treated, the weight and

age of the subject, the severity of the disease condition, the manner of administration and the like, which can readily be determined by one of ordinary skill in the art. The specific amount may vary depending on one or more of: the particular agent chosen, the target cell type, the location of the target cell in the subject, the dosing regimen to be followed, whether it is administered in combination with other compounds, timing of administration, and the physical delivery system in which it is carried.

The term "pharmaceutically acceptable carrier" refers to a substance that aids the administration of an agent (e.g., DNA nuclease, etc.) to a cell, an organism, or a subject. "Pharmaceutically acceptable carrier" refers to a carrier or excipient that can be included in a composition or formulation and that causes no significant adverse toxicological effect on the patient. Non-limiting examples of pharmaceutically acceptable carrier include water, NaCl, normal saline solutions, lactated Ringer's, normal sucrose, normal glucose, binders, fillers, disintegrants, lubricants, coatings, sweeteners, flavors and colors, and the like. One of skill in the art will recognize that other pharmaceutical carriers are useful in the present invention.

Variants, derivatives, analogues, and fragments

In addition to the specific proteins and nucleotides mentioned herein, the invention also encompasses variants, derivatives, and fragments thereof.

In the context of the invention, a "variant" of any given sequence is a sequence in which the specific sequence of residues (whether amino acid or nucleic acid residues) has been modified in such a manner that the polypeptide or polynucleotide in question retains at least one of its endogenous functions. A variant sequence can be obtained by addition, deletion, substitution, modification, replacement and/or variation of at least one residue present in the naturally occurring polypeptide or polynucleotide.

The term "derivative" as used herein in relation to proteins or polypeptides of the invention includes any substitution of, variation of, modification of, replacement of, deletion of and/or addition of one (or more) amino acid residues from or to the sequence, providing that the resultant protein or polypeptide retains at least one of its endogenous functions.

Typically, amino acid substitutions may be made, for example from 1, 2 or 3, to 10 or 20 substitutions, provided that the modified sequence retains the required activity or ability. Amino acid substitutions may include the use of non-naturally occurring analogues.

Proteins used in the invention may also have deletions, insertions or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent protein. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity and/or the amphipathic nature of the residues as long as the endogenous function is retained. For example, negatively charged amino acids include aspartic acid and glutamic acid; positively charged amino acids include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values include asparagine, glutamine, serine, threonine and tyrosine.

Conservative substitutions may be made, for example according to the table below. Amino acids in the same block in the second column and in the same line in the third column may be substituted for each other:

ALIPHATIC	Non-polar	G A P
		I L V
	Polar - uncharged	C S T M
		N Q
	Polar - charged	D E
		K R H
AROMATIC		F W Y

Typically, a variant may have a certain identity with the wild type amino acid sequence or the wild type nucleotide sequence.

In the present context, a variant sequence is taken to include an amino acid sequence which may be at least 50%, 55%, 65%, 75%, 85% or 90% identical, suitably at least 95%, 96% or 97% or 98% or 99% identical to the subject sequence. Although a variant can also be considered in terms of similarity (i.e. amino acid residues having similar chemical properties/functions), in the context of the present invention it is preferred to express in terms of sequence identity.

In the present context, a variant sequence is taken to include a nucleotide sequence which may be at least 50%, 55%, 65%, 75%, 85% or 90% identical, suitably at least 95%, 96% or 97% or 98% or 99% identical to the subject sequence. Although a variant can also be considered in terms of similarity, in the context of the present invention it is preferred to express it in terms of sequence identity.

Suitably, reference to a sequence which has a percent identity to any one of the SEQ ID NOs detailed herein refers to a sequence which has the stated percent identity over the entire length of the SEQ ID NO referred to.

Sequence identity comparisons can be conducted by eye, or more usually, with the aid of readily available sequence comparison programs. These commercially available computer programs can calculate percent identity between two or more sequences.

Percent identity may be calculated over contiguous sequences, i.e. one sequence is aligned with the other sequence and each amino acid or nucleotide in one sequence is directly compared with the corresponding amino acid or nucleotide in the other sequence, one residue at a time. This is called an “ungapped” alignment. Typically, such ungapped alignments are performed only over a relatively short number of residues.

Although this is a very simple and consistent method, it fails to take into consideration that, for example, in an otherwise identical pair of sequences, one insertion or deletion in the amino acid or nucleotide sequence may cause the following residues or codons to be put out of alignment, thus potentially resulting in a large reduction in percent identity when a global alignment is performed. Consequently, most sequence comparison methods are designed to produce optimal alignments that take into consideration possible insertions and deletions without penalising unduly the overall identity score. This is achieved by inserting “gaps” in the sequence alignment to try to maximise local identity.

However, these more complex methods assign “gap penalties” to each gap that occurs in the alignment so that, for the same number of identical amino acids or nucleotides, a sequence alignment with as few gaps as possible, reflecting higher relatedness between the two compared sequences, will achieve a higher score than one with many gaps. “Affine gap costs” are typically used that charge a relatively high cost for the existence of a gap and a smaller penalty for each subsequent residue in the gap. This is the most commonly used gap scoring system. High gap penalties will of course produce optimised alignments with fewer gaps. Most alignment programs allow the gap penalties to be modified. However, it is preferred to use the default values when using such software for sequence comparisons. For example when using the GCG Wisconsin Bestfit package the default gap penalty for amino acid sequences is -12 for a gap and -4 for each extension.

Calculation of maximum percent identity therefore firstly requires the production of an optimal alignment, taking into consideration gap penalties. A suitable computer program for carrying out

such an alignment is the GCG Wisconsin Bestfit package (University of Wisconsin, USA; Devereux et al. (1984) *Nucleic Acids Research* 12: 387). Examples of other software that can perform sequence comparisons include, but are not limited to, the BLAST package (see Ausubel et al. (1999) *ibid* – Ch. 18), FASTA (Atschul et al. (1990) *J. Mol. Biol.* 403-410), EMBOSS Needle (Madeira, F., et al., 2019. *Nucleic acids research*, 47(W1), pp.W636-W641) and the GENEWORKS suite of comparison tools. Both BLAST and FASTA are available for offline and online searching (see Ausubel et al. (1999) *ibid*, pages 7-58 to 7-60). However, for some applications, it is preferred to use the GCG Bestfit program. Another tool, BLAST 2 Sequences, is also available for comparing protein and nucleotide sequences (FEMS Microbiol. Lett. (1999) 174(2):247-50; FEMS Microbiol. Lett. (1999) 177(1):187-8).

Although the final percent identity can be measured, the alignment process itself is typically not based on an all-or-nothing pair comparison. Instead, a scaled similarity score matrix is generally used that assigns scores to each pairwise comparison based on chemical similarity or evolutionary distance. An example of such a matrix commonly used is the BLOSUM62 matrix (the default matrix for the BLAST suite of programs). GCG Wisconsin programs generally use either the public default values or a custom symbol comparison table if supplied (see the user manual for further details). For some applications, it is preferred to use the public default values for the GCG package, or in the case of other software, the default matrix, such as BLOSUM62.

Once the software has produced an optimal alignment, it is possible to calculate percent sequence identity. The software typically does this as part of the sequence comparison and generates a numerical result. The percent sequence identity may be calculated as the number of identical residues as a percentage of the total residues in the SEQ ID NO referred to.

“Fragments” are also variants and the term typically refers to a selected region of the polypeptide or polynucleotide that is of interest either functionally or, for example, in an assay. “Fragment” thus refers to an amino acid or nucleic acid sequence that is a portion of a full-length polypeptide or polynucleotide.

Such variants, derivatives, and fragments may be prepared using standard recombinant DNA techniques such as site-directed mutagenesis. Where insertions are to be made, synthetic DNA encoding the insertion together with 5' and 3' flanking regions corresponding to the naturally-occurring sequence either side of the insertion site may be made. The flanking regions will contain convenient restriction sites corresponding to sites in the naturally-occurring sequence so that the sequence may be cut with the appropriate enzyme(s) and the synthetic DNA ligated into the cut.

The DNA is then expressed in accordance with the invention to make the encoded protein. These methods are only illustrative of the numerous standard techniques known in the art for manipulation of DNA sequences and other known techniques may also be used.

The invention will be now illustrated by the following examples.

Examples

Materials and methods

Plasmids used as Cas9 templates

The plasmids used for AAV vectors production derived from a pAAV2.1 plasmid that contains the inverted terminal repeats of AAV serotype 2.

The AAV vector plasmid required to generate AAV-SpCas9 contains the hybrid liver promoter (HLP) and a synthetic pA sequence.

The AAV vector plasmid required to generate AAV-gRNA-donorDsRed contains: the U6 promoter, a specific gRNA and PAM sequences, and the chimeric gRNA scaffold; a splice acceptor signal, exon 14 of *mAlb*, the T2A linker, the DsRed coding sequence [CDS (NCBI ref. MK301207.1)], the Woodchuck hepatitis virus post-transcriptional regulatory element (WPRE), the bovine growth hormon polyA (BGH polyA), and a stop codon all surrounded by the inverted gRNA and PAM sequences.

The AAV vector plasmid required to generate AAV-gRNA-donorARSB contains: the U6 promoter, a specific gRNA and PAM sequences, and the chimeric gRNA scaffold; a splice acceptor signal, exon 14 of *mAlb*, the T2A linker, the human ARSB CDS (NCBI ref. NM_000046.5), the BGH polyA, and a stop codon all surrounded by the inverted gRNA and PAM sequences.

The AAV vector plasmid required to generate AAV-gRNA-Cas9 contains: the U6 promoter, a specific gRNA, and the chimeric gRNA scaffold; the hybrid liver promoter (HLP), spCas9 and a synthetic pA sequence.

The AAV vector plasmid required to generate AAV -donorFVIII contains: a splice acceptor signal, exon 14 of *mAlb*, the T2A linker, the human FVIII B-domain deleted codon optimized sequence (published in [33]), the BGH polyA, and a stop codon all surrounded by the inverted gRNA and PAM sequences.

The mouse albumin (*mAlb*) gRNAs (Tables 1, 3) were designed using the Benchling gRNA design tool (www.benchling.com), selecting the gRNAs with the best predicted on-target and off-target

scores, targeting intron 13 of *mAlb* or intron 12, 13 or 14 of human albumin (*hALB*). The scramble gRNA was designed to not align with any sequences in the mouse genome.

AAV Vector Production and Characterization

AAV vectors were produced by the TIGEM AAV Vector Core by triple transfection of HEK293 cells followed by two rounds of CsCl₂ purification [34]. For each viral preparation, physical titers (GC/mL) were determined by averaging the titer achieved by dot-blot analysis [39] and by PCR quantification using TaqMan (Applied Biosystems, Carlsbad, CA, USA) [34]. The probes used for dot-blot and PCR analyses were designed to anneal with the IRBP promoter for the pAAV2.1-IRBP-SpCas9-spA vector, the HLP promoter for the pAAV2.1-HLP-SpCas9-spA vector and the bGHpA region for the donor DNA vectors. The length of probes varied between 200 and 700 bp.

Culture and Transfection of HEK293 Cells

HEK293 cells were maintained in DMEM containing 10% fetal bovine serum (FBS) and 2 mM L-glutamine (Gibco, Thermo Fisher Scientific, Waltham, MA, USA). Cells were plated in 6-well plates (1×10^6 cells/well), and transfected 16 hr later with the plasmids encoding for Cas9 and the different gRNAs and donor DNAs, using the calcium phosphate method (1 to 2 mg/ 1×10^6 cells); medium was replaced 4 hr later. Maximum material transfected was 3 μ g. In all cases, quantity of plasmid DNA was equilibrated between wells, using an empty vector when necessary.

Cytofluorimetric Analysis

HEK293 cells, plated in 6-well plates, were washed once with PBS, detached with trypsin 0.05% EDTA (Thermo Fisher Scientific, Waltham, MA USA), washed twice with PBS, and resuspended in sorting solution containing PBS, 5% FBS and 2.5 mM EDTA. Cells were analyzed on a BD FACS ARIA III (BD Biosciences, San Jose, CA, USA) equipped with BD FACSDiva software (BD Biosciences) using appropriate excitation and detection settings for EGFP and DsRed. Thresholds for fluorescence detection were set on untransfected cells, and a minimum of 10,000 cells/sample were analyzed. A minimum of 50,000 GFP⁺ or GFP⁺/DsRed⁺ cells/sample were sorted and used for DNA extraction.

Mouse liver cryosections and fluorescence imaging

To evaluate DsRed expression in the liver after HITI, C57BL/6 mice were injected at p1 and sacrificed one month after injection by cardiac perfusion. A small piece of each liver lobe was dissected and fixed in 4% paraformaldehyde overnight. After fixation, pieces were infiltrated with 15% sucrose overday and 30% sucrose overnight before being included in O.C.T. matrix (Kalttek, Padua, Italy) for cryo-sectioning. Liver cryosections were cut at 6 μ m thickness, distributed on slides

and mounted with Vectashield supplemented with DAPI (Vector Lab, Peterborough, UK, #H-1200). Then, cryosections were analyzed under a Confocal microscope LSM 700 (Leica Microsystems, Wetzlar, Germany) at 20x using appropriate excitation and detection settings.

Characterization of integration junctions

DNA was extracted from liver tissue using DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) according to manufacturer's protocol.

For the T7 cleavage assay, 100 ng of DNA were used for PCR amplification of the region comprising the Cas9 target site in the mouse *Alb* intron 13 using specific primers (Table 2), which generate a PCR product of 652bp. PCR products were examined by T7 endonuclease I assay according to manufacturer's recommendations. Briefly, DNA was de-annealed and re-annealed by a slow temperature gradient with NEBuffer 2 (New England BioLabs, Ipswich, MA, USA) in a thermocycler. Samples were then incubated at 37 °C for 30 minutes with 1 µL of T7 Endonuclease (NEB, #M0302L) and analyzed in a 2% agarose gel. PCR products were also used for Sanger sequencing (Eurofins Genomics, Ebersberg, Germany) and then processed using the SYNTHOGO software (<https://ice.synthego.com/#/>) for analysis of the indel frequency.

To detect the integration of the donor into the 3' mouse Albumin locus (end of intron 13), 100 ng of the extracted DNA was used for PCR amplification of the HITI junctions using specific primers (Table 2). PCR products were analyzed in a 2% agarose gel and further cloned into PCR-Blunt II-TOPO (Invitrogen, Carlsbad, CA, USA). Single clones were then used for Sanger sequencing (Eurofins Genomics, Ebersberg, Germany) to characterize the integration of the donor.

Plasma collection and F8 assays

Nine parts of blood were collected by retro-orbital withdrawal into one part of buffered trisodium citrate 0.109M (5T31.363048; BD, Franklin Lakes, NJ, USA). Blood plasma was collected after samples centrifugation at 3000 rpm at 4°C for 15 minutes.

To evaluate F8 activity chromogenic assay was performed on plasma samples using a Coatest® SP4 FVIII-kit (K824094; Chromogenix, Werfen, Milan, Italy) according to manufacturer's instructions. Standard curve was generated by serial dilution of commercial human F8 (Refacto, Pfizer). Results are expressed as International Units (IU) per deciliter (dl).

RESULTS

Example 1

HITI mediated DsRed integration in newborn wild type mice at the mouse 3' *Alb* (*mAlb*) locus.

Inventors performed *in vivo* experiments to knock-in the reporter *DsRed* transgene at the 3' *mAlb* locus in wild type newborn mice as proof of concept (Fig. 1A). For this purpose, three different AAV8 vector were generated: one vector encoding for *SpCas9* under the expression of the hybrid liver promoter (HLP), a vector containing the HITI donor *DsRed* coding sequence (CDS) and a vector containing either the U6-gRNA or U6-scRNA expression cassette. Specifically, the donor DNA cassette includes a synthetic splicing acceptor signal (SAS), the last Albumin exon (ex 14) and the coding sequence for DsRed followed by the T2A sequence. Sequences are provided below.

Wild-type (WT) C57BL/6 mice were divided into two different treatment groups (gRNA or scRNA) and received a mixture of vectors at 1:1 ratio via the temporal vein at post-natal day 1 (p1). To integrate the *DsRed* CDS at the 3' *mAlb* locus, the gRNA group was injected with the vector encoding for *SpCas9* and the vector carrying the HITI donor together with the U6 promoter and the gRNA sequence. As a negative control, the scRNA group was treated following the same experimental scheme, but the vector carrying the HITI donor contained the U6-scRNA expression cassette. Animals were sacrificed 4 weeks post-injection and DNA was extracted from liver samples to evaluate the *SpCas9* cleavage efficiency and to PCR amplify the integration junctions using specific primers (Table 2). As expected, the *SpCas9* cleavage only occurred in gRNA-treated animals, and not in the scRNA group (Fig. 1B). Moreover, PCR analysis showed the 5' and the 3'-junction products at the targeting site in gRNA-treated animals, but not in scRNA-treated animals (Fig. 1C). Consistent with these results, microscopy images of liver cryo-sections revealed that DsRed was highly expressed in gRNA-treated animals but completely absent in scRNA-treated controls (Fig. 1D). All together, these data indicated that the HITI is a suitable platform for *DsRed* integration and expression at the 3' *mAlb* locus.

Example 2

HITI mediated ARSB delivery in newborn MPSVI mice.

Next, inventors tested if HITI at the 3' *mAlb* locus in newborn mice results in stable and therapeutically relevant levels of *Arylsulfatase B* (*ARSB*), the lysosomal hydrolase defective in the lysosomal storage disease mucopolysaccharidosis type VI (MPS VI). Since *ARSB* is secreted in the bloodstream and can be non-invasively measured, it can be used as readout of liver transduction [35]. *ARSB* deficiency results in abnormal glycosaminoglycan (GAG) storage and urinary secretion, which is a useful biomarker of MPS VI [36]. Inventors generated AAV vectors carrying the donor DNA cassette as described above: including a synthetic splicing acceptor signal (SAS), the last Albumin exon (ex 14) and the coding sequence for human *ARSB* (*hARSB*) followed by the T2A sequence, as well as a gRNA expression cassette for either the gRNA or the scramble sequence as control (Fig.

2A). The gRNA donor vector or the scrRNA vector were systemically co-delivered in combination with the HLP-SpCas9 vector (Fig. 2A) in neonatal MPS VI mice (p1-2). Serum ARSB activity was measured in gRNA-treated MPS VI mice at levels that were higher than normal littermates (Fig. 2B) and remained stable over time up to one year of age. Serum ARSB activity in scramble treated or untreated MPS VI mice was undetectable. Importantly, while no significant differences in urinary GAGs were observed between scramble and gRNA-treated groups at p60, AAV-HITI-mediated ARSB expression was able to normalize urinary GAGs from p90 up to p360 (Fig. 2C).

Dose-response of AAV-HITI

Inventors tested three doses of AAV-HITI to treat *Arylsulfatase B* (*Arsb*)^{-/-} mouse model of Mucopolysaccharidosis type VI (MPSVI) by integrating a donor DNA carrying the promoter less coding sequence of ARSB at the 3' *Albumin* locus. Animals were administered with three doses of AAV-HITI at 1-2 days old: 1.2E+14 total genome copies (GC)/Kg (high dose or HD) – 3.9E+13 total GC/Kg (medium dose or MD) and 1.2E+13 total GC/Kg (low dose or LD). Preliminary results from inventors' ARSB immune assay on serum samples show that the HD and the MD achieve supraphysiological levels of the secreted active ARSB at 30 days of age (Figure 7). The LD also induced secretion of active ARSB at variable levels (Figure7).

Example 3

HITI mediated F8 codopV3 integration in newborn hemophilic mice at the mouse 3' *Alb* (*mAlb*) locus.

Inventors performed *in vivo* experiments to knock-in the *F8 CodopV3* transgene at the 3' *mAlb* locus in hemophilic newborn mice. For this purpose, three different AAV8 vector were generated: two vectors encoding for *SpCas9* under the expression of the hybrid liver promoter (HLP) carrying the U6-gRNA or U6-scrRNA expression cassette and one vector containing the HITI donor *F8CodopV3* coding sequence (CDS). Specifically, the donor DNA cassette includes a synthetic splicing acceptor signal (SAS), the last Albumin exon (ex 14) followed by the T2A sequence and the coding sequence for F8 CodopV3 (Figure 3A). Hemophilic mice were divided into two different treatment groups (gRNA or scrRNA) and received a mixture of vectors at 1:1 ratio via the temporal vein at post-natal day 1 (p1). To integrate the *F8* CDS at the 3' *mAlb* locus, the gRNA group was injected with the vector encoding for *SpCas9* together with the U6 promoter and the gRNA sequence and the vector carrying the HITI donor. As a negative control, the scrRNA group was treated following the same experimental scheme, where the vector carrying the *SpCas9* has the U6-scrRNA expression cassette. Blood plasma samples were collected 4 weeks following vector administration.

F8 activity was monitored using the functional chromogenic assay and showed that F8 activity levels was 20% compared to unaffected controls (Figure 3B).

gRNA sequences

Tables

gRNA	Sequence of the gRNA	Position	ON target score	Off target score	Sequence of target sequence in direction 5'-3'
1	5'GTATTTAATAGGCAGCA GTG-3' (SEQ ID NO: 2)	Intron 13 (position 283, minus strand)	78.3	64.2	5'- CACTGCTGCCTATTA AATAC-3' [SEQ ID N.1] (5'->3', plus strand)

Table 1 gRNA sequence at intron13 of murine albumin locus

Primer name	Primer sequence	PCR product size (bp)
<i>Alb</i> intron 13 indel Fwd	5'-TGGATACATGTTGCAAGGCTGC-3' [SEQ ID N.3]	652
<i>Alb</i> intron 13 indel Rev	5'-GGCGTCTTTGCATCTAGTGACA-3' [SEQ ID N.4]	
<i>Alb</i> HITI 5' junction Fwd	5'-CACGTGGTCAGGTGTAGCTC-3' [SEQ ID N.5]	196
<i>Alb</i> HITI 5' junction Rev	5'-TGGAGAGAAAGGCAAAGTGGA-3' [SEQ ID N.6]	
<i>Alb</i> HITI 3' junction Fwd	5'-CAGCAAGGGGGAGGATTGG-3' [SEQ ID N.7]	169

<i>Alb</i> HITI 3' junction Rev	5'-GAAACATTTTCAGGGCAAGGT-3' [SEQ ID N.8]	
---------------------------------	---	--

Table 2 Primer used for the 3'mAlbumin locus.

gRNAs with the best predicted on-target and off-target scores targeting the ninth, eleventh, twelfth, thirteenth or fourteenth intron of human albumin (*hALB*) have been designed and are reported in Table 3.

gRNA	Sequence of the gRNA	Position	ON target score	Off target score	Sequences target sequence in 5'-3' direction
1	5'- AATCTCTGGACGGAAGCT CA-3' (SEQ ID NO: 10)	Intron 13 (position 456, minus strand)	66,4	41,9	5'- TGAGCTTCCGTCCAGAGA TT-3' [SEQ ID N.9] (5'->3', plus strand)
2	5'- ACAGTATGGCACAATAGA GC-3' (SEQ ID NO: 12)	Intron 13 (position 173, minus strand)	53,5	44,4	5'- GCTCTATTGTGCCATACT GT-3' [SEQ ID N.11] (5'->3', plus strand)

3	5'- ACACTACATAACGTGATG AG-3'	Intron 12 (position 927, plus strand)	65,3	85,0	5'- ACACTACATAACGTGATG AG-3' [SEQ ID N.13] (3'->5', minus strand)
4	5'- AAATAGTTTAGAATAGTG GT-3' (SEQ ID NO: 14)	Intron 14 (position 123, minus strand)	66,3	57,3	5'- ACCACTATTCTAACTAT TT-3' [SEQ ID N.15] (5'->3', plus strand)
5	5'- GTGGGCTGTAATCATCGT CT-3' (SEQ ID NO: 16)	Intron 12 (position 538, plus strand)	58,3	46,9	5'- GTGGGCTGTAATCATCGT CT -3' (5'->3', plus strand)
6	5'- TATTGGCAGTCAAGGCCC CG-3' (SEQ ID NO: 17)	Intron 11 (position 152, plus strand)	N.A.	N.A.	5'- TATTGGCAGTCAAGGCCC CG -3' (5'->3', plus strand)

7	5'- TCGAATGTATTGTGACAG AG-3' (SEQ ID NO: 18)	Intron 9 (position 733, plus strand)	71.0	41.1	5'- TCGAATGTATTGTGACAG AG-3' (5'→3', plus strand)
---	--	--	------	------	--

Table 3: gRNAs targeting human Albumin. Positions are referred to the first nucleotide of each intron. ON and OFF target scores are prediction calculated using Benchling. N.A.: not available
gRNAs are designed on either strand of genomic DNA and are indicated in 5'-3' orientation
SEQ IDs are indicated in the 5'-3' orientation.

Serum albumin levels

Serum albumin levels were collected from blood samples at p360 from treated and control mice (Figure 4) with the ELISA Kit (Abcam, 108791, Cambridge, UK) following the manufacturer's instructions. Serum albumin levels were found to be similar independently of the group of treatment meaning that inventors' AAV-HITI doesn't affect the expression of the endogenous protein.

Alfa-fetoprotein levels

Elevated alfa-fetoprotein (AFP) levels have been reported to be associated with hepatocellular carcinoma (HHC) in mice (Ferla et al., Molecular Therapy: Methods & Clinical Development 2021). inventors measured AFP levels in serum samples at p360 from treated and control mice using the mouse Alfa-Fetoprotein/AFP Quantikine Elisa Kit (R&D Systems, Minneapolis, MN, USA), following the manufacturer's instructions. Mouse AFP levels were found to be increased in AAV-HITI gRNA treated mice but not in scRNA and controls (Figure5).

Off-target analysis

To investigate potential chromosomal aberration (as translocation events) mediated by the on-target site (OMT) inventors performed CAST-seq analysis, a technique previously described (Turchiano et al., 2021) on inventors' AAV-HITI gRNA liver DNA samples while AAV-HITI scRNA and untreated liver DNA samples were used as controls. CAST-seq analysis data indicate that inventors' AAV-HITI gRNA samples present deletions events at the on-target site while no OMT where found (Figure6).

Example 4

Selection of gRNAs targeting the 3'human *Albumin* (*ALB*) locus

Inventors chose one gRNA targeting intron 13 of *ALB* and eight gRNAs targeting introns 9 and from 11 to 13 of *ALB* using Benchling and/or CHOPCHOP softwares (Table 4). The *in-silico* selection was based on i) low number of predicted off-targets and ii) high efficiency at targeting the desired locus. Plasmids encoding for Cas9-EGFP under the CBh promoter and one of the selected gRNAs or the scRNA under the human U6 promoter were transfected in HEPA 1-6 cells or HEK293 cells to target the *Alb* locus or the *ALB* locus, respectively. HEPA 1-6 cells were transfected with 1 µg of plasmidic DNA using Lipofectamine LTX (Thermo Fisher Scientific, Waltham, MA, USA) while HEK293 cells were transfected with 1 µg of plasmidic DNA using calcium phosphate. DNA was extracted from sorted cells expressing Cas9-EGFP and the genomic region recognized by the gRNA was PCR-amplified. The PCR product was digested with the T7 enzyme (Neb, Ipswich, MA, USA) to detect Cas9-mediated INDELS. The same PCR product was also Sanger sequenced to perform quantification of INDELS using the ICE software from Synthego. gRNA0 (*Alb* intron 13), and gRNA3 and 5 (*ALB* intron 12) induce high Cas9-mediated INDELS while lower levels were detected using gRNA2 (*ALB* intron 13) and no INDELS were detected with either gRNA1 (*ALB* intron 13) and gRNA4 (*ALB* intron 14) (Table 4 and Figure 8). The allelic variation frequency of the sequence recognized by gRNAs targeting the *ALB* locus was analyzed using the human genome aggregated database (gnomAD) version 3.1.2 for selected gRNAs. The highest detected allelic variation frequency is 1 SNP every 10³ alleles (gRNA 3 and 6) and, importantly, no variant is present in homozygosity.

Integration efficiency at the 3'*Alb* locus and the 3'*ALB* locus

Inventors evaluated the HITI-mediated integration efficiency at the 3'*Alb* locus and at the 3'*ALB* locus by generating HITI donors flanked by the inverted sequences of gRNA0 to integrate at the 3'*Alb* locus or gRNA3 or 5 to integrate at the 3'*ALB* locus. The donors encode for a synthetic splicing acceptor signal, Exon 14 of *Alb* or Exons 13-14 of *ALB* linked with a T2A skipping peptide to the fluorescent reporter DsRed coding sequence. inventors used Lipofectamine LTX (Thermo Fisher Scientific, Waltham, MA, USA) to transfect HEPA1-6 cells with 1 µg of plasmidic DNA encoding Cas9-EGFP and gRNA0 and 1 µg of plasmidic DNA encoding the donor DNA flanked by gRNA0; Human Hepatoma cell line 7 (HUH7) were transfected similarly with 1 µg of plasmid encoding Cas9-EGFP and gRNA 3 + 1 µg of plasmid encoding the HITI donor flanked by gRNA3, or 1 µg of plasmid encoding Cas9-EGFP and gRNA 5 + 1 µg of plasmid encoding the HITI donor flanked by gRNA5. Cells transfected with the HITI donors and plasmidic DNA encoding Cas9-EGFP and scRNA were used to normalize DsRed fluorescence and quantify productive HITI donor integration only.

Fluorescence-activated cells sorting analysis show gRNA0 and gRNA3 induce productive integration of the HITI donor at the 3' *Alb* locus and the 3' *ALB* locus, respectively (Figure9).

Gene	ID gRNA	gRNA sequence	SEQ ID NO:	INDEL (% $\pm$ SEM)	Allelic variation frequency
<i>Alb</i> Intron 13	gRNA 0	5'- GTATTTAATAGGCAGCAGTGTGG- 3'	54	58.7 $\pm$ 6.2	/
<i>ALB</i> Intron 13	gRNA 1	5'- AATCTCTGGACGGAAGCTCACGG- 3'	92	Not detected	Not performed
<i>ALB</i> Intron 13	gRNA 2	5'- ACAGTATGGCACAATAGAGCAGG- 3'	93	20.7 $\pm$ 4.4	Not performed
<i>ALB</i> Intron 12	gRNA 3	5'- ACACTACATAACGTGATGAGAGG- 3'	94	54.0 $\pm$ 5.8	One SNP every 10 ³ -10 ⁵ alleles. 5 possible SNPs
<i>ALB</i> Intron 14	gRNA 4	5'- AAATAGTTTAGAATAGTGGTCCG- 3'	95	Not detected	Not performed
<i>ALB</i> Intron 12	gRNA 5	5'- GTGGGCTGTAATCATCGTCTAGG- 3'	96	54.3 $\pm$ 3.0	One SNP every 10 ⁴ -10 ⁵ alleles.

					5 possible SNPs
<i>ALB</i> Intron 11	gRNA 6	5'- TATTGGCAGTCAAGGCCCG <u>GAGG</u> - 3'	97	Ongoing	One SNP every 10 ³ - 10 ⁶ alleles. 4 possible SNPs
<i>ALB</i> intron 9	gRNA 7	5'- TCGAATGTATTGTGACAGAG <u>CGG</u> - 3'	98	Ongoing	One SNP every 10 ⁴ - 10 ⁶ alleles. 11 possible SNPs

Table 4: gRNAs targeting the 3' murine(*Alb*) or human (*ALB*) *Albumin* locus. The table shows all gRNAs (the PAM sequence is underlined) and the targeted intron of *Alb* or *ALB*, the % of INDEL shown as mean $\pm$ standard error of mean (n = 3 independent experiments), and the allelic variation frequency. SEM = standard error of mean

EXAMPLE 5

Precision of the AAV-HITI platform at the 3'mAlb locus.

To evaluate the precision of inventors' AAV-HITI strategy inventors performed several molecular analysis. First, inventors studied the cutting efficiency (indel %) of inventors' selected gRNA. Illumina-seq NGS analysis were performed on genomic DNA extracted from livers of AAV-HITI gRNA or AAV-HITI scRNA treated MPSVI mice. inventors found 29 % of indels only in AAV-HITI gRNA treated mice (Figure 10A). Moreover, inventors also evaluated if inventors could find portion of AAV vector genome integrated at the on-target site upon cas9 induced double strand-breaks. Using the full AAV vector genome of either the HITI donor DNA or the Cas9 as reference, inventors aligned the reads generated from the Illumina-seq NGS experiment. inventors were able to align reads covering different portion of the given AAV vector genome with most of the reads covering the ITRs regions (Figure 10B). Next, we evaluated HITI mediated integration in comparison to ITRs-mediated integration. For this purpose, we generated a donor DNA with same structure of our HITI gRNA

donor DNA but not flanked by the inverted gRNA sites at its 5' and 3' extremities (ITR donor). This construct was next produced as AAV8 and used in vivo. Wild-type mice were injected by temporal vein at p1-2 with a mixture of AAV-Cas9 and AAV-ITR donor containing the Ds-Red coding sequence (as previously described for the HITI donor DNA). In parallel, a second group mice was injected with a combination of AAV-Cas9 and AAV-HITI gRNA donor. Both groups were sacrificed 1-month post-treatment, DNA was extracted from the liver of all the animals treated and was used for further molecular analysis.

We PCR amplified both the 5' and 3' junction's sites between the inserted donor DNA and the endogenous locus with specific primers (**Table 2**).

As visible in the agarose gel (figure 10C) we were able to PCR amplify the junctions bands (5' and 3') of the expected size, in accordance to the donor (HITI or ITR donor) received. Interestingly, in HITI treated mice we also observed an upper but fainter band, at the same size of the one observed in the AAV-ITR donor gRNA treated samples. Sanger sequencing analysis, performed on both the upper and the expected size bands revealed that in HITI-donor DNA treated animals, donor DNA integration is also occurring through ITR.

Next, inventors assessed potential gRNA off-target activity. To this aim inventors selected the top 10 predicted off-target (**Table5**) using CRISPOR. NGS analysis performed on PCR bands obtained from liver genomic DNA, for each the selected off-target locus (in both gRNA and scRNA treated samples) resulted in very low or undetectable off-target editing events (Figure 10D).

