

VARIANT IGF2 CONSTRUCTS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of priority to U.S. Provisional Application 5 62/913,677, filed on October 10, 2019, and U.S. Provisional Application 62/929,054, filed on October 31, 2019, each of which is incorporated by reference herein in its entirety.

BACKGROUND

[0002] Genetic disorders arise via heritable or de novo mutations occurring in gene coding regions of the genome. In some cases, such genetic disorders are treated by administration of a protein that replaces a protein encoded by the gene mutated in the individual having the genetic disorder or by administration of a gene therapy vector encoding such a protein. Such treatment has challenges however, as the administered protein or the protein encoded by the gene therapy vector does not always result in the protein reaching the organs, cells, or organelle where it is needed. Proteins having improved intracellular targeting (e.g., to lysosomes), and gene therapy vectors encoding them, are desired. 15

SUMMARY

[0003] In certain aspects, there are provided nucleic acid constructs comprising: (a) a nucleic acid sequence encoding a therapeutic protein, and (b) a nucleic acid sequence encoding a variant IGF2 (vIGF2) peptide. In some embodiments, the vIGF2 peptide has an amino acid sequence that is at least 90, 95, 96, 97, 98 or 99% identical to an IGF2 variant peptide of Table 3. In some embodiments, the vIGF2 peptide comprises an amino acid sequence that is at least 90, 95, 96, 97, 98 or 99% identical to an IGF2 variant peptide selected from the group consisting of SEQ ID NO:90-123 of Table 3. In some embodiments, the vIGF2 peptide further comprises a linker having a sequence that is at least 90, 95, 96, 97, 98 or 99% identical to a sequence selected from the group consisting of SEQ ID NOs: 181-188. In some embodiments, the vIGF2 peptide has decreased or no affinity for the insulin receptor and IGFR1 as compared to native IGF2 peptide. In some embodiments, the vIGF2 peptide has increased affinity for the CI-MPR as compared to native IGF2 peptide. In some embodiments, the vIGF2 peptide confers improved expression and/or secretion of a fusion protein, compared to a native IGF2 peptide. In some embodiments, the vIGF2 peptide is capable of facilitating uptake of the 30

therapeutic protein into a lysosome in a cell. In some embodiments, the therapeutic protein is capable of replacing a defective or deficient protein associated with a genetic disorder in a subject having the genetic disorder. In some embodiments, the genetic disorder is a lysosomal storage disorder. In some embodiments, the genetic disorder is selected from the group

5 consisting of aspartylglucosaminuria, CLN1, CLN2 C cystinosis, Fabry disease, Gaucher disease type I, Gaucher disease type II, Gaucher disease type III, Pompe disease, Tay Sachs disease, Sandhoff disease, metachromatic leukodystrophy, mucopolipidosis type I, mucopolipidosis type II, mucopolipidosis type III, mucopolipidosis type IV, Hurler disease, Hunter disease, Sanfilippo disease type A, Sanfilippo disease type B, Sanfilippo disease type C, Sanfilippo

10 disease type D, Morquio disease type A, Morquio disease type B, Maroteau-Lamy disease, Sly disease, Niemann-Pick disease type A, Niemann-Pick disease type B, Niemann-Pick disease type C1, Niemann-Pick disease type C2, Schindler disease type I, Schindler disease type II, adenosine deaminase severe combined immunodeficiency (ADA-SCID), chronic granulomatous disease (CGD), and neuronal ceroid lipofuscinosis. In some embodiments, the

15 genetic disorder is Pompe disease. In some embodiments, the genetic disorder is neuronal ceroid lipofuscinosis. In some embodiments, the therapeutic protein comprises an enzyme selected from the group consisting of alpha-galactosidase (A or B), beta-galactosidase, beta-hexosaminidase (A or B), galactosylceramidase, arylsulfatase (A or B), beta-glucocerebrosidase, glucocerebrosidase, lysosomal acid lipase, lysosomal enzyme acid sphingomyelinase,

20 formylglycine-generating enzyme, iduronidase (e.g., alpha-L), acetyl-CoA:alpha-glucosaminide N-acetyltransferase, glycosaminoglycan alpha-L-iduronohydrolase, heparan N-sulfatase, N-acetyl-alpha-D-glucosaminidase (NAGLU), iduronate-2-sulfatase, galactosamine-6-sulfate sulfatase, N-acetylgalactosamine-6-sulfatase, N-sulfoglucosamine sulfohydrolase, glycosaminoglycan N-acetylgalactosamine 4-sulfatase beta-glucuronidase, hyaluronidase, alpha-

25 N-acetyl neuraminidase (sialidase), ganglioside sialidase, phosphotransferase, alpha-glucosidase, alpha-D-mannosidase, beta-D-mannosidase, aspartylglucosaminidase, alpha-L-fucosidase, battenin, palmitoyl protein thioesterases, and other Batten-related proteins (e.g., ceroid-lipofuscinosis neuronal protein 6), or an enzymatically active fragment thereof. In some embodiments, the therapeutic protein is alpha-glucosidase, or an enzymatically active fragment

30 thereof. In some embodiments, the therapeutic protein is palmitoyl protein thioesterase 1 (PPT1). In some embodiments, the therapeutic protein is tripeptidyl peptidase 1 (TPP1). In some embodiments, the therapeutic protein is aspartylglucosaminidase. In some embodiments,

the therapeutic protein is NAGLU (SEQ ID NO:54). In some embodiments, the therapeutic protein is the mature peptide of NAGLU, corresponding to amino acids 24-743 of SEQ ID NO:54 that remain after removal of the native signal peptide (SEQ ID NO:180). In some embodiments, the nucleic acid construct further comprises a translation initiation sequence. In some embodiments, the translation initiation sequence comprises a Kozak sequence. In some embodiments, the vIGF2 encoding nucleic acid sequence is 5' to the nucleic acid sequence encoding a therapeutic protein. In some embodiments, the vIGF2 encoding nucleic acid sequence is 3' to the nucleic acid sequence encoding a therapeutic protein. In some embodiments, the nucleic acid construct further comprises a linker sequence encoding a linker peptide between the vIGF2 nucleotide sequence and the nucleic acid sequence encoding a therapeutic protein. In some embodiments, the linker peptide comprises SEQ ID NO: 181-188. In some embodiments, the nucleic acid construct is a virus vector. In some embodiments, the virus vector is an adenovirus vector, an adeno-associated virus (AAV) vector, a retrovirus vector, a lentivirus vector, a pox virus vector, a vaccinia virus vector, an adenovirus vector, or a herpes virus vector.

[0004] In additional aspects, there are provided pharmaceutical compositions comprising a therapeutically effective amount of any one of the nucleic acid constructs provided herein a pharmaceutically acceptable carrier or excipient. In some embodiments, the excipient comprises a non-ionic, low-osmolar compound, a buffer, a polymer, a salt, or a combination thereof.

[0005] In further aspects, there are provided methods for treating a genetic disorder comprising administering to a subject in need thereof any one of the nucleic acid constructs provided herein or any one of the pharmaceutical compositions provided herein. In some embodiments, genetic disorder is a lysosomal storage disorder. In some embodiments, the genetic disorder is selected from the group consisting of aspartylglucosaminuria, Batten disease, cystinosis, Fabry disease, Gaucher disease type I, Gaucher disease type II, Gaucher disease type III, Pompe disease, Tay Sachs disease, Sandhoff disease, metachromatic leukodystrophy, mucopolipidosis type I, mucopolipidosis type II, mucopolipidosis type III, mucopolipidosis type IV, Hurler disease, Hunter disease, Sanfilippo disease type A, Sanfilippo disease type B, Sanfilippo disease type C, Sanfilippo disease type D, Morquio disease type A, Morquio disease type B, Maroteau-Lamy disease, Sly disease, Niemann-Pick disease type A, Niemann-Pick disease type B, Niemann-Pick disease type C1, Niemann-Pick disease type C2, Schindler disease type I,

Schindler disease type II, adenosine deaminase severe combined immunodeficiency (ADA-SCID), chronic granulomatous disease (CGD), and neuronal ceroid lipofuscinosis (Batten disease). In some embodiments, the genetic disorder is Pompe disease. In some embodiments, the genetic disorder is neuronal ceroid lipofuscinosis. In some embodiments, the genetic disorder is Aspartylglucosaminuria. In some embodiments, the administering is performed intrathecally, intraocularly, intravitreally, retinally, intravenously, intramuscularly, intraventricularly, intracerebrally, intracerebellarly, intracerebroventricularly, intraparenchymally, subcutaneously, or a combination thereof. In some embodiments, the administering is performed intrathecally.

- 10 **[0006]** In additional aspects, there are provided pharmaceutical compositions comprising any one of the gene therapy vectors provided herein and a pharmaceutically acceptable carrier or excipient for use in treating a genetic disorder. In further aspects, there are provided pharmaceutical composition comprising any one of the nucleic acid constructs provided herein and a pharmaceutically acceptable carrier or excipient for use in preparation of a medicament
- 15 for treatment of a genetic disorder. In some embodiments, the genetic disorder is a lysosomal storage disorder. In some embodiments, the genetic disorder is selected from the group consisting of aspartylglucosaminuria, Batten disease, cystinosis, Fabry disease, Gaucher disease type I, Gaucher disease type II, Gaucher disease type III, Pompe disease, Tay Sachs disease, Sandhoff disease, metachromatic leukodystrophy, mucopolipidosis type I, mucopolipidosis type II, mucopolipidosis type III, mucopolipidosis type IV, Hurler disease, Hunter disease, Sanfilippo disease type A, Sanfilippo disease type B, Sanfilippo disease type C, Sanfilippo disease type D, Morquio disease type A, Morquio disease type B, Maroteau-Lamy disease, Sly disease, Niemann-Pick disease type A, Niemann-Pick disease type B, Niemann-Pick disease type C1, Niemann-Pick disease type C2, Schindler disease type I, Schindler disease type II,
- 25 adenosine deaminase severe combined immunodeficiency (ADA-SCID), chronic granulomatous disease (CGD), and neuronal ceroid lipofuscinosis. In some embodiments, the genetic disorder is Pompe disease. In some embodiments, the genetic disorder is neuronal ceroid lipofuscinosis. In some embodiments, the genetic disorder is Aspartylglucosaminuria. In some embodiments, the composition is formulated for administration intrathecally,
- 30 intraocularly, intravitreally, retinally, intravenously, intramuscularly, intraventricularly, intracerebrally, intracerebellarly, or subcutaneously. In some embodiments, the composition is formulated for administration intrathecally.

[0007] In additional aspects there are provided nucleic acids encoding a fusion protein having an amino acid sequence at least 90, 95, 96, 97, 98 or 99 % identical to a sequence selected from the group consisting of SEQ ID NO:47-53. In some embodiments, the nucleic acid is at least 85, 90, 95, 96, 97, 98 or 99 % identical to a sequence selected from the group consisting of SEQ ID NO: 60-67.

[0008] In further aspects, there are provided pharmaceutical composition comprising any one of the above nucleic acids and a pharmaceutically acceptable carrier or excipient. In some embodiments, the excipient comprises a non-ionic, low-osmolar compound, a buffer, a polymer, a salt, or a combination thereof.

[0009] In further aspects pharmaceutical composition comprising the fusion protein having an amino acid sequence at least 90, 95, 96, 97, 98 or 99 % identical to a sequence selected from the group consisting of SEQ ID NO: 47-53 and SEQ ID NO: 60-67, and a pharmaceutically acceptable carrier or excipient. In some embodiments, the excipient comprises a non-ionic, low-osmolar compound, a buffer, a polymer, a salt, or a combination thereof.

[0010] In additional aspects, there are provided gene therapy vectors comprising a nucleic acid encoding an amino acid sequence at least 90, 95, 96, 97, 98 or 99 % identical to a sequence selected from the group consisting of SEQ ID NO: 47-53 and SEQ ID NO: 60-67; and a nucleic acid encoding an amino acid sequence at least 90, 95, 96, 97, 98 or 99 % identical to a sequence selected from the group consisting of SEQ ID NO:106, 109, 111, 119, 120 and 121.

In some embodiments, the gene therapy vector is a virus vector. In some embodiments, the virus vector is an adenovirus vector, an adeno-associated virus (AAV) vector, a retrovirus vector, a lentivirus vector, a pox virus vector, a vaccinia virus vector, an adenovirus vector, or a herpes virus vector and a pharmaceutically acceptable carrier or excipient. In some embodiments, the excipient comprises a non-ionic, low-osmolar compound, a buffer, a polymer, a salt, or a combination thereof.

[0011] In additional aspects, there are provided methods of treating CLN1/PPT1 disease or CLN2/TPP1 disease comprising administering to a subject in need thereof a therapeutically effective amount of any one of the nucleic acids herein, any one of the fusion proteins herein, any one of the gene therapy vectors herein, or any one of the pharmaceutical compositions herein. In some embodiments, the administering is performed intrathecally, intraocularly, intravitreally, retinally, intravenously, intramuscularly, intraventricularly, intracerebrally,

intracerebellarly, intracerebroventricularly, intraparenchymally, subcutaneously, or a combination thereof.

[0012] . In some embodiments, the nucleic acid has a nucleic acid sequence at least 85, 90, 95, 96, 97, 98 or 99 % identical to a sequence selected from the group consisting of SEQ ID

5 NO:189-250.

[0013] In additional aspects there are provided pharmaceutical compositions comprising any one of the nucleic acids herein a pharmaceutically acceptable carrier or excipient. In some embodiments, the excipient comprises a non-ionic, low-osmolar compound, a buffer, a polymer, a salt, or a combination thereof.

10 **[0014]** In some embodiments, there are provided a variant IGF2 (vIGF2) peptide that is at least 95, 96, 97, 98 or 99 % identical to a sequence selected from the group consisting of SEQ ID NO: 90-103.

[0015] In some embodiments, the variant IGF2 (vIGF2) peptide is at least 98% identical to at least one sequence selected from SEQ ID NOs:106, 109, 111, 119, 120, 121. In some

15 embodiments, the vIGF2 peptide is at least 95, 96, 97, 98 or 99 % identical to SEQ ID NO:120 or 121.

[0016] In some embodiments, there are provided a fusion protein comprising a variant vIGF2 peptide and a therapeutic protein having an amino acid sequence at least 95, 96, 97, 98 or 99

20 % identical to a sequence selected from the group consisting of SEQ ID NO:4, amino acid residues 21-306 of SEQ ID NO:4, amino acid residues 28-306 of SEQ ID NO:4, SEQ ID NO: 8, SEQ ID NO:46, and SEQ ID NO:54.

[0017] In some embodiments, the fusion protein has an amino acid sequence at least 95, 96, 97, 98 or 99 % identical to a sequence selected from the group consisting of SEQ ID NO:60-

25 67, SEQ ID NO:47-53 and SEQ ID NO:54-59. In some embodiments, the fusion protein further comprises a lysosomal cleavage peptide. In some embodiments, the lysosomal cleavage peptide has SEQ ID NO:188. In some embodiments the vIGF2 peptide is N terminal to the therapeutic protein. In some embodiments, the vIGF2 peptide is C terminal to the therapeutic protein.

30 **[0018]** In some embodiments, the fusion protein comprises a signal sequence. In some embodiments, the signal sequence has an amino acid sequence at least 95, 96, 97, 98 or 99 % identical to a sequence selected from the group consisting of SEQ ID NO:169-180.

[0019] In some embodiments, the therapeutic protein is PPT1 or an enzymatically active fragment thereof, TPP1 or an enzymatically active fragment thereof, or NAGLU or enzymatically active fragment thereof.

[0020] In some embodiments, the fusion protein is taken up by target cells more efficiently
5 than the corresponding protein lacking the vIGF2 peptide. In some embodiments, the fusion protein is taken up by cells in the brain. In some embodiments the fusion protein is taken up by neuronal cells. In some embodiments the fusion protein is taken up by glial cells.

[0021] Provided herein are also pharmaceutical composition comprising fusion proteins having a vIGF2 peptide and a therapeutic protein, along with a pharmaceutically acceptable carrier or
10 excipient. Methods of treating a lysosomal storage disorder, comprising administering such pharmaceutical compositions to a subject in need thereof are also provided herein. In some embodiments, the lysosomal storage disorder is selected from the group consisting of CLN1/PPT1 disease, CLN2/TPP1 disease, and Sanfilippo Type B disease. In some embodiments, the fusion protein or pharmaceutical composition comprising the fusion protein
15 is administered intrathecally, intraocularly, intravitreally, retinally, intravenously, intramuscularly, intraventricularly, intracerebrally, intracerebellarly, intracerebroventricularly, intraperenchymally, subcutaneously, or a combination thereof.

[0022] In some embodiments, administering the pharmaceutical composition prevents/reduces or reverses accumulation of autofluorescent storage material (ASM) in the brain. In some
20 embodiments, administering the pharmaceutical composition prevents/reduces or reverses elevation of glial fibrillary acidic protein (GFAP) in the brain. In some embodiments, administering the pharmaceutical composition prevents/reduces or reverses accumulation of autofluorescent storage material (ASM) in the cortex or thalamus. In some embodiments, administering the pharmaceutical composition prevents/reduces or reverses elevation of glial
25 fibrillary acidic protein (GFAP) brain cortex or thalamus.

[0023] Further provided herein are nucleic acids encoding a fusion protein comprising vIGF2 and a therapeutic protein, wherein the nucleic acid is at least 85, 90, 95, 96, 97, 98 or 99% identical to a sequence selected from the group consisting of SEQ ID NO:189-250.

[0024] In additional aspects, there are provided pharmaceutical compositions comprising any
30 one of the fusion proteins herein and a pharmaceutically acceptable carrier or excipient. In some embodiments, the excipient comprises a non-ionic, low-osmolar compound, a buffer, a polymer, a salt, or a combination thereof.

[0025] In further aspects, there are provided gene therapy vectors comprising a nucleic acid encoding an amino acid sequence at least 90% identical to SEQ ID NO: 51. In some embodiments, the gene therapy vector is a virus vector. In some embodiments, the virus vector is an adenovirus vector, an adeno-associated virus (AAV) vector, a retrovirus vector, a
5 lentivirus vector, a pox virus vector, a vaccinia virus vector, an adenovirus vector, or a herpes virus vector.

[0026] In additional aspects, there are provided pharmaceutical compositions comprising any one of the gene therapy vectors provided herein and a pharmaceutically acceptable carrier or excipient. In some embodiments, the excipient comprises a non-ionic, low-osmolar compound,
10 a buffer, a polymer, a salt, or a combination thereof.

[0027] In another aspect, there are provided nucleic acid constructs comprising: (a) a nucleic acid sequence encoding a therapeutic protein, and (b) a nucleic acid sequence encoding a variant IGF2 (vIGF2) peptide that is at least 95, 96, 97, 98 or 99% identical to at least one sequence selected from SEQ ID NO: 90-103. In some aspects, the vIGF2 peptide has an amino
15 acid sequence that is at least 95, 96, 97, 98 or 99% identical to an IGF2 variant peptide selected from SEQ ID NOs:106, 109, 111, 119, 120, 121. In some embodiments, the vIGF2 peptide comprises an amino acid sequence that is at least 95, 96, 97, 98 or 99% identical to an IGF2 variant peptide selected from the group consisting of SEQ ID NO:120 and SEQ ID NO:121.

[0028] In some aspects, the nucleic acid further comprises a sequence encoding a linker having
20 a sequence that is at least 95, 96, 97, 98 or 99% identical to a sequence selected from the group consisting of SEQ ID NOs: 181-188. In some embodiments, the vIGF2 peptide is capable of increasing expression and/or secretion of a therapeutic protein compared to a vIGF2 peptide having the amino acid sequence of SEQ ID NO:80. . In some embodiments, the vIGF2 peptide has increased affinity for the CI-MPR as compared to a vIGF2 peptide having the
25 amino acid sequence of SEQ ID NO:80. In some embodiments, the vIGF2 peptide is capable of improving uptake of the therapeutic protein into a target cell, such as a human brain cell. In some embodiments, the human brain cell is a neuronal cell or a glial cell.

[0029] In certain aspects, the therapeutic protein is capable of replacing a defective or deficient protein associated with a genetic disorder in a subject having the genetic disorder. In some
30 embodiments, the genetic disorder is a lysosomal storage disorder. In some embodiments, the genetic disorder is selected from the group consisting of aspartylglucosaminuria, neuronal ceroid lipofuscinosis, CLN1/PPT1 disease, CLN2/PPT1 disease, cystinosis, Fabry disease,

Gaucher disease type I, Gaucher disease type II, Gaucher disease type III, Pompe disease, Tay Sachs disease, Sandhoff disease, metachromatic leukodystrophy, mucopolipidosis type I, mucopolipidosis type II, mucopolipidosis type III, mucopolipidosis type IV, Hurler disease, Hunter disease, Sanfilippo disease type A, Sanfilippo disease type B, Sanfilippo disease type C,

5 Sanfilippo disease type D, Morquio disease type A, Morquio disease type B, Maroteau-Lamy disease, Sly disease, Niemann-Pick disease type A, Niemann-Pick disease type B, Niemann-Pick disease type C1, Niemann-Pick disease type C2, Schindler disease type I, Schindler disease type II, adenosine deaminase severe combined immunodeficiency (ADA-SCID), and neuronal ceroid lipofuscinosis. In some embodiments, the genetic disorder is selected from the

10 group consisting of CLN1/PPT1 disease, CLN2/PPT1 disease, Pompe disease and MPS IIIB disease. In some aspects, the genetic disorder is CLN1/PPT1 disease or CLN2/PPT1 disease.

[0030] In some aspects, the therapeutic protein comprises a human enzyme selected from the group consisting of alpha-galactosidase (A or B), beta-galactosidase, beta-hexosaminidase (A or B), galactosylceramidase, arylsulfatase (A or B), beta-glucocerebrosidase, glucocerebrosidase,

15 lysosomal acid lipase, lysosomal enzyme acid sphingomyelinase, formylglycine-generating enzyme, iduronidase (e.g., alpha-L), acetyl-CoA:alpha-glucosaminide N-acetyltransferase, glycosaminoglycan alpha-L-iduronohydrolase, heparan N-sulfatase, N-acetyl-alpha-D-glucosaminidase (NAGLU), iduronate-2-sulfatase, galactosamine-6-sulfate sulfatase, N-acetylgalactosamine-6-sulfatase, N-sulfoglucosamine sulfohydrolase, glycosaminoglycan N-

20 acetylgalactosamine 4-sulfatase, beta-glucuronidase, hyaluronidase, alpha-N-acetylneuraminidase (sialidase), ganglioside sialidase, phosphotransferase, alpha-glucosidase, alpha-D-mannosidase, beta-D-mannosidase, aspartylglucosaminidase, alpha-L-fucosidase, battenin, PPT1, TPP1, and other Batten-related proteins (e.g., ceroid-lipofuscinosis neuronal protein 6), or an enzymatically active fragment thereof. In some embodiments, the therapeutic protein is a

25 human lysosomal enzyme or an enzymatically active fragment thereof. In some embodiments, the human lysosomal enzyme is alpha-glucosidase, PPT1, TPP1, or NAGLU.

[0031] In some aspects, the nucleic acid construct further comprises a sequence encoding a signal peptide. In some embodiments, the signal peptide is a sequence selected from the group consisting of SEQ ID NO:169-180. In some embodiments, the vIGF2 encoding nucleic acid

30 sequence is 5' to the nucleic acid sequence encoding a therapeutic protein. In other embodiments, the vIGF2 encoding nucleic acid sequence is 3' to the nucleic acid sequence encoding a therapeutic protein.

[0032] Further provided herein are gene therapy vectors comprising the nucleic acids described herein. In some embodiments, the gene therapy vector is a virus vector. In some embodiments, the virus vector is an adenovirus vector, an adeno-associated virus (AAV) vector, a retrovirus vector, a lentivirus vector, a pox virus vector, a vaccinia virus vector, an adenovirus vector, or a herpes virus vector.

[0033] In some aspects, the nucleic acid constructs herein are in a plasmid or bacterial artificial chromosome. In some embodiments, the nucleic acids constructs described herein are in a host cell.

[0034] There are further provided pharmaceutical compositions, comprising a therapeutically effective amount of the nucleic acid constructs described herein, or gene therapy vectors comprising the nucleic acid constructs described herein, along with a pharmaceutically acceptable carrier or excipient. In some embodiments, the excipient comprises a non-ionic, low-osmolar compound, a buffer, a polymer, a salt, or a combination thereof.

[0035] Further provided herein are methods for treating a genetic disorder comprising administering to a subject in need thereof the nucleic acid constructs, gene therapy vectors and/or pharmaceutical composition described herein. In some embodiments, the genetic disorder is a lysosomal storage disorder. In some embodiments, the genetic disorder is selected from the group consisting of aspartylglucosaminuria, neuronal ceroid lipofuscinosis, CLN1/PPT1 disease, CLN2/PPT1 disease, cystinosis, Fabry disease, Gaucher disease type I, Gaucher disease type II, Gaucher disease type III, Pompe disease, Tay Sachs disease, Sandhoff disease, metachromatic leukodystrophy, mucopolipidosis type I, mucopolipidosis type II, mucopolipidosis type III, mucopolipidosis type IV, Hurler disease, Hunter disease, Sanfilippo disease type A, Sanfilippo disease type B, Sanfilippo disease type C, Sanfilippo disease type D, Morquio disease type A, Morquio disease type B, Maroteau-Lamy disease, Sly disease, Niemann-Pick disease type A, Niemann-Pick disease type B, Niemann-Pick disease type C1, Niemann-Pick disease type C2, Schindler disease type I, Schindler disease type II, adenosine deaminase severe combined immunodeficiency (ADA-SCID), and chronic granulomatous disease (CGD). In some embodiments, the genetic disorder is Batten's disease, such as CLN1/PPT1 disease or CLN2/TPP1 disease. In some embodiments, the genetic disorder is Pompe disease or Sanfilippo disease type B.

[0036] In some embodiments, the administering is performed intrathecally, intraocularly, intravitreally, retinally, intravenously, intramuscularly, intraventricularly, intracerebrally,

intracerebellarly, intracerebroventricularly, intraparenchymally, subcutaneously, or a combination thereof.

[0037] In some aspects, administering the nucleic acid, gene therapy vector, fusion protein, or pharmaceutical composition prevents/reduces or reverses accumulation of autofluorescent storage material (ASM) in the brain. In some embodiments, administering the nucleic acid, gene therapy vector fusion protein, or pharmaceutical composition prevents/reduces or reverses elevation of glial fibrillary acidic protein (GFAP) in the brain. In some embodiments, administering the nucleic acid, gene therapy vector, fusion protein, or pharmaceutical composition prevents/reduces or reverses accumulation of autofluorescent storage material (ASM) in the cortex or thalamus. In some aspects, administering the nucleic acid, gene therapy vector, fusion protein, or pharmaceutical composition prevents/reduces or reverses elevation of glial fibrillary acidic protein (GFAP) brain cortex or thalamus.

[0038] In some aspects, the nucleic acid encodes a fusion protein having a sequence at least 95, 96, 97, 98 or 99% identical to a sequence selected from the group consisting of SEQ ID NO:60-67. In some embodiments, the nucleic acid encodes a fusion protein having a sequence at least 98% identical to a sequence selected from the group consisting of SEQ ID NO:47-53.

[0039] In some aspects, the nucleic acid encodes a fusion protein comprising: (a) an amino acid sequence at least 95, 96, 97, 98 or 99% identical to a sequence selected from the group consisting of SEQ ID NO:106, 109, 111, 119, 120 and 121; and (b) an amino acid sequence at least 95, 96, 97, 98 or 99% identical to a sequence selected from the group consisting of SEQ ID NO:4, residues 21-306 of SEQ ID NO:4, residues 28-306 of SEQ ID NO:4, SEQ ID NO: 8, and SEQ ID NO:46. In some embodiments, the nucleic acid encodes a vIGF2 at least 95, 96, 97, 98, or 99 % identical to SEQ ID NO:120 and 121. In some embodiments, the nucleic acid encodes a fusion protein comprising: (a) at least one of SEQ ID NO:106, 109, 111, 119, 120 or 121; and (b) at least one of SEQ ID NO:4, residues 21-306 of SEQ ID NO:4, residues 28-306 of SEQ ID NO:4, SEQ ID NO: 8, and SEQ ID NO:46.residues 28-306 of SEQ ID NO:4, SEQ ID NO: 8, and SEQ ID NO:46.

[0040] In some embodiments, the nucleic acid further encodes a lysosomal cleavage peptide.

[0041] In some aspects, the fusion protein has a sequence at least 95, 96, 97, 98, or 99 % identical to at least one of SEQ ID NO:60-67 and SEQ ID NO:47-53. In some embodiments, the fusion protein comprises at least one of SEQ ID NO:60-67 and SEQ ID NO:47-53. In

some embodiments the fusion protein consists or consists essentially of SEQ ID NO:60-67 and SEQ ID NO:47-53.

[0042] In additional aspects, there are provided methods of treating a lysosomal storage disease comprising administering to a subject in need thereof a therapeutically effective

5 amount of any one of the nucleic acids herein, any one of the fusion proteins herein, any one of the gene therapy vectors herein, or any one of the pharmaceutical compositions herein. In some embodiments, the administering is performed intrathecally, intraocularly, intravitreally, retinally, intravenously, intramuscularly, intraventricularly, intracerebrally, intracerebellarly, intracerebroventricularly, intraparenchymally, subcutaneously, or a combination thereof.

10 **[0043]** In further aspects, there are provided methods of treating Batten disease, including CLN1/PPT1 disease and CLN2/TPP1 disease comprising administering to a subject in need thereof a therapeutically effective amount of any one of the nucleic acids herein, any one of the fusion proteins herein, any one of the gene therapy vectors herein, or any one of the pharmaceutical compositions herein. In some embodiments, the administering is performed

15 intrathecally, intraocularly, intravitreally, retinally, intravenously, intramuscularly, intraventricularly, intracerebrally, intracerebellarly, intracerebroventricularly, intraparenchymally, subcutaneously, or a combination thereof.

[0044] In additional aspects, there are provided pharmaceutical compositions comprising any one of the nucleic acids provided herein and a pharmaceutically acceptable carrier or excipient.

20 In some embodiments, the excipient comprises a non-ionic, low-osmolar compound, a buffer, a polymer, a salt, or a combination thereof.

[0045] In additional aspects, there are provided pharmaceutical compositions comprising any one of the gene therapy vectors provided herein and a pharmaceutically acceptable carrier or excipient.

25 **[0046]** In additional aspects, there are provided fusion proteins comprising: (a) a lysosomal enzyme, and (b) a variant IGF2 (vIGF2) peptide, wherein the vIGF2 peptide comprises an amino acid sequence that is at least 95, 96, 97, 98, or 99 % identical to an IGF2 variant peptide of Table 3. In some embodiments, the vIGF2 peptide comprises an amino acid sequence that is at least 95, 96, 97, 98, or 99 % identical to an IGF2 variant peptide selected from the group

30 consisting of SEQ ID NO: 69-131. . In some embodiments, the vIGF2 peptide comprises an amino acid sequence that is at least 95, 96, 97, 98, or 99 % identical to an IGF2 variant peptide selected from the group consisting of SEQ ID NO: 90-123. In some embodiments, the vIGF2

has been modified to replace residues 31-38 of wildtype IGF2 with four glycine residues (Δ 31-38GGGG). In some embodiments, the vIGF2 has been further modified by a V43L mutation. In some embodiments, the vIGF2 has been further modified to replace the serine in position 50 with an acidic residue (aspartic or glutamic acid). In some aspects, the vIGF2 has the sequence of SEQ ID NO:120 or 121.

[0047] In some embodiments, the vIGF2 peptide further comprises a linker having a sequence that is at least 95, 96, 97, 98, or 99 % identical to a sequence selected from the group consisting of SEQ ID NOs: 181-188. In some embodiments, the linker is cleavable. In some embodiments, the vIGF2 peptide has decreased or no affinity for the insulin receptor and IGFR1 as compared to native IGF2 peptide. In some embodiments, the vIGF2 peptide has increased affinity for the CI-MPR as compared to native IGF2 peptide. In some embodiments, the vIGF2 peptide is capable of facilitating uptake of the lysosomal enzyme into a lysosome in a cell. In some embodiments, the lysosomal enzyme is capable of replacing a defective or deficient protein associated with a lysosomal storage disorder. In some embodiments, the lysosomal storage disorder is selected from the group consisting of aspartylglucosaminuria, Batten disease, cystinosis, Fabry disease, Gaucher disease type I, Gaucher disease type II, Gaucher disease type III, Pompe disease, Tay Sachs disease, Sandhoff disease, metachromatic leukodystrophy, mucopolipidosis type I, mucopolipidosis type II, mucopolipidosis type III, mucopolipidosis type IV, Hurler disease, Hunter disease, Sanfilippo disease type A, Sanfilippo disease type B, Sanfilippo disease type C, Sanfilippo disease type D, Morquio disease type A, Morquio disease type B, Maroteau-Lamy disease, Sly disease, Niemann-Pick disease type A, Niemann-Pick disease type B, Niemann-Pick disease type C1, Niemann-Pick disease type C2, Schindler disease type I, Schindler disease type II, adenosine deaminase severe combined immunodeficiency (ADA-SCID), chronic granulomatous disease (CGD), and neuronal ceroid lipofuscinosis. In some embodiments, the lysosomal storage disorder is Pompe disease. In some embodiments, the lysosomal storage disorder is neuronal ceroid lipofuscinosis. In some embodiments, the lysosomal enzyme comprises an enzyme selected from the group consisting of alpha-galactosidase (A or B), β -galactosidase, β -hexosaminidase (A or B), galactosylceramidase, arylsulfatase (A or B), β -glucocerebrosidase, glucocerebrosidase, lysosomal acid lipase, lysosomal enzyme acid sphingomyelinase, formylglycine-generating enzyme, iduronidase (e.g., alpha-L), acetyl-CoA:alpha-glucosaminide N-acetyltransferase, glycosaminoglycan alpha-L-iduronohydrolase, heparan N-sulfatase, N-acetyl- α -D-

glucosaminidase (NAGLU), iduronate-2-sulfatase, galactosamine-6-sulfate sulfatase, N-sulfolglucosamine sulfohydrolase, N-acetylgalactosamine-6-sulfatase, glycosaminoglycan N-acetylgalactosamine 4-sulfatase, β -glucuronidase, hyaluronidase, alpha-N -acetyl neuraminidase (sialidase), ganglioside sialidase, phosphotransferase, alpha-glucosidase, alpha-D-mannosidase, beta-D-mannosidase, aspartylglucosaminidase, alpha-L-fucosidase, battenin, palmitoyl protein thioesterases, and other Batten-related proteins (e.g., ceroid-lipofuscinosis neuronal protein 6), or an enzymatically active fragment thereof. In some embodiments, the lysosomal enzyme is alpha-glucosidase, or an enzymatically active fragment thereof. In some embodiments, the lysosomal enzyme is a palmitoyl protein thioesterase. In some embodiments, the lysosomal enzyme is tripeptidyl peptidase 1. In some embodiments, the lysosomal enzyme is aspartylglucosaminidase.

[0048] Additionally, provided herein are pharmaceutical compositions comprising a therapeutically effective amount of any one of the fusion proteins provided herein and a pharmaceutically acceptable carrier or excipient.

INCORPORATION BY REFERENCE

[0049] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[0050] The patent application file contains at least one drawing executed in color. Copies of this patent application with color drawing(s) will be provided by the Office upon request and payment of the necessary fee. An understanding of the features and advantages of the present disclosure will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the disclosure are utilized, and the accompanying drawings of which:

[0051] FIG. 1 shows affinity chromatography using immobilized CI-MPR was used to determine the proportion of GAA that is able to interact with the CI-MPR through phosphorylated oligosaccharides. The first peak is the material that flows through column indicating that it does not have phosphorylated glycans. The later peak is the material able to bind the immobilized CI-MPR. It is eluted with an increasing gradient of M6P. M6P reveals that GAA contains both M6P-containing and -lacking fractions. Since binding the CI-MPR is

the mandatory first step for receptor-mediated endocytosis, only the rhGAA fraction that binds the CI-MPR is capable of efficient cellular uptake.

[0052] FIG. 2 shows structure of the CI-MPR including the different binding domains for the IGF2 and for mono- and bis-phosphorylated oligosaccharides.

5 **[0053] FIG. 3** shows the sequence and structure of the mature, human IGF2 peptide. Site specific amino acid substitutions are proposed to influence binding to other receptors and serum proteins.

[0054] FIG. 4 shows binding of the wild-type IGF2 (wtIGF2) peptide to CI-MPR as measured by surface plasmon resonance

10 **[0055] FIG. 5** shows binding of the variant IGF2 (vIGF2) peptide binding to CI-MPR as measured by surface plasmon resonance.

[0056] FIG. 6 shows benefit of adding vIGF2 to alglucosidase alfa to increase the binding to the IGF2/CI-MPR.

15 **[0057] FIG. 7** shows the benefit of adding a vIGF2 to recombinant human N-acetyl- α -D-glucosaminidase (rhNAGLU) to increase the binding to the IGF2/CI-MPR.

[0058] FIG. 8 shows binding of wildtype human IGF2 to insulin receptor.

[0059] FIG. 9 shows no detectable binding of vIGF2 to insulin receptor.

[0060] FIG. 10 shows binding of wildtype IGF2 to insulin-like growth factor 1 receptor.

20 **[0061] FIG. 11** shows decreased binding of vIGF2 peptide to insulin-like growth factor 1 receptor, as compared to wildtype IGF2.

[0062] FIG. 12 shows two examples of gene therapy expression cassettes encoding Natural hGAA and Engineered hGAA. Natural hGAA is poorly phosphorylated, and unable to efficiently bind the CI-MPR. Engineered hGAA has element added for improved CIMPR binding (vIGF2), a 2GS linker is incorporate to allow for greater interaction of vIGF2-GAA protein with CI-MPR, and a BiP signal peptide to improve secretion.

25 **[0063] FIG. 13** shows a Western blot of palmitoyl-protein thioesterase 1 (PPT1) from cells expressing recombinant human PPT1 (PPT1-1), recombinant human PPT1 having a vIGF2 targeting domain (PPT1-2) and recombinant human PPT1 having a vIGF2 targeting domain and a BiP signal sequence (PPT1-29). Protein expression can be influenced by the variant of IGF used.

30 **[0064] FIG. 14** shows binding of PPT1 constructs to CI-MPR.

[0065] FIG. 15 shows GAA activity in conditioned media of CHO cells expressing engineered or natural hGAA.

[0066] FIG. 16 shows the study design of a 4-week mouse study of gene therapy in a GAA knockout mouse.

5 [0067] FIG. 17 shows GAA plasma activity in untreated wild type ("Normal") mice or GAA knockout mice treated with gene therapy vectors or vehicle as indicated.

[0068] FIG. 18 shows GAA levels measured in untreated wild type ("Normal") mice or GAA knockout mice treated with gene therapy vectors or vehicle as indicated.

10 [0069] FIG. 19 shows cell surface receptor CI-MPR binding of rhGAA from plasma samples obtained from treated mice as indicated.

[0070] FIG. 20 shows GAA activity, and quad glycogen histopathology score for tibialis anterior of untreated wild type ("Normal") mice or GAA knockout mice treated with gene therapy vectors or vehicle as indicated.

15 [0071] FIG. 21 shows glycogen PAS of tibialis anterior from untreated wild type mice or GAA knockout mice treated with gene therapy vectors or vehicle as indicated.

