(54) Title: HUMAN ALPHA GALACTOSIDASE A CODING SEQUENCE FOR THE TREATMENT OF FABRY DISEASE

(57) Abstract: The present invention relates to an optimized coding sequence for the expression of human alpha galactosidase A
(GLA) which can be useful for the treatment of Fabry disease. Nucleic acid constructs, vectors, plasmids, host cells and pharmaceutical
compositions including such sequence are also objects of the invention as well as medical uses thereof.

WO 2024/184376 A1

HUMAN ALPHA GALACTOSIDASE A CODING SEQUENCE FOR THE TREATMENT OF FABRY DISEASE

FIELD OF THE INVENTION

The present invention relates to a coding sequence for human alpha galactosidase A (GLA) which can be useful for the treatment of Fabry disease. Vectors, plasmids, host cells and pharmaceutical compositions including such sequence are also objects of the invention.

BACKGROUND OF THE INVENTION

Fabry disease

Fabry disease (OMIM 301500) (FD) is a rare lysosomal disorder often reminded of as Anderson-Fabry disease after the researchers who brought light to it (Anderson, 1898). It may be also called alpha-galactosidase A deficiency or GLA deficiency, angiokeratoma corporis diffusum, diffuse angiokeratoma, ceramide trihexosidosis, and hereditary dystopic lipidosis. It is an X-linked inborn error of metabolism which occurs due to the deficiency of Alpha galactosidase destined to catabolize Globotrioasylceramide or Gb3, instead progressively accumulates it in the lysosomes (Brady et al., 1967; Zarate and Hopkin, 2008). Lysosomal storage disorders (LSD) are a group of over 70 metabolic inherited disorders with a frequency of 1 in 5000 patients collectively though individually they stand rare in the population. Fabry disease is the second most common LSD and topping the list is the Gaucher disease. Others include Niemann-Pick disease, Hunter Syndrome, Pompe disease, and Tay-Sach disease (Parkinson-Lawrence et al., 2010).

FD is caused by mutations in the GLA gene. According to fabry-database.org as of 2019, 1993 mutations have been reported in the gene including 1518 missense and 297 nonsense accounting for a large percentage of the mutations (Eng and Desnick, 1994; Garman and Garboczi, 2004; Saito et al., 2011). However, it is crucial to note that not all of them lead to a defective GLA protein but more than 800 mutations have already been held responsible for the condition (Eng et al., 1993). Physiologically, the Galactosidase alpha enzyme is active in the lysosome where it functions to degrade globotriaosylceramide or Gb3 lipid molecules to smaller particles as part of the lysosome recycling process. In the diseased condition, the mutated GLA gene translates to a dysfunctional protein, leading to the deficiency of alpha-Galactosidase enzyme restricting the cleavage of Gb3 molecules and hence results in the accumulation of these lipid moieties in cells of various organs (Aerts et al., 2008; Askari et al., 2007; Auray-Blais et al., 2008; van der Tol et al., 2014).

Fabry disease (FD) affects most organs in a non-specific manner, some with common symptoms to

1

make the disease untraceable and hence unnoticed to a large extent. Suffering patients experience burning sensations and episodes of pain in the limbs, dark red spots on the skin, their eyes undergo impairment, there is a loss of hearing and ability to sweat. FD comes along with various gastrointestinal and cardiac complexities. There are two variants of FD: the classical variant or classical FD, and the atypical or late-onset FD. The classical form of FD objectifies patients with mutations causing minimal or mostly no residual enzyme activity affecting the quality of life from an early age and reducing life expectancy. Children with the classical FD form suffer intolerance to heat, skin abnormalities, gastrointestinal problems, and burning pain in the limbs known as Fabry crises which often extend to renal and cardiac symptoms towards adulthood (Eng et al., 2007; Mehta et al., 2004).

The late-onset variant of Fabry disease relates to patients with residual α-Gal activity (3-30%). This variant is generally tagged in with organ-specific manifestations in adulthood like the heart. Renal and cardiac variants comprise a major percentage of late-onset FD patients (Desnick, 2013).

Epidemiology

Fabry disease is a rare lysosomal storage disorder with an incidence of 1:8,454 to 1:117,000 in males, which is however highly underestimated. Since the disease involves multiple organs and varied clinical manifestations with variables like age and sex, it often remains underdiagnosed in a common population. The true prevalence of the disease should be studied regionally due to its pan-ethnic nature, which creates an alarming inconsistency in the number of cases globally. In Taiwan, a newborn screening predicts an unexpected 1 in 1,500 FD affected males, whereas in Italy the numbers reach up to 1 in 3,100 males (Spada et al., 2006). Another study in the Netherlands involving pre-natal and post-natal diagnosis accounts for 1 in 500,000 FD patients. Similar numbers were diagnosed in the UK; 1 in 300,000 as well. This discrepancy has led to a misjudged prevalence figures in the community (Germain, 2010).

Diagnosis

Due to complex symptoms of the disease which vary from each individual the diagnosis is often delayed and mostly neglected. It is crucial and perchance rare for a general practitioner to identify a patient's symptoms as those of Fabry disease. In case the patient is suspected to be suffering from Fabry disease, genetic and enzymatic evaluations are done. Clinical tests for the activity of the α-Galactosidase enzyme in plasma are done for both classical and non-classical FD. Males suffering from the classical form of FD indicate 1-3% of enzyme activity whereas in patients with late-onset

2

FD activity values range from 3-30% in which case the confirmed diagnosis calls for genetic evaluation. Genetic testing confirms the type of mutation in the enzyme locus and its repercussions. This type of genetic testing is recommended for females since enzymatic assay is most likely to indicate residual activity. Furthermore, genetic counselling or familial studies have helped diagnose Fabry disease. In families with this X-linked disorder, prenatal evaluation is done by the analysis of amniotic fluids and gene sequencing for suspected mutations (Bernardes et al., 2020; Laney and Fernhoff, 2008).

Available Treatments

Currently, Enzyme Replacement Therapy (ERT) (Bengtsson et al., 2003), and more recently, Chaperon Therapy products have seen some success in the treatment of FD. The ERT therapeutics (Ioannou et al., 2001), Shire's Replagal (Agalsidase alpha) (Pastores, 2007), and Sanofi's Fabrazyme (Agalsidase beta) (Germain et al., 2015; Lubanda et al., 2009) cost, roughly, $300.000-315.000/year. The annual cost of Galafold, the first oral chaperone therapy capsule from Amicus Therapeutics, is similar to ERT (Markham, 2016; Moran, 2018), but it can only target a sub-group of patients. The cost of these treatments is a concerning factor that encourages the need for novel and cheaper treatments that can be affordable by common people (Motabar et al., 2010).

Therefore, there is still the need of a treatment for Fabry disease which is efficient and more cost convenient with respect to current therapies.

SUMMARY OF THE INVENTION

It has now been found a codon optimized cDNA sequence expressing human GLA able to efficiently produce in vivo the enzyme. In particular, the use of this sequence advantageously improves GLA enzyme translation of about 4-6 fold over the wild-type sequence coding for GLA.

The codon optimization was initially done using online codon-optimization tools and then it was manually adjusted for further optimization. For example, potential cryptic splice sites and alternative reading frames, that may result in undesired polypeptides able to stimulate the immune system against the transduced cells, were removed. With this further step of optimization, the safety and the efficiency of the gene therapy treatment was increased, in particular by reducing the generation of undesired variants that may trigger an immune response against the transduced cells. Thus, the sequences of the invention result in a safer transgene with increased efficiency when administered to a subject.

It is an object of the invention a nucleic acid comprising or consisting of the following sequence of SEQ ID N.1:

ATGCAGCTGCGCAACCCCGAGCTGCACCTGGGCTGCGCCCTGGCCCTGCGCTTCCTGGCCCTGGTCAGCTGG

GACATCCCCGGCGCCCGCGCCCTGGACAACGGCCTGGCCCGCACCCCCACCATGGGCTGGCTGCACTGGGA

GCGCTTCATGTGCAACCTGGACTGCCAGGAGGAGCCCGACAGCTGCATCAGCGAGAAGCTGTTTATGGAGAT

GGCCGAGCTGATGGTCAGCGAGGGCTGGAAGGACGCCGGCTACGAGTACCTGTGCATCGACGACTGCTGGA

TGGCCCCCCAGCGCGACAGCGAGGGCCGCCTGCAGGCCGACCCCCAGCGCTTCCCCCACGGAATCCGCCAG

CTGGCCAACTACGTGCACAGCAAGGGCCTGAAGCTGGGCATCTACGCCGACGTGGGCAACAAGACCTGCGC

CGGCTTCCCCGGCAGCTTCGGCTACTACGACATCGACGCCCAGACCTTCGCCGACTGGGGCGTGGACCTGCT

GAAGTTCGACGGCTGCTACTGCGACAGCCTGGAGAACCTGGCCGACGGCTACAAGCACATGAGCCTGGCCC

TGAACCGCACCGGCCGCAGCATCGTGTACAGCTGCGAGTGGCCCCTGTATATGTGGCCCTTCCAGAAGCCCAA

CTACACCGAGATCCGCCAGTACTGCAACCACTGGCGCAACTTCGCCGACATCGACGACAGCTGGAAGAGCAT

CAAGAGCATCCTGGACTGGACCAGCTTCAACCAGGAGCGCATCGTGGACGTGGCCGGCCCCGGCGGCTGGA

ACGACCCCGACATGCTGGTGATCGGCAACTTCGGCCTGAGCTGGAACCAGCAGGTGACCCAGATGGCCCTGT

GGGCCATTATGGCCGCCCCCCTGTTTATGAGCAACGACCTGCGCCACATCAGCCCCCAGGCCAAGGCCCTGCT

GCAGGACAAGGACGTGATCGCTATCAACCAGGACCCCCTGGGCAAGCAGGGCTACCAGCTGCGCCAGGGCG

ACAACTTCGAGGTCTGGGAGCGCCCCCCTGAGCGGCCTGGCCTGGGCCGTGGCTATGATCAACCGCCAGGAG

ATCGGCGGCCCCCGCAGCTACACCATCGCCGTGGCCAGCCTGGGCAAGGGCGTGGCCTGCAACCCCGCCTG

CTTCATCACCCAGCTGCTGCCCGTGAAGCGCAAGCTGGGCTTCTACGAGTGGACCAGCCGCCTGCGCAGCCA

CATCAACCCCACCGGCACCGTGCTGCTGCAGCTGGAGAACACAATGCAGATGAGCCTGAAGGACCTGCTGTA

A [SEQ ID N.1],

or comprising or consisting of a sequence having at least 99.5% of identity with the sequence of SEQ ID N.1.

Preferably, said nucleic acid comprises or consists of a sequence having 99.5, 99.6, 99.7, 99.8, 99.9 or 100% identity with the sequence of SEQ ID N.1.

It is also an object of the invention a nucleic acid comprising or consisting of the following sequence with SEQ ID N.2:

ATGCAGTTGAGAAACCCAGAGCTCCACCTGGGCTGTGCCCTGGCACTGAGGTTCCTGGCCCTTGTGAGCTGG

GATATCCCTGGGGCCAGGGCCTTGGACAACGGCTTGGCCCGCACCCCCACAATGGGCTGGCTGCACTGGGA

ACGCTTTATGTGCAATCTGGACTGCCAGGAGGAGCCTGACAGCTGTATCAGCGAGAAGCTCTTTATGGAGATG

GCAGAGCTGATGGTGTCTGAGGGATGGAAGGACGCCGGCTACGAATACCTGTGCATTGACGATTGCTGGATG

GCTCCACAGAGGGACTCAGAAGGACGCCTGCAGGCTGATCCCCAGAGATTCCCCCATGGAATCCGCCAGCTG

GCCAACTATGTGCACAGCAAAGGCCTGAAGCTGGGCATCTACGCCGACGTGGGCAACAAGACCTGTGCTGG

CTTCCCTGGCTCCTTTGGATATTACGATATCGACGCTCAGACCTTTGCTGACTGGGGAGTGGATCTCCTCAAGT

TTGACGGCTGCTACTGTGACTCTCTGGAAAACCTGGCAGATGGCTACAAGCACATGTCCCTGGCTCTGAACAG

AACAGGCCGCAGCATTGTGTACAGCTGCGAGTGGCCCCTGTATATGTGGCCCTTCCAGAAGCCCAACTACACA

GAGATCAGGCAGTACTGCAACCACTGGAGGAACTTTGCCGACATTGACGACTCCTGGAAATCTATCAAGTCTA

TCCTGGATTGGACATCCTTCAACCAAGAGCGGATCGTGGACGTGGCTGGACCTGGAGGCTGGAATGATCCAG

ATATGCTGGTGATTGGAAACTTCGGGCTGTCTTGGAACCAGCAGGTCACTCAGATGGCGCTGTGGGCCATCAT

GGCCGCCCCCCTCTTTATGAGCAACGACCTGCGCCACATTTCTCCTCAAGCCAAGGCCCTGCTCCAGGACAAG

GACGTCATCGCCATTAATCAGGATCCTCTGGGGAAGCAGGGCTACCAGCTTAGACAGGGAGACAATTTTGAG

GTGTGGGAGAGGCCTCTCTCTGGACTTGCCTGGGCTGTGGCTATGATCAACCGGCAGGAAATTGGTGGCCCC

CGCTCCTACACCATTGCTGTTGCCTCCTTGGGCAAGGGCGTGGCCTGTAACCCTGCCTGCTTCATCACCCAGCT

CCTGCCTGTGAAGAGAAAACTGGGATTCTACGAGTGGACCAGCCGGCTGCGGAGCCACATCAATCCCACCGG

CACCGTGCTGCTTCAGCTGGAGAACACCATGCAGATGTCACTGAAAGATCTGCTGTGA [SEQ ID N.2]

or of the following sequence with SEQ ID N.3:

ATGCAGCTCCGCAACCCAGAGCTCCATCTTGGGTGTGCTCTCGCTCTTCGATTCCTTGCACTGGTCAGTTGGG

ATATCCCGGGAGCTAGAGCTTTGGATAACGGTCTCGCACGCACTCCCACAATGGGATGGCTTCACTGGGAGC

GATTTATGTGCAACCTGGACTGCCAGGAAGAGCCGGATAGCTGTATATCTGAGAAGCTTTTTATGGAGATGGC

GGAATTGATGGTCAGTGAAGGCTGGAAAGACGCGGGCTACGAATATCTCTGTATCGACGATTGTTGGATGGC

ACCACAACGCGATAGCGAAGGCAGGCTCCAGGCTGATCCACAGAGGTTTCCCCACGGAATACGACAGCTGG

CTAACTATGTGCACAGCAAGGGCCTCAAACTGGGAATCTACGCTGACGTGGGCAATAAGACGTGCGCCGGTT

TCCCGGGGTCTTTCGGTTACTACGACATTGACGCCCAAACTTTTGCTGACTGGGGTGTGGATCTTCTCAAGTT

TGACGGCTGTTACTGCGACTCCCTCGAAAATTTGGCTGATGGTTACAAGCACATGTCTCTTGCCTTGAATCGCA

CCGGCCGCTCCATCGTGTACTCTTGCGAGTGGCCGTTGTATATGTGGCCCTTTCAAAAACCGAACTACACAGA

AATAAGACAGTATTGCAACCACTGGAGAAACTTCGCTGATATCGACGATAGCTGGAAATCTATTAAATCTATTCT

TGATTGGACGAGTTTTAATCAAGAGCGAATTGTGGACGTTGCGGGGCCGGGAGGGTGGAACGACCCCGATA

TGCTGGTTATCGGAAATTTTGGCCTTTCCTGGAATCAGCAGGTTACCCAGATGGCCCTGTGGGCTATTATGGCC

GCTCCACTCTTCATGAGCAATGATTTGCGCCACATCAGTCCACAAGCGAAGGCTCTCTTGCAGGATAAGGATG

TGATTGCTATCAACCAAGATCCGCTGGGCAAGCAGGGGTATCAGTTGAGACAAGGAGATAACTTCGAAGTTT

GGGAGCGGCCCCTGAGTGGTTTGGCCTGGGCAGTGGCGATGATAAATCGACAAGAAATAGGAGGACCCAG

GAGTTATACTATTGCTGTAGCATCCCTTGGGAAAGGTGTCGCGTGTAACCCCGCTTGTTTTATTACACAACTGCT

GCCTGTTAAGAGAAAACTGGGCTTTTACGAGTGGACCTCTCGGCTCAGATCCCACATCAACCCGACAGGCAC

CGTTCTTCTGCAACTGGAGAATACGATGCAGATGAGCCTCAAGGACTTGTTGTAA [SEQ ID N.3],

or comprising or consisting of a sequence having at least 99.5% of identity with the sequence of SEQ ID N.2 or SEQ ID N.3.

In an embodiment, said nucleic acid comprises or consists of a sequence having 99.5, 99.6, 99.7, 99.8, 99.9 or 100% identity with the sequence of SEQ ID N.2.

In an embodiment, said nucleic acid comprises or consists of a sequence having 99.5, 99.6, 99.7, 99.8, 99.9 or 100% identity with the sequence of SEQ ID N.3.

Said nucleic acid is typically an isolated nucleic acid.

It is also an object of the invention said nucleic acid for the *in vivo* expression of human alpha galactosidase A (GLA), preferably in a cell, more preferably in a human cell. In this embodiment, said nucleic acid preferably comprises or consists of a sequence with SEQ ID N. 1 or 2.

It is also an object of the invention the use of said nucleic acid for the *in vitro* expression of human alpha galactosidase A (GLA), preferably in a cell, more preferably in a human cell. In this embodiment, said nucleic acid preferably comprises or consists of a sequence with SEQ ID N.3.

Said nucleic acid can be used to express the GLA enzyme in vitro or in vivo, in particular by using suitable vectors for cell transfection.

A further object of the invention is a nucleic acid construct comprising:

- a promoter sequence and
- a coding sequence of the alpha-galactosidase A (GLA) gene under control of said promoter, wherein said coding sequence is the nucleic acid above described, preferably comprising or consisting of a sequence with SEQ ID N.1, SEQ ID N.2 or SEQ ID N.3.

Preferably, said promoter sequence is operably linked to the 5'end portion of said coding sequence.

The promoter can be for example a ubiquitous promoter or a tissue-specific promoter.

In an embodiment, said promoter is a promoter specific for liver expression, preferably the human alpha 1-antitrypsin (hAAT) promoter.

In an exemplary embodiment, said promoter is the human alpha 1-antitrypsin (hAAT) promoter and a nucleic acid molecule comprising the hAAT promoter, a transcription start site and a portion of the first exon comprises or consists of a sequence with SEQ ID N. 4.

In a particular embodiment, the promoter is associated to an enhancer sequence, preferably an enhancer derived from apolipoprotein E gene, preferably the human ApoE control region.

Said nucleic acid construct may be an expression cassette comprising the nucleic acid of the invention operably linked to one or more expression control sequences or other sequences improving the expression of a transgene, as known in the art.

Preferably, said nucleic acid construct is inserted in a vector.

It is an object of the invention a vector comprising the nucleic acid or the nucleic acid construct above defined, said vector preferably comprising a nucleic acid comprising or consisting of a sequence with SEQ ID N.1, SEQ ID N.2 or SEQ ID N.3.

Preferably said vector is a viral vector. Examples of viral vectors are adenoviral vectors, adeno-associated viral (AAV) vectors, herpes viral vectors, retroviral vectors, lentiviral vectors, and baculoviral vectors vaccinia viruses, foamy viruses, cytomegaloviruses, Semliki forest virus, poxviruses, RNA virus vector and DNA virus vector. Preferably the viral vector derives from non-pathogenic parvovirus such as adeno-associated virus (AAV), retrovirus such as gammaretrovirus, spumavirus and lentivirus, adenovirus, poxvirus and an herpes virus. More preferably, the viral vector is selected from the group consisting of: adenoviral vectors, lentiviral vectors, retroviral vectors and adeno associated viral vectors (AAV).

In another embodiment, the vector is a non-viral vector such as 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).

In an embodiment, said viral vector further comprises one or more of:

-   a 5' inverted terminal repeat (ITR) sequence of AAV, preferably localized at the 5' end of the promoter;

-   an enhancer element, preferably localized at the 5' end of the promoter;

-   a promoter sequence;

-   a Kozak sequence, preferably localized at the 5' end of the GLA-coding sequence and operably linked to said sequence;

-   a transcription termination sequence preferably localized at the 3'end of the GLA coding sequence;

7

- a 3' inverted terminal repeat (ITR) sequence of AAV, preferably localized at the 3' end of the transcription termination sequence.