Table 5

gRNA specificity	Region	Position	(5'-3') gRNA sequence + PAM	SEQ ID NO:	Mismatches	Off-target CDF score
ON-target	Intron: albumin	chr5: 90622727-90622747:-	GTATTTAATAGGCAGCAGTG TGG	54	-	-
OFF-target 1	Intron: Rik	chr5:151333345-151333367:-	TTACTTAATAAGCAGCAGTG TGG * * *	99	3	0.647
OFF-target 2	Intron: Lrr 1	chr12:69224137-69224159:-	GTTTTTAAAAAGCAGAAGTG GGG * * * *	100	4	0.646
OFF-target 3	Intergenic: Ppp1r3c-Tnks2	chr19:36774036-36774058:-	TTATCTAATAGACAGCAATG CGG * * * *	101	4	0.646
OFF-target 4	Intron: Zim2	chr7:6660686-6660708:+	GAATTTGATAGACAGCAGTG GGG * * *	102	3	0.557
OFF-target 5	Intron: Slc39a12	chr2:14426612-14426634:+	GTATTTAAGAGGCAGCAGTT TGG ** *	103	3	0.476
OFF-target 6	Intron: Gsted	chr3:132751735-132751757:-	AAATTTGATTGGCAGCAGTG TGG ** * *	104	4	0.474
OFF-target 7	Intron: Kcnc1	chr7:46060812-46060834:-	GTATTTAAAAGGCTGAAGTA AGG * * * *	105	4	0.464
OFF-target 8	Intergenic: Rik/Lhfpl3-Lhfpl3	chr5:23268978-23269000:+	ATATTCAAGTGGCAGCAGTG AGG * * **	106	4	0.446
OFF-target 9	Intron: Dpyd	chr3:119141669-119141691:+	ATATTTAATAGGCAACATTT AGG * * * *	107	4	0.395
OFF-target 10	Intron: HCn2	chr10:79561901-79561923:+	GGATTCAAGTAGGCAGCAGTTGGG * * * *	108	4	0.392

SEQUENCES

Sequences of above Example 1

5'-ITR

*CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGGCGACCTT
TGGTCGCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATC
ACTAGGGGTTTCCT* [SEQ ID N.110]

gRNA sequence for murine Albumin intron 13

CACTGCTGCCTATTAAATAC [SEQ ID N.1]

scRNA sequence

gactcgcgcgagtcgaggag [SEQ ID N.111]

Inverted gRNA sequence for murine Albumin intron 13 without PAM

CACTGCTGCCTATTAAATAC [SEQ ID N.1]

Inverted gRNA sequence for murine Albumin intron 13 + PAM sequence (underlined)

CCACACTGCTGCCTATTAAATAC [SEQ ID N.20]

Splice acceptor sequence

GATAGGCACCTATTGGTCTTACTGACATCCACTTTGCCTTTCTCTCCACAG [SEQ ID N.21]

Exon 14 murine Albumin

GGTCCAAACCTTGTCCTAGATGCAAAGACGCCTTAGCC [SEQ ID N.22]

Thosea asigna virus 2A (T2A) skipping peptide

**GGAAGCGGAGAGGGCAGAGGAAGTCTGCTAACATGCGGTGACGTCGAGGAGAAT
CCTGGACCT** [SEQ ID N.23]

Discosoma Red (DsRed) coding sequence

ATGGATAGCACTGAGAACGTCATCAAGCCCTTCATGCGCTTCAAGGTGCACATGGAGGGGCTC
CGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCAAGCCCTACGAGGGGCAC
CCAGACCGCCAAGCTGCAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCT
GTCCCCCAGTTCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCCGAC
TACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAAGTTCGAGGACG
GCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCACCTTCATCTACCACGT
GAAGTTCATCGGCGTGAAGTTCCTCCGACGGCCCCGTAATGCAGAAGAAGACTCTGGGC
TGGGAGCCCTCCACCGAGCGCCTGTACCCCGCGACGGCGTGCTGAAGGGCGAGATCCAC
AAGGCGCTGAAGCTGAAGGGCGGCGGCCACTACCTGGTGGAGTTCAAGTCAATCTACATGG
CCAAGAAGCCCGTGAAGCTGCCCCGGCTACTACTACGTGGACTCCAAGCTGGACATCACCTC
CCACAACGAGGACTACACCGTGGTGGAGCAGTACGAGCGCGCCGAGGCCCCGCCACCACCT
GTTCCAGTAG [SEQ ID N.24]

Woodchuck hepatitis virus post-transcriptional regulatory element (WPRE)

aatcaacctctggattacaaaatttgtaaagattgactggattcttaactatgttgctccttttacgctatgtggatacgcctttaatgcctttgtatc
atgctattgcttcccgatggccttcattttctcctccttgataaatcctgggtgctgtctctttatgaggagttgtggcccggtgtcaggcaacgtggc
gtggtgtgcactgtgtttgctgacgcaacccccactgggtggggcattgccaccacctgtcagctcctttccgggactttcgctttccccctccctat

tgccacggcggaactcatcgccgcctgccttgcctgctggacaggggctcggctgttgggcactgacaattccgtggtgttcggggaa
atcatcgctcttcttgcctgctgcctgtgttgcacctggattctgcgcgggacgtccttctgctacgtcccttcggccctcaatccagcggacc
tcttcccgccgctgctgcgggctctgcggcctcttccgcgtcttcg [SEQ ID N.25]

Bovine growth hormone Poly A (BGH pA)

GCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCCTCCCCCGTGCCTTCC
TTGACCCTGGAAGGTGCCACTCCCCTGTCCTTTCCTAATAAAATGAGGAAATTGCATC
GCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAG
GGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGA [SEQ ID N.26]

Human U6 promoter

ttcccatgattccttcataattgcatatacgatacaaggcgttagagagataattggaatttaattgactgtaaacacaaagatattagtaaa
aatacgtgacgtagaaaglaataatttcttgggtagttgcagttttaaataatgtaataatggactatcatatgcttacgtaacttgaaggt
atttcgatttcttggctttatatacttcttggaaaggacg. [SEQ ID N.27]

Chimeric gRNA scaffold

GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAA
GTGGCACCGAGTCGGTGC [SEQ ID N.28]

3'-ITR

AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAG
 GCCGGGCGACCAAAGGTCGCCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCG
 AGCGAGCGCGCAG [SEQ ID N.29]

Construct p1492_pTIGEM_mAlb3'HITIdonor(SAS_albex14_T2A_dsRED_bGHpA)+
U6gRNA mAlb3'

CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCCGGGCGTCGGGCGACCTT
TGGTCGCCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGGGAGTGGCCAACTCCATC
ACTAGGGGTTCTTGCTAGTGCTAGCGGCGCGCCTCTAGCCACACTGCTGCCTATTAAAT
ACGATAGGCACCTATTGGTCTTACTGACATCCACTTTGCCTTTCTCTCCACAGGGT
CCAAACCTTGTCCTAGATGCAAAGACGCCTTAGCCGGAAGCGGAGAGGGCAGAGG
AAGTCTGCTAACATGCGGTGACGTCGAGGAGAATCCTGGACCTATGGATAGCACTG
AGAACGTCATCAAGCCCTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCA
CGAGTTCGAGATCGAGGGCGAGGGCGAGGGCAAGCCCTACGAGGGGCACCCAGACCGCCAA
GCTGCAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCCAGTTC
CAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCCGACTACAAGAAGCTGT

CCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGA
CCGTGACCCAGGACTCCTCCCTGCAGGACGGCACCTTCATCTACCACGTGAAGTTCATCGGC
GTGAACTTCCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACTCTGGGGCTGGGAGCCCTCCA
CCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAGATCCACAAGGCGCTGAAGC
TGAAGGGCGGGCGGCCACTACCTGGTGGAGTTCAAGTCAATCTACATGGCCAAGAAGCCCGT
GAAGCTGCCCCGGCTACTACTACGTGGACTCCAAGCTGGACATCACCTCCCACAACGAGGACT
ACACCGTGGTGGAGCAGTACGAGCGCGCCGAGGCCCGCCACCACCTGTTCCAGTAGGATC
 Caatcaacctctggattacaaaatttgtgaaagattgactggtattcttaactatgttgctccttttacgctatgtggatacgtgctttaatgcctttgta
 tcatgctattgcttcccgatggctttcattttctcctctgtataaatcctgggtgctgtctctttatgaggagttgtggccggtgtcaggcaacgtgg
 cgtggtgtgcactgtgtttgtgacgcaacccccactggttggggcattgccaccacctgtcagctcctttccgggactttcgctttccccctcccta
 ttgccacggcggaactcatcgccgcctgccttggccgctgctggacaggggctcggctgttgggcactgacaattccgtggtgtgtcggggaa
 atcatgctcctttccttggctgctcgcctgtgttgcacctggattctgcgcgggacgtccttctgctacgtcccttcggccctcaatccagcggacc
 ttcttcccggcgctgctgcgggctctgcggcctcttccgctcttcgAGATCTGCCTCGACTGTGCCTTCTAGTTG
CCAGCCATCTGTTGTTTGCCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCC
CACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATT
CTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATA
GCAGGCATGCTGGGGACCACACTGCTGCCTATTAAATACGAGCTCTTGTCGAGGTCGA
 CATTAAATGAAGGGCGAATTCAATTTGAGGGCCTAAttcccatgattccttcataattgcatatacgatacaa
 ggcgttagagagataattggaattaatltgactgtaaacacaaagatattagtaaaaatacgtgacgtagaaagtaataatttcttgggtta
 gtttgcagttttaaaaattatgtttttaaattggactatcatatgcttacgtaacttgaaagtatttcgatttcttggctttatatacttgtggaaagga
 cgAAACACCGTATTTAATAGGCAGCAGTGGTTTTAGAGCTAGAAATAGCAAGTTAAAAT
AAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTGTTTTAG
 AGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTTTTTAGCGCGTGCGCCAATTCT
 GCAGACAAATGGCTCTAGAGGTACCAATTGAGGAACCCCTAGTGATGGAGTTGGCCAC
 TCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCCGACGC
 CCGGGCTTTGCCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAG [SEQ ID N.30]

Construct p1496_pTIGEM_mAlb3'HITIdonor(SAS_albex14_T2A_dsRED_bGHpA)+
 U6scrambleRNA mAlb3'

CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCCGGGCGTCGGGCGACCTT
TGGTCGCCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACCTCCATC
ACTAGGGGTTCTGCTAGTGCTAGCGGCGCGCCTCTAGCCACACTGCTGCCTATTAAAT
ACGATAGGCACCTATTGGTCTTACTGACATCCACTTTGCCTTTCTCTCCACAGGGT
CCAAACCTTGTCCTAGATGCAAAGACGCCTTAGCCGGAAGCGGAGAGGGCAGAGG

AAGTCTGCTAACATGCGGTGACGTCGAGGAGAATCCTGGACCTATGGATAGCACTG
AGAACGTCATCAAGCCCTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCA
CGAGTTCGAGATCGAGGGCGAGGGCGAGGGCAAGCCCTACGAGGGCACCCAGACCGCCAA
GCTGCAGGTGACCAAGGGCGGCCCTGCCCCTTCGCCTGGGACATCCTGTCCCCCAGTTC
CAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCCGACTACAAGAAGCTGT
CCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGA
CCGTGACCCAGGACTCCTCCCTGCAGGACGGCACCTTCATCTACCACGTGAAGTTCATCGGC
GTGAACTTCCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACTCTGGGCTGGGAGCCCTCCA
CCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAGATCCACAAGGCGCTGAAGC
TGAAGGGCGGCGGCCACTACCTGGTGGAGTTCAAGTCAATCTACATGGCCAAGAAGCCCGT
GAAGCTGCCCCGGCTACTACTACGTGGACTCCAAGCTGGACATCACCTCCCACAACGAGGACT
ACACCGTGGTGGAGCAGTACGAGCGCGCCGAGGCCCGCCACCACCTGTTCCAGTAGGATC
 Caatcaacctctggattacaaaatttgtgaaagattgactggtattcttaactatgttgctccttttacgctatgtggatacgtgctttaatgcctttgta
 tcatgctattgcttcccgtatggctttcattttctctcctgtataaatcctggtgctgtctctttatgaggagtgtggcccgtgtcaggcaacgtgg
 cgtggtgtgactgtgtttgtgacgcaacccccactggttggggcattgccaccacgtgcagctcctttccgggactttcgtttccccctcccta
 ttgccacggcggaactcatcgccgcctgccttgcctgctgacaggggctcggctgttgggcactgacaattcgtggtgtgtcggggaa
 atcatcgtcctttccttggtgctgcctgtgttgcacctggattctgcgcgggacgtccttctgctacgtcccttcggccctcaatccagcggacc
 ttcttcccgggctgctgcccgtctgcggcctcttccgcgtcttcgAGATCTGCCTCGACTGTGCCTTCTAGTTG
CCAGCCATCTGTTGTTTGCCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCC
CACTGTCCTTTCCTAATAAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATT
CTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATA
GCAGGCATGCTGGGGACCACACTGCTGCCTATTAATAACGAGCTCTTGTCGAGGTCTGA
 CCTGACCTCGAGGttcccatgattcctcatattgcatatacgatacaaggctgttagagagataaattggaattaatttgactgtaaa
 cacaagatattagtagtaaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaattatgttttaaatggactatcatat
 gcttaccgtaactgaaagtatttcgatttcttggctttatatacttgtggaaaggacgAAACACCGgactcgcgcgagtcgaggagG
TTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAG
TGGCACCGAGTCGGTGCTTTTTTTGTTTTAGAGCTAGAAATAGCAAGTCTAGAGGTACCA
 ATTGAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCAC
 TGAGGCCGGGCGACCAAAGGTGCGCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTG
 AGCGAGCGAGCGCGCAG [SEQ ID N.31]

Sequences of above Example 2

5'-ITR

**CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCCGGGCGTCGGGCGACCTT
TGGTCGCCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATC
ACTAGGGGTTTCCT** [SEQ ID N.110]

gRNA sequence for murine Albumin intron 13 (orientation 5'-3' on plus strand)

CACTGCTGCCTATTAAATAC [SEQ ID N.1]

scRNA sequence

gactcgcgcgagtcgaggag [SEQ ID N.111]

Inverted gRNA sequence for murine Albumin intron 13 without PAM

CACTGCTGCCTATTAAATAC [SEQ ID N.1]

Inverted gRNA sequence for murine Albumin intron 13 + PAM sequence

CCACACTGCTGCCTATTAAATAC [SEQ ID N.20]

Splice acceptor sequence

GATAGGCACCTATTGGTCTTACTGACATCCACTTTGCCTTTCTCTCCACAG [SEQ ID N.21]

Exon 14 murine Albumin

GGTCCAAACCTTGTCCTAGATGCAAAGACGCCTTAGCC [SEQ ID N.22]

Thosea asigna virus 2A (T2A) skipping peptide

**GGAAGCGGAGAGGGCAGAGGAAGTCTGCTAACATGCGGTGACGTCGAGGAGAAT
CCTGGACCT** [SEQ ID N.23]

ARSB coding sequence

atgggtccgcgcggcgcgagcttgcggcgaggccccggacctcgcggtgctcctcccgtcgtcctccgctgctgctgctgttg
ttggcgccgcccgggctcggcgccggggccagccggccgccccacctggtcttcttctgctggcagacgacctaggctggaacgacgtcggt
tccacggctcccgcacccgcacgcccacctggacgcgctggcgccggcggggtgctcctggacaactactacacgcagccgctgtgcac
gccgtcgcggagccagctgctcactggccgctaccagatccgtacaggtttacagcaccataataatctggccctgtcagcccagctgtgttcctc
tggatgaaaaactcctgccccagctcctaaaagaagcaggttatactacccatattggtcggaataatggcacctgggaatgtaccggaaagaatg
cctccaacccgcccaggatttgatacctactttggatatctcctgggtagtgaagattatttcccatgaacgctgtacattaattgacgctctgaat
gtcacacgatgtgctcttgattttcgagatggcgaagaagttgcaacaggatataaaaatatgtattcaacaacatattcaccaaaagggtatag
ccctcataactaacatccaccagagaagcctctgtttctctaccttgcctcctcagctgtgcatgagccccttcaggtccctgaggaatacttgaag
ccatatgactttatccaagacaagaacaggcatcactatgcaggaatggtgtcccttatggatgaagcagtaggaaatgtcactgcagcttataaa

agcagtgggctctggaacaacacggtgttcattttctacagataacggagggcagactttggcaggggtaataactggcccccttcgaggaa
 gaaaatggagcctgtgggaaggaggcgctccgaggggtgggctttgtggcaagccccttgctgaagcagaagggcgtaagaaccgggagc
 tcatccacatctctgactggctgccaacactcgtgaagctggccaggggacacaccaatggcacaagcctctggatggcttcgacgtgtggaa
 aaccatcagtgaaggaagcccatccccagaattgagctgctgcataatattgacccgaacttcgtggactcttcaccgtgtcccaggaacagca
 tggctccagcaaaggatgactcttctctccagaatattcagcctttaacacatctgtccatgctgcaattagacatggaaattggaaactcctcag
 ggctacccaggtgtggttactggttccctccaccgtctcaatacaatgtttctgagataccctcatcagaccaccaaccaagaccctctggctct
 ttgatattgatcgggaccctgaagaaagacatgacctgtccagagaatatcctcacatcgtcacaagctcctgtcccgctacagttctaccata
 aacactcagtcctcggtacttccctgcacaggacccccgctgtgatcccaaggccactggggtgtggggcccttgatgtag [SEQ ID
 N.33]

Bovine growth hormone poli A (BGH pA)

GCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCCTCCCCCGTGCCTTCC
TTGACCCTGGAAGGTGCCACTCCCCTGTCCTTTCCTAATAAAATGAGGAAATTGCATC
GCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAG
GGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGA [SEQ ID N.26]

Human U6 promoter

ttcccatgattccttcataatttgcatalacgatacaaggctgttagagagataattggaatttaattgactglaaacacaaagatattagtaaa
aatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaataattgttttaaaatggactatcatatgcttacgtaacttgaaggt
atttcgatttcttggctttatatacttgttggaaggacg [SEQ ID N.27]

Chimeric gRNA scaffold

GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAA
GTGGCACCGAGTCGGTGC [SEQ ID N.28]

3'-ITR

AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAG
 GCCGGGCGACCAAAGGTCGCCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCG
 AGCGAGCGCGCAG [SEQ ID N.29]

Construct p1479_pTIGEM_mAlb3'HITIdonor(SAS_albex14_T2A_ARSB_bGHpA)+
U6gRNA mAlb3'

CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGGCGACCTT
TGGTCGCCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACCTCCATC
ACTAGGGGGTTTCCTgttagttaatgattaacccgccatgctacttatctacgtagccatgctctaggaagatcggaattcgccctaaact
 agtaacggccgccagtgtgctggaattcaggtgacctcgagttcccatgattccttcataatttgcatalacgatacaaggctgttagagagat

aattggaattaaattgactgtaaacacaaagatattagtaaaaaacgtgacgtagaaagtaataatttcttgggtagtttgcagtttataat
 tatgttttaaaatggactatcatatgcttaccgtaacttgaagatatttgcatttcttggccttataatcttgggaaaggacgAAACACCG
 TATTTAATAGGCAGCAGTGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGT
 CCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGTCTTTTTTgttttagagctagaaatagcaagctc
 gacgagctcctgaattctgcagatatccatcacactggcggttaagctagcactagtaacggccgagtgctgtgaattcgcccttCCAC
 ACTGCTGCCTATTAAATACtccccagcatgctgtattgttctcccaatcctcccccttgcgtgctgccccaccccccccc
 agaatagaatgacacctactcagacaatgcgatgcaatttctcattttattaggaaaggacagtgaggagtggcaccttcagggtcaaggaagg
 cacgggggaggggcaaacaacagatggctggcaactagaaggcacagtcgaggcagatctactagaatcgataagcttgattcgagctacat
 ccaagggccccacccccagtgcccttgggatcacagcgggggtcctgtgcagggaagtacacggggactgagtgttatggtagaactgta
 ggcgggacaggagctttgtgacgatgtgaggatattctctggacaggtcatgtcttctcagggtcccgatcaatatcaaagagccagagggtct
 tgggtgggtgggtctgatgagggtatctcagaacattgtattgagacgggtggagggaaccagtaaccacagcctgggtagcccgtgaggagt
 ccaatttccatgtctaattgcagcatggacagatgtgttaaaggctgaatattctggaagagaagagtcaccttctgctggagccatgctgttctg
 gacacgggtgaagagtcacgaagttcgggtcaatattatgcagcagctcaattctgggggatgggtctccttactgatggtttccacacgtcga
 agccatccagaggctttgtgccattgggtgtgtccctggccagcttcacagagtggtggcagccagtcagagatgtggatgagctcccggttctca
 cgcccttctgcttcagcaaggggcttggcacaagcccacccctcggacgcctccttccacaggtccatttcttctcgaaggggcccagttat
 tacccttgcacaaggtctgcccctcgttatctgtagaaaagatgaacaccggtgtgttccagagcccactgcttttaaagctgcagtgcatttct
 actgcttcatccataaggagacaccattctgcatagtgatgcctgttcttcttggataaagtcatatggcttcaagtattcctcagggacctgaag
 gggctcatgcacagactggagagcaaggtagagaaacagaggcttctctggtggatggtagttatgagggtatagcccttttggatgaatgtt
 tgttgaatacatattttatatcctgttgaacttcttgcacatctcgaaaatcaagagcacatcgtgtgacattcagagcgtcaattaatgtacagcgt
 catgggaataataatcttactaccaggagatatccaaagtaggtatcaaatcctcgcggggttgaaggcattcttccgggtacattcccaggt
 gccatttccgaccatattgggtagtataacctgcttctttaggagctggggcaggagtttccatccagagggaacacagctgggctgacagggcc
 agattatttgggtgctgtaaacctgtacggatctggtagcggccagtgagcagctgggtccgcgacggcgtgcacagcgggtgcgtgtagtagtt
 gtccaggagcaccccgccggccgagcgcgtccaggtgcggcgtgcggatgcgggagccgtggaagccgacgtcgttccagcctaggtc
 gtctgccagcaagaagaccaggtggggcgccgggtggccccggcgcccgagcccgccggcgccaacaacagcagcagcagcagcgg
 gaggacgacggggaggagcagccgccgaggtccggggcctcggggcaagctcgccgcgcccgcgggacccataggtccaggattctcct
 cgacgtcaccgcatgttagcagacttctctgccctctccgcttccGGCTAAGGCGTCTTTGCATCTAGTGACAAGG
 TTTGGACCctgttgagagaaaggcaaagtggatgtcagtaagaccaataggtgcctatcCCACACTGCTGCCTATTAA
 ATACAAGGGCgaattctgcagatatccatcacactggcggtcTCGAGttaaggcggaattcccgataaggatcttcttagagcatg
 gctacgtagataagtagcatggcgggttaatcattaactacaAGGAACCCCTAGTGATGGAGTTGGCCACTCCCT
 CTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCACGCCCCGGG
 CTTTGCCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAG [SEQ ID N.34]

Construct p1480_pTIGEM_mAlb3'HITIdonor(SAS_albex14_T2A_ARSB_bGHpA)+
U6scrRNA mAlb3'

CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGACCTT
 TGGTCGCCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATC
 ACTAGGGGTTCTTgtagttaatgattaacccgccatgctacttatctacgtagccatgctctaggaagatcggaattcgcccttaaact
 agtaacggccgccagtgtgctggaattcaggctgacctcgagttcccatgattccttcataatttgcataacgatacaaggcgttagagagat
 aattggaaattaatttgactgtaaacacaaagatattagtaaaaaacgtgacgtagaagaataaatttctgggtagtttgcagttttaaatt
 tatgttttaaaatggactatcatatgcttaccgtaacttgaaagtatttcgatttcttggctttatatacttgggaaggacgAAACACCg
 GACTCGCGCGAGTCGAGGAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAG
 TCCGTTATCAACTTGAAAAAGTGGCACCCAGTCTCGGTGCTTTTTTgttttagagctagaaatagcaagct
 cgagcacctgaattctgcagatatccatcacactggcggttaagctagcactagtaacggccgccagtgtgctggaattcgcccttCCACA
 CTGCTGCCTATTAAATACtcccagcatgectgtattgtcttccaatctcccccttgtgtctgccccacccccccccag
 aatagaatgacacctactcagacaatgcgatgcaatttctcattttattaggaaaggacagtgggagtggcaccttcagggtcaaggaaggca
 cggggggaggggcaaacaacagatggctggcaactagaaggcacagtgcaggcgagatctactagaatcgataagcttgattcgagctacatcc
 aagggccccacacccagtgcccttgggatcacagcgggggtcctgtgcagggaagtacacggggactgagtgtttatggtagaactgtagg
 cgggacaggagctttgtacgatgtgaggatattctctggacaggtcatgtcttcttcagggtcccgatcaatatcaaagagccagagggtcttg
 gttggtgggtctgatgagggtatctcagaaacattgtattgagacgggtggagggaaccagtaaccacagcctgggtagcccgtagggagttcc
 aatttccatgtctaattgcagcatggacagatgtgttaaaggctgaatattctggaagagaagagtcaccttctgctggagccatgctgttctgga
 cacggtgaagagtcacgaagttcgggtcaatattatgcagcagctcaattctgggggatgggcttccttactgatggtttccacacgtcgaag
 ccatccagaggctttgtgccattggtgtgtccctggccagcttcacagtggtggcagccagtcagagatgtggatgagctcccggttcttcag
 ccttctgcttcagcaaggggttgcacaaagcccacccctcgacgcctcctccacaggtccattttcttctcgaagggggcagttatta
 cccctgcccaggctgcccctcggttatctgtagaaaagatgaacaccgtgtgttccagagcccactgcttttaagctgcagtgcatttctac
 tgcttcatccataagggacaccattctgcatagtgatgcctgttcttcttggataaagtcataatggcttcaagtattcctcaggacctgaagg
 gctcatgcacagactggagagcaaggtagagaaacagaggcttctctggtggatggttagttatgagggtatagcccttttggtgaatatgtttgt
 tgaatacatattttatctctgttgaacttcttcccatctcgaataaagagcacatcgtgtgacattcagagcgtcaattaatgtacagcgttca
 tgggaataataatcttactaccaggagatatcaaagtaggtatcaaatcctcggggggttgaaggcatttctccggtacattcccagggtgc
 catttccgaccatatgggtagtataacctgcttcttttaggagctggggcaggagttttcatccagaggaaacacagctgggctgacaggggccag
 attatttgggtgctgtaaacctgtacggatctggtagcggccagtgagcagctggctccgcgacggcgtgcacagcggctgcgtgtagtagtgc
 caggagcaccgcccggccgagcgcgtccaggtgcggcgtgcggatgcgggagccgtggaagccgacgtcgttccagcctaggtcgtc
 tgccagcaagaagaccaggtggggcgccggctggccccggcgcccgagcccgcgccgccaacaacagcagcagcagcagcgggag
 gacgacggggaggagcagccgcccagggtccggggcctcggggcaagctcggcgccgcgcgggaccataggtccaggattctcctcga
 cgtcacccgatgttagcagacttctctgcccctcctcctccGGCTAAGGCGTCTTTGCATCTAGTGACAAGGTT
 TGGACCctgtggagagaaaggcaaaagtggatgtcagtaagaccaataggtgcctatcCCACACTGCTGCCTATTAAAT
 ACAAGGGCgaattctgcagatatccatcacactggcgccTCGAGttaaggcggaattcccagataaggtatctcctagagcatggct
 acgtagataagtagcatggcggttaatacctaactacaAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCT

CTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCCGACGCCCGGGCT
TTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAG [SEQ ID N.35]

Sequences of the above Example 3

5'-ITR

***CTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGCCCGGGCGTCGGGCGACCTT
TGGTCGCCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATC
ACTAGGGGTTCT*** [SEQ ID N.110]

Inverted gRNA sequence for murine Albumin intron 13 without PAM

CACTGCTGCCTATTAAATAC [SEQ ID N.1]

Inverted gRNA sequence for murine Albumin intron 13 + PAM sequence (underlined)

CCACACTGCTGCCTATTAAATAC [SEQ ID N.20]

Splice acceptor sequence

GATAGGCACCTATTGGTCTTACTGACATCCACTTTGCCTTTCTCTCCACAG [SEQ ID N.21]

Exon 14 murine Albumin

GGTCCAAACCTTGTCCTAGATGCAAAGACGCCTTAGCC [SEQ ID N.22]

Thosea asigna virus 2A (T2A) skipping peptide

**GGAAGCGGAGAGGGCAGAGGAAGTCTGCTAACATGCGGTGACGTCGAGGAGAAT
CCTGGACCT** [SEQ ID N.23]

F8_CodopV3 coding sequence

ATGCAGATTGAGCTGAGCACCTGCTTCTTCCTGTGCCTGCTGAGGTTCTGCTTCTCTGCC
ACCAGGAGATACTACCTGGGGGCTGTGGAGCTGAGCTGGGACTACATGCAGTCTGACC
TGGGGGAGCTGCCTGTGGATGCCAGGTTCCCCCCCAGAGTGCCCAAGAGCTTCCCCTTC
AACACCTCTGTGGTGTACAAGAAGACCCTGTTTGTGGAGTTCACTGACCACCTGTTCAA
CATTGCCAAGCCCAGGCCCCCCTGGATGGGCCTGCTGGGCCCCACCATCCAGGCTGAG
GTGTATGACACTGTGGTGATCACCTGAAGAACATGGCCAGCCACCCTGTGAGCCTGC
ATGCTGTGGGGGTGAGCTACTGGAAGGCCTCTGAGGGGGCTGAGTATGATGACCAGAC
CAGCCAGAGGGAGAAGGAGGATGACAAGGTGTTCCCTGGGGGCAGCCACACCTATGT
GTGGCAGGTGCTGAAGGAGAATGGCCCCATGGCCTCTGACCCCTGTGCCTGACCTAC
AGCTACCTGAGCCATGTGGACCTGGTGAAGGACCTGAACTCTGGCCTGATTGGGGCCC

TGCTGGTGTGCAGGGAGGGCAGCCTGGCCAAGGAGAAGACCCAGACCCTGCACAAGT
TCATCCTGCTGTTTGCTGTGTTTGATGAGGGCAAGAGCTGGCACTCTGAAACCAAGAAC
AGCCTGATGCAGGACAGGGATGCTGCCTCTGCCAGGGCCTGGCCCAAGATGCACACTG
TGAATGGCTATGTGAACAGGAGCCTGCCTGGCCTGATTGGCTGCCACAGGAAGTCTGT
GTACTGGCATGTGATTGGCATGGGCACCAACCCTGAGGTGCACAGCATCTTCCTGGAG
GGCCACACCTTCCTGGTCAGGAACCACAGGCAGGCCAGCCTGGAGATCAGCCCCATCA
CCTTCCTGACTGCCCAGACCCTGCTGATGGACCTGGGCCAGTTCCTGCTGTTCTGCCAC
ATCAGCAGCCACCAGCATGATGGCATGGAGGCCTATGTGAAGGTGGACAGCTGCCCTG
AGGAGCCCCAGCTGAGGATGAAGAACAATGAGGAGGCTGAGGACTATGATGATGACC
TGACTGACTCTGAGATGGATGTGGTGAGGTTTGATGATGACAACAGCCCCAGCTTCAT
CCAGATCAGGTCTGTGGCCAAGAAGCACCCCAAGACCTGGGTGCACTACATTGCTGCT
GAGGAGGAGGACTGGGACTATGCCCCCTGGTGCTGGCCCCTGATGACAGGAGCTACA
AGAGCCAGTACCTGAACAATGGCCCCCAGAGGATTGGCAGGAAGTACAAGAAGGTCA
GGTTCATGGCCTACACTGATGAAACCTTCAAGACCAGGGAGGCCATCCAGCATGAGTC
TGGCATCCTGGGCCCCCTGCTGTATGGGGAGGTGGGGGACACCCTGCTGATCATCTTCA
AGAACCAGGCCAGCAGGCCCTACAACATCTACCCCCATGGCATCACTGATGTGAGGCC
CCTGTACAGCAGGAGGCTGCCCAAGGGGGTGAAGCACCTGAAGGACTTCCCCATCCTG
CCTGGGGAGATCTTCAAGTACAAGTGGACTGTGACTGTGGAGGATGGCCCCACCAAGT
CTGACCCCAGGTGCCTGACCAGATACTACAGCAGCTTTGTGAACATGGAGAGGGACCT
GGCCTCTGGCCTGATTGGCCCCCTGCTGATCTGCTACAAGGAGTCTGTGGACCAGAGG
GGCAACCAGATCATGTCTGACAAGAGGAATGTGATCCTGTTCTCTGTGTTTGATGAGA
ACAGGAGCTGGTACCTGACTGAGAACATCCAGAGGTTCCCTGCCCAACCCTGCTGGGGT
GCAGCTGGAGGACCCTGAGTTCCAGGCCAGCAACATCATGCACAGCATCAATGGCTAT
GTGTTTGACAGCCTGCAGCTGTCTGTGTGCCTGCATGAGGTGGCCTACTGGTACATCCT
GAGCATTGGGGCCCAGACTGACTTCCTGTCTGTGTTCTTCTCTGGCTACACCTTCAAGC
ACAAGATGGTGTATGAGGACACCCTGACCCTGTTCCCTTCTCTGGGGAGACTGTGTTT
ATGAGCATGGAGAACCCTGGCCTGTGGATTCTGGGCTGCCACAACCTCTGACTTCAGGA
ACAGGGGCATGACTGCCCTGCTGAAAGTCTCCAGCTGTGACAAGAACAACACTGGGGACTA
CTATGAGGACAGCTATGAGGACATCTCTGCCTACCTGCTGAGCAAGAACAATGCCATT
GAGCCCAGGAGCTTCAGCCAGAATGCCACTAATGTGTCTAACAACAGCAACACCAGCA
ATGACAGCAATGTGTCTCCCCCAGTGCTGAAGAGGCACCAGAGGGAGATCACCAGGAC
CACCTGCAGTCTGACCAGGAGGAGATTGACTATGATGACACCATCTCTGTGGAGATG
AAGAAGGAGGACTTTGACATCTACGACGAGGACGAGAACCAGAGCCCCAGGAGCTTC
CAGAAGAAGACCAGGCACTACTTCATTGCTGCTGTGGAGAGGCTGTGGGACTATGGCA

TGAGCAGCAGCCCCCATGTGCTGAGGAACAGGGCCCAGTCTGGCTCTGTGCCCCAGTT
CAAGAAGGTGGTGTTCAGGAGTTCACTGATGGCAGCTTCACCCAGCCCCTGTACAGA
GGGGAGCTGAATGAGCACCTGGGCCTGCTGGGCCCCTACATCAGGGCTGAGGTGGAGG
ACAACATCATGGTGACCTTCAGGAACCAGGCCAGCAGGCCCTACAGCTTCTACAGCAG
CCTGATCAGCTATGAGGAGGACCAGAGGCAGGGGGCTGAGCCCAGGAAGAACTTTGT
GAAGCCCAATGAAACCAAGACCTACTTCTGGAAGGTGCAGCACCATGGCCCCCACC
AAGGATGAGTTTGACTGCAAGGCCTGGGCCTACTTCTCTGATGTGGACCTGGAGAAGG
ATGTGCACTCTGGCCTGATTGGCCCCCTGCTGGTGTGCCACACCAACACCCTGAACCT
GCCCATGGCAGGCAGGTGACTGTGCAGGAGTTTGCCCTGTTCTTCACCATCTTTGATGA
AACCAAGAGCTGGTACTTCACTGAGAACATGGAGAGGAACTGCAGGGCCCCCTGCAAC
ATCCAGATGGAGGACCCACCTTCAAGGAGAACTACAGGTTCCATGCCATCAATGGCT
ACATCATGGACACCCTGCCTGGCCTGGTGTATGGCCCAGGACCAGAGGATCAGGTGGTA
CCTGCTGAGCATGGGCAGCAATGAGAACATCCACAGCATCCACTTCTCTGGCCATGTG
TTCATGTGAGGAAGAAGGAGGAGTACAAGATGGCCCTGTACAACCTGTACCCTGGGG
TGTTTGAGACTGTGGAGATGCTGCCCAGCAAGGCTGGCATCTGGAGGGTGGAGTGCCT
GATTGGGGAGCACCTGCATGCTGGCATGAGCACCTGTTCTTGGTGTACAGCAACAAG
TGCCAGACCCCCCTGGGCATGGCCTCTGGCCACATCAGGGACTTCCAGATCACTGCCTC
TGGCCAGTATGGCCAGTGGGCCCCCAAGCTGGCCAGGCTGCACTACTCTGGCAGCATC
AATGCCTGGAGCACCAAGGAGCCCTTCAGCTGGATCAAGGTGGACCTGCTGGCCCCCA
TGATCATCCATGGCATCAAGACCCAGGGGGCCAGGCAGAAGTTCAGCAGCCTGTACAT
CAGCCAGTTCATCATCATGTACAGCCTGGATGGCAAGAAGTGGCAGACCTACAGGGGC
AACAGCACTGGCACCCCTGATGGTGTTCCTTTGGCAATGTGGACAGCTCTGGCATCAAGC
ACAACATCTTCAACCCCCCATCATTGCCAGATACATCAGGCTGCACCCACCCACTAC
AGCATCAGGAGCACCTGAGGATGGAGCTGATGGGCTGTGACCTGAACAGCTGCAGCA
TGCCCCTGGGCATGGAGAGCAAGGCCATCTCTGATGCCAGATCACTGCCAGCAGCTA
CTTACCAACATGTTTGCCACCTGGAGCCCCAGCAAGGCCAGGCTGCACCTGCAGGGC
AGGAGCAATGCCTGGAGGGCCCCAGGTCAACAACCCCAAGGAGTGGCTGCAGGTGGAC
TTCCAGAAGACCATGAAGGTGACTGGGGTGACCACCCAGGGGGTGAAGAGCCTGCTG
ACCAGCATGTATGTGAAGGAGTTCCTGATCAGCAGCAGCCAGGATGGCCACCAGTGGA
CATGCAGATTGAGCTGAGCACCTGCTTCTTCTTCTGTGCCTGCTGACCTGTTCTTCCAGAA
TGGCAAGGTGAAGGTGTTCCAGGGCAACCAGGACAGCTTCACCCCTGTGGTGAACAGC
CTGGACCCCCCCTGCTGACCAGATACCTGAGGATTCACCCCAAGAGCTGGGTGCACC
AGATTGCCCTGAGGATGGAGGTGCTGGGCTGTGAGGCCAGGACCTGTACTGA [SEQ ID
N.36]

Synthetic polyadenylation Signal

tcgcgaataaaagatctttattttcattagatctgtgtgttggtttttgtgtgatgcagc [SEQ ID N.37]

3' ITR

AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAG
GCCGGGCGACCAAAGGTCGCCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCG
AGCGAGCGCGCAG [SEQ ID N.29]

Construct p1493_pTIGEM_mAlb3'HITIdonor(SAS_albex14_T2A_CodopV3_pA)

***CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGACCTT
TGGTCGCCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATC
ACTAGGGGTTCTGCTAGGCTAGCGGCGCGCCTCTAGCCACACTGCTGCCTATTAAATA
CGATAGGCACCTATTGGTCTTACTGACATCCACTTTGCCTTTCTCTCCACAGGGTC
CAAACCTTGTCACTAGATGCAAAGACGCCTTAGCCGGAAGCGGAGAGGGCAGAGGA
AGTCTGCTAACATGCGGTGACGTCGAGGAGAATCCTGGACCTATGCAGATTGAGCT
GAGCACCTGCTTCTTCCTGTGCCTGCTGAGGTTCTGCTTCTCTGCCACCAGGAGATACT
ACCTGGGGGCTGTGGAGCTGAGCTGGGACTACATGCAGTCTGACCTGGGGGAGCTGCC
TGTGGATGCCAGGTTCCCCCCCAGAGTGCCCAAGAGCTTCCCCTTCAACACCTCTGTGG
TGTACAAGAAGACCCTGTTTGTGGAGTTCACTGACCACCTGTTCAACATTGCCAAGCCC
AGGCCCCCCTGGATGGGCCTGCTGGGCCCCACCATCCAGGCTGAGGTGTATGACACTG
TGGTGATCACCTGAAGAACATGGCCAGCCACCCTGTGAGCCTGCATGCTGTGGGGGT
GAGCTACTGGAAGGCCTCTGAGGGGGCTGAGTATGATGACCAGACCAGCCAGAGGGA
GAAGGAGGATGACAAGGTGTTCCCTGGGGGCAGCCACACCTATGTGTGGCAGGTGCTG
AAGGAGAATGGCCCCATGGCCTCTGACCCCCTGTGCCTGACCTACAGCTACCTGAGCC
ATGTGGACCTGGTGAAGGACCTGAACTCTGGCCTGATTGGGGCCCTGCTGGTGTGCAG
GGAGGGCAGCCTGGCCAAGGAGAAGACCCAGACCCTGCACAAGTTCATCCTGCTGTTT
GCTGTGTTTGATGAGGGCAAGAGCTGGCACTCTGAAACCAAGAACAGCCTGATGCAGG
ACAGGGATGCTGCCTCTGCCAGGGCCTGGCCCAAGATGCACACTGTGAATGGCTATGT
GAACAGGAGCCTGCCTGGCCTGATTGGCTGCCACAGGAAGTCTGTGTACTGGCATGTG
ATTGGCATGGGCACCACCCCTGAGGTGCACAGCATCTTCCTGGAGGGGCCACACCTTCCT
GGTCAGGAACCACAGGCAGGCCAGCCTGGAGATCAGCCCCATCACCTTCCTGACTGCC
CAGACCCTGCTGATGGACCTGGGCCAGTTCCTGCTGTTCTGCCACATCAGCAGCCACCA
GCATGATGGCATGGAGGCCTATGTGAAGGTGGACAGCTGCCCTGAGGAGCCCCAGCTG
AGGATGAAGAACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAG
ATGGATGTGGTGAGGTTTGATGATGACAACAGCCCCAGCTTCATCCAGATCAGGTCTG***