[0072] FIG. 22 shows hGAA immunohistochemistry of tibialis anterior from untreated wild type mice or GAA knockout mice treated with gene therapy vectors or vehicle as indicated.

20 [0073] FIG. 23 shows brain GAA activity, brain glycogen, and spinal cord glycogen histopathology scoring for brain and spinal cord from untreated wild type ("Normal") mice or GAA knockout mice treated with gene therapy vectors or vehicle as indicated.

[0074] FIG. 24 shows glycogen PAS of brain from untreated wild type mice or GAA knockout mice treated with gene therapy vectors or vehicle as indicated.

25 [0075] FIG. 25 shows hGAA immunohistochemistry of brainstem and choroid plexus from untreated wild type ("Normal") mice or GAA knockout mice treated with gene therapy vectors or vehicle as indicated.

[0076] FIG. 26 shows glycogen PAS of spinal cord from untreated wild type mice or GAA knockout mice treated with gene therapy vectors or vehicle as indicated.

[0077] FIG. 27 shows hGAA immunohistochemistry of spinal cord from untreated wild type mice or GAA knockout mice treated with gene therapy vectors or vehicle as indicated.

30 [0078] FIG. 28 shows quadriceps GAA activity and glycogen histopathology scoring from untreated wild type ("Normal") mice or GAA knockout mice treated with gene therapy vectors or vehicle as indicated.

[0079] **FIG. 29** shows glycogen luxol/PAS for quadriceps from untreated wild type mice or GAA knockout mice treated with gene therapy vectors or vehicle as indicated.

[0080] **FIG. 30** shows hGAA immunohistochemistry of quadriceps from untreated wild type mice or GAA knockout mice treated with gene therapy vectors or vehicle as indicated.

5 [0081] **FIG. 31** shows triceps GAA activity and histopathology scoring for untreated wild type (“Normal”) mice or GAA knockout mice treated with gene therapy vectors or vehicle as indicated.

[0082] **FIG. 32** shows glycogen luxol/PAS of triceps from untreated wild type mice or GAA knockout mice treated with gene therapy vectors or vehicle as indicated.

10 [0083] **FIG. 33** shows hGAA immunohistochemistry of triceps from untreated wild type mice or GAA knockout mice treated with gene therapy vectors or vehicle as indicated.

[0084] **FIG. 34** shows engineered and wild type PPT1 binding to CIMPR.

[0085] **FIG. 35** shows engineered and wild type TPP1 binding to CIMPR.

[0086] **FIG. 36** shows engineered and wild type AGA binding to CIMPR.

15 [0087] **FIG. 37** shows engineered and wild type GLA binding to CIMPR.

[0088] **FIG. 38** shows a Western blot of GAA from cells expressing various mutant vIGF2-GAA constructs in conditioned media.

[0089] **FIG. 39** shows secretion of new IGF2-GAA variants relative to vIGF2-GAA, constructs from Western blot of **FIG. 38**.

20 [0090] **FIG. 40** shows CI-MPR binding of various vIGF2-GAA constructs.

[0091] **FIG. 41** shows B_{max} and K_d values for CIMPR binding of various vIGF2-GAA constructs.

[0092] **FIG. 42** shows CI-MPR binding of various vIGF2-GAA constructs.

25 [0093] **FIG. 43** shows B_{max} and K_d values for CIMPR binding of various vIGF2-GAA constructs.

[0094] **FIG. 44** shows CI-MPR binding of various vIGF2-GAA constructs.

[0095] **FIG. 45** shows B_{max} and K_d values for CIMPR binding of various vIGF2-GAA constructs.

[0096] **FIG. 46** shows cell uptake for various vIGF2-GAA constructs.

30 [0097] **FIG. 47** shows cell uptake for various vIGF2-GAA constructs.

[0098] **FIG. 48** shows various vIGF2 peptides binding to CI-MPR or IGF2R.

[0099] **FIG. 49** shows PPT1 in conditioned media quantified by Western blot.

- [00100] FIG. 50 shows PPT1 in conditioned media quantified by Western blot.
- [00101] FIG. 51 shows PPT1 in conditioned media quantified by activity.
- [00102] FIG. 52 shows correlation between PPT1 Western blot quantification versus activity quantification.
- 5 [00103] FIG. 53 shows binding of PPT1 constructs to CI-MPR.
- [00104] FIG. 54 shows a structure diagram of selected PPT1 constructs.
- [00105] FIG. 55 shows Western blot of PPT1 secreted into conditioned media.
- [00106] FIG. 56 shows processing of PPT1 in the cell by Western blot.
- [00107] FIG. 57 shows PPT1 in conditioned media quantified by Western blot.
- 10 [00108] FIG. 58 shows relative PPT1 activity.
- [00109] FIG. 59 shows binding of PPT1 constructs to CI-MPR.
- [00110] FIG. 60 shows binding of PPT1 constructs to CI-MPR.
- [00111] FIG. 61 shows an alignment of variants of IGF2-GAA (1: vIGF2; 2: vIGF2-17; 3: IGF2-20; and 4: IGF2-22).
- 15 [00112] FIG. 62 shows additional PPT1 constructs.
- [00113] FIG. 63 (A) shows expression of PPT1 constructs, normalized to wild-type, untagged PPT1 (construct 100), as measured by the band intensity on a Western Blot. The average intensity for four replicate transfections is shown for each sample with standard deviation error bars. (B) shows PPT1 expression/secretion of PPT1 in media, normalized to
- 20 wild-type, as measured by the band intensity on a Western Blot.
- [00114] FIG 64 shows uptake into rat cortical neurons of PPT1 constructs as measured by immunofluorescence. (A) shows neuronal uptake of purified PPT1-101 and PPT1-104. (B) shows neuronal uptake of PPT-1 constructs from media (not purified).
- [00115] FIG. 65 shows additional NAGLU constructs.
- 25 [00116] FIG. 67 (A) shows expression of NAGLU constructs, normalized to wild-type, untagged PPT1 (construct 100), as measured by the band intensity on a Western Blot. The average intensity for four replicate transfections is shown for each sample with standard deviation error bars. (B) shows PPT1 expression/secretion of PPT1 in media, normalized to
- wild-type, as measured by the band intensity on a Western Blot.
- 30 [00117] FIG. 68 shows expression of TPP1 constructs, normalized to wild-type, untagged TPP1, as measured by the band intensity on a Western Blot.
- [00118] FIG. 69 shows CIMPR binding of TPP1 constructs.

[00119] FIG. 70 shows human CLN1 transgene expression as detected by RT-qPCR.

[00120] FIG. 71-72 show brain autofluorescent storage material (ASM) accumulation, a correlate of lysosomal dysfunction.

[00121] FIG. 73 shows Glial Fibrillary Acidic Protein (GFAP), a correlate of
5 astrogliosis and neuroinflammation.

DETAILED DESCRIPTION

[00122] Provided herein are novel, engineered IGF2 peptides with enhanced properties, including enhanced expression, secretion and CIMPR binding. Further provided herein are
10 fusion proteins and nucleic acids encoding fusion proteins comprising novel IGF2 peptides and lysosomal enzymes with enhanced properties, such as increased CIMPR binding and improved expression and secretion. The fusion proteins and nucleic acid constructs provided herein are useful both for enzyme replacement therapies and for gene therapies to treat lysosomal storage disorders

15 [00123] Gene therapy for single gene genetic disorders presents a potential one-time treatment for diseases and disorders, some of which have devastating symptoms that can appear early in life and sometimes lead to life-long disability. Genetic disorders, such as neurological disorders or lysosomal storage disorders, are often treated with enzyme replacement therapies which administer to the patient a therapeutic protein that is an active
20 form of the protein that is defective or deficient in the disease or disorder state. However, there are challenges for current therapies, including frequent treatments, development of an immune response to the therapeutic protein, and difficulty targeting the therapeutic protein to the affected tissue, cell, or subcellular compartment. Gene therapy offers advantages including a reduced number of treatments and long-lasting efficacy.

25 [00124] Provided herein are fusion proteins for administration as enzyme replacement therapy or encoded by vectors for gene therapy vectors that offer improvements to enzyme replacement therapy or gene therapy, such as providing more therapeutic protein where it is needed, thus improving treatment efficacy. Such challenges are addressed herein by improving expression and cellular uptake or delivery and intracellular or subcellular targeting of
30 therapeutic proteins. Specific tools or components provided herein include but are not limited to signal peptides (e.g., binding immunoglobulin protein (BiP) and Gaussia signal peptides) for increasing secretion and peptides that increase endocytosis of the therapeutic protein (e.g.,

peptides that bind to the CI-MPR with high affinity for increasing cellular uptake and lysosomal delivery). Such peptides are fused to therapeutic proteins encoded by gene therapy vectors. In some embodiments, the peptides are IGF2 (Insulin Like growth factor 2) peptides or variants thereof. Gene therapy vectors provided herein are contemplated to comprise, in
5 some embodiments, a nucleic acid encoding a therapeutic protein fused to a peptide that bind to the CI-MPR with high affinity for optimizing efficacy of gene therapy.

[00125] Gene therapy constructs for enzyme replacement gene therapy were designed. A translation initiation sequence, including, but not limited to a Kozak sequence or an IRES sequence, such as CrPV IRES, located at the 5' end of the construct, followed by a nucleic acid
10 encoding a signal peptide selected from one or more of a GAA signal peptide, a nucleic acid encoding an anti-trypsin inhibitor, and a nucleic acid encoding BiP sequence. These are followed by a nucleic acid encoding a cell targeting domain which can be a vIGF-2, a HIRMAb, or a TfRMAb or other cell targeting peptide or protein. The gene therapy construct further comprises a nucleic acid encoding a linker and a nucleic acid encoding a corrective
15 enzyme or enzymatically active fragment thereof, wherein the linker connects the cell targeting domain to the corrective enzyme, or enzymatically active fragment thereof. Suitable corrective enzymes include but are not limited to alpha-glucosidase (GAA), alpha-galactosidase (GLA), iduronidase (IDUA), iduronate-2-sulfatase (IDS), PPT1, TPP1, NAGLU, or enzymatically active fragments thereof, and other enzymes found deficient in an individual.

20 Intracellular targeting of therapeutic proteins

[00126] N-linked carbohydrates of most lysosomal proteins are modified to contain a specialized carbohydrate structure called mannose 6-phosphate (M6P). M6P is the biological signal that enables transport of lysosomal proteins to lysosomes via membrane-bound M6P
25 receptors. Enzyme replacement therapies for lysosomal storage disorders utilize M6P receptors for uptake and delivery of therapeutic proteins to lysosomes. Certain therapeutics do not utilize M6P receptors including Cerezyme® and other versions of recombinant human GCCase, utilize the mannose receptor that is able to bind terminal mannose on protein glycans and deliver to the lysosome. A problem facing certain enzyme replacement therapeutics is there are low amounts of M6P present on the enzyme therapeutic which necessitate higher
30 doses to reach therapeutic efficacy. This leads to substantially longer infusion times, higher probability of developing immune responses to the therapeutic, and higher drug demand, requiring increased protein manufacturing resulting in increased costs.

[00127] The CI-MPR captures M6P-containing lysosomal enzymes from circulation. The receptor has distinct binding domains for M6P and insulin-like growth factor (domains 1-3 and 7-9, see FIG. 2) and therefore is also known as the IGF2/Mannose-6-phosphate receptor or IGF2/CI-MPR. This receptor can be utilized for targeting M6P- or IGF2- or IGF2 variant-
 5 containing enzyme replacement therapeutics. Binding affinity of this receptor for these ligands including insulin-like growth factor is provided in Table 1. Notably, IGF2 peptide has a higher binding affinity for CI-MPR than mono- or bis-phosphorylated oligosaccharides.

Table 1: Ligands for CI-MPR	
Ligand	Binding Affinity (Apparent K_d; nM)
IGF2	0.03-0.2
[Leu27]IGF2	0.05
Bis-M6P	2
Beta-galactosidase	20
Pentamannose-M6P	6,000
Free M6P	7,000

[00128] Accordingly, in some embodiments, it is desired to design improved variant
 10 IGF2 (vIGF2) peptides for making therapeutic fusion proteins that have increased stability, CI-MPR binding, cellular uptake and lysosomal localization, for example in treating diseases such as lysosomal storage diseases.

[00129] In some embodiments, the variant vIGF2 has improved binding to CI-MPR which is responsible for cellular uptake and delivery of IGF2 to lysosomes for degradation.
 15 Some variant IGF2 peptides have decreased affinity for insulin-like growth factor receptor 1 (IGF1R). In some embodiments, IGF2 has decreased or no affinity for integrins. In some embodiments, the IGF2 also has decreased or no affinity for at least one insulin-like growth factor binding proteins (IGFBP1-6). In some embodiments, the IGF2 variants have decreased or no binding to heparin. In some embodiments, the IGF2 variants

[00130] A goal in designing a vIGF2 peptide would be to improve the biophysical
 20 properties of the vIGF2 and enhance binding to CI-MPR/cellular uptake and lysosomal delivery, while minimizing the other functions. Accordingly, vIGF2 peptides may (1) improve stability/solubility of vIGF2; (2) attenuate binding affinity to IR/IGF1R/integrins; and (3) improve binding affinity to CI-MPR. In some embodiments, vIGF2 peptides are designed

using structure guided rational design, identifying crucial versus dispensable residues, point mutations and truncations. In some embodiments, vIGF2 peptides are designed using in silico computational experiments comprising systemic mutational studies to determine if a given mutation affects stability and affinity to various binding partners, alanine scanning mutagenesis (NAM), and/or improving IGF2 solubility, bioavailability, and/or reducing immunogenicity. In some embodiments, vIGF2 peptides are designed via directed evolution based on split-GFP assays. In some embodiments, vIGF2 peptides are designed via directed evolution based on phage display.

[00131] In some embodiments, vIGF2 peptides are designed using in silico

computational experiments comprising systemic mutational studies to determine if a given mutation affects stability of the IGF2 peptide. In some embodiments, the stability of the peptide with the mutation is the same as or increased as compared to the wild type IGF2.

[00132] In some embodiments, vIGF2 peptides are designed to reduce binding to

integrin. In some embodiments, vIGF2 peptides with reduced binding to integrin comprise mutations R24E/R34E, R24E/R37E/R38E, R34E/R37E/R38E, R24E/R37E, R24E/R38E, or R24E/R34E/R37E/R38E. In some embodiments, vIGF2 peptides have reduced binding to integrin and heparin, such as mutation of residues R37, R38, or R40.

[00133] In some embodiments, mutations T16I, T16V, T16L, T16F, T16Y, or T16W increase binding of vIGF2 to CI-MPR. In some embodiments, mutations T16V or T16Y

increase binding of vIGF2 to CI-MPR. In some embodiments, mutations at D23, for example, D23K or D23R, increase binding of vIGF2 to CI-MPR. In some embodiments, mutations at F19, such as F19W, increase binding of vIGF2 to CI-MPR. In some embodiments, mutations at S50, such as S50D or S50E, increase binding of vIGF2 to CI-MPR. In some embodiments vIGF2 having mutations D23K and S50E have increased binding to CI-MPR. In some embodiments, vIGF2 having mutations Δ 1-4, E6R, Y27L, and K65R have increased binding to CI-MPR. In some embodiments, vIGF2 having mutations Δ 33-40, D23R, F26E, and S50E have increased binding to CI-MPR.

[00134] In some embodiments, vIGF2 peptides are designed to have reduced IGFR1

binding. In some embodiments, mutations that affect IGF1R binding are on the different face of IGF2 compared to mutations that affect CI-MPR binding. In some embodiments, F26, Y27, and V43 are important for binding to IGF1R. In some embodiments, vIGF2 peptides having a mutation of S29N, R34_GS, S39_PQ, R34_GS/S39_PQ, S29N/S39_PQ, or S29N/S39PQ,

R43_GS have decreased binding to insulin receptor and IGF1R. In some embodiments, a vIGF2 peptide having a mutation of S39_PQ (PQ insertion after S39) has decreased binding to the insulin receptor and IGF1R. In some embodiments, vIGF2 peptides having mutations at G11, V14, L17, G25, F26, Y27, F28, S29, R30, P31, A32, S33, V35, S36, R37, S39, G41, I42, V43, E44, F48, T53, Y59, C60, or A61 have reduced binding to IGF1R. In some
 5 embodiments, vIGF2 peptides having mutations at G10, L13, V14, L17, F26, Y27, F28, S29, R30, P31, A32, S33, V35, G41, I42, V43, T58, or Y59 have reduced binding to IGF1R. In some embodiments, vIGF2 peptides having mutations V14D/F26A/F28R/V43D have reduced binding to IGF1R. In some embodiments, vIGF2 peptides having mutations F26S, Y27L, or
 10 V43L have reduced binding to IGF1R and/or insulin receptor.

[00135] In some embodiments, vIGF2 peptides have a deletion in the C domain (e.g., residues 32-41, SRVSRRSR) causing the vIGF2 peptides to have reduced binding to IGF1R, insulin receptor, heparin, and integrin. In some embodiments the vIGF2 peptides have the mutation Δ 1-4, E6R, Δ 30-39. In some embodiments, the vIGF2 peptides have the mutation
 15 Δ 1-4, E6R, Δ 33-40.

[00136] In some embodiments, vIGF2 peptides have mutations to decrease its instability index. In some embodiments, mutations of IGF2 peptides with increased stability include R38G, R38G/E45W, R38G/E45W/S50G, P31G/R38G/E45W/S50G, or L17N/P31G/R38G/E45W/S50G. In some embodiments, mutations of IGF2 peptides with
 20 increased stability include R38G, R38G/E45W, R38G/E45W/S50G, P31G/R38G/E45W/S50G, L17N/P31G/R38G/E45W/S50G, L17N/P31G/R38G/E45W/S50G/S66G, L17N/P31G/R38G/E45W/S50G/A64M/S66G, or S5L/L17N/P31G/R38G/E45W/S50G/A64M/S66G.

[00137] In some embodiments, vIGF2 peptides are mutated to reduce aggregation. In some
 25 embodiments, residues prone to aggregation include residues 17-21 (LQFVC), 41-49 (GIVEECCFR), or 53-62 (LALLETYCAT). In some embodiments, vIGF2 peptides are mutated at F26, Y59, Y27, V14, A1, or L8 to reduce aggregation.

[00138] In some embodiments, vIGF2 peptides are designed to have reduced binding to IGFBP. In some embodiments, vIGF2 peptides have the mutations L8A, V20A, or L56A. In
 30 some embodiments, vIGF2 peptides having mutations at E6, L8, R24, G25, F26, Y27, or F28 have reduced binding to IGFBP4. In some embodiments, vIGF2 peptides having mutations at T7, G10, V14, V43, E44, C47, or F48 have reduced binding to IGFBP4. In some

embodiments, vIGF2 peptides having mutations at E6 or L8 have reduced binding to IGFBP4. In some embodiments, vIGF2 peptides having mutations E6Q or T7A have reduced binding to human serum binding protein. In some embodiments, vIGF2 peptides having mutations Q18Y or F19L have reduced binding to human serum binding protein. In some embodiments, vIGF2 peptides having mutations at E6Q, T7A, Q18Y, or F19L have reduced binding to human serum binding protein.

[00139] In some embodiments, vIGF2 peptides have been modified to replace residues 31-38 with GGGG (vIGF2 Δ 31-38GGGG), and some of these vIGF2 peptides further contain a V43L and an S50E or S50D mutation. (SEQ ID NO:s 120-121). In some embodiments, vIGF2 peptides that are at least 95% identical to SEQ ID NO:s 120-121 enhance expression and/or secretion of a therapeutic protein. In some embodiments, the therapeutic protein is PPT1 or TPP1 or an enzymatically active fragment thereof.

Therapeutic Fusion Proteins for Gene Therapy

[00140] Therapeutic fusion proteins produced from gene therapy vectors are provided herein. In some embodiments the fusion protein is secreted by cells transduced with the gene therapy vector encoding the fusion protein. In some embodiments, the transduced cells are within a tissue or organ (e.g., liver). Once secreted from a cell, the fusion protein is transported through a patient's vascular system and reaches the tissue of interest. In some embodiments, the therapeutic fusion protein is engineered to have improved secretion. In some embodiments, the fusion protein comprises a signal peptide for improving the secretion level as compared to the corresponding therapeutic protein or a fusion protein comprising the therapeutic protein having only a native signal peptide.

[00141] The provided gene therapy vectors are, in some embodiments, engineered to improve delivery of the therapeutic protein. For example, in some instances gene therapy may not achieve the intended treatment by merely generating a sufficient amount of a therapeutic protein in the body of the patient if an insufficient amount of the therapeutic protein is delivered *into* the cells in need of the therapeutic protein, due to, for example, physical and/or biological barriers that impede distribution of the therapeutic protein to the site where needed. As such, even if a gene therapy is capable of flooding blood or a tissue, to a point of saturation, with a high concentration of a therapeutic protein, the gene therapy may not be sufficiently therapeutic. Additionally, non-productive clearance pathways may remove the vast majority of

the therapeutic protein. Even if the therapeutic protein is transported out of the vasculature to the interstitial space within the tissue (e.g., muscle fibers), adequate therapeutic effects are not assured. For effective treatment of lysosomal storage disorders, a therapeutically effective amount of the therapeutic protein must undergo cellular endocytosis and lysosomal delivery to result in a meaningful efficacy. The present disclosure addresses these issues by providing gene therapy vectors encoding fusion proteins comprising a peptide that enables endocytosis of the therapeutic protein into a target cell for treatment resulting in efficacious treatment. In some embodiments, the peptide that enables endocytosis is a peptide that binds the CI-MPR. In some embodiments, the peptide that binds the CI-MPR is a vIGF2 peptide. Recombinantly expressed GLA was known to be well phosphorylated and thus bind to the CIMPR, but surprisingly, GLA expressed in mice is under-phosphorylated and does not bind well to the CIMPR. Therefore, GLA for use in gene therapy unexpectedly requires additional engineering to enhance CIMPR binding (such as the IGF2 tag).

[00142] Provided herein are gene therapy vectors encoding fusion proteins comprising a peptide that enables endocytosis the therapeutic protein into a target cell for treatment. In some embodiments, the gene therapy vectors encode fusion proteins comprising a therapeutic protein and a peptide that binds the CI-MPR. Such fusion proteins when expressed from a gene therapy vector target therapeutic proteins, such as enzyme replacement therapeutics, to the cells where they are needed, increase delivery into or cellular uptake by such cells and target the therapeutic protein to a subcellular location (e.g., a lysosome). In some embodiments, the peptide is an IGF2 peptide or variant thereof, which can target a therapeutic protein to the lysosome. Fusion proteins herein also, in some embodiments, further comprise a signal peptide that increases secretion, such as a BiP signal peptide or a Gaussia signal peptide. In some embodiments, fusion proteins comprise a linker sequence. In some embodiments, nucleic acids encoding fusion proteins herein, comprise internal ribosomal entry sequences.

Uh

Therapeutic Fusion Proteins for Enzyme Replacement Therapy

[00143] Therapeutic fusion proteins produced for enzyme replacement therapy are provided herein. The provided fusion proteins are, in some embodiments, engineered to improve delivery of the therapeutic protein. For example, in some instances fusion protein may not achieve the intended treatment if an insufficient amount of the therapeutic fusion protein is delivered *into* the cells in need of the therapeutic protein, due to, for example, physical and/or

biological barriers that impede distribution of the therapeutic protein to the site where needed. Even if the therapeutic protein is transported out of the vasculature to the interstitial space within the tissue (e.g., muscle fibers), adequate therapeutic effects are not assured. For effective treatment of lysosomal storage disorders, a therapeutically effective amount of the therapeutic protein must undergo cellular endocytosis and lysosomal delivery to result in a meaningful efficacy. The present disclosure addresses these issues by providing fusion proteins comprising a peptide that enables endocytosis of the therapeutic protein into a target cell for treatment resulting in efficacious treatment. In some embodiments, the peptide that enables endocytosis is a peptide that binds the CI-MPR. In some embodiments, the peptide that binds the CI-MPR is a vIGF2 peptide.

[00144] Provided herein are fusion proteins comprising a peptide that enables endocytosis the therapeutic protein into a target cell for treatment. In some embodiments, the fusion proteins comprises a peptide that binds the CI-MPR. Such fusion proteins are used as enzyme replacement therapeutics, have increased delivery into or cellular uptake by cells needing such proteins and target the therapeutic protein to a subcellular location (e.g., a lysosome). In some embodiments, the peptide is an IGF2 peptide or variant thereof, which can target a therapeutic protein to the lysosome.

[00145] Therapeutic proteins for enzyme replacement therapy or gene therapy comprising a vIGF2 peptide are provided herein. Exemplary proteins are provided in Table 2 below.

Table 2: Amino Acid Sequences		SEQ ID NO
Natural hGAA	MGVRHPPCSHRL LAVCALVSLATAALLGHILLHDFLLVPRELSGSSP VLEETHPAHQQGASRPGPRDAQAHPGRPRAVPTQCDVPPNSRFDCA PDKAITQEQCEARGCCYIPAKQGLQGAQMGPWCFFPPSYPSYKLE NLSSSEMGYTATLTRTTPTFFPKDILTLRLDVMMETENRLHFTIKDP ANRRYEVPLETPHVHSRAPSPLYSVEFSEEPFGVIVRRQLDGRVLLN TTVAPLFFADQFLQLSTSLPSQYITGLAEHLSPLMLSTSWTRITLWNR DLAPTPGANLYGSHPFYLALEDGGSAGHVFLNNSNAMDVVLPQSPA LSWRSTGGILDVYIFLGPEPKSVVQQYLDVVGYPFMPYPYWGGLGFHL CRWGYSSAITRQVVENMTRAHFPLDVQWNDLDYMDSRRDFTFN KDGFRDFPAMVQELHQGGRRYMMIVDPAISSSGPAGSYRPYDEGLR	1

	RGVFITNETGQPLIGKVWPGSTAFPDFTNPTALAWWEDMVAEFHD QVPFDGMWIDMNEPSNFIRGSEDGCPNNELENPPYVPGVVGGTLQA ATICASSHQFLSTHYNLHNLYGLTEAIAASHRALVKARGTRPFVISRST FAGHGRYAGHWTGDVWSSWEQLASSVPEILQFNLLGVPLVGADVC GFLGNTSEELCVRWTQLGAFYPFMRNHNSLLSLPQEPYSFSEPAQQ AMRKALTLRYALLPHLYTLFHQAHVAGETVARPLFLEFPKDSSTWT VDHQLLWGEALLITPVLQAGKAEVTGYFPLGTWYDLQTVPVEALG SLPPPPAAPREPAIHSEGQWVTLPAPLDTINVHLRAGYIIPLQGPGLT TTESRQQPMALAVALT KGGEARGELFWDDGESLEVLERGAYTQVIF LARNNTIVNELVRVTSEGAGLQLQKVTVLGVATAPQQVLSNGVPVS NFTYSPDTKVLDICVSLLMGEQFLVSWC	
Engineered hGAA (BiP- vIGF2- GAA)	MKLSLVAAML LLLS AARASRTL CGGELVDTLQFVCGDRGFLFSRPA SRVSRRSRGIVEECCFRSCDLALLE TYCATPARSEG GGGSGGGGSRP GPRDAQAH PGRPRAVPTQCDVPPNSRFDCAPDKAITQE QCEARGCC YIPAKQGLQGAQMGPWCFFPPSYPSYKLENLSSEM GYTATLTRT TPTFFPKDILTLRLDVM METENRLHFTIKDPANRRYEVPLET PHVHS RAPSPLYSVEFSEEPFGVIVRRQLDGRVLLNTTVAPLFFADQFLQLST SLPSQYITGLAEHLSPLMLSTSWTRITLWNRDLAPTPGANLYGSHPF YLALEDGGS AHGVFLN SNAMDVVLQPSPALSWRSTGGILDVYIFL GPEPKSVVQQYLDVVGYPFMP PYWGLGFHL CRWGYSSAITRQVV ENMTRAHFPLDVQWNDLDYMD SRRDFTFNKDGFRDFPAMVQELH QGGRRYMMIVDPAISSSGPAGSYR PYDEGLRRGVFITNETGQPLIGK VWPGSTAFPDFTNPTALAWWEDMVAEFHDQVPFDGMWIDMNEPS NFIRGSEDGCPNNELENPPYVPGVVGGTLQAATICASSHQFLSTHYN LHNLYGLTEAIAASHRALVKARGTRPFVISRSTFAGHGRYAGHWTGD VWSSWEQLASSVPEILQFNLLGVPLVGADVCGFLGNTSEELCVRWT QLGAFYPFMRNHNSLLSLPQEPYSFSEPAQQAMRKALTLRYALLPH LYTLFHQAHVAGETVARPLFLEFPKDSSTWTVDHQLLWGEALLITP VLQAGKAEVTGYFPLGTWYDLQTVPVEALGSLPPPPAAPREPAIHSE GQWVTLPAPLDTINVHLRAGYIIPLQGPGLTTTESRQQPMALAVALT KGGEARGELFWDDGESLEVLERGAYTQVIFLARNNTIVNELVRVTS EGAGLQLQKVTVLGVATAPQQVLSNGVPVSNFTYSPDTKVLDICVS	2

	LLMGEQFLVSWC	
hGAA Δ1-60	SRPGPRDAQAHPGRPRAVPTQCDVPPNSRFDCAPDKAITQEQCEAR GCCYIPAKQGLQGAQMGPWCFFPPSYPSYKLENLSSSEMGYTATL TRTTPTFFPKDILTRLDDVMMETENRLHFTIKDPANRRYEVPLETPH VHSRAPSPLYSVEFSEEPFGVIVRRQLDGRVLLNTTVAPLFFADQFL QLSTSLPSQYITGLAEHLSPLMLSTSWTRITLWNRDLAPTPGANLYG SHPFYLALEDGGSAGHVFLLSNAMDVVLQSPALSWRSTGGILDV YIFLGPEPKSVVQQYLDVVGYPFMPPYWGLGFHLCRWGYSSAITR QVVENMTRAHFPLDVQWNDLDYMDSRRDFTFNKDGFRDFPAMVQ ELHQGGRRYMMIVDPAISSSGPAGSYRPHYDEGLRRGVFITNETGQPL IGKVWPGSTAFPDFTNPTALAWWEDMVAEFHDQVPFDGMWIDMN EPSNFIRGSEDGCPNNELENPPYVPGVVGGTLQAATICASSHQFLST HYNLHNLYGLTEAIAASHRALVKARGTRPFVISRSTFAGHGRYAGHW TGDVWSSWEQLASSVPEILQFNLLGVPLVGADVCGFLGNTSEELCV RWTQLGAFYPFMRNHNSLLSLPQEPYSFSEPAQQAMRKALTRYAL LPHLYTLFHQAHVAGETVARPLFLEFPKDSSTWTVDHQLLWGEALL ITPVLQAGKAEVTGYFPLGTWYDLQTVPEALGSLPPPPAAPREPAI HSEGGWVTLAPLDTINVHLRAGYIIPLQGPGLTTTESRQQPMALAV ALTKGGEARGELFWDDGESLEVLERGAYTQVIFLARNTIVNELVR VTSEGAGLQLQKVTVLGVATAPQQVLSNGVPVSNFTYSPDTKVLDI CVSLLMGEQFLVSWC	3
wt- palmitoyl- protein thioesterase 1 (PPT1)	MASPGCLWLLAVALLPWTCASRALQHLDPPAPLPLVIWHGMGDSC CNPLSMGAIKKMVEKKIPGIYVLSLEIGKTLMEDVENSFFLNVNSQV TTVCQALAKDPKLQQGYNAMGFSQGGQFLRAVAQRCPSPPMINLIS VGGQHQQGVFGLPRCPGESSHICDFIRKTLNAGAYSKVQERLVQAE YWHDPKEDVYRNHSIFLADINQERGINESYKKNLMALKKFVMVKF LNDIVDPVDSEWFGFYRSGQAKETIPLQETSLEYTQDRLGLKEMDN AGQLVFLATEGDHLQLSEWFYAHIPFLG	4
PPT1-2 (vIGF2- PPT1)	MASPGCLWLLAVALLPWTCASRALQHLSRTLCCGGELVDTLQFVCG DRGFLFSRPASRVSRRSRGIVECCFRSCDLALLETYCATPARSEGG GGSGGGGSRPRAVPTQDPPAPLPLVIWHGMGDSCCNPLSMGAIKK MVEKKIPGIYVLSLEIGKTLMEDVENSFFLNVNSQVTTVCQALAKDP	5

	KLQQGYNAMGFSQGGQFLRAVAQRCPSPPMINLISVGGQHQG VFGLPRCPGESSHICDFIRKTLNAGAYSKVVQERLVQAEYWHDPKEDVYRNHSIFLADINQERGINESYKKNLMALKKFVMVKFLNDSIVDPVDSEWFGFYRSGQAKETIPLQETSLYTQDRLGLKEMDNAGQLVFLATEGDHLQLSEEWFYAHIIPFLG	
PPT1-29 (BiP2aa- vIGF2- PPT1)	MKLSLVAAMLLLLWVALLLLSAARAAASRTLCTGGELVDTLQFVCGDRGFLFSRPASRVSRRSRGIVEECCFRSCDLALLETYCATPARSEGGSGGGGSRPRAVPTQDPPAPLPLVIWHGMGDSCCNPLSMGAIKKMVEKKIPGIYVLSLEIGKTLMEDVENSFFLNVNVSQVTTVCQALAKDPKLQQGYNAMGFSQGGQFLRAVAQRCPSPPMINLISVGGQHQG VFGLPRCPGESSHICDFIRKTLNAGAYSKVVQERLVQAEYWHDPKEDVYRNHSIFLADINQERGINESYKKNLMALKKFVMVKFLNDSIVDPVDSEWFGFYRSGQAKETIPLQETSLYTQDRLGLKEMDNAGQLVFLATEGDHLQLSEEWFYAHIIPFLG	6
PPT1 engineered	MASPGCLWLLAVALLPWTCASRALQHLDPPAPLPLVIWHGMGDSCCNPLSMGAIKKMVEKKIPGIYVLSLEIGKTLMEDVENSFFLNVNVSQVTTVCQALAKDPKLQQGYNAMGFSQGGQFLRAVAQRCPSPPMINLISVGGQHQG VFGLPRCPGESSHICDFIRKTLNAGAYSKVVQERLVQAEYWHDPKEDVYRNHSIFLADINQERGINESYKKNLMALKKFVMVKFLNDSIVDPVDSEWFGFYRSGQAKETIPLQETSLYTQDRLGLKEMDNAGQLVFLATEGDHLQLSEEWFYAHIIPFLGRPRAVPTQGGSGSGSTS SRRTLCTGGELVDTLQFVCGDRGFLFSRPASRVSRRSRGIVEECCFRECDLALLETYCATPARSE	7
TPP1 wildtype	MGLQACLLGLFALILSGKCSYSPEPDQRRTLPPGWVSLGRADPEEELSLTFALRQQNVERLSELVQAVSDPSSPQYGKYLTENVADLVRPSPLTLHTVQKWLLAAGAQAQKCHSVITQDFLTCWLSIRQAELLPLGAEFHHYVGGPTETHVVRSPHPYQLPQALAPHVDFVGGHLHRFPPTSSLRQRPEPQVTGTVGLHLGVTPSVIRKRYNLTSQDVGSGTSNNSQACAQFLEQYFHDSDLAQFMRLFNGFAHQASVARVVGQQGRGRAGIEASLDVQYLMSAGANISTWVYSSPGRHEGQEPFLQWMLLSNESALPHVHTVS YGDDDSLSSAYIQRVNTELMKAAARGLTLLFASGDSGAGCWSVS GRHQFRPTFPASSPYVTTVGGTSFQEPFLITNEIVDYISGGGFSNVFPR	8

	PSYQEEAVTKFLSSSPHLPPSSYFNASGRAYPDVAALSDGYWVVS RVPIPWVSGTSASTPVFGGILSLINEHRILSGRPPLGFLNPRLYQQHG AGLFDVTRGCHESCLDEEVEGQGFCSGPGWDPVTGWGTPNFPALL KTLLNP	
TPP1 engineered	MGLQACLLGLFALILSGKCSRTLCCGELVDTLQFVCGDRGFLFSRPA SRVSRRSRGIVEECCFRSCDLALLETYCATPARSEGGGGSGGGSRP RAVPTQSYSPEPDQRRTLPPGWVSLGRADPEEELSLTFALRQQNVER LSELVQAVSDPSSPQYGKYLLENVADLVRPSPLTLHTVQKWLLAA GAQKCHSVITQDFLTCWLSIRQAELLPGAEEFHHYVGGPTETHVVR SPHPYQLPQALAPHVDFVGGHLHRFPPTSSLRQRPEPQVTGTVGLHLG VTPSVIRKRYNLTSQDVGSGTSNNSQACAQFLEQYFHDSDLAQFMR LFGGNFAHQASVARVVGQQGRGRAGIEASLDVQYLMSAGANISTW VYSSPGRHEGQEPFLQWLMLLSNESALPHVHTVSYGDDDEDSLSSAY IQRVNTELMKAAARGLTLLFASGDSGAGCWSVSGRHQFRPTFPASS PYVTTVGGTSFQEPFLITNEIVDYISGGGFSNVFPRPSYQEEAVTKFL SSSPHLPPSSYFNASGRAYPDVAALSDGYWVVSNRVPIPWVSGTSAS TPVFGGILSLINEHRILSGRPPLGFLNPRLYQQHGAGLFDVTRGCHES CLDEEVEGQGFCSGPGWDPVTGWGTPNFPALLKTLLNP	9
AGA wildtype	MARKSNLPVLLVPFLLCQALVRCSSPLPLVVNTWPFKNATEAAWR ALASGGSALDAVESGCAMCEREQCDGSVGFGGSPDELGETTLDAMI MDGTTMDVGAVGDLRRIKNAIGVARKVLEHTTHTLLVGESATTFA QSMGFINEDLSTTASQALHSDWLARNCQPNYWRNVIPDPSKYCGPY KPPGILKQDIPHIKETEDDRGHDTIGMVVIHKTGHIAAGTSTNGIKFK IHGRVGDSPGAGAYADDTAGAAAATGNGDILMRFLPSYQAVEY MRRGEDPTIACQKVISRIQKHFPEFFGAVICANVTGSYGAACNKLST FTQFSFMVYNSEKNQPTEEKVDCI	10
AGA engineered (N-terminal fusion)	MARKSNLPVLLVPFLLCQALVRCSSRTLCCGELVDTLQFVCGDRGFL FSRPASRVSRRSRGIVEECCFRSCDLALLETYCATPARSEGGGGSGG GGSRPRAVPTQSSPLPLVVNTWPFKNATEAAWRALASGGSALDAV ESGCAMCEREQCDGSVGFGGSPDELGETTLDAMIMDGTTMDVGAV GDLRRIKNAIGVARKVLEHTTHTLLVGESATTFAQSMGFINEDLSTT ASQALHSDWLARNCQPNYWRNVIPDPSKYCGPYKPPGILKQDIPHI	11