Preferably the vector comprises in a 5'-3' direction:

- an AAV 5'-inverted terminal repeat (5'-ITR) sequence;

5 - an enhancer sequence;

- a promoter sequence;

- a Kozak sequence;

- the GLA coding sequence above described under control of said promoter;

- a transcription termination sequence; and

10 - an AAV 3'-inverted terminal repeat (3'-ITR) sequence.

Preferably said enhancer element is an enhancer derived from apolipoprotein E gene.

Preferably said transcription termination sequence is a poly-adenylation signal sequence, preferably the human hemoglobin beta a poly-adenylation signal (HHB polyA).

Preferably the ITRs derive from the same virus serotype or from different virus serotypes, preferably

15 the virus is an AAV, preferably of serotype 2.

Optionally, the vector further comprises an intron, preferably the human hemoglobin beta-derived synthetic intron (HBB2). Said intron is preferably operably linked to the 3' of the promoter sequence.

All such elements are known in the field and their sequences can be obtained by the skilled person according to the general knowledge in the field.

20 In an exemplary embodiment, said enhancer element is an ApoE control region-enhancer and it comprises or consists of a sequence with SEQ ID N. 5.

In an exemplary embodiment, said transcription termination sequence is human hemoglobin beta a poly-adenylation signal (HHB polyA) and it comprises or consists of a sequence with SEQ ID N. 6.

In an exemplary embodiment, said 5'ITR sequence derives from AAV2 and it comprises or consists

25 of a sequence with SEQ ID N. 7.

In an exemplary embodiment, said 3'ITR sequence derives from AAV2 and it comprises or consists of a sequence with SEQ ID N. 8.

In an exemplary embodiment, said intron is the human hemoglobin beta-derived synthetic intron (HBB2) and it comprises or consists of a sequence with SEQ ID N. 9.

Preferably, the vector can be obtained as disclosed in Ronzitti et al. 2016, by substituting the cDNA coding for UGT1A1 with the nucleotide sequence coding for GLA herein described.

5   A further object of the invention is a plasmid comprising:

-    a promoter sequence,

-    a coding sequence of the alpha-galactosidase A (GLA) gene under control of said promoter, wherein said coding sequence is the nucleic acid above described, preferably comprising or consisting of a sequence with SEQ ID N.1, SEQ ID N.2 or SEQ ID N.3.

10   Said plasmid may further comprise one or more of:

-    a 5' inverted terminal repeat (ITR) sequence of AAV, preferably localized at the 5' end of the promoter;

-    an enhancer element, preferably localized at the 5' end of the promoter;

-    a promoter sequence;

15   -   a Kozak sequence, preferably localized at the 5' end of the GLA-coding sequence and operably linked to said sequence;

-    a transcription termination sequence preferably localized at the 3'end of the GLA coding sequence;

-    a 3' inverted terminal repeat (ITR) sequence of AAV, preferably localized at the 3' end of the
20   transcription termination sequence.

Preferably the plasmid comprises in a 5'-3' direction:

- an AAV 5'-inverted terminal repeat (5'-ITR) sequence;

- an enhancer sequence;

- a promoter sequence;

25   - a Kozak sequence;

- the GLA coding sequence above described under control of said promoter;

- a nucleotide sequence of a transcription termination sequence; and

- an AAV 3'-inverted terminal repeat (3'-ITR) sequence.

Optionally, the plasmid further comprises an intron, preferably the human hemoglobin beta-derived synthetic intron (HBB2). Said intron is preferably operably linked to the 3' of the promoter sequence.

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

In a preferred embodiment, the plasmid comprises or consists of a sequence with SEQ ID N.10.

A further object of the invention is the use of said plasmid for the generation of a vector according to the invention. Preferably said vector is an AAV vector or a lentivirus vector. Preferably, it is a vector as described above. Preferably the adeno-associated virus is from the serotype 8.

In an embodiment the nucleic acid or the invention is included in an integrative vector, i.e. a vector for targeted integration of a therapeutic cDNA into the albumin locus. Integrative vectors are known in the field, see for example Barzel et al., 2015; De Caneva et al., 2019; Porro et al., 2017.

In this embodiment, the vector comprises at least:

- coding sequence of the alpha-galactosidase A (GLA) gene, wherein said coding sequence is the nucleic acid above described, preferably comprising or consisting of a sequence with SEQ ID N.1, SEQ ID N.2 or SEQ ID N.3, and
- one or more albumin sequences.

Said albumin sequences can be from albumin genes of any origin, preferably they are from mouse or human albumin gene. Preferably, they are genomic albumin sequences, preferably they comprise coding and/or non-coding sequences of the albumin gene, more preferably they comprise exons 13, 14 and/or 15, and/or introns 12, 13, and/or 14, and/or fragments thereof, of mouse or human albumin gene. In an embodiment, the albumin sequences comprise both coding and non-coding sequences, preferably they comprise a fragment of intron 12, exon 13, intron 13, exon 14, intron 14, and exon 15. Such elements may be comprised in the same albumin sequence or in separate albumin sequences.

Preferably, the vector comprises the GLA coding sequence within a first and a second albumin genomic sequences.

In an embodiment, said vector further comprises one or more of:

- a 5' inverted terminal repeat (ITR) sequence of AAV, preferably localized at the 5' end of the first albumin sequence;

- a ribosomal skipping sequence, preferably localized between the first albumin sequence and the GLA coding sequence, preferably it is in frame with the albumin sequence;

5   - a 3' inverted terminal repeat (ITR) sequence of AAV, preferably localized at the 3' end of the second albumin sequence.

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

Preferably said vector comprises also a protospacer-adjacent motif (PAM) sequence, preferably a

10   PAM8 sequence, preferably located within an albumin sequence.

In an embodiment, said viral vector comprises:

- an AAV 5'-inverted terminal repeat (5'-ITR) sequence;

- a first genomic albumin sequence;

- a ribosomal skipping sequence, preferably P2A,

15   - a GLA coding sequence comprising or consisting of a sequence with SEQ ID N.1, SEQ ID N.2 or SEQ ID N.3;

- a second genomic albumin sequence comprising a PAM sequence;

- an AAV 3'-inverted terminal repeat (3'-ITR) sequence.

Preferably, the GLA coding sequence is deprived of the first codon, i.e. it has the sequence of SEQ ID

20   N.1, 2 or 3 without the initial ATG, in order to reduce the risk of off-target expression.

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

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

All such elements are known in the field and their sequences can be obtained by the skilled person according to the general knowledge in the field.

25   In an exemplary embodiment, said 5'ITR sequence derives from AAV2 and it comprises or consists of a sequence with SEQ ID N. 11.

In an exemplary embodiment, said first genomic albumin sequence derives from mouse albumin and it comprises or consists of a sequence with SEQ ID N. 12.

In an exemplary embodiment, said ribosomal skipping sequence is P2A and it comprises or consists of a sequence with SEQ ID N. 13.

In an exemplary embodiment, said second genomic albumin sequence derives from mouse albumin and it comprises or consists of a sequence with SEQ ID N. 14.

In an exemplary embodiment, said 3'ITR sequence derives from AAV2 and it comprises or consists of a sequence with SEQ ID N. 19.

Preferably, the vector can be obtained as disclosed in De Caneva et al., 2019, by substituting the cDNA coding for UGT1A1 with the nucleotide sequence coding for GLA herein described.

In an embodiment, said vector is administered together with a vector expressing a nuclease. Said vector expressing a nuclease comprises a nucleic acid coding for a nuclease.

Therefore, a pharmaceutical composition comprising a vector according to the invention and a vector expressing a nuclease is also within the scope of the invention.

Preferably, said nucleic acid coding for a nuclease is a DNA construct comprising a nucleic acid coding for Cas9 or spCas9 or SaCas9 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 syntethic polyA (sh polyA). All such elements are well known in the art and may have conventional nucleotide sequences.

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, Cpfl, Casl2b (C2cl), Casl3a (C2c2), Cas3, Csf1, Casl3b (C2c6), and C2c3 or variants thereof such as SaCas9, VQR-Cas9-HF1 or dcas9.

In an embodiment, said vector expressing a nuclease further comprises a guide RNA. A guide RNA is a RNA sequence that hybridizes to a targeting sequence and guides the nuclease to the site of cut. 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. Preferably, the guide RNA is a single guide RNA (sgRNA) that consists of both the crRNA and tracrRNA as a single construct.

An exemplary vector coding for a nuclease and a guide RNA is disclosed in De Caneva et al., 2019.

12

In another embodiment, said vector is administered together with one or more drugs enhancing gene targeting rate.

Therefore, a pharmaceutical composition comprising a vector according to the invention and one or more drugs enhancing gene targeting rate is also within the scope of the invention.

5   A further object of the invention is a plasmid comprising:

-    coding sequence of the alpha-galactosidase A (GLA) gene wherein said coding sequence is the nucleic acid above described, preferably comprising or consisting of a sequence with SEQ ID N.1, SEQ ID N.2 or SEQ ID N.3, and

-    one or more albumin sequences, as above described.

10   In an embodiment, said plasmid comprises:

-    an AAV 5'-inverted terminal repeat (5'-ITR) sequence;

-    a first genomic albumin sequence;

-    a ribosomal skipping sequence, preferably P2A,

-    a GLA coding sequence comprising or consisting of a sequence with SEQ ID N.1, 2 or 3;

15   -   a second genomic albumin sequence comprising a PAM sequence;

-    an AAV 3'-inverted terminal repeat (3'-ITR) sequence.

Preferably, the GLA coding sequence is deprived of the first codon, i.e. it has the sequence of SEQ ID N.1, 2 or 3 without the initial ATG, in order to reduce the risk of off-target expression.

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

20   The plasmid usually further comprises backbone elements which are typically required for the large scale plasmid production in bacteria, such as bacterial origin of replication, bacterial promoter, antibiotic resistance gene.

In a preferred embodiment, the plasmid comprises or consists of a sequence with SEQ ID N.15.

A further object of the invention is the use of said plasmid for the generation of a vector according

25   to the invention. Preferably said vector is an AAV vector or a lentivirus vector. Preferably, it is a vector as described above.

A further object of the invention is a viral particle containing the viral vector as above defined.

It is also an object of the invention a host cell transformed with the vector according to the invention.

In an embodiment, said host cell comprises the vector of the invention as an episomal vector, i.e. the nucleic acid of the invention is not integrated in the genome of the host cell.

In another embodiment, said host cell comprises the nucleic acid of the invention within its genome and expresses said nucleic acid.

5   It is also an object of the invention a pharmaceutical composition comprising the nucleic acid, the nucleic acid construct, the vector or the host cell according to the invention and at least one pharmaceutically acceptable vehicle and/or excipient.

It is an object of the invention the nucleic acid, the nucleic acid construct, the vector or the host cell or the pharmaceutical composition of the invention for use as a medicament.

10 It is an object of the invention the nucleic acid, the nucleic acid construct, the vector or the host cell or the pharmaceutical composition of the invention for use in gene therapy.

Preferably, said nucleic acid, nucleic acid construct, vector or host cell or pharmaceutical composition is for use in the treatment of Fabry disease.

Preferably, the vector is systemically delivered.

15 Preferably, the vector is administered through intravenous injection or retro-orbital injection.

In an embodiment, the vector comprises a promoter specific for liver expression. In this embodiment, the transgene is specifically expressed in the liver thereby treating the disease by liver production of the GLA enzyme. Advantageously, the increase of the GLA enzyme in the liver leads to a lowering of the circulating levels of lyso-Gb3 in plasma and to a complete or almost complete

20 clearance of lyso-Gb3 in the tissues.

It is also an object of the invention a method for treating Fabry disease comprising administering to a subject in need thereof an effective amount of the nucleic acid or of the nucleic acid construct or of the vector or of the host cell or of the pharmaceutical composition according to the invention.

In an embodiment, the integrative vector as above described is administered to a neonatal subject

25 affected by Fabry disease in order to treat the disease at its early stages.

The invention also provides a method for increasing expression of alpha-galactosidase A comprising administering to a subject in need thereof the nucleic acid or nucleic acid construct according to the invention, the vector according to the invention, the host cell according to the invention or the pharmaceutical composition according to the invention.

The invention also provides a method for lowering the circulating level of lyso-Gb3 comprising administering to a subject in need thereof the nucleic acid or nucleic acid construct, the vector, the viral particle, the host cell or the pharmaceutical composition according to the invention.

DETAILED DESCRIPTION OF THE INVENTION

5   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".

10   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

15   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.

As used herein, the terms "nucleic acid" and "polynucleotide sequence" and "construct" refer to a deoxyribonucleotide or ribonucleotide polymer in either single- or double-stranded form, and

20   unless otherwise limited, would encompass known analogs of natural nucleotides that can function in a similar manner as naturally-occurring nucleotides. The polynucleotide sequences include both full-length sequences as well as shorter sequences derived from the full-length sequences. It is understood that a particular polynucleotide sequence includes the degenerate codons of the native sequence or sequences which may be introduced to provide codon preference in a specific host cell.

25   The polynucleotide sequences falling within the scope of the subject invention further include sequences which specifically hybridize with the sequences coding for a peptide of the invention. The polynucleotide includes both the sense and antisense strands as either individual strands or in the duplex.

Included in the present invention are also nucleic acid sequences derived from the nucleotide

30   sequences herein mentioned, e.g. functional fragments, mutants, variants, derivatives, analogues, and sequences having a % of identity of at least 99.5% 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.

The term "inverted terminal repeat" means sequences which are repeated at both ends of a nucleotide sequence in the opposite orientation (reverse complementary).

The term "Fabry disease" (FD) means a rare lysosomal disorder identified in the Online Mendelian Inheritance in Man (OMIM) database with the ID OMIM 301500 and characterized by one or more mutations in the alpha-galactosidase A gene leading to a deficiency of the enzyme alpha galactosidase A. Alpha-galactosidase A deficiency, GLA deficiency, Anderson-Fabry disease, angiokeratoma corporis diffusum, diffuse angiokeratoma, ceramide trihexosidosis, and hereditary dystopic lipidosis are synonyms.

The term "alpha-galactosidase A" or "GLA" herein refers to the human alpha-galactosidase A enzyme, for example encoded by wild type cDNA having the sequence of SEQ ID N.16.

The term "codon optimized" herein means that a codon that expresses a bias for humans is changed to a synonym codon that does not express a bias for humans. The change in codon does not result in a change of the amino acid in the encoded protein.

Figures

Figure 1. Comparative assessment of codon-optimized sequences *in vitro*. A- Experimental design for the assessment of various codon-optimized sequences created. These sequences were cloned into an episomal vector which was transfected in HuH7 cells via lipofectamine 2000. 48 hours post-transfection the cell extract and the supernatant were collected as samples for analysis. B- An enzymatic reaction was performed to evaluate the relative enzymatic activity of the codon-optimized versions and the wild-type hGLA sequence. This data represents the relative hGLA enzyme activity with three individual replicate assays (n=2) where black circles and gray squares indicate cell extract and cell supernatant respectively in each assay.

Figure 2. Comparative assessment of codon-optimized sequences *in vitro*. A-Western blot assay was done for cell extract and supernatant using hGLA specific antibody. B- The bar graph shows a quantitative representation of three technical replicates (n=2). Mean±SEM (*p<0.05). Cell extract and supernatant are indicated by black and gray bars, respectively.

Figure 3. *In vivo* studies in WT C57Bl/6 mice. A- Scheme of the liver-specific episomal AAV vector. The WT and CO02 hGLA cDNA variante are trenascribed under the control of a liver specific

promoter. B- C57BL6/WT mice (P30) were treated with AAV8_hGLA episomal vectors (pSMD2 hGLA_WT, AAV8_pSMD2_CO02 and AAV8_pSMD2_CO03) via retro-orbital injections with a dose of 3E+12vg/kg. The liver was harvested, and blood was collected 3 weeks post-injection (P51). Untreated WT mice were considered as the control. Blood was processed to obtain plasma while protein was extracted from liver tissues. C- An enzyme assay was done to determine GLA activity by initiating a reaction between the enzyme-containing samples; plasma (1:10,000) (red bars) and liver proteins (3ng) (blue bars) and the substrate (4-methylumbelliferyl α-D-galactopyranoside -4MUG) for 1 hour at 37°C. The released fluorescent product; 4-methylumbelliferone (4MU) was measured with excitation at 365nm and emission at 450nm. 4MU concentrations (2pmoles to 100pmoles) were used to obtain a standard curve to determine enzyme activity (EA). One unit of enzyme activity was defined as the amount of enzyme releasing 1 nmol of 4-MU per milligram/milliliter of enzyme and in an hour. Liver and plasma are indicated by black and gray bars, respectively.

Figure 4. *In vivo* studies in WT C57Bl/6 mice. A and B- Western blot analysis was done for all treated and untreated animals using liver proteins (blue bars) and plasma (red bars) as well. HSP70 (for liver) and mIgG (plasma) were used as reference genes. The quantification analysis indicated higher protein expression levels in case pSMD2_hGLA_CO02 and pSMD2_hGLA_CO03 when compared to pSMD2_hGLA_WT in the liver as well as plasma which suggests the better efficiency of these two codon-optimized sequences *in vivo*. Liver and plasma are indicated by black and gray bars, respectively.

Figure 5. Treatment of Fabry mice with liver-specific episomal AAV vectors. A- Mice were treated at P30 with AAV8_pSMD2_WT or AAV8_pSMD2_CO02 at different AAV doses, ranging from 3.0E11 vg/kg to 3.0E13 vg/kg, and sacrificed at P150 (M5). B and C- Plasma GLA activity (B) and lyso-Bb3 determination (C). Untreated WT and untreated MUT Fabry KO mice are used as controls. MUT Fabry mice treated with ERT (Replagal, 1 mg/kg weekly, i.v., for two months) were included as a comparison of efficacy with available treatments. D- Determination of viral genome copies in the liver. E - Western blot of liver extracts and quantification of the blot.

Figure 6. Treatment of Fabry mice with liver-specific episomal AAV vectors. A- Determination of GLA activity in tissues. B- Determination of lyso-Gb3 accumulation in tissues. MUT Fabry mice treated with ERT (Replagal, 1 mg/kg weekly, i.v., for two months) were included as a comparison of efficacy with available treatments.

Figure 7. Treatment of Fabry mice with liver gene targeting approach. A- Mice were treated at P5

with AAV8-Alb-hGLA WT or AAV8-Alb-hGLA CO02, plus the AA8-pX602 vector, expressing the SaCas9 and the sgRNA, at two different AAV doses, and sacrificed at P150 (M5). B and C- Plasma GLA activity (B) and lyso-Bb3 determination (C). Untreated WT and untreated MUT Fabry KO mice are used as controls. MUT Fabry mice treated with ERT (Replagal, 1 mg/kg weekly, i.v., for two months) were included as a comparison of efficacy with available treatments. D- Determination of the rate of targeted integration by ddPCR in the liver. E-Western blot of liver extracts and quantification of the blot.

Figure 8. Treatment of Fabry mice with liver gene targeting approach. A- Determination of GLA activity in tissues. B- Determination of lyso-Gb3 accumulation in tissues.

GENE THERAPY

During the past decade, gene therapy has been applied to the treatment of disease in hundreds of clinical trials. Various tools have been developed to deliver genes into human cells. In the present invention the vectors may be administered to a patient. A skilled worker would be able to determine appropriate dosage range.

Gene therapy may be directed to a specific tissue in order to correct a genetic defect.

In the present invention, a galactosidase A deficiency is corrected by the reduction in lyso-Gb3 levels in plasma and tissue achieved through expression of the GLA enzyme.

REGULATORY ELEMENTS

The subject invention also concerns a vector that can include regulatory elements that are functional in the intended host cell in which the vector 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 vector 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-trascriptional 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 or the vector of the invention can comprise a promoter sequence operably linked to the 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. retinal cell or liver cell).

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), thyroxine binding globulin (TBG) promoter20, OAT chimeric CMV/chicken beta-actin promoter (CBA) and the truncated form of CBA (smCBA) promoter (US8298818 and Light-Driven Cone Arrestin Translocation in Cones of Postnatal Guanylate Cyclase-1 Knockout Mouse Retina Treated with AAVGC1), Rhodopsin promoter (NG_009115, bp 4205-5010), Interphotoreceptor retinoid binding protein promoter (NG_029718.1, bp 4777-5011), vitelliform macular dystrophy 2 promoter (NG_009033.1, bp 4870-5470), PR-specific human G protein-coupled receptor kinase 1 (hGRK1; AY327580.1 bp1793-2087 or bp 1793-1991) (Haire et al. 200622; U.S. Patent No. 8,298,818). However any suitable promoter known in the art may be used.