TGGCCAAGAAGCACCCCAAGACCTGGGTGCACTACATTGCTGCTGAGGAGGAGGACTG
GGACTATGCCCCCTGGTGCTGGCCCCCTGATGACAGGAGCTACAAGAGCCAGTACCTG
AACAATGGCCCCCAGAGGATTGGCAGGAAGTACAAGAAGGTCAGGTTCATGGCCTAC
ACTGATGAAACCTTCAAGACCAGGGAGGCCATCCAGCATGAGTCTGGCATCCTGGGGCC
CCCTGCTGTATGGGGAGGTGGGGGACACCCTGCTGATCATCTTCAAGAACCAGGCCAG
CAGGCCCTACAACATCTACCCCCATGGCATCACTGATGTGAGGCCCCTGTACAGCAGG
AGGCTGCCCCAAGGGGGTGAAGCACCTGAAGGACTTCCCCATCCTGCCTGGGGAGATCT
TCAAGTACAAGTGGACTGTGACTGTGGAGGATGGCCCCACCAAGTCTGACCCCAGGTG
CCTGACCAGATACTACAGCAGCTTTGTGAACATGGAGAGGGACCTGGCCTCTGGCCTG
ATTGGCCCCCTGCTGATCTGCTACAAGGAGTCTGTGGACCAGAGGGGCAACCAGATCA
TGTCTGACAAGAGGAATGTGATCCTGTTCTCTGTGTTTGATGAGAACAGGAGCTGGTAC
CTGACTGAGAACATCCAGAGGTTCCCTGCCCCAACCTGCTGGGGTGCAGCTGGAGGACC
CTGAGTTCCAGGCCAGCAACATCATGCACAGCATCAATGGCTATGTGTTTGACAGCCT
GCAGCTGTCTGTGTGCCTGCATGAGGTGGCCTACTGGTACATCCTGAGCATTGGGGCCC
AGACTGACTTCCTGTCTGTGTTCTTCTCTGGCTACACCTTCAAGCACAAAGATGGTGTAT
GAGGACACCCTGACCCTGTTCCCCTTCTCTGGGGAGACTGTGTTTCATGAGCATGGAGA
ACCCTGGCCTGTGGATTCTGGGCTGCCACAACCTCTGACTTCAGGAACAGGGGGCATGAC
TGCCCTGCTGAAAGTCTCCAGCTGTGACAAGAACACTGGGGACTACTATGAGGACAGC
TATGAGGACATCTCTGCCTACCTGCTGAGCAAGAACAATGCCATTGAGCCCAGGAGCT
TCAGCCAGAATGCCACTAATGTGTCTAACAACAGCAACACCAGCAATGACAGCAATGT
GTCTCCCCCAGTGCTGAAGAGGCACCAGAGGGAGATCACCAGGACCACCCTGCAGTCT
GACCAGGAGGAGATTGACTATGATGACACCATCTCTGTGGAGATGAAGAAGGAGGAC
TTTGACATCTACGACGAGGACGAGAACCAGAGCCCCAGGAGCTTCCAGAAGAAGACC
AGGCACTACTTCATTGCTGCTGTGGAGAGGCTGTGGGACTATGGCATGAGCAGCAGCC
CCCATGTGCTGAGGAACAGGGCCCAGTCTGGCTCTGTGCCCCAGTTCAAGAAGGTGGT
GTTCCAGGAGTTCACTGATGGCAGCTTACCCAGCCCCTGTACAGAGGGGAGCTGAAT
GAGCACCTGGGCCTGCTGGGGCCCTACATCAGGGCTGAGGTGGAGGACAACATCATGG
TGACCTTCAGGAACCAGGCCAGCAGGCCCTACAGCTTCTACAGCAGCCTGATCAGCTA
TGAGGAGGACCAGAGGCAGGGGGCTGAGCCCAGGAAGAACTTTGTGAAGCCCAATGA
AACCAAGACCTACTTCTGGAAGGTGCAGCACCATGGCCCCCACCAGGATGAGTTT
GACTGCAAGGCCTGGGCCTACTTCTCTGATGTGGACCTGGAGAAGGATGTGCACTCTG
GCCTGATTGGCCCCCTGCTGGTGTGCCACACCAACACCCTGAACCCTGCCCATGGCAG
GCAGGTGACTGTGCAGGAGTTTGCCTGTTCTTACCATCTTTGATGAAACCAAGAGCT
GGTACTTCACTGAGAACATGGAGAGGAACTGCAGGGCCCCCTGCAACATCCAGATGGA

GGACCCACCTTCAAGGAGAACTACAGGTTCCATGCCATCAATGGCTACATCATGGAC
 ACCCTGCCTGGCCTGGTGATGGCCCAGGACCAGAGGATCAGGTGGTACCTGCTGAGCA
 TGGGCAGCAATGAGAACATCCACAGCATCCACTTCTCTGGCCATGTGTTCACTGTGAG
 GAAGAAGGAGGAGTACAAGATGGCCCTGTACAACCTGTACCCTGGGGTGTGTTGAGACT
 GTGGAGATGCTGCCCAGCAAGGCTGGCATCTGGAGGGTGGAGTGCCTGATTGGGGAGC
 ACCTGCATGCTGGCATGAGCACCTGTTCCTGGTGTACAGCAACAAGTGCCAGACCCC
 CCTGGGCATGGCCTCTGGCCACATCAGGGACTTCCAGATCACTGCCTCTGGCCAGTATG
 GCCAGTGGGCCCCCAAGCTGGCCAGGCTGCACTACTCTGGCAGCATCAATGCCTGGAG
 CACCAAGGAGCCCTTCAGCTGGATCAAGGTGGACCTGCTGGCCCCCATGATCATCCAT
 GGCATCAAGACCCAGGGGGCCAGGCAGAAGTTCAGCAGCCTGTACATCAGCCAGTTCA
 TCATCATGTACAGCCTGGATGGCAAGAAGTGGCAGACCTACAGGGGCAACAGCACTGG
 CACCCTGATGGTGTCTTTGGCAATGTGGACAGCTCTGGCATCAAGCACAAACATCTTCA
 ACCCCCCCATCATTGCCAGATACATCAGGCTGCACCCCAACCACTACAGCATCAGGAG
 CACCCTGAGGATGGAGCTGATGGGCTGTGACCTGAACAGCTGCAGCATGCCCCTGGGC
 ATGGAGAGCAAGGCCATCTCTGATGCCCAGATCACTGCCAGCAGCTACTTCACCAACA
 TGTTTGCCACCTGGAGCCCCAGCAAGGCCAGGCTGCACCTGCAGGGCAGGAGCAATGC
 CTGGAGGCCCCAGGTCAACAACCCCAAGGAGTGGCTGCAGGTGGACTTCCAGAAGACC
 ATGAAGGTGACTGGGGTGACCACCCAGGGGGTGAAGAGCCTGCTGACCAGCATGTATG
 TGAAGGAGTTCCTGATCAGCAGCAGCCAGGATGGCCACCAGTGGACATGCAGATTGAG
 CTGAGCACCTGCTTCTTCCTGTGCCTGCTGACCTGTTCTTCCAGAATGGCAAGGTGAAG
 GTGTTCCAGGGCAACCAGGACAGCTTCACCCCTGTGGTGAACAGCCTGGACCCCCCCC
 TGCTGACCAGATACCTGAGGATTACACCCCAAGAGCTGGGTGCACCAGATTGCCCTGAG
 GATGGAGGTGCTGGGCTGTGAGGCCAGGACCTGTACTGAtcgcggaataaaagatctttatttcattagat
 ctgtgtgttggtttttgtgtgatgcgcCCACACTGCTGCCTATTAAATACCCCCAATTGAGGAACCCCT
 AGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGAC
 CAAAGGTCGCCCCGACGCCCGGGCTTTGCCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCG
 CAG [SEQ ID N.38]

Construct_p1139_pAAV2.1._HLP_SpCas9(HA)_spA

Promoter sequence is underlined

Cas9/Cas9-2a-GFP

ataacaatttcacacaggaaacagctatgaccatgattacgccagatttaattaaggctgcgcgctcgctcgctcactgaggccgcccgggcaa
 agccccgggctcgggcgacctttggtcgccccggcctcagtgagcgagcgagcgcgagagaggagtgggccaactccatcactaggggtt
 cctttagttaatgattaacccgcatgctacttatctacgtagccatgctctaggaagatcggaattcgcccGGAATTCGCCCTTAA

gcggccgcaagcCTTAAGTGTTTGCTGCTTGCAATGTTTGCCCATTTTAGGGTGGACACAGGA
CGCTGTGGTTTCTGAGCCAGGGGGCGACTCAGATCCCAGCCAGTGGACTTAGCCCCTG
TTTGCTCCTCCGATAACTGGGGTGACCTTGGTTAATATTCACCAGCAGCCTCCCCGTT
GCCCCCTCTGGATCCACTGCTTAAATACGGACGAGGACAGGGCCCTGTCTCCTCAGCTTC
AGGCACCACCCTGACCTGGGACAGTGAATCACCGGTacCTGCTTTTGCTCGCTTGGAT
CCCCGGTGCCACCATGTccggtgccaccatgtaccatacagatgtccagattacgcttcgccgaagaaaaagcgcaagggtcga
agcgtccgacaagaagtacagcatcggcctggacatcggcaccaactctgtgggctgggcccgtgatcaccgacgagtacaagggtcccagc
aagaaattcaagggtcgtgggcaacaccgaccggcacagcatcaagaagaacctgatcggagccctgctgttcgacagcggcgaaacagccg
aggccacccggctgaagagaaccgccagaagaagatacaccagacggaagaaccggatctgctatctgcaagagatcttcagcaacgagat
ggccaagggtggacgacagcttctccacagactggaagagtccttctgtggaagaggataagaagcacgagcggcaccccatcttcggca
acatcgtggacgaggtggcctaccacgagaagtaccccacatctaccacctgagaagaaactgggtggacagcaccgacaaggccgacct
gcggctgatctatctggccctggccacatgatcaagttccggggccacttctgatcgaaggcgacctgaaccccgacaacagcgacgtgga
caagctgttcacagctggtgcagacctacaaccagctgttcgaggaaaaccccatcaacgccagcggcgtggacgccaaggccatcctgtc
tgcagactgagcaagagcagacggctggaatatctgatcggcagctgcccggcgagaagaagaatggcctgttcggcaacctgattgcc
tgagcctgggctgaccccaactcaagagcaacttcgacctggccgaggatgccaaactgcagctgagcaaggacacctacgacgacga
cctggacaacctgctggccagatcggcgaccagtacggcgacctgtttctggccgccaagaacctgtccgacgccatcctgctgagcgacat
cctgagagtgaacaccgagatcaccaaggccccctgagcgcctctatgatcaagagatacagcagcaccaccaggacctgacctgctga
aagctctcgtcggcgacgagctgcctgagaagtacaaagagattttctcgaccagagcaagaacggctacgccggctacattgacggcgga
gccagccaggaagagttctacaagttcatcaagccatcctggaaaagatggacggcaccgaggaactgctcgtgaagctgaacagagagg
acctgctgcgggaagcagcggaccttcgacaacggcagcatccccaccagatccacctgggagagctgcacgccattctgcggcggcagga
agattttaccattcctgaaggacaaccgggaaaagatcgagaagatcctgacctccgcatcccctactacgtgggcccctctggccagggga
aacagcagattcgctggatgaccagaaagagcagaggaaaccatcacccccctggaacttcgaggaagtgggtggacaaggcgcttcgccc
agagcttcacgagcggatgaccaacttcgataagaacctgccaacgagaaggtgctgccaagcacagcctgctgtacgagtacttcaccg
tgtataacgagctgaccaaagtgaatacgtgaccgaggggaatgagaaagcccgccttctgagcggcgagcagaaaaaggccatcgtgga
cctgctgttcaagaccaaccggaaagtaccgtgaagcagctgaaagaggactactcaagaaaatcgagtgttcgactccgtggaatctcc
ggcgtggaagatcggttcaacgcctccctgggcacataccacgatctgctgaaaattatcaaggacaaggacttctggacaatgaggaaaac
gaggacattctggaagatatcgtgctgacctgacactgtttgaggacagagagatgatcgaggaaacggctgaaaacctatgccacctgttcg
acgacaaaagtgatgaagcagctgaagcggcgagatacaccggctggggcaggctgagccggaagctgatcaacggcatccgggacaag
cagtccggcaagacaatcctggatttctgaagtccgacggcttcgccaacagaaactcatgcagctgatccacgacgacagcctgaccttaa
agaggacatccagaaagcccaggtgtccggccaggcgatagcctgcacgagcacattgccaatctggccggcagccccgccattaagaag
ggcatcctgcagacagtgaagggtggtggacgagctcgtgaaagtgtatggcgccgacacaagccccgagaacatcgtgatgaaatggccagag
agaaccagaccaccagaaggacagaagaacagccgcgagagaatgaagcgatcgaagaggcatcaaagagctggcgagccagat
cctgaaagaacacccctggaaaacacccagctgcagaacgagaagctgtacctgtactacctgcagaatggcggggatgtacgtggacc
aggaactggacatcaaccggctgtccgactacgatgtggaccatactgtccctcagagctttctgaaggacgactccatcgacaacaaggtgct

gaccagaagcgcacaagaaccggggcaagagcgacaacgtgccctccgaagaggtcgtgaagaagatgaagaactactggcggcgagctgct
gaacgccaaagctgattaccagagaaagttcgacaatctgaccaaggccgagagagggcgccctgagcgaaactggataaggccggcttcac
aagagacagctggtggaaacccggcagatcacaagcacgtggcacagatcctggactcccggatgaacactaagtacgacgagaatgaca
agctgatccgggaagtgaagtgatcacctgaagtccaagctggtgtccgatttccggaaggatttccagttttacaaagtgcgcgagatcaac
aactaccaccacgcccacgacgcctacctgaacgccctcgtgggaaccggccctgatcaaaaagtaccctaagctggaaagcgagttcgtgtac
ggcgactacaagggtgtacgacgtgcggaagatgatcgccaagagcgagcaggaaatcggaaggctaccgccaagtacttcttacagcaa
catcatgaactttttcaagaccgagattaccctggccaacggcgagatccggaagcgccctctgatcgagacaaacggcgaaaccggggaga
tcgtgtgggataaaggccgggattttgccaccgtgcggaagtgctgagcatgcccgaagtgaatatcgtgaaaaagaccgaggtgcagaca
ggcggcttcagcaaagagtctatcctgccaaagaggaacagcgataagctgatcgccagaaagaaggactgggaccctaagaaagtacggcg
gcttcgacagccccaccgtggcctattctgtgctggtggtggccaaagtggaaaaggcgcaagtccaagaaactgaagagtgtgaaagagctgc
tggggatcaccatcatggaaagaagcagcttcgagaagaatcccatcgacttttggaaaggcgaagggtacaaagaagtgaaaaaggacctg
atcatcaagctgcctaagtactccctgttcgagctggaaaacggccggaagagaatgctggcctctgccggcgaaactgcagaagggaacga
actggccctgccctccaaatatgtgaacttctgtacctggccagccactatgagaagctgaagggtcccccgaggataatgagcagaaaca
gctgttttggaaacagcacaagcactacctggacgagatcatcgagcagatcagcgagtttccaagagagtatcctggccgacgctaattctg
gacaaagtgtgtccgcctacaacaagcaccgggataagcccatcagagagcaggccgagaatatcatccacctgtttacctgaccaatctg
ggagcccctgccgccttcaagtactttgacaccaccatcgaccggaagaggtacaccagcaccaaagaggtgctggacgccacctgatcca
ccagagcatcaccggcctgtacgagacacggatcgacctgtctcagctgggagggcgacagccccaagaagaagagaaaaggtggaggccag
ctaagaattcaataaaagatctttattttcattagatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt
ctgcgcgctcgtcgtcactgaggccggggcgaccaaaggctgcccgcgcccgggctttgccggggcgccctcagtgcgcgagcgagcg
cgcagctgcctgcagggggcgccctgatgcggtattttctccttacgcacatctgtgcggtatttcacaccgcatacgtcaaagcaaccatagtacgcgc
cctgtagcggcgcatgaagcgcggcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt
cttcccttctttctcgccacgttcgccgggtttccccgtcaagctctaatacgggggtcccttaggggtccgatttagtgccttacggcacctga
ccccaaaaaacttgattgggtgatggttcacgtatgtggccatcgccctgatagacggtttttgccctttgacgttgaggtccacgttcttaata
gtggactcttgttcaaaactggaacaacactcaaccctatctcgggctattctttgatttataagggattttgccgatttcggcctatttggttaaaaaat
gagctgatttaacaaaaatttaacgcgaatttaacaaaatattaacgtttacaatttatggtgcactctcagtacaatctgctctgatgccgatagtt
aagccagccccgacaccgcgaacaccgcgtgacgcgcctgacgggctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt
ccgggagctgcagtgtcagaggtttaccgtcatccgaaacgcgcgagacgaaagggcctcgtgatacgcctattttataggttaagtca
tgataataatggtttcttagacgtcaggtggcacttttcgggaaatgtgcgcggaaccctattttgttttttctaaatacattcaaatatgtatccg
ctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgccttattccctttttgcgg
cattttgccttctgtttttgctcaccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtggttacatgaactgga
tctcaacagcggtaagatccttgagagtttcgccccgaagaacgtttccaatgatgagcacttttaaagtctgctatgtggcgcggtattatccc
gtattgacggcggaagagcaactcggtcggcgatacactattctcagaatgacttggtgagtactaccagtcacagaaaagcatcttacg
gatggcatgacagtaagagaattatgcagtgtgccataaccatgagtataactgcggccaacttacttctgacaacgatcggaggaccga
aggagctaaccgctttttgcacaacatgggggatcatgtaactgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacga

gcgtgacaccacgatgcctgtagcaatggcaacaacggtgcgcaaaactattaaactggcgaaactacttactctagcttcccggcaacaattaatag
actggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctgggttattgctgataaatctggagccggtgagcgt
ggaagccgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggat
gaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaagttactcatatatactttagattgatttaa
acttcatttttaattttaaaaggatctaggtgaagatcctttttgataatctcatgacaaaaatcccttaacgtgagtttctgtccactgagcgtcagacc
ccgtagaaaagatcaaaggatcttctgagatccttttttctgcgcgtaatctgctgcttgcaacaaaaaaaccaccgctaccagcgggtggttgt
ttgccggatcaagagctaccaactcttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttag
gccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaactctgttaccagtggctgctgccagtggcgataagtcgtgtcttac
cgggttggaactcaagacgatagttaccggataaggcgacgcagggtcgggctgaacgggggggttcgtgcacacagcccagcttgagcgaacg
acctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaaggagaaaggcggacaggtatccggtaaagcg
gcagggtcggaaacaggagagcgcacgaggagcgttcagggggaaacgcctggatatcttatagtctgtcgggttcgccacctctgacttg
agcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggcctttttacggttccttggccttttctggccttt
tgctcacatgttcttctgctgttatcccctgattctgtggataaccgtattaccgcttttgagtgagctgataccgctcggcgagccgaacgaccg
agcgcagcagtgagtgagcgaaggagcgggaagagcgcccaatacgcgaaccgcctctccccgcgcgttggccgattcattaatgcagctg
gcacgcagaggttcccactggaaagcgggcagtgagcgcaacgcgaattaatgtgagttagctcactcattaggcaccacaggccttacacttta
tgcttccggctcgatgttgtgtggaattgtgagcgg [SEQ ID N.43]

HITI 3'mALb f8 (Haemophilia A)

p1498_pAAV_HLP_SpCas9 + U6 3'malb_gRNA (5,1kb)

5'ITR

CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCTGGGCGACCTTTGGTCGCCCCGG
CCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCCT (SEQ ID
NO:110)

Additional AAV sequences

Tgtagttaatgattaacccgccatgctacttatctacgtagagctcttgtcgaggtcgacatttaaataagggcgaattcaattt (SEQ ID NO: 44)

U6 expression cassette gRNA

Gagggcctatttccatgattccttcatatttgcataacgatacaaggctgtagagagataattggaattaatttgactgtaaacacaaaga
tattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaattatgttttaaattggactatcatatgcttaccg
taacttgaaagtatttcgatttcttggctttatataatcttgtggaaaggacgaaacaccgtatttaataggcagcagtggttttagagctagaaa
tagcaagttaaataaggctagtcggttatcaacttgaaaaagtggcaccgagtcggtgctttttttagagcta (SEQ ID NO:45)

HLP PROMOTER

TGTTTGCTGCTTGAATGTTTGCCCATTTTAGGGTGGACACAGGACGCTGTGGTTTCTGAGCCAGGGGGCGACTCAGATCCCAGCCAGTGGACTTAGCCCTGTTTGCTCCTCCGATAACTGGGGTGACCTTGGTTAATATTACACAGCAGCCTCCCGGTTGCCCCTCTGGATCCACTGCTTAAATACGGACGAGGACAGGGCCCTGTCTCCTCAGCTTCAGGCACCACTGACCTGGGACAGTGAAT (SEQ ID NO:46)

spCas9

atgtacccatacagatgttccagattacgcttcgccgaagaaaaagcgcaaggtcgaagcgtccgacaagaagtacagcatcggcctggacatcggcac
 caactctgtgggctgggcccgtgatcaccgacgagtacaaggtgcccagcaagaaattcaaggtgctgggcaacaccgaccggcacagcatcaagaag
 aacctgatcggagccctgtgttcgacagcggcgaaacagccgaggccacccggctgaagagaaccgccaagaagaatacaccagacgggaagaac
 cggatctgctatctgcaagagatcttcagcaacgagatggccaaggtggacgacagcttcttccacagactggaagagtccttctggtggaagaggat
 aagaagcacgagcggcaccctatcttggcaacatcgtggacgaggtggcctaccacgagaagtacccacatctaccacctgagaaagaaactggt
 ggacagcacggacaaggccgacctgcggctgatctatctggccctggcccatatgatcaagttccggggccattctctgatcaggggacgacctgaacccc
 gacaacagcgacgtggacaagctgttcatccagctggtgacagctacaaccagctgttcgaggaaaaccccatcaacgccagcggcgtggacgcca
 ggccatctgtctgcccagactgagcaagagcagacggctggaaaatctgatcggccagctgcccggcgagaagaagaatggcctgttcggaaacctga
 ttgccctgagcctgggctgaccccaacttcaagacaaacttcagactggcggaggatgccaaactgcagctgagcaaggacacctacgacgacgacc
 tggacaacctgctggcccagatcggcgaccagtacgccgacctgttctggccgccaagaacctgtccgacgcatctgtgagcgacatctgagagt
 gaacaccgagatcaccaagccccctgagcgctctatgatcaagagatacgcagcagcaccaccaggacctgacctgctgaaagctctctgctgccc
 agcagctgctgagaagtacaaagagattttcttcgaccagagcaagaacggctacccggctacattgacggcgagccagccaggaagagttctac
 aagttcatcaagccatcttgaaaaagatggacggcaccgaggaactgctcgtgaagctgaacagagaggacctgtcgggaagcagcggaccttcg
 acaacggcagcatccccaccagatccacctgggagagctgcacgccattctgcggcgaggaagattttaccattctgaaggacaacccgggaaa
 agatcgagaagatcctgacctccgcatcccactactcgtgggccccttggccaggggaaacagcagattcgctggatgaccagaaagagcgaggaa
 accatcacccctggaaacttcgaggaagtgtggacaaggcgcttcccccagagcttcatcgagcggatgaccaactcgataagaacctgccaaac
 gagaaggtgctgcccacagcctgtgtacgagtacttaccgtgtataacgagctgaccaaagtgaatacgtgacggagggaatgagaagcc
 cgccttctgagcggcgagcagaaaaaggccatcgtggacctgtgttcaagaccaaccggaaaagtaccgtgaagcagctgaaagaggactacttca
 agaaaatcgagtgcttcgactccgtggaaatctccggcgtggaagatcggttcaacgcctccctgggacataccacgactctgtgaaaaattatcaagg
 acaaggacttctggacaatgaggaaaacgaggacattctggaagatatcgtgctgacctgacactgtttgaggacagagagatgatcgaggaaacgg
 ctgaaaacctatgccacctgttcgacgacaagtgatgaagcagctgaacggcgagatacaccggctggggcaggctgagccggaagctgatca
 acggcatccgggacaagcagctccggcaagacaatcctggatttctgaagtcgacggcttcgccaacagaaacttcatgcagctgatccacgacgaca
 gcctgaccttaaaaggagacatccagaaagcccagggtgtccggccagggcgatagcctgcacgagcacattgccaatctggccggcagccccgccatta
 agaagggcatcctgcagacagtgaaaggtgtggacgagctcgtgaaagtgtatgggcccgcacaagcccagaaatcgtgatcgaatggccagaga
 gaaccagaccaccagaaggagcagaagaacagccgcgagagaatgaagcggatcgaagagggcatcaaagagctgggagccagatcctgaaag
 aacaccccgctgaaaaacaccagctgcagaacgagaagctgtacttctacgtcagaatggcggggatgtgactgtgaccaggaactggacatc
 aaccggctgtccgactacgatgtggaccatatcgtgcctcagagctttctgaaggcagactccatcgacaacaaggtgctgaccagaagcgacaagaac
 cggggcaagagcgacaacgtgccctccgaagaggtcgtgaagaagatgaagaactactggcgagctgctgaacgccaagctgattaccagagaa
 agttcgacaatctgaccaaggccgagagaggcggcctgagcgaaatggataaggccggttcatcaagagacagctggtggaacccggcagatcac
 aaagcagctggcacagatcctggactccggatgaacactaagtacgacgagaatgacaagctgatccgggaagtgaagtgatcacctgaagtcc
 aagctggtgtccgatttccggaaggatttccagttttacaaagtgcgcgagatcaacaactaccaccacgcccacgacgcctacctgaacgccgtcgtgg
 gaaccgccctgatcaaaaagtacctaagctggaaagcgagttcgtgtacggcgactacaaggtgtacgacgtgcggaagatgatcgccaagagcga
 gcaggaatcggaaggctaccgcaagtaacttcttacagcaacatcatgaacttttcaagaccgagattacctggccaacggcgagatccggaa
 gcggcctctgatcgagacaaacggcgaaacggggagatcgtgtgggataaggccgggattttgccaccgtcggaagtgctgagcatgccccaaag
 tgaatatcgtgaaaaagaccgaggtgcagacaggcggcttcagaaagagtctatctgcccgaaggaacagcgataagctgatcgccagaaagaa
 ggactgggaccctaagaagtacggcggcttcgacagccccaccgtggcctattctgtgctggtggtggccaaagtggaaaaggcgcaagtccaagaaac
 tgaagagtgtgaaagagctgctggggatcacctcatgtgaaagaagcagcttcgagaagaatcccatcgactttctggaaagccaagggtacaaaga
 agtgaaaaaggacctgatcatcaagctgcctaagtactccctgttcgagctggaaaacggccggaagagaatgtggcctctgcccggcgaactgcaga
 agggaaacgaactggcctgccctccaaatatgtgaacttctgtacctggccagccactatgagaagctgaagggtccccgaggataatgagcaga
 aacagctgtttgtggaacagcacaagcactacctggacgagatcatcgagcagatcagcgagtttccaagagagtgatcctggccgacgctaactctgg
 acaaagtgtgtccgctacaacaagcaccgggataagcccatcagagagcaggccgagaatatcatccacctgtttacctgaccaatctgggagccc
 ctgccgcttcaagtactttgacaccaccatcgaccggaagaggtacaccagaccaaagaggtgctggacgccacctgatccaccagagcatcacgg
 gcctgtacgagacagggatcgacctgtctcagctgggaggcgacagcccaagaagaagagaaaggtggaggccagctaag (SEQ ID NO: 47)

Synthetic PolyA

Aattcaataaaaagatctttattttcattagatctgtgtgtttttgtgtgcqgccc (SEQ ID NO: 48)

3'ITR

Aggaaccctagtgtatggaattgacactccctctctgcgcgctcgtcgtcactgaagccggcgacccaagggtcagccgacgcccaggtttgcccggcgccctcagtgaagcgaagcgaagcgcgcag (SEQ ID NO:29)

complete sequence p1498_pAAV_HLP_SpCas9 + U6 3'malb_gRNA (5,1kb)

CTGCGCGCTCGCTCGCTCACTGAGGCCGCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGGTCGCCCCG
CCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCTTgtagttaatga
ttaaccgccatgctacttatctacgtagagctcttgcgaggtcgacatttaaatgaagggcgaattcaattttagggcctatttcccatgat
tcctcatatttgcataacgatacaaggctgtagagagataattggaattaatttgaactgtaaacacaaagatattagtacaaaatacgtga
cgtagaaagtaataatttcttgggttagtttgcagttttaaataattgttttaaattggactatcatatgcttaccgtaacttgaaagtatttgcatt
tcttggctttatatacttcttggaaaggacgaaacacccgtattttaaaggcagcagtggttttagagctagaaatagcaagttaaaataagg
ctagtccgttatcaacttgaaaaagtggcaccgagtcggtgctttttttagagctagaaatagcaagttaaaataaggctagtccgttt
ttagcgcgtgcgccaattctgcagacaaatggctctagaggtaccaatttacgtagctaagTGTGTGCTGCTTGCAATGTTTGCCC
ATTTTAGGGTGGACACAGGACGCTGTGTTTCTGAGCCAGGGGGCGACTCAGATCCCAGCCAGTGGACTTA
GCCCCTGTTTGCTCCTCCGATAACTGGGGTGACCTTGTTAATATTACCAGCAGCCTCCCCGTTGCCCTC
TGATCCACTGCTTAAATACGACGAGGACAGGGCCTGTCTCCTCAGCTTCAGGCACCACCACTGACCTG
GGACAGTGAATcaccggtggtacctgcttttgcctgcttggatccccggtgccaccatgtaccatacgtattccagattacgcttcgcc
gaagaaaaagcgcaaggtcgaagcgtccgacaagaagtaacagcatcggcctggacatcggcaccaactctgtgggtcgggctgatca
ccgacgagtacaaggtgccagcaagaattcaaggtgctgggcaacaccgaccggcacagcatcaagaagaacctgatcggagccctg
ctgttcgacagcggcgaaacagcgggaccccggtgaagagaaccgccagaagaagatacaccagacggaagaaccggtatctgct
atctgcaagagatcttcagcaacgagatggccaaggtggacgacagcttcttcacagactggaagagtccttctggtggaagaggata
agaagcacgagcggcaccccatcttcggcaacatcgtggacgaggtggcctaccacgagaagtacccacatctaccacctgagaaga
aactggtggacagcaccgacaaggccgacctgcggctgatctatctggcctggccacatgatcaagtccggggccacttctgatcgag
ggcgacctgaacccgacaacagcgacgtggacaagctgttcatccagctggtgcagacctacaaccagctgttcgaggaaaacccatc
aacgccagcggcgtggacgccaaggccatctgtctgcagactgagcaagagcagacggctgaaaatctgatcgccagctgcccggc
gagaagaagaatggcctgttcggaacactgattgcctgagcctgggctgaccccaactcaagagcaacttcgacctggcggaggtg
ccaaactgcagctgagcaaggacacctacgacgacacctggacaacctgctggccagatcggcgaccagtacgccacctgtttctggc
cgccaagaacctgtccgacgcatctgtgagcgacatctgagagtgaacaccgagatcaccaaggccccctgagcgctctatgatc
aagagatacgcagcagcaccacagacctgacctgctgaaagctctctgctggcagcagctgctgagaagtacaaagagattttcttc
gaccagagcaagaacggctacgcccgtacattgacggcgagccagccaggaagagtttacaagtcatcaagccatcttgaaaa
gatggacggcaccgaggaactgctcgtgaagtgaacagagaggacctgctgcggaagcagcggaccttcgacaacggcagcatcccc
accagatccacctgggagagctgcacgccattctcgcgccgaggaagattttaccattctgaaggacaacggggaaaagatcgaga
agatcctgaccttcgcatccctactacgtgggctctgcccaggggaaacagcagattcgctggatgaccagaaagagcgaggaaa
ccatccccctggaacttcgaggaagtgttggaagaaggcgcttccgcccagagcttcatcgagcggatgaccaacttcgataagaacct
gccaacgagaaggtgctgccaagcacagcctgctgtacgagtacttcacctgtataacgagctgacaaaagtgaatacgtgaccga
gggaatgagaaagcccgttctgagcggcgagcagaaaaaggccatcgtggacctgctgttcaagaccaaccggaaagtaccgtga
agcagctgaaagaggactacttaagaaaaatcgagtgttcgactcctggaaatctccggcgtggaagatcggttcaacgcctccctggg
cacataccacgatctgtgaaaattatcaaggacaaggacttctggacaatgaggaaaacgaggacattctggaagatatcgtgtgac
cctgacactgtttgaggacagagagatgatcgaggaacggctgaaaacctatgccacctgttcgacgacaaagtgtgaagcagctgaa
gcggcggagatacaccggctggggcaggctgagccggaagctgatcaacggcatccgggacaagcagtcgggcaagacaatcctggatt
tcctgaagtccgacggcttcgcaacagaaacttcagctgatccacgacgacagctgacctttaaagaggacatccagaaagccc
agggtgtccggccaggcgatagcctgcacgagcacattgccaatctggccggcagccccgcattaagaagggtcctgacagacgtga
agggtgtgacgagctcgtgaaagtgtggccggcacaagcccagaaacatcgtgatcgaatggccagagagaaccagaccacca
gaagggcagagaagaacagccgcgagagaatgaagcggatcgaagagggtcacaagagctgggagccagatcctgaaagaacacc
cgtggaacacccagctgcagaacgagaagctgtacctgtactacctgcagaatgggcccggatgtacgtggaccaggaactggacat
caaccggctgtccgactacgatgtggacatctcgtgcctcagagctttctgaaggacgactccatcgacaacaaggtgtgaccagaagc
gacaagaaccggggcaagagcgacaacgtgcctccgaagaggtcgtgaagaagtgaagaactactggcggcagctgctgaaccca
agctgattaccagagaaagtgcacaatctgaccaaggccgagagaggcgctgagcgaactggataaggccgcttcatcaagaga

cagctggtggaaaccggcgagatcacaaagcacgtggcacagatcctggactccgggatgaacactaagtagcagcagagaatgacaagct
gatccgggaagtgaagtgatcacccctgaagccaagctggtgtccgatttccggaaggatttccagttttacaaagtcgcgcgagatcaac
aactaccaccacgcccacgacgcctacctgaacccgtcgtgggaaccgcccctgatcaaaaagtaccctaagctggaaagcaggttctgtg
acggcgactacaagggtgtacgacgtgcggaagatgatcgcaagagcgcagcaggaaatcggcaaggctaccgccaagtacttcttctaca
gcaacatcatgaacttttcaagaccgagattaccctggccaacggcgagatccggaagcggcctctgatcgagacaaacggcgaaaccg
gggagatcgtgtgggataagggcgggattttgccaccgtgcggaaagtgtgagcatgccccagtgaatatcgtgaaaaagaccgagg
tgcagacaggcggcttcagcaaagagtctatcctgccaaagaggaacagcgataagctgatcgccagaaagaaggactgggaccctaag
aagtacggcggcttcgacagccccaccgtggcctatttctgtgctggtggcgaagtggaaaagggcaagtccaagaaactgaagagt
gtgaaagagctgctggggatcacatcatggaaagaagcagcttcgagaagaatccatcgactttctggaagccaagggctacaaaga
agtgaaaaaggacctgatcatcaagctgcctaagtactccctgttcgagctggaaaacggccggaagagaatgtctggcctctgccggcga
actgcagaagggaaaacgaactggccctgccctccaaatatgtgaacttctgtacctggccagccactatgagaagctgaagggctcccc
gaggataatgagcagaaacagctgtttgtggaacagcacaaagcactacctggacgagatcatcgagcagatcagcgagtttccaagag
agtgatcctggccgacgctaattctggacaaagtgtgtccgcctacaacaagcaccgggataagcccatcagagagcaggccgagaatat
catccacctgtttacctgaccaatctgggagcccctgccgccttaagtaactttgacaccaccatcgaccggaagaggtacaccagcacca
aagaggtgctggacgccaccctgatccaccagagcatcaccggcctgtacgagacacggatcgacctgtctcagctgggaggcgacagcc
ccaagaagaagagaaaggtggaggccagctaaggaattcaataaaaagatctttattttcattagatctgtgtgttqgtttttgtgtgcgc
cgcaggaacccctagtgtatgaagttggccactccctctctgcgcgctcgtctcgtcactgaagggccggcgaccaaaaggtgcgccgacg
ccccggctttgccggggcgccctcagtgaagcgaagcgaagcgcga (SEQ ID NO:49)

p1500//pAAV_ HLP_SpCas9 + U6 3'malb_scRNA

5'ITR

CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGGTCGCCCCGG
CCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCCT (SEQ ID
NO:110)

Additional AAV sequences

Tgtagttaatgattaacccgccatgctacttatctacgtagagctcttgctcgaggctcgac (SEQ ID NO:50)

U6 expression cassette scRNA

Ctgacctcgagtttcccatgattccttcataatttgcataacgatacaaggctgttagagagataattggaattaatttgactgtaaacacaaa
gatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttttaaaatggactatcatatgcttac
cgtaacttgaaagtatttcgatttcttggctttatatacttgtggaaaggacgaaacaccggactcgcgcgagtcgaggaggttttagagcta
gaaatagcaagttaaaataaggctagtcggttatcaacttgaaaaagtggcaccgagtcggtgctttttgttttagagctagaaatagcaag
(SEQ ID NO:51)

HLP PROMOTER

TGTTTGCTGCTTGAATGTTTGCCCATTTTAGGGTGGACACAGGACGCTGTGGTTTCTGAGCCAGGGGGCGACTCAGATCCCAGCCAGTGGACTTAGCCCTGTTTGCTCCTCCGATAACTGGGGTGACCTTGGTTAATATTACCAGCAGCCTCCCGGTTGCCCCTCTGGATCCACTGCTTAAATACGGACGAGGACAGGGCCCTGTCTCCTCAGCTTCAGGCACCACTGACCTGGGACAGTGAAT (SEQ ID NO:46)

spCas9

atgtaccatac gatgttcagattacgcttcgccgaagaaaaagcgcaaggctgaagcgtccgacaagaagtacagcatcggcctggacatcggcac
caactctgtgggctgggccgtgatcccgacgagtacaagggtcccagcaagaaattcaagggtctgggcaacaccgaccggcacagcatcaagaag

Synthetic PolyA

3' ITR

*Gcaggaaccctagtgatggagttggccactccctctctgcgcgctcgctcactgaggccgggaccaaaggctgcccgcgcccgggtt
tgccggggcgccctcagtgaagcgagcgagcgcgag (SEQ ID NO:29)*