	KETEDDRGHDITIGMVVIHKTGHIAAGTSTNGIKFKIHGRVGDSPiPG AGAYADDTAGAAAATGNGDILMRFLPSYQAVEYMRRGEDPTIACQ KVISRIQKHFPEFFGAVICANVTGSYGAAACNKLSTFTQFSFMVYNSE KNQPTEEKVDCI	
GLA wildtype	MQLRNPELHLGCALALRFLALVSWDIPGARALDNGLARTPTMGWL HWERFMCNLDCQEEDSCISEKLFMEMAELMVSEGWKDAGYEYL CIDDCWMAQRDSEGRQLQADPQRFPHGIRQLANYVHSGKLKLGIIYA DVGKNTCAGFPGSFGYYDIDAQTFADWGVDLLKFDGCYCDLENL ADGYKHMSLALNRTGRSIVYSCEWPLYMWPFQKPNYTEIRQYCNH WRNFADIDDSWKSISILDWTSFNQERIVDVAGPGGWNDPDMLVIG NFGLSWNQQVTQMALWAIMAAPLFMSNDLRHISPAKALLQDKD VIAINQDPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQEIG GPRSYTIAVASLGKGVACNPACFITQLLPVKKRLGFYEWTSLRSHI NPTGTVLLQLENTMQMSLKDLL	12
GLA engineered	MQLRNPELHLGCALALRFLALVSWDIPGARALDNGLARTPTMGWL HWERFMCNLDCQEEDSCISEKLFMEMAELMVSEGWKDAGYEYL CIDDCWMAQRDSEGRQLQADPQRFPHGIRQLANYVHSGKLKLGIIYA DVGKNTCAGFPGSFGYYDIDAQTFADWGVDLLKFDGCYCDLENL ADGYKHMSLALNRTGRSIVYSCEWPLYMWPFQKPNYTEIRQYCNH WRNFADIDDSWKSISILDWTSFNQERIVDVAGPGGWNDPDMLVIG NFGLSWNQQVTQMALWAIMAAPLFMSNDLRHISPAKALLQDKD VIAINQDPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQEIG GPRSYTIAVASLGKGVACNPACFITQLLPVKKRLGFYEWTSLRSHI NPTGTVLLQLENTMQMSLKDLLYIPAKQGLQGAQMGPGGGGSGG GGSRITLCGGELVDTLQFVCGDRGFLFSRPASRVSRRSRGIVEECCFR SCDLALLETYCATPARSE	13
BiP-vIGF2- 17-2GS- GAA	MKLSLVAAMLLLLSAARASRTLCGGELVDTLQFVCGDRGFLFSRPA SRVSRRSRGIVEECCFRECDLALLETYCATPARSEGGGGSGGGGSRP GPRDAQAHPGRPRAVPTQCDVPPNSRFDCAPDKAITQEQCEARGCC YIPAKQGLQGAQMGPWCFFPPSYPSYKLENLSSEMGYTATLTRT TPTFFPKDILTLRLDVMMETENRLHFTIKDPANRRYEVPLETPHVHS RAPSPLYSVEFSEEPFGVIVRRQLDGRVLLNTTVAPLFFADQFLQLST	14

	SLPSQYITGLAEHLSPLMLSTSWTRITLWNRDLAPTPGANLYGSHPF YLALEDGGSAGHVFLLSNAMDVVLQSPALSWRSTGGILDVYIFL GPEPKSVVQQYLDVVGYPFMPYWGLGFHLCRWGYSSTAITRQVV ENMTRAHFPLDVQWNDLDYMDSRDFTFNKDGFRDFPAMVQELH QGGRRYMMIVDPAISSSGPAGSYRPFYDEGLRRGVFITNETGQPLIGK VWPGSTAFPDFTNPTALAWWEDMVAEFHDQVPFDGMWIDMNEPS NFIRGSEDGCPNNELENPPYVPGVVGGTLQAATICASSHQFLSTHYN LHNLYGLTEAIAASHRALVKARGTRPFVISRSTFAGHGRYAGHWTGD VWSSWEQLASSVPEILQFNLLGVPLVGADVCGFLGNTSEELCVRWT QLGAFYPFMRNHNSLLSLPQEPYSFSEPAQQAMRKALTRYALLPH LYTLFHQAHVAGETVARPLFLEFPKDSSTWTVDHQLLWGEALLITP VLQAGKAEVTGYFPLGTWYDLQTVPVEALGSLPPPPAAPREPAIHSE GQWVTLPAPLDTINVHLRAGYIIPLQGPGLTTTESRQQPMALAVALT KGGEARGELFWDDGESLEVLERGAYTQVIFLARNNTIVNELVRVTS EGAGLQLQKVTVLGVATAPQQVLSNGVPVSNFTYSPDTKVLDICVS LLMGEQFLVSWC	
BiP-vIGF2- 20-2GS- GAA	MKLSLVAAML LLLSAARASRTL CGGELVDTLQFVCGDRGFLFSRPA SRVSRRSRGILECCFRSCDLALLE TYCATPARSEGGGGSGGGGSRP GPRDAQAHPGRPRAVPTQCDVPPNSRFDCAPDKAITQE QCEARGCC YIPAKQGLQGAQMGPWCFFPPSYPSYKLENLSSEMGYTATLTRT TPTFFPKDILTLRLDVMMETENRLHFTIKDPANRRYEVPLETPHVHS RAPSPLYSEVFSEEPFGVIVRRQLDGRVLLNTTVAPLFFADQFLQLST SLPSQYITGLAEHLSPLMLSTSWTRITLWNRDLAPTPGANLYGSHPF YLALEDGGSAGHVFLLSNAMDVVLQSPALSWRSTGGILDVYIFL GPEPKSVVQQYLDVVGYPFMPYWGLGFHLCRWGYSSTAITRQVV ENMTRAHFPLDVQWNDLDYMDSRDFTFNKDGFRDFPAMVQELH QGGRRYMMIVDPAISSSGPAGSYRPFYDEGLRRGVFITNETGQPLIGK VWPGSTAFPDFTNPTALAWWEDMVAEFHDQVPFDGMWIDMNEPS NFIRGSEDGCPNNELENPPYVPGVVGGTLQAATICASSHQFLSTHYN LHNLYGLTEAIAASHRALVKARGTRPFVISRSTFAGHGRYAGHWTGD VWSSWEQLASSVPEILQFNLLGVPLVGADVCGFLGNTSEELCVRWT QLGAFYPFMRNHNSLLSLPQEPYSFSEPAQQAMRKALTRYALLPH	15

	LYTLFHQAHVAGETVARPLFLEFPKDSSTWTVDHQLLWGEALLITP VLQAGKAEVTGYFPLGTWYDLQTVPVEALGSLPPPPAAPREPAIHSE GQWVTLPAPLDTINVHLRAGYIIPLQGPGLTTTESRQQPMALAVALT KGGGEARGELFWDDGESLEVLERGAYTQVIFLARNNTIVNELVRVTS EGAGLQLQKVTVLGVATAPQQVLSNGVPVSNFTYSPDTKVLDICVS LLMGEQFLVSWC	
BiP-vIGF2- 22-2GS- GAA	MKLSLVAAML LLLSAARASRTL CGGELVDTLQFVCGDRGFLFSRG GGGSRGIVEECCFRSCDLALLE TYCATPARSEG GGGSGGGGSRPGPR DAQAHPGRPRAVPTQCDVPPNSRFD CAPDKAITQE QCEARGCCYIP AKQGLQGAQMGPWCFFPPSYPSYKLENLSSEMGYTATLTRTTPT FFPKDILTLRLDVM METENRLHFTIKDPANRRYEVPLETPHVHSRAP SPLYSV EFSEEPFGVIVRRQLDGRVLLNTTVAPLFFADQFLQLSTSLP SQYITGLAEHLSPLMLSTSWTRITLWNRDLAPTGANLYGSHPFYLA LEDGGS AHGVFLLNSNAMDVVLQPSPALSWRSTGGILDVYIFLGPEP KSVVQQYLDVVGYPFMP PYWGLGFHL CRWGYSSAITRQVVENM TRAHFPLDVQWNDLDYMDSR RDFTFNKDGFRDFPAMVQELHQGG RRYMMIVDPAISSSGPAGSYR PYDEGLRRGVFITNETGQPLIGKVWP GSTAFPDFTNPTALAWWEDMVAEFHDQVPFDGMWIDMNEPSNFIR GSEDGCPNNELENPPYVPGVVG GTLQAATICASSHQFLSTHYNLHN LYGLTEAIASHRALVKARGTRPFVISRSTFAGHG RYAGHWTGDVWS SWEQLASSVPEILQFNLLGVPLVGADVCGFLGNTSEELCVRWTQLG AFYPPMRNHNSLLSLPQEPYSFSEPAQQAMRKALTRYALLPHLYT LFHQAHVAGETVARPLFLEFPKDSSTWTVDHQLLWGEALLITPVLQ AGKAEVTGYFPLGTWYDLQTVPVEALGSLPPPPAAPREPAIHSEGG WVTLPAPLDTINVHLRAGYIIPLQGPGLTTTESRQQPMALAVALT KGGGEARGELFWDDGESLEVLERGAYTQVIFLARNNTIVNELVRVTSE GAGLQLQKVTVLGVATAPQQVLSNGVPVSNFTYSPDTKVLDICVSL LMGEQFLVSWC	16
PPT1-3 (BiP-PPT1)	MKLSLVAAML LLLSAARADPPAPLPLVIWHGMGDSCCNPLSMGAI KKMVEKKIPGIYVLSLEIGKTLMEDVENSFFLN VNSQVTTVCQALA KDPKLQQGYNAMGFSQGGQFLRAVAQRCPSP PMINLISVGGQHQG VFGLPRCPGESSHICDFIRKTLNAGAYS KVVQERLVQA EYWH DPIKE	17

	DVYRNHSIFLADINQERGINESYKKNLMALKKFVMVKFLNDSIVDP VDSEWFGFYRSGQAKETIPLQETSLYTQDRLGLKEMDNAGQLVFLA TEGDHLQLSEEWFYAHIIIPFLG	
PPT1-4 (BiP-vIGF2- PPT1)	MKLSLVAAMLLLLSAARASRTLCCGELVDTLQFVCGDRGFLFSRPA SRVSRRSRGIVEECCFRSCDLALLETYCATPARSEGSGGGSGGGSRP RAVPTQDPPAPLPLVIWHGMGDSCCNPLSMGAIKKMVEKKIPGIYV LSLEIGKTLMEDVENSFFLNVSQVTTVCQALAKDPKLQQGYNAM GFSQGGQFLRAVAQRCPSPPMINLISVGGQHQQGVFGLPRCPGESSHI CDFIRKTLNAGAYSKVQERLVQAEYWHDPIKEDVYRNHSIFLADI NQERGINESYKKNLMALKKFVMVKFLNDSIVDPVDSEWFGFYRSG QAKETIPLQETSLYTQDRLGLKEMDNAGQLVFLATEGDHLQLSEEW FYAHIIIPFLG	18
PPT1-5 (wt- PPT1- vIGF2)	MASPGCLWLLAVALLPWTCASRALQHLDPPAPLPLVIWHGMGDSC CNPLSMGAIKKMVEKKIPGIYVLSLEIGKTLMEDVENSFFLNVSQV TTVCQALAKDPKLQQGYNAMGFSQGGQFLRAVAQRCPSPPMINLIS VGGQHQQGVFGLPRCPGESSHICDFIRKTLNAGAYSKVQERLVQAE YWHDPIKEDVYRNHSIFLADINQERGINESYKKNLMALKKFVMVKF LNDIVDPVDSEWFGFYRSGQAKETIPLQETSLYTQDRLGLKEMDN AGQLVFLATEGDHLQLSEEWFYAHIIIPFLGRPRAVPTQGGGGSGGG GSRTLCCGELVDTLQFVCGDRGFLFSRPAASRVSRRSRGIVEECCFRS CDLALLETYCATPARSE	19
PPT1-9 (wt- PPT1)	MASPGCLWLLAVALLPWTCASRALQHLDPPAPLPLVIWHGMGDSC CNPLSMGAIKKMVEKKIPGIYVLSLEIGKTLMEDVENSFFLNVSQV TTVCQALAKDPKLQQGYNAMGFSQGGQFLRAVAQRCPSPPMINLIS VGGQHQQGVFGLPRCPGESSHICDFIRKTLNAGAYSKVQERLVQAE YWHDPIKEDVYRNHSIFLADINQERGINESYKKNLMALKKFVMVKF LNDIVDPVDSEWFGFYRSGQAKETIPLQETSLYTQDRLGLKEMDN AGQLVFLATEGDHLQLSEEWFYAHIIIPFLG	20
PPT1-10 (wt-PPT1- vIGF2_2)	MASPGCLWLLAVALLPWTCASRALQHLDPPAPLPLVIWHGMGDSC CNPLSMGAIKKMVEKKIPGIYVLSLEIGKTLMEDVENSFFLNVSQV TTVCQALAKDPKLQQGYNAMGFSQGGQFLRAVAQRCPSPPMINLIS VGGQHQQGVFGLPRCPGESSHICDFIRKTLNAGAYSKVQERLVQAE	21

	YWHDPIKEDVYRNHSIFLADINQERGINESYKKNLMALKKFVMVKF LNSIVDPVDSEWFGFYRSGQAKETIPLQETSLYTQDRLGLKEMDN AGQLVFLATEGDHLQLSEEWFYAHIPFLGRPRAVPTQGGSGSGSTS SRTLCCGGELVDTLQFVCGDRGFLFSRPASRVSRRSRGIVEECCFRSC DLALLETYCATPARSE	
PPT1-11 (BiP- PPT1_2)	MKLSLVAAMLLLLSAARASRALQHLDPPAPLPLVIWHGMGDSCCN PLSMGAIKKMVEKKIPGIYVLSLEIGKTLMEDVENSFFLNVNSQVTT VCQALAKDPKLQQGYNAMGFSQGGQFLRAVAQRCPSPPMINLISV GGQHQQGVFGLPRCPGESSHICDFIRKTLNAGAYSKVVQERLVQAEY WHDPIKEDVYRNHSIFLADINQERGINESYKKNLMALKKFVMVKFL NDSIVDPVDSEWFGFYRSGQAKETIPLQETSLYTQDRLGLKEMDNA GQLVFLATEGDHLQLSEEWFYAHIPFLG	22
PPT1-12 (BiPaa- PPT1_2)	MKLSLVAAMLLLLSAARASRALQHLDPPAPLPLVIWHGMGDSCCN PLSMGAIKKMVEKKIPGIYVLSLEIGKTLMEDVENSFFLNVNSQVTT VCQALAKDPKLQQGYNAMGFSQGGQFLRAVAQRCPSPPMINLISV GGQHQQGVFGLPRCPGESSHICDFIRKTLNAGAYSKVVQERLVQAEY WHDPIKEDVYRNHSIFLADINQERGINESYKKNLMALKKFVMVKFL NDSIVDPVDSEWFGFYRSGQAKETIPLQETSLYTQDRLGLKEMDNA GQLVFLATEGDHLQLSEEWFYAHIPFLG	23
PPT1-13 (BiPaa- PPT1)	MKLSLVAAMLLLLSAARAAADPPAPLPLVIWHGMGDSCCNPLSMG AIKKMVEKKIPGIYVLSLEIGKTLMEDVENSFFLNVNSQVTTVCQAL AKDPKLQQGYNAMGFSQGGQFLRAVAQRCPSPPMINLISVGGQH GVFGLPRCPGESSHICDFIRKTLNAGAYSKVVQERLVQAEYWHDPI KEDVYRNHSIFLADINQERGINESYKKNLMALKKFVMVKFLNDSIV DPVDSEWFGFYRSGQAKETIPLQETSLYTQDRLGLKEMDNAGQLV FLATEGDHLQLSEEWFYAHIPFLG	24
PPT1-14 (BiP1- vIGF2- PPT1)	MKLSLVAAMLLLLSLVAAMLLLLSAARASRTLCCGGELVDTLQFVC GDRGFLFSRPASRVSRRSRGIVEECCFRSCDLALLETYCATPARSEG GGGSGGGGSRPRAVPTQDPPAPLPLVIWHGMGDSCCNPLSMGAIKK MVEKKIPGIYVLSLEIGKTLMEDVENSFFLNVNSQVTTVCQALAKDP KLQQGYNAMGFSQGGQFLRAVAQRCPSPPMINLISVGGQHQQGVFG LPRCPGESSHICDFIRKTLNAGAYSKVVQERLVQAEYWHDPIKEDV	25

	YRNHSIFLADINQERGINESYKKNLMALKKFVMVKFLNDSIVDPVD SEWFGFYRSGQAKETIPLQETSLYTQDRLGLKEMDNAGQLVFLATE GDHLQLSEEFYAHIIIPFLG	
PPT1-15 (BiP1aa- vIGF2- PPT1)	MKLSLVAAMLLLLSLVAAMLLLLSAARAAASRTLCCGELVDLTLQF VCGDRGFLFSRPASRVSRRSRGIVECCFRSCDLALLETYCATPARS EGGGGSGGGGSRPRAVPTQDPPAPLPLVIWHGMGDSCCNPLSMGAI KKMVEKKIPGIYVLSLEIGKTLMEDVENSFFLNVNSQVTTVCQALA KDPKLQQGYNAMGFSQGGQFLRAVAQRCPSPPMINLISVGGQHQQG VFGLPRCPGESSHICDFIRKTLNAGAYSKVVQERLVQAEYWHDPIKE DVYRNHSIFLADINQERGINESYKKNLMALKKFVMVKFLNDSIVDP VDSEWFGFYRSGQAKETIPLQETSLYTQDRLGLKEMDNAGQLVFLA TEGDHLQLSEEFYAHIIIPFLG	26
PPT1-16 (BiP1aa- PPT1_2)	MKLSLVAAMLLLLSLVAAMLLLLSAARAAASRALQHLDPPAPLPL VIWHGMGDSCCNPLSMGAIKKMVEKKIPGIYVLSLEIGKTLMEDVE NSFFLNVNSQVTTVCQALAKDPKLQQGYNAMGFSQGGQFLRAVAQ RCPSPPMINLISVGGQHQQGVFGLPRCPGESSHICDFIRKTLNAGAYSK VVQERLVQAEYWHDPIKEDVYRNHSIFLADINQERGINESYKKNLM ALKKFVMVKFLNDSIVDPVDSEWFGFYRSGQAKETIPLQETSLYTQ DRLGLKEMDNAGQLVFLATEGDHLQLSEEFYAHIIIPFLG	27
PPT1- 17(wt- PPT1-C6S)	MASPGSLWLLAVALLPWTCASRALQHLDPPAPLPLVIWHGMGDSC CNPLSMGAIKKMVEKKIPGIYVLSLEIGKTLMEDVENSFFLNVNSQV TTVCQALAKDPKLQQGYNAMGFSQGGQFLRAVAQRCPSPPMINLIS VGGQHQQGVFGLPRCPGESSHICDFIRKTLNAGAYSKVVQERLVQAE YWHDPKEDVYRNHSIFLADINQERGINESYKKNLMALKKFVMVKF LNDIVDPVDSEWFGFYRSGQAKETIPLQETSLYTQDRLGLKEMDN AGQLVFLATEGDHLQLSEEFYAHIIIPFLG	28
PPT1-18 (BiP2aa- PPT1)	MKLSLVAAMLLLLWVALLLSAARAAASRALQHLDPPAPLPLVIW HGMGDSCCNPLSMGAIKKMVEKKIPGIYVLSLEIGKTLMEDVENSF FLNVNSQVTTVCQALAKDPKLQQGYNAMGFSQGGQFLRAVAQRC PSPPMINLISVGGQHQQGVFGLPRCPGESSHICDFIRKTLNAGAYSKVV QERLVQAEYWHDPIKEDVYRNHSIFLADINQERGINESYKKNLMAL KKFVMVKFLNDSIVDPVDSEWFGFYRSGQAKETIPLQETSLYTQDR	29

	LGLKEMDNAGQLVFLATEGDHLQLSEEWFYAHIIIPFLG	
PPT1-19 (GaussiaAA -PPT1_2)	MGVKVLFALICIAVAEAAAASRALQHLDPPAPLPLVIWHGMGDSCCN PLSMGAIKKMVEKKIPGIYVLSLEIGKTLMEDVENSFFLNVNSQVTT VCQALAKDPKLQQGYNAMGFSQGGQFLRAVAQRCPSPPMINLISV GGQHQQGVFGLPRCPGESSHICDFIRKTLNAGAYSKVVQERLVQAEY WHDPIKEDVYRNHSIFLADINQERGINESYKKNLMALKKFVMVKFL NDSIVDPVDSEWFGFYRSGQAKETIPLQETSLYTQDRLGLKEMDNA GQLVFLATEGDHLQLSEEWFYAHIIIPFLG	30
PPT1-20 (GaussiaAA -vIGF2- PPT1)	MGVKVLFALICIAVAEAAAASRTLCCGGLVDTLQFVCGDRGFLFSRP ASRVSRRSRGIVEECCFRSCDLALLETYCATPARSEGGGGSGGGGSR PRAVPTQDPPAPLPLVIWHGMGDSCCNPLSMGAIKKMVEKKIPGIY VLSLEIGKTLMEDVENSFFLNVNSQVTTVCQALAKDPKLQQGYNA MGFSQGGQFLRAVAQRCPSPPMINLISVGGQHQQGVFGLPRCPGESS HICDFIRKTLNAGAYSKVVQERLVQAEYWHDPIKEDVYRNHSIFLA DINQERGINESYKKNLMALKKFVMVKFLNDSIVDPVDSEWFGFYRS GQAKETIPLQETSLYTQDRLGLKEMDNAGQLVFLATEGDHLQLSEE WFYAHIIIPFLG	31
PPT1-21 (ppt2ss- PPT1)	MLGLWGQRLPAAWVLLLLPFLPLLLADPPAPLPLVIWHGMGDSC CNPLSMGAIKKMVEKKIPGIYVLSLEIGKTLMEDVENSFFLNVNSQV TTVCQALAKDPKLQQGYNAMGFSQGGQFLRAVAQRCPSPPMINLIS VGGQHQQGVFGLPRCPGESSHICDFIRKTLNAGAYSKVVQERLVQAE YWHDPIKEDVYRNHSIFLADINQERGINESYKKNLMALKKFVMVKF LNSIVDPVDSEWFGFYRSGQAKETIPLQETSLYTQDRLGLKEMDN AGQLVFLATEGDHLQLSEEWFYAHIIIPFLG	32
PPT1-22 (ppt2ss- PPT1_2)	MLGLWGQRLPAAWVLLLLPFLPLLLASRALQHLDPPAPLPLVIWH GMGDSCCNPLSMGAIKKMVEKKIPGIYVLSLEIGKTLMEDVENSFFL NVNSQVTTVCQALAKDPKLQQGYNAMGFSQGGQFLRAVAQRCPS PMINLISVGGQHQQGVFGLPRCPGESSHICDFIRKTLNAGAYSKVVQ RLVQAEYWHDPIKEDVYRNHSIFLADINQERGINESYKKNLMALKK FVMVKFLNDSIVDPVDSEWFGFYRSGQAKETIPLQETSLYTQDRLGL KEMDNAGQLVFLATEGDHLQLSEEWFYAHIIIPFLG	33
PPT1-23	MASPSCLWLLAVALLPWSCAARALGHLDPPAPLPLVIWHGMGDSC	34

(consensusS S-PPT1)	CNPLSMGAIKKMVEKKIPGIYVLSLEIGKTLMEDVENSFFLNVNSQV TTVCQALAKDPKLQQGYNAMGFSQGGQFLRAVAQRCPSPPMINLIS VGGQHQQGVFGLPRCPGESSHICDFIRKTLNAGAYSKVVQERLVQAE YWHDPIKEDVYRNHSIFLADINQERGINESYKKNLMALKKFVMVKF LNDSIVDPVDSEWFGFYRSGQAKETIPLQETSLYTQDRLGLKEMDN AGQLVFLATEGDHLQLSEEWFYAHIPFLG	
PPT1-24 (consensus- PPT1)	MASPSCLWLLAVALLPWSCAARALGHLDPPAPLPLVIWHGMGDSC CNPLSMGAIKKMVEKKIPGIYVLSLEIGKTLMEDVENSFFLNVNSQV TTVCQILAKDPKLQQGYNAMGFSQGGQFLRAVAQRCPSPPMINLIS VGGQHQQGVFGLPRCPGESSHICDFIRKTLNAGAYSKAVQERLVQAE YWHDPIKEDVYRNHSIFLADINQERGVNESYKKNLMALKKFVMVK FLNDSIVDPVDSEWFGFYRSGQAKETIPLQETTLYTQDRLGLKEMD KAGQLVFLATEGDHLQLSEEWFYAHIPFLE	35
PPT1-25 (wt-PPT1 L283C H300C)	MASPGCLWLLAVALLPWTCASRALQHLDPPAPLPLVIWHGMGDSC CNPLSMGAIKKMVEKKIPGIYVLSLEIGKTLMEDVENSFFLNVNSQV TTVCQALAKDPKLQQGYNAMGFSQGGQFLRAVAQRCPSPPMINLIS VGGQHQQGVFGLPRCPGESSHICDFIRKTLNAGAYSKVVQERLVQAE YWHDPIKEDVYRNHSIFLADINQERGINESYKKNLMALKKFVMVKF LNDSIVDPVDSEWFGFYRSGQAKETIPLQETSLYTQDRLGLKEMDN AGQLVFCATEGDHLQLSEEWFYACIIPFLG	36
PPT1-26 (wt-PPT1 G113C L121C)	MASPGCLWLLAVALLPWTCASRALQHLDPPAPLPLVIWHGMGDSC CNPLSMGAIKKMVEKKIPGIYVLSLEIGKTLMEDVENSFFLNVNSQV TTVCQALAKDPKLQQGYNAMCFSQGGQFCRAVAQRCPSPPMINLIS VGGQHQQGVFGLPRCPGESSHICDFIRKTLNAGAYSKVVQERLVQAE YWHDPIKEDVYRNHSIFLADINQERGINESYKKNLMALKKFVMVKF LNDSIVDPVDSEWFGFYRSGQAKETIPLQETSLYTQDRLGLKEMDN AGQLVFLATEGDHLQLSEEWFYAHIPFLG	37
PPT1-27 (wt-PPT1 A171C A183C)	MASPGCLWLLAVALLPWTCASRALQHLDPPAPLPLVIWHGMGDSC CNPLSMGAIKKMVEKKIPGIYVLSLEIGKTLMEDVENSFFLNVNSQV TTVCQALAKDPKLQQGYNAMGFSQGGQFLRAVAQRCPSPPMINLIS VGGQHQQGVFGLPRCPGESSHICDFIRKTLNAGCYSKVVQERLVQCE YWHDPIKEDVYRNHSIFLADINQERGINESYKKNLMALKKFVMVKF	38

	LNDSIVDPVDSEWFGFYRSGQAKETIPLQETSLYTQDRLGLKEMDN AGQLVFLATEGDHLQLSEEWFYAHIPFLG	
PPT1-28 (BiP2aa- PPT1)	MKLSLVAAMLLLLWVALLLLSAARAAASRALQHLDPPAPLPLVIW HGMGDSCCNPLSMGAIKKMVEKKIPGIYVLSLEIGKTLMEDVENSF FLNVNSQVTTVCQALAKDPKLQQGYNAMGFSQGGQFLRAVAQRC PSPPMINLISVGGQHQQGVFGLPRCPGESSHICDFIRKTLNAGAYSKVV QERLVQAEYWHDPKEDVYRNHSIFLADINQERGINESYKKNLMAL KKFVMVKFLNDSIVDPVDSEWFGFYRSGQAKETIPLQETSLYTQDR LGLKEMDNAGQLVFLATEGDHLQLSEEWFYAHIPFLG	39
PPT1-31 (BiP1- vIGF2-PPT1)	MKLSLVAAMLLLLSLVAAMLLLLSAARASRTLCCGGELVDTLQFVC GDRGFLFSRPASRVSRRSRGIVEECCFRSCDLALLETYCATPARSE GGGSGGGGSRPRAVPTQDPPAPLPLVIWHGMGDSCCNPLSMGAIKK MVEKKIPGIYVLSLEIGKTLMEDVENSFFLNVNSQVTTVCQALAKDP KLQQGYNAMGFSQGGQFLRAVAQRCPSPPMINLISVGGQHQQGVFG LPRCPGESSHICDFIRKTLNAGAYSKVVQERLVQAEYWHDPKEDV YRNHSIFLADINQERGINESYKKNLMALKKFVMVKFLNDSIVDPVD SEWFGFYRSGQAKETIPLQETSLYTQDRLGLKEMDNAGQLVFLATE GDHLQLSEEWFYAHIPFLG	40
PPT1-32 (wt-PPT1- vIGF2-32)	MASPGCLWLLAVALLPWTCASRALQHLDPPAPLPLVIWHGMGDSC CNPLSMGAIKKMVEKKIPGIYVLSLEIGKTLMEDVENSFFLNVNSQV TTVCQALAKDPKLQQGYNAMGFSQGGQFLRAVAQRCPSPPMINLIS VGGQHQQGVFGLPRCPGESSHICDFIRKTLNAGAYSKVVQERLVQAE YWHDPKEDVYRNHSIFLADINQERGINESYKKNLMALKKFVMVKF LNDSIVDPVDSEWFGFYRSGQAKETIPLQETSLYTQDRLGLKEMDN AGQLVFLATEGDHLQLSEEWFYAHIPFLGRPRAVPTQGGSGSGSTS SSRTLCCGGELVDTLQFVCGDRGFLFSRGGGSRGILEECCFRECDLA LLETYCATPARSE	41
PPT1-33 (wt-PPT1- vIGF2-8Q)	MASPGCLWLLAVALLPWTCASRALQHLDPPAPLPLVIWHGMGDSC CNPLSMGAIKKMVEKKIPGIYVLSLEIGKTLMEDVENSFFLNVNSQV TTVCQALAKDPKLQQGYNAMGFSQGGQFLRAVAQRCPSPPMINLIS VGGQHQQGVFGLPRCPGESSHICDFIRKTLNAGAYSKVVQERLVQAE YWHDPKEDVYRNHSIFLADINQERGINESYKKNLMALKKFVMVKF	42

	LNDSIVDPVDSEWFGFYRSGQAKETIPLQETSLYTQDRLGLKEMDN AGQLVFLATEGDHLQLSEEWFYAHIIIPFLGRPRAVPTQGGSGSGSTS SSRTLCCGGELVDTLQFVCGDRGFLFSRPASRVSRRSRGIVEECCFRE CDLALLETYCATPARSE	
PPT1-34 (wt-PPT1- vIGF2-8Q)	MASPGCLWLLAVALLPWTCASRALQHLDPPAPLPLVIWHGMGDSC CNPLSMGAIKKMVEKKIPGIYVLSLEIGKTLMEDVENSFFLNVSQV TTVCQALAKDPKLQQGYNAMGFSQGGQFLRAVAQRCPSPPMINLIS VGGQHQG VFGLPRCPGESSHICDFIRKTLNAGAYSKVVQERLVQAE YWHDPIKEDVYRNHSIFLADINQERGINESYKKNLMALKKFFVMVKF LNDSIVDPVDSEWFGFYRSGQAKETIPLQETSLYTQDRLGLKEMDN AGQLVFLATEGDHLQLSEEWFYAHIIIPFLGRPRAVPTQGGSGSGSTS SSRTLCCGGELVDTLQFVCGDRGFLFSRPASRVSRRSRGIVEECCFRE CDLALLETYCATPARSE	43
PPT1-35 (wt-PPT1- vIGF2-8Q)	MASPGCLWLLAVALLPWTCASRALQHLDPPAPLPLVIWHGMGDSC CNPLSMGAIKKMVEKKIPGIYVLSLEIGKTLMEDVENSFFLNVSQV TTVCQALAKDPKLQQGYNAMGFSQGGQFLRAVAQRCPSPPMINLIS VGGQHQG VFGLPRCPGESSHICDFIRKTLNAGAYSKVVQERLVQAE YWHDPIKEDVYRNHSIFLADINQERGINESYKKNLMALKKFFVMVKF LNDSIVDPVDSEWFGFYRSGQAKETIPLQETSLYTQDRLGLKEMDN AGQLVFLATEGDHLQLSEEWFYAHIIIPFLGRPRAVPTQGGSGSGSTS SSRTLCCGGELVDTLQFVCGDRGFLFSRPASRVSRRSRGIVEECCFRE CDLALLETYCATPARSE	44
Human TPP1 Propeptide	SYSPEPDQRRTLPPGWVSLGRADPEEELSLTFALRQQNVERLSELVQ AVSDPSSPQYGKYLTLNVADLVRPSPLTLHTVQKWLLAAGAQKC HSVITQDFLTCWLSIRQAELLPGAEFHHYVGGPTETHVVRSPHPYQ LPQALAPHVDFVGGGLHRFPPTSSLRQRPEPQVTGTVG	45
Human TPP1 Mature Peptide	LHLGVTPSVIRKRYNLTSQDVGSGTSNNSQACAQFLEQYFHDSDLA QFMRLFGGNFAHQASVARVVGQQGRGRAGIEASLDVQYLMSAGA NISTWVYSSPGRHEGQEPFLQWLMLLSNESALPHVHTVSYGDEDS LSSAYIQRVNTELMKAAARGLTLLFASGDSGAGCWSVSGRHQFRPT FPASSPYVTTVGGTSFQEPFLITNEIVDYISGGGFSNVFPRPSYQEEAV TKFLSSSPHLPPSSYFNASGRAYPDVAALSDGYWVVSNRVPIPWVS	46

	GTSASTPVFGGILSLINEHRILSGRPPLGFLNPRLYQQHGAGLFDVTR GCHESCLDEEVEGQGFCSGPGWDPVTGWGTPNFPALLKTLLNP	
pSvelte001- Native TPP1 Signal Peptide – vIGF2 – GS linker – Lyso Cleave – TPP1 propeptide – TPP1 mature peptide	MGLQACLLGLFALILSGKCSRTLCCGELVDTLQFVCGDRGFLFSRPA SRVSRRSRGIVEECCFRSCDLALLETYCATPARSEGGGGSGGGGSRP RAVPTQSYSPEPDQRRTLPPGWVSLGRADPEEEELSLTFALRQQNVER LSELVQAVSDPSSPQYGKYLTLENVADLVRPSPLTLHTVQKWLLAA GAQKCHSVITQDFLTCWLSIRQAELLPLGAEFHHYVGGPTETHVVR SPHPYQLPQALAPHVDFVGGHLHRFPPTSSLRQRPEPQVTGTVGLHLG VTPSVIRKRYNLTSQDVSGTSNNSQACAQFLEQYFHDSDLAQFMR LFGGNFAHQASVARVVGQQGRGRAGIEASLDVQYLMSAGANISTW VYSSPGRHEGQEPFLQWLMLLSNESALPHVHTVSYGDDDEDSLSSAY IQRVNTELMKAAARGLTLLFASGDSGAGCWSVSGRHQFRPTFPASS PYVTTVGGTSFQEPFLITNEIVDYISGGGFSNVFPRPSYQEEAVTKFL SSSPHLPPSSYFNASGRAYPDVAALSDGYWVVSNRVPIPWVSGTSAS TPVFGGILSLINEHRILSGRPPLGFLNPRLYQQHGAGLFDVTRGCHES CLDEEVEGQGFCSGPGWDPVTGWGTPNFPALLKTLLNPG	47
Svelte057 – Native TPP1 Signal Peptide – vIGF2v17 – GS linker – Lyso Cleave – TPP1 propeptide – TPP1 mature peptide	MGLQACLLGLFALILSGKCSRTLCCGELVDTLQFVCGDRGFLFSRPA SRVSRRSRGIVEECCFRECDLALLETYCATPARSEGGGGSGGGGSRP RAVPTQASYSPEPDQRRTLPPGWVSLGRADPEEEELSLTFALRQQNV ERLSELVQAVSDPSSPQYGKYLTLENVADLVRPSPLTLHTVQKWLL AAGAQKCHSVITQDFLTCWLSIRQAELLPLGAEFHHYVGGPTETHV VRSPHPYQLPQALAPHVDFVGGHLHRFPPTSSLRQRPEPQVTGTVGLH LGVTPSVIRKRYNLTSQDVSGTSNNSQACAQFLEQYFHDSDLAQF MRLFGGNFAHQASVARVVGQQGRGRAGIEASLDVQYLMSAGANIS TWVYSSPGRHEGQEPFLQWLMLLSNESALPHVHTVSYGDDDEDSLSS AYIQRVNTELMKAAARGLTLLFASGDSGAGCWSVSGRHQFRPTFPA SSPYVTTVGGTSFQEPFLITNEIVDYISGGGFSNVFPRPSYQEEAVTKF LSSSPHLPPSSYFNASGRAYPDVAALSDGYWVVSNRVPIPWVSGTS ASTPVFGGILSLINEHRILSGRPPLGFLNPRLYQQHGAGLFDVTRGCH ESCLDEEVEGQGFCSGPGWDPVTGWGTPNFPALLKTLLNPG	48
pSvelte059	MGLQACLLGLFALILSGKCSRTLCCGELVDTLQFVCGDRGFLFSRG	49

-Native TPP1 Signal Peptide – vIGF2v22 – GS linker – Lyso Cleave – TPP1 propeptide – TPP1 mature peptide	GGGSRGIVEECCFRSCDLALLETYCATPARSEGGGGSGGGGSRPRA VPTQASYSPEPDQRRTLPPGWVSLGRADPEEELSLTFALRQQNVERL SELVQAVSDPSSPQYGKYLTLENVADLVRPSPLTLHTVQKWLLAAG AQKCHSVITQDFLTCWLSIRQAELLPLGAEFHHYVGGPTETHVVRSP HPYQLPQALAPHVDFVGGGLHRFPPTSSLRQRPEPQVTGTVGLHLGV TPSVIRKRYNLTSQDVGSGTSNNSQACAQFLEQYFHDSDLAQFMRL FGGNFAHQASVARVVGQQGRGRAGIEASLDVQYLMSAGANISTWV YSSPGRHEGQEPFLQWLMLLSNESALPHVHTVSYGDDDEDSLSSAYI QRVNTELMKAAARGLTLLFASGDSGAGCWSVSGRHQFRPTFPASSP YVTTVGGTSFQEPFLITNEIVDYISGGGFSNVFPRPSYQEEAVTKFLS SSPHLPPSSYFNASGRAYPDVAALSDGYWVVSNRVPIPWVSGTSAS TPVFGGILSLINEHRILSGRPPLGFLNPRLYQQHGAGLFDVTRGCHES CLDEEVEGQGFCSGPGWDPVTGWGTPNFPALLKTLLNPG	
pSvelte060 - Native TPP1 Signal Peptide – vIGF2v24– GS linker – Lyso Cleave – TPP1 propeptide – TPP1 mature peptide	MGLQACLLGLFALILSGKCSRTLCCGELVDTLQFVCGDRGFLFSRG GGGSRGILEECCFRSCDLALLETYCATPARSEGGGGSGGGGSRPRA VPTQASYSPEPDQRRTLPPGWVSLGRADPEEELSLTFALRQQNVERL SELVQAVSDPSSPQYGKYLTLENVADLVRPSPLTLHTVQKWLLAAG AQKCHSVITQDFLTCWLSIRQAELLPLGAEFHHYVGGPTETHVVRSP HPYQLPQALAPHVDFVGGGLHRFPPTSSLRQRPEPQVTGTVGLHLGV TPSVIRKRYNLTSQDVGSGTSNNSQACAQFLEQYFHDSDLAQFMRL FGGNFAHQASVARVVGQQGRGRAGIEASLDVQYLMSAGANISTWV YSSPGRHEGQEPFLQWLMLLSNESALPHVHTVSYGDDDEDSLSSAYI QRVNTELMKAAARGLTLLFASGDSGAGCWSVSGRHQFRPTFPASSP YVTTVGGTSFQEPFLITNEIVDYISGGGFSNVFPRPSYQEEAVTKFLS SSPHLPPSSYFNASGRAYPDVAALSDGYWVVSNRVPIPWVSGTSAS TPVFGGILSLINEHRILSGRPPLGFLNPRLYQQHGAGLFDVTRGCHES CLDEEVEGQGFCSGPGWDPVTGWGTPNFPALLKTLLNPG	50
pSvelte061- Native TPP1 Signal Peptide – vIGF2v30 –	MGLQACLLGLFALILSGKCSRTLCCGELVDVLQFVCGRRGFLFSRP ASRVSRRSRGIVEECCFRDCDLALLETYCATPARSEGGGGSGGGGS RPRAVPTQASYSPEPDQRRTLPPGWVSLGRADPEEELSLTFALRQQN VERLSELVQAVSDPSSPQYGKYLTLENVADLVRPSPLTLHTVQKWL LAAGAQAQKCHSVITQDFLTCWLSIRQAELLPLGAEFHHYVGGPTETH	51