In a preferred embodiment, the promoter is a promoter specific for liver expression, preferably the human alpha 1-antitrypsin (hAAT) promoter.

Promoters can be incorporated into a construct or a vector using standard techniques known in the art. Multiple copies of promoters or multiple promoters can be used in a construct or a vector 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.

INTRON

19

Introns may be included in a nucleic acid construct or in a vector. In particular, an intron placed between the promoter and the coding sequence increases mRNA stability and protein production, thereby increasing transgene expression.

Any suitable intron may be used, the selection of which may be readily made by the skilled person.

A preferred intron is hemoglobin beta-derived synthetic intron (HBB2).

POLYADENYLATION SEQUENCE

The vector 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 preferred embodiments the polyadenylation sequence is the human hemoglobin beta poly-adenylation signal (HHB polyA) or a fragment thereof that retains the natural function of the polyadenylation sequence.

POST-TRANSCRIPTIONAL REGULATORY ELEMENTS

The vector 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.

KOZAK SEQUENCE

The vector of the present invention may comprise a Kozak sequence. Suitably, the GLA-coding 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, Kozak25).

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.

Any ribosomal skipping sequence may be utilized within the meaning of the present invention. A preferred one is P2A. Ribosomal skipping peptides, for example 2A peptides, are preferably localized between the albumin sequence 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 $(Y)_n$NYAG, 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 $(Y)_n$NCAG, where n is 10-20, or a variant with at least 90% or at least 95% sequence identity.

ALBUMIN SEQUENCES

In some embodiments, the vector of the invention also comprises albumin sequences. Albumin sequences are genomic sequences coding for albumin. Preferably they are from mouse or human genome.

An albumin genomic sequence comprises at least one intron and at least one exon. Preferably it comprises exons 13, 14 and/or 15 and adjacent introns, or fragments thereof.

In an embodiment, the vector of the invention comprises a first albumin sequence which is upstream of the GLA coding sequence and comprises exon 13 and a portion of exon 14 and adjacent introns, or fragments thereof, and a second albumin sequence which is downstream of the GLA coding sequence and comprises a portion of exon 14, exon 15 and adjacent introns. This allows to introduce the GLA coding sequence in the albumin gene of the genome of the transfected cell.

Preferably, the vector comprises the P2A peptide in frame with the first albumin sequence.

GENOME INSERTION SITES

In the embodiment of an integrative vector, the site of the double-strand break (DSB) can be introduced specifically by any suitable technique, for example using a CRISPR/Cas9 system.

In this embodiment, the nuclease is directed to the insertion sites preferably by a guide RNA.

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, tracRNA 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 Csnl and Csxl2), CasIO, , Cpfl, C2c3, C2c2 and C2clCsyl, Csy2, Csy3, Csel, Cse2, Cscl, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmrl, Cmr3, Cmr4, Cmr5, Cmr6, Cpfl, Csbl, Csb2, Csb3, Csxl7, Csxl4, CsxlO, Csxl6, CsaX, Csx3, Csxl, Csxl5, Csfl, 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(l):58-66). Type II Cas nucleases include, but are not limited to, CasI, 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 NBCI Ref. Seq. No. NP 269215, and the amino acid sequence of Streptococcus thermophilus wild-type Cas9 polypeptide is set forth, e.g., in NBCI 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, Nitratifractor , 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 RuvCl 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, Fokl, linked to dCas9. The Fokl-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

5   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

10   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

15   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

20   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

25   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

30   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 Cpfl 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.

VECTORS

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

Such vector may therefore contain any of the elements above described. In particular, it can comprise, besides the nucleic acid coding for GLA, 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.

5   Particularly preferred viruses according to the present invention are 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

10   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

15   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

20   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.

25   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 concatamer 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 AVV 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), aVβ5 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 coinfected 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 AAV8 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 plasmid 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, the nucleic acid with SEQ ID N.1 in the present case, 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. Other methods known in the art for the generation of vectors are equally suitable.

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, DH5α, 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. The cell may be for example a 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

The present invention also provides a pharmaceutical composition for treating an individual by gene therapy, wherein the composition comprises a therapeutically effective amount of the vector of the present invention comprising the therapeutic GLA coding sequence herein described or a viral particle produced by or obtained from the same.

Pharmaceutical compositions within the meaning of the present invention comprise the vector or the host cell of the invention optionally in combination with at least one 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.

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 earned out on cells of such human and non- human species.

KITS

The subject invention also concerns kits comprising the viral vector or the host cells 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.

MEDICAL USES AND METHODS OF TREATMENT

In one aspect the invention provides the nucleic acid, vector, cell, kit or composition of the invention for use as a medicament.

5   In another aspect the invention provides the nucleic acid, vector, cell, kit or composition of the invention for use in treatment of Fabry disease.

Typically, an ordinary skilled clinician will determine the actual dosage which will be most suitable for an individual subject and it will vary with the age, weight and response of the particular individual and administration route. A dose range between 1x10e9 and 1x10e15 genome copies of each vector/kg,

10   preferentially between 1x10e11 and 1x10e13 genome copies of each vector/kg are expected to be effective in humans.

Dosage regimes and effective amounts to be administered can be determined by ordinarily skilled clinicians. Administration may be in the form of a single dose or multiple doses, preferably as a single dose. General methods for performing gene therapy using polynucleotides, expression constructs, and vectors

15   are known in the art (see, for example, Gene Therapy: Principles and Applications, Springer Verlag 199926; and U.S. Patent Nos. 6,461 ,606; 6,204,251 and 6,106,826).

The vector for the use according to the present invention may be used alone or in combination with other treatments or components of the treatment. For example, it can be used together with other treatments which might be helpful for Fabry disease, for example enzyme replacement therapies or oral chaperone

20   therapies.

In an embodiment, the vector for the use of the invention is delivered to the liver, for example by intra-venous injection.

In an embodiment, the vector is an integrative vector as above described and is administered to a neonatal or young subject together with a vector expressing a nuclease, such as Cas9, or together with one or more

25   drugs enhancing gene targeting rate, such as fludarabine. This advantageously allows to treat Fabry disease at its early stages by integrating the functioning GLA coding sequence in the genome of the subject.

In another embodiment, the vector is a non-integrative vector and is administered to a subject of any age.

Any suitable delivery method is contemplated to be used for delivering the compositions of the disclosure.

33

When an integrative vector is used, the individual components of the 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.

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.

The invention will be now described by the following examples.

EXAMPLES

To evaluate the enhanced translatability of the optimized cDNA, *in vitro* and *in vivo* tests have been done, in the following order:

1 - *In silico*: codon optimisation of the human a-galactosidase A (GLA) cDNA

2 - Cloning of the optimised variants into a liver-specific expression vector

3 - *In vitro* test in HUH7 human liver cell line

4 - Production of AAV stocks of the most active constructs

5 - *In vivo* test in WT C57 Bl/6 male juvenile (P30) mice

6 - *In vivo* test in Fabry male juvenile (P30) mice

The *in vivo* tests were done in WT and in Fabry mice (B6;129-Gla^{tm1kul}/J Fabry KO mouse model, Jackson lab stock number 003535). Fabry mice are from The Jackson Laboratory, Maine USA; wild type mice are from ENVIGO, Milan, Italy. Two therapeutic strategies were tested, both based on the delivery of the DNA mediated by adeno-associated virus (AAV): non-integrative and integrative approaches. The main disease markers were the enzymatic activity and the decrease in the accumulation of lyso-Gb3, the toxic metabolite, in plasma and target tissues.

Example 1

Comparative study for codon-optimized hGLA plasmids *in vitro*

Construction of Codon optimized hGLA sequences

To increase the expression levels of the hGLA cDNA, the wild-type sequence was codon-optimized.

Transcript 201 of the wild-type hGLA sequence was taken as the mother sequence and different variants were generated using online codon-optimization tools: the IDT Codon Optimization tool (https://eu.idtdna.com/CodonOpt), JCAT or Java Codon Adaptation Tool (http://www.jcat.de/) and COOL or Codon Optimization On-Line by National University of Singapore.

The online codon-optimized sequences generated were then assessed for parameters like the GC content, CpG islands, cryptic splicing sites, and the presence of alternative OFRs.

These sequences were manually modified to obtain better values for the parameters in consideration and four sequences were finally selected, named hGLA_CO01, hGLA_CO02, hGLA_CO03, hGLA_CO04. To further improve translation efficacy, the region of the start codon was modified following the Kozak sequence consensus. To facilitate cloning into the AAV vector, NheI and SalI cloning sites were introduced. The cDNAs were synthesized by a company. In the first phase, these hGLA sequences were cloned into a non-integrative pSMD2-ApoE-hAAT (ampicillin) liver-specific episomal plasmid for *in vitro* and *in vivo* assessment having the following structure:

5' SalI--Kozak seq—ATG—hGLA--STOP--NheI—EcoR1 3'

GLA enzyme assay

The enzyme assay was optimized exploiting the substrate 4-methylumbelliferyl α-D Galactopyranoside (4-MUG) which when reacts with the enzyme Alpha-galactosidase A (GLA) results in a fluorescent product; 4-Methylumbelliferone (4MU) which can be analyzed at a wavelength of 365 nm (excitation) and 450 nm (emission). Therefore, 2.46 mM substrate in a citrate-phosphate buffer (pH 4.5) was incubated with 25 µl samples containing the GLA enzyme for 1h at 37°C on a shaker, and later the reaction was stopped using 200 mM NaOH-Glycine (pH 10.4). 4MU standards were made ranging from 2 pmoles to 100 pmoles to obtain a standard curve. The samples were measured at 365-450 nm and analyzed for enzyme activity (nmoles/mg/h)

To determine whether the codon optimized hGLA cDNA sequences were functional and more efficient than the wild-type hGLA counterpart, pSMD2_hGLA plasmids, both codon-optimized and wild-type versions, were transfected into HuH7 cells using Lipofectamine (to normalize for transfection efficiency, a plasmid expressing EGFP was co-transfected). Untreated cells were considered Mock control for the assay. Proceeding an incubation period of 48h, cell supernatant was collected, and cell extract was harvested in PBS. Proteins were extracted from the cells and supernatant. 5 µg of cell's proteins and 5 µl of cell supernatant were used to react with 4MUG for 1h and later the reaction was stopped, and fluorescence was measured (Figure 1A).

It was evident with three technical replicates of the assay and two biological replicates each that the codon-optimized hGLA sequences were more efficient than the wild-type sequence, showing higher levels of GLA enzyme activity (excluding CO04; Figure 1B). Moreover, it was interesting to observe that the GLA present in the supernatant (4-fold higher) was superior in terms of activity than the cell extract (2-fold higher), giving evidence that a larger proportion of the GLA is in the secreted form.

To confirm the presence of the hGLA protein in the samples, expression analysis was required. Therefore, SDS-PAGE gels were run for cell extract and supernatant separately. Western blot analysis was done to compare protein levels of codon-optimized hGLA and wildtype sequences in both cases, using eGFP as control for transfection efficiency. (Figure 2A) Anti-GLA antibody (1:3000; 49 KDa; Sino Biological Cat# 12078-R001) was used to obtain hGLA specific bands.

Figure 2A represents the western blots analysis where it is evident that codon-optimized sequences express 5-6 fold higher protein that wild-type in the supernatant where ~2-fold higher expression in the cell extract (n=2). These data coincide with the information obtained from the enzyme assay and therefore it can be claimed with confidence that codon-optimized hGLA sequences were more efficient than the wild-type hGLA sequence (excluding CO04) (Figure 2).

Example 2

Comparative study for codon-optimized hGLA vectors *in vivo*

With the confidence in the results of the *in vitro* assays where the codon-optimized sequences demonstrated better efficiency than wild-type both in terms of protein expression and enzyme activity, it was necessary to validate the same *in vivo.*

For this, AAV8-pSMD2-hGLA virus stocks were prepared for codon-optimized hGLA CO02, and CO03 as well as for wild-type hGLA (Figure 3A). The vector contains the alpha1-AAT liver-specific gene promoter.

The wild type human GLA cDNA has sequence with SEQ ID N. 16. Codon optimized CO2 sequence has SEQ ID N.1; codon optimized CO3 sequence has SEQ ID N.2.

One-month-old (post-natal day 30, P30) C57BL6/WT mice (n=5) were treated with a high dose of the episomal vectors (3.0E+12 vg/kg) via retro-orbital injections, to expect much higher hGLA expression than the endogenous mGLA. Untreated animals were considered as control. These mice were sacrificed three weeks post-injection (P51). The liver was harvested and processed for protein extraction and quantified by Bradford's assay and blood was collected for plasma analysis (Figure

3B).

To determine GLA activity, the fluorescent enzyme assay was performed with 3 ng liver proteins and plasma diluted 1:10000 in PBS.

An impressive increase in the activity levels (nmole/mg/hr for liver and nmoles/ml/hr for plasma) of hGLA enzyme was observed in the case for CO02 and CO03 in the liver as well as in plasma when compared to the wild type hGLA treated samples (Figure 3C). CO02 and CO03 levels reach up to 6000-8000 nmoles/ml/hr in plasma while the wild-type values reside to ~2000 nmoles/ml/hr. In case of liver tissue, the codon-optimized sequences (CO02 and CO03) activity rose up to 2000-3000 nmoles/mg/hr from that of 1000 nmoles/mg/hr in the wild-type sequence (Figure 3C).

For a more confident result, hGLA protein expression was analyzed by western blot (Figure 4). Plasma and liver proteins from treated animals as well as untreated control groups were run on an SDS-PAGE gel and blotted on a nitrocellulose membrane (Figure 4A). GLA specific antibody was used along with HSP70 (liver) and mIgG (plasma) for normalization. According to the quantification done from the bands obtained, hGLA CO02 and hGLA CO03 indicate an increase in the relative expression by ~1-fold as compared to wild-type sequence in both liver tissue and plasma which followed the same trend as their enzyme activity (Figure 4B).

Considering the data obtained from the enzyme assay and the western blot it was concluded that CO02 and CO03 had better potentials than wildtype in C57BL/6 mice. CO02 was given preference for subsequent experiments based on its notable increase in the enzyme activity levels.

Example 3

AAV-mediated liver-specific episomal gene therapy in Fabry KO mice

Fabry mouse model [B6;129-Gla^{tm1kul}/J]

The B6;129-Gla^{tm1kul}/J Fabry KO mouse model was bought from The Jackson laboratory, Maine, USA, and established and housed according to international guidelines approved by the animal facility in ICGEB, Trieste receiving standard chow and water ad libitum. The colony was propagated to obtain hemizygous males for experimentation and wild-type males as controls.

Results

In this study the WT and CO2 hGLA cDNAs were delivered to one-month old (P30) Fabry mice by retro-orbital injection. The AAV vector is the one described in Figure 3A, in which the hGLA cDNA is

transcribed under the control of a liver-specific promoter (Ronzitti et al., 2016). The wild type pSMD2 vector comprising the hGLA wild type cDNA has sequence with SEQ ID N.18 and the pSMD2 CO2 hGLA vector comprising the hGLA CO02 cDNA has sequence with SEQ ID N.10.

The two Gene therapy non-integrative vectors AAV8 pSMD2 hGLA WT or AAV8 pSMD2 hGLA CO02 were injected at doses ranging from 3.0E11 to 3.0E13vg/kg. Untreated mutants and untreated wild-type animals were used as control groups. All animals were weighed and bled by sub-mandibular vein puncture every 30 days to extract plasma. At 5 months of age (P150) all the animals were sacrificed, and liver, kidney, and heart were harvested, blood was also collected for plasma (Figure 5B). MUT Fabry mice treated with ERT (Replagal, 1 mg/kg weekly, i.v., for two months) were included as a comparison of efficacy with available treatments. Enzyme activity levels in mice treated with ERT were in the range of those of WT mice, but significantly lower than in mice treated with the CO2 hGLA cDNA.

Using the plasma collected at different time points of the treatment, an enzyme assay was done to determine the efficacy of the vectors. 1:10,000 dilution of plasma in PBS was used to set up the assay for 1 hour (nmoles/ml/hr). Animals treated with episomal vectors AAV8 pSMD2 hGLA WT and AAV8 pSMD2 hGLA CO02 displayed supraphysiological levels of enzyme activity reaching ~10000x higher values than the wild-type animals in the groups treated with the higher AAV doses (Figure 5A).

GLA activity persisted for 4 months post-treatment with a decrease in the first month in the case of pSMD2 hGLA CO02 due to excessive production of GLA and ER stress combined with an elevation in the inflammation markers.

We observed the complete clearance of lyso-Gb3 in plasma in case of animals treated with both vectors at the highest AAV dose (Figure 5C). We observed a clear increase in the GLA activity in the groups treated with the AAV8 pSMD2 hGLA CO02 vector (66x, 74x, and 93x, in the 1.0E13 vg/kg, 3.0E12 vg/kg, and 3.0E11 vg/kg, respectively; Figure 5A). Mass spec analysis of plasma lyso-Gb3accumulation also showed an impressive difference between the WT and CO2-treated groups. While the treatment with the 3.0E11 vg/kg dose resulted in a ~50% decrease in the lyso-Gb3 levels in the group treated with the WT hGLA cDNA, the reduction was about 90% in the hGLA CO2-treated animals (Figure 5C). Importantly, vector copy number in liver were similar in WT and CO2-treated groups (Figure 5D). Western blot of liver extracts and quantification of the blot were also carried out (Figure 5E).

Distribution and uptake of GLA by tissues was also considered by measuring GLA activity and Lyso-Gb3 analysis in Liver, kidney, and heart. Like plasma, at all doses tested and with both WT and CO2-vectors, all tissues showed supraphysiological GLA activity when compared to untreated WT animals (Figure 6A). The activity in the tissues of the animals treated with the CO2 hGLA cDNA presented higher enzymatic activity, more evident in the groups treated with the lower AAV doses, due probably to the oversaturation of the hGLA enzyme-capture mechanism present in the tissues, in the groups of animals treated with the higher AAV doses.

Lyso-Gb3 was completely cleared in the groups treated with the highest dose, both for WT and CO2-treated animals (Figure 6B). In the groups treated with the lowest AAV dose (3.0E11 vg/kg) we observed a complete or almost complete clearance of lyso-Gb3 in the tissues of mice treated with the AAV8 pSMD2 hGLA CO02 vector (98%, 90%, and 85% in liver, kidney and heart, respectively, Figure 6B), while in the mice treated with the AAV8 pSMD2 hGLA WT vector we observed a decrease in lyso-Gb3 accumulation of 65%, 40%, and 40% in liver, kidney, and heart, respectively (Figure 6B). The clearance of lysoGb3 in the animals treated with the hGLA CO02 cDNA was significantly more efficient than in mice treated with the standard ERT doses, for all AAV doses tested.

Example 4

Cloning hGLA into an integrative vector and treating early-onset Fabry disease by targeting neonatal Fabry KO animals

Considering the results obtained from the *in vitro* and *in vivo* data where we compared the codon-optimized constructs, it was evident that hGLA_CO02 has the potential to work better than the wild-type construct.

Therefore, we tested another therapeutic approach based on the targeted integration of the therapeutic cDNA into the albumin locus (Barzel et al., 2015; De Caneva et al., 2019; Porro et al., 2017). Thus, the wild-type vector and the CO02 version of hGLA cDNA were cloned into an integrative vector; pAB288 backbone flanked by albumin homology arms on both ends, a P2A peptide sequence, and a Modified PAM8 sequence. The vector has ITR regions to be packaged into AAV vectors. Plasmid DNA was prepared of the cloned vectors and AAV8 viral vectors were prepared by the AAV facility in ICGEB, Trieste. The wild type pAB288 vector comprising the hGLA wild type cDNA has sequence with SEQ ID N.17 and the pAB288 CO2 hGLA vector comprising the hGLA CO02 cDNA has sequence with SEQ ID N.15.

In order to treat a disease in its early stages or neonatal stage, the integrative approach is crucial

and can proof effective. Therefore, an experiment was designed where neonate Fabry KO male mice at P5 were treated with donor vector WT as well as CO02 coupled with Cas9 (5:1) at two different doses i.e., 3E+13vg/kg and 1E+14vg/kg. Animals were sacrificed at 5-months of age. Liver, kidney, and heart were harvested at sacrifice, along with blood for plasma collected at intermediate time points for GLA enzyme activity and lyso-Gb3 MS analysis (Figure 7A). MUT Fabry mice treated with ERT (Replagal, 1 mg/kg weekly, i.v., for two months) were included as a comparison of efficacy with available treatments.