96

CTGCGCTCGCTCGCTCACTGAGGCCGCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGGTCGCCCCG
CCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCTT
tgtagttaatgattaacccgccatgctacttatctacgtagagctcttgtcgaggctgacctgacctcgagtttccatgattccttcattttgc
atatacgatacaaggctgttagagagataattggaattaatttgactgtaaacacaaagatatttagtcaaaaatcgtgacgtagaaagtaa
taatttcttggtagtttgagttttaaattatgttttaaatggactatcatatgcttacgtaactgaaagtatttcgatttcttggctttata
tatcttgtggaaaggacgaaacaccggactcgcgcgagtcgaggaggttttagagctagaaatagcaagttaaaataaggctagtcggttat
caacttgaaaaagtggcaccgagtcggtgctttttgttttagagctagaaatagcaagtcagaggtaccaatttacgtagctaagTGTTT
GCTGCTTGCAATGTTTGCCATTTTAGGGTGGACACAGGACGCTGTGTTTCTGAGCCAGGGGGCGACTCA
GATCCCAGCCAGTGGACTTAGCCCCCTGTTTGCTCCTCCGATAACTGGGGTGACCTTGTTAATATTACAG
CAGCCTCCCCGTTGCCCTCTGGATCCACTGCTTAATAACGACGAGGACAGGGCCCTGTCTCCTCAGCTT
CAGGCCACCACCTGACCTGGGACAGTGAATcaccggtggtacctgctttgctcgcttggatccccggtgccaccatgtaccc
atacgatgttccagattacgcttcgccgaagaaaaagcgcaaggctgaagcgtccgacaagaagtacagcatcggcctggacatcggcac
caactctgtgggctgggctgacacccgagtcacaaggtgccagcaagaattcaaggtgctgggaacaccgaccggcacagcat
caagaagaacctgatcgagccctgctgttcgacagcggcgaaacagccgaggccaccggctgaagagaaccgccagaagaagatac
accagacggaagaaccgatctgctatctgaagagatcttcagcaacgagatggccaaggtggacgacagcttcttcacagactggaa
gagtccttctgtggaagaggataagaagcacgagcggcaccatcttcggcaacatcgtggacgaggtggcctaccacgagaagtac
cccaccatctaccacctgagaaagaaactgggtggacagcaccgacaaggccgacctgcggtgatctatctggccctggcccacatgatca
agttccggggccacttctgatcgagggcgacctgaacccccgacaacagcgacgtggacaagctgttcatccagctggtgcagacctaaa
ccagctgttcgaggaaccccatcaacgccagcggcgtggacgccaaggccatcctgtctgccagactgagcaagagcagacggctgga
aaatctgatcgccagctgcccggcgagaagaagaatggcctgttcggaacactgattgccctgagcctgggctgaccccaactcaag
agcaacttcgacctggcggaggtgccaactgcagctgagcaaggacacctacgacgacacctggacaacctgctggcccagatcggc
gaccagtacccgacctgtttctggccgccaagaacctgtccgacgccatcctgctgagcgacatcctgagagtgaacaccgagatcacca
aggccccctgagcgcctctatgatcaagagatacgcagcagcaccaccaggacctgacctgctgaaagctctcgtgcggcagcagctgcc
tgagaagtacaaagagattttcttcgaccagagcaagaacggctacgcccgtacattgacggcggagccagccaggaagagtctacaa
gttcatcaagcccatcctggaaaagatggacggcaccgaggaactgctcgtgaagctgaacagagaggacctgctgcggaagcagcggga
ccttcgacaacggcagcatccccaccagatccacctgggagagctgcacgccattctgcggcggcaggaagattttaccattcctgaag
gacaaccgggaaaagatcgagaagatcctgacctccgcatcccctactacgtgggcccctgaggcaggggaaacagcagattcgctgg
atgaccagaaagagcgaggaaacctacccccctggaacttcgaggaaagtggtggacaagggcgcttcgcccagagcttcatcgagcg
gatgaccaacttcgataagaacctgcccaacgagaaggtgctgccaagcacagcctgctgtacgagtacttcacctgtataacgagctg
accaaagtgaatacgtgaccgaggggaatgagaaagccccttctgagcggcgagcagaaaaaggccatcgtggacctgctgttcaa
gaccaaccggaaagtgacctgaagcagctgaaagaggactattcaagaaaatcgagtgccttcgactcctggaaatctccggcgtgga
agatcgggtcaacgcctccctgggacataccagatctgctgaaaattatcaaggacaaggacttctggacaatgaggaaaacgagga
cattctggaagatatcgtgctgacctgacactgtttgaggacagagagatgatcgaggaacggctgaaaacctatgccacctgttcgac
gacaaagtgatgaagcagctgaagcggcgagatacaccggctggggcaggctgagccggaagctgatcaacggcatccgggacaagc
agtccggcaagacaatcctggatttctgaagtccgacggcttcgccaacagaaacttcagctgatccacgacgacagcctgacctt
aaaggagcatccagaaagcccaggtgtccggcaggggcagatagcctgcacgagcacattgccaatctggccggcagccccgccattaag
aagggcatcctgcagacagtgaaagtggtggacgagctcgtgaaagtgatgggcccgcacaagcccagaaacatcgtgatcgaatggc
cagagagaaccagaccaccagaaggacagaagaacagccgcgagagaatgaagcggatcgaagagggcatcaaagagctgggca
gccagatcctgaaagaacccccgtggaacacccagctgcagaacgagaagctgtacctgtactacctgcagaatgggcccggatgt
acgtggaccaggaactggacatcaaccggctgctcgactacgatgtggaccatatcgtgcctcagagcttctgaaggacgactccatcga
caacaaggtgctgaccagaagcgacaagaacccggggcaagagcgacaacgtgccctccgaagaggtcgtgaagaagatgaagaacta
ctggcggcagctgctgaacgccaagctgattaccagagaaagttcgacaatctgaccaaggccgagagaggcgccctgagcgaactgg
ataaggccggcttcatcaagagacagctggtggaacccggcagatcacaagcacgtggcacagatcctggactcccggatgaacact
aagtacgacgagaatgacaagctgatccgggaagtgaagtgatcacctgaagtccaagctggtgtccgatttccggaaggatttccag
ttttacaaagtgcgcgagatcaacaactaccaccagcccacgacgctacctgaacgccgtcgtgggaaccgccctgatcaaaaagtacc
ctaagctggaaagcgagttcgtgtacggcgactacaaggtgtacgacgtgcggaagatgatcgccaagagcgagcaggaatcggcaag
gctaccgccaagtacttcttacagcaacatcatgaacttttcaagaccgagattaccctggccaacggcgagatccggaagcggcctct
gatcgagacaaacggcgaaacccggggagatcgtgtgggataaggggccgggattttgccacctgcggaagtgctgagcatgccccag

[illegible]

**P1617_ pTIGEM_HITI 3'malb CodopV3 HITI donor
5'ITR**

CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCTGGGCGACCTTTGGTCGCCCCGG
CCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCCT (SEQ ID
NO:110)

Inverted gRNA +pam site
GTATTTAATAGGCAGCAGTGTGG (SEQ ID NO: 54)

Synthetic splicing Acceptor
GATAGGCACCTATTGGTCTTACTGACATCCACTTTGCCTTTCTCTCCACAG (SEQ ID NO:21)

mAlbumin exon 14
 ggtccaaaccttgctactagatgcaagacgccttagcc (SEQ ID NO:22)

T2A sequence
Ggaagcggagagggcagaggaagtctgctaacatgcggtgacgtcgaggagaatcctggacct (SEQ ID NO:23)

Codopv3

atgcagattgagctgagcacctgcttcttctgtgcctgctgaggttctgcttcttctgccaccaggagatactacctgggggctgtggagctgag
ctgggactacatgcagcttgacctgggggagctgcctgtggatgccaggttccccccagagtgcccaagagcttccccttcaacacctctgtg
gtgtacaagaagaccctgtttgtggagttcactgaccacctgttcaacattgccaaagccaggccccctggatgggctgtctgggccccacc
atccaggctgaggtgtatgacactgtggtgatcacctgaagaacatggccagccacctgtgagcctgcatgctgtgggggtgagctactgg
aaggcctctgaggggggctgagtatgatgaccagaccagccagaggggagaaggaggatgacaaggtgttccctgggggcagccacacctat
gtgtggcaggtgctgaaggagaatggccccatggcctctgacccccctgtgcctgacctacagctacctgagccatgtggacctggtgaagga
cctgaactctggcctgattggggccctgctggtgtgcagggaggggcagcctggccaaggagaagaccagacctgcacaagttcatcctgc
tgtttgctgttttgatgagggcaagagctggcactctgaaaccaagaacagcctgatgcaggacaggggatgctgcctctgccagggcctgg
ccaagatgcacacttgaatggctatgtgaacaggagcctgcctggcctgattggctgccacaggaagtctgtgtactggcatgtgattggc
atgggcaccaccccctgaggtgcacagcatcttctgaggggccacaccttctggtcaggaaccacaggcaggccagcctggagatcagccc
catcaccttctgactgccagacctgctgatggacctggggccagttcctgctgttctgccacatcagcagccaccagcatgatggcatggag
gcctatgtgaaggtggacagctgccctgaggagccccagctgaggatgaagaacaatgaggaggctgaggactatgatgatgacctgactg
actctgagatggatgtggtgaggtttgatgatgacaacagccccagcttcatccagatcaggtctgtggccaagaagcaccccaagacctgg
gtgcactacattgctgctgaggaggaggactgggactatgccccctggctgctggccctgatgacaggagctacaagagccagtacctgaa
caatggccccccagaggattggcaggaagtacaagaaggtcaggttcatggcctacactgatgaaaccttcaagaccaggggaggccatccag

catgagtctggcatcctgggccccctgctgtatggggagggtgggggacacccctgctgatcatcttcaagaaccaggccagcaggccctacaa
catctacccccatggcatcactgatgtgaggccccctgtacagcaggaggctgcccaagggggtgaagcacctgaaggacttccccatcctgc
ctggggagatcttcaagtacaagtggactgtgactgtggaggatggccccaccaagtctgacccaggtgcctgaccagatactacagcagc
tttgaacatggagaggggacctggcctctggcctgattggccccctgctgatctgctacaaggagtctgtggaccagaggggcaaccagatc
atgtctgacaagaggaatgtgatcctgttctctgtgtttgatgagaacaggagctggtacctgactgagaacatccagaggttctgccaac
cctgctggggtgcagctggaggacacctgagttccaggccagcaacatcatgcacagcatcaatggctatgtgtttgacagcctgcagctgtct
gtgtgctgcatgagggtggcctactggtacatcctgagcattggggcccagactgacttctgtctgtgttcttctggtacaccttcaagca
caagatggtgtatgaggacacccctgacctgttcccccttctggtggagactgtgttcatgagcatggagaacctggcctgtggattctgggc
tgccacaactctgacttcaggaacaggggcatgactgccctgctgaaagtctccagctgtgacaagaacactggggactactatgaggacag
ctatgaggacatctctgcctacctgctgagcaagaacaatgccattgagcccaggagcttcagccagaatgccactaatgtgtctaacaaca
gcaacaccagcaatgacagcaatgtgtctccccagtgctgaagaggcaccagagggagatcaccaggaccacctgcagctgaccagg
aggagattgactatgatgacacatctctgtggagatgaagaaggaggactttgacatctacgacgaggacgagaaccagagccccagga
gcttcagaagaagaccaggcactacttcattgctgctgtggagaggctgtgggactatggcatgagcagcagcccccatgtgctgaggaaac
agggcccagctgtgctctgtgccccagttcaagaaggtggtgttccaggagttcactgatggcagcttcacccagcccctgtacagaggggag
ctgaatgagcacctgggctgtggtggccccctacatcagggtgaggtggaggacaacatcatggtgaccttcaggaaccaggccagcaggc
cctacagcttctacagcagcctgatcagctatgaggaggaccagaggcagggggctgagcccaggaagaactttgtgaagcccaatgaaac
caagacctacttctggaaggtgcagcaccacatggccccaccaaggatgagtttactgcaaggcctgggctacttctctgatgtggacct
ggagaaggatgtgactctggcctgattggccccctgctggtgtgccacaccaacacctgaaacctgcccagggcaggtgactgtgtgc
aggagtttgcctgttcttaccatcttggatgaaaccaagagctggtacttactgagaacatggagaggaactgcagggccccctgcaaca
tccagatggaggacccaccttcaaggagaactacaggttccatgccatcaatggctacatcatggacacctgctggcctggtgatggccc
aggaccagaggatcaggtgttacctgctgagcatgggcagcaatgagaacatccacagcatccacttctggtccatgtgttactgtgagg
aagaaggaggagtacaagatggcctgtacaacctgtacctggtgtgttggagactgtggagatgtgcccagcaaggctggcatctggag
ggtggagtgcctgattggggagcacctgcatgctggcatgagcacctgttctggtgtacagcaacaagtggcagacccccctgggcatgg
cctctggccacatcagggacttccagatcactgcctctggccagtatggccagtgggcccccaagctggccaggctgcactacttggcagca
tcaatgcctggagcaccaaggagcccttcagctggatcaagggtggacctgtggcccccatgatcatccatggcatcaagaccagggggcc
aggcagaagtgcagcagcctgtacatcagccagttcatcatcatgtacagcctggatggcaagaagtggcagacctacaggggcaacagca
ctggcacctgatggtgttcttggcaatgtggacagctctggcatcaagcacaacatcttcaaccccccatcattgccagatacatcaggct
gcacccaccactacagcatcaggagcacctgaggatggagctgatgggctgtgacctgaacagctgcagcatgccctgggcatggag
agcaaggccatctctgatgccagatcactgccagcagctacttcaccaacatgtttgccacctggagccccagcaaggccaggctgcacct
gcagggcaggagcaatgcctggaggccccaggtcaacaacccaaggagtggctgcaggtggacttccagaagacctgaaggtagtgg
ggtgaccacccagggggtgaagagcctgctgaccagcatgtatgtgaaggagtctctgatcagcagcagccaggatggccaccagtggacc
ctgttcttccagaatggcaaggtgaaggtgttcagggcaaccaggacagcttcacccctgtggtgaacagcctggacccccctgctgacc
agatacctgaggattcacccccagagctgggtgcaccagattgcctgaggatggaggtgctgggctgtgaggccaggacctgtac (SEQ
ID NO: 55)

3XFLAG

GACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGACTACAAGGATGACGATGACAAGTGA
(SEQ ID NO:56)

Synthetic polyA

Tcgcaataaaaagatctttattttcattagatctgtgtgttggtttttgtgtgatgcagc (SEQ ID NO:37)

inverted gRNA AND PAM

gtatttaataggcagcagtggtg (SEQ ID NO:54)

ADDITIONAL AAV SEQUENCES

GAGCTCTGTGTCGAGGTCGACATTTAAATGAATTCCAATTG (SEQ ID NO:57)

3'itr

Aggaacccttagtgatggagttggccactccctctctgcgcgctcgtcgtcactgaggccggcgaccaaaggtcgcccgacgcccgggctttgcccggcgccctcagtgagcgagcgagcgcgag (SEQ ID NO:29)

CTGCGCGCTCGCTCGCTACTGAGGCCGCCCCGGGCAAAGCCCCGGGCGTCGGGCGACCTTTGGTCGCCCCG
CCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCTTGctagcGTATT
TAATAGGCAGCAGTGTGGGATAGGCACCTATTGGTCTTACTGACATCCACTTTGCCTTTCTCTCCACAGggtc
caaaccttgtcactagatgcaagagcgcttagccggaagcggagagggcagaggaagtctgctaacatcggtgacgtcgaggagaa
tcctggacctatgcagattgagctgagcacctgcttctcctgtgcctgctgaggttctgcttctgcccaccaggagatactacctgggggctg
tggagctgagctgggactacatgcagtctgacctgggggagctgcctgtggatgccaggttccccccagagtgcccaagacttcccttca
acacctctgtggtgtacaagaagaccctgtttgtggagttcactgaccacctgttcaacattgccaagcccaggccccctggatgggcctgct
gggccccaccatccaggctgaggtgtatgacactgtggtgatcacctgaagaacatggccagccacctgtgagcctgcatgctgtggggg
tgagctactggaaggcctctgaggggctgagtatgatgaccagaccagccagagggagaaggaggatgacaaggtgttccctgggggca
gccacacctatgtgtggcaggtgtgaaggagaatggccccatggcctctgacccctgtgctgacctacagctacctgagccatgtggacc
tgggtgaaggacctgaactctggcctgattggggcctgctggtgtgcagggagggcagcctggccaaggagaagaccagacctgcacaa
gttcatcctgctgtttgctgtgtttgatgagggcaagagctggcactctgaaccaagaacagcctgatgcaggacagggatgtgcctctgc
cagggcctggccaagatgcacactgtgaatggctatgtgaacaggagcctgcctggcctgattggctgccacaggaagtctgtgtactggc
atgtgattggcatgggcaccacccctgaggtgcacagcatcttctggaggggccacaccttctggtcaggaaccacaggcaggccagcctg
gagatcagccccatcaccttctgactgccagacctgctgatggacctgggacgttctgctgttctgccacatcagcagccaccagcatg
atggcatggaggcctatgtgaaggtggacagctgcctgaggagccccagctgaggatgaagaacaatgaggaggtgaggactatgatg
atgacctgactgactctgagatggatgtggtgaggtttgatgatgacaacagcccagcttcatccagatcaggtctgtggccaagaagcacc
ccaagacctgggtgactacattgctgctgaggaggaggactgggactatgccccctgggtgctggccccctgatgacaggagctacaagagc
cagtacctgaacaatggccccagaggattggcaggaagtacaagaaggtcaggttcatggcctacactgatgaaacctcaagaccaggg
aggccatccagcatgagctggcatcctgggccccctgctgtatggggaggtgggggacacccctgctgatcatcttcaagaaccaggccagc
aggccctacaacatctacccccatggcatcactgatgtgaggccccctgtacagcaggaggctgccaagggggtgaagcacctgaaggactt
ccccatcctgctggggagatcttcaagtacaagtggactgtgactgtggaggatggccccaccaagtctgacccaggtgctgaccagat
actacagcagctttgtgaacatggagagggacctggcctctggcctgattggccccctgctgatctgctacaaggagtctgtggaccagaggg
gcaaccagatcatgtctgacaagaggatgtgatcctgttctctgtgtttgatgagaacaggagctggtacctgactgagaacatccagaggt
tcctgccaacccctgctggggtgcagctggaggacctgagttccaggccagcaacatcatgcacagcatcaatggctatgtgtttgacagcc
tgcagctgtctgtgtcctgcatgaggtggcctactggtacatcctgagcattggggcccagactgacttctgtctgttcttctgtgctaca
ccttcaagcacaagatgggtgatgaggacacctgacctgttcccccttcttggggagactgtgttcatgagcatggagaacctggcctgtg
gattctgggctgccacaactctgacttcaggaacaggggcatgactgcctgtgaaagtctccagctgtgacaagaacactggggactact
atgaggacagctatgaggacatctctgcctacctgctgagcaagaacaatgccattgagcccaggagcttcagccagaatgccactaatgtg
tctaacaacagcaacaccagcaatgacagcaatgtgtctccccagtgctgaagaggcaccagagggagatcaccaggaccacctgcagt
ctgaccaggaggagattgactatgatgacacctctctgtggagatgaagaaggaggactttgacatctacgacgaggacgagaaccagag
ccccaggagcttcagaagaagaccaggcactacttcattgctgctgtggagaggctgtgggactatggcatgagcagcagccccatgtgc
tgaggaacagggccagctgtgctgtgtgccccagttcaagaaggtggtgttcaggagttcactgatggcagcttcacccagcccctgtaca
gaggggagctgaatgagcacctgggctgtgggccccctacatcagggtgaggtggaggacaacatcatggtgaccttcaggaaccaggc
cagcaggccctacagcttctacagcagcctgatcagctatgaggaggaccagaggcagggggctgagcccaggaagaactttgtgaagccc
aatgaaaccaagacctacttctggaaggtgcagcaccacatggccccaccaaggatgagtttactgcaaggcctgggctacttctctgat
gtggacctggagaaggatgtgcaactgtgctgattggccccctgctggtgtgccacaccaacacctgaacctgcccattggcaggcaggt
gactgtgcaggagtttccctgttcttaccatctttgatgaaaccaagagctggtacttctactgagaacatggagaggaactgcaggggccc
ctgcaacatccagatggaggacccccaccttcaaggagaactacaggttccatgccatcaatggctacatcatggacacctgcctggcctggt
gatggcccaggaccagaggatcaggtggtacctgtgagcatgggcagcaatgagaacatccacagcatccacttctgtggccatgtgttca
ctgtgaggaagaaggaggagtacaagatggccctgtacaacctgtaccctggggtgtttgagactgtggagatgtgcccagcaaggctggc
atctggagggtggagtgcctgattggggagcacctgcatgctggcatgagcacctgttctggtgtacagcaacaagtgccagacccccctg
ggcatggcctctggccacatcagggacttcagatcactgcctctggccagtatggccagtgggcccccaagctggccaggctgcactactct
ggcagcatcaatgcctggagcaccaaggagcccttcagctggatcaaggtggacctgctggccccatgatcatccatggcatcaagacca

gggggcccaggcagaagttcagcagcctgtacatcagccagttcatcatcatgtacagcctggatggcaagaagtggcagacctacaggggc
aacagcactggcaccctgatggtgttctttggcaatgtggacagctctggcatcaagcacaacatcttcaaccccccatcattgccagatac
atcaggctgcacccaccactacagcatcaggagcaccctgaggatggagctgatgggctgtgacctgaacagctgcagcatgccctggg
catggagagcaaggccatctctgatgccagatcactgccagcagctacttcaccaacatgtttgccacctggagccccagcaaggccaggc
tgcacctgcagggcaggagcaatgcctggaggccccaggtaacaacccaaggagtggctgcaggtggacttcagaagaccatgaagg
tgactggggtgaccacccagggggtgaagagcctgctgaccagcatgtatgtgaaggagttcctgatcagcagcagccaggatggccacca
gtggacctgttcttcagaatggcaaggtgaaggtgttcagggaaccaggacagcttcacccctgtggtgaacagcctggacccccct
gctgaccagatacctgaggattacccccagagctgggtgcaccagattgcctgaggatggaggtgctgggctgtgaggcccaggacctgt
acGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGACTACAAGGATGACGATGACAAGTGA
tcgcgaataaaagatctttatcttcattagatctgtgtgttgggtttttgtgtgatgcagcgtatttaataggcagcagtggtggGAGCTCTT
GTCGAGGTGCGACATTAAATGAATTCCAATTGaggaaccctagtgtgatggagtggccactccctctctgcgcgctcgctcgct
cactgaggccgggacgacaaaggtcggcgacggcggtttgcccgggcgccctcagtgcgagcgagcgcgag (SEQ ID
NO:58)

Sequences of above Example 4**HITI 3'human ALb**

Construct p939_ pCbh-SpCas9(BB)-2A-GFP+Scramble gRNA

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGGAATT
 AATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTT
 GCAGTTTTTAAAATTATGTTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTCGATTTCTTGG
 CTTTATATATCTTGTGGAAAGGACGAAACACC (SEQ ID NO: 59)

Human U6 promoter

Gactcgcgcgagtcgaggag (SEQ ID NO:111)

Scramble RNA

GTTTTAGAGCTAGAAATAGCAAGTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC
GGTGCTTTTTT (SEQ ID NO:60)

Chimeric gRNA scaffold

cgttacataacttacggtaaattggcccgctggctgaccgccaacgacccccgccattgacgtcaatagtaacgccaatagggactttcca
 ttgacgtcaatgggtggagtatTTacggtaaactgccacttggcagtagcatcaagtgtatcatatgccaagtagccccctattgacgtcaat
 gacggtaaattggcccgctggcattgtgccagtagcatgacctatgggactttctacttggcagtagcatctacgtattagtcacgcctattac
 catggtcgaggtgagccccacgttctgcttactctccccatctccccccctccccaccccccaattttgtattttattttttaattattttgtgc
 agcgatgggggcgggggggggggggggcgcgcgccaggcggggcggggcgaggggcggggcgaggcgagaggtgc
 ggcggcagccaatcagagcggcgcgctccgaaagtcttctttatggcgaggcgggcgggcgggcgccctataaaaagcgaagcgcgcg
 cgggcgggagtcgctgcgacgctgccttcgccccgtgccccgctccgcccgcctcgcgccgccccgggctctgactgaccgcttact
 cccacaggtgagcgggcgggacggcccttctctccgggctgtaattagctgagcaagaggtaagggttaagggttggttggtgggg
 tattaatgtttaattacctggagcacctgcctgaaatcacttttttcaggttg (SEQ ID NO:61)

CBH promoter

ATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAG
(SEQ ID NO: 62)

3XFlag tag

ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGACAAGAAGTACAGCATCG
GCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAA
ATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTTCGACA
GCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACC
GGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCACAGACTG
GAAGAGTCCTTCTGGTGGAAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACG
AGGTGGCCTACCACGAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACAA
GGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCACATGATCAAGTTCGGGGCCACTTCTGATCGAGG

GCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTTCATCCAGCTGGTGCAGACCTACAACCAGCTG
TTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGA
GCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTTCGGAAACCTGAT
TGCCCTGAGCCTGGGCCTGACCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAACTGCAGC
TGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCAGATCGGCGACCAGTACGCCGACCT
GTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCA
CCAAGGCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAA
GCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGC
CGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATG
GACGGCACCCGAGGAAGTCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGAC
AACGGCAGCATCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTC
CCATTCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCGCATCCCCTACTACGTGGGCC
CTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCTGGAA
CTTCGAGGAAGTGGTGGACAAGGGCGCTTCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAG
AACCTGCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCT
GACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGC
CATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAG
AAAATCGAGTGCTTCGACTCCGTGGAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATA
CCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAA
GATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCA
CCTGTTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCG
GAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCTGAAGTCCGACGGCT
TCGCCAACAGAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCC
CAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGA
AGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAGTGATGGGCGGCACAAAGCCCGAGA
ACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAA
TGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAACA
CCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGA
ACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACT
CCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGA
GGTCTGTAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTC
GACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCGCGCTTCATCAAGAGACAGC
TGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGA
CGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCTGAAGTCCAAGCTGGTGTCCGATTTCCGG
AAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACCTACCACCACGCCACGACGCCTACCTGAACGCC
GTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGG
TGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTT
CTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTC
TGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGA
AAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGA
GTCTATCCTGCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTAC
GGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAGGGCAAGTCCA
AGAACTGAAGAGTGTAAGAGCTGCTGGGGATCACCATCATGGAAGAAGCAGCTTCGAGAAGAATCC
CATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACT
CCCTGTTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGA
ACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCC
CGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAAGCACTACCTGGACGAGATCATCGAGCAG
ATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAA

GCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAG
 CCCCTGCCGCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAGAGGTGCTG
 GACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGG
 CGACAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAGAAAAAG (SEQ ID NO:112)

5'Nuclear localization signal + SpCas9 + 3'Nuclear localization signal

GGCAGTGGAGAGGGCAGAGGAAGTCTGCTAACATGCGGTGACGTGAGGAGAATCCTGGCCCA (SEQ ID
 NO: 63)

Thosea Asigna Virus T2A skipping peptide

GTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAA
 CGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTT
 CATCTGCACCACCGGCAAGCTGCCCGTGGCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGT
 GCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCGAAGGCTACGTC
 CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGC
 GACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCAC
 AAGCTGGAGTACAACATAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAG
 GTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAAC
 ACCCCCATCGGCGACGGCCCCGTGCTGCTGCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAA
 AGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTCGTGACCGCCGCGGGGATCACTCTCGGC
 ATGGACGAGCTGTACAAGGAATTCTAA (SEQ ID NO:64)

EGFP fusion protein

Ctagagctcgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgccccctccccctgccttcttgaccctggaaggtgcca
 ctcccactgtcctttcctaataaaatgaggaaattgcatcgattgtctgagtaggtgtcattctattctgggggtgggggtggggcaggacag
 caagggggaggattgggaagagaatagcaggcatgctgggga (SEQ ID NO:65)

BGH polIA

AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCA
AAGGTCGCCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG
 (SEQ ID NO:66)

3'ITR

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGGAATT
 AATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTT
 GCAGTTTTAAATATGTTTTAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTCGATTTCTTGG
 CTTTATATATCTTGTGGAAGGACGAAACACCGgactcgcgagtcgaggagGTTTTAGAGCTAGAAATAGCAA
 GTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTGTTTTAGAGCTA
 GAAATAGCAAGTTAAAATAAGGCTAGTCCGTTTTTAGCGCGTGCGCCAATTCTGCAGACAAATGGCTCTAG
 AGGTACCggttacataacttacggtaaatggcccgctggctgaccgccaacgacccccgccattgacgtcaatagtaacgccaatag
 ggacttttcattgacgtcaatgggtggagtatttacggtaaactgccacttggcagtagcatcaagtgtatcatatgccaagtacgccccctatt
 gacgtcaatgacggtaaatggcccgctggcattgtgccagtagcatgacctatgggactttctacttggcagtagcatctacgtatttagtcat
 cgctattaccatggtcgaggtgagccccacgttctgcttactctccccatctccccccctccccaccccccaattttgtattttatttttaatt

tat t t t t g t g c a g c g a t g g g g g c g g g g g g g g g g g g g g g c g c g c g c c a g g c g g g g c g g g g c g g g g c g a g g g g c g g g g c g g g g c g a g g c g
g a g a g g t g c g g c g g c a g c c a a t c a g a g c g g c g c t c c g a a g t t t c t t t a t g g c g a g g c g g c g g c g g c c t a t a a a a g c g a
a g c g c g c g g c g g g a g t c g t g c g a c g t g c t t c g c c c g t g c c c g t c c g c c c g c c t c g c c g c c c g c c c g g c t g a c t g a
c c g c g t t a c t c c c a c a g g t g a g c g g g c g g a c g g c c t t c t c c t c g g g c t g t a a t t a g c t g a g c a a g a g g t a a g g g t t a a g g g a t g g t t g
t t g t g g g g t a t t a a t g t t t a a t t a c c t g g a g c a c c t g c c t g a a a t c a c t t t t t c a g g t t g g A C C G T G C C A C C A T G G A C T A T A A G
G A C C A C G A C G G A G A C T A C A A G G A T C A T G A T A T T G A T T A C A A G A C G A T G A C G A T A A G A T G G C C C C A A A G
A A G A A G C G G A A G G T C G G T A T C C A C G G A G T C C C A G C A G C C G A C A A G A A G T A C A G C A T C G G C C T G G A C A T C G
G C A C C A A C T C T G T G G G C T G G G C C G T G A T C A C C G A C G A G T A C A A G G T G C C C A G C A A G A A A T T C A A G G T G C T
G G G C A A C A C C G A C C G G C A C A G C A T C A A G A A G A A C C T G A T C G G A G C C C T G C T G T T C G A C A G C G G C G A A A C A
G C C G A G G C C A C C C G G C T G A A G A G A A C C G C C A G A A G A A G A T A C A C C A G A C G G A A G A A C C G G A T C T G C T A T C
T G C A A G A G A T C T T C A G C A A C G A G A T G G C C A A G G T G G A C G A C A G C T T C T T C C A C A G A C T G G A A G A G T C C T T C
C T G G T G G A A G A G G A T A A G A A G C A C G A G C G G C A C C C C A T C T T C G G C A A C A T C G T G G A C G A G G T G G C C T A C C
A C G A G A A G T A C C C C A C C A T C T A C C A C C T G A G A A G A A A C T G G T G G A C A G C A C C G A C A A G G C C G A C C T G C G
G C T G A T C T A T C T G G C C C T G G C C C A C A T G A T C A A G T T C C G G G G C C A C T T C C T G A T C G A G G G C G A C C T G A A C C C
C G A C A A C A G C G A C G T G G A C A A G C T G T T C A T C C A G C T G G T G C A G A C C T A C A A C C A G C T G T T C G A G G A A A A C C
C C A T C A A C G C C A G C G G C G T G G A C G C C A A G G C C A T C C T G T C T G C C A G A C T G A G C A A G A G C A G A C G G C T G G A
A A A T C T G A T C G C C C A G C T G C C C G G C G A G A A G A A G A A T G G C C T G T T C G G A A A C C T G A T T G C C C T G A G C C T G G
G C C T G A C C C C A A C T T C A A G A G C A A C T T C G A C C T G G C C G A G G A T G C C A A C T G C A G C T G A G C A A G G A C A C C
T A C G A C G A C G A C C T G G A C A A C C T G C T G G C C C A G A T C G G C G A C C A G T A C G C C G A C C T G T T T C T G G C C G C C A A
G A A C C T G T C C G A C G C C A T C C T G C T G A G C G A C A T C C T G A G A G T G A A C A C C G A G A T C A C C A A G G C C C C C C T G A
G C G C C T C T A T G A T C A A G A G A T A C G A C G A G C A C C A C C A G G A C C T G A C C C T G C T G A A A G C T C T C G T G C G G C A G
C A G C T G C C T G A G A A G T A C A A A G A G A T T T T C T T C G A C C A G A G C A A G A A C G G C T A C G C C G G C T A C A T T G A C G G
C G G A G C C A G C C A G G A A G A G T T C T A C A A G T T C A T C A A G C C C A T C C T G G A A A A G A T G G A C G G C A C C G A G G A A
C T G C T C G T G A A G C T G A A C A G A G A G G A C C T G C T G C G G A A G C A G C G G A C C T T C G A C A A C G G C A G C A T C C C C C
A C C A G A T C C A C C T G G G A G A G C T G C A C G C C A T T C T G C G G C G G C A G G A A G A T T T T A C C C A T T C C T G A A G G A C
A A C C G G G A A A A G A T C G A G A A G A T C C T G A C C T T C C G C A T C C C C T A C T A C G T G G G C C C T C T G G C C A G G G G A A A
C A G C A G A T T C G C C T G G A T G A C C A G A A A G A G C G A G G A A A C C A T C A C C C C T G G A A C T T C G A G G A A G T G G T G
G A C A A G G G C G C T T C C G C C C A G A G C T T C A T C G A G C G G A T G A C C A A C T T C G A T A A G A A C C T G C C C A A C G A G A A
G G T G C T G C C C A A G C A C A G C C T G C T G T A C G A G T A C T T C A C C G T G T A A C G A G C T G A C C A A A G T G A A A T A C G
T G A C C G A G G G A A T G A G A A A G C C C G C C T T C C T G A G C G G C G A G C A G A A A A G G C C A T C G T G G A C C T G C T G T T
C A A G A C C A A C C G G A A A G T G A C C G T G A A G C A G C T G A A A G A G G A C T A C T T C A A G A A A T C G A G T G C T T C G A C
T C C G T G G A A A T C T C C G G C G T G G A A G A T C G G T T C A A C G C C T C C C T G G G C A C A T A C C A C G A T C T G C T G A A A A T
T A T C A A G G A C A A G G A C T T C C T G G A C A A T G A G G A A A A C G A G G A C A T T C T G G A A G A T A T C G T G C T G A C C C T G
A C A C T G T T T G A G G A C A G A G A G A T G A T C G A G G A A C G G C T G A A A A C C T A T G C C C A C C T G T T C G A C G A C A A A G
T G A T G A A G C A G C T G A A G C G G C G G A G A T A C A C C G G C T G G G G C A G G C T G A G C C G G A A G C T G A T C A A C G G C A
T C C G G G A C A A G C A G T C C G G C A A G A C A A T C C T G G A T T T C C T G A A G T C C G A C G G C T T C G C C A A C A G A A A C T T C
A T G C A G C T G A T C C A C G A C G A C A G C C T G A C C T T T A A G A G A G A C A T C C A G A A A G C C C A G G T G T C C G G C C A G G
G C G A T A G C C T G C A C G A G C A C A T T G C C A A T C T G G C C G G C A G C C C C G C C A T T A A G A A G G G C A T C C T G C A G A C A
G T G A A G G T G G T G G A C G A G C T C G T G A A A G T G A T G G G C C G G C A C A A G C C C G A G A A C A T C G T G A T C G A A A T G
G C C A G A G A G A A C C A G A C C A C C C A G A A G G G A C A G A A G A A C A G C C G C G A G A G A A T G A A G C G G A T C G A A G A
G G G C A T C A A A G A G C T G G G C A G C C A G A T C C T G A A A G A A C A C C C C G T G G A A A A C A C C C A G C T G C A G A A C G A G
A A G C T G T A C C T G T A C C T G C A G A A T G G G C G G G A T A T G T A C G T G G A C C A G G A A C T G G A C A T C A A C C G G C T
G T C C G A C T A C G A T G T G G A C C A T A T C G T G C C T C A G A G C T T T C T G A A G G A C G A C T C C A T C G A C A A C A A G G T G C
T G A C C A G A A G C G A C A A G A A C C G G G G C A A G A G C G A C A A C G T G C C C T C C G A A G A G G T C G T G A A G A A G A T G A
A G A A C T A C T G G C G G C A G C T G C T G A A C G C C A A G C T G A T T A C C C A G A G A A A G T T C G A C A A T C T G A C C A A G G C C
G A G A G A G G C G G C C T G A G C G A A C T G G A T A A G G C C G G C T T C A T C A A G A G A C A G C T G G T G G A A A C C C G G C A G
A T C A C A A A G C A C G T G G C A C A G A T C C T G G A C T C C C G A T G A A C A C T A A G T A C G A C G A G A A T G A C A A G C T G A T
C C G G G A A G T G A A A G T G A T C A C C C T G A A G T C C A A G C T G G T G T C C G A T T T C C G G A A G G A T T T C A G T T T T A C A

AAGTGC GCGAGATCAACA ACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTG
 ATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGA
 TGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAAC
 TTTTTC AAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGA
 AACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCAA
 GTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGA
 ACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCAC
 CGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAGGGCAAGTCCAAGAACTGAAGAGTGTGAAA
 GAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAA
 GGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCTGAGCTGGAAAACG
 GCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATA
 TGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCGAGGATAATGAGCAGAAAC
 AGCTGTTTGTGGAACAGCACAAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGA
 GTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGGATAAGCCCATCAG
 AGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTT
 TGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAGAGGTGCTGGACGCCACCCTGATCCACCAG
 AGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGCGACAAAAGGCCGGCGGCCA
CGAAAAAGGCCGGCCAGGCCAAAAAAGAAAAAGGAATTCGGCAGTGGAGAGGGCAGAGGAAGTCTGCTAA
CATGCGGTGACGTCGAGGAGAATCCTGGCCAGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGC
CCATCCTGGTCTGAGCTGGACGGCGACGTAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCG
ATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCCGTGCCCTGGCCCACC
CTCGTGACCACCCTGACCTACGGCGTGCACTGCTTCAGCCGCTACCCCGACCATGAAGCAGCACGACTT
CTTCAAGTCCGCCATGCCCCAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTAC
AAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGAC
TTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACAGCCACAACGTCTATATC
ATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAAGTTCAGATCCGCCACAACATCGAGGACGGCAG
CGTGCACTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCCGACAAC
CACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGG
AGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGAATTCTAActagagctcgctgat
 cagcctcgactgtgccttctagttgccagccatctgtgtttgcccctccccgtgccttcctgaccctggaaggtgccactcccactgtcctttc
 ctaataaaatgaggaaattgcatcgattgtctgagtaggtgtcattctattctggggggtgggggtggggcaggacagcaagggggaggatt
 gggaagagaatagcaggcatgtctggggaGCGGCCGCAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCG
CGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTGCCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCA
GTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG (SEQ ID NO:67)

Construct p1526_pCbh-SpCas9(BB)-2A-GFP + 3' HUMAN albumin gRNA1

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGGAATT
 AATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTT
 GCAGTTTTTAAATTATGTTTTAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTGATTTCTTGG
 CTTTATATATCTTGTGGAAGGACGAAACACC (SEQ ID NO:59)

Human U6 promoter

Aatctctggacggaagctca (SEQ ID NO:10)

gRNA1 human albumin

GT TTTAGAGCTAGAAATAGCAAGTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC
GGTGCTTTTTT (SEQ ID NO:60)