GS linker – Lyso Cleave – TPP1 propeptide – TPP1 mature peptide	VVRSPHPYQLPQALAPHVDFVGGGLHRFPPTSSLRQRPEPQVTGTVGL HLGVTPSVIRKRYNLTSQDVGSGTSNNSQACAQFLEQYFHDSDLAQ FMRLFGGNFAHQASVARVVGQQGRGRAGIEASLDVQYLMSAGANI STWVYSSPGRHEGQEPFLQWLMLLSNESALPHVHTVSYGDDSDLS SAYIQRVNTELMKAAARGLTLLFASGDSGAGCWSVSGRHQFRPTFP ASSPYVTTVGGTSFQEPFLITNEIVDYISGGGFSNVFPRPSYQEEAVT KFLSSPHLPPSSYFNASGRAYPDVAALSDGYWVVSNRVPIPWVSG TSASTPVFGGILSLINEHRILSGRPPLGFLNPRLYQQHGAGLFDVTRG CHESCLDEEVEGQGFCSGPGWDPVTGWGTPNFPALLKTLLNPG*	
pSvelte062- Native TPP1 Signal Peptide – vIGF2v31 – GS linker – Lyso Cleave – TPP1 propeptide – TPP1 mature peptide	MGLQACLLGLFALILSGKCSRTLCCGELVDTLQFVCGDRGFLFSRG GGGSRGILEECCFRDCDLALLETYCATPARSEG GGGSGGGSGRPR VPTQASYSPEPDQRRTLPPGWVSLGRADPEEELSLTFALRQQNVERL SELVQAVSDPSSPYGKYLTLENVADLVRPSPLTLHTVQKWLLAAG AQKCHSVITQDFLTCWLSIRQAELLPLGAEFHHYVGGPTETHVVRSP HPYQLPQALAPHVDFVGGGLHRFPPTSSLRQRPEPQVTGTVGLHLGV TPSVIRKRYNLTSQDVGSGTSNNSQACAQFLEQYFHDSDLAQFMRL FGGNFAHQASVARVVGQQGRGRAGIEASLDVQYLMSAGANISTWV YSSPGRHEGQEPFLQWLMLLSNESALPHVHTVSYGDDSDLS SAYI QRVNTELMKAAARGLTLLFASGDSGAGCWSVSGRHQFRPTFPASSP YVTTVGGTSFQEPFLITNEIVDYISGGGFSNVFPRPSYQEEAVTKFLS SSPHLPPSSYFNASGRAYPDVAALSDGYWVVSNRVPIPWVSGTSAS TPVFGGILSLINEHRILSGRPPLGFLNPRLYQQHGAGLFDVTRGCHES CLDEEVEGQGFCSGPGWDPVTGWGTPNFPALLKTLLNPG	52
pSvelte063- Native TPP1 Signal Peptide – vIGF2v32 – GS linker – Lyso Cleave – TPP1 propeptide –	MGLQACLLGLFALILSGKCSRTLCCGELVDTLQFVCGDRGFLFSRG GGGSRGILEECCFRECDLALLETYCATPARSEG GGGSGGGSGRPR VPTQASYSPEPDQRRTLPPGWVSLGRADPEEELSLTFALRQQNVERL SELVQAVSDPSSPYGKYLTLENVADLVRPSPLTLHTVQKWLLAAG AQKCHSVITQDFLTCWLSIRQAELLPLGAEFHHYVGGPTETHVVRSP HPYQLPQALAPHVDFVGGGLHRFPPTSSLRQRPEPQVTGTVGLHLGV TPSVIRKRYNLTSQDVGSGTSNNSQACAQFLEQYFHDSDLAQFMRL FGGNFAHQASVARVVGQQGRGRAGIEASLDVQYLMSAGANISTWV YSSPGRHEGQEPFLQWLMLLSNESALPHVHTVSYGDDSDLS SAYI	53

TPP1 mature peptide	QRVNTELMKAAARGLTLLFASGDSGAGCWSVSGRHQFRPTFPASSP YVTTVGGTSFQEPFLITNEIVDYISGGGFSNVFPRPSYQEEAVTKFLS SSPHLPSSYFNASGRAYPDVAALSDGYWVVSNRVPIPWVSGTSAS TPVFGGILSLINEHRILSGRPPLGFLNPRLYQQHGAGLFDVTRGCHES CLDEEVEGQGFCSGPGWDPVTGWGTPNFPALLKTLLNPG	
Wt- NAGLU	MEAVAVAAAVGVLLLAGAGGAAGDEAREAAAVRALVARLLGPGP AADFSVSVERALAAKPGLDTYSLGGGAARVRVRGSTGVAAAAGL HRYLRDFCGCHVAWSGSQLRLPRPLPAVPGELTEATPNRYRYYYQN VCTQSYSFVWWDWARWEREIDWMALNGINLALAWSGQEAIWQR VYLALGLTQAEINEFFTGPAFLAWGRMGNLHTWDGPLPPSWHIKQL YLQHRVLDQMRSFGMTPVLPFAFAGHVPEAVTRVFPQVNVTKMGS WGHFNCSYSCSFLLAPEDPIFPIIGSLFLRELIKEFGTDHIYGADTFNE MQPPSSEPSYLAATTAVYEAMTAVDTEAVWLLQGWLWFQHQPQF WGPAQIRAVLGAVPRGRLLVLDLFAESQPVYTRTASFQGQPFIWCM LHNFGGNHGLFGALEAVNGGPEAARLFPNSTMVGTMapeGISQN EVVYSLMAELGWRKDPVPDLAAWVTSFAARRYGVSHPDAGAAWR LLRSVYNCSGEACRGHNRSPVRRPSLQMNTSIWYNRSDVFEAWR LLTSAPSLATSPAFRYDLLDLTRQAVQELVSLYYEEARSAYLSKEL ASLLRAGGVLAYELLPALDEVLASDSRFLGSLWLEQARAAAVSEAE ADFYEQNSRYQLTLWGPEGNILDYANKQLAGLVANYYTTPRWRLFL EALVDSVAQGIPFQQHQFDKNVFQLEQAFVLSKQRYPSQPRGDTV LAKKIFLKYYPVWVAGSW	54
Wt NAGLU- HPC4	MEAVAVAAAVGVLLLAGAGGAAGDEAREAAAVRALVARLLGPGP AADFSVSVERALAAKPGLDTYSLGGGAARVRVRGSTGVAAAAGL HRYLRDFCGCHVAWSGSQLRLPRPLPAVPGELTEATPNRYRYYYQN VCTQSYSFVWWDWARWEREIDWMALNGINLALAWSGQEAIWQR VYLALGLTQAEINEFFTGPAFLAWGRMGNLHTWDGPLPPSWHIKQL YLQHRVLDQMRSFGMTPVLPFAFAGHVPEAVTRVFPQVNVTKMGS WGHFNCSYSCSFLLAPEDPIFPIIGSLFLRELIKEFGTDHIYGADTFNE MQPPSSEPSYLAATTAVYEAMTAVDTEAVWLLQGWLWFQHQPQF WGPAQIRAVLGAVPRGRLLVLDLFAESQPVYTRTASFQGQPFIWCM LHNFGGNHGLFGALEAVNGGPEAARLFPNSTMVGTMapeGISQN	55

	EVVYSLMAELGWRKDPVPDLAAWVTSFAARRYGVSHPDAGAAWR LLLRSVYNCSGEACRGHNRSPLVRRPSLQMNTSIWYNRSDVFEAWR LLLTSAPSLATSPAIFYDLDDLTRQAVQELVSLYYEEARSAYLSKEL ASLLRAGGVLAYELLPALDEVLASDSRFLGSLWLEQARAAAVSEAE ADFYEQNSRYQLTLWGPEGNILDYANKQLAGLVANYYTTPRWRLFL EALVDSVAQGIPFQQHQFDKNVFQLEQAFVLSKQRYPSQPRGDTV LAKKIFLKYYPWWAGSWGLEVLFGQPEDQVDPRLIDGK	
vIGF2- NAGLU- HPC4	MEAVAVAAAVGVLLLAGAGGAAGDASRTLGGELVDTLQFVCGD RGFLFSRPASRVSRRSRGIVEECCFRSCDLALLETYCATPARSEGGG GSGGGGSRPRAVPTQAEAREAAAVRALVARLLGPGPAADFSVSVE RALAAKPGLDTYSLGGGGAARVRVRGSTGVAAAAGLHRYLRDFC GCHVAWSGSQLRLPRPLPAVPGELTEATPNRYRYYYQNVCTQSYFV WWDWARWEREIDWMALNGINLALAWSGQEAIWQRVYLALGLTQ AEINEFFTGAFLAWGRMGNLHTWDGPLPPSWHIKQLYLQHRVLD QMRSGMTPVLPAPAGHVPEAVTRVFPQVNVTKMGSWGWHFNCSYS CSFLLAPEDPIFPIIGSLFLRELIFEGTDHIYGADTFNEMQPPSSEPSY LAAATTAVYEAMTAVDTEAVWLLQGWLFQHQPFQWGPQIRAVL GAVPRGRLVLDLFAESQPVYTRTASFQGGQPFWCMLHNFNGNHGL FGALEAVNGGPEAARLFPNSTMVGTGMAPEGISQNEVVYSLMAEL GWRKDPVPDLAAWVTSFAARRYGVSHPDAGAAWRLLRSVYNCS GEACRGHNRSPLVRRPSLQMNTSIWYNRSDVFEAWRLLTSAPSLA TSPAIFYDLDDLTRQAVQELVSLYYEEARSAYLSKELASLLRAGGV LAYELLPALDEVLASDSRFLGSLWLEQARAAAVSEAEADFYEQNSR YQLTLWGPEGNILDYANKQLAGLVANYYTTPRWRLFLALVDSVAQ GIPFQQHQFDKNVFQLEQAFVLSKQRYPSQPRGDTVDLAKKIFLKY YPRWWAGSWGLEVLFGQPEDQVDPRLIDGK	56
vIGF2-17- NAGLU	MEAVAVAAAVGVLLLAGAGGAAGDASRTLGGELVDTLQFVCGD RGFLFSRPASRVSRRSRGIVEECCFRECDLALLETYCATPARSEGGG GSGGGGSRPRAVPTQAEAREAAAVRALVARLLGPGPAADFSVSVE RALAAKPGLDTYSLGGGGAARVRVRGSTGVAAAAGLHRYLRDFC GCHVAWSGSQLRLPRPLPAVPGELTEATPNRYRYYYQNVCTQSYFV WWDWARWEREIDWMALNGINLALAWSGQEAIWQRVYLALGLTQ	57

	AEINEFFTGAFLAWGRMGNLHTWDGPLPPSWHIKQLYLQHRVLD QMRSFGMTPVLPAPAGHVPEAVTRVFPQVNVTKMGSWGHNCSYS CSFLLAPEDPIFPIIGSLFLRELIKEFGTDHIYGADTFNEMQPPSSEPSY LAAATTAVYEAMTAVDTEAVWLLQGWLQFQHPQFWGPAQIRAVL GAVPRGRLLVLDLFAESQPVTYRTASFQGPFIWCMLHNFGGNHGL FGALEAVNGGPEAARLFPNSTMVGTGMAPEGISQNEVVYSLMAEL GWRKDPVPDLAAWVTSFAARRYGVSHPDAGAAWRLLLRSVYNCS GEACRGHNRSLVRRPSLQMNTSIWYNRSDVFEAWRLLLTSAPSLA TSPAFRYDLLDLTRQAVQELVSLYEEARSAYLSKELASLLRAGGV LAYELLPALDEVLASDSRFLGSLWLEQARAAAVSEAEADFYEQNSR YQLTLWGPEGNILDYANKQLAGLVANYTTPRWRLFLEALVDSVAQ GIPFQQHQFDKNVFQLEQAFVLSKQRYPSQPRGDTVDLAKKIFLKY YPRWVAGSWGLEVLFGQPEDQVDPRLIDGK	
vIGF2-31- NAGLU- HPC4	MEAVAVAAAVGVLLLAGAGGAAGDASRTLCCGGELVDTLQFVCGD RGFLFSRGGGSRGILEECCFRDCDLALLETCATPARSEGSGGSGG GGSRPRAVPTQAEAREAAAVRALVARLLGPGPAADFSVSVERALA AKPGLDITYSLGGGGAARVRVRGSTGVAAAAGLHRYLRDFCGCHV AWSGSQLRLPRPLPAVPGELEATPNRYRYYYQNVCTQSYSFVWWD WARWEREIDWMALNGINLALAWSGQEAIWQRVYLALGLTQAEINE FFTGAFLAWGRMGNLHTWDGPLPPSWHIKQLYLQHRVLDQMRSF GMTPVLPAPAGHVPEAVTRVFPQVNVTKMGSWGHNCSYSFLL APEDPIFPIIGSLFLRELIKEFGTDHIYGADTFNEMQPPSSEPSYAAA TTAVYEAMTAVDTEAVWLLQGWLQFQHPQFWGPAQIRAVLGAVP RGRLLVLDLFAESQPVTYRTASFQGPFIWCMLHNFGGNHGLFGAL EAVNGGPEAARLFPNSTMVGTGMAPEGISQNEVVYSLMAELGWRK DPVPDLAAWVTSFAARRYGVSHPDAGAAWRLLLRSVYNCSGEACR GHNRSLVRRPSLQMNTSIWYNRSDVFEAWRLLLTSAPSLATSPAFR YDLLDLTRQAVQELVSLYEEARSAYLSKELASLLRAGGVLAYELL PALDEVLASDSRFLGSLWLEQARAAAVSEAEADFYEQNSRYQLTL WGPEGNILDYANKQLAGLVANYTTPRWRLFLEALVDSVAQGIPFQ QHQFDKNVFQLEQAFVLSKQRYPSQPRGDTVDLAKKIFLKYPRW VAGSWGLEVLFGQPEDQVDPRLIDGK	58

vIGF2-32-NAGLU	MEAVAVAAAVGVLLLAGAGGAAGDASRTLCTGGELVDTLQFVCGD RGFLFSRGGGSRGILEECCFRECDLALLETYCATPARSEGGGGSGG GGSRPRAVPTQAEAREAAAVRALVARLLGPGPAADFSVVERALA AKPGLDTYSLGGGGAARVRVRGSTGVAAAAGLHRYLRDFCGCHV AWSGSQLRLPRPLPAVPGELTEATPNRYRYQNVCTQSYSFVWWD WARWEREIDWMALNGINLALAWSGQEIWQRVYLALGLTQAEINE FFTGPAFLAWGRMGNLHTWDGPLPPSWHIKQLYLQHRVLDQMRSF GMTPVLPFAFAGHVPEAVTRVFPQVNVTKMGSWGHFNCSYSCSFL APEDPIFIIGSLFLRELIKEFGTDHIYGADTFNEMQPPSSEPSYLA TTAVYEAMTAVDTEAVWLLQGWLQFQHPQFWGPAQIRAVLGAVP RGRLLVLDLFAESQPVYTRTASFQGGQPFWCMLHNFNGHGLFGAL EAVNGGPEAARLFPNSTMVGTGMAPEGISQNEVVYSLMAELGWRK DPVPDLAAWVTSFAARRYGVSHPDAGAAWRLLLRSVYNCSGEACR GHNRSPLVRRPSLQMNTSIWYNRSDVFCAWRLLLTSAPSLATSPA FRYDLLDLTRQAVQELVSLYEEARSAYLSKELASLLRAGGVLAYELL PALDEVLASDSRFLGSLWLEQARAAAVSEAEADFYEQNSRYQLTL WGPEGNILDYANKQLAGLVANYYTTPRWRLFLEALVDSVAQGIPFQ QHQFDKNVFQLEQAFVLSKQRYPSQPRGDTVDLAKKIFLKYYP RWVAGSWGLEVLFQGPEDQVDPRLIDGK	59
PPT1-101	MKLSLVAAML LLLWVALLLSAARAAASRTLCTGGELVDTLQFVCG DRGFLFSRGGGSRGILEECCFRDCDLALLETYCATPARSEGGGGSG GGGSRPRAVPTQDPPAPLPLVIWHGMGDSCCNPLSMGAIKKMVEK KIPGIYVLSLEIGKTLMEDVENSFFLNVNSQVTTVCQALAKDPKLQQ GYNAMGFSQGGQFLRAVAQRCPSPPMINLISVGGQHQQGVFGLPRCP GESSHICDFIRKTLNAGAYSKVVQERLVQAEYWHDPKEDVYRNHS IFLADINQERGINESYKKNLMALKKFVMVKFLNDSIVDPVDSEWFG FYRSGQAKETIPLQETSLYTQDRLGLKEMDNAGQLVFLATEGDHLQ LSEEWFYAHIIPFLG	60
PPT1-104	MASPGSLWLLAVALLPWTCASRALQHLDPPAPLPLVIWHGMGDSC CNPLSMGAIKKMVEKKIPGIYVLSLEIGKTLMEDVENSFFLNVNSQV TTVCQALAKDPKLQQGYNAMGFSQGGQFLRAVAQRCPSPPMINLIS VGGQHQQGVFGLPRCPGESSHICDFIRKTLNAGAYSKVVQERLVQAE	61

	YWHDPIKEDVYRNHSIFLADINQERGINESYKKNLMALKKFVMVKF LNDIVDPVDSEWFGFYRSGQAKETIPLQETSLYTQDRLGLKEMDN AGQLVFLATEGDHLQLSEEWFYAHIPFLGRPRAVPTQGGSGSGSTS SSRTLCCGELVDTLQFVCGDRGFLFSRGGGSGRGILEECCFRECDLA LLETYCATPARSE	
PPT1-112	MASPGSLWLLAVALLPWTCASRALQHLAASRTLCCGELVDTLQFV CGDRGFLFSRGGGSGRGILEECCFRDCDLALLETYCATPARSEGGG GSGGGGSRPRAVPTQDPPAPLPLVIWHGMGDSCCNPLSMGAIKKM VEKKIPGIYVLSLEIGKTLMEDVENSFFLVNSQVTTVCQALAKDPK LQQGYNAMGFSQGGQFLRAVAQRCPSPMINLISVGGQHQQGVFGLP RCPGESSHICDFIRKTLNAGAYSKVVQERLVQAEYWHDPIKEDVYR NHSIFLADINQERGINESYKKNLMALKKFVMVKFLNDIVDPVDSE WFGFYRSGQAKETIPLQETSLYTQDRLGLKEMDNAGQLVFLATEGD HLQLSEEWFYAHIPFLG	62
PPT1-114	MASPGSLWLLAVALLPWTCASRALQHLAASRTLCCGELVDTLQFV CGDRGFLFSRGGGSGRGILEECCFRECDLALLETYCATPARSEGGGG SGGGGSRPRAVPTQDPPAPLPLVIWHGMGDSCCNPLSMGAIKKMVE KKIPGIYVLSLEIGKTLMEDVENSFFLVNSQVTTVCQALAKDPKLQ QQGYNAMGFSQGGQFLRAVAQRCPSPMINLISVGGQHQQGVFGLPRC PGESSHICDFIRKTLNAGAYSKVVQERLVQAEYWHDPIKEDVYRNH SIFLADINQERGINESYKKNLMALKKFVMVKFLNDIVDPVDSEWF GFYRSGQAKETIPLQETSLYTQDRLGLKEMDNAGQLVFLATEGDHL QLSEEWFYAHIPFLG	63
PPT1-115	MASPGSLWLLAVALLPWTCASRALQHLAADPPAPLPLVIWHGMGD SCCNPLSMGAIKKMVEKKIPGIYVLSLEIGKTLMEDVENSFFLVNS QVTTVCQALAKDPKLQQGYNAMGFSQGGQFLRAVAQRCPSPMIN LISVGGQHQQGVFGLPRCPGESSHICDFIRKTLNAGAYSKVVQERLVQ AEYWHDPIKEDVYRNHSIFLADINQERGINESYKKNLMALKKFVMV KFLNDIVDPVDSEWFGFYRSGQAKETIPLQETSLYTQDRLGLKEM DNAGQLVFLATEGDHLQLSEEWFYAHIPFLGRPRAVPTQGGSGSGS TSSRTLCCGELVDTLQFVCGDRGFLFSRGGGSGRGILEECCFRECD LALLETYCATPARSE	64

PPT1-116	MASPGSLWLLAVALLPWTCASRALQH LAADPPAPLPLVIWHGMGD SCCNPLSMGAIKKMVEKKIPGIYVLSLEIGKTLMEDVENSFFLNVNS QVTTVCQALAKDPKLQQGYNAMGFSQGGQFLRAVAQRCPSPPMIN LISVGGQHQQGVFGLPRCPGESSHICDFIRKTLNAGAYSKVVQERLVQ AEYWHDPIKEDVYRNHSIFLADINQERGINESYKKNLMALKKFVMV KFLNDSIVDPVDSEWFGFYRSGQAKETIPLQETSLYTQDRLGLKEM DNAGQLVFLATEGDHLQLSEEWFYAHIPFLGRPRAVPTQGGGGSG SGGGSSRTLCCGELVDTLQFVCGDRGFLFSRGGGSGRGILEECCFR ECDLALLETYCATPARSE	65
PPT1-117	MASPGSLWLLAVALLPWTCASRALQHLDPPAPLPLVIWHGMGDSC CNPLSMGAIKKMVEKKIPGIYVLSLEIGKTLMEDVENSFFLNVNSQV TTVCQALAKDPKLQQGYNAMGFSQGGQFLRAVAQRCPSPPMINLIS VGGQHQQGVFGLPRCPGESSHICDFIRKTLNAGAYSKVVQERLVQAE YWHDPIKEDVYRNHSIFLADINQERGINESYKKNLMALKKFVMVKF LNSIVDPVDSEWFGFYRSGQAKETIPLQETSLYTQDRLGLKEMDN AGQLVFLATEGDHLQLSEEWFYAHIPFLGRPRAVPTQGGGGSGSG GGGSSRTLCCGELVDTLQFVCGDRGFLFSRGGGSGRGILEECCFREC DLALLETYCATPARSE	66
PPT1-118	MASPGSLWLLAVALLPWTCASRALQH LAADPPAPLPLVIWHGMGD SCCNPLSMGAIKKMVEKKIPGIYVLSLEIGKTLMEDVENSFFLNVNS QVTTVCQALAKDPKLQQGYNAMGFSQGGQFLRAVAQRCPSPPMIN LISVGGQHQQGVFGLPRCPGESSHICDFIRKTLNAGAYSKVVQERLVQ AEYWHDPIKEDVYRNHSIFLADINQERGINESYKKNLMALKKFVMV KFLNDSIVDPVDSEWFGFYRSGQAKETIPLQETSLYTQDRLGLKEM DNAGQLVFLATEGDHLQLSEEWFYAHIPFLGRPRAVPTQGGGGSG GGGSSRTLCCGELVDTLQFVCGDRGFLFSRGGGSGRGILEECCFREC DLALLETYCATPARSE	67

[00146] Components of fusion proteins provided herein are further described below.
Peptides that bind CI-MPR (e.g., vIGF2 Peptides)

[00147] Provided herein are peptides that bind CI-MPR. Fusion proteins comprising
5 such peptides and a therapeutic protein, when expressed from a gene therapy vector, target the

therapeutic protein to the cells where it is needed, increase cellular uptake by such cells and target the therapeutic protein to a subcellular location (e.g., a lysosome). In some embodiments, the peptide is fused to the N-terminus of the therapeutic peptide. In some embodiments, the peptide is fused to the C-terminus of the therapeutic protein. In some

5 embodiments, the peptide is a vIGF2 peptide. Some vIGF2 peptides maintain high affinity binding to CI-MPR while their affinity for IGF1 receptor, insulin receptor, and IGF binding proteins (IGFBP) is decreased or eliminated. Some vIGF2 peptides increase affinity of binding to CI-MPR. Thus, some variant IGF2 peptides are substantially more selective and have reduced safety risks compared to wt IGF2. vIGF2 peptides herein include those having the

10 amino acid sequence of SEQ ID NO: 31, 120 and 121. Variant IGF2 peptides further include those with variant amino acids at positions 6, 26, 27, 31-38, 43, 48, 49, 50, 54, 55, or 65 compared to wt IGF2 (SEQ ID NO: 68). In some embodiments, the vIGF2 peptide has a sequence having one or more substitutions from the group consisting of E6R, F26S, Y27L, V43L, F48T, R49S, S50E, S50I, A54R, L55R, and K65R. In some embodiments, the vIGF2

15 peptide has a sequence having a substitution of E6R. In some embodiments, the vIGF2 peptide has a sequence having a substitution of F26S. In some embodiments, the vIGF2 peptide has a sequence having a substitution of Y27L. In some embodiments, the vIGF2 peptide has a sequence having a substitution of V43L. In some embodiments, the vIGF2 peptide has a sequence having a substitution of F48T. In some embodiments, the vIGF2 peptide has a

20 sequence having a substitution of R49S. In some embodiments, the vIGF2 peptide has a sequence having a substitution of S50I. In some embodiments, the vIGF2 peptide has a sequence having a substitution of S50E. In some embodiments, the vIGF2 peptide having a sequence having a substitution of S50E has increased binding to the CI-MPR. In some

25 embodiments, the vIGF2 peptide has a sequence having a substitution of A54R. In some embodiments, the vIGF2 peptide has a sequence having a substitution of L55R. In some embodiments, the vIGF2 peptide has a sequence having a substitution of K65R. In some

30 embodiments, the vIGF2 peptide has a sequence having a substitution of E6R, F26S, Y27L, V43L, F48T, R49S, S50I, A54R, and L55R. In some embodiments, the vIGF2 peptide has an N-terminal deletion. In some embodiments, the vIGF2 peptide has an N-terminal deletion of one amino acid. In some embodiments, the vIGF2 peptide has an N-terminal deletion of two amino acids. In some embodiments, the vIGF2 peptide has an N-terminal deletion of three amino acids. In some embodiments, the vIGF2 peptide has an N-terminal deletion of four

amino acids. In some embodiments, the vIGF2 peptide has an N-terminal deletion of four amino acids and a substitution of E6R, Y27L, and K65R. In some embodiments, the vIGF2 peptide has an N-terminal deletion of four amino acids and a substitution of E6R and Y27L. In some embodiments, the vIGF2 peptide has an N-terminal deletion of five amino acids. In some embodiments, the vIGF2 peptide has an N-terminal deletion of six amino acids. In some embodiments, the vIGF2 peptide has an N-terminal deletion of seven amino acids. In some embodiments, the vIGF2 peptide has an N-terminal deletion of seven amino acids and a substitution of Y27L and K65R. In some embodiments, B_{max} for CIMPR binding by SEQ ID NO:83 is enhanced compared to SEQ ID NO:80.

10

Table 3: IGF2 Amino Acid Sequences (variant residues are underlined)		
Peptide	Sequence	SEQ ID NO
Wildtype	AYRPSETLCGGELVDTLQFVCGDRGFYFSRP ASRVSRRSRGIVEECCFRSCDLALLETYCATP AKSE	68
F26S	AYRPSETLCGGELVDTLQFVCGDRG <u>S</u> YFSRP ASRVSRRSRGIVEECCFRSCDLALLETYCATP AKSE	69
Y27L	AYRPSETLCGGELVDTLQFVCGDRG <u>F</u> LFSRPA SRVSRRSRGIVEECCFRSCDLALLETYCATPA KSE	70
V43L	AYRPSETLCGGELVDTLQFVCGDRGFYFSRP ASRVSRRSRGIL <u>E</u> ECCFRSCDLALLETYCATP AKSE	71
F48T	AYRPSETLCGGELVDTLQFVCGDRGFYFSRP ASRVSRRSRGIVEECC <u>T</u> RSCDLALLETYCATP AKSE	72
R49S	AYRPSETLCGGELVDTLQFVCGDRGFYFSRP ASRVSRRSRGIVEECCF <u>S</u> SCDLALLETYCATP AKSE	73
S50I	AYRPSETLCGGELVDTLQFVCGDRGFYFSRP	74

	ASRVSRRSRGIVEECCFR <u>I</u> CDLALLETYCATPA KSE	
A54R	AYRPSETLCGGELVDTLQFVCGDRGFYFSRP ASRVSRRSRGIVEECCFRSCDL <u>R</u> LLETYCATP AKSE	75
L55R	AYRPSETLCGGELVDTLQFVCGDRGFYFSRP ASRVSRRSRGIVEECCFRSCDLA <u>R</u> LLETYCATP AKSE	76
F26S, Y27L, V43L, F48T, R49S, S50I, A54R, L55R	AYRPSETLCGGELVDTLQFVCGDRG <u>S</u> LFSRPA SRVSRRSRGIL <u>E</u> ECCT <u>S</u> ICDL <u>R</u> RLETYCATPAK SE	77
Δ1-6, Y27L, K65R	TLCGGELVDTLQFVCGDRG <u>F</u> LFSRPASRVSR RGIVEECCFRSCDLALLETYCATP <u>A</u> RSE	78
Δ1-7, Y27L, K65R	LCGGELVDTLQFVCGDRG <u>F</u> LFSRPASRVSR RGIVEECCFRSCDLALLETYCATP <u>A</u> RSE	79
Δ1-4, E6R, Y27L, K65R	<u>S</u> R <u>T</u> LCGGELVDTLQFVCGDRG <u>F</u> LFSRPASRV RRSRGIVEECCFRSCDLALLETYCATP <u>A</u> RSE	80
Δ1-4, E6R, Y27L	<u>S</u> R <u>T</u> LCGGELVDTLQFVCGDRG <u>F</u> LFSRPASRV RRSRGIVEECCFRSCDLALLETYCATPAKSE	81
E6R	AYRPS <u>R</u> TLCGGELVDTLQFVCGDRGFYFSRP ASRVSRRSRGIVEECCFRSCDLALLETYCATP AKSE	82
Δ1-4, E6R, Y27L, S50E, K65R	SRTLCGGELVDTLQFVCGDRGFLFSRPASRV RRSRGIVEECCFRECDLALLETYCATPARSE	83
Cleavable IGF2 variant-N terminal	GGGGSGGGSRPRAVPTQ	84
Cleavable IGF2 variant-C terminal	YIPAKQGLQGAQMGPGGGGSGGGG	85
Cleavable IGF2 variant-C terminal	RPRAVPTQGGSGSGSTSS	86

Cleavable IGF2 variant-N terminal	SRTLCCGGELVDTLQFVCGDRGFLFSRPASRVSR RRSRGIVEECCFRSCDLALLETYCATPARSEG GGGSGGGGSRPRAVPTQ	87
Cleavable IGF2 variant-C terminal	YIPAKQGLQGAQMGPGGGGSGGGGSRTLC GGELVDTLQFVCGDRGFLFSRPASRVSRRSRG IVEECCFRSCDLALLETYCATPARSE	88
Cleavable IGF2 variant-C terminal	RPRAVPTQGGSGSGSTSSSRTLCCGGELVDTLQ FVCGDRGFLFSRPASRVSRRSRGIVEECCFRE CDLALLETYCATPARSE	89
vIGF2-1 (vIGF2_1_NGGW GMG)	SRTLCCGGELVDTN Q FVCGDRGFLFSRGASRV SRGSRGIVEWCCFRGCDLALLETYCATPMRG E	90
vIGF2-2 (vIGF2_2_GGWG MG)	SRTLCCGGELVDTLQFVCGDRGFLFSRGASRV SRGSRGIVEWCCFRGCDLALLETYCATPMRG E	91
vIGF2-3 (vIGF2_3_NGGG MG)	SRTLCCGGELVDTN Q FVCGDRGFLFSRGASRV SRGSRGIVEECCFRGCDLALLETYCATPMRG E	92
vIGF2-4 (vIGF2_4_Δ32-41, 53aa)	SRTLCCGGELVDTLQFVCGDRGFLFSRPIVEEC CFRSCDLALLETYCATPARSE	93
vIGF2-5 (vIGF2 Δ 30-39, 53aa)	SRTLCCGGELVDTLQFVCGDRGFLFSRGIVEEC CFRSCDLALLETYCATPARSE	94
vIGF2-6 (vIGF2 Δ 33-40, 55aa)	SRTLCCGGELVDTLQFVCGDRGFLFSRPAGIVE ECCFRSCDLALLETYCATPARSE	95
vIGF2-7 (vIGF2 Δ 30- 39/V14D/F28R/V4 3D/F26A)	SRTLCCGGELDDTLQFVCGDRGALRSRGIDEE CCFRSCDLALLETYCATPARSE	96
vIGF2-8 (vIGF2_8_REE)	SRTLCCGGELVDTLQFVCGRRGELFSRPASRVSR RRSRGIVEECCFRECDLALLETYCATPARSE	97
vIGF2-9	SRTLCCGGELVDTLQFVCGRRGELFSRPAGIVE	98

(vIGF2_9_ Δ 30-39-REE; vIGF2)	ECCFR <u>E</u> CDLALLETYCATPARSE	
vIGF2-10 (vIGF2_1Q; vIGF2 D23R)	SRTLCCGGELVDTLQFVCG <u>R</u> RGFLFSRPASRV RRSRGIVEECCFRSCDLALLETYCATPARSE	99
vIGF2-11 (vIGF2_2Q; vIGF2 F19W)	SRTLCCGGELVDTLQ <u>W</u> VCGDRGFLFSRPASRV SRRSRGIVEECCFRSCDLALLETYCATPARSE	100
vIGF2-12 (vIGF2_3Q; vIGF2 T16W)	SRTLCCGGELVD <u>W</u> LQFVCGDRGFLFSRPASRV SRRSRGIVEECCFRSCDLALLETYCATPARSE	101
vIGF2-13 (vIGF2_4Q; vIGF2 D23K)	SRTLCCGGELVDTLQFVCG <u>K</u> RGFLFSRPASRV RRSRGIVEECCFRSCDLALLETYCATPARSE	102
vIGF2-14 (vIGF2_5Q; vIGF2 T16Y)	SRTLCCGGELVD <u>Y</u> LQFVCGDRGFLFSRPASRV RRSRGIVEECCFRSCDLALLETYCATPARSE	103
vIGF2-15 (vIGF2_6Q; vIGF2 F26E)	SRTLCCGGELVDTLQFVCGDRG <u>E</u> LFSRPASRV RRSRGIVEECCFRSCDLALLETYCATPARSE	104
vIGF2-16 (vIGF2_7Q; vIGF2 T16V)	SRTLCCGGELVD <u>V</u> LQFVCGDRGFLFSRPASRV RRSRGIVEECCFRSCDLALLETYCATPARSE	105
vIGF2-17 (vIGF2_8Q; vIGF2 S50E)	SRTLCCGGELVDTLQFVCGDRGFLFSRPASRV RRSRGIVEECCFR <u>E</u> CDLALLETYCATPARSE	106
vIGF2-18 (vIGF2_9Q; vIGF2 S50D)	SRTLCCGGELVDTLQFVCGDRGFLFSRPASRV RRSRGIVEECCFR <u>D</u> CDLALLETYCATPARSE	107
vIGF2-19 (vIGF2 F26S V43L)	SRTLCCGGELVDTLQFVCGDRGSLFSRPASRV RRSRGILEECCFRSCDLALLETYCATPARSE	108
vIGF2-20 (vIGF2	SRTLCCGGELVDTLQFVCGDRGFLFSRPASRV	109

V43L)	RRSRGILEECCFRSCDLALLETYCATPARSE	
vIGF2-21 (vIGF2_ESRE; vIGF2 V14E F26S F28R V43E)	SRTLCCGGELEDTLQFVCGDRGSLRSRPASRV RRSRGIEEECCFRSCDLALLETYCATPARSE	110
vIGF2-22 (vIGF2 Δ 31-38GGGG)	SRTLCCGGELVDTLQFVCGDRGFLFSRGGGG RGIVEECCFRSCDLALLETYCATPARSE	111
vIGF2-23 (vIGF2 Δ 30-40GGGG)	SRTLCCGGELVDTLQFVCGDRGFLFSGGGGSGIV EECCFRSCDLALLETYCATPARSE	112
vIGF2-24 (vIGF2 Δ 31-38GGGG V43L)	SRTLCCGGELVDTLQFVCGDRGFLFSRGGGG RGILEECCFRSCDLALLETYCATPARSE	113
vIGF2-25 (vIGF2 L8A)	SRTACGGELVDTLQFVCGDRGFLFSRPASRV RRSRGIVEECCFRSCDLALLETYCATPARSE	114
vIGF2-26 (vIGF2 R6Q T7A L8A)	SQAACGGELVDTLQFVCGDRGFLFSRPASRV SRRSRGIVEECCFRSCDLALLETYCATPARSE	115
vIGF2-27 (vIGF2 R24E R34E R37E R38E)	SRTLCCGGELVDTLQFVCGDEGFLFSRPASEVS EESRGIVEECCFRSCDLALLETYCATPARSE	116
vIGF2-28 (vIGF2 R24E R34E)	SRTLCCGGELVDTLQFVCGDEGFLFSRPASEVS RRSRGIVEECCFRSCDLALLETYCATPARSE	117
vIGF2-29 (vIGF2 D23R S40D)	SRTLCCGGELVDTLQFVCGRRGFLFSRPASRV RRSRGIVEECCFRDCDLALLETYCATPARSE	118
vIGF2-30 (vIGF2 T16V D23R S50D)	SRTLCCGGELVDVLQFVCGRRGFLFSRPASRV RRSRGIVEECCFRDCDLALLETYCATPARSE	119
vIGF2-31 (vIGF2 Δ 31-38GGGG V43L S50D)	SRTLCCGGELVDTLQFVCGDRGFLFSRGGGG RGILEECCFRDCDLALLETYCATPARSE	120
vIGF2-32 (vIGF2 Δ 31-38GGGG)	SRTLCCGGELVDTLQFVCGDRGFLFSRGGGG RGILEECCFRECDLALLETYCATPARSE	121