Plasma GLA activity level was higher than in WT mice in all treated mice. Importantly, mice treated with the CO02 donor vector presented an increase in GLA activity, more evident in the groups treated with the lower dose (about 10x increase; Figure 7B). In both treated groups (low and high dose), GLA activity levels were significantly higher than in the mice treated with the standard ERT treatment. Mass spec determination of lyso-Gb3 accumulation on plasma showed the complete correction at the higher AAV dose for both cDNAs. In the animals treated with the lower dose, reduction was complete for the CO02 vector, while in the group treated with the WT cDNA lyso-Gb3 levels presented a 75% reduction, compared to untreated mutant mice (Figure 7C). Determination of the integration rate by ddPCR indicated similar values for both donor constructs (Figure 7D). Quantification of the GLA protein by Western blot analysis showed 3-4 times higher GLA levels in the liver of animals treated with the CO02 vector, confirming the increased translatability of this codon optimised version, compared to the WT sequence (Figure 7E).

GLA activity and lys-Gb3 accumulation were determined in tissues of treated mice (Figure 8). Activity in tissues was in all cases higher in the mice treated with the CO02 hGLA cDNA (Figure 8A). In line with this results, we observed complete clearance of lyso-Gb3 accumulation in mice treated with both doses of the CO02 hGLA cDNA donor vector, while clearance was not complete in the animals treated with the WT hGLA donor vector (Figure 8B). The clearance of lysoGb3 was more efficient in the mice treated with the CO02 hGLA cDNA than in those treated with ERT, confirming the increased therapeutic efficacy of the CO02 hGLA cDNA.

SEQUENCES

Human GLA cDNA wild type

ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTGCGCTTCGCTTCCTGGCCCTCGTTTCCTGGG
ACATCCCTGGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCCTACCATGGGCTGGCTGCACTGGGAG
CGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCAGATTCCTGCATCAGTGAGAAGCTCTTCATGGAGATG

GCAGAGCTCATGGTCTCAGAAGGCTGGAAGGATGCAGGTTATGAGTACCTCTGCATTGATGACTGTTGGATG
GCTCCCCAAAGAGATTCAGAAGGCAGACTTCAGGCAGACCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTG
GCTAATTATGTTCACAGCAAAGGACTGAAGCTAGGGATTTATGCAGATGTTGGAAATAAAACCTGCGCAGGCT
TCCCTGGGAGTTTTGGATACTACGACATTGATGCCCAGACCTTTGCTGACTGGGGAGTAGATCTGCTAAAATTT
GATGGTTGTTACTGTGACAGTTTGGAAAATTTGGCAGATGGTTATAAGCACATGTCCTTGGCCCTGAATAGGA
CTGGCAGAAGCATTGTGTACTCCTGTGAGTGGCCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAA
ATCCGACAGTACTGCAATCACTGGCGAAATTTTGCTGACATTGATGATTCCTGGAAAAGTATAAAGAGTATCTT
GGACTGGACATCTTTTAACCAGGAGAGAATTGTTGATGTTGCTGGACCAGGGGGTTGGAATGACCCAGATAT
GTTAGTGATTGGCAACTTTGGCCTCAGCTGGAATCAGCAAGTAACTCAGATGGCCCTCTGGGCTATCATGGCT
GCTCCTTTATTCATGTCTAATGACCTCCGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGT
AATTGCCATCAATCAGGACCCCTTGGGCAAGCAAGGGTACCAGCTTAGACAGGGAGACAACTTTGAAGTGTG
GGAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCGGCAGGAGATTGGTGGACCTCGCTCT
TATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAATCCTGCCTGCTTCATCACACAGCTCCTCCC
TGTGAAAAGGAAGCTAGGGTTCTATGAATGGACTTCAAGGTTAAGAAGTCACATAAATCCCACAGGCACTGTT
TTGCTTCAGCTAGAAAATACAATGCAGATGTCATTAAAAGACTTACTT**TAA** [SEQ ID N.16]

Human GLA cDNA Codon optimized 02 (CO02)

ATGCAGCTGCGCAACCCCGAGCTGCACCTGGGCTGCGCCCTGGCCCTGCGCTTCCTGGCCCTGGTCAGCTGG
GACATCCCCGGCGCCCGCGCCCTGGACAACGGCCTGGCCCGCACCCCCACCATGGGCTGGCTGCACTGGGA
GCGCTTCATGTGCAACCTGGACTGCCAGGAGGAGCCCGACAGCTGCATCAGCGAGAAGCTGTTTATGGAGAT
GGCCGAGCTGATGGTCAGCGAGGGCTGGAAGGACGCCGGCTACGAGTACCTGTGCATCGACGACTGCTGGA
TGGCCCCCCAGCGCGACAGCGAGGGCCGCCTGCAGGCCGACCCCAGCGCTTCCCCCACGGAATCCGCCAG
CTGGCCAACTACGTGCACAGCAAGGGCCTGAAGCTGGGCATCTACGCCGACGTGGGCAACAAGACCTGCGC
CGGCTTCCCCGGCAGCTTCGGCTACTACGACATCGACGCCCAGACCTTCGCCGACTGGGGCGTGGACCTGCT
GAAGTTCGACGGCTGCTACTGCGACAGCCTGGAGAACCTGGCCGACGGCTACAAGCACATGAGCCTGGCCC
TGAACCGCACCGGCCGCAGCATCGTGTACAGCTGCGAGTGGCCCCTGTATATGTGGCCCTTCCAGAAGCCCAA
CTACACCGAGATCCGCCAGTACTGCAACCACTGGCGCAACTTCGCCGACATCGACGACAGCTGGAAGAGCAT
CAAGAGCATCCTGGACTGGACCAGCTTCAACCAGGAGCGCATCGTGGACGTGGCCGGCCCCGGCGGCTGGA
ACGACCCCGACATGCTGGTGATCGGCAACTTCGGCCTGAGCTGGAACCAGCAGGTGACCCAGATGGCCCTGT
GGGCCATTATGGCCGCCCCCCTGTTTATGAGCAACGACCTGCGCCACATCAGCCCCCAGGCCAAGGCCCTGCT
GCAGGACAAGGACGTGATCGCTATCAACCAGGACCCCCTGGGCAAGCAGGGCTACCAGCTGCGCCAGGGCG
ACAACTTCGAGGTCTGGGAGCGCCCCCTGAGCGGCCTGGCCTGGGCCGTGGCTATGATCAACCGCCAGGAG
ATCGGCGGCCCCGCAGCTACACCATCGCCGTGGCCAGCCTGGGCAAGGGCGTGGCCTGCAACCCCGCCTG
CTTCATCACCCAGCTGCTGCCCGTGAAGCGCAAGCTGGGCTTCTACGAGTGGACCAGCCGCCTGCGCAGCCA
CATCAACCCCACCGGCACCGTGCTGCTGCAGCTGGAGAACACAATGCAGATGAGCCTGAAGGACCTGCT**GTA
A** [SEQ ID N. 1]

Human GLA cDNA codon optimized 03 (CO03)

ATGCAGTTGAGAAACCCAGAGCTCCACCTGGGCTGTGCCCTGGCACTGAGGTTCCTGGCCCTTGTGAGCTGG

GATATCCCTGGGGCCAGGGCCTTGGACAACGGCTTGGCCCGCACCCCCACAATGGGCTGGCTGCACTGGGA

ACGCTTTATGTGCAATCTGGACTGCCAGGAGGAGCCTGACAGCTGTATCAGCGAGAAGCTCTTTATGGAGATG

GCAGAGCTGATGGTGTCTGAGGGATGGAAGGACGCCGGCTACGAATACCTGTGCATTGACGATTGCTGGATG

GCTCCACAGAGGGACTCAGAAGGACGCCTGCAGGCTGATCCCCAGAGATTCCCCCATGGAATCCGCCAGCTG

GCCAACTATGTGCACAGCAAAGGCCTGAAGCTGGGCATCTACGCCGACGTGGGCAACAAGACCTGTGCTGG

CTTCCCTGGCTCCTTTGGATATTACGATATCGACGCTCAGACCTTTGCTGACTGGGGAGTGGATCTCCTCAAGT

TTGACGGCTGCTACTGTGACTCTCTGGAAAACCTGGCAGATGGCTACAAGCACATGTCCCTGGCTCTGAACAG

AACAGGCCGCAGCATTGTGTACAGCTGCGAGTGGCCCCTGTATATGTGGCCCTTCCAGAAGCCCAACTACACA

GAGATCAGGCAGTACTGCAACCACTGGAGGAACTTTGCCGACATTGACGACTCCTGGAAATCTATCAAGTCTA

TCCTGGATTGGACATCCTTCAACCAAGAGCGGATCGTGGACGTGGCTGGACCTGGAGGCTGGAATGATCCAG

ATATGCTGGTGATTGGAAACTTCGGGCTGTCTTGGAACCAGCAGGTCACTCAGATGGCGCTGTGGGCCATCAT

GGCCGCCCCCCTCTTTATGAGCAACGACCTGCGCCACATTTCTCCTCAAGCCAAGGCCCTGCTCCAGGACAAG

GACGTCATCGCCATTAATCAGGATCCTCTGGGGAAGCAGGGCTACCAGCTTAGACAGGGAGACAATTTTGAG

GTGTGGGAGAGGCCTCTCTCTGGACTTGCCTGGGCTGTGGCTATGATCAACCGGCAGGAAATTGGTGGCCCC

CGCTCCTACACCATTGCTGTTGCCTCCTTGGGCAAGGGCGTGGCCTGTAACCCTGCCTGCTTCATCACCCAGCT

CCTGCCTGTGAAGAGAAACTGGGATTCTACGAGTGGACCAGCCGGCTGCGGAGCCACATCAATCCCACCGG

CACCGTGCTGCTTCAGCTGGAGAACACCATGCAGATGTCACTGAAAGATCTGCTGTGA [SEQ ID N.2]

Human GLA cDNA codon optimized 01 (CO01)

ATGCAGCTCCGCAACCCAGAGCTCCATCTTGGGTGTGCTCTCGCTCTTCGATTCCTTGCACTGGTCAGTTGGG

ATATCCCGGGAGCTAGAGCTTTGGATAACGGTCTCGCACGCACTCCCACAATGGGATGGCTTCACTGGGAGC

GATTTATGTGCAACCTGGACTGCCAGGAAGAGCCGGATAGCTGTATATCTGAGAAGCTTTTTATGGAGATGGC

GGAATTGATGGTCAGTGAAGGCTGGAAAGACGCGGGCTACGAATATCTCTGTATCGACGATTGTTGGATGGC

ACCACAACGCGATAGCGAAGGCAGGCTCCAGGCTGATCCACAGAGGTTTCCCCACGGAATACGACAGCTGG

CTAACTATGTGCACAGCAAGGGCCTCAAACTGGGAATCTACGCTGACGTGGGCAATAAGACGTGCGCCGGTT

TCCCGGGGGTCTTTCGGTTACTACGACATTGACGCCCAAACTTTTGCTGACTGGGGTGTGGATCTTCTCAAGTT

TGACGGCTGTTACTGCGACTCCCTCGAAAATTTGGCTGATGGTTACAAGCACATGTCTCTTGCCTTGAATCGCA

CCGGCCGCTCCATCGTGTACTCTTGCGAGTGGCCGTTGTATATGTGGCCCTTTCAAAAACCGAACTACACAGA

AATAAGACAGTATTGCAACCACTGGAGAAACTTCGCTGATATCGACGATAGCTGGAAATCTATTAAATCTATTCT

TGATTGGACGAGTTTTAATCAAGAGCGAATTGTGGACGTTGCGGGGCCGGGAGGGTGGAACGACCCCGATA

TGCTGGTTATCGGAAATTTTGGCCTTTCCTGGAATCAGCAGGTTACCCAGATGGCCCTGTGGGCTATTATGGCC

GCTCCACTCTTCATGAGCAATGATTTGCGCCACATCAGTCCACAAGCGAAGGCTCTCTTGCAGGATAAGGATG

TGATTGCTATCAACCAAGATCCGCTGGGCAAGCAGGGGTATCAGTTGAGACAAGGAGATAACTTCGAAGTTT

GGGAGCGGCCCCTGAGTGGTTTGGCCTGGGCAGTGGCGATGATAAATCGACAAGAAATAGGAGGACCCAG

GAGTTATACTATTGCTGTAGCATCCCTTGGGAAAGGTGTCGCGTGTAACCCCGCTTGTTTTATTACACAACTGCT

GCCTGTTAAGAGAAAACTGGGCTTTTACGAGTGGACCTCTCGGCTCAGATCCCACATCAACCCGACAGGCAC

CGTTCTTCTGCAACTGGAGAATACGATGCAGATGAGCCTCAAGGACTTGTTGTAA [SEQ ID N.3]

pSMD2 vector elements:

5' AAV2 ITR
Ctgcgcgctcgctcgctcactgaggccgcccgggcaaagcccgggcgtcgggcgacctttggtcgcccggcctcagtgagcgagcgagcgcg
cagagagggagtggccaactccatcactaggggttcct [SEQ ID N.7]

ApoE control region-enhancer
Aaggctcagaggcacacaggagtttctgggctcaccctgcccccttccaacccctcagttcccatcctccagcagctgtttgtgtgctgcctctga
agtccacactgaacaaacttcagcctactcatgtccctaaaatgggcaaacattgcaagcagcaaacagcaaacacacagccctccctgcctg
ctgaccttggagctggggcagaggtcagagacctctctgggcccatgccacctccaacatccactcgacccccttggaatttcggtggagagga
gcagaggttgtcctggcgtggtttaggtagtgtgagaggg [SEQ ID N.5]

hAAT promoter and first exon
GATCTTGCTACCAGTGGAACAGCCACTAAGGATTCTGCAGTGAGAGCAGAGGGCCAGCTAAGTGGTACTCTC
CCAGAGACTGTCTGACTCACGCCACCCCCTCCACCTTGGACACAGGACGCTGTGGTTTCTGAGCCAGGTACA
ATGACTCCTTTCGGTAAGTGCAGTGGAAGCTGTACACTGCCCAGGCAAAGCGTCCGGGCAGCGTAGGCGGG
CGACTCAGATCCCAGCCAGTGGACTTAGCCCCTGTTTGCTCCTCCGATAACTGGGGTGACCTTGGTTAATATTC
ACCAGCAGCCTCCCCCGTTGCCCCTCTGGATCCACTGCTTAAATACGGACGAGGACAGGGccctgtctcctcagctt
caggcaccaccactgacctgggacagtgaat [SEQ ID N.4]

Modified human hemoglobin beta intron (HBB2)
Gtacacatattgaccaaatcagggtaattttgcatttgtaattttaaaaaatgctttcttcttttaatatactttttttgtttatcttatttctaatactttc
cctaatctctttctttcagggcaatattgatacaatgtatcttgcctctttgcaccattctaaagaataacagtgataatttctgggttaaggcaata
gcaatatttctgcatataaatatttctgcatataaattgtaactgatgtaagaggtttcatattgctaatagcagctacaatccagctaccattctg
cttttatttttctggttgggataaggctggattattctgagtccaagctaggcccttttgctaatcttgttcatacctcttatcttcctcccacagctcct
gggcaacctgctggtctctctgctggcccatcactttggcaaag [SEQ ID N.9]

HBB2 poly A
Gaattcacccaccagtgcaggctgcctatcagaaagtggtggctggtgtggctaatgccctggcccacaagtatcactaagctcgctttcttg
ctgtccaatttctattaaaggttcctttgttccctaagtccaactactaaactggggggatattatgaagggccttgagcatctggattctgcctaat
aaaaaacatttattttcattgcaatgatgtatttaaattatttctgaatatttttactaaaaagggaatgtgggaggtcagtgcatttaaaacataa
agaaatgaagagctagttcaaaccttgggaaaatacactatatcttaaactccatgaaagaaggtgaggctgcaaacagctaatgcacattg
gcaacagcccctgatgcctatgccttattcatccctcagaaaaggattcaagtagaggcttgatttggaggttaaagtttggctatgctgtatttt
acattacttattgttttagctgtcctcatgaatgtctttcactacccatttgcttatcctgcatctctcagccttgactccactcagttctcttgcttag
agataccacctttcccctgaagtgttccttccatgttttacggcgagatggtttctcctcgcctggccactcagccttagttgtctctgttgtcttata
gaggtctacttgaagaaggaaaaacaggggggcatggtttgactgtcctgtgagcccttcttccctgcctcccccactcacagtgacccggaatc

[SEQ ID N.6]

3' AAV2 ITR
Aggaacccctagtgatggagttggccactccctctctgcgcgctcgctcgctcactgaggccgggcgaccaaaggtcgcccgacgcccgggct

5 ttgcccgggcggcctcagtgagcgagcgagcgcgc [SEQ ID N.8]

Modified pAB288 vector elements:

5' AAV2 ITR

10 Ttggccactccctctctgcgcgctcgctcgctcactgaggccgggcgaccaaaggtcgcccgacgcccgggctttgcccgggcggcctcagtg
agcgagcgagcgcgcagagagggagtggccaactccatcactaggggttcct [SEQ ID N.11]

mouse albumin-left homology arm
tgtacataggaggttcgaaccctgctgaagggagaggttccaatactacaaaatgtagcgggatattgtcatcacctttgggggacatgtcatca

15 tggtccccagacagagttacaaaactcatcccctacacagcactatgtctctggtactgtttgttctacagatgtcaacaacagaggcccagcca
tctcctattgcttggcttgtcagtctttctagcctccccattattaatttcaaatggggcaggtgttaggagggcaaaaatccacatattaagtgca
aagcctttcaggagatttcctgaaactagacaaaacccgtgtgactggcatcgattattctatttgatctagctagtcctagcaaagtgacaact
gctactcccctcctacacagccaagattcctaagttggcagtggcatgcttaatcctcaaagccaaagtacttggctccaagatttatagcctta
aactgtggcctcacattccttcctatcttactttcctgcactggggtaaatgtctccttgctcttcttgctttctgtcctactgcagGGCTCTTGCT

20 GAGCTGGTGAAGCACAAGCCCAAGGCTACAGCGGAGCAACTGAAGACTGTCATGGATGACTTTGCACAGTT
CCTGGATACATGTTGCAAGGCTGCTGACAAGGACACCTGCTTCTCGACTGAGgtcagaaacgtttttgcattttgacgat
gttcagtttccattttctgtgcacgtggtcaggtgtagctctctggaactcacacactgaataactccaccaatctagatgttgttctctacgtaact
gtaatagaaactgacttacgtagctttta attttta ttttctgccacactgctgcctattaaatacctattatcactatttggtttcaaatttgtgacac
agaagagcatagttagaaatacttgcaaagcctagaatcatgaactcatttaaaccttgccctgaaatgtttcttttt gaattgagttattttacac

25 atgaatggacagttaccattatatatctgaatcatttcacattccctcccatggcctaacaacagtttatcttcttattttgggcacaacagatgtca
gagagcctgctttaggaattctaagtagaactgtaattaagcaatgcaaggcacgtacgtttactatgtcattgcctatggctatgaagtgcaaa
tcctaACAGTCCTGCTAATACTTTTCtaacatccatcatttctttgttttcagGGTCCAAACCTTGTCACTAGATGCAAAGAC
GCCTTAGCC [SEQ ID N.12]

30 P2A
Ggaagcggcgccaccaatttcagcctgctgaaacaggccggcgacgtggaagagaaccctggccct [SEQ ID N.13]

mouse albumin-modified right homology arm
ttagccTAAacacatcacaaccacaaccttctcaggtaactatacttgggacttaaaaaacataatcataatcattttttcctaaaacgatcaag

35 actgataaccatttgacaagagccatacagacaagcaccagctggcTCTCGAGCGTCTTCACGTATGGTCATCagtttgggttccat
ttgtagataagaaactgaacatataaaggtctaggttaatgcaatttacacaaaaggagaccaaaccagggagagaaggaaccaaaattaa
aaattcaaaccagagcaaaggagttagccctggttttgctctgacttacatgaaccactatgtggagtcctccatgttagcctagtcaagcttatc
ctctggatgaagttgaaaccatatgaaggaatatttgggggggtgggtcaaaacagttgtgtatcaatgattccatgtggtttgacccaatcattct
gtgaatccatttcaacagaagatacaacgggttctgtttcataataagtgatccacttccaaatttctgatgtgccccatgctaagctttaacaga