Chimeric gRNA scaffold

Cgttacataacttacggtaaattggccgcctggctgaccgccaacgacccccgccattgacgtcaatagtaacgccaatagggactttcca
 ttgacgtcaatgggtggagattttacggtaaactgccacttggcagtagcatcaagtgtatcatatgccaagtacgcccctattgacgtcaat
 gacggtaaattggccgcctggcatttgcccagtagcatgaccttatgggactttctacttggcagtagcatctacgtattagtcacgtattac
 catggtcgaggtgagccccacgttctgcttactctcccatctccccccctccccacccaattttgtattatttttttaattattttgtgc
 agcgtatggggcgggggggggggggggcgcgccaggcgggcgggcgggcgaggggcgggcgggcgaggcgaggagaggtgc
 ggcggcagccaatcagagcggcgcgctccgaaagtttcttttatggcgaggcgggcgggcgggcgccctataaaaagcgaagcgcgcg
 cggcgggagtcgctgcgacgtccttcgccccgtccccgctccgcccgcctcgccgcccggcctgactgaccgcttact
 cccacaggtgagcggcggggacggcccttctctccgggtgtaattagctgagcaagaggttaagggtttaagggatggttggtgggg
 tattaatgtttaattacctggagcacctgctgaaatcacttttttcaggttg (SEQ ID NO:61)

CBH promoter

ATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAG
(SEQ ID NO:62)

3XFlag tag

ATGGCCCCAAAGAAGAAGCGGAAGGTCTGGTATCCACGGAGTCCCAGCAGCCGACAAGAAGTACAGCATCG
 GCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAA
 ATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTTCGACA
 GCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACC
 GGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCACAGACTG
 GAAGAGTCCTTCTGCTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACG
 AGGTGGCCTACCACGAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACAA
 GGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCACATGATCAAGTTCCGGGGCCACTTCTGATCGAGG
 GCGACCTGAACCCGACAACAGCGACGTGGACAAGCTGTTTCATCCAGCTGGTGCAGACCTACAACAGCTG
 TTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGA
 GCAGACGGCTGGAAAATCTGATCGCCAGCTGCCGGCGAGAAGAAGAATGGCCTGTTTCGGAACCTGAT
 TGCCCTGAGCCTGGGCCTGACCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAACTGCAGC
 TGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCAGATCGGCGACCAGTACGCCGACCT
 GTTTCTGGCCGCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCA
 CCAAGGCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACAGGACCTGACCCTGCTGAAA
 GCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGC
 CGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATG
 GACGGCACCGAGGAAGTCTGCTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGAC
 AACGGCAGCATCCCCACAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTTA
 CCCATTCTGAAGGACAACCGGGAAGATCGAGAAGATCCTGACCTTCGCATCCCCTACTACGTGGGCC
 CTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCTGGAA
 CTTTCGAGGAAGTGGTGGACAAGGGCGCTTCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAG
 AACCTGCCAACGAGAAGGTGCTGCCAACGACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCT
 GACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGC
 CATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAG
 AAAATCGAGTGCTTCGACTCCGTGGAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATA

CCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAA
 GATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCA
 CCTGTTGACGACAAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCG
 GAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCTGAAGTCCGACGGCT
 TCGCCAACAGAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCC
 CAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGA
 AGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGA
 ACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAA
 TGAAGCGGATCGAAGAGGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAACA
 CCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGA
 ACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACT
 CCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGA
 GGTCTGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTC
 GACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGC
 TGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGA
 CGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCTGAAGTCCAAGCTGGTGTCCGATTTCCGG
 AAGGATTTCCAGTTTTACAAAGTGC GCGAGATCAACAACCTACCACACGCCACGACGCCTACCTGAACGCC
 GTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGG
 TGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTT
 CTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTC
 TGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGA
 AAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGA
 GTCTATCCTGCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTAC
 GCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAGGGCAAGTCCA
 AGAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCC
 CATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACT
 CCCTGTTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGA
 ACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCC
 CGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAAGCACTACCTGGACGAGATCATCGAGCAG
 ATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAA
 GCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAG
 CCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAGAGGTGCTG
 GACGCCACCCTGATCCACCAGAGCATACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGG
 CGACAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCCAAAAAGAAAAAG (SEQ ID NO: 68)

5'Nuclear localization signal + SpCas9 + 3'Nuclear localization signal

GGCAGTGGAGAGGGCAGAGGAAGTCTGCTAACATGCGGTGACGTGAGGAGAATCCTGGCCCA (SEQ ID NO: 63)

Thosea Asigna Virus T2A skipping peptide

GTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCATCCTGGTCGAGCTGGACGGCGACGTAA
 CGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTT
 CATCTGCACCACCGGCAAGCTGCCCGTGGCCACCTCGTGACCACCTGACCTACGGCGTGCAGT
 GCTTCAGCCGCTACCCCGACCACATGAAGCAGCAGACTTCTTCAAGTCCGCCATGCCGAAGGCTACGTC
 CAGGAGCGCACCATCTTCTTCAAGGACGACGGCACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGC

GACACCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCAC
AAGCTGGAGTACAACACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAG
GTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAAC
ACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAA
AGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGC
ATGGACGAGCTGTACAAGGAATTCTAA

EGFP fusion protein

CTagagctcgtgatcagcctcgactgtgccttctagtgtccagccatctgttgtttgccctccccgtgccttccttgaccctggaaggtgcca
ctcccactgtcctttcctaataaaatgaggaaattgcatcgcatgtctgagtaggtgtcattctattctggggggtgggggtggggcaggacag
caaggggggaggattgggaagagaatagcaggcatgctgggga (SEQ ID NO:65)

BGH poliA

AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCA
AAGGTCGCCCCGACGCCCGGGCTTTGCCCCGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG
 (SEQ ID NO:66)

3'ITR

GAGGGCCTATTCCCATGATTCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGGAATT
AATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAAGTAATAATTTCTTGGGTAGTTT
GCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTGATTTCTTGG
CTTTATATATCTTGTGGAAGGACGAAACACCGaatctctggacggaagctcaGTTTTAGAGCTAGAAATAGCAAG
TTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTGTTTTAGAGCTAG
AAATAGCAAGTTAAAATAAGGCTAGTCCGTTTTTAGCGCGTGCGCCAATTCTGCAGACAAATGGCTCTAGA
GGTACCggttacataacttacggtaaatggcccgctggctgaccgccaacgaccccgccattgacgtcaatagtaacgccaatagg
gactttccattgacgtcaatgggtggagtatttacggttaaactgccacttggcagtagcatcaagtgtatcatatgccaagtacgccccctattg
acgtcaatgacggtaaatggcccgctggcattgtgccagtagcatgacctatgggactttcctacttggcagtagcatctacgtatttagtcatc
gctattaccatgggtcgaggtgagccccacgttctgcttcaactctccccatctccccccctccccaccccccaattttgtatttatttttttaatt
atthttgtgcagcgatggggggcgggggggggggggggggcgcgcgccaggcgggggcgggggcgggggcgaggggcgggggcgaggcggg
agaggtgcgggcgagccaatcagagcgggcgcgctccgaaaagtttcttttatggcgaggcgggcgggcgggcgccctataaaaagcgaa
cgcgcgggcgggcgggagtcgtgcgacgtgccttcgccccgtgccccgtccgcccgcgctcgcgccgccccgccccggctctgactgacc
gcgttactcccacaggtgagcgggcgggagcgcccttctcctccgggtgtaattagctgagcaagaggttaaggggttaagggatggttgggt
gggtgggggtattaatgtttaattacctggagcacctgcctgaaatcacttttttcaggttgACCGGTGCCACCATGGACTATAAGG
ACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAGATGGCCCCAAAGA
AGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGACAAGAAGTACAGCATCGGCCTGGACATCGG
CACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAAATTCAAGGTGCTG
GGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTTGACAGCGGCGAAACAG
CCGAGGCCACCCGGCTGAAGAGAACC GCCAGAAGAAGATACACCAGACGGAAGAACC GGATCTGCTATCT
GCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCC
TGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCA
CGAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACAAGGCCGACCTGCGG
CTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCGGGGGCCACTTCTGATCGAGGGCGACCTGAACCC
GACAACAGCGACGTGGACAAGCTGTTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTTCGAGGAAAACCC
CATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAA
AATCTGATCGCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGG

CCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAACTGCAGCTGAGCAAGGACACCT
ACGACGACGACCTGGACAACCTGCTGGCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAG
AACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAG
CGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGC
AGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGC
GGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAAC
TGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCA
CCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCTGAAGGACA
ACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGGCCCTCTGGCCAGGGGAAAC
AGCAGATTGCTGCTGATGACCAGAAAGAGCGAGGAAACCATCACCCCTGGAACCTTCGAGGAAGTGGTGG
ACAAGGGCGCTTCCGCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCAACGAGAAG
GTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGT
GACCGAGGGGAATGAGAAAAGCCCGCCTTCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTT
AAGACCAACCGGAAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACT
CCGTGGAAATCTCCGGCGTGGAAGATCGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATT
ATCAAGGACAAGGACTTCTGACAAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGA
CACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCACCTGTTTCGACGACAAAGT
GATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCAT
CCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCTGAAGTCCGACGGCTTCGCCAACAGAACTTCA
TGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCAGGTGTCCGGCCAGGG
CGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAG
TGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGG
CCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAG
GGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAACACCCAGCTGCAGAACGAGA
AGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACCTGGACATCAACCGGCT
GTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGC
TGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCTGTAAGAAGATGA
AGAATACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAAGTTTCGACAATCTGACCAAGGCC
GAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAG
ATCACAAGCAGCTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGAT
CCGGGAAGTGAAAGTGATCACCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACA
AAGTGCGCGAGATCAACAATAACACCACGCCCACGACGCCTACCTGAACGCCGTCTGTGGGAACCGCCCTG
ATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGA
TGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTACAGCAACATCATGAAC
TTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGA
AACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAA
GTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGA
ACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCAC
CGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAGGGCAAGTCCAAGAACTGAAGAGTGTGAAA
GAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCATCGACTTTCTGGAAGCCAA
GGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTTCGAGCTGGAAAACG
GCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACCTGGCCCTGCCCTCCAAATA
TGTGAACCTTCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAAC
AGCTGTTTGTGGAACAGCACAAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGA
GTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGGATAAGCCCATCAG
AGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTT
TGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAGAGGTGCTGGACGCCACCCTGATCCACCAG
AGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGCGACAAAAGGCCGGCGGCCA

CGAAAAAGGCCGCCAGGCCAAAAAAGAAAAAGGAATTCGGCAGTGGAGAGGGCAGAGGAAGTCTGCTAA
CATGCGGTGACGTCGAGGAGAATCCTGGCCAGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGC
CCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCG
ATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCCTGGCCCTGGCCACC
CTCGTGACCACCCTGACCTACGGCGTGCACTGCTTCAGCCGCTACCCCGACCCACATGAAGCAGCAGCACTT
CTTCAAGTCCGCCATGCCCAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTAC
AAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGAC
TTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACAGCCACAACGTCTATATC
ATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACCTCAAGATCCGCCACAACATCGAGGACGGCAG
CGTGCACTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCCGACAAC
CACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGG
AGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGAATTCTAActagagctcgctgat
cagcctcgactgtgccttctagttgccagccatctgtgtttgcccctccccgtgccttccttgaccctggaaggtgccactcccactgtcctttc
ctaataaaatgaggaaattgcatgcattgtctgagtaggtgtcattctattctggggggtggggtggggcaggacagcaagggggaggatt
gggaagagaatagcaggcatgtggggaGCGGCCGCAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCG
CGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTGCGCCGACGCCCCGGGCTTTGCCCGGGCGGCCTCA
GTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG (SEQ ID NO: 69)

Construct p1530_pCbh-SpCas9(BB)-2A-GFP + 3' HUMAN albumin gRNA2

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGGAATT
AATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTT
GCAGTTTTTAAATTATGTTTTTAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTTCGATTTCTTGG
CTTTATATATCTTGTGGAAAGGACGAAACACC (SEQ ID NO:59)

Human U6 promoter

Acagtatggcacaatagagc (SEQ ID NO:12)

gRNA2 human albumin

GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC
GGTGCTTTTTT (SEQ ID NO:60)

Chimeric gRNA scaffold

cgttacataacttacggtaaatggcccgctggctgaccgccaacgacccccgccattgacgtcaatagtaacgccaatagggactttcca
ttgacgtcaatgggtggagtatttacggtaaatgcccacttggcagtagcatcaagtgtatcatatgccaagtacgccccctattgacgtcaat
gacggtaaatggcccgctggcattgtgcccagtagcatgaccttatgggactttcctacttggcagtagcatctacgtattagtcacgtattac
catggtcgaggtgagccccacgttctgcttactctccccatctccccccccctcccccccccaattttgtattttatttttttaattattttgtgc
agcgatggggggcgggggggggggggggggcgcgcgccaggcgggggcgggggcgagggggcgggggcgagggcgagagaggtgc
ggcggcagccaatcagagcgggcgcgctccgaaagtttccttttatggcgaggcgggcgggcgggcgccctataaaaaagcgaagcgcgcg
cgggcgggagtcgctgcgacgctgccttcgccccgtgccccgctccgcccgcgctcgcgccgccccggccttgactgaccgcgttact
cccacaggtgagcgggcgggacggcccttctctccgggctgtaattagctgagcaagaggttaagggttaagggttggttggtgggg
tattaatgtttaattacctggagcacctgcctgaaatcacttttttcaggttgg

CBH promoter

**ATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAG
(SEQ ID NO:62)**

3XFlag tag

ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGACAAGAAGTACAGCATCG
GCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAA
ATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTTCGACA
GCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACC
GGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCACAGACTG
GAAGAGTCCTTCCTGGTGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACG
AGGTGGCCTACCACGAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACAA
GGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCACATGATCAAGTTCCGGGGCCACTTCTGATCGAGG
GCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTTCATCCAGCTGGTGCAGACCTACAACCAGCTG
TTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGA
GCAGACGGCTGGAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTTCGGAACCTGAT
TGCCCTGAGCCTGGGCTGACCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAACTGCAGC
TGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCAGATCGGCGACCAGTACGCCGACCT

GTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCA
CCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAA
GCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGC
CGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATG
GACGGCACCGAGGAAGTCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGAC
AACGGCAGCATCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTT
CCATTCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCC
CTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCTGGAA
CTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAG
AACCTGCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCT
GACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCTTCTGAGCGGCGAGCAGAAAAAGGC
CATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAG
AAAATCGAGTGCTTCGACTCCGTGGAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATA
CCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCTGGACAATGAGGAAAACGAGGACATTCTGGAA
GATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCA
CCTGTTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCG
GAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCTGAAGTCCGACGGCT
TCGCCAACAGAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCC
CAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGA
AGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAGTGATGGGCCGGCACAAGCCCGAGA
ACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAA
TGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAACA
CCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGA
ACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACT
CCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCAGGGGCAAGAGCGACAACGTGCCCTCCGAAGA
GGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTC
GACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGC
TGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGA
CGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCTGAAGTCCAAGCTGGTGTCCGATTTCCGG
AAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACCTACCACCACGCCACGACGCCTACCTGAACGCC
GTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGG
TGACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTT
CTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTC
TGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGA
AAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGA
GTCTATCCTGCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTAC
GGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCA
AGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCC
CATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACT
CCCTGTTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGA
ACTGGCCCTGCCCTCAAATATGTGAACTTCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCC
CGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAAGCACTACCTGGACGAGATCATCGAGCAG
ATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACA
GCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAG
CCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAGAGGTGCTG
GACGCCACCCTGATCCACCAGAGCATACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGG
CGACAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAG (SEQ ID NO: 70)

5'Nuclear localization signal + SpCas9 + 3'Nuclear localization signal

GGCAGTGGAGAGGGCAGAGGAAGTCTGCTAACATGCGGTGACGTGAGGAGAATCCTGGCCCA (SEQ ID NO: 63)

Thosea Asigna Virus T2A skipping peptide

GTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAA
CGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTT
CATCTGCACCACCGGCAAGCTGCCCCTGCCCTGGCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGT
GCTTCAGCCGCTACCCCGACCACATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGC
GACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCAC
AAGCTGGAGTACAACACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAG
GTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGACGCTCGCCGACCACTACCAGCAGAAC
ACCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAA
AGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGC
ATGGACGAGCTGTACAAGGAATTCTAA

EGFP fusion protein

Ctagagctcgctgatcagcctcgactgtgccttctagttgccagccatctgtgtttgcccctccccctgccttcttgaccctggaaggtgcc
ctcccactgtcctttcctaataaaatgaggaaatgcatcgattgtctgagtaggtgtcattctattctgggggtgggggtggggcaggacag
caagggggaggattgggaagagaatagcaggcatgctgggga (SEQ ID NO:65)

BGH poliA

AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCA
AAGGTCGCCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG
(SEQ ID NO:66)

3'ITR

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGGAATT
AATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTT
GCAGTTTTAAATATGTTTTAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTGATTTCTTGG
CTTTATATATCTTGTGGAAGGACGAAACACCGacagtatggcacaatagagcGTTTTAGAGCTAGAAATAGCAA
GTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTGTTTAGAGCTA
GAAATAGCAAGTTAAATAAGGCTAGTCCGTTTTTAGCGCGTGCGCCAATTCTGCAGACAAATGGCTCTAG
AGGTACCggttacataacttacggtaaatggccgcctggctgaccgccaacgacccccgccattgacgtcaatagtaacgccaatag
ggactttcattgacgtcaatgggtggagtatttacggtaaactgcccacttggcagtagcatcaagtgtatcatatgccaagtacgccccatt
gacgtcaatgacggtaaatggccgcctggcattgtgcccagtagcatgacctatgggactttcctacttggcagtagcatctacgtattgcat
cgctattaccatggctgaggtgagcccccacgttctgcttacttccccatctccccccccctccccaccccccaattttgtattttatttttaatt
tattttgtgcagcgatggggcgggggggggggggggcgcgccaggcgggcgggcgggcgagggcgggcgggcgagggcg
gagaggtgcgggcgagccaatcagagcgggcgcgctccgaaagtcttctttatggcgaggcgggcgggcgggcgccctataaaaagcga
agcgcgggcgggcgggagtcgctgacgctgccttcgccccgtgccccgctccgcccgcctcgcgcccgccccggcctgactga
ccgcttactccacaggtgagcgggcgggacggcccttctcctccgggctgtaattagctgagcaagaggttaagggttggttg
ttggtggggtattaatgtttaattacctggagcacctgcctgaaatcacttttttcaggttgACCGGTGCCACCATGGACTATAAG

GACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAGATGGCCCCAAAG
AAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGACAAGAAGTACAGCATCGGCCTGGACATCG
GCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAAATTCAGGTGCT
GGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTTCGACAGCGGCGAAACA
GCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATC
TGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTC
CTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGACGAGGTGGCCTACC
ACGAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACAAGGCCGACCTGCG
GCTGATCTATCTGGCCCTGGCCCATGATCAAGTTCGGGGGCCACTTCTGATCGAGGGCGACCTGAACCC
CGACAACAGCGACGTGGACAAGCTGTTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACC
CCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGA
AAATCTGATCGCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTTCGAAACCTGATTGCCCTGAGCCTGG
GCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAACTGCAGCTGAGCAAGGACACC
TACGACGACGACCTGGACAACCTGCTGGCCAGATCGGCGACCAAGTACGCCGACCTGTTTCTGGCCGCCAA
GAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATACCAAGGCCCCCTGA
GCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAG
CAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGG
CGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAA
CTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCC
ACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTACCCATTCTGAAGGAC
AACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAA
CAGCAGATTTCGCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCTGGAACCTCGAGGAAGTGGTG
GACAAGGGCGCTTCGCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCAACGAGAA
GGTGCTGCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAGTGAAATACG
TGACCGAGGGAATGAGAAAGCCCGCCTTCTGAGCGGCGAGCAGAAAAAGGCCATCGTGACCTGCTGTT
CAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGAC
TCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAT
TATCAAGGACAAGGACTTCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTG
ACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCACCTGTTTCGACGACAAAG
TGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCA
TCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAACTTC
ATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGG
GCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACA
GTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCGGCACAAGCCCGAGAACATCGTGATCGAAATG
GCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGA
GGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAACACCCAGCTGCAGAACGAG
AAGCTGTACCTGTACTACCTGCAGAAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCT
GTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGC
TGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGA
AGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTTCGACAATCTGACCAAGGCC
GAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAG
ATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGAT
CCGGGAAGTGAAAGTGATCACCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACA
AAGTGCGCGAGATCAACAACCTACCACCACGCCCACGACGCCTACCTGAACGCCGTCTGTGGGAACCGCCCTG
ATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGA
TGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTACAGCAACATCATGAAC
TTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGA
AACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAA

GTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGA
ACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCAC
CGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGAAAAAGGGCAAGTCCAAGAACTGAAGAGTGTGAAA
GAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAA
GGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTTCGAGCTGGAAAACG
GCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATA
TGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCGAGGATAATGAGCAGAAAC
AGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGA
GTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGGATAAGCCCATCAG
AGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTT
TGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAGAGGTGCTGGACGCCACCCTGATCCACCAG
AGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGCGACAAAAGGCCGGCGGCCA
CGAAAAAGGCCGGCCAGGCCAAAAAAGAAAAAGGAATTCGGCAGTGGAGAGGGCAGAGGAAGTCTGCTAA
CATGCGGTGACGTCGAGGAGAATCCTGGCCAGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGC
CCATCCTGGTCGAGCTGGACGGCGACGTAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCG
ATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCCGTGCCCTGGCCACC
CTCGTGACCACCCTGACCTACGGCGTGAGTGCTTACGCCGCTACCCGACCACATGAAGCAGCACGACTT
CTTCAAGTCCGCCATGCCCCAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTAC
AAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGAC
TTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACAGCCACAACGTCTATATC
ATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAG
CGTGCACTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCCGACAAC
CACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGG
AGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGAATTCTAActagagctcgctgat
cagcctcgactgtgccttctagttgccagccatctgtgtttgcccctccccctgccttccttgaccctggaaggtgccactcccactgtccttcc
ctaataaaatgaggaaattgcatcgattgtctgagtaggtgtcattctattctggggggtgggggtggggcaggacagcaagggggaggatt
gggaagagaatagcaggcatgctggggaGCGGCCGCAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCG
CGCTCGCTCGCTCACTGAGGCCGGGCGACCAAGGTGCCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCA
GTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG (SEQ ID NO:71)

Construct p1531_pCbh-SpCas9(BB)-2A-GFP + 3' HUMAN albumin gRNA3

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGGAATT
AATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTT
GCAGTTTTTAAATTATGTTTTTAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTTCGATTTCTTGG
CTTTATATATCTTGTGGAAAGGACGAAACACC (SEQ ID NO:59)

Human U6 promoter

Acactacataacgtgatgag (SEQ ID NO:13)

gRNA3 human albumin

GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC
GGTGCTTTTTT (SEQ ID NO:60)

Chimeric gRNA scaffold

cgttacataacttacggtaaatggcccgctggctgaccgccaacgacccccgccattgacgtcaatagtaacgccaatagggactttcca
ttgacgtcaatgggtggagtatttacggtaaatgcccacttggcagtagcatcaagtgtatcatatgccaagtacgccccctattgacgtcaat
gacggtaaatggcccgctggcattgtgcccagtagcatgaccttatgggactttcctacttggcagtagcatctacgtattagtcacgtattac
catggtcgaggtgagccccacgttctgcttactctccccatctccccccctccccaccccccaattttgtattttattttttaattattttgtgc
agcgatgggggcgggggggggggggggggcgcgcgccaggcgggggcgggggcgagggggcgggggcgagggcgagagaggtgc
ggcggcagccaatcagagcgggcgcgctccgaaagtttccttttatggcgaggcgggcgggcgggcgccctataaaaaagcgaagcgcgcg
cgggcgggagtcgctgcgacgctgccttcgccccgtgccccgctccgcccgcgctcgcgccccgccccggccttgactgaccgcgttact
cccacaggtgagcgggcgggagggcccttctctccgggctgtaattagctgagcaagaggtaagggtttaagggatggttggttggtggg
tattaatgtttaattacctggagcacctgcctgaaatcacttttttcaggttgg

CBH promoter

**ATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAG
(SEQ ID NO:62)**

3XFlag tag

ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGACAAGAAGTACAGCATCG
GCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAA
ATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTTCGACA
GCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACC
GGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCACAGACTG
GAAGAGTCCTTCCTGGTGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACG
AGGTGGCCTACCACGAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACAA
GGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCACATGATCAAGTTCCGGGGCCACTTCTGATCGAGG
GCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTTCATCCAGCTGGTGCAGACCTACAACCAGCTG
TTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGA
GCAGACGGCTGGAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTTCGGAACCTGAT
TGCCCTGAGCCTGGGCTGACCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAACTGCAGC
TGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCAGATCGGCGACCAGTACGCCGACCT

GTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCA
CCAAGGCCCCCCTGAGCGCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAA
GCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGC
CGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATG
GACGGCACCGAGGAAGTCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGAC
AACGGCAGCATCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTT
CCATTCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCC
CTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCTGGAA
CTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAG
AACCTGCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCT
GACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCTTCTGAGCGGCGAGCAGAAAAAGGC
CATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAG
AAAATCGAGTGCTTCGACTCCGTGGAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATA
CCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCTGGACAATGAGGAAAACGAGGACATTCTGGAA
GATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCA
CCTGTTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCG
GAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCTGAAGTCCGACGGCT
TCGCCAACAGAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCC
CAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGA
AGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAGTGATGGGCGGCACAAGCCCGAGA
ACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAAGAGCCGCGAGAGAA
TGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAACA
CCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGA
ACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACT
CCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCAGGGGCAAGAGCGACAACGTGCCCTCCGAAGA
GGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTC
GACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCGCGCTTCATCAAGAGACAGC
TGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGA
CGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCTGAAGTCCAAGCTGGTGTCCGATTTCCGG
AAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACCTACCACCACGCCACGACGCCTACCTGAACGCC
GTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGG
TGACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTT
CTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTC
TGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGA
AAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGA
GTCTATCCTGCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTAC
GGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCA
AGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCC
CATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACT
CCCTGTTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGA
ACTGGCCCTGCCCTCAAATATGTGAACTTCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCC
CGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAAGCACTACCTGGACGAGATCATCGAGCAG
ATCAGCGAGTTCTCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAA
GCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAG
CCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAGAGGTGCTG
GACGCCACCCTGATCCACCAGAGCATACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGG
CGACAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAG (SEQ ID NO: 72)

5'Nuclear localization signal + SpCas9 + 3'Nuclear localization signal

GGCAGTGGAGAGGGCAGAGGAAGTCTGCTAACATGCGGTGACGTGAGGAGAATCCTGGCCCA (SEQ ID NO: 63)

Thosea Asigna Virus T2A skipping peptide

GTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAA
CGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTT
CATCTGCACCACCGGCAAGCTGCCCCTGCCCTGGCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGT
GCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGC
GACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCAC
AAGCTGGAGTACAACACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAG
GTGAACCTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGACGCTCGCCGACCACTACCAGCAGAAC
ACCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAA
AGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGC
ATGGACGAGCTGTACAAGGAATTCTAA

EGFP fusion protein

Ctagagctcgctgatcagcctcgactgtgccttctagttgccagccatctgtgtttgcccctccccctgccttcttgaccctggaaggtgcc
ctcccactgtcctttcctaataaaatgaggaaatgcatcgattgtctgagtaggtgtcattctattctgggggtgggggtggggcaggacag
caagggggaggattgggaagagaatagcaggcatgctgggga (SEQ ID NO:65)

BGH poliA

AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCA
AAGGTCGCCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG
(SEQ ID NO:66)

3'ITR

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGGAATT
AATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTT
GCAGTTTTAAATATGTTTTAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTGATTTCTTGG
CTTTATATATCTTGTGGAAGGACGAAACACCGacactacataacgtgatgagGTTTTAGAGCTAGAAATAGCAAG
TTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTGTTTTAGAGCTAG
AAATAGCAAGTTAAATAAGGCTAGTCCGTTTTTAGCGCGTGCGCCAATTCTGCAGACAAATGGCTCTAGA
GGTACCggttacataacttacggtaaatggcccgctggctgaccgccaacgacccccgcccattgacgtcaatagtaacccaatagg
gactttccattgacgtcaatgggtggagtatttacggtaaactgcccacttggcagtagcatcaagtgtatcatatgccaaagtacgccccctattg
acgtcaatgacggtaaatggcccgctggcattgtgcccagtagacattatgggactttcctacttggcagtagcatctacgtattagtcac
gctattaccatggctgaggtgagccccacgttctgcttactctccccatctccccccctccccaccccccaattttgtattttatttttaatt
attttgtcagcgatggggcgggggggggggggggcgcgccaggcgggcgggcgggcgaggggcgggcgggcgaggcgg
agaggtgcgggcgagccaatcagagcgggcgcgctccgaaagtttcctttatggcgaggcgggcgggcgggcgccctataaaaagcgaa
gcgcgcgggcgggagtcgctgacgctgccttcgccccgtgccccgctccgcccgcctcgccgcccgggctctgactgacc
gcgttactccacaggtgagcgggcgggacggcccttctcctcgggctgtaattagctgagcaagaggttaagggttaagggtggttggtt
gggtgggtattaatgtttaattacctggagcacctgcctgaaatcacttttttcagggttgACCGGTGCCACCATGGACTATAAGG

ACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAGATGGCCCCAAAGA
AGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGACAAGAAGTACAGCATCGGCCTGGACATCGG
CACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAAATTCAAGGTGCTG
GGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAG
CCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCT
GCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCC
TGGTGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCA
CGAGAAGTACCCACCATCTACCACTGAGAAAGAACTGGTGGACAGCACCGACAAGGCCGACCTGCGG
CTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCGGGGGCCACTTCTGATCGAGGGCGACCTGAACCCC
GACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCC
CATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAA
AATCTGATCGCCCAGCTGCCCCGGCGAGAAGAAGAATGGCCTGTTGGAACCTGATTGCCCTGAGCCTGGG
CCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAACTGCAGCTGAGCAAGGACACCT
ACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCACTACGCCGACCTGTTTCTGGCCGCCAAG
AACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAGGCCCCCTGAG
CGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGC
AGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGC
GGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGGCCATCCTGGAAAAGATGGACGGCACCGAGGAAC
TGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCA
CCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTACCCATTCTGAAGGACA
ACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAAC
AGCAGATTGCTCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCTGGAACCTTCGAGGAAGTGGTGG
ACAAGGGCGCTTCCGCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCAACGAGAAG
GTGCTGCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGT
GACCGAGGGAATGAGAAAGCCCGCCTTCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTC
AAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACT
CCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATT
ATCAAGGACAAGGACTTCTTGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGA
CACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCACCTGTTCGACGACAAAGT
GATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCAT
CCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCTGAAGTCCGACGGCTTCGCCAACAGAACTTCA
TGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGG
CGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAG
TGAAGGTGGTGGACGAGCTCGTGAAGTGATGGGCCGGCACAAGCCCCGAGAACATCGTGATCGAAATGG
CCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAG
GGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGA
AGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACCTGGACATCAACCGGCT
GTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGC
TGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCTGTAAGAAGATGA
AGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTGACAATCTGACCAAGGCC
GAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAG
ATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGAT
CCGGGAAGTGAAAGTGATCACCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACA
AAGTGCGCGAGATCAACAACCTACCACCACGCCCACGACGCCTACCTGAACGCCGTCTGTTGGAACCGCCCTG
ATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGA
TGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTACAGCAACATCATGAAC
TTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGA
AACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAA

GTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGA
ACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCAC
CGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAACTGAAGAGTGTGAAA
GAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAA
GGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTTCGAGCTGGAAAACG
GCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATA
TGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCGAGGATAATGAGCAGAAAC
AGCTGTTTGTGGAACAGCACAAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGA
GTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGGATAAGCCCATCAG
AGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTT
TGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAGAGGTGCTGGACGCCACCCTGATCCACCAG
AGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGCGACAAAAGGCCGGCGGCCA
CGAAAAAGGCCGGCCAGGCCAAAAAAGAAAAAGGAATTCGGCAGTGGAGAGGGCAGAGGAAGTCTGCTAA
CATGCGGTGACGTCGAGGAGAATCCTGGCCAGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGC
CCATCCTGGTCGAGCTGGACGGCGACGTAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCG
ATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCCGTGCCCTGGCCCACC
CTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTT
CTTCAAGTCCGCCATGCCCCAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTAC
AAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGAC
TTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACAGCCACAACGTCTATATC
ATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAG
CGTGCACTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCCGACAAC
CACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGG
AGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGAATTCTAActagagctcgctgat
cagcctcgactgtgccttctagttgccagccatctgtgtttgcccctccccctgccttccttgaccctggaaggtgccactcccactgtccttc
ctaataaaatgaggaaattgcatcgattgtctgagtaggtgtcattctattctggggggtgggggtggggcaggacagcaagggggaggatt
gggaagagaatagcaggcatgctggggaGCGGCCGCAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCG
CGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTGCCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCA
GTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG (SEQ ID NO:73)

Construct p1532_pCbh-SpCas9(BB)-2A-GFP + 3' HUMAN albumin gRNA4

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGGAATT
AATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTT
GCAGTTTTTAAATTATGTTTTTAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTTCGATTTCTTGG
CTTTATATATCTTGTGGAAAGGACGAAACACC (SEQ ID NO:59)

Human U6 promoter

Aaatagtttagaatagtggt (SEQ ID NO:14)

gRNA4 human albumin

GTTTATAGAGCTAGAAATAGCAAGTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC
GGTGCTTTTTT (SEQ ID NO:60)

Chimeric gRNA scaffold

cgttacataacttacggtaaatggcccgctggctgaccgccaacgacccccgccattgacgtcaatagtaacgccaatagggactttcca
ttgacgtcaatgggtggagtatttacggtaaatgcccacttggcagtagcatcaagtgtatcatatgccaagtacgccccctattgacgtcaat
gacggtaaatggcccgctggcattgtgcccagtagcatgaccttatgggactttcctacttggcagtagcatctacgtattagtcacgtattac
catggtcgaggtgagccccacgttctgcttactctccccatctccccccctccccaccccccaattttgtattttatttttttaattattttgtgc
agcgatggggggcggggggggggggggggcgcgcgccaggcggggcgggggcgaggggcgggggcgggggcgaggcgagaggtgc
ggcgccagccaatcagagcgggcgcgctccgaaagtttccttttatggcgaggcgggcgggcgggcgccctataaaaaagcgaagcgcgcg
cgggcgggagtcgctgcgacgctgccttcgccccgtgccccgctccgcccgcgctcgcgccgccccggccttgactgaccgcgttact
cccacaggtgagcgggcgggacggcccttctctccgggctgtaattagctgagcaagaggttaagggttaagggttggttggtgggg
tattaatgtttaattacctggagcacctgcctgaaatcacttttttcaggttgg

CBH promoter

**ATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAG
(SEQ ID NO:62)**

3XFlag tag

ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGACAAGAAGTACAGCATCG
GCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAA
ATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTTCGACA
GCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACC
GGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCACAGACTG
GAAGAGTCCTTCCTGGTGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACG
AGGTGGCCTACCACGAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACAA
GGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCACATGATCAAGTTCCGGGGCCACTTCTGATCGAGG
GCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTTCATCCAGCTGGTGCAGACCTACAACCAGCTG
TTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGA
GCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTTCGGAACCTGAT
TGCCCTGAGCCTGGGCTGACCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAACTGCAGC
TGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCAGATCGGCGACCAAGTACGCCGACCT

GTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCA
CCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAA
GCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGC
CGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATG
GACGGCACCGAGGAAGTCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGAC
AACGGCAGCATCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTT
CCATTCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCC
CTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCTGGAA
CTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAG
AACCTGCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCT
GACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCTTCTGAGCGGCGAGCAGAAAAAGGC
CATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAG
AAAATCGAGTGCTTCGACTCCGTGGAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATA
CCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCTGGACAATGAGGAAAACGAGGACATTCTGGAA
GATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCA
CCTGTTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCG
GAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCTGAAGTCCGACGGCT
TCGCCAACAGAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCC
CAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGA
AGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAGTGATGGGCCGGCACAAGCCCGAGA
ACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAA
TGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAACA
CCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGA
ACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACT
CCATCGACAACAAGTGCTGACCAGAAGCGACAAGAACCAGGGGCAAGAGCGACAACGTGCCCTCCGAAGA
GGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTC
GACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGC
TGGTGGAAACCCGGCAGATCACAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGA
CGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCTGAAGTCCAAGCTGGTGTCCGATTTCCGG
AAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACCTACCACCACGCCACGACGCCTACCTGAACGCC
GTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGG
TGACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTT
CTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTC
TGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGA
AAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGA
GTCTATCCTGCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTAC
GGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCA
AGAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCC
CATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACT
CCCTGTTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGA
ACTGGCCCTGCCCTCCAAATATGTGAACTTCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCC
CGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAAGCACTACCTGGACGAGATCATCGAGCAG
ATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACA
GCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAG
CCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAGAGGTGCTG
GACGCCACCCTGATCCACCAGAGCATACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGG
CGACAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAG (SEQ ID NO:109)

5'Nuclear localization signal + SpCas9 + 3'Nuclear localization signal

GGCAGTGGAGAGGGCAGAGGAAGTCTGCTAACATGCGGTGACGTCGAGGAGAATCCTGGCCCA (SEQ ID NO: 63)

Thosea Asigna Virus T2A skipping peptide

GTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAA
CGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTT
CATCTGCACCACCGGCAAGCTGCCCCTGCCCTGGCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGT
GCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGC
GACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCAC
AAGCTGGAGTACAACACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAG
GTGAACCTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGACGCTCGCCGACCACTACCAGCAGAAC
ACCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCAGTCCGCCCTGAGCAA
AGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGC
ATGGACGAGCTGTACAAGGAATTCTAA

EGFP fusion protein

Ctagagctcgctgatcagcctcgactgtgccttctagttgccagccatctgtgtttgcccctccccctgccttcttgaccctggaaggtgcc
ctcccactgtcctttcctaataaaatgaggaaatgcatcgattgtctgagtaggtgtcattctattctgggggtgggggtggggcaggacag
caagggggaggattgggaagagaatagcaggcatgctgggga (SEQ ID NO:65)

BGH poliA

AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCA
AAGGTCGCCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG
(SEQ ID NO:66)