V43L S50E)		
vIGF2-33 (vIGF2-N1)	SRTLCCGGELVDTLQFVCGDRGFLFRLPSRPVS RHSHRRSRGIVEECCFQRCNLALLETYCATPARSE	122
vIGF2-34 (vIGF2-N1 V43L S50E)	SRTLCCGGELVDTLQFVCGDRGFLFRLPSRPVS RHSHRRSRGILEECCFQECNLALLETYCATPARSE	123
vIGF2-1 R38G	SRTLCCGGELVDTLQFVCGDRGFLFSRPASRV <u>R</u> GSRGIVEECCFRSCDLALLETYCATPARSE	124
vIGF2-2 R38G, E45W	SRTLCCGGELVDTLQFVCGDRGFLFSRPASRV <u>R</u> GSRGIVE <u>W</u> CCFRSCDLALLETYCATPARSE	125
vIGF2-3 R38G, E45W, S50G	SRTLCCGGELVDTLQFVCGDRGFLFSRPASRV <u>R</u> GSRGIVE <u>W</u> CCFR <u>G</u> CDLALLETYCATPARSE	126
vIGF2-4 P31G, R38G, E45W, S50G	SRTLCCGGELVDTLQFVCGDRGFLFSR <u>G</u> ASRV SR <u>G</u> SRGIVE <u>W</u> CCFR <u>G</u> CDLALLETYCATPARSE	127
vIGF2-5 L17N, P31G, R38G, E45W, S50G	SRTLCCGGELVDT <u>N</u> QFVCGDRGFLFSR <u>G</u> ASRV SR <u>G</u> SRGIVE <u>W</u> CCFR <u>G</u> CDLALLETYCATPARSE	128
vIGF2-6 L17N, P31G, R38G, E45W, S50G, S66G	SRTLCCGGELVDT <u>N</u> QFVCGDRGFLFSR <u>G</u> ASRV SR <u>G</u> SRGIVE <u>W</u> CCFR <u>G</u> CDLALLETYCATPAR <u>G</u> E	129
vIGF2-7 L17N, P31G, R38G, E45W, S50G, A64M, S66G	SRTLCCGGELVDT <u>N</u> QFVCGDRGFLFSR <u>G</u> ASRV SR <u>G</u> SRGIVE <u>W</u> CCFR <u>G</u> CDLALLETYCATP <u>M</u> <u>R</u> <u>G</u> E	130
vIGF2-8 S5L, L17N, P31G, R38G, E45W, S50G, A64M, S66G	<u>L</u> RRTLCCGGELVDT <u>N</u> QFVCGDRGFLFSR <u>G</u> ASRV SR <u>G</u> SRGIVE <u>W</u> CCFR <u>G</u> CDLALLETYCATP <u>M</u> <u>R</u> <u>G</u> E	131

Table 4: IGF2 DNA Coding Sequences		
Peptide	DNA Sequence	SEQ ID NO
Mature WT IGF2	GCTTACCGCCCCAGTGAGACCCTGTGCGGC GGGGAGCTGGTGGACACCCTCCAGTTCGTC TGTGGGGACCGCGGCTTCTACTTCAGCAGG CCCGCAAGCCGTGTGAGCCGTCGCAGCCGT GGCATCGTTGAGGAGTGCTGTTTCCGCAGC TGTGACCTGGCCCTCCTGGAGACGTA CTGT GCTACCCCCGCCAAGTCCGAG	132
vIGF2 Δ 1-4, E6R, Y27L, K65R	TCTAGAACACTGTGCGGAGGGGAGCTTGTA GAACTCTTCAGTTCGTGTGTGGAGATCGC GGGTTCCCTCTTCTCTCGCCCCGCTTCCAGAG TTTCACGGAGGTCTAGGGGTATAGTAGAGG AGTGTTGTTTCAGGTCCTGTGACTTGGCGCT CCTCGAGACCTATTGCGCGACGCCAGCCAG GTCCGAA	133
vIGF2 Δ 1-4, E6R, Y27L, S50E, K65R	TCTAGAACACTGTGCGGAGGGGAGCTTGTA GAACTCTTCAGTTCGTGTGTGGAGATCGC GGGTTCCCTCTTCTCTCGCCCCGCTTCCAGAG TTTCACGGAGGTCTAGGGGTATAGTAGAGG AGTGTTGTTTCAGGGAGTGACTTGGCGC TCCTCGAGACCTATTGCGCGACGCCAGCCA GGTCCGAA	134
vIGF2-1 (vIGF2_1_NGGW GMG)	TCTAGAACACTGTGCGGAGGGGAGCTTGTA GAACTAACCAGTTCGTGTGTGGAGATCGC GGGTTCCCTCTTCTCTCGCGGCGCTTCCAGAG TTTCACGGGGCTCTAGGGGTATAGTAGAGT GGTGTTGTTTCAGGGGCTGTGACTTGGCGC TCCTCGAGACCTATTGCGCGACGCCAATGA GGGGCGAA	135

vIGF2-2 (vIGF2_2_GGWG MG)	TCTAGAACACTGTGCGGAGGGGAGCTTGTA GACACTCTTCAGTTCGTGTGTGGAGATCGC GGGTTCCCTCTTCTCTCGCGGCGCTTCCAGAG TTTCACGGGGCTCTAGGGGTATAGTAGAGT GGTGTGTTTTCAGGGGCTGTGACTTGCGCGC TCCTCGAGACCTATTGCGCGACGCCAATGA GGGGCGAA	136
vIGF2-3 (vIGF2_3_NGGG MG)	TCTAGAACACTGTGCGGAGGGGAGCTTGTA GACACTAACCAGTTCGTGTGTGGAGATCGC GGGTTCCCTCTTCTCTCGCGGCGCTTCCAGAG TTTCACGGGGCTCTAGGGGTATAGTAGAGG AGTGTGTTTTCAGGGGCTGTGACTTGCGCGC TCCTCGAGACCTATTGCGCGACGCCAATGA GGGGCGAA	137
vIGF2-4 (vIGF2_4_Δ32-41, 53aa)	TCTAGAACACTGTGCGGAGGGGAGCTTGTA GACACTCTTCAGTTCGTGTGTGGAGATCGC GGGTTCCCTCTTCTCTCGCCCCATAGTAGAGG AGTGTGTTTTCAGGTCCTGTGACTTGCGCGCT CCTCGAGACCTATTGCGCGACGCCAGCCAG GTCCGAA	138
vIGF2-5 (vIGF2 Δ 30-39, 53aa)	TCTAGAACACTGTGCGGAGGGGAGCTTGTA GACACTCTTCAGTTCGTGTGTGGAGATCGC GGGTTCCCTCTTCTCTAGGGGTATAGTAGAG GAGTGTGTTTTCAGGTCCTGTGACTTGCGCGC TCCTCGAGACCTATTGCGCGACGCCAGCCA GGTCCGAA	139
vIGF2-6 (vIGF2 Δ 33-40, 55aa)	TCTAGAACACTGTGCGGAGGGGAGCTTGTA GACACTCTTCAGTTCGTGTGTGGAGATCGC GGGTTCCCTCTTCTCTCGCCCCGCTGGTATAG TAGAGGAGTGTTGTTTTCAGGTCCTGTGACTT GGCGCTCCTCGAGACCTATTGCGCGACGCC AGCCAGGTCCGAA	140

vIGF2-7 (vIGF2 Δ 30-39/V14D/F28R/V43D/F26A)	TCTAGAACACTGTGCGGAGGGGAGCTTGAC GAACTCTTCAGTTCGTGTGTGGAGATCGC GGGGCCCTCAGATCTAGGGGTATAGACGAG GAGTGTGTTTCAGGTCCTGTGACTTGCGCG TCCTCGAGACCTATTGCGCGACGCCAGCCA GGTCCGAA	141
vIGF2-8 (vIGF2_8_REE)	TCTAGAACACTGTGCGGAGGGGAGCTTGTA GAACTCTTCAGTTCGTGTGTGGAAGACGC GGGGAGCTCTTCTCTCGCCCCGCTTCCAGA GTTTCACGGAGGTCTAGGGGTATAGTAGAG GAGTGTGTTTCAGGGAGTGTGACTTGCGCG CTCCTCGAGACCTATTGCGCGACGCCAGCC AGGTCCGAA	142
vIGF2-9 (vIGF2_9_ Δ 30-39-REE; vIGF2 Homerun)	TCTAGAACACTGTGCGGAGGGGAGCTTGTA GAACTCTTCAGTTCGTGTGTGGAAGACGC GGGGAGCTCTTCTCTCGCCCCGCTGGTATA GTAGAGGAGTGTTGTTTCAGGGAGTGTGAC TTGGCGCTCCTCGAGACCTATTGCGCGACG CCAGCCAGGTCCGAA	143
vIGF2-10 (vIGF2_1Q; vIGF2 D23R)	TCTAGAACACTGTGCGGAGGGGAGCTTGTA GAACTCTTCAGTTCGTGTGTGGACGTCGC GGGTTCTCTTCTCTCGCCCCGCTTCCAGAG TTTCACGGAGGTCTAGGGGTATAGTAGAGG AGTGTGTTTCAGGTCCTGTGACTTGCGCGCT CCTCGAGACCTATTGCGCGACGCCAGCCAG GTCCGAA	144
vIGF2-11 (vIGF2_2Q; vIGF2 F19W)	TCTAGAACACTGTGCGGAGGGGAGCTTGTA GAACTCTTCAGTGGGTGTGTGGAGATCGC GGGTTCTCTTCTCTCGCCCCGCTTCCAGAG TTTCACGGAGGTCTAGGGGTATAGTAGAGG AGTGTGTTTCAGGTCCTGTGACTTGCGCGCT CCTCGAGACCTATTGCGCGACGCCAGCCAG	145

	GTCCGAA	
vIGF2-12 (vIGF2_3Q; vIGF2 T16W)	TCTAGAACACTGTGCGGAGGGGAGCTTGTA GACTGGCTTCAGTTCGTGTGTGGAGATCGC GGGTTCTCTTCTCTCGCCCCGCTTCCAGAG TTTCACGGAGGTCTAGGGGTATAGTAGAGG AGTGTTGTTTCAGGTCCTGTGACTTGGCGCT CCTCGAGACCTATTGCGCGACGCCAGCCAG GTCCGAA	146
vIGF2-13 (vIGF2_4Q; vIGF2 D23K)	TCTAGAACACTGTGCGGAGGGGAGCTTGTA GACACTCTTCAGTTCGTGTGTGGAAAGCGC GGGTTCTCTTCTCTCGCCCCGCTTCCAGAG TTTCACGGAGGTCTAGGGGTATAGTAGAGG AGTGTTGTTTCAGGTCCTGTGACTTGGCGCT CCTCGAGACCTATTGCGCGACGCCAGCCAG GTCCGAA	147
vIGF2-14 (vIGF2_5Q; vIGF2 T16Y)	TCTAGAACACTGTGCGGAGGGGAGCTTGTA GACTATCTTCAGTTCGTGTGTGGAGATCGC GGGTTCTCTTCTCTCGCCCCGCTTCCAGAG TTTCACGGAGGTCTAGGGGTATAGTAGAGG AGTGTTGTTTCAGGTCCTGTGACTTGGCGCT CCTCGAGACCTATTGCGCGACGCCAGCCAG GTCCGAA	148
vIGF2-15 (vIGF2_6Q; vIGF2 F26E)	TCTAGAACACTGTGCGGAGGGGAGCTTGTA GACACTCTTCAGTTCGTGTGTGGAGATCGC GGGGAGCTCTTCTCTCGCCCCGCTTCCAGA GTTTCACGGAGGTCTAGGGGTATAGTAGAG GAGTGTTGTTTCAGGTCCTGTGACTTGGCGC TCCTCGAGACCTATTGCGCGACGCCAGCCA GGTCCGAA	149
vIGF2-16 (vIGF2_7Q; vIGF2 T16V)	TCTAGAACACTGTGCGGAGGGGAGCTTGTA GACGTTCTTCAGTTCGTGTGTGGAGATCGC GGGTTCTCTTCTCTCGCCCCGCTTCCAGAG	150

	TTTCACGGAGGTCTAGGGGTATAGTAGAGG AGTGTTGTTTCAGGTCCTGTGACTTGGCGCT CCTCGAGACCTATTGCGCGACGCCAGCCAG GTCCGAA	
vIGF2-17 (vIGF2_8Q; vIGF2 S50E)	TCTAGAACACTGTGCGGAGGGGAGCTTGTA GACACTCTTCAGTTCGTGTGTGGAGATCGC GGGTTCCCTCTTCTCTCGCCCCGCTTCCAGAG TTTCACGGAGGTCTAGGGGTATAGTAGAGG AGTGTTGTTTCAGGGAGTGTGACTTGGCGC TCCTCGAGACCTATTGCGCGACGCCAGCCA GGTCCGAA	151
vIGF2-18 (vIGF2_9Q; vIGF2 S50D)	TCTAGAACACTGTGCGGAGGGGAGCTTGTA GACACTCTTCAGTTCGTGTGTGGAGATCGC GGGTTCCCTCTTCTCTCGCCCCGCTTCCAGAG TTTCACGGAGGTCTAGGGGTATAGTAGAGG AGTGTTGTTTCAGGGACTGTGACTTGGCGC TCCTCGAGACCTATTGCGCGACGCCAGCCA GGTCCGAA	152
vIGF2-19 (vIGF2 F26S V43L)	TCTAGAACACTGTGCGGAGGGGAGCTTGTA GACACTCTTCAGTTCGTGTGTGGAGATCGC GGGAGCCTCTTCTCTCGCCCCGCTTCCAGA GTTTCACGGAGGTCTAGGGGTATACTGGAG GAGTGTTGTTTCAGGTCCTGTGACTTGGCGC TCCTCGAGACCTATTGCGCGACGCCAGCCA GGTCCGAA	153
vIGF2-20 (vIGF2 V43L)	TCTAGAACACTGTGCGGAGGGGAGCTTGTA GACACTCTTCAGTTCGTGTGTGGAGATCGC GGGTTCCCTCTTCTCTCGCCCCGCTTCCAGAG TTTCACGGAGGTCTAGGGGTATACTGGAGG AGTGTTGTTTCAGGTCCTGTGACTTGGCGCT CCTCGAGACCTATTGCGCGACGCCAGCCAG GTCCGAA	154

vIGF2-21 (vIGF2_ESRE; vIGF2 V14E F26S F28R V43E)	TCTAGAACACTGTGCGGAGGGGAGCTTGAG GACACTCTTCAGTTCGTGTGTGGAGATCGC GGGAGCCTCAGATCTCGCCCCGCTTCCAGA GTTTCACGGAGGTCTAGGGGTATAGAGGAG GAGTGTTGTTTCAGGTCCTGTGACTTGGCGC TCCTCGAGACCTATTGCGCGACGCCAGCCA GGTCCGAA	155
vIGF2-22 (vIGF2 Δ 31-38GGGG)	TCTAGAACACTGTGCGGAGGGGAGCTTGTA GACACTCTTCAGTTCGTGTGTGGAGATCGC GGGTTTCCTCTTCTCTCGCGGAGGTGGAGGT TCTAGGGGTATAGTAGAGGAGTGTTGTTTC AGGTCCTGTGACTTGGCGCTCCTCGAGACC TATTGCGCGACGCCAGCCAGGTCCGAA	156
vIGF2-23 (vIGF2 Δ 30-40GGGG)	TCTAGAACACTGTGCGGAGGGGAGCTTGTA GACACTCTTCAGTTCGTGTGTGGAGATCGC GGGTTTCCTCTTCTCTGGTGGAGGTTCTGGTA TAGTAGAGGAGTGTTGTTTCAGGTCCTGTG ACTTGGCGCTCCTCGAGACCTATTGCGCGA CGCCAGCCAGGTCCGAA	157
vIGF2-24 (vIGF2 Δ 31-38GGGG V43L)	TCTAGAACACTGTGCGGAGGGGAGCTTGTA GACACTCTTCAGTTCGTGTGTGGAGATCGC GGGTTTCCTCTTCTCTCGCGGAGGTGGAGGT TCTAGGGGTATACTGGAGGAGTGTTGTTTC AGGTCCTGTGACTTGGCGCTCCTCGAGACC TATTGCGCGACGCCAGCCAGGTCCGAA	158
vIGF2-25 (vIGF2 L8A)	TCTCAGGCCGCGTGCGGAGGGGAGCTTGTA GACACTCTTCAGTTCGTGTGTGGAGATCGC GGGTTTCCTCTTCTCTCGCCCCGCTTCCAGAG TTTCACGGAGGTCTAGGGGTATAGTAGAGG AGTGTTGTTTCAGGTCCTGTGACTTGGCGCT CCTCGAGACCTATTGCGCGACGCCAGCCAG GTCCGAA	159

vIGF2-26 (vIGF2 R6Q T7A L8A)	TCTCAGGCCGCGTGCGGAGGGGAGCTTGTA GACACTCTTCAGTTCGTGTGTGGAGATCGC GGGTTCCCTCTTCTCTCGCCCCGCTTCCAGAG TTTCACGGAGGTCTAGGGGTATAGTAGAGG AGTGTTGTTTCAGGTCCTGTGACTTGCGCGCT CCTCGAGACCTATTGCGCGACGCCAGCCAG GTCCGAA	160
vIGF2-27 (vIGF2 R24E R34E R37E R38E)	TCTAGAACACTGTGCGGAGGGGAGCTTGTA GACACTCTTCAGTTCGTGTGTGGAGATGAG GGGTTCCCTCTTCTCTCGCCCCGCTTCCGAGG TTTCAGAGGAATCTAGGGGTATAGTAGAGG AGTGTTGTTTCAGGTCCTGTGACTTGCGCGCT CCTCGAGACCTATTGCGCGACGCCAGCCAG GTCCGAA	161
vIGF2-28 (vIGF2 R24E R34E)	TCTAGAACACTGTGCGGAGGGGAGCTTGTA GACACTCTTCAGTTCGTGTGTGGAGATGAG GGGTTCCCTCTTCTCTCGCCCCGCTTCCGAGG TTTCACGGAGGTCTAGGGGTATAGTAGAGG AGTGTTGTTTCAGGTCCTGTGACTTGCGCGCT CCTCGAGACCTATTGCGCGACGCCAGCCAG GTCCGAA	162
vIGF2-29 (vIGF2 D23R S40D)	TCTAGAACACTGTGCGGAGGGGAGCTTGTA GACACTCTTCAGTTCGTGTGTGGAAGACGC GGGTTCCCTCTTCTCTCGCCCCGCTTCCAGAG TTTCACGGAGGTCTAGGGGTATAGTAGAGG AGTGTTGTTTCAGGGACTGTGACTTGCGCGC TCCTCGAGACCTATTGCGCGACGCCAGCCA GGTCCGAA	163
vIGF2-30 (vIGF2 T16V D23R S50D)	TCTAGAACACTGTGCGGAGGGGAGCTTGTA GACGTGCTTCAGTTCGTGTGTGGAAGACGC GGGTTCCCTCTTCTCTCGCCCCGCTTCCAGAG TTTCACGGAGGTCTAGGGGTATAGTAGAGG	164

	AGTGTGTTTTCAGGGACTGTGACTTGGCGC TCCTCGAGACCTATTGCGCGACGCCAGCCA GGTCCGAA	
vIGF2-31 (vIGF2 Δ 31-38GGGG V43L S50D)	TCTAGAACACTGTGCGGAGGGGAGCTTGTA GACACTCTTCAGTTCGTGTGTGGAGATCGC GGGTTCTTCTTCTCTCGCGGAGGTGGAGGT TCTAGGGGTATACTGGAGGAGTGTGTTTC AGGGACTGTGACTTGGCGCTCCTCGAGACC TATTGCGCGACGCCAGCCAGGTCCGAA	165
vIGF2-32 (vIGF2 Δ 31-38GGGG V43L S50E)	TCTAGAACACTGTGCGGAGGGGAGCTTGTA GACACTCTTCAGTTCGTGTGTGGAGATCGC GGGTTCTTCTTCTCTCGCGGAGGTGGAGGT TCTAGGGGTATACTGGAGGAGTGTGTTTC AGGGAGTGTGACTTGGCGCTCCTCGAGACC TATTGCGCGACGCCAGCCAGGTCCGAA	166
vIGF2-33 (vIGF2- N1)	TCTAGAACACTGTGCGGAGGGGAGCTTGTA GACACTCTTCAGTTCGTGTGTGGAGATCGC GGGTTCTTGTTTCGATTGCCGTCCAGGCCCG TGTCCTGGCACAGTCACCGCAGGTCAAGGG GGATAGTTGAAGAATGTTGCTTTCAGAGGT GTAATTTGGCGCTCCTCGAGACCTATTGCG CGACGCCAGCCAGGTCCGAA	167
vIGF2-34 (vIGF2- N1 V43L S50E)	TCTAGAACACTGTGCGGAGGGGAGCTTGTA GACACTCTTCAGTTCGTGTGTGGAGATCGC GGGTTCTTGTTTCGATTGCCGTCCAGGCCCG TGTCCTGGCACAGTCACCGCAGGTCAAGGG GGATACTGGAAGAATGTTGCTTTCAGGAGT GTAATTTGGCGCTCCTCGAGACCTATTGCG CGACGCCAGCCAGGTCCGAA	168

[00148] Provided herein are gene therapy constructs useful in treating a disorder further comprising an internal ribosome entry sequence (IRES) for increasing gene expression by bypassing the bottleneck of translation initiation. Suitable internal ribosomal entry sequences for optimizing expression for gene therapy include but are not limited to a cricket paralysis virus (CrPV) IRES, a picornavirus IRES, an Aphthovirus IRES, a Kaposi's sarcoma-associated herpesvirus IRES, a Hepatitis A IRES, a Hepatitis C IRES, a Pestivirus IRES, a Cripavirus IRES, a Rhopalosiphum padi virus IRES, a Merek's disease virus IRES, and other suitable IRES sequences. In some embodiments, the gene therapy construct comprises a CrPV IRES. In some embodiments, the CrPV IRES has a nucleic acid sequence of

10 AAAAATGTGATCTTGCTTGTAATACAATTTTGAGAGGTTAATAAATTACAAGTAG
TGCTATTTTTGTATTTAGGTTAGCTATTTAGCTTTACGTTCCAGGATGCCTAGTGGC
AGCCCCACAATATCCAGGAAGCCCTCTCTGCGGTTTTTCAGATTAGGTAGTCGAAA
AACCTAAGAAATTTACCTGCT (SEQ ID NO: 191). In some embodiments, the CrPV
IRES sequence is at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least
15 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 191.

Signal Peptides

[00149] Gene therapy constructs provided herein, in some embodiments, further comprise a signal peptide, which improves secretion of the therapeutic protein from the cell transduced with the gene therapy construct. The signal peptide in some embodiments
20 improves protein processing of therapeutic proteins and facilitates translocation of the nascent polypeptide-ribosome complex to the ER and ensuring proper co-translational and post-translational modifications. In some embodiments, the signal peptide is located (i) in between the translation initiation sequence and the therapeutic protein or (ii) a downstream position of the therapeutic protein. Signal peptides useful in gene therapy constructs include but are not
25 limited to binding immunoglobulin protein (BiP) signal peptide from the family of HSP70 proteins (e.g., HSPA5, heat shock protein family A member 5) and Gaussia signal peptides, and variants thereof. These signal peptides have ultrahigh affinity to the signal recognition particle. Examples of BiP and Gaussia amino acid sequences are provided in Table 5 below. In some embodiments, the signal peptide has an amino acid sequence that is at least 90, 95, 96,
30 97, 98 or 99% identical to a sequence selected from the group consisting of SEQ ID NOs: 169-180. In some embodiments, the signal peptide differs from a sequence selected from the group consisting of SEQ ID NOs: 169-180 by 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1

amino acid. In some embodiments, the native signal peptide, referred to interchangeably herein as the “endogenous signal peptide” of a lysosomal protein is used.

Table 5: Signal Peptide Sequences		
Signal Peptide	Amino Acid Sequence	SEQ ID NO:
Native human BiP	MKLSLVAAML LLLSAARA	169
Modified BiP-1	MKLSLVAAML LLLSLVAAML LLLSAARA	170
Modified BiP-2	MKLSLVAAML LLLWVAL LLLSAARA	171
Modified BiP-3	MKLSLVAAML LLLSLVAL LLLSAARA	172
Modified BiP-4	MKLSLVAAML LLLALVAL LLLSAARA	173
Gaussia	MGVKVLFALICIAVAEA	174
Native PPT1 Signal Peptide (eSP)	MASPGCLWLLAVALLPWTCASRALQHL	175
Native PPT1 Signal Peptide (eSP AA)	MASPGCLWLLAVALLPWTCASRALQH LAA	176
Native PPT1 Signal Peptide C6S (eSP C6S)	MASPGSLWLLAVALLPWTCASRALQHL	177
Native PPT1 Signal Peptide C6S (eSP C6S AA)	MASPGSLWLLAVALLPWTCASRALQH LAA	178
Native TPP1 Signal Peptide	MGLQACLLGLFALILSGKC	179
Native NAGLU Signal Peptide	MEAVAVAAAVGVLLLAGAGGAAGD	180

- 5 **[00150]** The BiP signal peptide-signal recognition particle (SRP) interaction facilitates translocation to the ER. This interaction is illustrated in **FIG 20**.

[00151] The Gaussia signal peptide is derived from the luciferase from *Gaussia princeps* and directs increased protein synthesis and secretion of therapeutic proteins fused to this signal peptide. In some embodiments, the Gaussia signal peptide has an amino acid sequence that is

at least 90, 95, 96, 97, 98 or 99% identical to SEQ ID NO: 174. In some embodiments, the signal peptide differs from SEQ ID NO: 174 by 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer, or 1 amino acid.

5 *Linker*

[00152] Gene therapy constructs provided herein, in some embodiments, comprise a linker between the targeting peptide and the therapeutic protein. Such linkers, in some embodiments, maintain correct spacing and mitigate steric clash between the vIGF2 peptide and the therapeutic protein. Linkers, in some embodiments, comprise repeated glycine residues, repeated glycine-serine residues, and combinations thereof. In some embodiments, the linker consists of 5-20 amino acids, 5-15 amino acids, 5-10 amino acids, 8-12 amino acids, or about 5, 6, 7, 8, 9, 10, 11, 12 or 13 amino acids. Suitable linkers for gene therapy and enzyme replacement therapy constructs herein include but are not limited to those provided in Table 6 below.

Table 6: Linker Sequences		
GS Linkers	Sequence	SEQ ID NO:
	GGGGS	181
	GGGS	182
	GGGSGGGGS	183
	GGGSGGGGS	184
	GGSGSGSTS	185
	GGGSGGGGS	186
	GGGSGSGGGGS	187
Lysosomal cleavage linker	RPRAVPTQA	188

15 *Translation Initiation Sequence*

[00153] Gene therapy constructs provided herein comprise a nucleic acid having a translation initiation sequence, such as a Kozak sequence which aids in initiation of translation of the mRNA. Kozak sequences contemplated herein have a consensus sequence of (gcc)RccATGG where a lowercase letter denotes the most common base at the position and the base varies, uppercase letters indicate highly conserved bases that only vary rarely change. R

indicates that a purine (adenine or guanine) is always observed at that position. The sequence in parentheses (gcc) is of uncertain significance. In some embodiments, the Kozak sequence comprises the sequence AX_1X_2ATGA , wherein each of X_1 and X_2 is any nucleotide. In some embodiments, X_1 comprises A. In some embodiments, X_2 comprises G. In some
5 embodiments, the Kozak sequence comprises a nucleic acid sequence at least 85% identical to AAGATGA. In some embodiments, the Kozak sequence differs from the sequence of AAGATGA by one or two nucleotides. In some embodiments, Kozak sequences provided herein have a sequence of AAGATGA. In some embodiments the Kozak sequence comprises a nucleic acid sequence at least 85% identical to GCAAGATG. In some embodiments the
10 Kozak sequence differs from the sequence of GCAAGATG by one or two nucleotides. In some embodiments, the Kozak sequence comprises GCAAGATG. In some embodiments the Kozak sequence comprises a nucleic acid sequence at least 85% identical to CACCATG. In some embodiments the Kozak sequence differs from the sequence of CACCATG by one or two nucleotides. In some embodiments, the Kozak sequence comprises CACCATG.

15 *Therapeutic Protein*

[00154] Gene therapy constructs provided herein comprise a nucleic acid encoding a therapeutic protein for treating a genetic disorder due to a genetic defect in an individual resulting in an absent or defective protein. The therapeutic protein expressed from the gene therapy construct replaces the absent or defective protein. Therapeutic proteins, therefore, are
20 chosen based on the genetic defect in need of treatment in an individual. In some embodiments, the therapeutic protein is a structural protein. In some embodiments, the therapeutic protein is an enzyme. In some embodiments, the therapeutic protein is a regulatory protein. In some embodiments, the therapeutic protein is a receptor. In some embodiments, the therapeutic protein is a peptide hormone. In some embodiments, the therapeutic protein is
25 a cytokine or a chemokine.

[00155] In some embodiments, gene therapy constructs herein encode an enzyme, such as an enzyme having a genetic defect in an individual with a lysosomal storage disorder. In some embodiments, gene therapy constructs encode a lysosomal enzyme, such as a glycosidase, a protease, or a sulfatase. In some embodiments, enzymes encoded by gene
30 therapy constructs provided herein include but are not limited to α -D-mannosidase; N-aspartyl- β -glucosaminidase; β -galactosidase; ceramidase; fucosidase; galactocerebrosidase; arylsulfatase A; N-acetylglucosamine-1-phosphotransferase; iduronate sulfatase; N-

acetylglucosaminidase; acetyl-CoA: α -glucosaminide acetyltransferase; N-acetylglucosamine 6-sulfatase; β -glucuronidase; hyaluronidase; sialidase; sulfatase; sphingomyelinase; acid β -mannosidase; cathepsin K; 3-hexosaminidase A; β -hexosaminidase B; α -N-acetylgalactosaminidase; sialin; hexosaminidase A; beta-glucosidase; α -iduronidase; α -galactosidase A; β -glucocerebrosidase; lysosomal acid lipase; glycosaminoglycan alpha-L-iduronohydrolase; iduronate-2-sulfatase; N-acetylgalactosamine-6-sulfatase; glycosaminoglycan N-acetylgalactosamine 4-sulfatase; alpha-glucosidase; heparan sulfamidase; gp-91 subunit of NADPH oxidase; adenosine deaminase; cyclin dependent kinase like 5; and palmitoyl protein thioesterase 1. In some embodiments, enzymes encoded by gene therapy constructs provided herein comprise alpha-glucosidase. In some embodiments, the therapeutic protein is associated with a genetic disorder selected from the group consisting of cystic fibrosis, alpha- and beta-thalassemias, sickle cell anemia, Marfan syndrome, fragile X syndrome, Huntington's disease, hemochromatosis, Congenital Deafness (nonsyndromic), Tay-Sachs, Familial hypercholesterolemia, Duchenne muscular dystrophy, Stargardt disease, Usher syndrome, choroideremia, achromatopsia, X-linked retinoschisis, hemophilia, Wiskott-Aldrich syndrome, X-linked chronic granulomatous disease, aromatic L-amino acid decarboxylase deficiency, recessive dystrophic epidermolysis bullosa, alpha 1 antitrypsin deficiency, Hutchinson-Gilford progeria syndrome (HGPS), Noonan syndrome, X-linked severe combined immunodeficiency (X-SCID).

20

[00156] *Gene Therapy Vector Examples*

Gene Therapy Vectors and Compositions

[00157] Provided herein are gene therapy vectors in which a nucleic acid, such as a DNA, encoding a therapeutic fusion protein, such as a vIGF2 fusion, optionally having a signal peptide. The gene therapy vector optionally comprises an internal ribosomal entry sequence. Vectors derived from retroviruses such as the lentivirus are suitable tools to achieve long-term gene transfer since they allow long-term, stable integration of a transgene and its propagation in daughter cells. Lentiviral and adeno-associated viral vectors have the added advantage over vectors derived from onco-retroviruses such as murine leukemia viruses in that they are capable of transducing non-proliferating cells, such as hepatocytes and neurons. They also have the added advantage of low immunogenicity.

30

[00158] Exemplary gene therapy vectors herein encode therapeutic proteins and therapeutic fusion proteins comprising a vIGF2 peptide. Nucleic acids encoding exemplary fusion protein amino acid sequences are provided in Table 7 below.

5

Table 7: DNA Sequences		
Construct	DNA Sequence	SEQ ID NO
Kozak-hGAA (Natural GAA)	GCAAGATGGGAGTGAGGCACCCGCCCTGCTCCACCGGCTCCTG GCCGTCTGCGCCCTCGTGTCTTGGCAACCGCTGCACTCCTGGGG CACATCCTACTCCATGATTTCTGCTGGTTCCCCGAGAGCTGAGT GGCTCCTCCCCAGTCCTGGAGGAGACTACCCAGCTCACCAGCA GGGAGCCAGTAGACCAGGGCCCCGGGATGCCAGGCACACCCC GGCCGTCCCAGAGCAGTGCCACACAGTGCGACGTCCCCCCCCAA CAGCCGCTTCGATTGCGCCCCTGACAAGGCCATCACCCAGGAAC AGTGCGAGGCCCCGCGGCTGTTGCTACATCCCTGCAAAGCAGGGG CTGCAGGGAGCCCAGATGGGGCAGCCCTGGTGCTTCTTCCCACC CAGCTACCCCAGCTACAAGCTGGAGAACCTGAGCTCCTCTGAAA TGGGCTACACGGCCACCCTGACCCGTACCACCCCCACCTTCTTCC CCAAGGACATCCTGACCCTGCGGCTGGACGTGATGATGGAGACT GAGAACCGCCTCCACTTCACGATCAAAGATCCAGCTAACAGGCG CTACGAGGTGCCCTTGGAGACCCCGCATGTCCACAGCCGGGCAC CGTCCCCACTCTACAGCGTGGAGTTCTCCGAGGAGCCCTTCGGG GTGATCGTGCGCCGGCAGCTGGACGGCCGCGTGCTGCTGAACAC GACGGTGGCGCCCCCTGTTCTTTGCGGACCAGTTCCTTCAGCTGTC CACCTCGCTGCCCTCGCAGTATATCACAGGCCTCGCCGAGCACCT CAGTCCCCTGATGCTCAGCACCAGCTGGACCAGGATCACCCCTGT GGAACCGGGACCTTGCGCCACGCCCCGGTGCGAACCTCTACGGG TCTCACCCCTTTCTACCTGGCGCTGGAGGACGGCGGGTTCGGCACA CGGGGTGTTCTGCTAAACAGCAATGCCATGGATGTGGTCCTGC AGCCGAGCCCTGCCCTTAGCTGGAGGTGACAGGTGGGATCCTG GATGTCTACATCTTCTGCGGCCAGAGCCCAAGAGCGTGGTGCA	189

	GCAGTACCTGGACGTTGTGGGATACCCGTTTCATGCCGCCATACTG GGGCCTGGGCTTCCACCTGTGCCGCTGGGGCTACTCCTCCACCGC TATACCCGCCAGGTGGTGGAGAACATGACCAGGGCCCCACTTCC CCCTGGACGTCCAGTGGAACGACCTGGACTACATGGACTCCCGG AGGGACTTCACGTTCAACAAGGATGGCTTCCGGGACTTCCCGGC CATGGTGCAGGAGCTGCACCAGGGCGGCCGGCGCTACATGATGA TCGTGGATCCTGCCATCAGCAGCTCGGGCCCTGCCGGGAGCTAC AGGCCCTACGACGAGGGTCTGCGGAGGGGGGTTTTTCATCACCAA CGAGACCGGCCAGCCGCTGATTGGGAAGGTATGGCCCGGGTCCA CTGCCTTCCCCGACTTCACCAACCCACAGCCCTGGCCTGGTGGG AGGACATGGTGGCTGAGTTCCATGACCAGGTGCCCTTCGACGGC ATGTGGATTGACATGAACGAGCCTTCCAACCTTCATCAGGGGCTCT GAGGACGGCTGCCCCAACAATGAGCTGGAGAACCACCCCTACGT GCCTGGGGTGGTTGGGGGGACCCCTCCAGGCGGCCACCATCTGTG CCTCCAGCCACCAGTTTCTCTCCACACACTACAACCTGCACAACC TCTACGGCCTGACCGAAGCCATCGCCTCCCACAGGGCGCTGGTG AAGGCTCGGGGGACACGCCCATTTGTGATCTCCCGCTCGACCTTT GCTGGCCACGGCCGATACGCCGGCCACTGGACGGGGGACGTGTG GAGCTCCTGGGAGCAGCTCGCCTCCTCCGTGCCAGAAATCCTGC AGTTTAACCTGCTGGGGGTGCCTCTGGTTCGGGGCCGACGTCTGC GGCTTCCTGGGCAACACCTCAGAGGAGCTGTGTGTGCGCTGGAC CCAGCTGGGGGCCTTCTACCCCTTCATGCGGAACCACAACAGCC TGCTCAGTCTGCCCCAGGAGCCGTACAGCTTCAGCGAGCCGGCC CAGCAGGCCATGAGGAAGGCCCTCACCTGCGCTACGCACTCCT CCCCACCTCTACACACTGTTCCACCAGGCCACGTCGCGGGGG AGACCGTGGCCCGGCCCTCTTCCTGGAGTTCCCCAAGGACTCTA GCACCTGGACTGTGGACCACCAGCTCCTGTGGGGGGAGGCCCTG CTCATCACCCAGTGCTCCAGGCCGGAAGGCCGAAGTGACTGG CTACTTCCCCTTGGGCACATGGTACGACCTGCAGACGGTGCCAGT AGAGGCCCTTGGCAGCCTCCCACCCACCTGCAGCTCCCCGTG AGCCAGCCATCCACAGCGAGGGGCAGTGGGTGACGCTGCCGGCC CCCCTGGACACCATCAACGTCCACCTCCGGGCTGGGTACATCATC	
--	---	--

	CCCCTGCAGGGCCCTGGCCTCACAACCACAGAGTCCCGCCAGCA GCCCATGGCCCTGGCTGTGGCCCTGACCAAGGGTGGGGAGGCCC GAGGGGAGCTTTTCTGGGACGATGGAGAGAGCCTGGAAGTGCTG GAGCGAGGGGCCTACACACAGGTCATCTTCCTGGCCAGGAATAA CACGATCGTGAATGAGCTGGTACGTGTGACCAGTGAGGGAGCTG GCCTGCAGCTGCAGAAGGTGACTGTCCTGGGCGTGGCCACGGCG CCCCAGCAGGTCTCTCCAACGGTGTCCCTGTCTCCAACCTCACC TACAGCCCCGACACCAAGGTCCTGGACATCTGTGTCTCGCTGTTG ATGGGAGAGCAGTTTCTCGTCAGCTGGTGTTAG	
Kozak BiP- vIGF2-GAA (“Engineered hGAA”)	GCAAGATGAAGCTCTCCCTGGTGGCCGCGATGCTGCTGCTC AGCGCGGCGCGGGCCTCTAGAACACTGTGCGGAGGGGAGCTTGT AGACACTCTTCAGTTCGTGTGTGGAGATCGCGGGTTCCTCTTCTC TCGCCCCGCTTCCAGAGTTTCACGGAGGTCTAGGGGTATAGTAG AGGAGTGTTGTTTCAGGTCTGTGACTTGGCGCTCCTCGAGACCT ATTGCGCGACGCCAGCCAGGTCCGAAGGGGGCGGTGGCTCAGGT GGTGGAGGTAGCAGACCAGGGCCCCGGGATGCCAGGCACACC CCGGCCGTCCCAGAGCAGTGCCACACAGTGCGACGTCCCCCCC AACAGCCGCTTCGATTGCGCCCCCTGACAAGGCCATCACCCAGGA ACAGTGCGAGGGCCCGCGGCTGTTGCTACATCCCTGCAAAGCAGG GGCTGCAGGGAGCCCAGATGGGGCAGCCCTGGTGCTTCTTCCCA CCCAGCTACCCAGCTACAAGCTGGAGAACCTGAGCTCCTCTGA AATGGGCTACACGGCCACCCTGACCCGTACCACCCCCACCTTCTT CCCCAGGACATCCTGACCCTGCGGCTGGACGTGATGATGGAGA CTGAGAACCGCCTCCACTTCACGATCAAAGATCCAGCTAACAGG CGCTACGAGGTGCCCTTGGAGACCCCGCATGTCCACAGCCGGGC ACCGTCCCCACTCTACAGCGTGGAGTTCTCCGAGGAGCCCTTCGG GGTGATCGTGCGCCGGCAGCTGGACGGCCGCGTGCTGCTGAACA CGACGGTGGCGCCCCTGTTCTTTGCGGACCAGTTCCTTCAGCTGT CCACCTCGCTGCCCTCGCAGTATATCACCGGCCTCGCCGAGCACC TCAGTCCCCTGATGCTCAGCACCAGCTGGACCAGGATCACCTGT GGAACCGGGACCTTGCGCCACGCCCCGGTGCGAACCTCTACGGG TCTCACCCCTTTCTACCTGGCGCTGGAGGACGGCGGGTTCGGCACA	190