40 atttatcttcttatgacaaagcagcctcctttgaaaatatagccaactgcacacagctatgttgatcaattttgtttataatcttgcagaagagaat
ttttttaaaatagggcaataatggaaggctttggcaaaaaaattgtttctccatgaaaacaaaaaacttattttttttattcaagcaaagaacct
atagacataaggctatttcaaaattatttcagtttttagaaagaattgaaagtttttgtagcattctgagaagacagctttcatttgtaatcataggta

atatgtaggtcctcagaaatggtgagacccctgactttgacacttggggactctgagggaccagtgatgaagagggcacaacttatatcacac
atgcacgagttggggtgagagggtgtcacaacatctatcagtgtgtcatctgcccaccaagtaacagatgtcagctaagactaggtcatgtgta
ggctgtctacaccagtgaaaatcgcaaaaagaatctaagaaattccacatttctagaaaataggtttggaaaccgtattccattttacaaagga
cacttacatttctctttttgttttccagGCTACCCTGAGAAAAAAGACATGAAGACTCAGGACTCATCTTTTCTGTTGGT
GTAAAATCAACACCCTAAGGAACACAAATTTCTTTAAACATTTGACTTCTTGTCTCTGTGCTGCAATTAATAAAA
AATGGAAAGAATCTACtctgtggttcagaactctatcttccaaaggcgcgcttcaccctagcagcctctttggctcagaggaatccctgcc
tttcctcccttcatctcagcagagaatgtagttccacatggg [SEQ ID N.14]

3' AAV2 ITR

Aggaaccccctagtgatggagttggccactccctctctgcgcgctcgctcgctcactgaggccgcccgggcaaagcccgggcgtcgggcgacct
ttggtcgcccggcctcagtgagcgagcgagcgcgcagagagggagtggccaa [SEQ ID N.19]

pAB hGLA WILD TYPE plasmid

ttggccactccctctctgcgcgctcgctcgctcactgaggccgggcgaccaaaggtcgcccgacgcccgggctttgcccgggcggcctcagtga
gcgagcgagcgcgcagagagggagtggccaactccatcactagggggttcctggaggggtggagtcgtgacgtaaagatctgatatcatgat
cgcgatgcattaattaagcggccgctgtacataggaggttcgaaccctgctgaagggagaggttccaatactacaaaatgtagcgggatattgt
catcacctttggggacatgtcatcatggtccccagacagagttacaaaactcatcccctacacagcactatgtctctggtactgtttgttctacag
atgtcaacaacagaggcccagccatctcctattgcttggcttgtcagtctttctagcctccccattattaatttcaaatggggcaggtgttaggag
ggcaaaaatccacatattaagtgcaaagcctttcaggagatttcctgaaactagacaaaacccgtgtgactggcatcgattattctatttgatct
agctagtcctagcaaagtgacaactgctactcccctcctacacagccaagattcctaagttggcagtggcatgcttaatcctcaaagccaaagt
tacttggctccaagatttatagccttaaactgtggcctcacattccttcctatcttactttcctgcactggggtaaatgtctccttgctcttcttgcttt
ctgtcctactgcagGGCTCTTGCTGAGCTGGTGAAGCACAAGCCCAAGGCTACAGCGGAGCAACTGAAGACTGTC
ATGGATGACTTTGCACAGTTCCTGGATACATGTTGCAAGGCTGCTGACAAGGACACCTGCTTCTCGACTGAGGg
tcagaaacgttttgcattttgacgatgttcagtttccattttctgtgcacgtggtcaggtgtagctctctggaactcacacactgaataactccacc
aatctagatgttgttctctacgtaactgtaatagaaactgacttacgtagcttttaattttttattttctgccacactgctgcctattaaatacctattat
cactatttggtttcaaatttgtgacacagaagagcatagttagaaatacttgcaaagcctagaatcatgaactcatttaaaccttgccctgaaat
gtttcttttgaattgagttattttacacatgaatggacagttaccattatatatctgaatcatttcacattccctcccatggcctaacaacagtttat
cttcttattttgggcacaacagatgtcagagagcctgctttaggaattctaagtagaactgtaattaagcaatgcaaggcacgtacgtttactatg
tcattgcctatggctatgaagtgcaaatcctaACAGTCCTGCTAATACTTTTCtaacatccatcatttctttgttttcagGGTCCAAAC
CTTGTCACTAGATGCAAAGACGCCTTAGCCggaagcggcgccaccaatttcagcctgctgaaacaggccggcgacgtggaagag
aaccctggccctGCTAGCCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTGCGCTTCGCTTCCTGGCC
CTCGTTTCCTGGGACATCCCTGGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCCTACCATGGGCTGG
CTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCAGATTCCTGCATCAGTGAGAAGCTC
TTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGCTGGAAGGATGCAGGTTATGAGTACCTCTGCATTGATG
ACTGTTGGATGGCTCCCCAAAGAGATTCAGAAGGCAGACTTCAGGCAGACCCTCAGCGCTTTCCTCATGGGA
TTCGCCAGCTGGCTAATTATGTTCACAGCAAAGGACTGAAGCTAGGGATTTATGCAGATGTTGGAAATAAAAC
CTGCGCAGGCTTCCCTGGGAGTTTTGGATACTACGACATTGATGCCCAGACCTTTGCTGACTGGGGAGTAGAT
CTGCTAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAATTTGGCAGATGGTTATAAGCACATGTCCTTGGC
CCTGAATAGGACTGGCAGAAGCATTGTGTACTCCTGTGAGTGGCCTCTTTATATGTGGCCCTTTCAAAAGCCCA
ATTATACAGAAATCCGACAGTACTGCAATCACTGGCGAAATTTTGCTGACATTGATGATTCCTGGAAAAGTATA

AAGAGTATCTTGGACTGGACATCTTTTAACCAGGAGAGAATTGTTGATGTTGCTGGACCAGGGGGTTGGAAT
GACCCAGATATGTTAGTGATTGGCAACTTTGGCCTCAGCTGGAATCAGCAAGTAACTCAGATGGCCCTCTGGG
CTATCATGGCTGCTCCTTTATTCATGTCTAATGACCTCCGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGG
ATAAGGACGTAATTGCCATCAATCAGGACCCCTTGGGCAAGCAAGGGTACCAGCTTAGACAGGGAGACAACT

5 TTGAAGTGTGGGAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCGGCAGGAGATTGGTG
GACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAATCCTGCCTGCTTCATCACA
CAGCTCCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACTTCAAGGTTAAGAAGTCACATAAATCCCA
CAGGCACTGTTTTGCTTCAGCTAGAAAATACAATGCAGATGTCATTAAAAGACTTACTTTAAtgagctAGCttagcc
TAAacacatcacaaccacaaccttctcaggtaactatacttgggacttaaaaaacataatcataatcatttttcctaaaacgatcaagactgat

10 aaccatttgacaagagccatacagacaagcaccagctggcTCTCGAGCGTCTTCACGTATGGTCATCAgtttggggttccatttgtag
ataagaaactgaacatataaaggtctaggttaatgcaatttacacaaaaggagaccaaaccagggagagaaggaaccaaaattaaaaattc
aaaccagagcaaaggagttagccctggttttgctctgacttacatgaaccactatgtggagtcctccatgttagcctagtcaagcttatcctctgg
atgaagttgaaaccatatgaaggaatatttgggggggtgggtcaaaacagttgtgtatcaatgattccatgtggtttgacccaatcattctgtgaa
tccatttcaacagaagatacaacgggttctgtttcataataagtgatccacttccaaatttctgatgtgccccatgctaagctttaacagaatttat

15 cttcttatgacaaagcagcctcctttgaaaatatagccaactgcacacagctatgttgatcaattttgtttataatcttgcagaagagaatttttta
aaataagggcaataatggaaggctttggcaaaaaaattgtttctccatatgaaaacaaaaaacttatttttttattcaagcaaagaacctataga
cataaggctatttcaaaattatttcagttttagaaagaattgaaagttttgtagcattctgagaagacagctttcatttgtaatcataggtaatatgt
aggtcctcagaaatggtgagacccctgactttgacacttggggactctgagggaccagtgatgaagagggcacaacttatatcacacatgcac
gagttgggggtgagagggtgtcacaacatctatcagtgtgtcatctgcccaccaagtaacagatgtcagctaagactaggtcatgtgtaggctgt

20 ctacaccagtgaaaatcgcaaaaagaatctaagaaattccacatttctagaaaataggtttggaaaccgtattccattttacaaaggacactta
catttctcttttttgttttccagGCTACCCTGAGAAAAAAAGACATGAAGACTCAGGACTCATCTTTTCTGTTGGTGTAAA
ATCAACACCCTAAGGAACACAAATTTCTTTAAACATTTGACTTCTTGTCTCTGTGCTGCAATTAATAAAAAATGG
AAAGAATCTACtctgtggttcagaactctatcttccaaaggcgcgcttcaccctagcagcctctcttggctcagaggaatccctgcctttcctcc
cttcatctcagcagagaatgtagttccacatgggactagtgtacacgcgtgatatcagatctgttacgtagataagtagcatggcgggttaatca

25 ttaactacaaggaaccccctagtgatggagttggccactccctctctgcgcgctcgctcgctcactgaggccgcccgggcaaagcccgggcgtcg
ggcgacctttggtcgcccggcctcagtgagcgagcgagcgcgcagagagggagtggccaaagcgcgcagctgcctgcaggtcgactctagag
gatccccgggtaccgagctcgaattcactggccgtcgttttacaacgtcgtgactgggaaaaccctggcgttacccaacttaatcgccttgcagc
acatccccctttcgccagctggcgtaatagcgaagaggcccgcaccgatcgcccttcccaacagttgcgcagcctgaatggcgaatggcgcct
gatgcggtattttctccttacgcatctgtgcggtatttcacaccgcatacgtcaaagcaaccatagtacgcgccctgtagcggcgcattaagcgc

30 ggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccttagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcg
ccggctttccccgtcaagctctaaatcggggggctccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgatttgggtg
atggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactgg
aacaacactcaactctatctcgggctattctttttgatttataagggattttgccgatttcggtctattggttaaaaaatgagctgatttaacaaaaa
tttaacgcgaattttaacaaaatattaacgtttacaattttatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagccccgac

35 acccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtg
tcagaggttttcaccgtcatcaccgaaacgcgcgagacgaaagggcctcgtgatacgcctatttttataggttaatgtcatgataataatggtttc
ttagacgtcaggtggcacttttcggggaaatgtgcgcggaacccctatttgtttattttttctaaatacattcaaatatgtatccgctcatgagacaa
taaccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattcccttttttgcggcattttgcct
tcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaac

40 agcggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattg
acgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatg
gcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaagg

agctaaccgctttttttgcacaacatggggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagc

gtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaataga

ctggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgt

gggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatg

aacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgatttaa

aacttcattttttaatttaaaaggatctaggtgaagatcctttttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtca

gaccccgtagaaaagatcaaaggatcttcttgagatccttttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggt

ggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgttcttctagtgtagcc

gtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagt

cgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacggggggttcgtgcacacagcccagcttgg

agcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatcc

ggtaagcggcagggtcggaacaggagagcgcacgagggagcttccaggggggaaacgcctggtatctttatagtcctgtcgggtttcgccacct

ctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggcctttttacggttcctggccttttg

ctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagcc

gaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcatta

atgcagctggcacgacaggtttcccgactggaaagcgggcagtgagcgcaacgcaattaatgtgagttagctcactcattaggcaccccaggc

tttacactttatgcttccggctcgtatgttgtgtggaattgtgagcggataacaatttcacacaggaaacagctatgaccatgattacgccaagct

tgcatgcctgcaggcagctgcgcgctcgaacttcatgcctgccgaccttccccaggtcacgatccggacggcgggtgagttcacattttarcagc

cggacgtgcaractccgctggtggtctaacgtcggttaggtccctttgaatcacgggacatatgttggtgttggaggt [SEQ ID N.17]

pAB hGLA CO02 plasmid

ttggccactccctctctgcgcgctcgctcgctcactgaggccgggcgaccaaaggtcgcccgacgcccgggctttgcccgggcggcctcagtga

gcgagcgagcgcgcagagagggagtggccaactccatcactagggggttcctggagggggtggagtcgtgacgtaaagatctgatatcatcgat

cgcgatgcattaattaagcggccgctgtacataggaggttcgaaccctgctgaagggagaggttccaatactacaaaatgtagcgggatattgt

catcacctttggggacatgtcatcatggtccccagacagagttacaaaactcatccccctacacagcactatgtctctggtactgtttgttctacag

atgtcaacaacagaggcccagccatctcctattgcttggcttgtcagtctttctagcctccccattattaatttcaaatggggcaggtgttaggag

ggcaaaaatccacatattaagtgcaaagcctttcaggagatttcctgaaactagacaaaacccgtgtgactggcatcgattattctatttgatct

agctagtcctagcaaagtgacaactgctactcccctcctacacagccaagattcctaagttggcagtggcatgcttaatcctcaaagccaaagt

tacttggctccaagatttatagccttaaactgtggcctcacattccttcctatcttactttcctgcactggggtaaatgtctccttgctcttcttgcttt

ctgtcctactgcagGGCTCTTGCTGAGCTGGTGAAGCACAAGCCCAAGGCTACAGCGGAGCAACTGAAGACTGTC

ATGGATGACTTTGCACAGTTCCTGGATACATGTTGCAAGGCTGCTGACAAGGACACCTGCTTCTCGACTGAGg

tcagaaacgttttttgcattttgacgatgttcagtttccattttctgtgcacgtggtcaggtgtagctctctggaactcacacactgaataactccacc

aatctagatgttgttctctacgtaactgtaatagaaactgacttacgtagcttttaattttattttctgccacactgctgcctattaaatacctattat

cactatttggtttcaaatttgtgacacagaagagcatagttagaaatacttgcaaagcctagaatcatgaactcatttaaaccttgccctgaaat

gtttcttttttgaattgagttattttacacatgaatggacagttaccattatatatctgaatcatttcacattccctcccatggcctaacaacagtttat

cttcttattttgggcacaacagatgtcagagagcctgctttaggaattctaagtagaactgtaattaagcaatgcaaggcacgtacgtttactatg

tcattgcctatggctatgaagtgcaaatcctaACAGTCCTGCTAATACTTTTCtaacatccatcatttctttgttttcagGGTCCAAAC

CTTGTCACTAGATGCAAAGACGCCTTAGCCggaagcggcgccaccaatttcagcctgctgaaacaggccggcgacgtggaagag

aaccctggccctGCTAGCcagctgcgcaacccccgagctgcacctgggctgcgccctggccctgcgcttcctggccctggtcagctgggacat

ccccggcgcccgcgccctggacaacggcctggcccgcacccccaccatgggctggctgcactgggagcgcttcatgtgcaacctggactgcc

aggaggagcccgacagctgcatcagcgagaagctgtttatggagatggccgagctgatggtcagcgagggctggaaggacgccggctacga

gtacctgtgcatcgacgactgctggatggcccccagcgcgacagcgagggccgcctgcaggccgaccccagcgcttcccccacggaatcc
gccagctggccaactacgtgcacagcaagggcctgaagctgggcatctacgccgacgtgggcaacaagacctgcgccggcttccccggcagc
ttcggctactacgacatcgacgcccagaccttcgccgactggggcgtggacctgctgaagttcgacggctgctactgcgacagcctggagaac
ctggccgacggctacaagcacatgagcctggccctgaaccgcaccggccgcagcatcgtgtacagctgcgagtggcccctgtatatgtggccc

ttccagaagcccaactacaccgagatccgccagtactgcaaccactggcgcaacttcgccgacatcgacgacagctggaagagcatcaagag
catcctggactggaccagcttcaaccaggagcgcatcgtggacgtggccggccccggcggctggaacgaccccgacatgctggtgatcggca
acttcggcctgagctggaaccagcaggtgacccagatggccctgtgggccattatggccgccccccctgtttatgagcaacgacctgcgccacat
cagcccccaggccaaggccctgctgcaggacaaggacgtgatcgctatcaaccaggaccccctgggcaagcagggctaccagctgcgccag
ggcgacaacttcgaggtctgggagcgccccctgagcggcctggcctgggccgtggctatgatcaaccgccaggagatcggcggcccccgcag

ctacaccatcgccgtggccagcctgggcaagggcgtggcctgcaaccccgcctgcttcatcacccagctgctgcccgtgaagcgcaagctggg
cttctacgagtggaccagccgcctgcgcagccacatcaaccccaccggcaccgtgctgctgcagctggagaacacaatgcagatgagcctga
aggacctgctgtaatgagctAGCttagccTAAacacatcacaaccacaaccttctcaggtaactatacttgggacttaaaaaacataatcat
aatcattttcctaaaacgatcaagactgataaccatttgacaagagccatacagacaagcaccagctggcTCTCGAGCGTCTTCACGT
ATGGTCATCagtttgggttccatttgtagataagaaactgaacatataaaggtctaggttaatgcaatttacacaaaaggagaccaaacca

gggagagaaggaaccaaaattaaaaattcaaaccagagcaaaggagttagccctggtttttgctctgacttacatgaaccactatgtggagtcc
tccatgttagcctagtcaagcttatcctctggatgaagttgaaaccatatgaaggaatatttggggggtgggtcaaaacagttgtgtatcaatga
ttccatgtggtttgacccaatcattctgtgaatccatttcaacagaagatacaacgggttctgtttcataataagtgatccacttccaaatttctgat
gtgccccatgctaagctttaacagaatttatcttcttatgacaaagcagcctcctttgaaaatatagccaactgcacacagctatgttgatcaattt
tgtttataatcttgcagaagagaatttttttaaaataggggcaataatggaaggctttggcaaaaaaattgtttctccatatgaaaacaaaaaactt

atttttttattcaagcaaagaacctatagacataaggctatttcaaaattatttcagttttagaaagaattgaaagtttttgtagcattctgagaaga
cagctttcatttgtaatcataggtaatatgtaggtcctcagaaatggtgagacccctgactttgacacttggggactctgagggaccagtgatga
agagggcacaacttatatcacacatgcacgagttgggggtgagagggtgtcacaacatctatcagtgtgtcatctgcccaccaagtaacagatgt
cagctaagactaggtcatgtgtaggctgtctacaccagtgaaaatcgcaaaaagaatctaagaaattccacatttctagaaaataggtttggaa
accgtattccatttacaaaggacacttacatttctcttttttgttttccagGCTACCCTGAGAAAAAAAGACATGAAGACTCAGGA

CTCATCTTTTCTGTTGGTGTAAAATCAACACCCTAAGGAACACAAATTTCTTTAAACATTTGACTTCTTGTCTCT
GTGCTGCAATTAATAAAAAATGGAAAGAATCTACtctgtggttcagaactctatcttccaaaggcgcgcttcaccctagcagcctc
tttggctcagaggaatccctgcctttcctcccttcatctcagcagagaatgtagttccacatgggactagtgtacacgcgtgatatcagatctgtta
cgtagataagtagcatggcgcggttaatcattaactacaaggaacccctagtgatggagttggccactccctctctgcgcgctcgctcgctcactg
aggccgcccgggcaaagcccgggcgtcggcgacctttggtcgcccggcctcagtgagcgagcgagcgcgcagagagggagtggccaaag

cgcgcagctgcctgcaggtcgactctagaggatccccgggtaccgagctcgaattcactggccgtcgttttacaacgtcgtgactgggaaaacc
ctggcgttacccaacttaatcgccttgcagcacatccccctttcgccagctggcgtaatagcgaagaggcccgcaccgatcgcccttcccaaca
gttgcgcagcctgaatggcgaatggcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcatacgtcaaagcaaccata
gtacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccttagcgcccgctcctt
tcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcggggggctccctttagggttccgatttagtgctttacggc

acctcgaccccaaaaaacttgatttgggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgt
tctttaatagtggactcttgttccaaactggaacaacactcaactctatctcgggctattcttttgatttataagggattttgccgatttcggtctatt
ggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaattttatggtgcactctcagtacaatctgct
ctgatgccgcatagttaagccagccccgacacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagaca
agctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgagacgaaagggcctcgtgatacgcctatttt

tataggttaatgtcatgataataatggtttcttagacgtcaggtggcacttttcggggaaatgtgcgcggaacccctatttgtttattttctaaata
cattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgt
cgcccttattcccttttttgcggcattttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgca

49

cgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagtt
ctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactc
accagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaactt
acttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggag

ctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactactt
actctagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttat
tgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctaca
cgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaa
gtttactcatatatactttagattgatttaaaacttcattttaatttaaaaggatctaggtgaagatcctttttgataatctcatgaccaaaatccct

taacgtgagttttcgttccactgagcgtcagacccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgc
aaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgca
gataccaaatactgttcttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgtta
ccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacg
ggggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcc

cgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccaggggggaaacgcctggt
atctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagca
acgcggcctttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctt
tgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcccaatacgcaaacc
gcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgactggaaagcgggcagtgagcgcaacgcaattaatgtg

agttagctcactcattaggcaccccaggctttacactttatgcttccggctcgtatgttgtgtggaattgtgagcggataacaatttcacacagga
aacagctatgaccatgattacgccaagcttgcatgcctgcaggcagctgcgcgctcgaacttcatgcctgccgaccttccccaggtcacgatcc
ggacggcgggtgagttcacattttarcagccggacgtgcaractccgctggtggtctaacgtcggttaggtcccttgaatcacgggacatatgtt
ggtgttggaggt [SEQ ID N.15]

pSMD2 hGLA WT

AGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGGTCGC
CCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCCTTGTAGTT
AATGATTAACCCGCCATGCTACTTATCTACGTAGCCATGCTCTAGAGGATCAAGGCTCAGAGGCACACAGGAG

TTTCTGGGCTCACCCTGCCCCCTTCCAACCCCTCAGTTCCCATCCTCCAGCAGCTGTTTGTGTGCTGCCTCTGA
AGTCCACACTGAACAAACTTCAGCCTACTCATGTCCCTAAAATGGGCAAACATTGCAAGCAGCAAACAGCAAA
CACACAGCCCTCCCTGCCTGCTGACCTTGGAGCTGGGGCAGAGGTCAGAGACCTCTCTGGGCCCATGCCACC
TCCAACATCCACTCGACCCCTTGGAATTTCGGTGGAGAGGAGCAGAGGTTGTCCTGGCGTGGTTTAGGTAGT
GTGAGAGGGGTACCCGGGGATCTTGCTACCAGTGGAACAGCCACTAAGGATTCTGCAGTGAGAGCAGAGGG

CCAGCTAAGTGGTACTCTCCCAGAGACTGTCTGACTCACGCCACCCCCTCCACCTTGGACACAGGACGCTGTG
GTTTCTGAGCCAGGTACAATGACTCCTTTCGGTAAGTGCAGTGGAAGCTGTACACTGCCCAGGCAAAGCGTC
CGGGCAGCGTAGGCGGGCGACTCAGATCCCAGCCAGTGGACTTAGCCCTGTTTGCTCCTCCGATAACTGGG
GTGACCTTGGTTAATATTCACCAGCAGCCTCCCCCGTTGCCCCTCTGGATCCACTGCTTAAATACGGACGAGGA
CAGGGCCCTGTCTCCTCAGCTTCAGGCACCACCACTGACCTGGGACAGTGAATAGATCCTGAGAACTTCAGG

GTGAGTCTATGGGACCCTTGATGTTTTCTTTCCCCTTCTTTTCTATGGTTAAGTTCATGTCATAGGAAGGGGAG
AAGTAACAGGGTACACATATTGACCAAATCAGGGTAATTTTGCATTTGTAATTTTAAAAAATGCTTTCTTCTTTT
AATATACTTTTTTGTTTATCTTATTTCTAATACTTTCCCTAATCTCTTTCTTTCAGGGCAATATTGATACAATGTATC

TTGCCTCTTTGCACCATTCTAAAGAATAACAGTGATAATTTCTGGGTTAAGGCAATAGCAATATTTCTGCATATA
AATATTTCTGCATATAAATTGTAACTGATGTAAGAGGTTTCATATTGCTAATAGCAGCTACAATCCAGCTACCATT
CTGCTTTTATTTTCTGGTTGGGATAAGGCTGGATTATTCTGAGTCCAAGCTAGGCCCTTTTGCTAATCTTGTTCA
TACCTCTTATCTTCCTCCCACAGCTCCTGGGCAACCTGCTGGTCTCTCTGCTGGCCCATCACTTTGGCAAAGCA
CGCGTCgacgccgccaccATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTGCGCTTCGCTTCCTG
GCCCTCGTTCCTGGGACATCCCTGGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCCTACCATGGGC
TGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCAGATTCTGCATCAGTGAGAAG
CTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGCTGGAAGGATGCAGGTTATGAGTACCTCTGCATTG
ATGACTGTTGGATGGCTCCCCAAAGAGATTCAGAAGGCAGACTTCAGGCAGACCCTCAGCGCTTTCCTCATG
GGATTCGCCAGCTGGCTAATTATGTTCACAGCAAAGGACTGAAGCTAGGGATTTATGCAGATGTTGGAAATAA
AACCTGCGCAGGCTTCCCTGGGAGTTTTGGATACTACGACATTGATGCCCAGACCTTTGCTGACTGGGGAGTA
GATCTGCTAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAATTTGGCAGATGGTTATAAGCACATGTCCTT
GGCCCTGAATAGGACTGGCAGAAGCATTGTGTACTCCTGTGAGTGGCCTCTTTATATGTGGCCCTTTCAAAAG
CCCAATTATACAGAAATCCGACAGTACTGCAATCACTGGCGAAATTTTGCTGACATTGATGATTCCTGGAAAAG
TATAAAGAGTATCTTGGACTGGACATCTTTTAACCAGGAGAGAATTGTTGATGTTGCTGGACCAGGGGGTTGG
AATGACCCAGATATGTTAGTGATTGGCAACTTTGGCCTCAGCTGGAATCAGCAAGTAACTCAGATGGCCCTCT
GGGCTATCATGGCTGCTCCTTTATTCATGTCTAATGACCTCCGACACATCAGCCCTCAAGCCAAAGCTCTCCTTC
AGGATAAGGACGTAATTGCCATCAATCAGGACCCCTTGGGCAAGCAAGGGTACCAGCTTAGACAGGGAGAC
AACTTTGAAGTGTGGGAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCGGCAGGAGATT
GGTGGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAATCCTGCCTGCTTCAT
CACACAGCTCCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACTTCAAGGTTAAGAAGTCACATAAAT
CCCACAGGCACTGTTTTGCTTCAGCTAGAAAATACAATGCAGATGTCATTAAAAGACTTACTTTAAtgagctAGC
TCGAGAGATCTGAATTCACCCCACCAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCT
GGCCCACAAGTATCACTAAGCTCGCTTTCTTGCTGTCCAATTTCTATTAAAGGTTCCTTTGTTCCCTAAGTCCAA
CTACTAAACTGGGGGATATTATGAAGGGCCTTGAGCATCTGGATTCTGCCTAATAAAAAACATTTATTTTCATTG
CAATGATGTATTTAAATTATTTCTGAATATTTTACTAAAAAGGGAATGTGGGAGGTCAGTGCATTTAAAACATAA
AGAAATGAAGAGCTAGTTCAAACCTTGGGAAAATACACTATATCTTAAACTCCATGAAAGAAGGTGAGGCTGC
AAACAGCTAATGCACATTGGCAACAGCCCCTGATGCCTATGCCTTATTCATCCCTCAGAAAAGGATTCAAGTAG
AGGCTTGATTTGGAGGTTAAAGTTTGGCTATGCTGTATTTTACATTACTTATTGTTTTAGCTGTCCTCATGAATGT
CTTTTCACTACCCATTTGCTTATCCTGCATCTCTCAGCCTTGACTCCACTCAGTTCTCTTGCTTAGAGATACCACC
TTTCCCCTGAAGTGTTCCTTCCATGTTTTACGGCGAGATGGTTTCTCCTCGCCTGGCCACTCAGCCTTAGTTGTC
TCTGTTGTCTTATAGAGGTCTACTTGAAGAAGGAAAAACAGGGGGCATGGTTTGACTGTCCTGTGAGCCCTTC
TTCCCTGCCTCCCCCACTCACAGTGACCCGGAATCCCTCGACATGGCAGTCTAGAGCATGGCTACGTAGATAA
GTAGCATGGCGGGTTAATCATTAACTACAAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCT
CGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAG
CGAGCGAGCGCGCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCC
TGAATGGCGAATGGCGATTCCGTTGCAATGGCTGGCGGTAATATTGTTCTGGATATTACCAGCAAGGCCGATA
GTTTGAGTTCTTCTACTCAGGCAAGTGATGTTATTACTAATCAAAGAAGTATTGCGACAACGGTTAATTTGCGT
GATGGACAGACTCTTTTACTCGGTGGCCTCACTGATTATAAAAACACTTCTCAGGATTCTGGCGTACCGTTCCT
GTCTAAAATCCCTTTAATCGGCCTCCTGTTTAGCTCCCGCTCTGATTCTAACGAGGAAAGCACGTTATACGTGCT
CGTCAAAGCAACCATAGTACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGC
GTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGC

CGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACC
CCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGAC
GTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTC
TTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGC
GAATTTTAACAAATATTAACGCTTACAATTTAAATATTTGCTTATACAATCTTCCTGTTTTTGGGGCTTTTCTGAT
TATCAACCGGGGTACATATGATTGACATGCTAGTTTTACGATTACCGTTCATCGATTCTCTTGTTTGCTCCAGACT
CTCAGGCAATGACCTGATAGCCTTTGTAGAGACCTCTCAAAAATAGCTACCCTCTCCGGCATGAATTTATCAGC
TAGAACGGTTGAATATCATATTGATGGTGATTTGACTGTCTCCGGCCTTTCTCACCCGTTTGAATCTTTACCTAC
ACATTACTCAGGCATTGCATTTAAAATATATGAGGGTTCTAAAAATTTTTATCCTTGCGTTGAAATAAAGGCTTC
TCCCGCAAAAGTATTACAGGGTCATAATGTTTTTGGTACAACCGATTTAGCTTTATGCTCTGAGGCTTTATTGCT
TAATTTTGCTAATTCTTTGCCTTGCCTGTATGATTTATTGGATGTTGGAATCGTCCTGATGCGGTATTTTCTCCTTA
CGCATCTGTGCGGTATTTCACACCGCATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCC
AGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGAC
AAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGACGAA
AGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTT
TTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGAC
AATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTA
TTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAA
GATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGC
CCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGC
CGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAA
AAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGG
CCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATG
TAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGC
CTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTA
ATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATT
GCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCC
TCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGA
TAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACT
TCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTT
TTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAA
TCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCT
TTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGC
CACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAG
TGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTG
AACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTG
AGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGG
AACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCC
ACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACG
CGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTG
TGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGT
CAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAAT

G [SEQ ID N.18]

pSMD2 hGLA CO02

AGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGGTCGC
CCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCCTTGTAGTT
AATGATTAACCCGCCATGCTACTTATCTACGTAGCCATGCTCTAGAGGATCAAGGCTCAGAGGCACACAGGAG
TTTCTGGGCTCACCCTGCCCCCTTCCAACCCCTCAGTTCCCATCCTCCAGCAGCTGTTTGTGTGCTGCCTCTGA
AGTCCACACTGAACAAACTTCAGCCTACTCATGTCCCTAAAATGGGCAAACATTGCAAGCAGCAAACAGCAAA
CACACAGCCCTCCCTGCCTGCTGACCTTGGAGCTGGGGCAGAGGTCAGAGACCTCTCTGGGCCCATGCCACC
TCCAACATCCACTCGACCCCTTGGAATTTCGGTGGAGAGGAGCAGAGGTTGTCCTGGCGTGGTTTAGGTAGT
GTGAGAGGGGTACCCGGGGATCTTGCTACCAGTGGAACAGCCACTAAGGATTCTGCAGTGAGAGCAGAGGG
CCAGCTAAGTGGTACTCTCCCAGAGACTGTCTGACTCACGCCACCCCCTCCACCTTGGACACAGGACGCTGTG
GTTTCTGAGCCAGGTACAATGACTCCTTTCGGTAAGTGCAGTGGAAGCTGTACACTGCCCAGGCAAAGCGTC
CGGGCAGCGTAGGCGGGCGACTCAGATCCCAGCCAGTGGACTTAGCCCCTGTTTGCTCCTCCGATAACTGGG
GTGACCTTGGTTAATATTCACCAGCAGCCTCCCCCGTTGCCCCTCTGGATCCACTGCTTAAATACGGACGAGGA
CAGGGCCCTGTCTCCTCAGCTTCAGGCACCACCACTGACCTGGGACAGTGAATAGATCCTGAGAACTTCAGG
GTGAGTCTATGGGACCCTTGATGTTTTCTTTCCCCTTCTTTTCTATGGTTAAGTTCATGTCATAGGAAGGGGAG
AAGTAACAGGGTACACATATTGACCAAATCAGGGTAATTTTGCATTTGTAATTTTAAAAAATGCTTTCTTCTTTT
AATATACTTTTTTGTTTATCTTATTTCTAATACTTTCCCTAATCTCTTTCTTTCAGGGCAATATTGATACAATGTATC
TTGCCTCTTTGCACCATTCTAAAGAATAACAGTGATAATTTCTGGGTTAAGGCAATAGCAATATTTCTGCATATA
AATATTTCTGCATATAAATTGTAACTGATGTAAGAGGTTTCATATTGCTAATAGCAGCTACAATCCAGCTACCATT
CTGCTTTTATTTTCTGGTTGGGATAAGGCTGGATTATTCTGAGTCCAAGCTAGGCCCTTTTGCTAATCTTGTTCA
TACCTCTTATCTTCCTCCCACAGCTCCTGGGCAACCTGCTGGTCTCTCTGCTGGCCCATCACTTTGGCAAAGCA
CGCgtcgacgccgccaccatgcagctgcgcaaccccgagctgcacctgggctgcgccctggccctgcgcttcctggccctggtcagctggga
catccccggcgcccgcgccctggacaacggcctggcccgcacccccaccatgggctggctgcactgggagcgcttcatgtgcaacctggactg
ccaggaggagcccgacagctgcatcagcgagaagctgtttatggagatggccgagctgatggtcagcgagggctggaaggacgccggctac
gagtacctgtgcatcgacgactgctggatggcccccagcgcgacagcgagggccgcctgcaggccgaccccagcgcttcccccacggaat
ccgccagctggccaactacgtgcacagcaagggcctgaagctgggcatctacgccgacgtgggcaacaagacctgcgccggcttccccggca
gcttcggctactacgacatcgacgcccagaccttcgccgactggggcgtggacctgctgaagttcgacggctgctactgcgacagcctggaga
acctggccgacggctacaagcacatgagcctggccctgaaccgcaccggccgcagcatcgtgtacagctgcgagtggcccctgtatatgtggc
ccttccagaagcccaactacaccgagatccgccagtactgcaaccactggcgcaacttcgccgacatcgacgacagctggaagagcatcaag
agcatcctggactggaccagcttcaaccaggagcgcatcgtggacgtggccggccccggcggctggaacgaccccgacatgctggtgatcgg
caacttcggcctgagctggaaccagcaggtgacccagatggccctgtgggccattatggccgccccccctgtttatgagcaacgacctgcgcca
catcagcccccaggccaaggccctgctgcaggacaaggacgtgatcgctatcaaccaggacccctgggcaagcagggctaccagctgcgc
cagggcgacaacttcgaggtctgggagcgccccctgagcggcctggcctgggccgtggctatgatcaaccgccaggagatcggcggccccg
cagctacaccatcgccgtggccagcctgggcaagggcgtggcctgcaaccccgcctgcttcatcacccagctgctgcccgtgaagcgcaagct
gggcttctacgagtggaccagccgcctgcgcagccacatcaaccccaccggcaccgtgctgctgcagctggagaacacaatgcagatgagcc
tgaaggacctgctgtaatgagctagcTCGAGAGATCTGAATTCACCCCACCAGTGCAGGCTGCCTATCAGAAAGTGGT
GGCTGGTGTGGCTAATGCCCTGGCCCACAAGTATCACTAAGCTCGCTTTCTTGCTGTCCAATTTCTATTAAAGG
TTCCTTTGTTCCCTAAGTCCAACTACTAAACTGGGGGATATTATGAAGGGCCTTGAGCATCTGGATTCTGCCTA
ATAAAAAACATTTATTTTCATTGCAATGATGTATTTAAATTATTTCTGAATATTTTACTAAAAAGGGAATGTGGGA

GGTCAGTGCATTTAAAACATAAAGAAATGAAGAGCTAGTTCAAACCTTGGGAAAATACACTATATCTTAAACTC
CATGAAAGAAGGTGAGGCTGCAAACAGCTAATGCACATTGGCAACAGCCCCTGATGCCTATGCCTTATTCATC
CCTCAGAAAAGGATTCAAGTAGAGGCTTGATTTGGAGGTTAAAGTTTGGCTATGCTGTATTTTACATTACTTAT
TGTTTTAGCTGTCCTCATGAATGTCTTTTCACTACCCATTTGCTTATCCTGCATCTCTCAGCCTTGACTCCACTCA

5   GTTCTCTTGCTTAGAGATACCACCTTTCCCCTGAAGTGTTCCTTCCATGTTTTACGGCGAGATGGTTTCTCCTCG
CCTGGCCACTCAGCCTTAGTTGTCTCTGTTGTCTTATAGAGGTCTACTTGAAGAAGGAAAAACAGGGGGCATG
GTTTGACTGTCCTGTGAGCCCTTCTTCCCTGCCTCCCCACTCACAGTGACCCGGAATCCCTCGACATGGCAGT
CTAGAGCATGGCTACGTAGATAAGTAGCATGGCGGGTTAATCATTAACTACAAGGAACCCCTAGTGATGGAGT
TGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCCGGGCT

10 TTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGC
CCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGATTCCGTTGCAATGGCTGGCGGTAATATTGTTCTG
GATATTACCAGCAAGGCCGATAGTTTGAGTTCTTCTACTCAGGCAAGTGATGTTATTACTAATCAAAGAAGTATT
GCGACAACGGTTAATTTGCGTGATGGACAGACTCTTTTACTCGGTGGCCTCACTGATTATAAAAACACTTCTCA
GGATTCTGGCGTACCGTTCCTGTCTAAAATCCCTTTAATCGGCCTCCTGTTTAGCTCCCGCTCTGATTCTAACGA

15 GGAAAGCACGTTATACGTGCTCGTCAAAGCAACCATAGTACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGG
GTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCC
TTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTA
GTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATA
GACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACA

20 CTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAG
CTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGCTTACAATTTAAATATTTGCTTATACAATCT
TCCTGTTTTTGGGGCTTTTCTGATTATCAACCGGGGTACATATGATTGACATGCTAGTTTTACGATTACCGTTCAT
CGATTCTCTTGTTTGCTCCAGACTCTCAGGCAATGACCTGATAGCCTTTGTAGAGACCTCTCAAAAATAGCTAC
CCTCTCCGGCATGAATTTATCAGCTAGAACGGTTGAATATCATATTGATGGTGATTTGACTGTCTCCGGCCTTTC

25 TCACCCGTTTGAATCTTTACCTACACATTACTCAGGCATTGCATTTAAAATATATGAGGGTTCTAAAAATTTTTAT
CCTTGCGTTGAAATAAAGGCTTCTCCCGCAAAAGTATTACAGGGTCATAATGTTTTTGGTACAACCGATTTAGC
TTTATGCTCTGAGGCTTTATTGCTTAATTTTGCTAATTCTTTGCCTTGCCTGTATGATTTATTGGATGTTGGAATC
GTCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATGGTGCACTCTCAGTACAATC
TGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTC

30 TGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTC
ATCACCGAAACGCGCGAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGG
TTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATT
CAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTA
TTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGC

35 TGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCG
GTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGC
GCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGT
TGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATA
ACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTT

40 TGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACG
ACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTAC
TCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCC

CTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCAC
TGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACG
AAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATA
TACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGAC
CAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGA
GATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCC
GGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTT
CTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCT
GTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGAT
AAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCG
AACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTAT
CCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTT
ATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCT
ATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTC
CTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGA
ACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCG
CGCGTTGGCCGATTCATTAATG [SEQ ID N.10]

BIBLIOGRAPHY

Aerts, J.M., Groener, J.E., Kuiper, S., Donker-Koopman, W.E., Strijland, A., Ottenhoff, R., van Roomen, C., Mirzaian, M., Wijburg, F.A., Linthorst, G.E., Vedder, A.C., Rombach, S.M., Cox-Brinkman, J., Somerharju, P., Boot, R.G., Hollak, C.E., Brady, R.O., and Poorthuis, B.J. (2008). Elevated globotriaosylsphingosine is a hallmark of Fabry disease. Proceedings of the National Academy of Sciences *105*, 2812-2817.

Anderson, A. (1898). A case of "Angeio-Keratoma.". Br J Dermatol *10*, 113-117.

Askari, H., Kaneski, C.R., Semino-Mora, C., Desai, P., Ang, A., Kleiner, D.E., Perlee, L.T., Quezado, M., Spollen, L.E., Wustman, B.A., and Schiffmann, R. (2007). Cellular and tissue localization of globotriaosylceramide in Fabry disease. Virchows Archiv *451*, 823-834.