3'ITR

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGGAATT
AATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTT
GCAGTTTTAAATATGTTTTAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTGATTTCTTGG
CTTTATATATCTTGTGGAAGGACGAAACACCGaaatagtttagaatagtggtGTTTTAGAGCTAGAAATAGCAAG
TTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTGTGTTAGAGCTAG
AAATAGCAAGTTAAATAAGGCTAGTCCGTTTTTAGCGCGTGCGCCAATTCTGCAGACAAATGGCTCTAGA
GGTACCggttacataacttacggtaaatggcccgctggctgaccgccaacgacccccgcccattgacgtcaatagtaacccaatagg
gactttccattgacgtcaatgggtggagtatttacggtaaactgcccacttggcagtagcatcaagtgtatcatatgccaaagtagccccctattg
acgtcaatgacggtaaatggcccgctggcattgtgcccagtagacattatgggactttcctacttggcagtagcatctacgtattagtcac
gctattaccatggctgaggtgagccccacgttctgcttactctccccatctccccccctccccaccccccaattttgtatttttttttaatt
atthttgtcagcgatggggcgggggggggggggggcgcgccaggcgggcgggcgggcgaggggcgggcgggcgaggcg
agaggtgaggcgagccaatcagagcgggcgcgctccgaaagtctcttttatggcgaggcgggcgggcgggcgccctataaaaagcgaa
gcgcgcgggcggggagtcgctgacgctgccttcgccccgtgccccgctccgcccgcctcgccgcccgggctctgactgacc
gcgttactccacaggtgagcgggcgggacggcccttctcctcgggctgtaattagctgagcaagaggttaagggttaagggtggtggtt
gggtgggtattaatgtttaattacctggagcacctgcctgaaatcacttttttcagggttgACCGGTGCCACCATGGACTATAAGG

ACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAGATGGCCCCAAAGA
AGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGACAAGAAGTACAGCATCGGCCTGGACATCGG
CACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAAATTCAAGGTGCTG
GGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAG
CCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCT
GCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTCCACAGACTGGAAGAGTCCTTCC
TGGTGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCA
CGAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACAAGGCCGACCTGCGG
CTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCGGGGGCCACTTCTGATCGAGGGCGACCTGAACCCC
GACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCC
CATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAA
AATCTGATCGCCCAGCTGCCCCGGCGAGAAGAAGAATGGCCTGTTGGAACCTGATTGCCCTGAGCCTGGG
CCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAACTGCAGCTGAGCAAGGACACCT
ACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCACTACGCCGACCTGTTTCTGGCCGCCAAG
AACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAGGCCCCCTGAG
CGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGC
AGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGC
GGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGGCCATCCTGGAAAAGATGGACGGCACCGAGGAAC
TGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCA
CCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTACCCATTCTGAAGGACA
ACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAAC
AGCAGATTGCTGCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCTGGAACCTTCGAGGAAGTGGTGG
ACAAGGGCGCTTCCGCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCAACGAGAAG
GTGCTGCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGT
GACCGAGGGAATGAGAAAGCCCGCCTTCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTC
AAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACT
CCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATT
ATCAAGGACAAGGACTTCTTGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGA
CACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCACCTGTTCGACGACAAAGT
GATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCAT
CCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCTGAAGTCCGACGGCTTCGCCAACAGAACTTCA
TGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGG
CGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAG
TGAAGGTGGTGGACGAGCTCGTGAAGTGATGGGCCGGCACAAGCCCCGAGAACATCGTGATCGAAATGG
CCAGAGAGAACCAGACCCAGAAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAG
GGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAACACCCAGCTGCAGAACGAGA
AGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACCTGGACATCAACCGGCT
GTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGC
TGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCTGTAAGAAGATGA
AGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTGACAATCTGACCAAGGCC
GAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAG
ATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGAT
CCGGGAAGTGAAAGTGATCACCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACA
AAGTGCGCGAGATCAACAACCTACCACCACGCCCACGACGCCTACCTGAACGCCGTCTGTTGGAACCGCCCTG
ATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGA
TGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAAC
TTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGA
AACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAA

GTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGA
ACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCAC
CGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGAAAAAGGGCAAGTCCAAGAACTGAAGAGTGTGAAA
GAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAA
GGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTTCGAGCTGGAAAACG
GCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATA
TGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCGAGGATAATGAGCAGAAAC
AGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGA
GTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGGATAAGCCCATCAG
AGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTT
TGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAGAGGTGCTGGACGCCACCCTGATCCACCAG
AGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGCGACAAAAGGCCGGCGGCCA
CGAAAAAGGCCGGCCAGGCCAAAAAAGAAAAAGGAATTCGGCAGTGGAGAGGGCAGAGGAAGTCTGCTAA
CATGCGGTGACGTCGAGGAGAATCCTGGCCAGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGC
CCATCCTGGTCGAGCTGGACGGCGACGTAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCG
ATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCCTGCCCTGGCCACC
CTCGTGACCACCCTGACCTACGGCGTGCACTGCTTACGCCGCTACCCCGACCATGAAGCAGCACGACTT
CTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTAC
AAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGAC
TTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACAGCCACAACGTCTATATC
ATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAG
CGTGCACTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAAC
CACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGG
AGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGAATTCTAActagagctcgctgat
cagcctcgactgtgccttctagttgccagccatctgtgtttgcccctccccctgccttccttgaccctggaaggtgccactcccactgtccttcc
ctaataaaatgaggaaattgcatcgattgtctgagtaggtgtcattctattctggggggtgggggtggggcaggacagcaagggggaggatt
gggaagagaatagcaggcatgctggggaGCGGCCGCAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCG
CGCTCGCTCGCTCACTGAGGCCGGGCGACCAAGGTGCCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCA
GTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG (SEQ ID NO: 74)

Construct p1556_pCbh-SpCas9(BB)-2A-GFP + 3' HUMAN albumin gRNA5

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGGAATT
AATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTT
GCAGTTTTTAAATTATGTTTTTAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTTCGATTTCTTGG
CTTTATATATCTTGTGGAAAGGACGAAACACC (SEQ ID NO:59)

Human U6 promoter

Gtgggctgtaatcatcgtct (SEQ ID NO:16)

gRNA5 human albumin

GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC
GGTGCTTTTTT (SEQ ID NO:60)

Chimeric gRNA scaffold

cgttacataacttacggtaaatggcccgctggctgaccgccaacgacccccgccattgacgtcaatagtaacgccaatagggactttcca
ttgacgtcaatgggtggagtatttacggtaaatgcccacttggcagtagcatcaagtgtatcatatgccaagtacgccccctattgacgtcaat
gacggtaaatggcccgctggcattgtgcccagtagcatgaccttatgggactttcctacttggcagtagcatctacgtattagtcacgtattac
catggtcgaggtgagccccacgttctgcttactctccccatctccccccctccccaccccccaattttgtattttatttttttaattattttgtgc
agcgatgggggcgggggggggggggggggcgcgcgccaggcgggggcgggggcgaggggcgggggcgggggcgaggcgagaggtgc
ggcggcagccaatcagagcgggcgcgctccgaaagtttccttttatggcgaggcgggcgggcgggcgccctataaaaaagcgaagcgcgcg
cgggcgggagtcgctgcgacgctgccttcgccccgtgccccgctccgcccgcctcgcgccgccccggcctctgactgaccgcgttact
cccacaggtgagcgggcgggacggcccttctctccgggctgtaattagctgagcaagaggttaagggttaagggttggttggtgggg
tattaatgtttaattacctggagcacctgcctgaaatcacttttttcaggttgg

CBH promoter

**ATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAG
(SEQ ID NO:62)**

3XFlag tag

ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGACAAGAAGTACAGCATCG
GCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAA
ATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTTCGACA
GCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACC
GGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCACAGACTG
GAAGAGTCCTTCCTGGTGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACG
AGGTGGCCTACCACGAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACAA
GGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCACATGATCAAGTTCCGGGGCCACTTCTGATCGAGG
GCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTTCATCCAGCTGGTGCAGACCTACAACCAGCTG
TTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGA
GCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTTCGGAACCTGAT
TGCCCTGAGCCTGGGCTGACCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAACTGCAGC
TGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCAGATCGGCGACCAGTACGCCGACCT

GTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCA
CCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAA
GCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGC
CGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATG
GACGGCACCGAGGAAGTCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGAC
AACGGCAGCATCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTT
CCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCC
CTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCTGGAA
CTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAG
AACCTGCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCT
GACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCTTCTGAGCGGCGAGCAGAAAAAGGC
CATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAG
AAAATCGAGTGCTTCGACTCCGTGGAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATA
CCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCTGGACAATGAGGAAAACGAGGACATTCTGGAA
GATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCA
CCTGTTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCG
GAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCTGAAGTCCGACGGCT
TCGCCAACAGAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCC
CAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGA
AGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAGTGATGGGCGGCACAAGCCCGAGA
ACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAAGAGCCGCGAGAGAA
TGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAACA
CCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGA
ACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACT
CCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCAGGGGCAAGAGCGACAACGTGCCCTCCGAAGA
GGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTC
GACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCGCGCTTCATCAAGAGACAGC
TGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGA
CGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCTGAAGTCCAAGCTGGTGTCCGATTTCCGG
AAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACCTACCACCACGCCACGACGCCTACCTGAACGCC
GTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGG
TGACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTT
CTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTC
TGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGA
AAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGA
GTCTATCCTGCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTAC
GGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCA
AGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCC
CATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACT
CCCTGTTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGGAAACGA
ACTGGCCCTGCCCTCAAATATGTGAACTTCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCC
CGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAAGCACTACCTGGACGAGATCATCGAGCAG
ATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACA
GCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAG
CCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAGAGGTGCTG
GACGCCACCCTGATCCACCAGAGCATACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGG
CGACAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAG (SEQ ID NO: 75)

5'Nuclear localization signal + SpCas9 + 3'Nuclear localization signal

GGCAGTGGAGAGGGCAGAGGAAGTCTGCTAACATGCGGTGACGTCGAGGAGAATCCTGGCCCA (SEQ ID NO: 63)

Thosea Asigna Virus T2A skipping peptide

GTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAA
CGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTT
CATCTGCAACCACGGCAAGCTGCCCCTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCACT
GCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGC
GACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCAC
AAGCTGGAGTACAACACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAG
GTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGACGCTCGCCGACCACTACCAGCAGAAC
ACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAA
AGACCCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGC
ATGGACGAGCTGTACAAGGAATTCTAA

EGFP fusion protein

Ctagagctcgctgatcagcctcgactgtgccttctagttgccagccatctgtgtttgccctccccgtgccttccttgaccctggaaggtgcca
ctccactgtcctttcctaataaaatgaggaaattgcatcgattgtctgagtaggtgtcattctattctggggggtgggggtggggcaggacag
caagggggaggattgggaagagaatagcaggcatgctgggga (SEQ ID NO:65)

BGH poliA

AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCA
AAGGTCGCCCCGACGCCCGGGCTTTGCCCCGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG
(SEQ ID NO:66)

3'ITR

GAGGGCCTATTTCCCATGATTCCCTTCATATTTGCATATACGATACAAGGCTGTAGAGAGATAATTGGAATT
AATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTT
GCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTGATTTCTTGG
CTTTATATATCTTGTTGAAAGGACGAAACACCGgtgggctgtaatcatcgtctGTTTTAGAGCTAGAAATAGCAAGT
TAAATAAGGCTAGTCCGTTATCAACTGAAAAAGTGGCACCGAGTCGGTGCTTTTTGTTTTAGAGCTAGA
AATAGCAAGTTAAAATAAGGCTAGTCCGTTTTTAGCGCGTGCGCCAATTCTGCAGACAAATGGCTCTAGAG
GTACCgttacataacttacggtaaatggccgcctggctgaccgccaacgacccccgccattgacgtcaatagtaacgccaataggga
ctttccattgacgtcaatgggtggagtatttacggtaaactgccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgac
gtcaatgacggtaaatggccgcctggcattgtgccagtacatgacctatgggactttcctacttggcagtacatctacgtattagtcatcgc
tattaccatggctgaggtgagccccacgttctgcttactctccccatctccccccctccccacccccaatttgtattttatttttaattatt
ttgtgcagcgatggggggcgggggggggggggggcgcgcgccaggcgggggcgggggcgggggcgagggggcgggggcgaggcgggag
agggtgcgggcggcagccaatcagagcggcgcgctccgaaagtttcttttatggcgaggcgggcgggcgggcgccctataaaaagcgaagc
gcgcggcgggcggggagtcgctgcgacgctgccttcgccccgtgccccgctccgcccgcgcctcgcgccgcccggcctctgactgaccgc
gttactcccacaggtgagcgggcgggacggcccttctctccgggctgtaattagctgagcaagaggtgaagggttaagggatggttggttg
tgggggtattaatgtttaattacctggagcacctgcctgaaatcacttttttcaggttgACCGGTGCCACCATGGACTATAAGGA

CCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAGATGGCCCCAAAGAA
GAAGCGGAAGGTCTGGTATCCACGGAGTCCCAGCAGCCGACAAGAAGTACAGCATCGGCCTGGACATCGGC
ACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAAATTCAAGGTGCTGG
GCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCCGACAGCGGCGAAACAGC
CGAGGCCACCCGGCTGAAGAGAACC GCCAGAAGAAGATACACCAGACGGAAGAACC GGATCTGCTATCTG
CAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCT
GGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCAC
GAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGC
TGATCTATCTGGCCCTGGCCACATGATCAAGTTCGGGGGCCACTTCTGATCGAGGGCGACCTGAACCCCG
ACAACAGCGACGTGGACAAGCTGTTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTTCGAGGAAAACCCC
ATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAA
ATCTGATCGCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTTCGGAAACCTGATTGCCCTGAGCCTGGGC
CTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAACTGCAGCTGAGCAAGGACACCTA
CGACGACGACCTGGACAACCTGCTGGCCAGATCGGCGACCAAGTACGCCGACCTGTTTCTGGCCGCCAAGA
ACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCTGAGC
GCCTCTATGATCAAGAGATACGACGAGCACCACAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCA
GCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCG
GAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACT
GCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCAC
CAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTACCCATTCTGAAGGACAA
CCGGGAAAAGATCGAGAAGATCCTGACCTTCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACA
GCAGATTGCTGCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCTGGAACCTTCGAGGAAGTGGTGG
CAAGGGCGCTTCCGCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAG
GTGCTGCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAGTGAAATACGT
GACCGAGGGAATGAGAAAGCCCGCCTTCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTT
AAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACT
CCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATT
ATCAAGGACAAGGACTTCTTGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGA
CACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCACCTGTTTCGACGACAAAGT
GATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCAT
CCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCTGAAGTCCGACGGCTTCGCCAACAGAACTTCA
TGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGG
CGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAG
TGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCCGAGAACATCGTGATCGAAATGG
CCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAG
GGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACCCCCGTGGAAAACACCCAGCTGCAGAACGAGA
AGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCT
GTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGC
TGACCAGAAGCGACAAGAACC GGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGA
AGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTTCGACAATCTGACCAAGGCC
GAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAG
ATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGAT
CCGGGAAGTGAAAGTGATCACCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACA
AAGTGCGCGAGATCAACAACCTACCACCACGCCCACGACGCCTACCTGAACGCCGTCTGGGAACCGCCCTG
ATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGA
TGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTACAGCAACATCATGAAC
TTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGA
AACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAA

GTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGA
ACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCAC
CGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGAAAAAGGGCAAGTCCAAGAACTGAAGAGTGTGAAA
GAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAA
GGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTTCGAGCTGGAAAACG
GCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATA
TGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCGAGGATAATGAGCAGAAAC
AGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGA
GTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGGATAAGCCCATCAG
AGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTT
TGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAGAGGTGCTGGACGCCACCCTGATCCACCAG
AGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGCGACAAAAGGCCGGCGGCCA
CGAAAAAGGCCGGCCAGGCCAAAAAAGAAAAAGGAATTCGGCAGTGGAGAGGGCAGAGGAAGTCTGCTAA
CATGCGGTGACGTCGAGGAGAATCCTGGCCAGTGAGCAAGGGCGAGGAGCTGTTACACGGGGTGGTGC
CCATCCTGGTCGAGCTGGACGGCGACGTAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCG
ATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCCGTGCCCTGGCCCACC
CTCGTGACCACCCTGACCTACGGCGTGCACTGCTTACGCCGCTACCCCGACCACATGAAGCAGCACGACTT
CTTCAAGTCCGCCATGCCCCAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTAC
AAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGAC
TTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACAGCCACAACGTCTATATC
ATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAG
CGTGCACTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCCGACAAC
CACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGG
AGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGAATTCTAActagagctcgctgat
cagcctcgactgtgccttctagttgccagccatctgtgtttgcccctccccctgccttccttgaccctggaaggtgccactcccactgtccttcc
ctaataaaatgaggaaattgcatcgattgtctgagtaggtgtcattctattctggggggtgggggtggggcaggacagcaagggggaggatt
gggaagagaatagcaggcatgctggggaGCGGCCGCAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCG
CGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTGCCCCGACGCCCCGGGCTTTGCCCGGGCGGCCTCA
GTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG (SEQ ID NO:76)

Construct p1545_pTIGEM_hALB3'HITIdonor(SAS_albex13_ex14_T2A_dsRED_WPRE_bGHpA)

**CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGGTCGCCCCGG
CCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCT (SEQ ID
NO:110)**

5'-ITR

CCACTCATCACGTTATGTAGTGT (SEQ ID NO: 77)

Inverted gRNA sequence for human Albumin intron 12 + PAM sequence

Gataggcacctattggtcttactgacatccactttgcctttctctccacag (SEQ ID NO:21)

Splice acceptor sequence

TGCACTTGTTGAGCTCGTGAAACACAAGCCCAAGGCAACAAAAGAGCAACTGAAAGCTGTTATGGATGATT
TCGCAGCTTTTGTAGAGAAGTGCTGCAAGGCTGACGATAAGGAGACCTGCTTTGCCGAGGAG (SEQ ID
NO:78)

Exon 13 human Albumin

Ggtaaaaaactgttgctgcaagtcaagctgccttaggctta (SEQ ID NO:79)

Exon 14 human Albumin

**GGAAGCGGAGAGGGCAGAGGAAGTCTGCTAACATGCGGTGACGTCGAGGAGAATCCTGGACCT (SEQ ID
NO:23)**

Thosea asigna virus 2A (T2A) skipping peptide

ATGGATAGCACTGAGAACGTCATCAAGCCCTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGG
CCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCAAGCCCTACGAGGGGCACCCAGACCGCCAAGCTGCA
GGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCCAGTTCCAGTACGGCTCCAAGG
TGTACGTGAAGCACCCCGCCGACATCCCCGACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAG
CGCGTGATGAACTTCGAGGACGGCGGCGTGTTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCACCT
TCATCTACCACGTGAAGTTCATCGGCGTGAAGTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACTCTGG
GCTGGGAGCCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAGATCCACAAGGCGCT
GAAGCTGAAGGGCGGCGGCCACTACCTGGTGGAGTTCAAGTCAATCTACATGGCCAAGAAGCCCGTGAAG
CTGCCCGGCTACTACTACGTGGACTCCAAGCTGGACATCACCTCCCAACGAGGACTACACCGTGGTGA
GCAGTACGAGCGCGCCGAGGCCCGCCACCACCTGTTCCAGTAG

Discosoma Red (DsRed) coding sequence

aatcaacctctggattacaaaatttgtaaagattgactgggtattcttaactatgttgctccttttacgctatgtggatacgctgctttaatgcctt
tgtatcatgctattgcttcccgatggctttcattttctcctcctgtataaatcctgggtgctgtctctttatgaggagttgtggcccggtgtcaggc
aacgtggcggtggtgtgactgtgtttgctgacgcaacccccactgggtggggcattgccaccacctgtcagctcctttccgggactttcgctttcc
ccctccctattgccacggcggaactcatcgccgcctgccttggccgctgctggacaggggctcggctgttgggcactgacaattccgtggtgtt
gtcggggaaatcatcgtcctttccttggtgctcgcctgtgttgccacctggattctgcgcgggacgtccttctgctacgtcccttcggccctcaa
tccagcggaccttccttcccgcggtcgtgccggctctgcggcctcttccgcgtcttcg

Woodchuck hepatitis virus post-transcriptional regulatory element (WPRE)

gcctcgactgtgccttctaagttqccaqccatctgttqtttqcccctccccgtgccttccttgaccctqaaaggtgccactcccactgtcctttc
ctaataaaatqaggaattqcatcgcattgtctgaagtaggtgtcattctattctqggggtqgggtqgggcaggacagcaaggggga
ggattqggaaqacaataqcaagcatgctqggga (SEQ ID NO:26)

Bovine growth hormone poli A (BGH pA)

CCACTCATCACGTTATGTAGTGT (SEQ ID NO:77)

Inverted gRNA sequence for human Albumin intron 12 + PAM sequence

**AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGGCTCGCTCGCTCACTGAGGCCGGGCGACC
AAAGGTCGCCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAG (SEQ ID
NO:29)**

3'-ITR

ctgcgcgctcgctcgctcactgaggccgcccgggcaaagcccgggcgctcgggcgacctttggctgcccggcctcagtgagcgagcgagcgc
gcagagagggagtgggccaactccatcactaggggttcctgctagtgctagcggcgcgccctctaCCACTCATCACGTTATGTAGTGT
gataggcacctattgggtcttactgacatccactttgcctttctctccacagTGCACTTGTTGAGCTCGTGAAACACAAGCCCAA
GGCAACAAAAGAGCAACTGAAAGCTGTTATGGATGATTTGCGAGCTTTTGTAGAGAAGTGCTGCAAGGCTG
ACGATAAGGAGACCTGCTTTGCCGAGGAGggtaaaaaacttgttgctgcaagtcaagctgccttaggcttaGGAAGCGGA
GAGGGCAGAGGAAGTCTGCTAACATGCGGTGACGTCGAGGAGAATCCTGGACCTATGGATAGCACTGAGA
ACGTCATCAAGCCCTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATC
GAGGGCGAGGGCGAGGGCAAGCCCTACGAGGGCACCCAGACCGCCAAGCTGCAGGTGACCAAGGGCGG
CCCCCTGCCCTTCGCTGGGACATCCTGTCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCC
CGCCGACATCCCCGACTACAAGAAGCTGTCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCG
AGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCACCTTCATCTACCACGTGAAG
TTCATCGGCGTGAACCTCCCTCCGACGGCCCCGTAATGCAGAAGAAGACTCTGGGCTGGGAGCCCTCCAC
CGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAGATCCACAAGGCGCTGAAGCTGAAGGGCGG
CGGCCACTACCTGGTGGAGTTCAAGTCAATCTACATGGCCAAGAAGCCCGTGAAGCTGCCCCGCTACTACT
ACGTGGACTCCAAGCTGGACATCACCTCCCACAACGAGGACTACACCGTGGTGGAGCAGTACGAGCGCGC
CGAGGCCCCGCCACCACCTGTTCCAGTAGgatccaatcaacctctggattacaaaatttgtgaaagattgactggtattcttaacta
tgttgctccttttacgctatgtggatacgtgctttaatgcctttgtatcatgtattgcttcccgtatggctttcattttctcctcctgtataaatcc
tggttgctgtctctttatgaggagttgtggccgtgtcaggcaacgtggcgtggtgtgactgtgtttgctgacgcaacccccactggtgggg
cattgccaccacgtgtcagctcctttccgggactttcgctttccccctccctattgccacggcggaactcatcgccgcctgccttcccgcgtctg
gacaggggctcggtgttgggcactgacaattccgtggtgttgcggggaaatcatcgctccttccctggctgctgcctgtgttgccacctgga
ttctgcgcgggacgtccttctgctacgtcccttcggccctcaatccagcgaccttccctcccgcggcctgctgcccgtctgcggcctctccgc
gtcttcgagatctgcctcgactgtgccttctagtgtccagccatctgttgtttgcccctccccgtgccttccctgacctggaaggtgccactc
ccactgtcctttcctaataaaaatgaggaaattgcatcgattgtctgagtaggtgtcattctattctgggggtgggggtggggcaggaca
gcaagggggaggattgggaagacaatagcaggcatgctggggaCCACTCATCACGTTATGTAGTGTgctctgtcgaggaa
ttgAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGA
CCAAAGGTGCCCCGACGCCCCGGGCTTTGCCCGGGCGGCTCAGTGAGCGAGCGAGCGCGCAG (SEQ ID
NO: 80)

Construct p1615_pCbh-SpCas9(BB)-2A-GFP + 3' HUMAN albumin gRNA6

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGGAATT
AATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTT
GCAGTTTTTAAATTATGTTTTTAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTTCGATTTCTTGG
CTTTATATATCTTGTGGAAAGGACGAAACACC (SEQ ID NO:59)

Human U6 promoter

Tattggcagtcaaggcccg (SEQ ID NO:17)

gRNA6 human albumin

GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC
GGTGCTTTTTT (SEQ ID NO:60)

Chimeric gRNA scaffold

cgttacataacttacggtaaatggcccgctggctgaccgccaacgacccccgccattgacgtcaatagtaacgccaatagggactttcca
ttgacgtcaatgggtggagtatttacggtaaatgcccacttggcagttacatcaagtgtatcatatgccaagtacgccccctattgacgtcaat
gacggtaaatggcccgctggcattgtgcccagttacatgaccttatgggactttcctacttggcagttacatctacgtattagtcacgtattac
catggtcgaggtgagccccacgttctgcttactctccccatctccccccctccccacccaattttgtattttattttttaattttttgtgc
agcgatggggcgggggggggggggggcgcgcgccaggcgggggcgggggcgagggggcgggggcgagggcgagaggtgc
ggcgcgagccaatcagagcgggcgcgctccgaaagtttccttttatggcgaggcgggcgggcgggcgccctataaaaaagcgaagcgcgcg
cgggcgggagtcgctgcgacgtgccttcgccccgtgccccgctccgcccgcctcgcgcgccccggcctctgactgaccgcgttact
cccacaggtgagcgggcgggacggcccttctctccgggctgtaattagctgagcaagaggttaagggttaagggtggttggtgggg
tattaatgtttaattacctggagcacctgcctgaaatcacttttttcaggttgg

CBH promoter

**ATGGA CTATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAG
(SEQ ID NO:62)**

3XFlag tag

ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGACAAGAAGTACAGCATCG
GCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAA
ATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTTCGACA
GCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACC
GGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCACAGACTG
GAAGAGTCCTTCCTGGTGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACG
AGGTGGCCTACCACGAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACAA
GGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCACATGATCAAGTTCCGGGGCCACTTCTGATCGAGG
GCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTTCATCCAGCTGGTGCAGACCTACAACCAGCTG
TTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGA
GCAGACGGCTGGAATCTGATCGCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTTCGGAACCTGAT
TGCCCTGAGCCTGGGCTGACCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAACTGCAGC
TGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCAGATCGGCGACCAGTACGCCGACCT

GTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCA
CCAAGGCCCCCTGAGCGCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAA
GCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGC
CGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATG
GACGGCACCGAGGAAGTCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGAC
AACGGCAGCATCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTT
CCATTCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCC
CTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAACCATCACCCCTGGAA
CTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAG
AACCTGCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCT
GACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCTTCTGAGCGGCGAGCAGAAAAAGGC
CATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAG
AAAATCGAGTGCTTCGACTCCGTGGAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATA
CCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCTGGACAATGAGGAAAACGAGGACATTCTGGAA
GATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCA
CCTGTTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCG
GAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCTGAAGTCCGACGGCT
TCGCCAACAGAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCC
CAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGA
AGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAGTGATGGGCCGGCACAAGCCCGAGA
ACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAA
TGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAACA
CCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGA
ACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACT
CCATCGACAACAAGTGCTGACCAGAAGCGACAAGAACCAGGGGCAAGAGCGACAACGTGCCCTCCGAAGA
GGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTC
GACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGC
TGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGA
CGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCTGAAGTCCAAGCTGGTGTCCGATTTCCGG
AAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACCTACCACCACGCCACGACGCCTACCTGAACGCC
GTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGG
TGACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTT
CTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTC
TGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGA
AAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGA
GTCTATCCTGCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTAC
GGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCA
AGAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCC
CATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACT
CCCTGTTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGGAAACGA
ACTGGCCCTGCCCTCAAATATGTGAACTTCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCC
CGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAAGCACTACCTGGACGAGATCATCGAGCAG
ATCAGCGAGTTCTCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAA
GCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAG
CCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAGAGGTGCTG
GACGCCACCCTGATCCACCAGAGCATACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGG
CGACAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAG (SEQ ID NO:81)

5'Nuclear localization signal + SpCas9 + 3'Nuclear localization signal

GGCAGTGGAGAGGGCAGAGGAAGTCTGCTAACATGCGGTGACGTGAGGAGAATCCTGGCCCA (SEQ ID NO: 63)

Thosea Asigna Virus T2A skipping peptide

GTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGCAAGCTGCCGTGCCCTGGCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCGAAGGCTACGTCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCAC AAGCTGGAGTACAACACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACCTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGACGCTCGCCGACCACTACCAGCAGAACACCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGAATTCTAA

EGFP fusion protein

Ctagagctcgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctccccctgccttcttgaccctggaaggtgccactcccactgtcctttcctaataaaatgaggaaatgcatcgattgtctgagtaggtgtcattctattctgggggtgggggtggggcaggacagcaagggggaggattgggaagagaatagcaggcatgctgggga (SEQ ID NO:65)

BGH poliA

AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGGCGACCA AAGGTCGCCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG (SEQ ID NO:66)

3'ITR

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGGAATT AATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTT GCAGTTTTAAATATGTTTTAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTGATTTCTTGG CTTTATATATCTTGTGGAAGGACGAAACACCGTattggcagtcaaggccccgGTTTTAGAGCTAGAAATAGCAAG TTAATAAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTGTGTTTAGAGCTAG AAATAGCAAGTTAAATAAAGGCTAGTCCGTTTTTAGCGCGTGCGCCAATTCTGCAGACAAATGGCTCTAGA GGTACCggttacataacttacggtaaatggcccgctggctgaccgccaacgacccccgcccattgacgtcaatagtaacccaatagg gactttccattgacgtcaatgggtggagtatttacggtaaaactgcccacttggcagtagacatcaagtgtatcatatgccaaagtacgccccctattg acgtcaatgacggtaaatggcccgctggcattgtgcccagtagacattatgggactttcctacttggcagtagacatctacgtattagtcac gctattaccatggctgaggtgagccccacgttctgcttactctccccatctccccccctccccaccccccaattttgtattttatttttttaatt atttgtgcagcgatggggcgggggggggggggggcgcgccaggcgggcgggcgggcgaggggcgggcgggcgaggcgg agaggtgcgggcgagccaatcagagcgggcgcgctccgaaagtctcttttatggcgaggcgggcgggcgggcgccctataaaaagcgaa gcgcgcgggcggggagtcgctgacgctgccttcgccccgtgccccgctccgcccgcctcgccgcccgggctctgactgacc gcgttactccacaggtgagcgggcgggacggcccttctcctcgggctgtaattagctgagcaagaggttaagggttaagggtggttggtt ggtggggtattaatgtttaattacctggagcacctgcctgaaatcacttttttcagggttgACCGGTGCCACCATGGACTATAAGG

ACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAGATGGCCCCAAAGA
AGAAGCGGAAGGTTCGGTATCCACGGAGTCCCAGCAGCCGACAAGAAGTACAGCATCGGCCTGGACATCGG
CACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAAATTCAAGGTGCTG
GGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAG
CCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCT
GCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCC
TGGTGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCA
CGAGAAGTACCCCAACCTCTACCACTGAGAAAGAACTGGTGGACAGCACCGACAAGGCCGACCTGCGG
CTGATCTATCTGGCCCTGGCCACATGATCAAGTTCGGGGGCCACTTCTGATCGAGGGCGACCTGAACCCC
GACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCC
CATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAA
AATCTGATCGCCCAGCTGCCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGG
CCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAACTGCAGCTGAGCAAGGACACCT
ACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCACTACGCCGACCTGTTTCTGGCCGCCAAG
AACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAGGCCCCCTGAG
CGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGC
AGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGC
GGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGGCCATCCTGGAAAAGATGGACGGCACCGAGGAAC
TGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCA
CCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTACCCATTCTGAAGGACA
ACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAAC
AGCAGATTGCTCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCTGGAACCTTCGAGGAAGTGGTGG
ACAAGGGCGCTTCCGCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCAACGAGAAG
GTGCTGCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGT
GACCGAGGGAATGAGAAAGCCCGCCTTCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTC
AAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACT
CCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATT
ATCAAGGACAAGGACTTCTTGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGA
CACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCACCTGTTCGACGACAAAGT
GATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCAT
CCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCTGAAGTCCGACGGCTTCGCCAACAGAACTTCA
TGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGG
CGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAG
TGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCCGAGAACATCGTGATCGAAATGG
CCAGAGAGAACCAGACCAACCAGAAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAG
GGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAGAACACCCAGCTGCAGAACGAGA
AGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACCTGGACATCAACCGGCT
GTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGC
TGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCTGTAAGAAGATGA
AGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTTGACAATCTGACCAAGGCC
GAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAG
ATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGAT
CCGGGAAGTGAAAGTGATCACCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACA
AAGTGCGCGAGATCAACAACCTACCACCACGCCCACGACGCCTACCTGAACGCCGTCTGTTGGAACCGCCCTG
ATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGA
TGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTACAGCAACATCATGAAC
TTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGA
AACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAA

GTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGA
ACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCAC
CGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGAAAAAGGGCAAGTCCAAGAACTGAAGAGTGTGAAA
GAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAA
GGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTTCGAGCTGGAAAACG
GCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATA
TGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCGAGGATAATGAGCAGAAAC
AGCTGTTTGTGGAACAGCACAAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGA
GTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGGATAAGCCCATCAG
AGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTT
TGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAGAGGTGCTGGACGCCACCCTGATCCACCAG
AGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGCGACAAAAGGCCGGCGGCCA
CGAAAAAGGCCGGCCAGGCCAAAAAAGAAAAAGGAATTCGGCAGTGGAGAGGGCAGAGGAAGTCTGCTAA
CATGCGGTGACGTCGAGGAGAATCCTGGCCAGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGC
CCATCCTGGTCGAGCTGGACGGCGACGTAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCG
ATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCCGTGCCCTGGCCCACC
CTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTT
CTTCAAGTCCGCCATGCCCCAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTAC
AAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGAC
TTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACAGCCACAACGTCTATATC
ATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAG
CGTGCACTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCCGACAAC
CACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGG
AGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGAATTCTAActagagctcgctgat
cagcctcgactgtgccttctagttgccagccatctgtgtttgcccctccccctgccttccttgaccctggaaggtgccactcccactgtccttc
ctaataaaatgaggaaattgcatcgattgtctgagtaggtgtcattctattctggggggtgggggtggggcaggacagcaagggggaggatt
gggaagagaatagcaggcatgctggggaGCGGCCGCAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCG
CGCTCGCTCGCTCACTGAGGCCGGGCGACCAAGGTGCCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCA
GTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG (SEQ ID NO:82)

Construct p1616_pCbh-SpCas9(BB)-2A-GFP + 3' HUMAN albumin gRNA7

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGGAATT
AATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTT
GCAGTTTTTAAATTATGTTTTTAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTTCGATTTCTTGG
CTTTATATATCTTGTGGAAAGGACGAAACACC (SEQ ID NO:59)

Human U6 promoter

Tcgaatgtattgtgacagag (SEQ ID NO:18)

gRNA7 human albumin

GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC
GGTGCTTTTTT (SEQ ID NO:60)

Chimeric gRNA scaffold

cgttacataacttacggtaaatggcccgctggctgaccgccaacgacccccgccattgacgtcaatagtaacgccaatagggactttcca
ttgacgtcaatgggtggagtatttacggtaaatgccacttggcagtagcatcaagtgtatcatatgccaagtacgccccctattgacgtcaat
gacggtaaatggcccgctggcattgtgcccagtagcatgaccttatgggactttcctacttggcagtagcatctacgtattagtcacgtattac
catggtcgaggtgagccccacgttctgcttactctccccatctccccccctccccaccccccaattttgtattttattttttaattattttgtgc
agcgatgggggcgggggggggggggggggcgcgcgccaggcggggcggggcggggcgaggggcggggcgaggcgagaggtgc
ggcggcagccaatcagagcgggcgcgctccgaaagtttccttttatggcgaggcgggcgggcgggcgccctataaaaaagcgaagcgcgcg
cgggcgggagtcgctgcgacgctgccttcgccccgtgccccgctccgcccgcctcgcgccgccccggccttgactgaccgcgttact
cccacaggtgagcgggcgggacggcccttctctccgggctgtaattagctgagcaagaggtaagggtttaagggatggttggttggtggg
tattaatgtttaattacctggagcacctgcctgaaatcacttttttcaggttgg

CBH promoter

**ATGGA CTATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAG
(SEQ ID NO:62)**

3XFlag tag

ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGACAAGAAGTACAGCATCG
GCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAA
ATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTTCGACA
GCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACC
GGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCACAGACTG
GAAGAGTCCTTCCTGGTGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACG
AGGTGGCCTACCACGAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACAA
GGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCACATGATCAAGTTCCGGGGCCACTTCTGATCGAGG
GCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTTCATCCAGCTGGTGCAGACCTACAACCAGCTG
TTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGA
GCAGACGGCTGGAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTTCGGAACCTGAT
TGCCCTGAGCCTGGGCTGACCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAACTGCAGC
TGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCAGATCGGCGACCAGTACGCCGACCT

GTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCA
CCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAA
GCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGC
CGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATG
GACGGCACCGAGGAAGTCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGAC
AACGGCAGCATCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTT
CCATTCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCC
CTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAACCATCACCCCTGGAA
CTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAG
AACCTGCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCT
GACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCTTCTGAGCGGCGAGCAGAAAAAGGC
CATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAG
AAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATA
CCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCTGGACAATGAGGAAAACGAGGACATTCTGGAA
GATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCA
CCTGTTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCG
GAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCTGAAGTCCGACGGCT
TCGCCAACAGAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCC
CAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGA
AGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAGTGATGGGCCGGCACAAGCCCGAGA
ACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAAGACGCCGCGAGAGAA
TGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAACA
CCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGA
ACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACT
CCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCAGGGGCAAGAGCGACAACGTGCCCTCCGAAGA
GGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTC
GACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGC
TGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGA
CGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCTGAAGTCCAAGCTGGTGTCCGATTTCCGG
AAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACCTACCACCACGCCACGACGCCTACCTGAACGCC
GTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGG
TGACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTT
CTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTC
TGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGA
AAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGA
GTCTATCCTGCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTAC
GGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCA
AGAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCC
CATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACT
CCCTGTTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGA
ACTGGCCCTGCCCTCCAAATATGTGAACTTCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCC
CGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAAGCACTACCTGGACGAGATCATCGAGCAG
ATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACA
GCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAG
CCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAGAGGTGCTG
GACGCCACCCTGATCCACCAGAGCATACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGG
CGACAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAG (SEQ ID NO: 19)

5'Nuclear localization signal + SpCas9 + 3'Nuclear localization signal

GGCAGTGGAGAGGGCAGAGGAAGTCTGCTAACATGCGGTGACGTGAGGAGAATCCTGGCCCA (SEQ ID NO: 63)