	CGGGGTGTTCTGCTAAACAGCAATGCCATGGATGTGGTCCTGC AGCCGAGCCCTGCCCTTAGCTGGAGGTCGACAGGTGGGATCCTG GATGTCTACATCTTCTGGGCCCAGAGCCCAAGAGCGTGGTGCA GCAGTACCTGGACGTTGTGGGATACCCGTTTCATGCCGCCATACTG GGGCCTGGGCTTCCACCTGTGCCGCTGGGGCTACTCCTCCACCGC TATCACCCGCCAGGTGGTGGAGAACATGACCAGGGCCCCACTTCC CCCTGGACGTCCAGTGGAACGACCTGGACTACATGGACTCCCGG AGGGACTTCACGTTCAACAAGGATGGCTTCCGGGACTTCCCGGC CATGGTGCAGGAGCTGCACCAGGGCGGCCGGCGCTACATGATGA TCGTGGATCCTGCCATCAGCAGCTCGGGCCCTGCCGGGAGCTAC AGGCCCTACGACGAGGGTCTGCGGAGGGGGGTTTTTCATACCAA CGAGACCGGCCAGCCGCTGATTGGGAAGGTATGGCCCGGGTCCA CTGCCTTCCCCGACTTCACCAACCCACAGCCCTGGCCTGGTGGG AGGACATGGTGGCTGAGTTCCATGACCAGGTGCCCTTCGACGGC ATGTGGATTGACATGAACGAGCCTTCCAACCTTCATAGGGGGCTCT GAGGACGGCTGCCCCAACAATGAGCTGGAGAACCCACCCTACGT GCCTGGGGTGTTGGGGGGACCCCTCAGGCGGCCACCATCTGTG CCTCCAGCCACCAGTTTCTCTCCACACACTACAACCTGCACAACC TCTACGGCCTGACCGAAGCCATCGCCTCCCACAGGGCGCTGGTG AAGGCTCGGGGGACACGCCCATTTGTGATCTCCCGCTCGACCTTT GCTGGCCACGGCCGATACGCCGGCCACTGGACGGGGGACGTGTG GAGCTCCTGGGAGCAGCTCGCCTCCTCCGTGCCAGAAATCCTGC AGTTTAACTGCTGGGGGTGCCTCTGGTTCGGGGCCGACGTCTGC GGCTTCCTGGGCAACACCTCAGAGGAGCTGTGTGTGCGCTGGAC CCAGCTGGGGGCCTTCTACCCCTTCATGCGGAACCACAACAGCC TGCTCAGTCTGCCCCAGGAGCCGTACAGCTTCAGCGAGCCGGCC CAGCAGGCCATGAGGAAGGCCCTCACCTGCGCTACGCACTCCT CCCCACCTCTACACACTGTTCCACCAGGCCACGTCGCGGGGG AGACCGTGGCCCGGCCCTCTTCTGAGTTCCCCAAGGACTCTA GCACCTGGACTGTGGACCACCAGCTCCTGTGGGGGGAGGCCCTG CTCATCACCCAGTGCTCCAGGCCGGGAAGGCCGAAGTGACTGG CTACTTCCCCTTGGGCACATGGTACGACCTGCAGACGGTGCCAGT	
--	--	--

	AGAGGGCCCTTGGCAGCCTCCCACCCCCACCTGCAGCTCCCCGTG AGCCAGCCATCCACAGCGAGGGGCAGTGGGTGACGCTGCCGGCC CCCCTGGACACCATCAACGTCCACCTCCGGGCTGGGTACATCATC CCCCTGCAGGGCCCTGGCCTCACAACCACAGAGTCCCGCCAGCA GCCCATGGCCCTGGCTGTGGCCCTGACCAAGGGTGGGGAGGCCC GAGGGGAGCTGTTCTGGGACGATGGAGAGAGCCTGGAAGTGCTG GAGCGAGGGGCCTACACACAGGTCATCTTCTGGCCAGGAATAA CACGATCGTGAATGAGCTGGTACGTGTGACCAGTGAGGGAGCTG GCCTGCAGCTGCAGAAGGTGACTGTCCTGGGCGTGGCCACGGCG CCCCAGCAGGTCTCTCCAACGGTGTCCCTGTCTCCAACCTCACC TACAGCCCCGACACCAAGGTCCTGGACATCTGTGTCTCGCTGTTG ATGGGAGAGCAGTTTCTCGTCAGCTGGTGTTAG	
Cricket Paralysis Virus IRES (underlined) -BiP-vIGF2- GAA	<u>AAAAATGTGATCTTGCTTGTAATAACAATTTGAGAGGTTAATAAATTA</u> <u>CAAGTAGTGCTATTTTTGTATTTAGGTTAGCTATTTAGCTTTACGTTCC</u> <u>AGGATGCCTAGTGGCAGCCCCACAATATCCAGGAAGCCCTCTCTGCG</u> <u>GTTTTTCAGATTAGGTAGTCGAAAAACCTAAGAAATTTACCTGCTATG</u> AAGCTCTCCCTGGTGGCCGCGATGCTGCTGCTGCTCAGCGCGGC GCGGGCCTCTAGAACACTGTGCGGAGGGGAGCTTGTAGACACTC TTCAGTTCGTGTGTGGAGATCGCGGGTTCTCTTCTCTCGCCCCG CTTCCAGAGTTTCACGGAGGTCTAGGGGTATAGTAGAGGAGTGT TGTTTCAGGTCCTGTGACTTGCGCTCCTCGAGACCTATTGCGCG ACGCCAGCCAGGTCCGAAGGGGGCGGTGGCTCAGGTGGTGGAG GTAGCAGACCAGGGCCCCGGGATGCCCAGGCACACCCCGGCCGT CCCAGAGCAGTGCCCACACAGTGCGACGTCCCCCCCCAACAGCCG CTTCGATTGCGCCCCTGACAAGGCCATCACCCAGGAACAGTGCG AGGCCCGCGGCTGTTGCTACATCCCTGCAAAGCAGGGGCTGCAG GGAGCCCAGATGGGGCAGCCCTGGTGCTTCTTCCCACCCAGCTA CCCCAGCTACAAGCTGGAGAACCCTGAGCTCCTCTGAAATGGGCT ACACGGCCACCCTGACCCGTACCACCCCCACCTTCTTCCCCAAGG ACATCCTGACCCTGCGGCTGGACGTGATGATGGAGACTGAGAAC CGCCTCCACTTCACGATCAAAGATCCAGCTAACAGGCGCTACGA GGTGCCCTTGGAGACCCCGCATGTCCACAGCCGGGCACCGTCCC	191

	CACTCTACAGCGTGGAGTTCTCCGAGGAGCCCTTCGGGGTGATC GTGCGCCGGCAGCTGGACGGCCGCGTGCTGCTGAACACGACGGT GGCGCCCCTGTTCTTTGCGGACCAGTTCCTTCAGCTGTCCACCTC GCTGCCCTCGCAGTATATCACAGGCCTCGCCGAGCACCTCAGTCC CCTGATGCTCAGCACCACTGGACCAAGGATCACCTGTGGAACC GGGACCTTGCGCCACGCCCCGGTGCGAACCTCTACGGGTCTCAC CCTTTCTACCTGGCGCTGGAGGACGGCGGGTCGGCACACGGGGT GTTCTGCTAAACAGCAATGCCATGGATGTGGTCTGCAGCCGA GCCCTGCCCTTAGCTGGAGGTCGACAGGTGGGATCCTGGATGTC TACATCTTCTGGGCCCAGAGCCCAAGAGCGTGGTGCAGCAGTA CCTGGACGTTGTGGGATACCCGTTTCATGCCGCCATACTGGGGCCT GGGCTTCCACCTGTGCCGCTGGGGCTACTCCTCCACCGCTATCAC CCGCCAGGTGGTGGAGAACATGACCAGGGCCCACTTCCCCCTGG ACGTCCAGTGGAACGACCTGGACTACATGGACTCCCGGAGGGAC TTCACGTTCAACAAGGATGGCTTCCGGGACTTCCCGGCCATGGTG CAGGAGCTGCACCAGGGCGGCCGGCGCTACATGATGATCGTGGA TCCTGCCATCAGCAGCTCGGGCCCTGCCGGGAGCTACAGGCCCT ACGACGAGGGTCTGCGGAGGGGGGTTTTTCATCACCAACGAGACC GGCCAGCCGCTGATTGGGAAGGTATGGCCCGGGTCCACTGCCTT CCCCGACTTCACCAACCCACAGCCCTGGCCTGGTGGGAGGACA TGGTGGCTGAGTTCCATGACCAGGTGCCCTTCGACGGCATGTGG ATTGACATGAACGAGCCTTCCAACCTTCATCAGGGGCTCTGAGGA CGGCTGCCCCAACAATGAGCTGGAGAACCACCTACGTGCCTG GGGTGGTTGGGGGGACCCCTCCAGGCGGCCACCATCTGTGCCTCC AGCCACCAGTTTCTCTCCACACACTACAACCTGCACAACCTCTAC GGCCTGACCGAAGCCATCGCCTCCCACAGGGCGCTGGTGAAGGC TCGGGGGACACGCCCATTGTGATCTCCCGCTCGACCTTTGCTGG CCACGGCCGATACGCCGGCCACTGGACGGGGGACGTGTGGAGCT CCTGGGAGCAGCTCGCCTCCTCCGTGCCAGAAATCCTGCAGTTTA ACCTGCTGGGGGTGCCTCTGGTTCGGGGCCGACGTCTGCGGCTTCC TGGGCAACACCTCAGAGGAGCTGTGTGTGCGCTGGACCCAGCTG GGGGCCTTCTACCCCTTCATGCGGAACCACAACAGCCTGCTCAGT	
--	---	--

	CTGCCCCAGGAGCCGTACAGCTTCAGCGAGCCGGCCCCAGCAGGC CATGAGGAAGGCCCTCACCCTGCGCTACGCACTCCTCCCCACCT CTACACACTGTTCCACCAGGCCACGTGCGGGGGAGACCGTGG CCCGGCCCTCTTCCTGGAGTTCCCCAAGGACTCTAGCACCTGGA CTGTGGACCACCAGCTCCTGTGGGGGGAGGCCCTGCTCATCACC CCAGTGCTCCAGGCCGGGAAGGCCGAAGTGACTGGCTACTTCCC CTTGGGCACATGGTACGACCTGCAGACGGTGCCAGTAGAGGCCC TTGGCAGCCTCCCACCCCCACCTGCAGCTCCCCGTGAGCCAGCCA TCCACAGCGAGGGGCAGTGGGTGACGCTGCCGGCCCCCTGGAC ACCATCAACGTCCACCTCCGGGCTGGGTACATCATCCCCCTGCAG GGCCCTGGCCTCACAACCACAGAGTCCCGCCAGCAGCCCATGGC CCTGGCTGTGGCCCTGACCAAGGGTGGGGAGGCCCGAGGGGAGC TGTTCTGGGACGATGGAGAGAGCCTGGAAGTGCTGGAGCGAGGG GCCTACACACAGGTCATCTTCCTGGCCAGGAATAACACGATCGT GAATGAGCTGGTACGTGTGACCAGTGAGGGAGCTGGCCTGCAGC TGCAGAAGGTGACTGTCCTGGGCGTGGCCACGGCGCCCCAGCAG GTCCTCTCCAACGGTGTCCCTGTCTCCAACCTTCACCTACAGCCCC GACACCAAGGTCCTGGACATCTGTGTCTCGCTGTTGATGGGAGA GCAGTTTCTCGTCAGCTGGTGT TAG	
wt-PPT1 IDT codon optimized	ATGGCATCACCGGGTTGCCTCTGGTTGTTGGCCGTTGCGTTGCTT CCGTGGACATGTGCATCAAGAGCTCTTCAACATCTGGATCCCCCA GCTCCCCTGCCGCTCGTAATCTGGCACGGGATGGGGGATTTCATGT TGTAACCCGTTGTCAATGGGCGCGATAAAAAAGATGGTTGAAAA GAAGATTCCAGGCATCTACGTTCTGTCCCTGGAAATCGGTAAGA CACTGATGGAAGACGTGGAGAACTCCTTCTTTCTCAACGTCAATA GTCAGGTCACTACCGTCTGTCAAGCATTGGCAAAGGACCCTAAA CTTCAGCAGGGGTACAATGCGATGGGGTTTAGCCAGGGCGGACA GTTTCTTAGAGCCGTCGCACAGCGCTGTCCATCTCCCCGATGAT TAACCTTATATCTGTGCGGGGACAACACCAGGGTGTTTTTGGTCT TCCTCGCTGTCTGGTGAAAGCTCCACATCTGTGATTTTCATACG CAAAACGTTGAACGCAGGAGCTTATAGTAAAGTCGTCCAAGAAC GGCTTGTTCAAGCGGAGTATTGGCATGACCCAATAAAAGAAGAC	192

	GTTTATAGGAATCACTCTATCTTCTTGGCCGATATCAACCAAGAA CGCGGAATCAACGAAAGCTACAAAAAGAATCTTATGGCTCTCAA GAAATTTGTTATGGTGAAATTCCTTAATGACTCTATAGTAGATCC TGTCGATTGAGAATGGTTCGGGTTCTACAGGTCTGGCCAGGCGA AGGAGACTATTCCCCTCCAAGAAACGTCTCTCTATACACAAGAC AGACTCGGACTGAAAGAGATGGATAATGCGGGCCAGTTGGTCTT CTTGGCTACGGAAGGCGATCATCTCCAACCTCTCCGAAGAGTGGT TCTATGCCCATATAATCCCGTTCCTGGGCTAA	
PPT1-2 (wt- vIGF2- PPT1; Codon optimized by IDT codon optimization tool)	ATGGCATCCCCCGGATGTTTGTGGCTGCTGGCGGTTGCGCTTCTG CCATGGACGTGCGCCTCCCGAGCCCTCCAACACCTGTCCAGGAC ACTTTGCGGCGGAGAGTTGGTCGATACGCTTCAATTCGTGTGTGG GGATAGAGGCTTCCTTTTTTCTCGGCCCGCTAGCCGCGTGTCCCG AAGGTCCCGGGGTATCGTTGAGGAATGCTGTTTCCGGTCTGCG ATCTTGCACTGTTGGAGACATACTGTGCTACGCCTGCGAGAAGC GAGGGTGGAGGGGGTTCTGGAGGTGGAGGGAGCCGGCCTCGGG CGGTTCCCACCCAGGATCCTCCAGCTCCTCTGCCTCTGGTCATCT GGCATGGGATGGGGGACTCATGTTGTAACCCGCTGAGTATGGGG GCAATTAAAAAATGGTTGAAAAGAAAATTCAGGTATTTATGT CCTCTCTCTTGAAATCGGTAAGACACTTATGGAGGATGTGGA AACTCCTTTTTCTTAATGTCAATTCTCAGGTCACAACAGTTTGTCA GGCTCTGGCGAAGGATCCTAAGCTGCAGCAAGGCTACAACGCCA TGGGTTTTTCCCAGGGAGGCCAATTTCTCAGAGCGGTAGCTCAGC GATGTCCATCACCACCGATGATAAATCTGATCAGTGTGCGCGGA CAACACCAGGGAGTTTTTCGGGCTGCCCAGGTGTCCGGGGGAATC TAGTCACATATGTGACTTCATTCGCAAGACCCTTAACGCCGGCGC TACTCAAAGGTGGTTCAAGAACGGCTTGTGCAGGCTGAATACT GGCACGATCCCATCAAGGAAGATGTATATAGGAACCACAGTATC TTTCTGGCAGACATAAATCAGGAAAGGGGTATTAACGAAAGCTA CAAGAAAAATCTCATGGCCCTGAAGAAATTTGTAATGGTTAAGT TTTTGAACGATTCTATAGTAGATCCTGTTGACTCCGAGTGGTTTCG GGTTCTATCGATCTGGTCAAGCCAAGGAGACGATTCCGCTTCAG GAAACTTCACTGTACACACAGGATCGGCTGGGACTCAAGGAGAT	193

	GGACAATGCGGGCCAGTTGGTGTCTTCTGGCTACAGAGGGAGACC ATCTCCAGTTGAGTGAAGAATGGTTCTATGCACATATTATCCCAT TCCTCGGCTAA	
PPT1-29 (BiP2aa- vIGF2- PPT1; native human sequence)	ATGAAGCTCTCCCTGGTGGCCGCGATGCTGCTGCTGCTCTGGGTG GCACTGCTGCTGCTCAGCGCGGCGAGGGCCGCGAGTCGCAC GTTGTGTGGAGGTGAACTCGTCGACACCCTTCAGTTCGTATGTGG AGATCGCGGTTTCTCTTCTCACGCCAGCTTCCAGAGTTTCCCG AAGATCACGAGGAATAGTTGAGGAGTGCTGTTTTCTGGTCTTGTG ATCTGGCTCTCCTCGAGACTTATTGTGCTACGCCGGCCCGCTCTG AAGGAGGTGGTGGCAGTGGAGGAGGAGGGAGTCGGCCTAGGGC AGTCCCAACCCAGGACCCGCGGCGCCGCTGCCGTTGGTGATCT GGCATGGGATGGGAGACAGCTGTTGCAATCCCTTAAGCATGGGT GCTATTAAAAAATGGTGGAGAAGAAAATACCTGGAATTTACGT CTTATCTTTAGAGATTGGGAAGACCCTGATGGAGGACGTGGAGA ACAGCTTCTTCTTGAATGTCAATTCCCAAGTAACAACAGTGTGTC AGGCACTTGCTAAGGATCCTAAATTGCAGCAAGGCTACAATGCT ATGGGATTCTCCCAGGGAGGCCAATTTCTGAGGGCAGTGGCTCA GAGATGCCCTTCACCTCCCATGATCAATCTGATCTCGGTTGGGGG ACAACATCAAGGTGTTTTTGGACTCCCTCGATGCCCAGGAGAGA GCTCTCACATCTGTGACTTCATCCGAAAAACACTGAATGCTGGG GCGTACTCCAAAGTTGTTTCAAGAACGCCTCGTGCAAGCCGAATA CTGGCATGACCCCATAAAGGAGGATGTGTATCGCAACCACAGCA TCTTCTTGGCAGATATAAATCAGGAGCGGGGTATCAATGAGTCC TACAAGAAAAACCTGATGGCCCTGAAGAAGTTTGTGATGGTGAA ATTCCTCAATGATTCCATTGTGGACCCTGTAGATTCGGAGTGGTT TGGATTTTACAGAAGTGGCCAAGCCAAGGAAACCATTCCCTTAC AGGAGACCTCCCTGTACACACAGGACCGCCTGGGGCTAAAGGAA ATGGACAATGCAGGACAGCTAGTGTCTTCTGGCTACAGAAGGGGA CCATCTTCAGTTGTCTGAAGAATGGTTTTATGCCACATCATACC ATTCCTTGATGA	194
PPT1 engineered	ATGGCGTCGCCCCGGCTGCCTGTGGCTCTTGGCTGTGGCTCTCCTG CCATGGACCTGCGCTTCTCGGGCGCTGCAGCATCTGGACCCGCC	195

	GGCGCCGCTGCCGTTGGTGATCTGGCATGGGATGGGAGACAGCT GTTGCAATCCCTTAAGCATGGGTGCTATTAAAAAATGGTGGAG AAGAAAATACCTGGAATTTACGTCTTATCTTTAGAGATTGGGAA GACCCTGATGGAGGACGTGGAGAACAGCTTCTTCTTGAATGTCA ATTCCCAAGTAACAACAGTGTGTCAGGCACTTGCTAAGGATCCT AAATTGCAGCAAGGCTACAATGCTATGGGATTCTCCCAGGGAGG CCAATTTCTGAGGGCAGTGGCTCAGAGATGCCCTTACCTCCCAT GATCAATCTGATCTCGGTTGGGGGACAACATCAAGGTGTTTTTG ACTCCCTCGATGCCCAGGAGAGAGCTCTCACATCTGTGACTTCAT CCGAAAAACACTGAATGCTGGGGCGTACTCCAAAGTTGTTTCAGG AACGCCTCGTGCAAGCCGAATACTGGCATGACCCCATAAAGGAG GATGTGTATCGCAACCACAGCATCTTCTTGGCAGATATAAATCA GGAGCGGGGTATCAATGAGTCCTACAAGAAAAACCTGATGGCCC TGAAGAAGTTTGTGATGGTGAAATTCCTCAATGATTCCATTGTGG ACCCTGTAGATTTCGGAGTGGTTTGGATTTTACAGAAGTGGCCAA GCCAAGGAAACCATTCCCTTACAGGAGACCTCCCTGTACACACA GGACCGCCTGGGGCTAAAGGAAATGGACAATGCAGGACAGCTA GTGTTTCTGGCTACAGAAGGGGACCATCTTCAGTTGTCTGAAGA ATGGTTTTATGCCACATCATACCATTCCTTGGAAGACCTAGAGC AGTGCCTACGCAGGGAGGGAGTGGGAGTGGATCCACTTCATCCT CTAGAACACTGTGCGGAGGGGAGCTTGTAGACACTCTTCAGTTC GTGTGTGGAGATCGCGGGTTCCTCTTCTCTCGCCCCGCTTCCAGA GTTTCACGGAGGTCTAGGGGTATAGTAGAGGAGTGTGTTTCAG GGAGTGTGACTTGGCGCTCCTCGAGACCTATTGCGCGACGCCAG CCAGGTCCGAATGA	
TPP1 wildtype	ATGGGACTCCAAGCCTGCCTCCTAGGGCTCTTTGCCCTCATCCTC TCTGGCAAATGCAGTTACAGCCCGGAGCCCGACCAGCGGAGGAC GCTGCCCCCAGGCTGGGTGTCCCTGGGCCGTGCGGACCCTGAGG AAGAGCTGAGTCTCACCTTTGCCCTGAGACAGCAGAATGTGGAA AGACTCTCGGAGCTGGTGCAGGCTGTGTGCGGATCCCAGCTCTCCT CAATACGGAAAATACCTGACCCTAGAGAATGTGGCTGATCTGGT GAGGCCATCCCCACTGACCCTCCACACGGTGCAAAAATGGCTCT	196

	TGGCAGCCGGAGCCCAGAAGTGCCATTCTGTGATCACACAGGAC TTTCTGACTTGCTGGCTGAGCATCCGACAAGCAGAGCTGCTGCTC CCTGGGGCTGAGTTTCATCACTATGTGGGAGGACCTACGGAAAC CCATGTTGTAAGGTCCCCACATCCCTACCAGCTTCCACAGGCCTT GGCCCCCATGTGGACTTTGTGGGGGGACTGCACCGTTTCCCCC AACATCATCCCTGAGGCAACGTCCTGAGCCGCAGGTGACAGGGA CTGTAGGCCTGCATCTGGGGGTAACCCCTCTGTGATCCGTAAGC GATACAACTTGACCTCACAAGACGTGGGCTCTGGCACCAGCAAT AACAGCCAAGCCTGTGCCAGTTCCTGGAGCAGTATTTCCATGA CTCAGACCTGGCTCAGTTCATGCGCCTCTTCGGTGGCAACTTTGC ACATCAGGCATCAGTAGCCCGTGTGGTTGGACAACAGGGCCGGG GCCGGGCCGGGATTGAGGCCAGTCTAGATGTGCAGTACCTGATG AGTGCTGGTGCCAACATCTCCACCTGGGTCTACAGTAGCCCTGGC CGGCATGAGGGACAGGAGCCCTTCCTGCAGTGGCTCATGCTGCT CAGTAATGAGTCAGCCCTGCCACATGTGCATACTGTGAGCTATG GAGATGATGAGGACTCCCTCAGCAGCGCCTACATCCAGCGGGTC AACACTGAGCTCATGAAGGCTGCCGCTCGGGGTCTCACCTGCT CTTCGCCTCAGGTGACAGTGGGGCCGGGTGTTGGTCTGTCTCTGG AAGACACCAGTTCCGCCCTACCTTCCCTGCCTCCAGCCCCCTATGT CACCACAGTGGGAGGCACATCCTTCCAGGAACCTTTCCTCATCAC AAATGAAATTGTTGACTATATCAGTGGTGGTGGCTTCAGCAATGT GTTCCACGGCCTTCATACCAGGAGGAAGCTGTAACGAAGTTCC TGAGCTCTAGCCCCCACCTGCCACCATCCAGTTACTTCAATGCCA GTGGCCGTGCCTACCCAGATGTGGCTGCACTTTCTGATGGCTACT GGGTGGTCAGCAACAGAGTGCCCATTCATGGGTGTCCGGAACC TCGGCCTCTACTCCAGTGTTTGGGGGGATCCTATCCTTGATCAAT GAGCACAGGATCCTTAGTGGCCGCCCCCTCTTGGCTTTCTCAAC CCAAGGCTCTACCAGCAGCATGGGGCAGGACTCTTTGATGTAAC CCGTGGCTGCCATGAGTCCTGTCTGGATGAAGAGGTAGAGGGCC AGGGTTTCTGCTCTGGTCCTGGCTGGGATCCTGTAACAGGCTGGG GAACACCCAACTTCCCAGCTTTGCTGAAGACTCTACTCAACCCCT GA	
--	---	--

TPP1 engineered	ATGGGACTCCAAGCCTGCCTCCTAGGGCTCTTTGCCCTCATCCTC TCTGGCAAATGCTCTAGAACACTGTGCGGAGGGGAGCTTGTAGA CACTCTTCAGTTCGTGTGTGGAGATCGCGGGTTCCTCTTCTCTCG CCCCGCTTCCAGAGTTTCACGGAGGTCTAGGGGTATAGTAGAGG AGTGTGTTTTCAGGTCCTGTGACTTGCGCTCCTCGAGACCTATT GCGCGACGCCAGCCAGGTCCGAAGGAGGTGGTGGCAGTGGAGG AGGAGGGAGTAGACCTAGAGCAGTGCCTACGCAGAGTTACAGCC CGGAGCCCGACCAGCGGAGGACGCTGCCCCCAGGCTGGGTGTCC CTGGGCCGTGCGGACCCTGAGGAAGAGCTGAGTCTCACCTTTGC CCTGAGACAGCAGAATGTGGAAAGACTCTCGGAGCTGGTGCAGG CTGTGTCGGATCCCAGCTCTCCTCAATACGGAAAATACCTGACCC TAGAGAATGTGGCTGATCTGGTGAGGCCATCCCCACTGACCCTC CACACGGTGCAAAAATGGCTCTTGGCAGCCGGAGCCCAGAAAGTG CCATTCTGTGATCACACAGGACTTTCTGACTTGCTGGCTGAGCAT CCGACAAGCAGAGCTGCTGCTCCCTGGGGCTGAGTTTCATCACT ATGTGGGAGGACCTACGGAAACCCATGTTGTAAGGTCCCCACAT CCCTACCAGCTTCCACAGGCCTTGGCCCCCATGTGGACTTTGTG GGGGGACTGCACCGTTTTCCCCAACATCATCCCTGAGGCAACG TCCTGAGCCGCAGGTGACAGGGACTGTAGGCCTGCATCTGGGGG TAACCCCTCTGTGATCCGTAAGCGATACAACCTGACCTCACAAG ACGTGGGCTCTGGCACCAGCAATAACAGCCAAGCCTGTGCCCAG TTCCTGGAGCAGTATTTCCATGACTCAGACCTGGCTCAGTTCATG CGCCTCTTCGGTGGCAACTTTGCACATCAGGCATCAGTAGCCCGT GTGGTTGGACAACAGGGCCGGGGCCGGGCGGGGATTGAGGCCA GTCTAGATGTGCAGTACCTGATGAGTGCTGGTGCCAACATCTCCA CCTGGGTCTACAGTAGCCCTGGCCGGCATGAGGGACAGGAGCCC TTCCTGCAGTGGCTCATGCTGCTCAGTAATGAGTCAGCCCTGCCA CATGTGCATACTGTGAGCTATGGAGATGATGAGGACTCCCTCAG CAGCGCCTACATCCAGCGGGTCAACACTGAGCTCATGAAGGCTG CCGCTCGGGGTCTCACCTGCTCTTCGCCTCAGGTGACAGTGGGG CCGGGTGTTGGTCTGTCTCTGGAAGACACCAGTTCCGCCCTACCT TCCCTGCCTCCAGCCCCTATGTCACCACAGTGGGAGGCACATCCT	197
--------------------	--	-----

	TCCAGGAACCTTTCTCATCACAAATGAAATTGTTGACTATATCA GTGGTGGTGGCTTCAGCAATGTGTTCCACGGCCTTCATACCAGG AGGAAGCTGTAACGAAGTTCCTGAGCTCTAGCCCCACCTGCCA CCATCCAGTTACTTCAATGCCAGTGGCCGTGCCTACCCAGATGTG GCTGCACTTTCTGATGGCTACTGGGTGGTCAGCAACAGAGTGCC CATTCCATGGGTGTCCGGAACCTCGGCCTCTACTCCAGTGTTTGG GGGGATCCTATCCTTGATCAATGAGCACAGGATCCTTAGTGGCC GCCCCCTCTTGGCTTTCTCAACCCAAGGCTCTACCAGCAGCATG GGGCAGGACTCTTTGATGTAACCCGTGGCTGCCATGAGTCCTGTC TGGATGAAGAGGTAGAGGGCCAGGGTTTCTGCTCTGGTCCTGGC TGGGATCCTGTAACAGGCTGGGGAACACCCAACCTTCCCAGCTTT GCTGAAGACTCTACTCAACCCCTGA	
AGA wildtype	ATGGCGCGGAAGTCGAACCTTGCTGTGCTTCTCGTGCCGTTTCTG CTCTGCCAGGCCCTAGTGCGCTGCTCCAGCCCTCTGCCCCTGGTC GTCAAACTTGGCCCTTTAAGAATGCAACCGAAGCAGCGTGGAG GGCATTAGCATCTGGAGGCTCTGCCCTGGATGCAGTGGAGAGCG GCTGTGCCATGTGTGAGAGAGAGCAGTGTGACGGCTCTGTAGGC TTTGGAGGAAGTCCTGATGAACTTGGAGAAACCACACTAGATGC CATGATCATGGATGGCACTACTATGGATGTAGGAGCAGTAGGAG ATCTCAGACGAATTA AAAATGCTATTGGTGTGGCACGGAAAGTA CTGGAACATAACAACACACACTTTTAGTAGGAGAGTCAGCCAC CACATTTGCTCAAAGTATGGGGTTTATCAATGAAGACTTATCTAC CACTGCTTCTCAAGCTCTTCATTCAGATTGGCTTGCTCGGAATTG CCAGCCAAATTATTGGAGGAATGTTATACCAGATCCCTCAAAAT ACTGCGGACCCTACAAACCACCTGGTATCTTAAAGCAGGATATT CCTATCCATAAAGAAACAGAAGATGATCGTGGTCATGACACTAT TGGCATGGTTGTAATCCATAAGACAGGACATATTGCTGCTGGTA CATCTACAAATGGTATAAAATTCAAAATACATGGCCGTGTAGGA GACTCACCAATACCTGGAGCTGGAGCCTATGCTGACGATACTGC AGGGGCAGCCGCAGCCACTGGGAATGGTGATATATTGATGCGCT TCCTGCCAAGCTACCAAGCTGTAGAATACATGAGAAGAGGAGAA GATCCAACCATAGCTTGCCAAAAAGTGATTTCAAGAATCCAGAA	198

	GCATTTTCCAGAATTCTTTGGGGCTGTTATATGTGCCAATGTGAC TGGAAGTTACGGTGCTGCTTGCAATAAACTTTCAACATTTACTCA GTTTAGTTTCATGGTTTATAATTCCGAAAAAATCAGCCAACTGA GGAAAAAGTGGACTGCATCTAA	
AGA engineered (N-terminal fusion)	ATGGCGCGGAAGTCGAACTTGCCGTGTGCTTCTCGTGCCGTTTCTG CTCTGCCAGGCCCTAGTGCGCTGCTCTAGAACACTGTGCGGAGG GGAGCTTGTAGACACTCTTCAGTTCGTGTGTGGAGATCGCGGGTT CCTCTTCTCTCGCCCCGCTTCCAGAGTTTCACGGAGGTCTAGGGG TATAGTAGAGGAGTGTTGTTTCAGGTCCTGTGACTTGGCGCTCCT CGAGACCTATTGCGCGACGCCAGCCAGGTCCGAAGGAGGTGGTG GCAGTGGAGGAGGAGGGAGTAGACCTAGAGCAGTGCCTACGCA GTCCAGCCCTCTGCCCCCTGGTCGTCAACACTTGGCCCTTTAAGAA TGCAACCGAAGCAGCGTGGAGGGCATTAGCATCTGGAGGCTCTG CCCTGGATGCAGTGGAGAGCGGCTGTGCCATGTGTGAGAGAGAG CAGTGTGACGGCTCTGTAGGCTTTGGAGGAAGTCCTGATGAACT TGGAGAAACCACACTAGATGCCATGATCATGGATGGCACTACTA TGGATGTAGGAGCAGTAGGAGATCTCAGACGAATTA AAAATGCT ATTGGTGTGGCACGGAAAGTACTGGAACATACAACACACACACT TTAGTAGGAGAGTCAGCCACCACATTTGCTCAAAGTATGGGGT TTATCAATGAAGACTTATCTACCACTGCTTCTCAAGCTCTTCATT CAGATTGGCTTGCTCGGAATTGCCAGCCAAATTATTGGAGGAAT GTTATACCAGATCCCTCAAATACTGCGGACCCTACAAACCACC TGGTATCTTAAAGCAGGATATTCCTATCCATAAAGAAACAGAAG ATGATCGTGGTCATGACACTATTGGCATGGTTGTAATCCATAAGA CAGGACATATTGCTGCTGGTACATCTACAAATGGTATAAAATTC AAAATACATGGCCGTGTAGGAGACTCACCAATACCTGGAGCTGG AGCCTATGCTGACGATACTGCAGGGGCAGCCGCAGCCACTGGGA ATGGTGATATATTGATGCGCTTCCTGCCAAGCTACCAAGCTGTAG AATACATGAGAAGAGGAGAAGATCCAACCATAGCTTGCCAAAA AGTGATTTCAAGAATCCAGAAGCATTTTCCAGAATTCCTTTGGGGC TGTTATATGTGCCAATGTGACTGGAAGTTACGGTGCTGCTTGCAA TAACTTTCAACATTTACTCAGTTTAGTTTCATGGTTTATAATTCC	199

	GAAAAAAATCAGCCAACTGAGGAAAAAGTGGACTGCATCTAA	
GLA wildtype	ATGAAGCTCTCCCTGGTGGCCGCGATGCTGCTGCTGCTCAGCGCG GCGCGGGCCCTGGACAATGGATTGGCAAGGACGCCTACCATGGG CTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGG AAGAGCCAGATTCCCTGCATCAGTGAGAAGCTCTTCATGGAGATG GCAGAGCTCATGGTCTCAGAAGGCTGGAAGGATGCAGGTTATGA GTACCTCTGCATTGATGACTGTTGGATGGCTCCCCAAAGAGATTC AGAAGGCAGACTTCAGGCAGACCCTCAGCGCTTTCCTCATGGGA TTCGCCAGCTAGCTAATTATGTTACAGCAAAGGACTGAAGCTA GGGATTTATGCAGATGTTGGAAATAAAACCTGCGCAGGCTTCCC TGGGAGTTTTGGATACTACGACATTGATGCCCAGACCTTTGCTGA CTGGGGAGTAGATCTGCTAAAATTTGATGGTTGTTACTGTGACAG TTTGGAAAATTTGGCAGATGGTTATAAGCACATGTCCTTGGCCCT GAATAGGACTGGCAGAAGCATTGTGTACTCCTGTGAGTGGCCTC TTTATATGTGGCCCTTTCAAAGCCCAATTATACAGAAATCCGAC AGTACTGCAATCACTGGCGAAATTTTGCTGACATTGATGATTCCT GGAAAAGTATAAAGAGTATCTTGGACTGGACATCTTTTAACCAG GAGAGAATTGTTGATGTTGCTGGACCAGGGGGTTGGAATGACCC AGATATGTTAGTGATTGGCAACTTTGGCCTCAGCTGGAATCAGC AAGTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTAT TCATGTCTAATGACCTCCGACACATCAGCCCTCAAGCCAAAGCTC TCCTTCAGGATAAGGACGTAATTGCCATCAATCAGGACCCCTTG GGCAAGCAAGGGTACCAGCTTAGACAGGGAGACAACCTTTGAAGT GTGGGAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGAT AAACCGGCAGGAGATTGGTGGACCTCGCTCTTATACCATCGCAG TTGCTTCCCTGGGTAAAGGAGTGGCCTGTAATCCTGCCTGCTTCA TCACACAGCTCCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAA TGGACTTCAAGGTTAAGAAGTCACATAAATCCCACAGGCACTGT TTTGCTTCAGCTAGAAAATACAATGCAGATGTCATTAAGACTT ACTTTAA	200
GLA engineered	ATGAAGCTCTCCCTGGTGGCCGCGATGCTGCTGCTGCTCAGCGCG GCGCGGGCCCTGGACAATGGATTGGCAAGGACGCCTACCATGGG	201

CTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGG AAGAGCCAGATTCCTGCATCAGTGAGAAGCTCTTCATGGAGATG GCAGAGCTCATGGTCTCAGAAGGCTGGAAGGATGCAGGTTATGA GTACCTCTGCATTGATGACTGTTGGATGGCTCCCCAAAGAGATTC AGAAGGCAGACTTCAGGCAGACCCTCAGCGCTTTCCTCATGGGA TTCGCCAGCTAGCTAATTATGTTACAGCAAAGGACTGAAGCTA GGGATTTATGCAGATGTTGGAAATAAAACCTGCGCAGGCTTCCC TGGGAGTTTTGGATACTACGACATTGATGCCCAGACCTTTGCTGA CTGGGGAGTAGATCTGCTAAAATTTGATGGTTGTTACTGTGACAG TTTGGAAAATTTGGCAGATGGTTATAAGCACATGTCCTTGGCCCT GAATAGGACTGGCAGAAGCATTGTGTACTCCTGTGAGTGGCCTC TTTATATGTGGCCCTTTCAAAGCCCAATTATACAGAAATCCGAC AGTACTGCAATCACTGGCGAAATTTTGCTGACATTGATGATTCTC GGAAAAGTATAAAGAGTATCTTGGACTGGACATCTTTTAACCAG GAGAGAATTGTTGATGTTGCTGGACCAGGGGGTTGGAATGACCC AGATATGTTAGTGATTGGCAACTTTGGCCTCAGCTGGAATCAGC AAGTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTAT TCATGTCTAATGACCTCCGACACATCAGCCCTCAAGCCAAAGCTC TCCTTCAGGATAAGGACGTAATTGCCATCAATCAGGACCCCTTG GGCAAGCAAGGGTACCAGCTTAGACAGGGAGACAACCTTTGAAGT GTGGGAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGAT AAACCGGCAGGAGATTGGTGGACCTCGCTCTTATACCATCGCAG TTGCTTCCCTGGGTAAAGGAGTGGCCTGTAATCCTGCCTGCTTCA TCACACAGCTCCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAA TGGACTTCAAGGTTAAGAAGTCACATAAATCCCACAGGCACTGT TTTGCTTCAGCTAGAAAATACAATGCAGATGTCATTAAGACTT ACTTTACATCCCTGCAAAGCAGGGGCTGCAGGGAGCCCAGATGG GGCAGCCCGGGGGCGGTGGCTCAGGTGGTGGAGGTTCAAGAAC ACTGTGCGGAGGGGAGCTTGTAGACACTCTTCAGTTCGTGTGTG GAGATCGCGGGTTCCTCTCTCGCCCCGCTTCCAGAGTTTCAC GGAGGTCTAGGGGTATAGTAGAGGAGTGTTGTTTCAGGTCCTGT GACTTGGCGCTCCTCGAGACCTATTGCGCGACGCCAGCCAGGTC	
---	--

	CGAATAA	
BiP-vIGF2- 17-2GS- GAA	ATGAAGCTCTCCCTGGTGGCCGCGATGCTGCTGCTGCTCAGCGCG GCGCGGGCCTCTAGAACACTGTGCGGAGGGGAGCTTGTAGACAC TCTTCAGTTCGTGTGTGGAGATCGCGGGTTCCTCTTCTCTCGCCC CGCTTCCAGAGTTTCACGGAGGTCTAGGGGTATAGTAGAGGAGT GTTGTTTCAGGGAGTGTGACTTGGCGCTCCTCGAGACCTATTGCG CGACGCCAGCCAGGTCCGAAGGGGGCGGTGGCTCAGGTGGTGG AGGTAGCAGACCAGGGCCCCGGGATGCCCAGGCACACCCCGGCC GTCCCAGAGCAGTGCCACACAGTGCGACGTCCCCCAACAGC CGCTTCGATTGCGCCCCCTGACAAGGCCATCACCCAGGAACAGTG CGAGGGCCCGCGGCTGTTGCTACATCCCTGCAAAGCAGGGGGCTGC AGGGAGCCCAGATGGGGCAGCCCTGGTGCTTCTTCCCACCCAGC TACCCAGCTACAAGCTGGAGAACCTGAGCTCCTCTGAAATGGG CTACACGGCCACCCTGACCCGTACACCCCCACCTTCTTCCCCAA GGACATCCTGACCCTGCGGCTGGACGTGATGATGGAGACTGAGA ACCGCCTCCACTTCACGATCAAAGATCCAGCTAACAGGCGCTAC GAGGTGCCCTTGAGAGACCCCGCATGTCCACAGCCGGGCACCGTC CCCACTCTACAGCGTGAGTTCTCCGAGGAGCCCTTCGGGGTGTA TCGTGCGCCGGCAGCTGGACGGCCGCTGCTGCTGAACACGACG GTGGCGCCCCCTGTTCTTTGCGGACCAGTTCCTTCAGCTGTCCACC TCGCTGCCCTCGCAGTATATCACAGGCCTCGCCGAGCACCTCAGT CCCCTGATGCTCAGCACCAGCTGGACCAGGATCACCTGTGGAA CCGGGACCTTGCGCCACGCCCCGGTGCGAACCTCTACGGGTCTC ACCCTTTCTACCTGGCGCTGGAGGACGGCGGGTCGGCACACGGG GTGTTCTTGCTAAACAGCAATGCCATGGATGTGGTCCTGCAGCC GAGCCCTGCCCTTAGCTGGAGGTCGACAGGTGGGATCCTGGATG TCTACATCTTCTGGGCCCAGAGCCCAAGAGCGTGGTGCAGCAG TACCTGGACGTTGTGGGATAACCGTTCATGCCGCCATACTGGGGC CTGGGCTTCCACCTGTGCCGCTGGGGCTACTCCTCCACCGCTATC ACCCGCCAGGTGGTGGAGAACATGACCAGGGCCCACTTCCCCCT GGACGTCCAGTGGAACGACCTGGACTACATGGACTCCCGGAGGG ACTTCACGTTCAACAAGGATGGCTTCCGGGACTTCCCGGCCATG	202

	GTGCAGGAGCTGCACCAGGGCGGCCGGCGCTACATGATGATCGT GGATCCTGCCATCAGCAGCTCGGGCCCTGCCGGGAGCTACAGGC CCTACGACGAGGGTCTGCGGAGGGGGGTTTTCATCACCAACGAG ACCGGCCAGCCGCTGATTGGGAAGGTATGGCCCCGGGTCCACTGC CTTCCCCGACTTCACCAACCCACAGCCCTGGCCTGGTGGGAGG ACATGGTGGCTGAGTTCCATGACCAGGTGCCCTTCGACGGCATG TGGATTGACATGAACGAGCCTTCCAACCTTCATCAGGGGCTCTGA GGACGGCTGCCCCAACAATGAGCTGGAGAACCACCCCTACGTGC CTGGGGTGGTTGGGGGGACCCTCCAGGCGGCCACCATCTGTGCC TCCAGCCACCAGTTTCTCTCCACACACTACAACCTGCACAACCTC TACGGCCTGACCGAAGCCATCGCCTCCCACAGGGCGCTGGTGAA GGCTCGGGGGACACGCCCATTTGTGATCTCCCGCTCGACCTTTGC TGGCCACGGCCGATACGCCGGCCACTGGACGGGGGACGTGTGGA GCTCCTGGGAGCAGCTCGCCTCCTCCGTGCCAGAAATCCTGCAGT TTAACCTGCTGGGGGTGCCTCTGGTCGGGGCCGACGTCTGCGGCT TCCTGGGCAACACCTCAGAGGAGCTGTGTGTGCGCTGGACCCAG CTGGGGGCCTTCTACCCCTTCATGCGGAACCACAACAGCCTGCTC AGTCTGCCCCAGGAGCCGTACAGCTTCAGCGAGCCGGCCAGCA GGCCATGAGGAAGGCCCTCACCCCTGCGCTACGCACTCCTCCCC ACCTCTACACACTGTTCCACCAGGCCACGTCGCGGGGGAGACC GTGGCCCGGCCCTCTTCTGAGTTCCCCAAGGACTCTAGCACC TGGACTGTGGACCACCAGCTCCTGTGGGGGGAGGCCCTGCTCAT CACCCAGTGCTCCAGGCCGGGAAGGCCGAAGTGACTGGCTACT TCCCCTTGGGCACATGGTACGACCTGCAGACGGTGCCAGTAGAG GCCCTTGGCAGCCTCCCACCCCCACCTGCAGCTCCCCGTGAGCCA GCCATCCACAGCGAGGGGAGTGGGTGACGCTGCCGGCCCCCT GGACACCATCAACGTCCACCTCCGGGCTGGGTACATCATCCCCCT GCAGGGCCCTGGCCTCACAACCACAGAGTCCCGCCAGCAGCCCA TGGCCCTGGCTGTGGCCCTGACCAAGGGTGGGGAGGCCCGAGGG GAGCTGTTCTGGGACGATGGAGAGAGCCTGGAAGTGCTGGAGCG AGGGGCCTACACACAGGTCATCTTCTGGCCAGGAATAACACGA TCGTGAATGAGCTGGTACGTGTGACCAGTGAGGGAGCTGGCCTG	
--	--	--

	CAGCTGCAGAAGGTGACTGTCCTGGGCGTGGCCACGGCGCCCCA GCAGGTCCTCTCCAACGGTGTCCCTGTCTCCAATTACCTACAG CCCCGACACCAAGGTCCTGGACATCTGTGTCTCGCTGTTGATGGG AGAGCAGTTTCTCGTCAGCTGGTGTTAG	
BiP-vIGF2- 20-2GS- GAA	ATGAAGCTCTCCCTGGTGGCCGCGATGCTGCTGCTGCTCAGCGCG GCGCGGGCCTCTAGAACACTGTGCGGAGGGGAGCTTGTAGACAC TCTTCAGTTCGTGTGTGGAGATCGCGGGTTCTTCTCTCGCCC CGCTTCCAGAGTTTCACGGAGGTCTAGGGGTATACTGGAGGAGT GTTGTTTCAGGTCCTGTGACTTGCGCTCCTCGAGACCTATTGCG CGACGCCAGCCAGGTCCGAAGGGGGCGGTGGCTCAGGTGGTGG AGGTAGCAGACCAGGGCCCCGGGATGCCCAGGCACACCCCGGCC GTCCCAGAGCAGTGCCACACAGTGCGACGTCCCCCAACAGC CGCTTCGATTGCGCCCCTGACAAGGCCATCACCCAGGAACAGTG CGAGGCCCCGCGGCTGTTGCTACATCCCTGCAAAGCAGGGGCTGC AGGGAGCCCAGATGGGGCAGCCCTGGTGCTTCTTCCCACCCAGC TACCCAGCTACAAGCTGGAGAACCTGAGCTCCTCTGAAATGGG CTACACGGCCACCCTGACCCGTACCACCCCACTTCTTCCCCAA GGACATCCTGACCCTGCGGCTGGACGTGATGATGGAGACTGAGA ACCGCCTCCACTTCACGATCAAAGATCCAGCTAACAGGCGCTAC GAGGTGCCCTTGGAGACCCCGCATGTCCACAGCCGGGCACCGTC CCCACTCTACAGCGTGGAGTTCTCCGAGGAGCCCTTCGGGGTGA TCGTGCGCCGGCAGCTGGACGGCCGCGTGCTGCTGAACACGACG GTGGCGCCCCTGTTCTTTGCGGACCAGTTCCTTCAGCTGTCCACC TCGCTGCCCTCGCAGTATATCACAGGCCTCGCCGAGCACCTCAGT CCCCTGATGCTCAGCACCAGCTGGACCAGGATCACCTGTGGAA CCGGGACCTTGCGCCACGCCCCGGTGCGAACCTCTACGGGTCTC ACCCTTTCTACCTGGCGCTGGAGGACGGCGGGTCGGCACACGGG GTGTTCTTGCTAAACAGCAATGCCATGGATGTGGTCTCTGCAGCC GAGCCCTGCCCTTAGCTGGAGGTCGACAGGTGGGATCCTGGATG TCTACATCTTCTGGGCCAGAGCCCAAGAGCGTGGTGCAGCAG TACCTGGACGTTGTGGGATACCCGTTTCATGCCGCCATACTGGGGC CTGGGCTTCCACCTGTGCCGCTGGGGCTACTCCTCCACCGCTATC	203

	ACCCGCCAGGTGGTGGAGAACATGACCAGGGCCCACTTCCCCCT GGACGTCCAGTGGAACGACCTGGACTACATGGACTCCCGGAGGG ACTTCACGTTCAACAAGGATGGCTTCCGGGACTTCCCGGCCATG GTGCAGGAGCTGCACCAGGGCGGCCGGCGCTACATGATGATCGT GGATCCTGCCATCAGCAGCTCGGGCCCTGCCGGGAGCTACAGGC CCTACGACGAGGGTCTGCGGAGGGGGGTTTTTCATCACCAACGAG ACCGGCCAGCCGCTGATTGGGAAGGTATGGCCCGGGTCCACTGC CTTCCCCGACTTCACCAACCCACAGCCCTGGCCTGGTGGGAGG ACATGGTGGCTGAGTTCCATGACCAGGTGCCCTTCGACGGCATG TGGATTGACATGAACGAGCCTTCCAACCTTCATCAGGGGCTCTGA GGACGGCTGCCCCAACAATGAGCTGGAGAACCCACCCTACGTGC CTGGGGTGGTTGGGGGGACCCTCCAGGCGGCCACCATCTGTGCC TCCAGCCACCAGTTTCTCTCCACACACTACAACCTGCACAACCTC TACGGCCTGACCGAAGCCATCGCCTCCCACAGGGCGCTGGTGAA GGCTCGGGGGACACGCCCATTTGTGATCTCCCGCTCGACCTTTCG TGGCCACGGCCGATACGCCGGCCACTGGACGGGGGACGTGTGGA GCTCCTGGGAGCAGCTCGCCTCCTCCGTGCCAGAAATCCTGCAGT TTAACCTGCTGGGGGTGCCTCTGGTCGGGGCCGACGTCTGCGGCT TCCTGGGCAACACCTCAGAGGAGCTGTGTGTGCGCTGGACCCAG CTGGGGGGCCTTCTACCCCTTCATGCGGAACCACAACAGCCTGCTC AGTCTGCCCCAGGAGCCGTACAGCTTCAGCGAGCCGGCCAGCA GGCCATGAGGAAGGCCCTCACCTGCGCTACGCACTCCTCCCCC ACCTCTACACACTGTTCCACCAGGCCACGTCGCGGGGGAGACC GTGGCCCGGCCCTCTTCCTGGAGTTCCCCAAGGACTCTAGCACC TGGACTGTGGACCACCAGCTCCTGTGGGGGGAGGCCCTGCTCAT CACCCAGTGCTCCAGGCCGGGAAGGCCGAAGTGACTGGCTACT TCCCCTTGGGCACATGGTACGACCTGCAGACGGTGCCAGTAGAG GCCCTTGGCAGCCTCCCACCCCCACCTGCAGCTCCCCGTGAGCCA GCCATCCACAGCGAGGGGCAGTGGGTGACGCTGCCGGCCCCCCT GGACACCATCAACGTCCACCTCCGGGCTGGGTACATCATCCCCCT GCAGGGCCCTGGCCTCACAACCACAGAGTCCCGCCAGCAGCCCA TGGCCCTGGCTGTGGCCCTGACCAAGGGTGGGGAGGCCCGAGGG	
--	---	--

	GAGCTGTTCTGGGACGATGGAGAGAGCCTGGAAGTGCTGGAGCG AGGGGCCTACACACAGGTCATCTTCCTGGCCAGGAATAACACGA TCGTGAATGAGCTGGTACGTGTGACCAGTGAGGGAGCTGGCCTG CAGCTGCAGAAGGTGACTGTCCTGGGCGTGGCCACGGCGCCCCA GCAGGTCTCTCCAACGGTGTCCCTGTCTCCAACCTTCACCTACAG CCCCGACACCAAGGTCCTGGACATCTGTGTCTCGCTGTTGATGGG AGAGCAGTTTCTCGTCAGCTGGTGTAG	
BiP-vIGF2- 22-2GS- GAA	ATGAAGCTCTCCCTGGTGGCCGCGATGCTGCTGCTGCTCAGCGCG GCGCGGGCCTCTAGAACACTGTGCGGAGGGGAGCTTGTAGACAC TCTTCAGTTCGTGTGTGGAGATCGCGGGTTCCTCTTCTCTCGCGG AGGTGGAGGTTCTAGGGGTATAGTAGAGGAGTGTTGTTTCAGGT CCTGTGACTTGGCGCTCCTCGAGACCTATTGCGCGACGCCAGCCA GGTCCGAAGGGGGCGGTGGCTCAGGTGGTGGAGGTAGCAGACC AGGGCCCCGGGATGCCCAGGCACACCCCGGCCGTCCCAGAGCAG TGCCACACAGTGCGACGTCCCCCCCCAACAGCCGCTTCGATTGC GCCCCTGACAAGGCCATCACCCAGGAACAGTGCGAGGCCCGCGG CTGTTGCTACATCCCTGCAAAGCAGGGGCTGCAGGGAGCCCAGA TGGGGCAGCCCTGGTGCTTCTTCCCACCCAGCTACCCAGCTACA AGCTGGAGAACCTGAGCTCCTCTGAAATGGGCTACACGGCCACC CTGACCCGTACCACCCCCACCTTCTTCCCCAAGGACATCCTGACC CTGCGGCTGGACGTGATGATGGAGACTGAGAACCGCCTCCACTT CACGATCAAAGATCCAGCTAACAGGCGCTACGAGGTGCCCTTGG AGACCCCGCATGTCCACAGCCGGGCACCGTCCCCACTCTACAGC GTGGAGTTCTCCGAGGAGCCCTTCGGGGTGATCGTGCGCCGGCA GCTGGACGGCCGCGTGCTGCTGAACACGACGGTGGCGCCCCCTGT TCTTTGCGGACCAGTTCCTTCAGCTGTCCACCTCGCTGCCCTCGC AGTATATCACAGGCCTCGCCGAGCACCTCAGTCCCCTGATGCTCA GCACCAGCTGGACCAGGATCACCTGTGGAACCGGGACCTTGCG CCCACGCCCCGGTGCGAACCTCTACGGGTCTCACCTTTCTACCTG GCGCTGGAGGACGGCGGGTCGGCACACGGGGTGTTCTGCTAAA CAGCAATGCCATGGATGTGGTCCTGCAGCCGAGCCCTGCCCTTA GCTGGAGGTCGACAGGTGGGATCCTGGATGTCTACATCTTCCTG	204

	GGCCCAGAGCCCAAGAGCGTGGTGCAGCAGTACCTGGACGTTGT GGGATACCCGTTTCATGCCGCCATACTGGGGCCTGGGCTTCCACCT GTGCCGCTGGGGCTACTCCTCCACCGCTATCACCCGCCAGGTGGT GGAGAACATGACCAGGGCCCACTTCCCCCTGGACGTCCAGTGGA ACGACCTGGACTACATGGACTCCCGGAGGGACTTCACGTTCAAC AAGGATGGCTTCCGGGACTTCCCGGCCATGGTGCAGGAGCTGCA CCAGGGCGGGCGCGCTACATGATGATCGTGGATCCTGCCATCA GCAGCTCGGGCCCTGCCGGGAGCTACAGGCCCTACGACGAGGGT CTGCGGAGGGGGGTTTTTCATCACCACGAGACCGGCCAGCCGCT GATTGGGAAGGTATGGCCCGGGTCCACTGCCTTCCCCGACTTCAC CAACCCACAGCCCTGGCCTGGTGGGAGGACATGGTGGCTGAGT TCCATGACCAGGTGCCCTTCGACGGCATGTGGATTGACATGAAC GAGCCTTCCAACCTTCATCAGGGGCTCTGAGGACGGCTGCCCCAA CAATGAGCTGGAGAACCCACCCTACGTGCCTGGGGTGGTTGGGG GGACCCTCCAGGCGGCCACCATCTGTGCCTCCAGCCACCAGTTTC TCTCCACACACTACAACCTGCACAACCTCTACGGCCTGACCGAA GCCATCGCCTCCCACAGGGCGCTGGTGAAGGCTCGGGGGACACG CCCATTGTGATCTCCCGCTCGACCTTTGCTGGCCACGGCCGATA CGCCGGCCACTGGACGGGGGACGTGTGGAGCTCCTGGGAGCAGC TCGCCTCCTCCGTGCCAGAAATCCTGCAGTTTAACCTGCTGGGGG TGCCTCTGGTCGGGGCCGACGTCTGCGGCTTCCTGGGCAACACCT CAGAGGAGCTGTGTGTGCGCTGGACCCAGCTGGGGGCCTTCTAC CCCTTCATGCGGAACCACAACAGCCTGCTCAGTCTGCCCCAGGA GCCGTACAGCTTCAGCGAGCCGGCCCAGCAGGCCATGAGGAAGG CCCTCACCTGCGCTACGCACTCCTCCCCACCTCTACACACTGT TCCACCAGGCCCACGTGCGGGGGGAGACCGTGGCCCCGGCCCCCTC TTCCTGGAGTTCCCCAAGGACTCTAGCACCTGGACTGTGGACCAC CAGCTCCTGTGGGGGGGAGGCCCTGCTCATCACCCAGTGCTCCA GGCCGGGAAGGCCGAAGTGACTGGCTACTTCCCCTTGGGCACAT GGTACGACCTGCAGACGGTGCCAGTAGAGGCCCTTGGCAGCCTC CCACCCCCACCTGCAGCTCCCCGTGAGCCAGCCATCCACAGCGA GGGGCAGTGGGTGACGCTGCCGGCCCCCCTGGACACCATCAACG	
--	--	--

	TCCACCTCCGGGCTGGGTACATCATCCCCCTGCAGGGCCCTGGCC TCACAACCACAGAGTCCCGCCAGCAGCCCATGGCCCTGGCTGTG GCCCTGACCAAGGGTGGGGAGGCCCGAGGGGAGCTGTTCTGGGA CGATGGAGAGAGCCTGGAAGTGCTGGAGCGAGGGGCCTACACA CAGGTCATCTTCCTGGCCAGGAATAACACGATCGTGAATGAGCT GGTACGTGTGACCAGTGAGGGAGCTGGCCTGCAGCTGCAGAAGG TGACTGTCCTGGGCGTGGCCACGGCGCCCCAGCAGGTCCTCTCC AACGGTGTCCCTGTCTCCAACCTTACCTACAGCCCCGACACCAAG GTCCTGGACATCTGTGTCTCGCTGTTGATGGGAGAGCAGTTTCTC GTCAGCTGGTGTTAG	
PPT1-3 (BiP-PPT1; Codon optimized IDT)	ATGAAACTGTCTCTGGTTGCAGCAATGCTCTTGCTGTTGAGTGCG GCCCCGCGCGGATCCACCTGCTCCCCTGCCCCCTCGTTATATGGCAT GGCATGGGAGATTCTGTTGTAATCCCCTCAGCATGGGGGCCAT CAAAAAAATGGTGGAATAAAAAAATACCTGGCATATATGTACTCT CACTTGAAATCGGTAAGACCCTTATGGAAGACGTCGAAAATTCC TTCTTTTGAACGTGAACTCACAAGTTACGACCGTCTGTCAAGCT CTCGCGAAAGACCCTAAGCTCCAGCAAGGTTATAATGCAATGGG CTTCTCACAGGGAGGTCAAGTTCTTGCGAGCGGTAGCCCAGAGGT GTCCGTCTCCGCCAATGATCAACTTGATCTCAGTGGGGGGTCAGC ACCAAGGCGTTTTTGGACTCCCTAGATGCCCTGGAGAGAGCTCTC ACATTTGCGATTTTATACGGAAGACGCTGAATGCCGGCGCGTATT CAAAGGTCGTTCAAGAGCGACTCGTCCAGGCTGAATACTGGCAC GATCCGATTAAGGAAGACGTGTATCGAAACCATTCTATCTTTCTT GCCGACATTAACCAGGAGCGAGGGATCAACGAAAGTTATAAAA AAAACCTGATGGCACTCAAGAAATTTGTAATGGTTAAATTCCTG AACGATTCAATAGTTGATCCGGTGGATTCCGAGTGGTTCGGCTTC TACCGGTCCGGTCAGGCCAAGGAAACAATCCCATTGCAAGAAAC CAGTCTCTATACTCAGGACCGCCTGGGTCTGAAAGAAATGGACA ACGCTGGCCAACTTGTTTTTCTGGCAACGGAGGGTGATCACTTGC AGCTCTCTGAAGAATGGTTTTACGCACACATCATTCTTCCTTG GTTAA	205
PPT1-4	ATGAAGTTGTCCCTCGTAGCTGCAATGTTGCTGCTCCTCAGTGCA	206

(BiP-vIGF2-PPT1; Codon optimized IDT)	GCGCGGGCAAGTCGCACGTTGTGTGGAGGTGAACTCGTCGACAC CCTTCAGTTCGTATGTGGAGATCGCGGTTTCCTCTTCTCACGCCC AGCTTCCAGAGTTTCCCGAAGATCACGAGGAATAGTTGAGGAGT GCTGTTTTTCGGTCTTGTGATCTGGCTCTCCTCGAGACTTATTGTGC TACGCCGGCCCCGCTCTGAAGGAGGTGGTGGCAGTGGAGGAGGA GGGAGTCGGCCTAGGGCAGTCCCAACCCAGGATCCCCCAGCACC CCTCCCCCTGGTAATTTGGCATGGAATGGGTGATTCTGCTGTAA CCCACTCTCAATGGGGGCAATTAAGAAAAATGGTAGAGAAAAAG ATCCCTGGCATTATGTCTGTCACTCGAAATCGGTAAAACGCTC ATGGAGGACGTAGAAAACAGCTTTTTTCTGAATGTTAATTCACA GGTTACCACGGTCTGCCAAGCATTGGCAAAGGACCCGAAATTGC AACAAGGCTATAACGCGATGGGGTTCAGCCAAGGCGGGCAGTTT CTTCGAGCTGTGGCTCAGCGCTGCCCTTCCCCACCGATGATAAAT TTGATTAGCGTAGGGGGACAACATCAAGGGGTTTTTCGGTTTGCC AAGGTGTCCTGGCGAATCTTCACATATTTGCGACTTTATACGGAA GACCTTGAATGCGGGGGCGTATAGTAAAGTCGTCCAGGAACGGC TTGTCCAAGCTGAATACTGGCACGATCCCATCAAAGAAGATGTC TATCGGAATCACAGCATTTTTCTCGCCGACATAAACCAAGAACG CGGAATTAATGAGTCATACAAGAAGAACTTGATGGCACTTAAAA AATTTGTGATGGTTAAGTTTTTGAATGATAGTATCGTAGATCCCG TAGATAGTGAATGGTTTGGTTTCTATCGATCCGGACAGGCTAAA GAAACGATACCATTGCAGGAAACCTCTTTGTATACTCAAGATAG GTTGGGCCTCAAGGAGATGGATAATGCGGGGCAACTTGTCTTCC TCGCGACTGAGGGTGACCACCTCCAGCTCAGCGAGGAATGGTTT TACGCCACATCATTCCTTTCCTTGGTTAA	
PPT1-5 (wt-PPT1-vIGF2; Codon optimized IDT)	ATGGCAAGTCCAGGGTGTCTTTGGTTGCTCGCGGTTGCCTTGCTC CCTTGGACGTGCGCGTCCCGAGCCCTTCAACACCTCGATCCACCA GCCCCGCTTCCTCTCGTGATATGGCACGGCATGGGCGACAGTTGC TGCAATCCCTTGTCTATGGGCGCAATTA AAAAAGATGGTGGAAAA GAAAATCCCTGGTATCTACGTTTTGAGCCTCGAAATTGGGAAAA CGCTCATGGAGGATGTCGAGAACAGCTTCTTTCTTAACGTCAATT CCCAAGTTACCACGGTTTGTCAAGCCTTGGCGAAAGATCCCAAG	207

	CTTCAGCAAGGGTATAACGCTATGGGATTTAGCCAGGGCGGACA GTTCTGAGGGCGGTAGCACAGAGGTGTCCTAGTCCACCAATGA TAAATCTCATCTCAGTCGGGGGCCAGCACCAGGGCGTCTTCGGG CTTCCTCGATGCCCCGGCGAATCCAGCCACATATGTGACTTCATT AGAAAACTTTGAATGCAGGGGCCTACAGTAAAGTGGTTCAAGA ACGCCTGGTACAAGCAGAGTATTGGCATGACCCGATTAAGGAAG ATGTCTACAGAAATCACTCTATTTTTTTGGCGGACATCAATCAGG AACGAGGCATTAACGAGTCTTACAAGAAGAACCTGATGGCGCTG AAAAAGTTCGTCTATGGTCAAGTTCTTGAATGACTCCATTGTCTGAT CCTGTAGACAGCGAGTGGTTTGGCTTCTACAGGTCTGGTCAAGC AAAGGAGACAATACCACTTCAGGAAACCAGTCTCTATACACAAG ACAGACTGGGTTTGAAGGAAATGGACAATGCAGGCCAACTGGTA TTCCTGGCTACAGAGGGAGATCATCTTCAACTGAGCGAAGAGTG GTTTTATGCCACATAATCCCCCTTCTGGGAAGACCTAGAGCAGT GCCTACGCAGGGTGGTGGTGGCTCTGGAGGAGGAGGCTCCAGGA CTCTGTGTGGGGGCGAGCTGGTGGACACCTTGCAATTCGTGTGTG GCGACCGAGGATTTCTGTTCAGTCGACCTGCCTCAAGAGTAAGC CGGAGGAGTCGGGGGATCGTTGAAGAATGCTGTTTCCGGAGCTG CGACTTGGCGTTGCTCGAGACTTATTGTGCCACACCTGCAAGGA GTGAGTGA	
PPT1-9 (wt- PPT1; native human sequence)	ATGGCGTCGCCCCGGCTGCCTGTGGCTCTTGGCTGTGGCTCTCCTG CCATGGACCTGCGCTTCTCGGGCGCTGCAGCATCTGGACCCGCC GGCGCCGCTGCCGTTGGTGATCTGGCATGGGATGGGAGACAGCT GTTGCAATCCCTTAAGCATGGGTGCTATTAAAAAATGGTGGAG AAGAAAATACCTGGAATTTACGTCTTATCTTTAGAGATTGGGAA GACCCCTGATGGAGGACGTGGAGAACAGCTTCTTCTTGAATGTCA ATTCCCAAGTAACAACAGTGTGTCAGGCACTTGCTAAGGATCCT AAATTGCAGCAAGGCTACAATGCTATGGGATTCTCCAGGGAGG CCAATTTCTGAGGGCAGTGGCTCAGAGATGCCCTTACCTCCCAT GATCAATCTGATCTCGTTGGGGGACAACATCAAGGTGTTTTTGG ACTCCCTCGATGCCAGGAGAGAGCTCTCACATCTGTGACTTCAT CCGAAAAAACTGAATGCTGGGGCGTACTCCAAAGTTGTTTCAGG	208

	AACGCCTCGTGCAAGCCGAATACTGGCATGACCCCATAAAGGAG GATGTGTATCGCAACCACAGCATCTTCTTGGCAGATATAAATCA GGAGCGGGGTATCAATGAGTCCTACAAGAAAAACCTGATGGCCC TGAAGAAGTTTGTGATGGTGAAATTCCTCAATGATTCCATTGTGG ACCCTGTAGATTTCGGAGTGGTTTGGATTTTACAGAAGTGGCCAA GCCAAGGAAACCATTCCCTTACAGGAGACCTCCCTGTACACACA GGACCGCCTGGGGCTAAAGGAAATGGACAATGCAGGACAGCTA GTGTTTCTGGCTACAGAAGGGGACCATCTTCAGTTGTCTGAAGA ATGGTTTTATGCCACATCATACCATTCCCTTGGATGA	
PPT1-10 (wt-PPT1- vIGF2_2; Codon optimized IDT)	ATGGCATCACCGGGTTGCCTCTGGTTGTTGGCCGTTGCGTTGCTT CCGTGGACATGTGCATCAAGAGCTCTTCAACATCTGGATCCCCCA GCTCCCCTGCCGCTCGTAATCTGGCACGGGATGGGGGATTTCATGT TGTAACCCGTTGTCAATGGGCGCGATAAAAAAGATGGTTGAAAA GAAGATTCCAGGCATCTACGTTCTGTCCCTGGAAATCGGTAAGA CACTGATGGAAGACGTGGAGAACTCCTTCTTTCTCAACGTCAATA GTCAGGTCACCTACCGTCTGTCAAGCATTGGCAAAGGACCCTAAA CTTCAGCAGGGGTACAATGCGATGGGGTTTAGCCAGGGCGGACA GTTTCTTAGAGCCGTCGCACAGCGCTGTCCATCTCCCCCGATGAT TAACCTTATATCTGTCTGGGGGACAACACCAGGGTGTTTTTGGTCT TCCTCGCTGTCCTGGTGAAAGCTCCCACATCTGTGATTTTCATACG CAAAACGTTGAACGCAGGAGCTTATAGTAAAGTCGTCCAAGAAC GGCTTGTTCAAGCGGAGTATTGGCATGACCCAATAAAAGAAGAC GTTTATAGGAATCACTCTATCTTCTTGGCCGATATCAACCAAGAA CGCGGAATCAACGAAAGCTACAAAAAGAATCTTATGGCTCTCAA GAAATTTGTTATGGTGAAATTCCTTAATGACTCTATAGTAGATCC TGTCGATTCAGAATGGTTCGGGTTCTACAGGTCTGGCCAGGCGA AGGAGACTATTCCCCTCCAAGAAACGTCTCTCTATACACAAGAC AGACTCGGACTGAAAGAGATGGATAATGCGGGCCAGTTGGTCTT CTTGGCTACGGAAGGCGATCATCTCCAACCTCTCCGAAGAGTGGT TCTATGCCCATATAATCCCGTTCCTGGGCAGACCTAGAGCAGTGC CTACGCAGGGAGGGAGTGGGAGTGGATCCACTTCATCCAGGACT CTGTGTGGGGGCGAGCTGGTGGACACCTTGCAATTCGTGTGTGG	209

	CGACCGAGGATTTCTGTTCAGTCGACCTGCCTCAAGAGTAAGCC GGAGGAGTCGGGGGATCGTTGAAGAATGCTGTTTCCGGAGCTGC GACTTGCGTGTCTCGAGACTTATTGTGCCACACCTGCAAGGAGT GAATGA	
PPT1-11 (BiP- PPT1_2; Codon optimized IDT)	ATGAAGCTCTCCCTGGTGGCCGCGATGCTGCTGCTGCTCAGCGCG GCGCGGGCCTCAAGAGCTCTTCAACATCTGGATCCCCCAGCTCCC CTGCCGCTCGTAATCTGGCACGGGATGGGGGATTCATGTTGTAA CCCGTTGTCAATGGGCGCGATAAAAAAGATGGTTGAAAAGAAGA TTCCAGGCATCTACGTTCTGTCCCTGGAAATCGGTAAGACACTGA TGGAAGACGTGGAGAACTCCTTCTTTCTCAACGTCAATAGTCAG GTCACTACCGTCTGTCAAGCATTGGCAAAGGACCCTAAACTTCA GCAGGGGTACAATGCGATGGGGTTTAGCCAGGGCGGACAGTTTC TTAGAGCCGTCGCACAGCGCTGTCCATCTCCCCCGATGATTAACC TTATATCTGTCTGGGGGACAACACCAGGGTGTTTTTGGTCTTCCTC GCTGTCCTGGTGAAAGCTCCCACATCTGTGATTTTCATACGCAAAA CGTTGAACGCAGGAGCTTATAGTAAAGTCGTCCAAGAACGGCTT GTTCAAGCGGAGTATTGGCATGACCCAATAAAAGAAGACGTTTA TAGGAATCACTCTATCTTCTTGGCCGATATCAACCAAGAACGCG GAATCAACGAAAGCTACAAAAAGAATCTTATGGCTCTCAAGAAA TTTGTTATGGTGAAATTCCTTAATGACTCTATAGTAGATCCTGTC GATTCAGAATGGTTCGGGTCTACAGGTCTGGCCAGGCGAAGGA GACTATTCCCCTCCAAGAAACGTCTCTCTATACACAAGACAGACT CGGACTGAAAGAGATGGATAATGCGGGCCAGTTGGTCTTCTTGG CTACGGAAGGCGATCATCTCCAACCTCTCCGAAGAGTGGTTCTAT GCCCATATAATCCCGTTCCTGGGCTAA	210
PPT1-12 (BiPaa- PPT1_2; Codon optimized IDT)	ATGAAGCTCTCCCTGGTGGCCGCGATGCTGCTGCTGCTCAGCGCG GCGCGGGCCTCAAGAGCTCTTCAACATCTGGATCCCCCAGCTCCC CTGCCGCTCGTAATCTGGCACGGGATGGGGGATTCATGTTGTAA CCCGTTGTCAATGGGCGCGATAAAAAAGATGGTTGAAAAGAAGA TTCCAGGCATCTACGTTCTGTCCCTGGAAATCGGTAAGACACTGA TGGAAGACGTGGAGAACTCCTTCTTTCTCAACGTCAATAGTCAG GTCACTACCGTCTGTCAAGCATTGGCAAAGGACCCTAAACTTCA	211

	GCAGGGGTACAATGCGATGGGGTTTAGCCAGGGCGGACAGTTTC TTAGAGCCGTCGCACAGCGCTGTCCATCTCCCCGATGATTAACC TTATATCTGTGCGGGGACAACACCAGGGTGTTTTTGGTCTTCCTC GCTGTCCTGGTGAAAGCTCCACATCTGTGATTTCATACGCAAAA CGTTGAACGCAGGAGCTTATAGTAAAGTCGTCCAAGAACGGCTT GTTCAAGCGGAGTATTGGCATGACCCAATAAAAGAAGACGTTTA TAGGAATCACTCTATCTTCTTGGCCGATATCAACCAAGAACGCG GAATCAACGAAAGCTACAAAAAGAATCTTATGGCTCTCAAGAAA TTTGTTATGGTGAAATTCCTTAATGACTCTATAGTAGATCCTGTC GATTCAGAATGGTTCGGGTCTACAGGTCTGGCCAGGCGAAGGA GACTATTCCCCTCCAAGAAACGTCTCTCTATACACAAGACAGACT CGGACTGAAAGAGATGGATAATGCGGGCCAGTTGGTCTTCTTGG CTACGGAAGGCGATCATCTCCAACCTCTCCGAAGAGTGGTTCTAT GCCCATATAATCCCGTTCCTGGGCTAA	
PPT1-13 (BiPaa- PPT1; Codon optimized IDT)	ATGAAACTGTCTCTGGTTGCAGCAATGCTCTTGCTGTTGAGTGCG GCCC GCGCGGGCGGCCGATCCACCTGCTCCCCTGCCCTCGTTATA TGGCATGGCATGGGAGATTCCTGTTGTAATCCCCTCAGCATGGG GGCCATCAAAAAAATGGTGGAAAAAAAATACCTGGCATATATG TACTCTCACTTGAAATCGGTAAGACCCTTATGGAAGACGTCGAA AATTCCTTCTTTTTGAACGTGAACTCACAAGTTACGACCGTCTGT CAAGCTCTCGCGAAAGACCCTAAGCTCCAGCAAGGTTATAATGC AATGGGCTTCTCACAGGGAGGTCAGTTCTTGCGAGCGGTAGCCC AGAGGTGTCCGTCTCCGCCAATGATCAACTTGATCTCAGTGGGG GGTCAGCACCAAGGCGTTTTTGGACTCCCTAGATGCCCTGGAGA GAGCTCTCACATTTGCGATTTTATACGGAAGACGCTGAATGCCG GCGCGTATTCAAAGGTCGTTCAAGAGCGACTCGTCCAGGCTGAA TACTGGCACGATCCGATTAAGGAAGACGTGTATCGAAACCATTC TATCTTTCTTGCCGACATTAACCAGGAGCGAGGGATCAACGAAA GTTATAAAAAAACCTGATGGCACTCAAGAAATTTGTAATGGTT AAATTCCTGAACGATTCAATAGTTGATCCGGTGGATTCCGAGTG GTTTCGGCTTCTACCGGTCCGGTCAGGCCAAGGAAACAATCCCAT TGCAAGAAACCAGTCTCTATACTCAGGACCGCCTGGGTCTGAAA	212