Auray-Blais, C., Cyr, D., Ntwari, A., West, M.L., Cox-Brinkman, J., Bichet, D.G., Germain, D.P., Laframboise, R., Melançon, S.B., Stockley, T., Clarke, J.T.R., and Drouin, R. (2008). Urinary globotriaosylceramide excretion correlates with the genotype in children and adults with Fabry disease. Molecular genetics and metabolism *93*, 331-340.

Barzel, A., Paulk, N.K., Shi, Y., Huang, Y., Chu, K., Zhang, F., Valdmanis, P.N., Spector, L.P., Porteus, M.H., Gaensler, K.M., and Kay, M.A. (2015). Promoterless gene targeting without nucleases ameliorates haemophilia B in mice. Nature *517*, 360-364.

Bengtsson, B.-Å., Johansson, J.-O., Hollak, C., Linthorst, G., and FeldtRasmussen, U. (2003). Enzyme replacement in Anderson-Fabry disease. The Lancet *361*, 352.

Bernardes, T.P., Foresto, R.D., and Kirsztajn, G.M. (2020). Fabry disease: genetics, pathology, and treatment. Revista da Associação Médica Brasileira *66*, s10-s16.

Brady, R.O., Gal, A.E., Bradley, R.M., Martensson, E., Warshaw, A.L., and Laster, L. (1967). Enzymatic Defect in Fabry's Disease. New England Journal of Medicine *276*, 1163-1167.

De Caneva, A., Porro, F., Bortolussi, G., Sola, R., Lisjak, M., Barzel, A., Giacca, M., Kay, M.A., Vlahovicek, K., Zentilin, L., and Muro, A.F. (2019). Coupling AAV-mediated promoterless gene targeting to SaCas9 nuclease to efficiently correct liver metabolic diseases. JCI insight *5* (15): :e128863.

Desnick, R.J. (2013). Fabry Disease (α-Galactosidase A Deficiency). 8-11.

Eng, C., Resnick-Silverman, L.A., Niehaus, D.J., and Desnick, R.J. (1993). Nature and frequency of mutations in the alpha-galactosidase A gene that cause Fabry disease. American journal of human genetics *53*, 1186-1197.

Eng, C.M., and Desnick, R.J. (1994). Molecular basis of fabry disease: Mutations and polymorphisms in the human α-galactosidase A gene. Human mutation *3*, 103-111.

Eng, C.M., Fletcher, J., Wilcox, W.R., Waldek, S., Scott, C.R., Sillence, D.O., Breunig, F., Charrow, J., Germain, D.P., Nicholls, K., and Banikazemi, M. (2007). Fabry disease: Baseline medical characteristics of a cohort of 1765 males and females in the Fabry Registry. Journal of inherited metabolic disease *30*, 184-192.

Garman, S.C., and Garboczi, D.N. (2004). The Molecular Defect Leading to Fabry Disease: Structure of Human α-Galactosidase. Journal of Molecular Biology *337*, 319-335.

Germain, D.P. (2010). Fabry disease. Orphanet journal of rare diseases *5*, 30.

Germain, D.P., Charrow, J., Desnick, R.J., Guffon, N., Kempf, J., Lachmann, R.H., Lemay, R., Linthorst, G.E., Packman, S., Scott, C.R., Waldek, S., Warnock, D.G., Weinreb, N.J., and Wilcox, W.R. (2015). Ten-year outcome of enzyme replacement therapy with agalsidase beta in patients with Fabry disease. Journal of medical genetics *52*, 353-358.

Ioannou, Y.A., Zeidner, K.M., Gordon, R.E., and Desnick, R.J. (2001). Fabry Disease: Preclinical Studies Demonstrate the Effectiveness of α-Galactosidase A Replacement in Enzyme-Deficient Mice. The American Journal of Human Genetics *68*, 14-25.

Laney, D.A., and Fernhoff, P.M. (2008). Diagnosis of Fabry Disease via Analysis of Family History. Journal of Genetic Counseling *17*, 79-83.

Lubanda, J.-C., Anijalg, E., Bzdúch, V., Thurberg, B.L., Bénichou, B., and Tylki-Szymanska, A. (2009). Evaluation of a low dose, after a standard therapeutic dose, of agalsidase beta during enzyme replacement therapy in patients with Fabry disease. Genetics in Medicine *11*, 256-264.

Markham, A. (2016). Migalastat: First Global Approval. Drugs *76*, 1147-1152.

Mehta, A., Ricci, R., Widmer, U., Dehout, F., Garcia de Lorenzo, A., Kampmann, C., Linhart, A., Sunder-Plassmann, G., Ries, M., and Beck, M. (2004). Fabry disease defined: baseline clinical manifestations of 366 patients in the Fabry Outcome Survey. European journal of clinical investigation *34*, 236-242.

Moran, N. (2018). FDA approves Galafold, a triumph for Amicus. Nature biotechnology *36*, 913-913.

Motabar, O., Sidransky, E., Goldin, E., and Zheng, W. (2010). Fabry Disease – Current Treatment and New Drug Development. Current Chemical Genomics *4*, 50-56.

Parkinson-Lawrence, E.J., Shandala, T., Prodoehl, M., Plew, R., Borlace, G.N., and Brooks, D.A. (2010). Lysosomal Storage Disease: Revealing Lysosomal Function and Physiology. Physiology *25*, 102-115.

Pastores, G.M. (2007). Agalsidase alfa (Replagal) in the treatment of Anderson-Fabry disease. Biol Targets Ther *1*, 291-300.

Ronzitti, G., Bortolussi, G., van Dijk, R., Collaud, F., Charles, S., Leborgne, C., Vidal, P., Martin, S., Gjata, B., Sola, M.S., van Wittenberghe, L., Vignaud, A., Veron, P., Bosma, P.J., Muro, A.F., and Mingozzi, F. (2016). A translationally optimized AAV-UGT1A1 vector drives safe and long-lasting correction of Crigler-Najjar syndrome. Molecular therapy Methods & clinical development *3*, 16049.

Saito, S., Ohno, K., and Sakuraba, H. (2011). Fabry-database.org: database of the clinical phenotypes, genotypes and mutant α-galactosidase A structures in Fabry disease. Journal of Human Genetics *56*, 467-468.

Spada, M., Pagliardini, S., Yasuda, M., Tukel, T., Thiagarajan, G., Sakuraba, H., Ponzone, A., and Desnick, R.J. (2006). High Incidence of Later-Onset Fabry Disease Revealed by Newborn Screening*. The American Journal of Human Genetics *79*, 31-40.

van der Tol, L., Smid, B.E., Poorthuis, B.J.H.M., Biegstraaten, M., Deprez, R.H.L., Linthorst, G.E., and Hollak, C.E.M. (2014). A systematic review on screening for Fabry disease: prevalence of individuals with genetic variants of unknown significance. Journal of medical genetics *51*, 1-9.

Zarate, Y.A., and Hopkin, R.J. (2008). Fabry's disease. The Lancet *372*, 1427-1435.

CLAIMS

1. A nucleic acid comprising or consisting of a sequence having at least 99.5% of identity with the sequence of SEQ ID N.1, SEQ ID N.2 or SEQ ID N.3.

2. The nucleic acid according to claim 1 which comprises or consists of a sequence with SEQ ID N.1.

3. The nucleic acid according to claim 1 which comprises or consists of a sequence with SEQ ID N.2 or which comprises or consists of a sequence with SEQ ID N.3.

4. The nucleic acid according to any one of preceding claims for use for the *in vivo* expression of human alpha galactosidase A (GLA), preferably in a human cell.

5. A nucleic acid construct comprising:

- a promoter sequence and

- a coding sequence of the alpha-galactosidase A (GLA) gene under control of said promoter, wherein said coding sequence is the nucleic acid of any one of claims 1-3.

6. A vector comprising the nucleic acid of any one of claims 1-3 or the nucleic acid construct of claim 5, preferably said vector is a viral vector selected from the group consisting of: adenoviral vectors, lentiviral vectors, retroviral vectors and adeno associated viral vectors (AAV).

7. The vector according to claim 6 which further comprises one or more of:

- a 5' inverted terminal repeat (ITR) sequence of an AAV, preferably localized at the 5' end of the promoter;

- an enhancer element, preferably localized at the 5' end of the promoter;

- a promoter sequence;

- a Kozak sequence, preferably localized at the 5' end of the GLA-coding sequence of claim 1 or 2 and operably linked to said sequence;

- a transcription termination sequence preferably localized at the 3'end of the GLA coding sequence of any one of claims 1-3;

- a 3' inverted terminal repeat (ITR) sequence of an AAV, preferably localized at the 3' end of the transcription termination sequence,

59

said vector preferably comprising in a 5'-3' direction:

  - an AAV 5'-inverted terminal repeat (5'-ITR) sequence;

  - an enhancer sequence;

  - a promoter sequence;

  - a Kozak sequence;

  - the nucleic acid of any one of claims 1-3 as GLA-coding sequence;

  - a transcription termination sequence; and

  - an AAV 3'-inverted terminal repeat (3'-ITR) sequence.

8. The vector according to claim 7 wherein said enhancer element is an enhancer derived from apolipoprotein E gene, and/or said transcription termination sequence is a poly-adenylation signal sequence, preferably the human hemoglobin beta a poly-adenylation signal (HHB polyA), and/or the ITRs derive from the same virus serotype or from different virus serotypes, preferably the virus is an AAV, preferably of serotype 2, and/or the vector further comprises an intron, preferably the human hemoglobin beta-derived synthetic intron (HBB2), preferably operably linked to the 3' of the promoter sequence.

9. An integrative vector comprising at least:

- a coding sequence of the alpha-galactosidase A (GLA) gene, wherein said coding sequence is the nucleic acid of any one of claims 1-3, and

- one or more albumin sequences,

preferably said vector comprising:

- an AAV 5'-inverted terminal repeat (5'-ITR) sequence;

- a first genomic albumin sequence;

- a ribosomal skipping sequence, preferably P2A;

- the nucleic acid of any one of claims 1-3 as GLA-coding sequence;

- a second genomic albumin sequence comprising a protospacer-adjacent motif (PAM) sequence;

- an AAV 3'-inverted terminal repeat (3'-ITR) sequence.

10. A plasmid comprising:

- a promoter sequence, and

- a coding sequence of the alpha-galactosidase A (GLA) gene under control of said promoter, wherein said coding sequence is the nucleic acid of any one of claims 1-3,

preferably said plasmid comprising or consisting of a sequence with SEQ ID N.10.

11. A plasmid comprising:

- the coding sequence of the alpha-galactosidase A (GLA) gene, wherein said coding sequence is the nucleic acid of any one of claims 1-3, and

- one or more albumin sequences,

preferably said plasmid comprising or consisting of a sequence with SEQ ID N.15.

12. A host cell transformed with the vector according to anyone of claims 6-9 or with the plasmid according to anyone of claims 10-11.

13. A pharmaceutical composition comprising the nucleic acid of any one of claims 1-3 or the nucleic acid construct of claim 5 or the vector according to anyone of claims 6-9 or the plasmid according to anyone of claims 10-11 or the host cell according to claim 12 and at least one pharmaceutically acceptable vehicle and/or excipient.

14. The nucleic acid of any one of claims 1-3 or the nucleic acid construct of claim 5 or the vector according to anyone of claims 6-9 or the plasmid according to anyone of claims 10-11 or the host cell according to claim 12 or the pharmaceutical composition of claim 13 for use as a medicament.

15. The nucleic acid of any one of claims 1-3 or the nucleic acid construct of claim 5 or the vector according to anyone of claims 6-9 or the plasmid according to anyone of claims 10-11 or the host cell according to claim 12 or the pharmaceutical composition of claim 13 for use in gene therapy.

16. The nucleic acid of any one of claims 1-3 or the nucleic acid construct of claim 5 or the vector according to anyone of claims 6-9 or the plasmid according to anyone of claims 10-11 or the host cell according to claim 12 or the pharmaceutical composition of claim 13 for use in the treatment of Fabry disease.
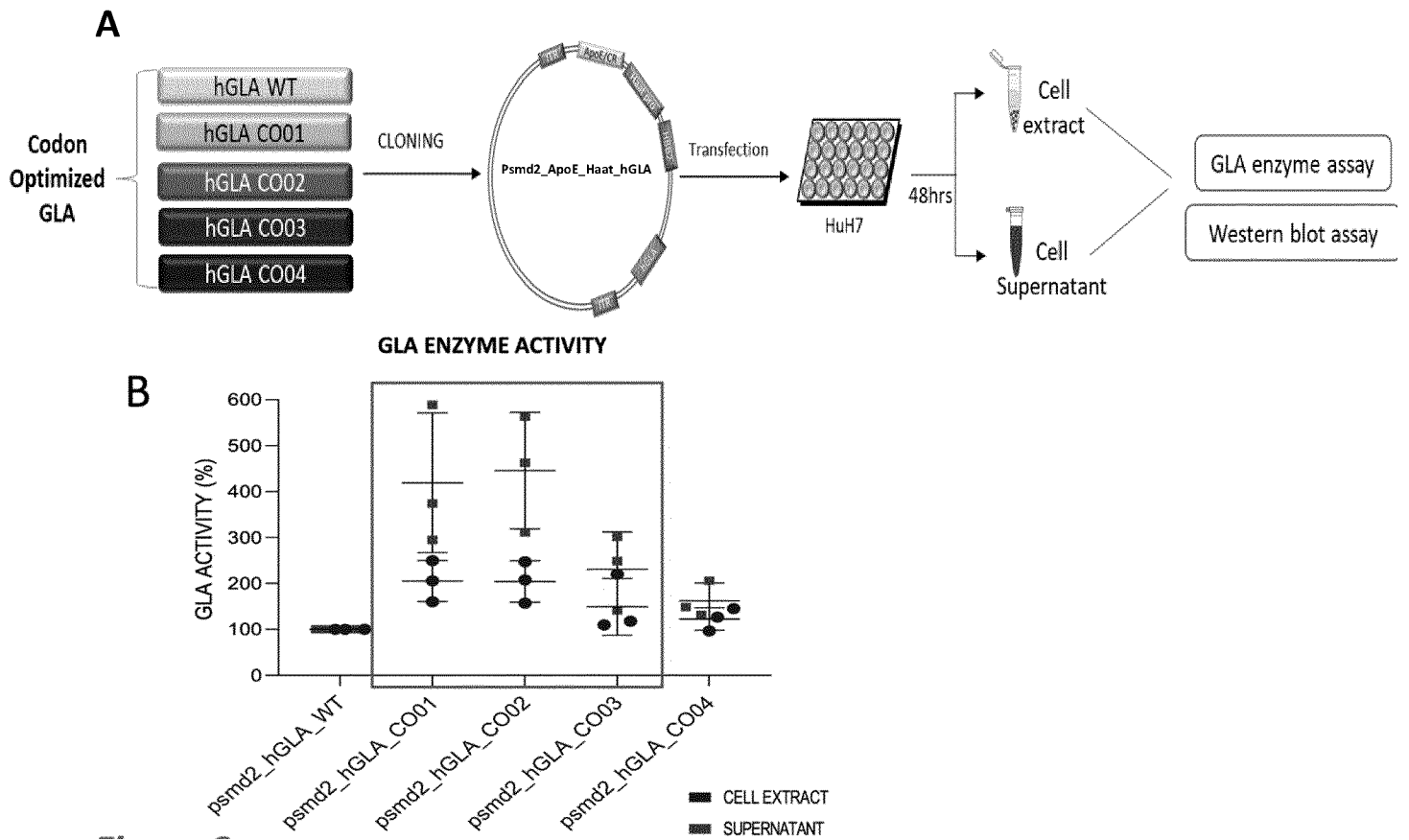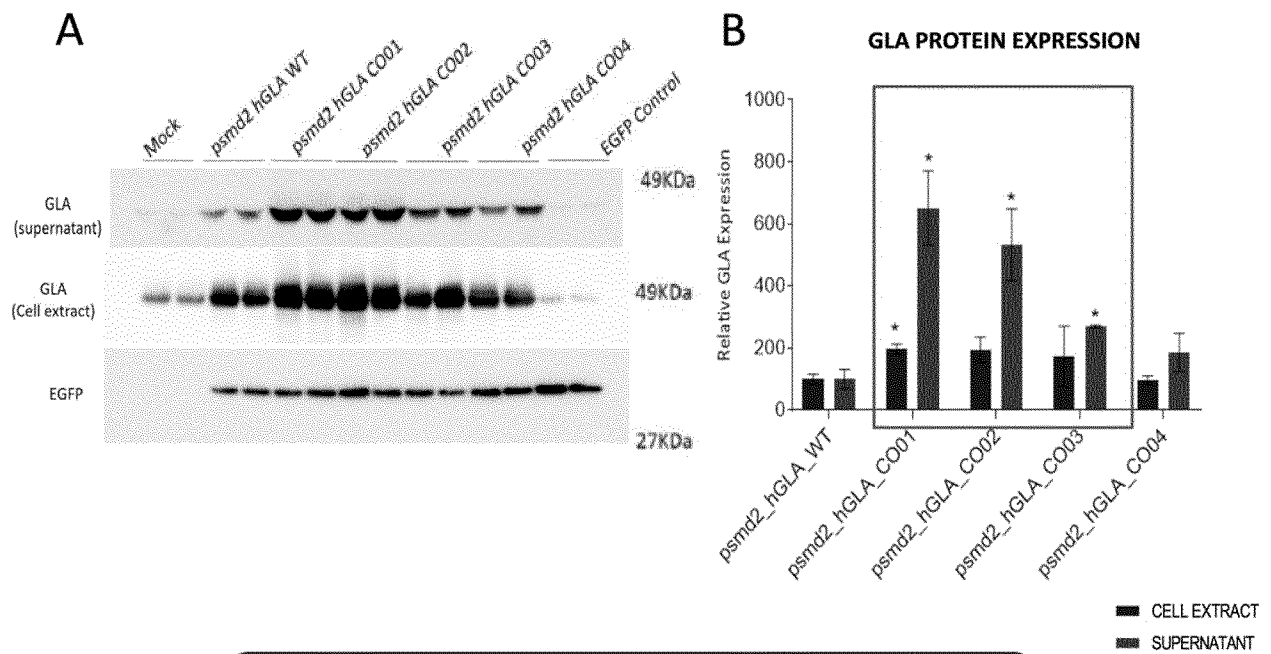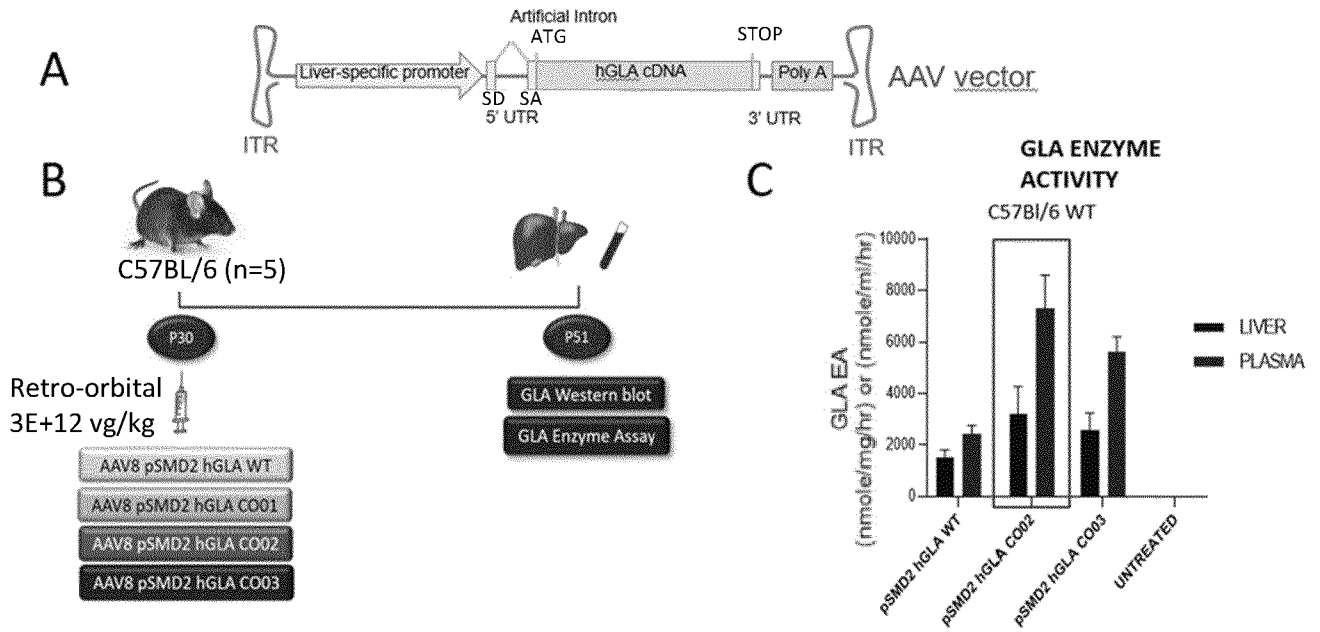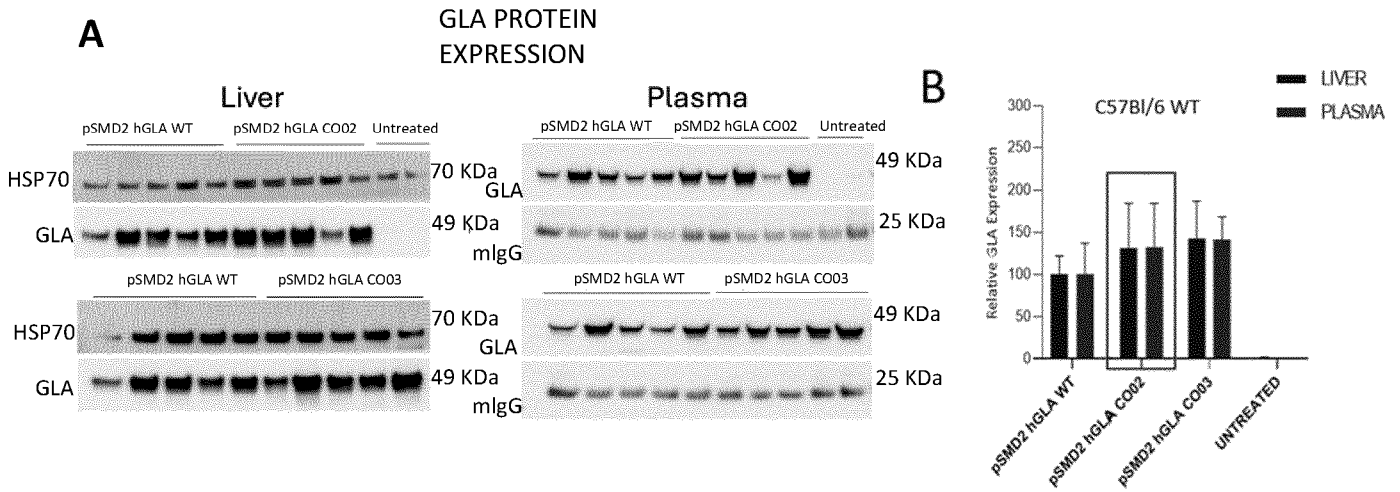
A