Thosea Asigna Virus T2A skipping peptide

GTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAA
CGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTT
CATCTGCACCACCGGCAAGCTGCCCCTGCCCTGGCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGT
GCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGC
GACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCAC
AAGCTGGAGTACAACACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAG
GTGAACCTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGACGCTCGCCGACCACTACCAGCAGAAC
ACCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAA
AGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGC
ATGGACGAGCTGTACAAGGAATTCTAA

EGFP fusion protein

Ctagagctcgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctccccctgccttcttgaccctggaaggtgcc
ctcccactgtcctttcctaataaaatgaggaaatgcatcgattgtctgagtaggtgtcattctattctgggggtgggggtggggcaggacag
caagggggaggattgggaagagaatagcaggcatgctgggga (SEQ ID NO:65)

BGH poliA

AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGGCGACCA
AAGGTCGCCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG
(SEQ ID NO:66)

3'ITR

GAGGGCCTATTTCCCATGATTCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGGAATT
AATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTT
GCAGTTTTAAATATGTTTTAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTGATTTCTTGG
CTTTATATATCTTGTGGAAGGACGAAACACCGTcgaatgtattgtgacagagGTTTTAGAGCTAGAAATAGCAAG
TTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTGTTTTAGAGCTAG
AAATAGCAAGTTAAATAAGGCTAGTCCGTTTTTAGCGCGTGCGCCAATTCTGCAGACAAATGGCTCTAGA
GGTACCggttacataacttacggtaaatggcccgctggctgaccgccaacgacccccgcccattgacgtcaatagtaacccaatagg
gactttccattgacgtcaatgggtggagtatttacggtaaactgcccacttggcagtagcatcaagtgtatcatatgccaaagtagccccctattg
acgtcaatgacggtaaatggcccgctggcattgtgcccagtagacattatgggactttcctacttggcagtagcatctacgtattagtcac
gctattaccatggctgaggtgagccccacgttctgcttactctccccatctccccccctccccaccccccaattttgtattttatttttaatt
atthttgtcagcgatggggcgggggggggggggggcgcgccaggcgggcgggcgggcgaggggcgggcgggcgaggcg
agaggtgcgggcgagccaatcagagcgggcgcgctccgaaagtctcttttatggcgaggcgggcgggcgggcgccctataaaaagcgaa
gcgcgcgggcgggagtcgctgacgctgccttcgccccgtgccccgctccgcccgcctcgccgcccgggctctgactgacc
gcgttactccacaggtgagcgggcgggacggcccttctcctccgggctgtaattagctgagcaagaggttaagggttaagggtggtggtt
gggtgggtattaatgtttaattacctggagcacctgcctgaaatcacttttttcagggttgACCGGTGCCACCATGGACTATAAGG

ACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAGATGGCCCCAAAGA
AGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGACAAGAAGTACAGCATCGGCCTGGACATCGG
CACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAAATTCAAGGTGCTG
GGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAG
CCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCT
GCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTCCACAGACTGGAAGAGTCCTTCC
TGGTGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCA
CGAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACAAGGCCGACCTGCGG
CTGATCTATCTGGCCCTGGCCACATGATCAAGTTCGGGGGCCACTTCTGATCGAGGGCGACCTGAACCCC
GACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCC
CATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAA
AATCTGATCGCCCAGCTGCCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGG
CCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAACTGCAGCTGAGCAAGGACACCT
ACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCACTACGCCGACCTGTTTCTGGCCGCCAAG
AACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAGGCCCCCTGAG
CGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGC
AGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGC
GGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGGCCATCCTGGAAAAGATGGACGGCACCGAGGAAC
TGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCA
CCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTACCCATTCTGAAGGACA
ACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAAC
AGCAGATTGCTGCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCTGGAACCTTCGAGGAAGTGGTGG
ACAAGGGCGCTTCCGCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCAACGAGAAG
GTGCTGCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGT
GACCGAGGGAATGAGAAAGCCCGCCTTCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTC
AAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACT
CCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATT
ATCAAGGACAAGGACTTCTTGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGA
CACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCACCTGTTCGACGACAAAGT
GATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCAT
CCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCTGAAGTCCGACGGCTTCGCCAACAGAACTTCA
TGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGG
CGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAG
TGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCCGAGAACATCGTGATCGAAATGG
CCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAG
GGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAGAACACCCAGCTGCAGAACGAGA
AGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACCTGGACATCAACCGGCT
GTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGC
TGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCTGTAAGAAGATGA
AGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTTGACAATCTGACCAAGGCC
GAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAG
ATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGAT
CCGGGAAGTGAAAGTGATCACCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACA
AAGTGCGCGAGATCAACAACCTACCACCACGCCCACGACGCCTACCTGAACGCCGTCTGTGGGAACCGCCCTG
ATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGA
TGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTACAGCAACATCATGAAC
TTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGA
AACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAA

GTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGA
 ACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCAC
 CGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAACTGAAGAGTGTGAAA
 GAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAA
 GGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTTCGAGCTGGAAAACG
 GCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGGAAACGAACTGGCCCTGCCCTCCAAATA
 TGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCGAGGATAATGAGCAGAAAC
 AGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGA
 GTGATCCTGGCCGACGCTAATCTGGACAAAGTGTGTCCGCCTACAACAAGCACCGGGGATAAGCCCATCAG
 AGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTT
 TGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAGAGGTGCTGGACGCCACCCTGATCCACCAG
 AGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGCGACAAAAGGCCGGCGGCCA
CGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGAATTTCGGCAGTGGAGAGGGCAGAGGAAGTCTGCTAA
CATGCGGTGACGTCGAGGAGAATCCTGGCCAGTGAGCAAGGGCGAGGAGCTGTTACCCGGGGTGGTGC
CCATCCTGGTCGAGCTGGACGGCGACGTAACCGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCG
ATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCCTGCCCTGGCCACC
CTCGTGACCACCTGACCTACGGCGTGAGTGCTTCAGCCGCTACCCCGACCATGAAGCAGCACGACTT
CTTCAAGTCCGCCATGCCCAGAGGTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTAC
AAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGAC
TTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACAGCCACAACGTCTATATC
ATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAG
CGTGACGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCCAGCGCCCCGTGCTGCTGCCCGACAAC
CACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGG
AGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGAATTCTAActagagctcgctgat
 cagcctcgactgtgccttctagtgtccagccatctgtgtttgccccccccctgccttccttgaccctggaaggtgccactcccactgtcctttc
 ctaataaaatgaggaaattgcatcgattgtctgagtaggtgtcattctattctgggggtgggggtggggcaggacagcaagggggaggatt
 gggaagagaatagcaggcatgtgtgggaGCGGCCGCAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCG
CGCTCGCTCGCTCACTGAGGCCGGGCGACCAAGGTGCCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCA
GTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG (SEQ ID NO: 83)

Sequences of the above Example 5

Sequence of the ITR donor DNA construct (donor DNA not flanked by the 5' and 3' inverted gRNA sites_p1547)

5' ITR

CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCCGGGCAAAGCCCCGGGCGTCGGGCGACCTTTGGTCGCCCCGG
 CCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACCTCCATCACTAGGGGTTCT (SEQ ID
 NO:110)

Additional aav sequences

Gctagtgttagc (SEQ ID NO: 84)

SAS (splicing acceptor signal)

Gataggcacctattgtcttactgacatccactttgcctttctctccacag (SEQ ID NO:21)

mouse albumin exon 14

ggtccaaacctgtcactagatgcaagacgccttagcc (SEQ ID NO:22)

T2A sequence

Ggaagcggagagggcagaggaagtctgctaucatgcggtgacgtcgaggagaatcctggacct (SEQ ID NO:23)

Atggatagcactgagaacgtcatcaagcccttcatgcgttcaagggtgcacatggagggtccgtgaacggccacgagttcagatcgaaggcgagggcgaggggcaagc
ccctacgagggcacccagaccgccaaagctgcaggtgaccaaggcgggcccccctgcccttcgcttgggacatcctgtccccccagttccagtlacggctccaaggtgtacgtga
agcaccgcccgacatccccgactacaagaagctgtccttccccgagggttcaagtgggagcgctgtatgaacttcgaggacgcgcgctgtgtgacgtgacccaggac
tctccctcgaggacggcaccttcatctaccacgtgaagttcatcggcgtgaacttccccctcgacggcccccgaatgcagaagaagactctgggctgggagccccccaccg
agcgccgtacccccgcgacggcgctgtcgaaggcgagatccacaaggcgctgaagctgaaggcgcgcgccactacctgtgtggagtccaagtcacatctacatgccc
gaagcccgtagaagctgcccggctactactacgtggactccaagctggacatcacctcccacaacgaggactacaccgtgtgtggagcagtlacgagcgcgccgagccgc
caccacctgttccagtag

Aatcaacacctctggattacaaaaatttgtgaagattgactggattctttaaactatgttgctccttttaccgctatgtggataccgctgctttaatgccctttgatcatgctattgcttcccgtatg
gctttcattttctcctccttgataaacctgggtgctgctctttatgaggagttgtggcccgtgtgcaggcaacgtggcgtggtgtgcactgtgttctgacgcaacccccactgggt
tggggcattgccaccacctgtcagctccttccgggaacttgcgttccccctccctattgccacggcggaaatcatcgccgcctgcttggccgctgctgacaggggctcgg
ctgttgggcactgacaattccgtgtgtgtcggggaaatcatcgcttcttctggctgctgcctgtgttgccacctggattctgcggcggacgtccttctgctacgtccctcgg
ccctcaatccagcggaccttccctcccgcgccctgctgccggctctcgggcctcttccgcgtcttcg

Agagctcttgcgagggtcgacatttaaatgaagggcgaattcaattt (SEQ ID NO: 85)

Ttagagctagaaatagcaagttaaaaataaggcctagtcctgttttagcgcgtgcgccaatctctgcagacaaatggctctagaggtaccaattg (SEQ ID NO:87)

Full sequence_p1547

CTTCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAAGCCGGGGCTCGGGCGACCTTTGGTCGCCCGG
CCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCCT

gctagtgtctagcgalaggcacctattggtcttactgacatccactttgccctttctccacagggtccaaacctgtgtaactagatgcaaaagacgccttagccgggaugcggaga
gggcagagggaagtctgctaacatgcggtgacgtcgaggagaatctggacctatggatagcactigagaacgtcatcaagcccttcagtcgcttcaaggtgcacatggaggg
gtctcgtgaacggccacgagttcgagatcgaggggcgaggggcaagccctacgaggggaccagaccgccaagctgcaggtgaccaaggggcgccccctgccc
ttcgcttgggacatccgtgtccccccagttccagtlacggctccaaggtgtacgtgaagcaccggccgacatccccgactacaagaagctgtccttccccgagggcttcaagt
ggagcgcgtgatgaacttcgaggacggcggcggtgtgtaccgtgaccagagactcctcctctgcaggacggcaccttcatctaccacgtgaagttcatcgcggtgaacttccc
tcgacggccccgtaatgcagaagaagactctgggcttggagccctccaccgagcgcctgtlaccggcgacggcggtgtgaaggcgagatccacaaggcgctgaag
ctgaaggcgcgcgccactacctggtggagtcaagtcaatctacatlgccaagaagcccggaagctgcccggctactactactgtgactccaagctggacatcacctccc
acaacgaggactacaccgtgtgtggagcagtlacgagcggcgaggggccgccaccacctgttccagtaggataccaatcaacctctgattacaaaatttgtgaagattgact
gtgatcttaactatgttgcctctttacgctatgtggatagcgtctcttaatgcccttgtatcatgctattgcttcccgtatgctttcatttttccctcctgtataaatctggtgtgtct
ctttatgaggagtttgtggccgtgtcaggcaacgtggcggtgtgtgtgcactgtgtttgctgacgcaacccccactggttggggcattggccaccacctgtcagctcctttccggga
cttctgctttccccctccctattgccacggcggaactatcgccgcctgcttggccgctgctgtgacaggggctcggtgtgtggcactgacaattccgtgtgtgtcggggaa
atcatcgctcctttccctggctgctcgccgtgtgtgccacctggattctgcgcgggacgtccttctgclacgtcccttcggccclcaatccagcggaccttccctcccggcgctgt
ggcggtctgcggcccttcccggtcttcgagatct**GCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCCT**
TCCCGTGCCTTCTCTTGACCCCTGGAAGGTGCCACTCCCACGTGCTCTTCTCAATAAAATGAGGAAAT
TGCATCGCATTGTCTGAGTAGGTGTCATCTATTTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGG
GGGAGGATTGGGAAGACAATAGCAGGCAATGCTGGGGagagctctgtcgaggctgacatttaaatgaaggcgcaattcaatttgag
ggcctatttcccatgattccttcatattgcatataccgatacaaggctgttagagagataattggaattaatttgactgtaaacacaaagatattagtacaaaatacgtgacgttagaaa
gtataattttctgggtagtgtgcagttttaaaattatgttttaaaatgactatcatatgcttaccgtlaactgaaaagtatttcgatttcttgcctttatatacttctgtggaaggacgaac

accgtatttaaataggcagcagtggttttagagctagaaatagcaagttaaaataaggctagtcggtatcaacttgaaaaagtgccaccgagtcgggtgctttttgtttagagctag
aaatagcaagttaaaataaggctagtcggttttagcgcgtgcgccaattctgcagacaaatggctctagaggtaccaattgaggaacccctagtcatggaattggccactcc
ctctctgcgcctcgtcctcactgagccggcgaccaaaggtcggccgacccggcgtttgccccggcgcctcagtgagcagcagcggcgaat (SEO
ID NO:88)

References

1. Cong, L., et al., *Multiplex genome engineering using CRISPR/Cas systems*. Science, 2013. **339**(6121): p. 819-23.
2. Jiang, W., et al., *RNA-guided editing of bacterial genomes using CRISPR-Cas systems*. Nat Biotechnol, 2013. **31**(3): p. 233-9.
3. Tu, Z., et al., *CRISPR/Cas9: a powerful genetic engineering tool for establishing large animal models of neurodegenerative diseases*. Mol Neurodegener, 2015. **10**: p. 35.
4. Nishiyama, J., T. Mikuni, and R. Yasuda, *Virus-Mediated Genome Editing via Homology-Directed Repair in Mitotic and Postmitotic Cells in Mammalian Brain*. Neuron, 2017. **96**(4): p. 755-768 e5.
5. Anguela, X.M., et al., *Robust ZFN-mediated genome editing in adult hemophilic mice*. Blood, 2013. **122**(19): p. 3283-7.
6. Barzel, A., et al., *Promoterless gene targeting without nucleases ameliorates haemophilia B in mice*. Nature, 2015. **517**(7534): p. 360-4.
7. Li, H., et al., *In vivo genome editing restores haemostasis in a mouse model of haemophilia*. Nature, 2011. **475**(7355): p. 217-21.
8. Sharma, R., et al., *In vivo genome editing of the albumin locus as a platform for protein replacement therapy*. Blood, 2015. **126**(15): p. 1777-84.
9. Bakondi, B., *In vivo versus ex vivo CRISPR therapies for retinal dystrophy*. Expert Rev Ophthalmol, 2016. **11**(6): p. 397-400.
10. Lackner, D.H., et al., *A generic strategy for CRISPR-Cas9-mediated gene tagging*. Nat Commun, 2015. **6**: p. 10237.
11. Suzuki, K., et al., *In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration*. Nature, 2016. **540**(7631): p. 144-149.
12. Brunetti-Pierri N, A.A., *Gene Therapy of Human Inherited Diseases*, in *The Metabolic and Molecular Bases of Inherited Diseases*, S. R, Editor. 2010, McGraw Hill: New York.
13. Ehrhardt, A., H. Xu, and M.A. Kay, *Episomal persistence of recombinant adenoviral vector genomes during the cell cycle in vivo*. J Virol, 2003. **77**(13): p. 7689-95.
14. E Neufeld, J.M., *The mucopolysaccharidoses*, in *The mucopolysaccharidoses*, A.B. CR Sriver, WS Sly, DM Valle, Editor. 2001, McGraw-Hill: New York (2001). p. 3421-3452.
15. Cotugno, G., et al., *Impact of age at administration, lysosomal storage, and transgene regulatory elements on AAV2/8-mediated rat liver transduction*. PLoS One, 2012. **7**(3): p. e33286.
16. Ferla, R., et al., *Similar therapeutic efficacy between a single administration of gene therapy and multiple administrations of recombinant enzyme in a mouse model of lysosomal storage disease*. Hum Gene Ther, 2014. **25**(7): p. 609-18.

17. Ferla, R., et al., *Gene therapy for mucopolysaccharidosis type VI is effective in cats without pre-existing immunity to AAV8*. Hum Gene Ther, 2013. **24**(2): p. 163-9.
18. Tessitore, A., et al., *Biochemical, pathological, and skeletal improvement of mucopolysaccharidosis VI after gene transfer to liver but not to muscle*. Mol Ther, 2008. **16**(1): p. 30-7.
19. Alliegro, M., et al., *Low-dose Gene Therapy Reduces the Frequency of Enzyme Replacement Therapy in a Mouse Model of Lysosomal Storage Disease*. Mol Ther, 2016. **24**(12): p. 2054-2063.
20. Ferla, R., et al., *Non-clinical Safety and Efficacy of an AAV2/8 Vector Administered Intravenously for Treatment of Mucopolysaccharidosis Type VI*. Mol Ther Methods Clin Dev, 2017. **6**: p. 143-158.
21. Cotugno, G., et al., *Long-term amelioration of feline Mucopolysaccharidosis VI after AAV-mediated liver gene transfer*. Mol Ther, 2011. **19**(3): p. 461-9.
22. Giugliani, R., et al., *Natural history and galsulfase treatment in mucopolysaccharidosis VI (MPS VI, Maroteaux-Lamy syndrome)--10-year follow-up of patients who previously participated in an MPS VI Survey Study*. Am J Med Genet A, 2014. **164A**(8): p. 1953-64.
23. Desnick, R.J. and E.H. Schuchman, *Enzyme replacement therapy for lysosomal diseases: lessons from 20 years of experience and remaining challenges*. Annu Rev Genomics Hum Genet, 2012. **13**: p. 307-35.
24. Neufeld, E.F., *Lysosomal storage diseases*. Annu Rev Biochem, 1991. **60**: p. 257-80.
25. Bowen, D.J., *Haemophilia A and haemophilia B: molecular insights*. Mol Pathol, 2002. **55**(2): p.127-44. Antonarakis, S.E., et al., *Molecular etiology of factor VIII deficiency in hemophilia A*. Adv Exp Med Biol, 1995. **386**: p. 19-34
26. Bunting, S., et al., *Gene Therapy with BMN 270 Results in Therapeutic Levels of FVIII in Mice and Primates and Normalization of Bleeding in Hemophilic Mice*. Mol Ther, 2018. **26**(2): p. 496-509
27. Rangarajan, S., et al., *AAV5-Factor VIII Gene Transfer in Severe Hemophilia A*. N Engl J Med, 2017. **377**(26): p. 2519-2530
28. Makris, M. *Gene therapy 1·0 in haemophilia: effective and safe, but with many uncertainties*. The Lancet Haematology (2020) doi:10.1016/S2352-3026(20)30035-1
29. Grieger, J. C. et al., *Packaging Capacity of Adeno-Associated Virus Serotypes: Impact of Larger Genomes on Infectivity and Postentry Steps*. J. Virol. (2005)
30. Dong, B. et al., *Characterization of genome integrity for oversized recombinant AAV vector*. Mol. Ther. (2010)
31. Hirsch, M. et al., *Little vector, big gene transduction: Fragmented genome reassembly of adeno-associated virus*. Molecular Therapy (2010)
32. Wu, Z. et al., *Effect of genome size on AAV vector packaging*. Mol. Ther. (2010)

33. McIntosh J, Lenting PJ, Rosales C, Lee D, Rabbanian S, Raj D, Patel N, Tuddenham EG, Christophe OD, McVey JH, Waddington S, Nienhuis AW, Gray JT, Fagone P, Mingozzi F, Zhou SZ, High KA, Cancio M, Ng CY, Zhou J, Morton CL, Davidoff AM, Nathwani AC. Therapeutic levels of FVIII following a single peripheral vein administration of rAAV vector encoding a novel human factor VIII variant. *Blood*. 2013 Apr 25;121(17):3335-44. doi: 10.1182/blood-2012-10-462200. Epub 2013 Feb 20. PMID: 23426947; PMCID: PMC3637010.
34. Doria, M., A. Ferrara, and A. Auricchio, *AAV2/8 vectors purified from culture medium with a simple and rapid protocol transduce murine liver, muscle, and retina efficiently*. *Hum Gene Ther Methods*, 2013. **24**(6): p. 392-8.
35. Ferla R, Claudiani P, Cotugno G, Saccone P, De Leonibus E, Auricchio A. Similar therapeutic efficacy between a single administration of gene therapy and multiple administrations of recombinant enzyme in a mouse model of lysosomal storage disease. *Hum Gene Ther*. 2014;25(7):609-618. doi:10.1089/hum.2013.213
36. Harmatz PR, Shediach R. Mucopolysaccharidosis VI: Pathophysiology, diagnosis and treatment. *Front Biosci - Landmark*. 2017;22(3):385-406. doi:10.2741/4490
37. Paddison PJ, Caudy AA, Bernstein E, Hannon GJ, Conklin DS. Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev*. 2002 Apr 15;16(8):948-58. doi: 10.1101/gad.981002. PMID: 11959843; PMCID: PMC152352.
38. Paul CP, Good PD, Winer I, Engelke DR. Effective expression of small interfering RNA in human cells. *Nat Biotechnol*. 2002 May;20(5):505-8. doi: 10.1038/nbt0502-505. PMID: 11981566.
39. Drittanti, L., et al., *High throughput production, screening and analysis of adeno-associated viral vectors*. *Gene Ther*, 2000. **7**(11): p. 924-9.

CLAIMS

1. A method of integrating an exogenous DNA sequence into a genome of a cell comprising contacting the cell with:
 - a) a donor nucleic acid comprising:
 - said exogenous DNA sequence,
 - optionally one or more albumin exons,wherein said donor nucleic acid is flanked at 5' and 3' by inverted targeting sequences;
 - b) a complementary strand oligonucleotide homologous to a targeting sequence and
 - c) a nuclease that recognizes said targeting sequence,wherein said targeting sequence is located at the 3' end of the albumin gene in a region selected from intron 9, intron 11, intron 12, intron 13 and intron 14 of said albumin gene.
2. The method according to claim 1 wherein said donor nucleic acid comprises one or more albumin exons and said exon is exon 13 and/or exon 14 or fragments thereof.
3. The method according to claim 1 wherein said donor nucleic acid comprises one or more albumin exons and said exon is exon 10 and/or exon 11 and/or exon 12 and/or exon 13 and/or exon 14 or fragments thereof.
4. The method according to any one of previous claims wherein said complementary strand oligonucleotide homologous to a targeting sequence is a guide RNA that hybridizes to a targeting sequence, or to its complementary strand, located within intron 9, intron 11, intron 12, intron 13 or intron 14 of the albumin gene, preferably said guide RNA being adjacent to a protospacer-adjacent motif (PAM) sequence.
5. The method according to anyone of claims 1-4 wherein said albumin gene is a human or murine gene.
6. The method according to anyone of claims 1-5 wherein said complementary strand oligonucleotide homologous to the targeting sequence is under the control of a promoter, preferably the U6 promoter.
7. The method according to anyone of claims 1-6 wherein said exogenous DNA sequence is a coding sequence of the Arylsulfatase B (ARSB) gene, preferably said ARSB coding sequence comprises or has essentially a sequence having at least 95% of identity to SEQ ID NO 33.

8. The method according to anyone of claims 1-6 wherein said exogenous DNA sequence is a coding sequence of the Factor 8 (F8) gene, preferably said F8 coding sequence comprises or has essentially a sequence having at least 95% of identity to SEQ ID NO 36 or 55.
9. The method according to anyone of claims 1-8 wherein said inverted targeting sequence is an inverted sequence with respect to a target sequence located at the 3' end of the albumin gene in a region selected from intron 9, intron 11, intron 12, intron 13 and intron 14, preferably said inverted targeting sequence being linked at its 3' to a protospacer-adjacent motif (PAM) sequence.
10. The method according to anyone of claims 1-9 wherein said donor nucleic acid further comprises one or more of:
 - a post-transcriptional regulatory element, preferably localized at the 3' end of the exogenous DNA sequence;
 - a transcription termination sequence preferably localized at the 3' end of the post-transcriptional regulatory element or at the 3' end of the exogenous DNA sequence;
 - a splice acceptor sequence, preferably localized at the 3' end of the donor nucleic acid, for example linked to an albumin exon, if present;
 - a ribosomal skipping sequence, preferably localized between the exogenous DNA sequence and the albumin exon(s).
11. The method according to claim 10 wherein said ribosomal-skipping sequence is a T2A, P2A, E2A, F2A, preferably T2A sequence and/or said post-transcriptional regulatory element is the Woodchuck hepatitis virus post-transcriptional regulatory element (WPRE) and/or said transcription termination sequence is a poly-adenylation signal sequence, preferably the bovine growth hormone polyA (BGH polyA).
12. The method according to anyone of claims 1-10 wherein the targeting sequence comprises or has essentially a sequence having at least 95% of identity to any one of SEQ ID N. 1-2, SEQ ID N. 9-18 or functional fragments thereof and/or the complementary strand oligonucleotide homologous to the targeting sequence comprises or has essentially a sequence having at least 95% of identity to any one of SEQ ID NO: 1-2, SEQ ID N. 9-18 or functional fragments thereof.
13. The method according to anyone of claims 1-12 wherein said donor nucleic acid comprises:
 - a first inverted targeting sequence with its protospacer-adjacent motif (PAM) sequence;

- a splice acceptor sequence;
 - one or more albumin exons, preferably one or more exons selected from exon 13 and 14;
 - a ribosomal skipping sequence, preferably T2A;
 - the exogenous DNA sequence, preferably the coding sequence of human ARSB gene;
 - a transcription termination sequence; ; and
 - a second inverted targeting sequence with its protospacer-adjacent motif (PAM) sequence.
14. The method according to anyone of claims 1-13 wherein said nuclease is preferably selected from: a CRISPR nuclease, a TALEN, a DNA-guided nuclease, a meganuclease, and a Zinc Finger Nuclease, preferably said nuclease is a CRISPR nuclease selected from the group consisting of: Cas9, Cpf1, Cas12b (C2c1), Cas13a (C2c2), Cas3, Csf1, Cas13b (C2c6), and C2c3 or variants thereof such as SaCas9 or VQR-Cas9-HF1.
15. The method according to anyone of claims 1-14 wherein the cell is contacted with a nucleic acid encoding for said nuclease, preferably said nucleic acid coding for said nuclease is under the control of a tissue specific promoter, e.g. a liver specific hybrid liver promoter (HLP).
16. The method according to anyone of claims 1-15 wherein said complementary strand oligonucleotide, said donor nucleic acid and a nucleic acid encoding the nuclease are comprised in a viral or non-viral vector, preferably said viral vector being selected from: an adeno-associated virus, a lentivirus, a retrovirus and an adenovirus.
17. The method according to anyone of claims 1-16 wherein the cell is selected from the group consisting of: liver cells, one or more of lymphocytes, monocytes, neutrophils, eosinophils, basophils, endothelial cells, epithelial cells, hepatocytes, osteocytes, platelets, adipocytes, cardiomyocytes, neurons, retinal cells, smooth muscle cells, skeletal muscle cells, spermatocytes, oocytes, and pancreas cells, induced pluripotent stem cells (iPSCs), stem cells, hematopoietic stem cells, hematopoietic progenitor stem cells, preferably the cell is an hepatocyte of a subject.
18. The cell obtainable by the method of claims 1-17 for use for the treatment of a diseases wherein both the mutant and wildtype alleles are replaced with a correct copy of the gene provided by the donor DNA or for use for the treatment of a recessive inherited and common disease due to loss-of-function, preferably said disease being selected from haemophilia, diabetes, Lysosomal storage diseases comprising mucopolysaccharidoses, such as MPSI, MPSII, MPSIIIA, MPSIIIB, MPSIIIC, MPSIVA, MPSIVB, MPSVI and MPSVII,

sphingolipidoses, such as Fabry's Disease, Gaucher Disease, Nieman-Pick Disease and GM1 Gangliosidosis, lipofuscinoses, such as Batten's Disease, and mucopolidoses; gyrate atrophy of the choroid and retina, adenylosuccinate deficiency, hemophilia A and B, ALA dehydratase deficiency, adrenoleukodystrophy,.

19. A system comprising:

a) a donor nucleic acid comprising:

- an exogenous DNA sequence,
- optionally one or more albumin exons,

wherein said donor nucleic acid is flanked at 5' and 3' by inverted targeting sequences;

b) a complementary strand oligonucleotide homologous to a targeting sequence and

c) a nuclease that recognizes said targeting sequence,

wherein said targeting sequence is located at the 3' end of the albumin gene in a region selected from intron 9, intron 11, intron 12, intron 13 and intron 14.

20. The system according to claim 19 wherein said nuclease is codified by a nucleic acid and said donor nucleic acid, said complementary strand oligonucleotide and said nucleic acid codifying for the nuclease are located on DNA constructs, preferably said donor nucleic acid and said complementary strand oligonucleotide are located on the same DNA construct while said nucleic acid codifying for the nuclease is located on a separate DNA construct.

21. The system according to any one of claims 19-20 wherein the donor nucleic acid and/or the exogenous DNA sequence and/or the targeting sequences and/or the complementary strand oligonucleotide and/or the nuclease are as defined in any one claims 1-16.

22. The system according to any one of claims 19-21 wherein the complementary strand oligonucleotide and/or the donor nucleic acid and/or the nucleic acid encoding the nuclease are comprised in one or more viral or non-viral vectors, preferably said viral vector being selected from: an adeno-associated virus, a retrovirus, an adenovirus and a lentivirus.

23. The system according to claim 22 wherein it comprises a first vector comprising a nucleic acid expressing a nuclease and a second vector comprising the donor nucleic acid and the complementary strand oligonucleotide homologous to the targeting sequence, wherein such elements are as defined in any one claims 1-16.

24. The system according to claim 22 wherein it comprises a first vector comprising the donor nucleic acid and a second vector comprising the complementary strand oligonucleotide homologous to a targeting sequence and the nucleic acid coding for the nuclease, wherein such elements are as defined in any one claims 1-16.
25. The system according to any one of claims 19-24 for use as a medicament.
26. A vector that comprises the donor nucleic acid and/or the complementary strand oligonucleotide homologous to the targeting sequence and/or a nucleic acid coding for a nuclease that recognizes the targeting sequence as defined in any one of claims 1-16.
27. The vector according to claim 26, wherein the vector is a viral vector, preferably a lentiviral vector or an adeno-associated vector, or a non-viral vector, preferably selected from a polymer-based, particle-based, lipid-based, peptide-based delivery vehicle or combinations thereof, such as cationic polymers, micelles, liposomes, exosomes, microparticles and nanoparticles including lipid nanoparticles (LNP).
28. The vector according to claim 26 or 27 further comprising 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, preferably the ITRs derive from the same virus serotype or from different virus serotypes, preferably the virus is an AAV, preferably of serotype 2.
29. A host cell comprising the system according to any one of claims 19-25, or a vector according to any of claims 26-28.
30. A viral particle that comprises the system according to any one of claims 19-25, or a vector according to any of claims 26-28.
31. The viral particle according to claim 30, wherein the viral particle comprises capsid proteins of an AAV.
32. The viral particle according to claim 31, wherein the viral particle comprises capsid proteins of an AAV of a serotype selected from one or more of the group consisting of AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8 AAV9 and AAV 10, preferably from the AAV2 or AAV8 serotype.
33. A pharmaceutical composition that comprises one of the following: system according to any one of claims 19-25, or one or more vectors according to any of claims 26-28, a host cell

- according to claim 29, a viral particle according to any of claims 30-32, and a pharmaceutically acceptable carrier.
34. A kit comprising : system according to any one of claims 19-25, or one or more vectors according to any of claims 26-28, a host cell according to claim 29, a viral particle according to any of claims 30-32 or a pharmaceutical composition according to claim 33 in one or more containers, optionally further comprising instructions or packaging materials that describe how to administer the nucleic acid construct, vector, host cell, viral particle or pharmaceutical composition to a patient.
35. The system according to any one of claims 19-25, or a vector according to any of claims 26-28, a host cell according to claim 29, a viral particle according to any of claims 30-32 or a pharmaceutical composition according to claim 33 for use as a medicament.
36. The system according to any one of claims 29-25, or a vector according to any of claims 26-28, a host cell according to claim 29, a viral particle according to any of claims 30-32 or a pharmaceutical composition according to claim 33 for use in the treatment of hepatic diseases, Lysosomal storage diseases comprising mucopolysaccharidoses, such as MPSI, MPSII, MPSIIIA, MPSIIIB, MPSIIIC, MPSIVA, MPSIVB, MPSVI and MPSVII, sphingolipidoses, such as Fabry's Disease, Gaucher Disease, Nieman-Pick Disease and GM1 Gangliosidosis, lipofuscinoses, such as (Batten's Disease, and mucolipidoses; other diseases where the liver can be used as a factory for production and/or secretion of therapeutic proteins, like diabetes, gyrate atrophy of the choroid and retina, adenylosuccinate deficiency, hemophilia A and B, ALA dehydratase deficiency, adrenoleukodystrophy.
37. Use of the system according to any one of claims 19-25, or a vector according to any of claims 26-28 for the production of viral particles.
38. A DNA construct comprising the donor nucleic acid and/or the complementary strand oligonucleotide homologous to a targeting sequence and/or a nucleic acid coding for a nuclease that recognizes said targeting sequence, as defined in any one of claims 1-16.

A

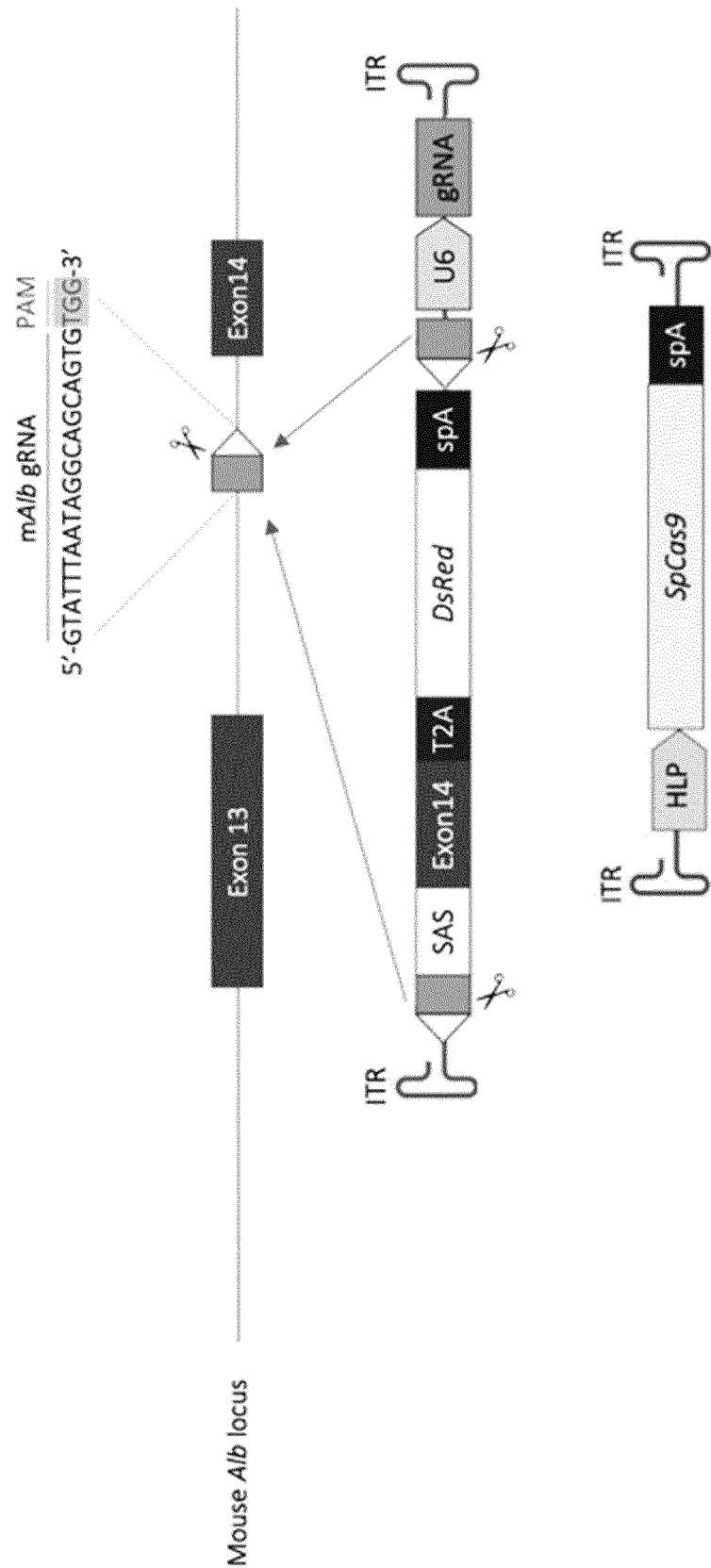
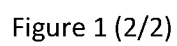


Figure 1 (1/2)



A

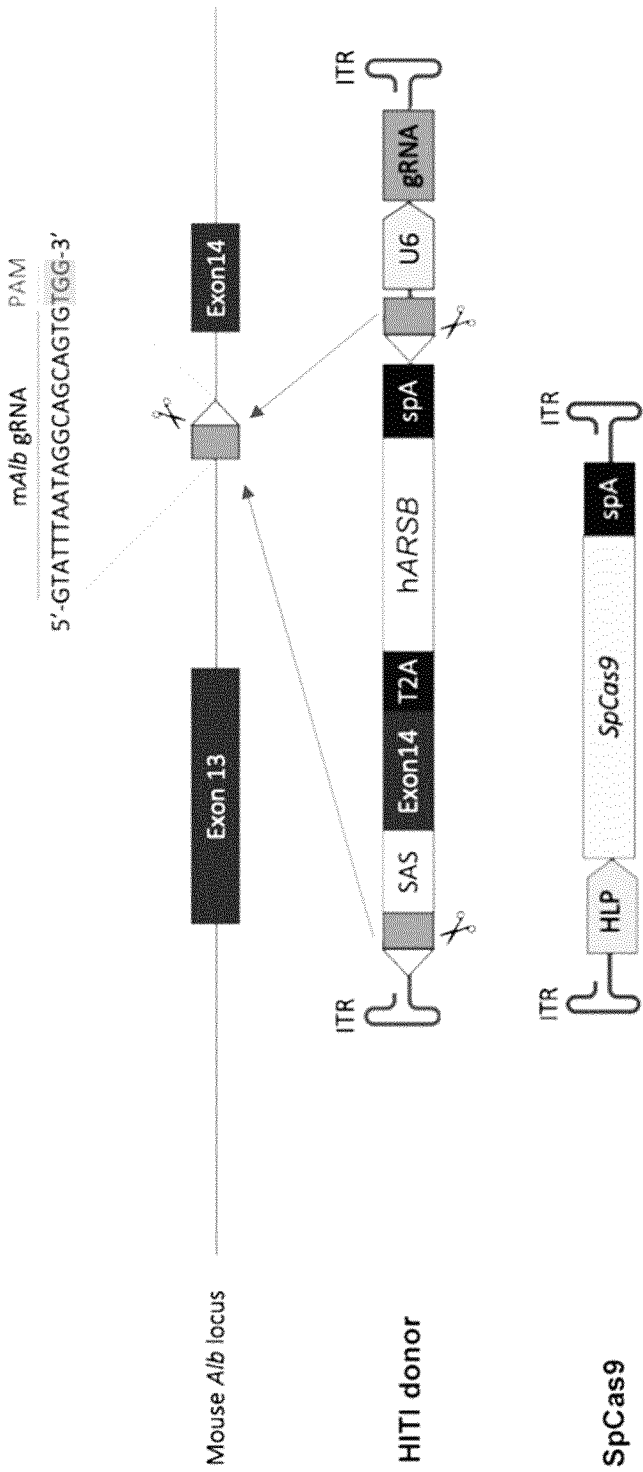


Figure 2 (1/2)

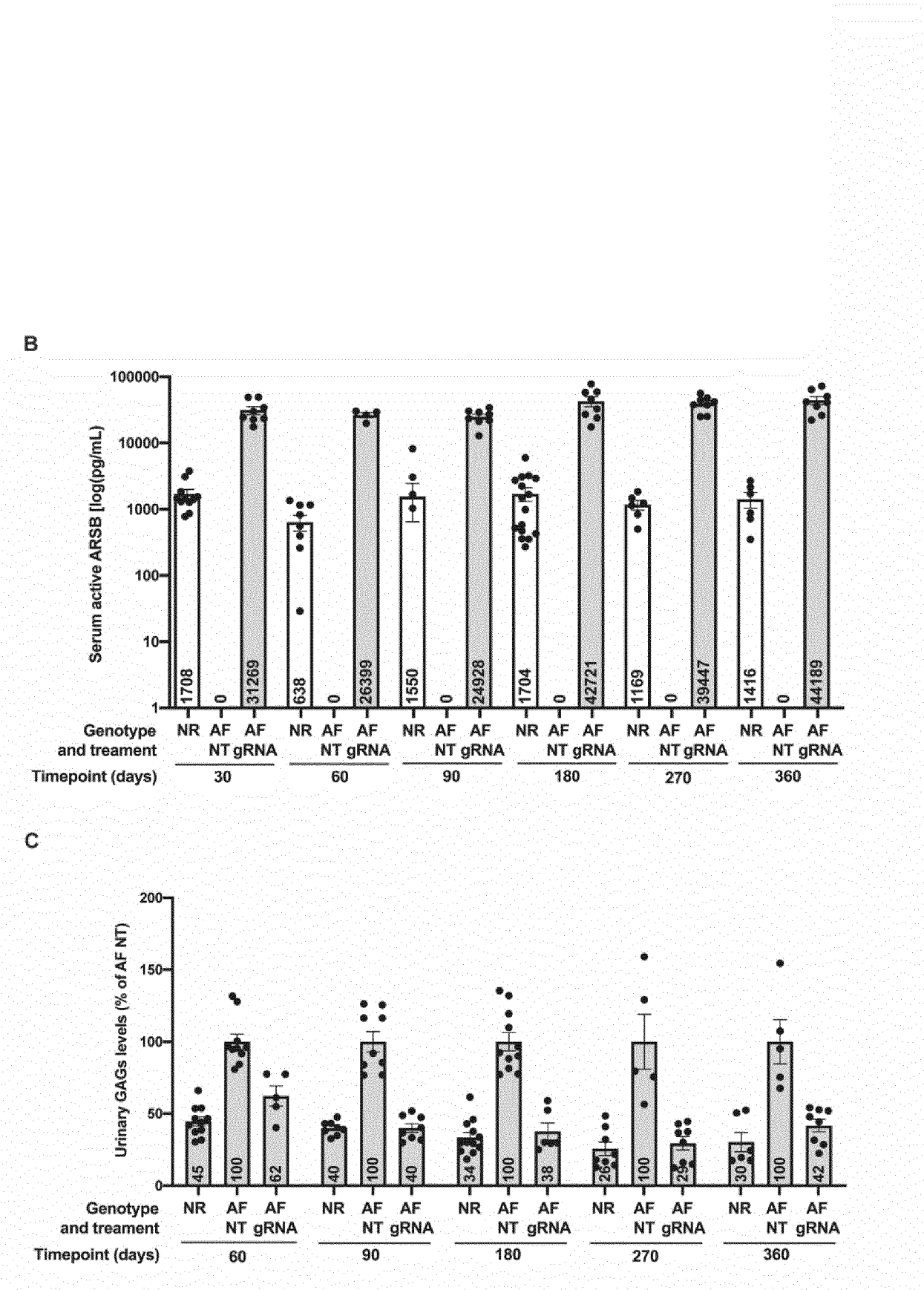


Figure 2 (2/2)

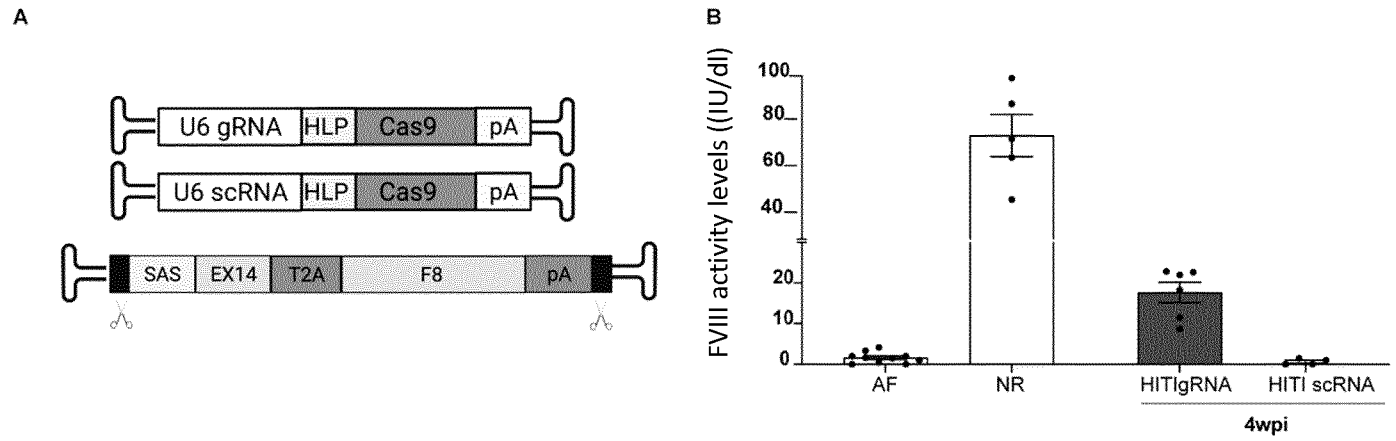


Figure 3

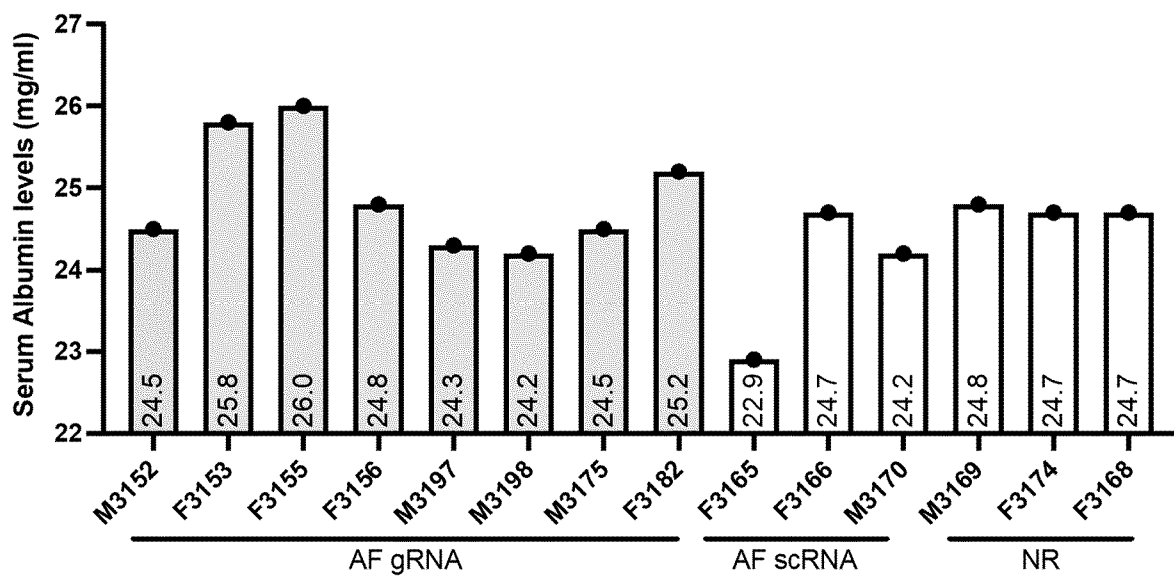


Fig. 4

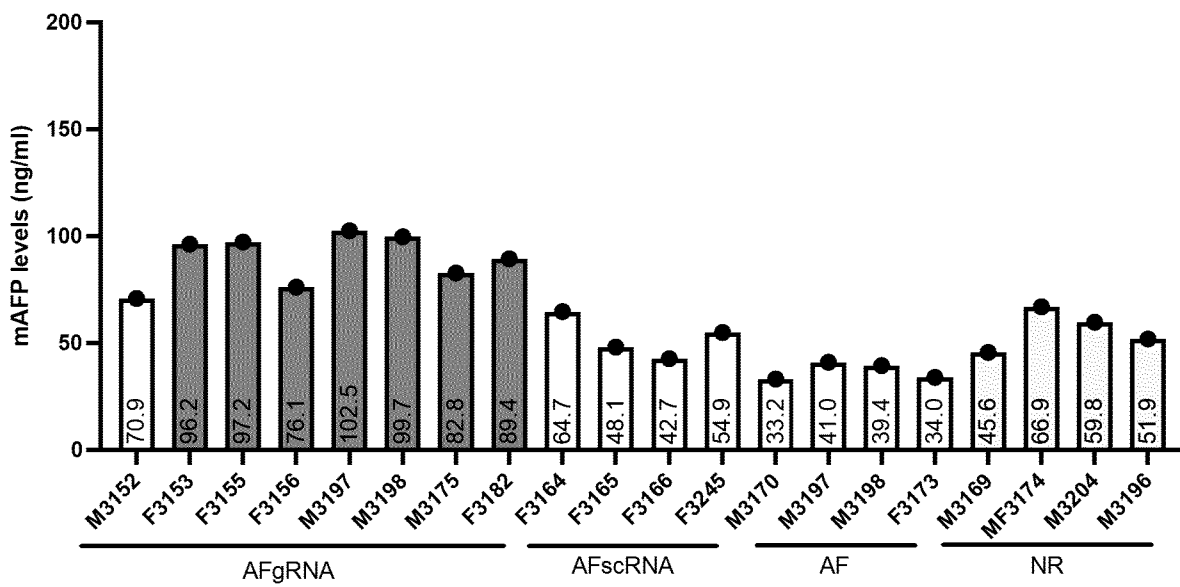


Fig. 5

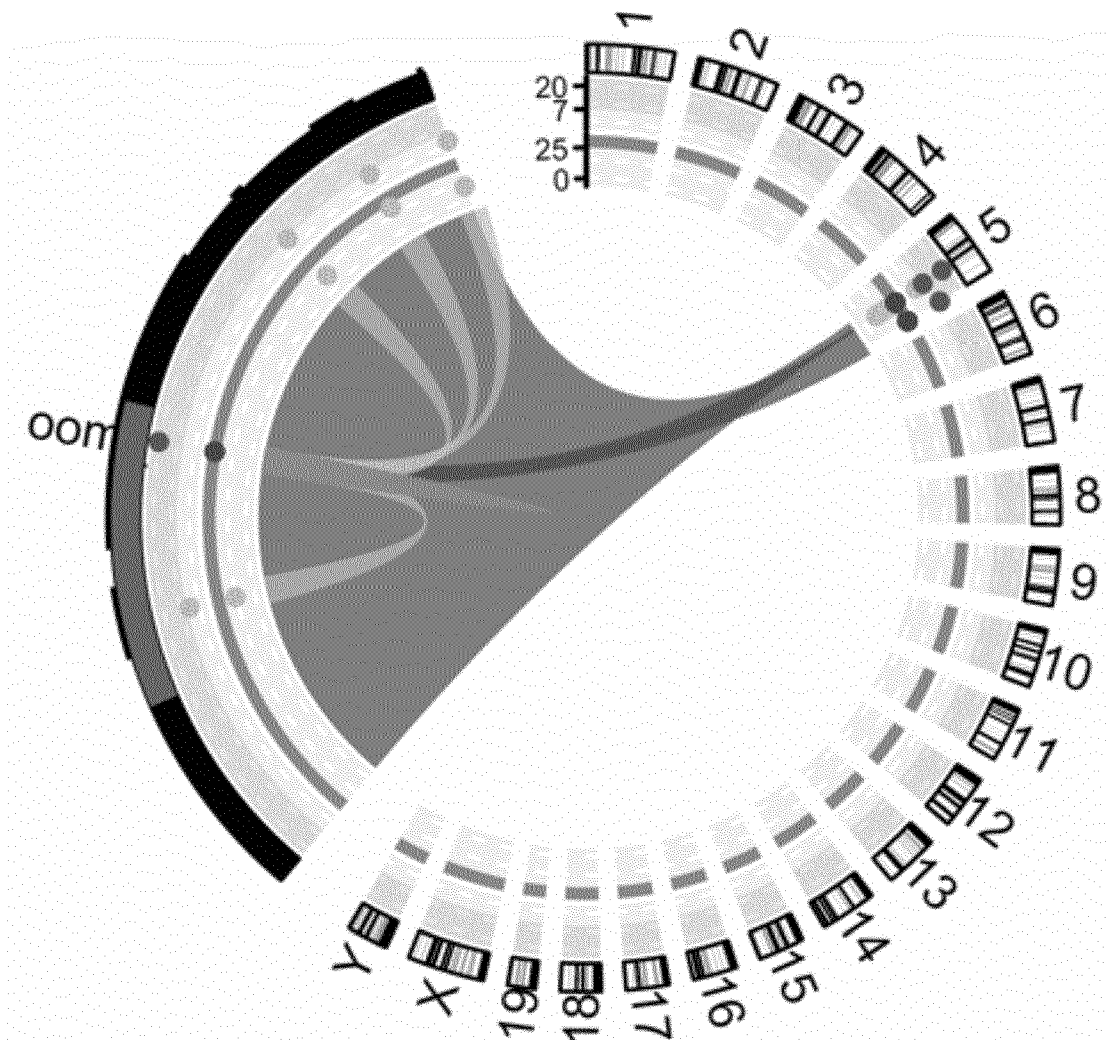


Fig. 6

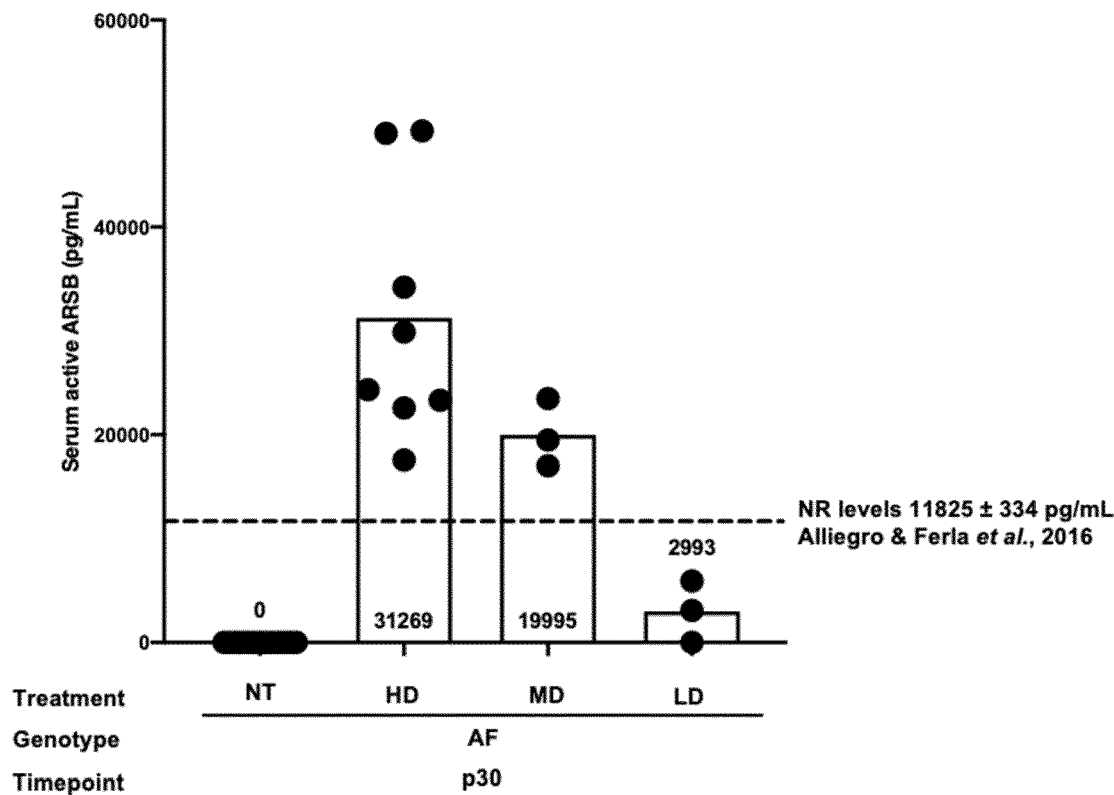


Fig. 7

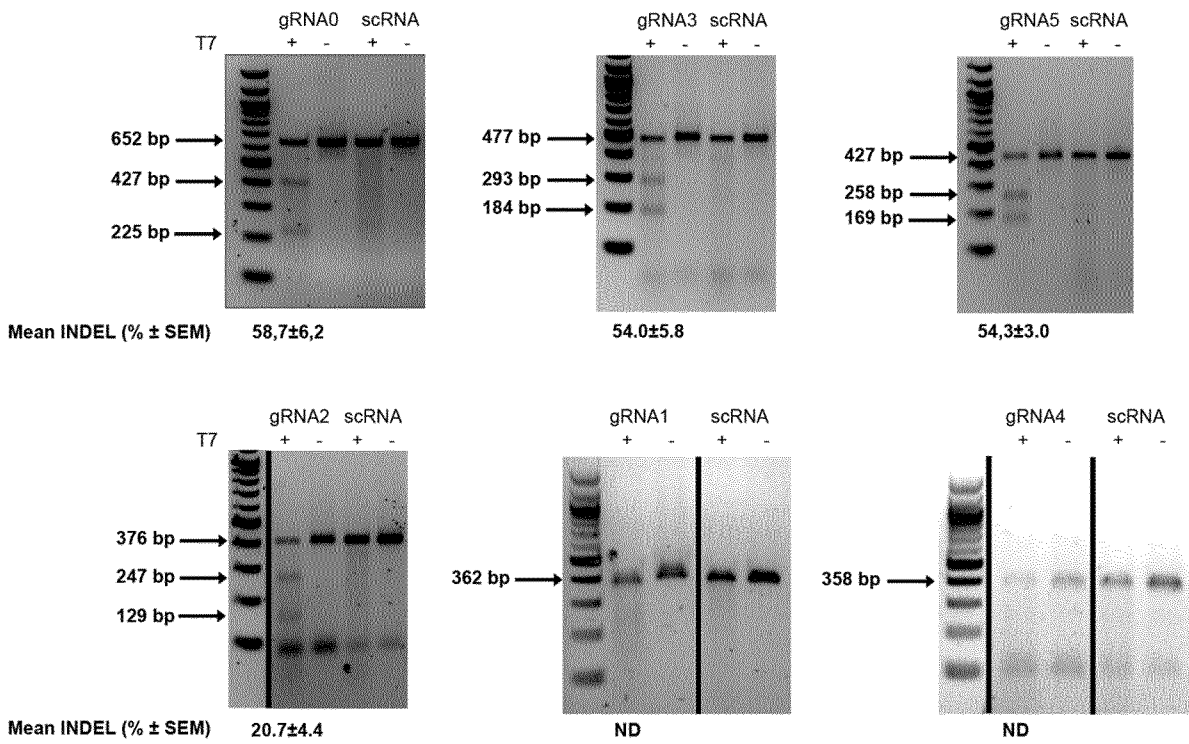


Fig. 8

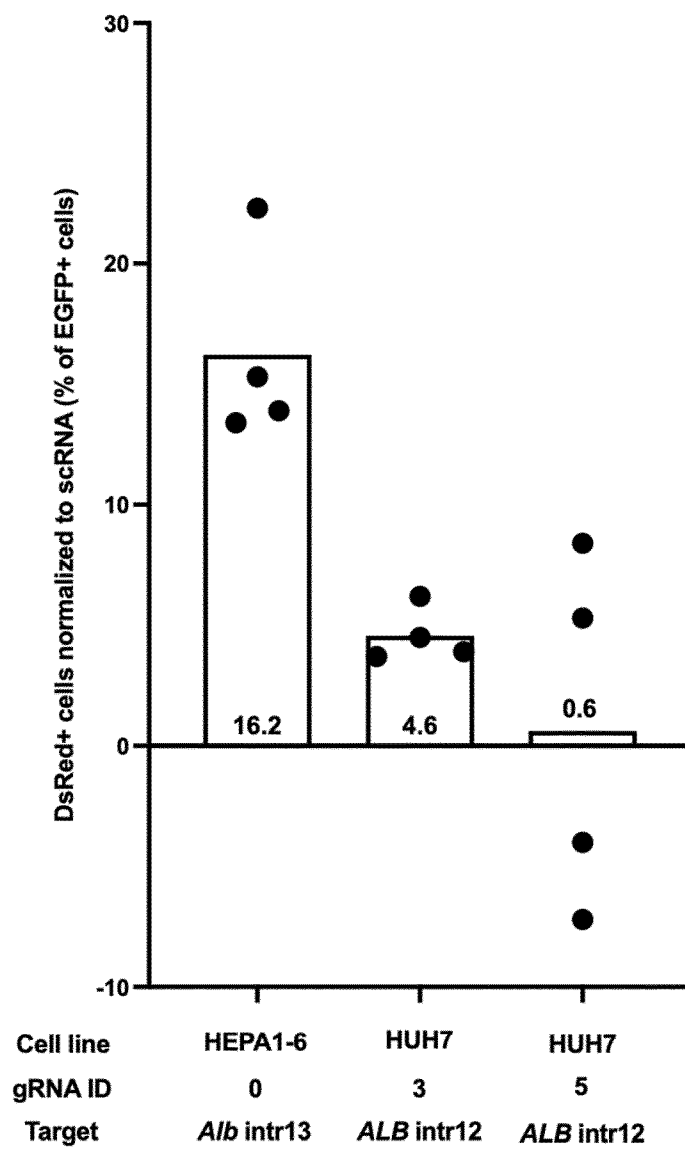


Fig. 9

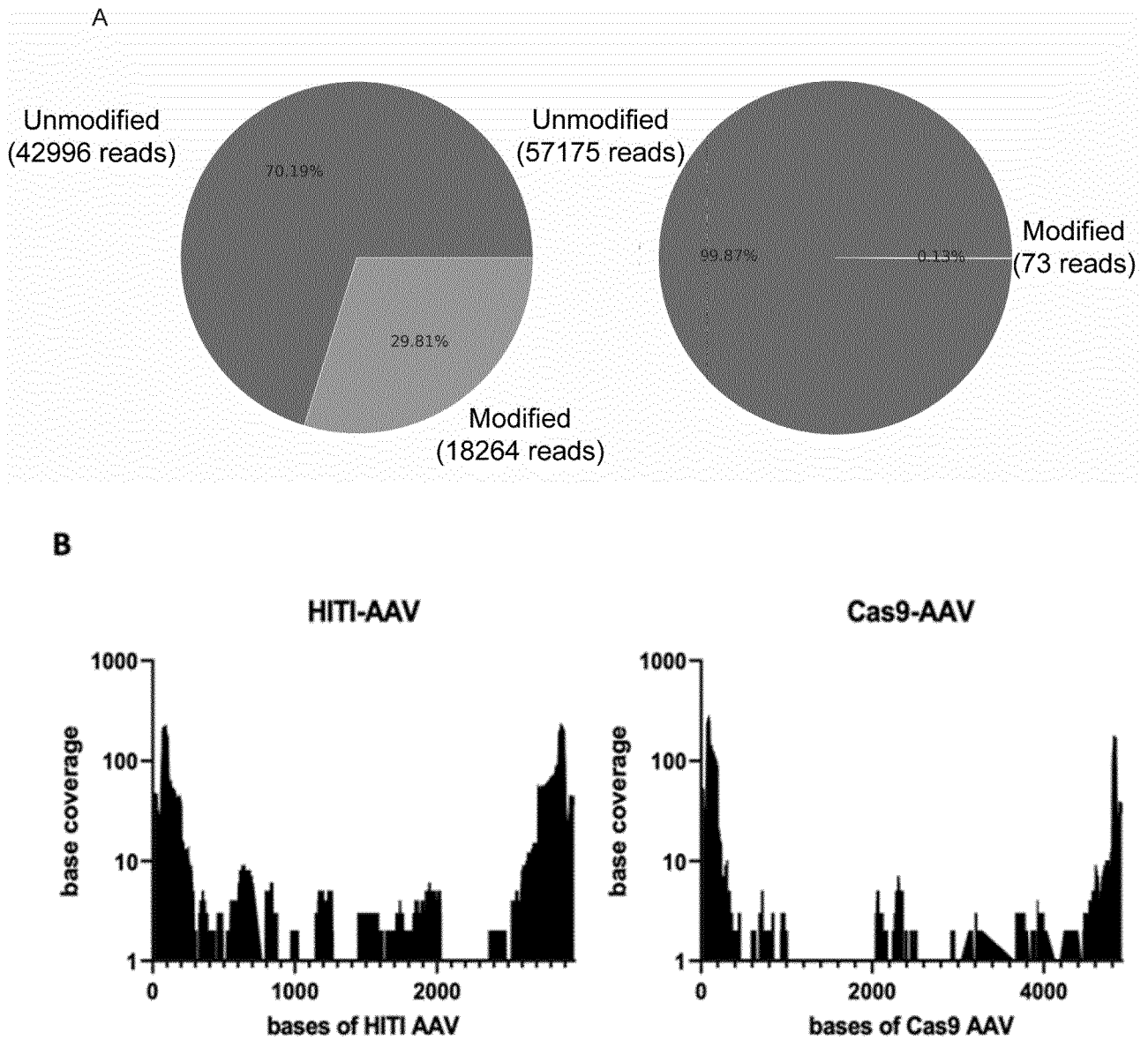
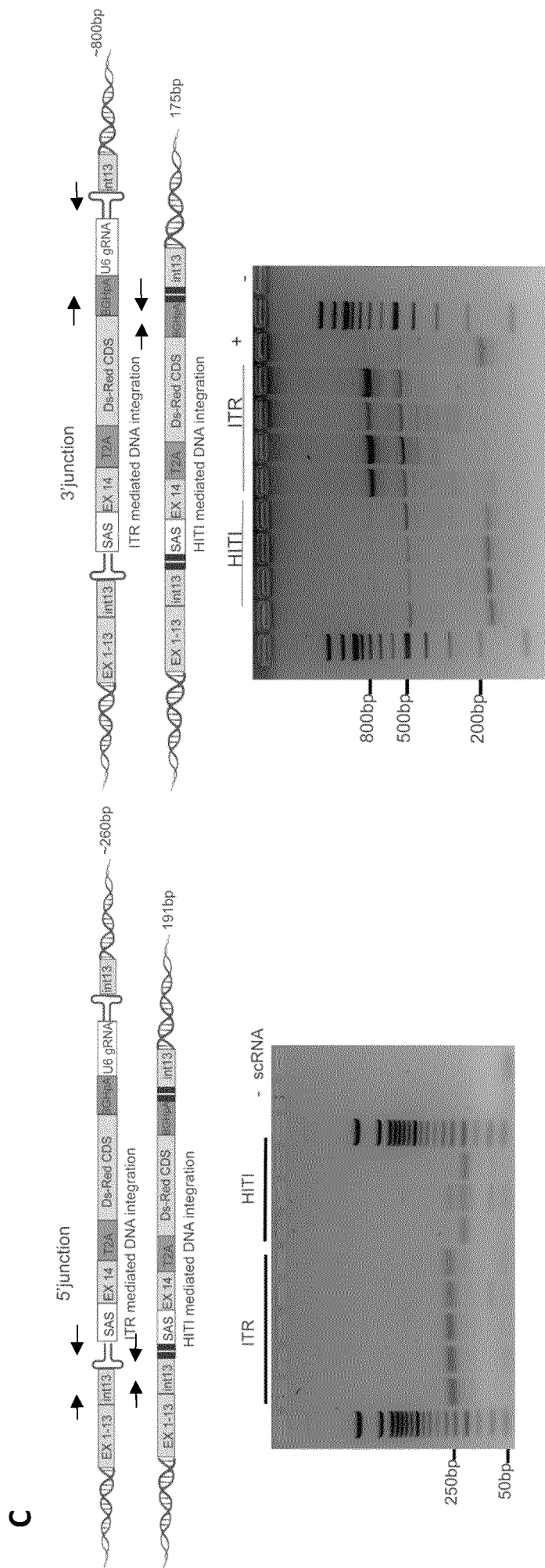


Fig. 10 (1/3)



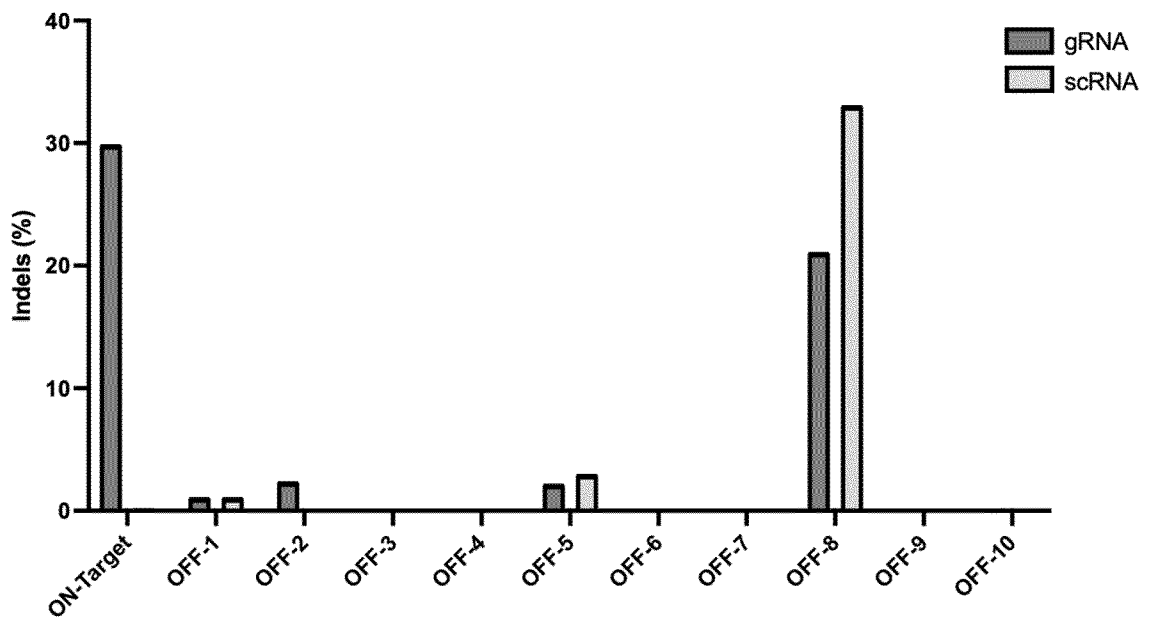
D

Fig. 10 (3/3)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/EP2023/061582

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
 - a. ☒ forming part of the international application as filed.
 - b. ☐ furnished subsequent to the international filing date for the purposes of international search (Rule 13*ter*.1(a)).

☐ accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.
2. ☐ With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.
3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2023/061582

A. CLASSIFICATION OF SUBJECT MATTER INV. A61K48/00 C12N15/11 C12N15/90 ADD.														
According to International Patent Classification (IPC) or to both national classification and IPC														
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) A61K C12N														
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched														
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) EPO-Internal														
C. DOCUMENTS CONSIDERED TO BE RELEVANT <table border="1"> <thead> <tr> <th>Category*</th> <th>Citation of document, with indication, where appropriate, of the relevant passages</th> <th>Relevant to claim No.</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>CN 112 741 906 A (UNIV EAST CHINA NORMAL; BIORAY LABORATORIES INC) 4 May 2021 (2021-05-04)</td> <td>1-6, 9-36, 38</td> </tr> <tr> <td>Y</td> <td>the whole document -----</td> <td>7, 8, 37</td> </tr> <tr> <td>Y</td> <td>WO 2020/079033 A1 (FOND TELETHON [IT]) 23 April 2020 (2020-04-23) cited in the application page 2, lines 14-17 page 4, lines 17-20 page 10, lines 28-29 ----- -/-</td> <td>7, 37</td> </tr> </tbody> </table>			Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.	X	CN 112 741 906 A (UNIV EAST CHINA NORMAL; BIORAY LABORATORIES INC) 4 May 2021 (2021-05-04)	1-6, 9-36, 38	Y	the whole document -----	7, 8, 37	Y	WO 2020/079033 A1 (FOND TELETHON [IT]) 23 April 2020 (2020-04-23) cited in the application page 2, lines 14-17 page 4, lines 17-20 page 10, lines 28-29 ----- -/-	7, 37
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.												
X	CN 112 741 906 A (UNIV EAST CHINA NORMAL; BIORAY LABORATORIES INC) 4 May 2021 (2021-05-04)	1-6, 9-36, 38												
Y	the whole document -----	7, 8, 37												
Y	WO 2020/079033 A1 (FOND TELETHON [IT]) 23 April 2020 (2020-04-23) cited in the application page 2, lines 14-17 page 4, lines 17-20 page 10, lines 28-29 ----- -/-	7, 37												
☒ Further documents are listed in the continuation of Box C. ☒ See patent family annex.														
* Special categories of cited documents : <table border="0"> <tr> <td style="vertical-align: top;"> "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed </td> <td style="vertical-align: top;"> "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance;; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance;; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family </td> </tr> </table>			"A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance;; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance;; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family										
"A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance;; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance;; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family													
Date of the actual completion of the international search 16 August 2023		Date of mailing of the international search report 30/08/2023												
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016		Authorized officer Knudsen, Henrik												

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2023/061582

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	ZHANG JIAN-PING ET AL: "Curing hemophilia A by NHEJ-mediated ectopic F8 insertion in the mouse", GENOME BIOLOGY, vol. 20, no. 1, 1 December 2019 (2019-12-01), pages 276-1, XP055980843, DOI: 10.1186/s13059-019-1907-9 Retrieved from the Internet: URL:https://genomebiology.biomedcentral.com/counter/pdf/10.1186/s13059-019-1907-9.pdf> page 11, left-hand column, paragraph 2 -----	8
X	CN 113 058 041 A (UNIV EAST CHINA NORMAL; BIORAY LABORATORIES INC) 2 July 2021 (2021-07-02) the whole document -----	1-5
X	WO 2020/206162 A1 (REGENERON PHARMA [US]) 8 October 2020 (2020-10-08) paragraph [0201]; claim 1 -----	1
A	WO 2018/013932 A1 (SALK INST FOR BIOLOGICAL STUDI [US]; TSUNEKAWA YUJI [JP]) 18 January 2018 (2018-01-18) paragraph [0017]; figure 5 -----	1
A	WO 2014/011237 A1 (SANGAMO BIOSCIENCES INC [US]) 16 January 2014 (2014-01-16) tables 3-4 -----	1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2023/061582

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
CN 112741906 A	04-05-2021	NONE	
WO 2020079033 A1	23-04-2020	CA 3116452 A1 CN 113348247 A EP 3867375 A1 JP 2022505139 A US 2022001030 A1 WO 2020079033 A1	23-04-2020 03-09-2021 25-08-2021 14-01-2022 06-01-2022 23-04-2020
CN 113058041 A	02-07-2021	NONE	
WO 2020206162 A1	08-10-2020	AU 2020256225 A1 BR 112021019512 A2 CA 3133361 A1 CL 2021002534 A1 CN 113727603 A CO 2021012676 A2 EP 3945800 A1 IL 286865 A JP 2022527809 A KR 20210148154 A SG 11202108451V A US 2020318136 A1 WO 2020206162 A1	02-09-2021 15-02-2022 08-10-2020 29-04-2022 30-11-2021 20-10-2021 09-02-2022 31-10-2021 06-06-2022 07-12-2021 29-09-2021 08-10-2020 08-10-2020
WO 2018013932 A1	18-01-2018	AU 2017295898 A1 CA 3030587 A1 EP 3485023 A1 JP 2019524098 A JP 2023092453 A US 2019225991 A1 WO 2018013932 A1	21-02-2019 18-01-2018 22-05-2019 05-09-2019 03-07-2023 25-07-2019 18-01-2018
WO 2014011237 A1	16-01-2014	AU 2013289206 A1 CA 2875618 A1 DK 2872625 T3 DK 3196301 T3 DK 3444342 T3 EP 2872625 A1 EP 3196301 A1 EP 3444342 A1 EP 3816281 A1 ES 2613691 T3 ES 2697912 T3 ES 2813080 T3 HK 1209788 A1 HU E030799 T2 HU E041553 T2 HU E051612 T2 JP 6225182 B2 JP 6587666 B2 JP 7218401 B2 JP 2015527881 A JP 2018038410 A JP 2019201672 A JP 2021168669 A JP 2023038332 A US 2014017212 A1	15-01-2015 16-01-2014 06-02-2017 14-01-2019 24-08-2020 20-05-2015 26-07-2017 20-02-2019 05-05-2021 25-05-2017 29-01-2019 22-03-2021 08-04-2016 28-06-2017 28-05-2019 01-03-2021 01-11-2017 09-10-2019 06-02-2023 24-09-2015 15-03-2018 28-11-2019 28-10-2021 16-03-2023 16-01-2014

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2023/061582

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
		US 2014112896 A1	24-04-2014
		US 2018110808 A1	26-04-2018
		US 2018214490 A1	02-08-2018
		US 2019224246 A1	25-07-2019
		US 2022331447 A1	20-10-2022
		WO 2014011237 A1	16-01-2014