GLA ENZYME ACTIVITY

B

Figure 1
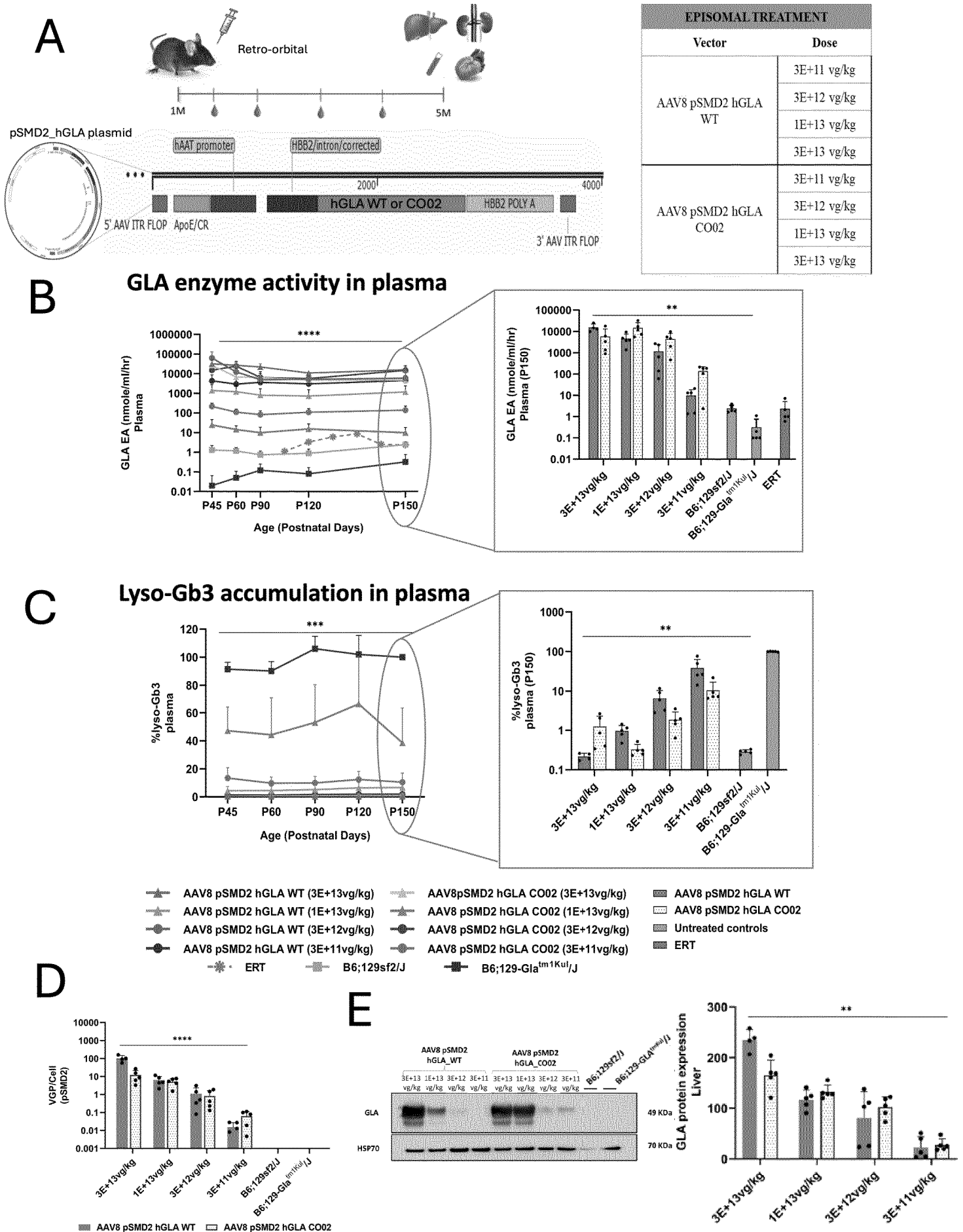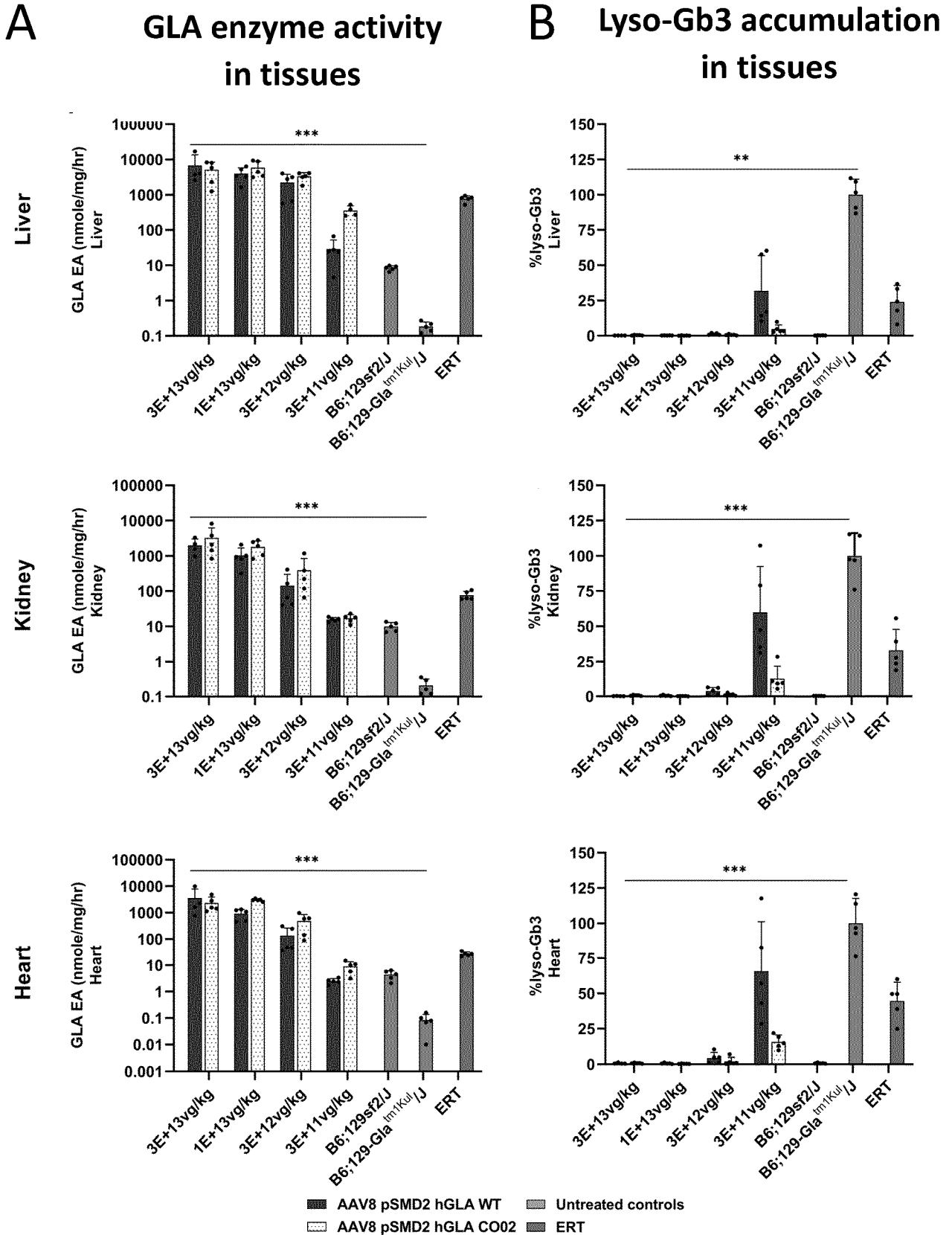
A

B
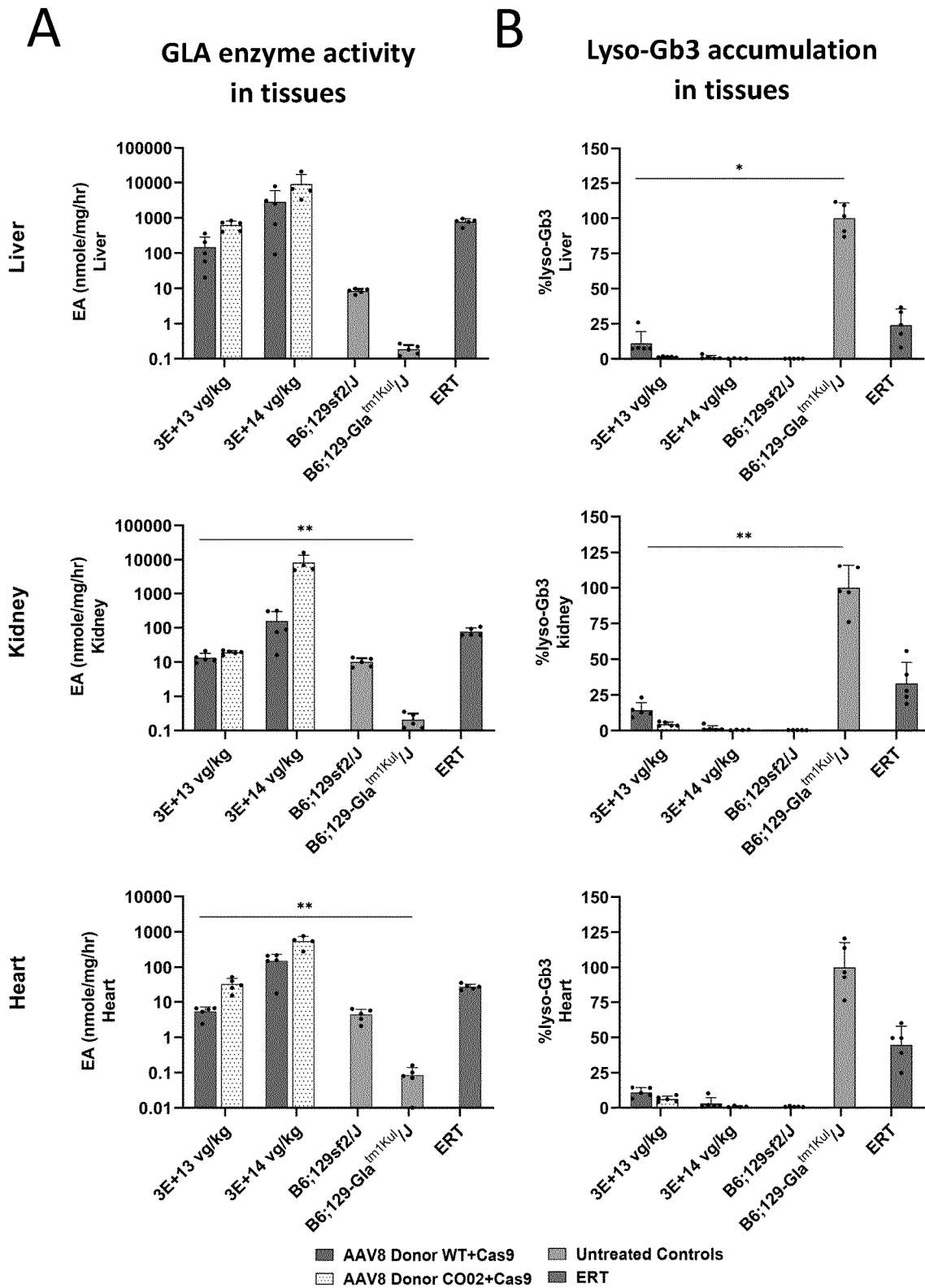
GLA PROTEIN EXPRESSION

Figure 2

Figure 3

Figure 4

Figure 5

4/6
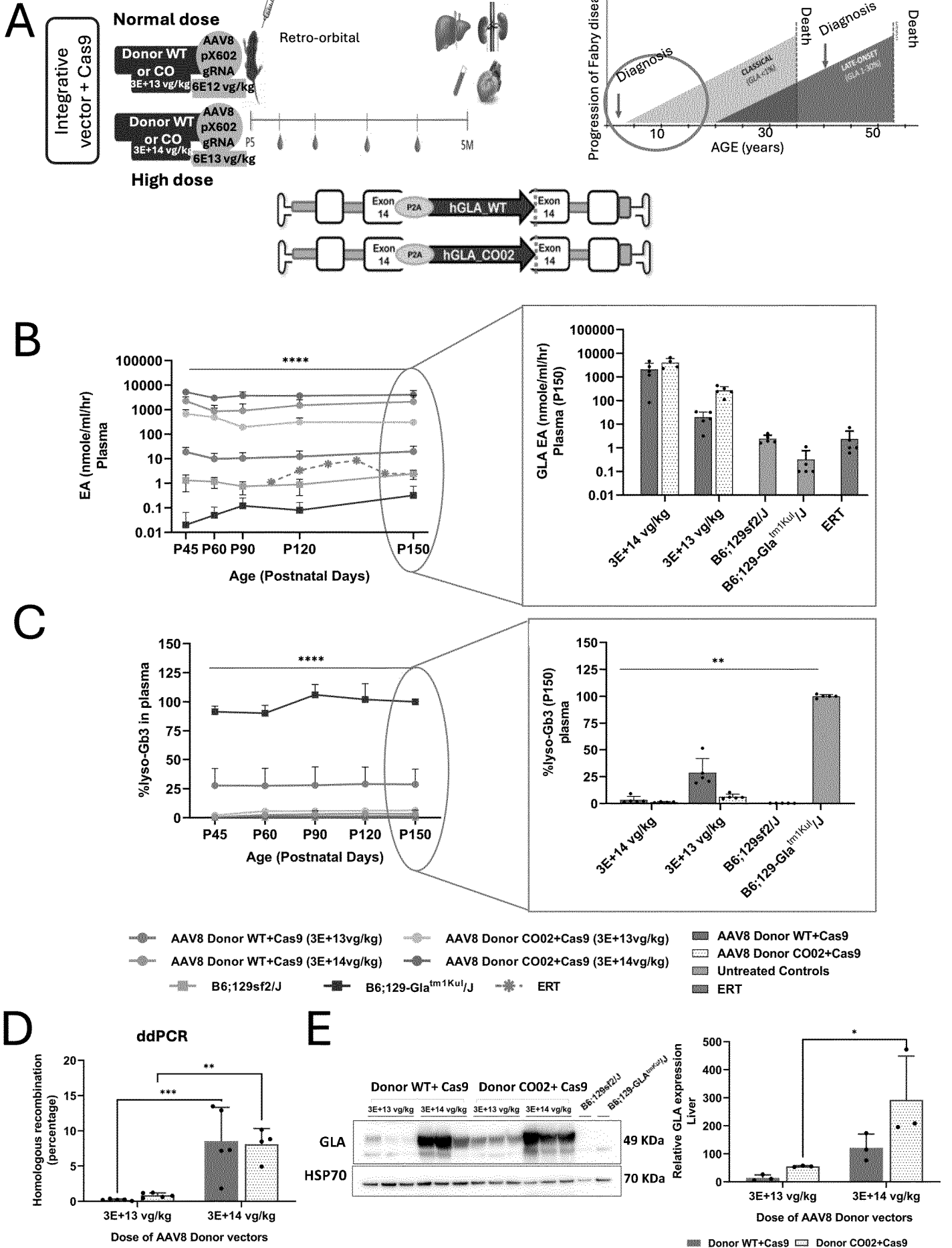
Figure 6

Figure 7

Figure 8

INTERNATIONAL SEARCH REPORT

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

A. CLASSIFICATION OF SUBJECT MATTER

INV. A61K38/47   C12N9/24   C12N15/86
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   C07K   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, Sequence Search

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 2023/028566 A2 (CANBRIDGE PHARMACEUTICALS INC [US]; LOGICBIO THERAPEUTICS INC [US]) 2 March 2023 (2023-03-02) the whole document in particular, figures 2,7; claims 1-88;page 43, line 21 to page 46, line 23 -----	1-16
Y	WO 2020/117898 A1 (ABEONA THERAPEUTICS INC [US]) 11 June 2020 (2020-06-11) claims 20,28; sequence 130 paragraph [0060] – paragraph [0061] paragraph [0067] paragraph [0076] – paragraph [0077] paragraph [0092] – paragraph [0097] -/--	1-16

[x] Further documents are listed in the continuation of Box C.   [x] See patent family annex.

* Special categories of cited documents :

"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	Date of mailing of the international search report
14 May 2024	27/05/2024

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Strobel, Andreas

2

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	-& DATABASE Geneseq [Online] 6 August 2020 (2020-08-06), "Codon optimized human alpha-galactosidase A (GLA) DNA CO2, SEQ 130.", XP093067091, retrieved from EBI accession no. GSN:BHW19125 Database accession no. BHW19125 the whole document -----	
Y	WO 2022/212720 A1 (SIGILON THERAPEUTICS INC [US]) 6 October 2022 (2022-10-06) claims 1-21; example 1; sequence 30 -----	1-16

2

Form PCT/ISA/210 (continuation of second sheet) (April 2005)

Patent document cited in search report		Publication date	Patent family member(s)			Publication date
WO 2023028566	A2	02-03-2023	AR	126870	A1	22-11-2023
			AU	2022332298	A1	14-03-2024
			CA	3229998	A1	02-03-2023
			TW	202330929	A	01-08-2023
			WO	2023028566	A2	02-03-2023
WO 2020117898	A1	11-06-2020	AU	2019391042	A1	10-06-2021
			BR	112021009913	A2	17-08-2021
			CA	3121177	A1	11-06-2020
			CN	113423434	A	21-09-2021
			EP	3890786	A1	13-10-2021
			IL	283546	A	29-07-2021
			JP	2022515338	A	18-02-2022
			KR	20220022107	A	24-02-2022
			US	2022090129	A1	24-03-2022
			WO	2020117898	A1	11-06-2020
WO 2022212720	A1	06-10-2022	CA	3214964	A1	06-10-2022
			CN	117241823	A	15-12-2023
			EP	4313119	A1	07-02-2024
			JP	2024513842	A	27-03-2024
			WO	2022212720	A1	06-10-2022