	GAAATGGACAACGCTGGCCAACTTGTTTTCTGGCAACGGAGGG TGATCACTTGCAGCTCTCTGAAGAATGGTTTTACGCACACATCAT TCCTTTCCTTGGTTAA	
PPT1-14 (BiP1- vIGF2- PPT1; Codon optimized IDT)	ATGAAGCTCAGTCTCGTGGCAGCTATGCTCCTCCTGCTGTCCCTG GTTGCGGCAATGTTGCTCTTGCTGAGCGCCGCGAGAGCAAGTCG CACGTTGTGTGGAGGTGAACTCGTCGACACCCTTCAGTTCGTATG TGGAGATCGCGGTTTCCTCTTCTCACGCCCAGCTTCCAGAGTTTC CCGAAGATCACGAGGAATAGTTGAGGAGTGCTGTTTTCGGTCTT GTGATCTGGCTCTCCTCGAGACTTATTGTGCTACGCCGGCCCGCT CTGAAGGAGGTGGTGGCAGTGGAGGAGGAGGGAGTCGGCCTAG GGCAGTCCCAACCCAGGATCCCCAGCACCCCTCCCCCTGGTAA TTTGGCATGGAATGGGTGATTCTGCTGTAACCCACTCTCAATGG GGGCAATTAAGAAAATGGTAGAGAAAAAGATCCCTGGCATTAT GTTCTGTCACTCGAAATCGGTAAAACGCTCATGGAGGACGTAGA AAACAGCTTTTTTCTGAATGTTAATTCACAGGTTACCACGGTCTG CCAAGCATTGGCAAAGGACCCGAAATTGCAACAAGGCTATAACG CGATGGGGTTCAGCCAAGGCGGGCAGTTTCTTCGAGCTGTGGCT CAGCGCTGCCCTTCCCCACCGATGATAAATTTGATTAGCGTAGGG GGACAACATCAAGGGGTTTTCGGTTTGCCAAGGTGTCTGGCGA ATCTTCACATATTTGCGACTTTATACGGAAGACCTTGAATGCGGG GGCGTATAGTAAAGTCGTCCAGGAACGGCTTGTCCAAGCTGAAT ACTGGCACGATCCCATCAAAGAAGATGTCTATCGGAATCACAGC ATTTTTCTCGCCGACATAAACCAAGAACGCGGAATTAATGAGTC ATACAAGAAGAACTTGATGGCACTTAAAAAATTTGTGATGGTTA AGTTTTTGAATGATAGTATCGTAGATCCCGTAGATAGTGAATGGT TTGGTTTCTATCGATCCGGACAGGCTAAAGAAACGATACCATTG CAGGAAACCTCTTTGTATACTCAAGATAGGTTGGGCCTCAAGGA GATGGATAATGCGGGGCAACTTGTCTTCCTCGCGACTGAGGGTG ACCACCTCCAGCTCAGCGAGGAATGGTTTTACGCCACATCATTC CTTTCCTTGGTTAA	213
PPT1-15 (BiP1aa-	ATGAAGCTCAGTCTCGTGGCAGCTATGCTCCTCCTGCTGTCCCTG GTTGCGGCAATGTTGCTCTTGCTGAGCGCCGCGAGAGCAGCAGC	214

vIGF2- PPT1; Codon optimized IDT)	TAGTCGCACGTTGTGTGGAGGTGAACTCGTCGACACCCTTCAGTT CGTATGTGGAGATCGCGGTTTCCTCTTCTCACGCCCAGCTTCCAG AGTTTCCC GAAGATCACGAGGAATAGTTGAGGAGTGCTGTTTTTC GGTCTTGTGATCTGGCTCTCCTCGAGACTTATTGTGCTACGCCGG CCCGCTCTGAAGGAGGTGGTGGCAGTGGAGGAGGAGGGAGTCG GCCTAGGGCAGTCCCAACCCAGGATCCCCCAGCACCCCTCCCC TGGTAATTTGGCATGGAATGGGTGATTCTGCTGTAACCCACTCT CAATGGGGGCAATTAAGAAAAATGGTAGAGAAAAAGATCCCTGG CATTTATGTTCTGTCACTCGAAATCGGTAAAACGCTCATGGAGGA CGTAGAAAACAGCTTTTTTCTGAATGTTAATTCACAGGTTACCAC GGTCTGCCAAGCATTGGCAAAGGACCCGAAATTGCAACAAGGCT ATAACGCGATGGGGTTCAGCCAAGGCGGGCAGTTTCTTCGAGCT GTGGCTCAGCGCTGCCCTTCCCCACCGATGATAAATTTGATTAGC GTAGGGGGACAACATCAAGGGGTTTTTCGGTTTGCCAAGGTGTCC TGGCGAATCTTCACATATTTGCGACTTTATACGGAAGACCTTGAA TGCGGGGGCGTATAGTAAAGTCGTCCAGGAACGGCTTGTTCCAAG CTGAATACTGGCACGATCCCATCAAAGAAGATGTCTATCGGAAT CACAGCATTTTTCTCGCCGACATAAACCAAGAACGCGGAATTAA TGAGTCATACAAGAAGAACTTGATGGCACTTAAAAAATTTGTGA TGGTTAAGTTTTTGAATGATAGTATCGTAGATCCCGTAGATAGTG AATGGTTTGGTTTCTATCGATCCGGACAGGCTAAAGAAACGATA CCATTGCAGGAAACCTCTTTGTATACTCAAGATAGGTTGGGCCTC AAGGAGATGGATAATGCGGGGCAACTTGTCTTCCTCGCGACTGA GGGTGACCACCTCCAGCTCAGCGAGGAATGGTTTTACGCCCACA TCATTCTTTCTTGGTTAA	
PPT1-16 (BiP1aa- PPT1_2; Codon optimized IDT)	ATGAAGCTCAGTCTCGTGGCAGCTATGCTCCTCCTGCTGTCCCTG GTTGCGGCAATGTTGCTCTTGCTGAGCGCCGCGAGAGCAGCCGC GTCAAGAGCTCTTCAACATCTGGATCCCCCAGCTCCCCTGCCGCT CGTAATCTGGCACGGGATGGGGGATTCATGTTGTAACCCGTTGTC AATGGGCGCGATAAAAAAGATGGTTGAAAAGAAGATTCCAGGC ATCTACGTTCTGTCCCTGGAAATCGGTAAGACACTGATGGAAGA CGTGGAGAACTCCTTCTTTCTCAACGTCAATAGTCAGGTCACTAC	215

	CGTCTGTCAAGCATTGGCAAAGGACCCTAAACTTCAGCAGGGGT ACAATGCGATGGGGTTTAGCCAGGGCGGACAGTTTCTTAGAGCC GTCGCACAGCGCTGTCCATCTCCCCGATGATTAACCTTATATCT GTCGGGGGACAACACCAGGGTGTTTTTGGTCTTCCTCGCTGTCCT GGTGAAAGCTCCCACATCTGTGATTTTCATACGCAAAACGTTGAA CGCAGGAGCTTATAGTAAAGTCGTCCAAGAACGGCTTGTTCAAG CGGAGTATTGGCATGACCCAATAAAGAAGACGTTTATAGGAAT CACTCTATCTTCTTGGCCGATATCAACCAAGAACGCGGAATCAA CGAAAGCTACAAAAAGAATCTTATGGCTCTCAAGAAATTTGTTA TGGTGAAATTCCTTAATGACTCTATAGTAGATCCTGTTCGATTGAG AATGGTTTCGGGTTCTACAGGTCTGGCCAGGCGAAGGAGACTATT CCCCTCCAAGAAACGTCTCTCTATACACAAGACAGACTCGGACT GAAAGAGATGGATAATGCGGGCCAGTTGGTCTTCTTGCTACGG AAGGCGATCATCTCCAACCTCTCCGAAGAGTGGTTCTATGCCATA TAATCCCGTTCCTGGGCTAA	
PPT1-17 (wt-PPT1- C6S; natural human sequence)	ATGGCGTCGCCCCGGCAGCCTGTGGCTCTTGGCTGTGGCTCTCCTG CCATGGACCTGCGCTTCTCGGGCGCTGCAGCATCTGGACCCGCC GGCGCCGCTGCCGTTGGTGATCTGGCATGGGATGGGAGACAGCT GTTGCAATCCCTTAAGCATGGGTGCTATTAAAAAATGGTGGAG AAGAAAATACCTGGAATTTACGTCTTATCTTTAGAGATTGGGAA GACCCTGATGGAGGACGTGGAGAACAGCTTCTTCTTGAATGTCA ATTCCCAAGTAACAACAGTGTGTCAGGCACCTTGCTAAGGATCCT AAATTGCAGCAAGGCTACAATGCTATGGGATTCTCCAGGGAGG CCAATTTCTGAGGGCAGTGGCTCAGAGATGCCCTTACCTCCCAT GATCAATCTGATCTCGGTTGGGGGACAACATCAAGGTGTTTTTGG ACTCCCTCGATGCCAGGAGAGAGCTCTCACATCTGTGACTTCAT CCGAAAAACACTGAATGCTGGGGCGTACTCCAAAGTTGTTTCAGG AACGCCTCGTGCAAGCCGAATACTGGCATGACCCCATAAAGGAG GATGTGTATCGCAACCACAGCATCTTCTTGGCAGATATAAATCA GGAGCGGGGTATCAATGAGTCCTACAAGAAAAACCTGATGGCCC TGAAGAAGTTTGTGATGGTGAAATTCCTCAATGATTCCATTGTGG ACCCTGTAGATTGCGAGTGGTTTGGATTTTACAGAAGTGGCCAA	216

	GCCAAGGAAACCATTCCCTTACAGGAGACCTCCCTGTACACACA GGACCGCCTGGGGCTAAAGGAAATGGACAATGCAGGACAGCTA GTGTTTCTGGCTACAGAAGGGGACCATCTTCAGTTGTCTGAAGA ATGGTTTTATGCCCACATCATACCATTCCCTTGGATGA	
PPT1-18 (BiP2aa- PPT1; Codon optimized IDT)	ATGAAGCTCTCCCTGGTGGCCGCGATGCTGCTGCTGCTCTGGGTG GCACTGCTGCTGCTCAGCGCGGCGAGGGCCGCGCTCAAGAGC TCTTCAACATCTGGATCCCCCAGCTCCCCTGCCGCTCGTAATCTG GCACGGGATGGGGGATTCATGTTGTAACCCGTTGTCAATGGGCG CGATAAAAAAGATGGTTGAAAAGAAGATTCCAGGCATCTACGTT CTGTCCCTGGAAATCGGTAAGACACTGATGGAAGACGTGGAGAA CTCCTTCTTTCTCAACGTCAATAGTCAGGTCACTACCGTCTGTCA AGCATTGGCAAAGGACCCTAAACTTCAGCAGGGGTACAATGCGA TGGGGTTTAGCCAGGGCGGACAGTTTCTTAGAGCCGTCGCACAG CGCTGTCCATCTCCCCCGATGATTAACCTTATATCTGTCGGGGGA CAACACCAGGGTGTTTTTGGTCTTCCTCGCTGTCCTGGTGAAAGC TCCCACATCTGTGATTTTCATACGCAAAACGTTGAACGCAGGAGC TTATAGTAAAGTCGTCCAAGAACGGCTTGTTCAAGCGGAGTATT GGCATGACCCAATAAAAGAAGACGTTTATAGGAATCACTCTATC TTCTTGCCGATATCAACCAAGAACGCGGAATCAACGAAAGCTA CAAAAAGAATCTTATGGCTCTCAAGAAATTTGTTATGGTGAAATT CCTTAATGACTCTATAGTAGATCCTGTGCTGATTGAGAAATGGTTCGG GTTCTACAGGTCTGGCCAGGCGAAGGAGACTATTCCTCCCAAG AAACGTCTCTCTATACACAAGACAGACTCGGACTGAAAGAGATG GATAATGCGGGCCAGTTGGTCTTCTTGGCTACGGAAGGCGATCA TCTCCAACCTCTCCGAAGAGTGTTTCTATGCCCATATAATCCCGTT CCTGGGCTAA	217
PPT1-19 (GaussiaAA -PPT1_2; Codon optimized IDT)	ATGGGTGTAAAGGTGTTGTTTCGCTCTTATCTGCATTGCCGTTGCA GAAGCTGCCGCGTCAAGAGCTCTTCAACATCTGGATCCCCCAGC TCCCCTGCCGCTCGTAATCTGGCACGGGATGGGGGATTCATGTTG TAACCCGTTGTCAATGGGCGCGATAAAAAAGATGGTTGAAAAGA AGATTCCAGGCATCTACGTTCTGTCCCTGGAAATCGGTAAGACA CTGATGGAAGACGTGGAGAACTCCTTCTTTCTCAACGTCAATAGT	218

	CAGGTCACTACCGTCTGTCAAGCATTGGCAAAGGACCCTAAACT TCAGCAGGGGTACAATGCGATGGGGTTTAGCCAGGGCGGACAGT TTCTTAGAGCCGTCGCACAGCGCTGTCCATCTCCCCGATGATTA ACCTTATATCTGTCGGGGGACAACACCAGGGTGTTTTTGGTCTTC CTCGCTGTCCTGGTGAAAGCTCCACATCTGTGATTCATACGCA AAACGTTGAACGCAGGAGCTTATAGTAAAGTCGTCCAAGAACGG CTTGTTCAAGCGGAGTATTGGCATGACCCAATAAAAGAAGACGT TTATAGGAATCACTCTATCTTCTTGCCGATATCAACCAAGAACG CGGAATCAACGAAAGCTACAAAAAGAATCTTATGGCTCTCAAGA AATTTGTTATGGTGAAATTCCTTAATGACTCTATAGTAGATCCTG TCGATTCAGAATGGTTCGGGTCTACAGGTCTGGCCAGGCGAAG GAGACTATTCCCCTCCAAGAAACGTCTCTCTATACACAAGACAG ACTCGGACTGAAAGAGATGGATAATGCGGGCCAGTTGGTCTTCT TGGCTACGGAAGGCGATCATCTCCAACCTCTCCGAAGAGTGGTTC TATGCCCATATAATCCCGTTCCTGGGCTAA	
PPT1-20 (GaussiaAA -vIGF2- PPT1; Codon optimized IDT)	ATGGGTGTAAAGGTGTTGTTTCGCTCTTATCTGCATTGCCGTTGCA GAAGCTGCAGCTAGTCGCACGTTGTGTGGAGGTGAACTCGTCGA CACCCCTCAGTTCGTATGTGGAGATCGCGGTTTCCTCTTCTCACG CCCAGCTTCCAGAGTTTCCCGAAGATCACGAGGAATAGTTGAGG AGTGCTGTTTTTCGGTCTTGTGATCTGGCTCTCCTCGAGACTTATTG TGCTACGCCGGCCCGCTCTGAAGGAGGTGGTGGCAGTGGAGGAG GAGGGAGTCGGCCTAGGGCAGTCCCAACCCAGGATCCCCCAGCA CCCCTCCCCCTGGTAATTTGGCATGGAATGGGTGATTCCTGCTGT AACCCACTCTCAATGGGGGCAATTAAGAAAATGGTAGAGAAAA AGATCCCTGGCATTATGTTCTGTCACTCGAAATCGGTAAAACGC TCATGGAGGACGTAGAAAACAGCTTTTTTCTGAATGTTAATTCAC AGGTTACCACGGTCTGCCAAGCATTGGCAAAGGACCCGAAATTG CAACAAGGCTATAACGCGATGGGGTTCAGCCAAGGCGGGCAGTT TCTTCGAGCTGTGGCTCAGCGCTGCCCTTCCCCACCGATGATAAA TTTGATTAGCGTAGGGGGGACAACATCAAGGGGTTTTTCGGTTTGCC AAGGTGTCCTGGCGAATCTTCACATATTTGCGACTTTATACGGAA GACCTTGAATGCGGGGGCGTATAGTAAAGTCGTCCAGGAACGGC	219

	TTGTCCAAGCTGAATACTGGCACGATCCCATCAAAGAAGATGTC TATCGGAATCACAGCATTTTTCTCGCCGACATAAACCAAGAACG CGGAATTAATGAGTCATACAAGAAGAACTTGATGGCACTTAAAA AATTTGTGATGGTTAAGTTTTTGAATGATAGTATCGTAGATCCCG TAGATAGTGAATGGTTTGGTTTCTATCGATCCGGACAGGCTAAA GAAACGATACCATTGCAGGAAACCTCTTTGTATACTCAAGATAG GTTGGGCCTCAAGGAGATGGATAATGCGGGGCAACTTGTCTTCC TCGCGACTGAGGGTGACCACCTCCAGCTCAGCGAGGAATGGTTT TACGCCACATCATTCCTTTCCTTGGTTAA	
PPT1-21 (ppt2ss- PPT1; Codon optimized IDT)	ATGCTGGGGCTCTGGGGGCAGCGGCTCCCCGCGGCGTGGGTCCT GCTTCTGTTGCCTTTCCTGCCGCTGCTGCTGCTTGCAGATCCCCCA GCTCCCCTGCCGCTCGTAATCTGGCACGGGATGGGGGATTCATGT TGTAACCCGTTGTCAATGGGCGCGATAAAAAAGATGGTTGAAAA GAAGATTCCAGGCATCTACGTTCTGTCCCTGGAAATCGGTAAGA CACTGATGGAAGACGTGGAGAACTCCTTCTTTCTCAACGTCAATA GTCAGGTCCTACCGTCTGTCAAGCATTGGCAAAGGACCCTAAA CTTCAGCAGGGGTACAATGCGATGGGGTTTAGCCAGGGCGGACA GTTTCTTAGAGCCGTCGCACAGCGCTGTCCATCTCCCCCGATGAT TAACCTTATATCTGTCTGGGGGACAACACCAGGGTGTTTTTGGTCT TCCTCGCTGTCCTGGTGAAAGCTCCCACATCTGTGATTTCATACG CAAAACGTTGAACGCAGGAGCTTATAGTAAAGTCGTCCAAGAAC GGCTTGTTCAAGCGGAGTATTGGCATGACCCAATAAAAGAAGAC GTTTATAGGAATCACTCTATCTTCTTGGCCGATATCAACCAAGAA CGCGGAATCAACGAAAGCTACAAAAAGAATCTTATGGCTCTCAA GAAATTTGTTATGGTGAAATTCCTTAATGACTCTATAGTAGATCC TGTCGATTCAGAATGGTTCGGGTTCTACAGGTCTGGCCAGGCGA AGGAGACTATTCCCCTCCAAGAAACGTCTCTCTATACACAAGAC AGACTCGGACTGAAAGAGATGGATAATGCGGGCCAGTTGGTCTT CTTGGCTACGGAAGGCGATCATCTCCAACCTCTCCGAAGAGTGGT TCTATGCCCATATAATCCCGTTCCTGGGCTAA	220
PPT1-22 (ppt2ss-	ATGCTGGGGCTCTGGGGGCAGCGGCTCCCCGCGGCGTGGGTCCT GCTTCTGTTGCCTTTCCTGCCGCTGCTGCTGCTTGCATCAAGAGC	221

PPT1_2; Codon optimized IDT)	TCTTCAACATCTGGATCCCCCAGCTCCCCTGCCGCTCGTAATCTG GCACGGGATGGGGGATTCATGTTGTAACCCGTTGTCAATGGGCG CGATAAAAAAGATGGTTGAAAAGAAGATTCCAGGCATCTACGTT CTGTCCCTGGAAATCGGTAAGACACTGATGGAAGACGTGGAGAA CTCCTTCTTTCTCAACGTCAATAGTCAGGTCACTACCGTCTGTCA AGCATTGGCAAAGGACCCTAACTTCAGCAGGGGTACAATGCGA TGGGGTTTAGCCAGGGCGGACAGTTTCTTAGAGCCGTCGCACAG CGCTGTCCATCTCCCCCGATGATTAACCTTATATCTGTCGGGGGA CAACACCAGGGTGTTTTTGGTCTTCCTCGCTGTCCTGGTGAAAGC TCCCACATCTGTGATTTTCATACGCAAAACGTTGAACGCAGGAGC TTATAGTAAAGTCGTCCAAGAACGGCTTGTTCAAGCGGAGTATT GGCATGACCCAATAAAAGAAGACGTTTATAGGAATCACTCTATC TTCTTGCCGATATCAACCAAGAACGCGGAATCAACGAAAGCTA CAAAAAGAATCTTATGGCTCTCAAGAAATTTGTTATGGTGAAATT CCTTAATGACTCTATAGTAGATCCTGTCGATTCAGAATGGTTCGG GTTCTACAGGTCTGGCCAGGCGAAGGAGACTATTCCCCTCCAAG AAACGTCTCTCTATACACAAGACAGACTCGGACTGAAAGAGATG GATAATGCGGGCCAGTTGGTCTTCTTGGCTACGGAAGGCGATCA TCTCCAACCTCTCCGAAGAGTGGTTCTATGCCCATATAATCCCGTT CCTGGGCTAA	
PPT1-23 (consensusS S-PPT1; Codon optimized IDT)	ATGGCAAGTCCTTCCTGTCTTTGGCTGCTGGCTGTTGCCTTGCTTC CTTGGTCTTGTGCGGCGCGGGCACTCGGCCATTTGGACCCACCAG CCCCACTGCCCTTGGTTATATGGCATGGAATGGGAGATAGTTGCT GTAATCCACTGAGCATGGGAGCCATAAAGAAAATGGTTGAGAAA AAAATACCGGGAATATATGTTCTGAGCCTGGAGATAGGTAAGAC ACTCATGGAAGACGTTGAAAACCTATTTTTTTTGAACGTGAATAG TCAAGTCACAACGGTCTGTCAAGCTCTGGCTAAAGATCCTAAGTT GCAACAGGGTTACAATGCGATGGGATTTAGTCAAGGTGGACAGT TCCTGCGGGCCGTCGCACAGAGGTGCCCCGAGTCCGCCAATGATA AATCTCATTTAGTAGGCGGACAACATCAGGGCGTGTTCCGGTCTT CCTCGCTGCCCCGGGTGAGTCTTCTCACATTTGCGATTCATACGC AAAACACTTAACGCGGGGGCTTACTCCAAGGTAGTTCAAGAAAG	222

	GCTCGTGCAGGCCGAATACTGGCATGATCCAATCAAAGAAGACG TCTATAGAAATCACTCTATATTCTTGGCCGACATCAACCAAGAGC GAGGTATAAATGAAAGTTACAAGAAAAACCTCATGGCTCTTAAA AAATTTGTTATGGTAAAATTTCTTAATGACTCTATCGTTGACCCG GTCGATAGTGAGTGGTTTGGGTTTTATAGGAGCGGACAGGCCAA AGAGACAATTCCGTTGCAGGAGACAAGTTTGTACACGCAGGATA GGCTTGGTCTTAAGGAGATGGACAACGCGGGCCAACCTTGATTT TTGGCTACTGAAGGTGATCACCTCCAATTGTCTGAAGAGTGGTTT TATGCGCATATTATTCCTTTCCTCGGCTAA	
PPT1-24 (consensus- PPT1; Codon optimized IDT)	ATGGCAAGCCCTTCCTGCCTCTGGTTGCTTGCTGTTGCTTTGCTTC CTTGGTCTTGCTGCAAGAGCACTTGGCCACCTTGATCCTCCTG CACCTCTCCCGCTCGTTATATGGCACGGCATGGGGGATAGCTGTT GTAATCCACTGTCAATGGGGGCTATTAAGAAAAATGGTGGAGAAG AAAATTCCGGGAATTTATGTGCTCTCCCTGGAGATAGGCAAAAC GCTTATGGAAGACGTGGAGAACAGTTTTTTTCTTAACGTAAATTC ACAGGTTACCACCGTCTGTCAAATTTTGGCCAAAGATCCCAAAC GCAACAAGGGTATAACGCTATGGGCTTCAGTCAAGGGGGTCAAT TTTTGAGGGCGGTTGCGCAACGCTGCCCTAGTCCGCCCATGATAA ACTTGATCAGTGTTGGGGGACAGCACCAGGGAGTATTTGGTCTG CCGAGGTGTCCAGGCGAGTCTTCACACATCTGTGACTTTATTCGC AAGACCTTGAACGCGGGCGCTTATTCCAAGGCTGTGCAGGAAAG GCTTGTGCAAGCGGAATATTGGCACGATCCTATAAAGGAAGATG TGTATCGCAACCACTCTATCTTCCTGGCGGATATCAATCAAGAAC GAGGAGTCAATGAGTCCTACAAGAAAAATCTGATGGCGCTTAAA AAGTTCGTAATGGTCAAGTTCCTGAATGACAGCATAGTAGATCC GGTGGATTCTGAATGGTTCGGATTCTACCGGTCAGGACAGGCCA AGGAGACAATCCCCCTTCAAGAGACGACCCTGTACACACAAGAT AGATTGGGACTGAAAGAAATGGATAAGGCCGGTCAATTGGTCTT CTTGGCCACAGAAGGGGACCATCTCCAACCTGAGTGAAGAATGGT TTTATGCACATATAATTCCCTTCCTGGAGTAA	223
PPT1-25 (wt-PPT1)	ATGGCATCACCGGGTTGCCTCTGGTTGTTGGCCGTTGCGTTGCTT CCGTGGACATGTGCATCAAGAGCTCTTCAACATCTGGATCCCCCA	224

L283C H300C; Codon optimized IDT)	GCTCCCCTGCCGCTCGTAATCTGGCACGGGATGGGGGATTCATGT TGTAACCCGTTGTCAATGGGCGCGATAAAAAAGATGGTTGAAAA GAAGATTCCAGGCATCTACGTTCTGTCCCTGGAAATCGGTAAGA CACTGATGGAAGACGTGGAGAACTCCTTCTTTCTCAACGTCAATA GTCAGGTCACTACCGTCTGTCAAGCATTGGCAAAGGACCCTAAA CTTCAGCAGGGGTACAATGCGATGGGGTTTAGCCAGGGCGGACA GTTTCTTAGAGCCGTCGCACAGCGCTGTCCATCTCCCCCGATGAT TAACCTTATATCTGTCTGGGGGACAACACCAGGGTGTTTTTGGTCT TCCTCGCTGTCTGGTGAAAGCTCCACATCTGTGATTTTCATACG CAAAACGTTGAACGCAGGAGCTTATAGTAAAGTCGTCCAAGAAC GGCTTGTTCAAGCGGAGTATTGGCATGACCCAATAAAAGAAGAC GTTTATAGGAATCACTCTATCTTCTTGCCGATATCAACCAAGAA CGCGGAATCAACGAAAGCTACAAAAAGAATCTTATGGCTCTCAA GAAATTTGTTATGGTGAAATTCCTTAATGACTCTATAGTAGATCC TGTCGATTCAGAATGGTTCGGGTTCTACAGGTCTGGCCAGGCGA AGGAGACTATTCCCCTCCAAGAAACGTCTCTCTATACACAAGAC AGACTCGGACTGAAAGAGATGGATAATGCGGGCCAGTTGGTCTT CTGCGCTACGGAAGGCGATCATCTCCAACCTCTCCGAAGAGTGGT TCTATGCCTGCATAATCCCGTTCCTGGGCTAA	
PPT1-26 (wt-PPT1 G113C L121C; Codon optimized IDT)	ATGGCATCACCGGGTTGCCTCTGGTTGTTGGCCGTTGCGTTGCTT CCGTGGACATGTGCATCAAGAGCTCTTCAACATCTGGATCCCCCA GCTCCCCTGCCGCTCGTAATCTGGCACGGGATGGGGGATTCATGT TGTAACCCGTTGTCAATGGGCGCGATAAAAAAGATGGTTGAAAA GAAGATTCCAGGCATCTACGTTCTGTCCCTGGAAATCGGTAAGA CACTGATGGAAGACGTGGAGAACTCCTTCTTTCTCAACGTCAATA GTCAGGTCACTACCGTCTGTCAAGCATTGGCAAAGGACCCTAAA CTTCAGCAGGGGTACAATGCGATGTGCTTTAGCCAGGGCGGACA GTTTTGCAGAGCCGTCGCACAGCGCTGTCCATCTCCCCCGATGAT TAACCTTATATCTGTCTGGGGGACAACACCAGGGTGTTTTTGGTCT TCCTCGCTGTCTGGTGAAAGCTCCACATCTGTGATTTTCATACG CAAAACGTTGAACGCAGGAGCTTATAGTAAAGTCGTCCAAGAAC GGCTTGTTCAAGCGGAGTATTGGCATGACCCAATAAAAGAAGAC	225

	GTTTATAGGAATCACTCTATCTTCTTGGCCGATATCAACCAAGAA CGCGGAATCAACGAAAGCTACAAAAAGAATCTTATGGCTCTCAA GAAATTTGTTATGGTGAAATTCCTTAATGACTCTATAGTAGATCC TGTCGATTCAGAATGGTTCGGGTTCTACAGGTCTGGCCAGGCGA AGGAGACTATTCCCCTCCAAGAAACGTCTCTCTATACACAAGAC AGACTCGGACTGAAAGAGATGGATAATGCGGGCCAGTTGGTCTT CTTGGCTACGGAAGGCGATCATCTCCAACCTCTCCGAAGAGTGGT TCTATGCCCATATAATCCCGTTCCTGGGCTAA	
PPT1-27 (wt-PPT1 A171C A183C; Codon optimized IDT)	ATGGCATCACCGGGTTGCCTCTGGTTGTTGGCCGTTGCGTTGCTT CCGTGGACATGTGCATCAAGAGCTCTTCAACATCTGGATCCCCCA GCTCCCCTGCCGCTCGTAATCTGGCACGGGATGGGGGATTTCATGT TGTAACCCGTTGTCAATGGGCGCGATAAAAAAGATGGTTGAAAA GAAGATTCCAGGCATCTACGTTCTGTCCCTGGAAATCGGTAAGA CACTGATGGAAGACGTGGAGAACTCCTTCTTTCTCAACGTCAATA GTCAGGTCACTACCGTCTGTCAAGCATTGGCAAAGGACCCTAAA CTTCAGCAGGGGTACAATGCGATGGGGTTTAGCCAGGGCGGACA GTTTCTTAGAGCCGTCGCACAGCGCTGTCCATCTCCCCCGATGAT TAACCTTATATCTGTCTGGGGGACAACACCAGGGTGTTTTTGGTCT TCCTCGCTGTCTGGTGAAAGCTCCCACATCTGTGATTTTCATACG CAAAACGTTGAACGCAGGATGCTATAGTAAAGTCGTCCAAGAAC GGCTTGTTCAATGCGAGTATTGGCATGACCCAATAAAAGAAGAC GTTTATAGGAATCACTCTATCTTCTTGGCCGATATCAACCAAGAA CGCGGAATCAACGAAAGCTACAAAAAGAATCTTATGGCTCTCAA GAAATTTGTTATGGTGAAATTCCTTAATGACTCTATAGTAGATCC TGTCGATTCAGAATGGTTCGGGTTCTACAGGTCTGGCCAGGCGA AGGAGACTATTCCCCTCCAAGAAACGTCTCTCTATACACAAGAC AGACTCGGACTGAAAGAGATGGATAATGCGGGCCAGTTGGTCTT CTTGGCTACGGAAGGCGATCATCTCCAACCTCTCCGAAGAGTGGT TCTATGCCCATATAATCCCGTTCCTGGGCTAA	226
PPT1-28 (BiP2aa- PPT1;	ATGAAACTTAGTCTCGTCGCAGCAATGTTGCTTCTCCTGTGGGTT GCCCTCCTGTTGCTCAGCGCAGCTAGGGCTGCTGCGTCTCGGGCG CTGCAGCATCTGGACCCGCCGGCGCCGCTGCCGTTGGTGATCTG	227

native human sequence)	GCATGGGATGGGAGACAGCTGTTGCAATCCCTTAAGCATGGGTG CTATTAAAAAATGGTGGAGAAGAAAATACCTGGAATTTACGTC TTATCTTTAGAGATTGGGAAGACCTGATGGAGGACGTGGAGAA CAGCTTCTTCTTGAATGTCAATTCCCAAGTAACAACAGTGTGTCA GGCACTTGCTAAGGATCCTAAATTGCAGCAAGGCTACAATGCTA TGGGATTCTCCCAGGGAGGCCAATTTCTGAGGGCAGTGGCTCAG AGATGCCCTTCACCTCCCATGATCAATCTGATCTCGGTTGGGGGA CAACATCAAGGTGTTTTTGGACTCCCTCGATGCCCAGGAGAGAG CTCTCACATCTGTGACTTCATCCGAAAAACACTGAATGCTGGGGC GTACTCCAAAGTTGTTTCAGGAACGCCTCGTGCAAGCCGAATACT GGCATGACCCCATAAAGGAGGATGTGTATCGCAACCACAGCATC TTCTTGGCAGATATAAATCAGGAGCGGGGTATCAATGAGTCCTA CAAGAAAAACCTGATGGCCCTGAAGAAGTTTGTGATGGTGAAAT TCCTCAATGATTCCATTGTGGACCCTGTAGATTCGGAGTGGTTTG GATTTTACAGAAGTGGCCAAGCCAAGGAAACCATTCCCTTACAG GAGACCTCCCTGTACACACAGGACCGCCTGGGGCTAAAGGAAAT GGACAATGCAGGACAGCTAGTGTCTTGGCTACAGAAGGGGACC ATCTTCAGTTGTCTGAAGAATGGTTTTATGCCCACATCATACCAT TCCTTGATGA	
PPT1-101	<u>ATGAAGCTCTCCCTGGTGGCCGCGATGCTGCTGCTGCTCTGGGTG</u> <u>GCACTGCTGCTGCTCAGCGCGGCGAGGGCCGCGCGTCTAGAAC</u> ACTGTGCGGAGGGGAGCTTGTAGACACTCTTCAGTTCGTGTGTG GAGATCGCGGGTTCCTCTTCTCTCGCGGAGGTGGAGGTTCTAGG GGTATACTGGAGGAGTGTGTTTCAGGGACTGTGACTTGCGCTC CTCGAGACCTATTGCGCGACGCCAGCCAGGTCCGAAGGAGGTGG TGGCAGTGGAGGAGGAGGGAGTCGGCCTAGGGCAGTCCCAACC CAGGACCCGCCGGCGCCGCTGCCGTTGGTGATCTGGCATGGGAT GGGAGACAGCTGTTGCAATCCCTTAAGCATGGGTGCTATTAAAA AAATGGTGGAGAAGAAAATACCTGGAATTTACGTCTTATCTTTA GAGATTGGGAAGACCCTGATGGAGGACGTGGAGAACAGCTTCTT CTTGAATGTCAATTCCCAAGTAACAACAGTGTGTCAGGCACTTGC TAAGGATCCTAAATTGCAGCAAGGCTACAATGCTATGGGATTCT	228

	CCCAGGGAGGCCAATTTCTGAGGGCAGTGGCTCAGAGATGCCCT TCACCTCCCATGATCAATCTGATCTCGGTTGGGGGACAACATCAA GGTGTTTTTGGACTCCCTCGATGCCCAGGAGAGAGCTCTCACATC TGTGACTTCATCCGAAAAACACTGAATGCTGGGGCGTACTCCAA AGTTGTTTCAGGAACGCCTCGTGCAAGCCGAATACTGGCATGACC CCATAAAGGAGGATGTGTATCGCAACCACAGCATCTTCTTGGCA GATATAAATCAGGAGCGGGGTATCAATGAGTCCTACAAGAAAAA CCTGATGGCCCTGAAGAAGTTTGTGATGGTGAAATTCCTCAATG ATTCCATTGTGGACCCTGTAGATTTCGGAGTGGTTTGGATTTTACA GAAGTGGCCAAGCCAAGGAAACCATTCCTTACAGGAGACCTCC CTGTACACACAGGACCGCCTGGGGCTAAAGGAAATGGACAATGC AGGACAGCTAGTGTCTTCTGGCTACAGAAGGGGACCATCTTCAGT TGTCTGAAGAATGGTTTTATGCCCACATCATACCATTCCTTGGAT GA	
PPT1-31 (BiP1- vIGF2- PPT1; native human sequence)	ATGAAGCTCAGTCTCGTGGCAGCTATGCTCCTCCTGCTGTCCCTG GTTGCGGCAATGTTGCTCTTGCTGAGCGCCGCGAGAGCAAGTCG CACGTTGTGTGGAGGTGAACTCGTCGACACCCTTCAGTTCGTATG TGGAGATCGCGGTTTCCTCTTCTCACGCCCAGCTTCCAGAGTTTC CCGAAGATCACGAGGAATAGTTGAGGAGTGCTGTTTTTCGGTCTT GTGATCTGGCTCTCCTCGAGACTTATTGTGCTACGCCGGCCCCGCT CTGAAGGAGGTGGTGGCAGTGGAGGAGGAGGGAGTCGGCCTAG GGCAGTCCCAACCCAGGACCCGCCGGCGCCGCTGCCGTTGGTGA TCTGGCATGGGATGGGAGACAGCTGTTGCAATCCCTTAAGCATG GGTGCTATTAAAAAAATGGTGGAGAAGAAAATACCTGGAATTTA CGTCTTATCTTTAGAGATTGGGAAGACCCTGATGGAGGACGTGG AGAACAGCTTCTTCTTGAATGTCAATTCCCAAGTAACAACAGTGT GTCAGGCACTTGCTAAGGATCCTAAATTGCAGCAAGGCTACAAT GCTATGGGATTCTCCCAGGGAGGCCAATTTCTGAGGGCAGTGGC TCAGAGATGCCCTTCACCTCCCATGATCAATCTGATCTCGGTTGG GGGACAACATCAAGGTGTTTTTGGACTCCCTCGATGCCCAGGAG AGAGCTCTCACATCTGTGACTTCATCCGAAAAACACTGAATGCT	229

	GGGGCGTACTCCAAAGTTGTTTCAGGAACGCCTCGTGCAAGCCGA ATACTGGCATGACCCCATAAAGGAGGATGTGTATCGCAACCACA GCATCTTCTTGGCAGATATAAATCAGGAGCGGGGTATCAATGAG TCCTACAAGAAAAACCTGATGGCCCTGAAGAAGTTTGTGATGGT GAAATTCCTCAATGATTCCATTGTGGACCCTGTAGATTTCGGAGTG GTTTGGATTTTACAGAAGTGGCCAAGCCAAGGAAACCATTCCCT TACAGGAGACCTCCCTGTACACACAGGACCGCCTGGGGCTAAAG GAAATGGACAATGCAGGACAGCTAGTGTCTTCTGGCTACAGAAGG GGACCATCTTCAGTTGTCTGAAGAATGGTTTTATGCCACATCAT ACCATTCCTTGGATGA	
PPT1-32 (wt-PPT1- vIGF2-32; native human sequence)	ATGGCGTCGCCCCGGCTGCCTGTGGCTCTTGGCTGTGGCTCTCCTG CCATGGACCTGCGCTTCTCGGGCGCTGCAGCATCTGGACCCGCC GGCGCCGCTGCCGTTGGTGATCTGGCATGGGATGGGAGACAGCT GTTGCAATCCCTTAAGCATGGGTGCTATTAAAAAATGGTGGAG AAGAAAATACCTGGAATTTACGTCTTATCTTTAGAGATTGGGAA GACCTGATGGAGGACGTGGAGAACAGCTTCTTCTTGAATGTCA ATTCCCAAGTAACAACAGTGTGTCAGGCACCTTGCTAAGGATCCT AAATTGCAGCAAGGCTACAATGCTATGGGATTCTCCAGGGAGG CCAATTTCTGAGGGCAGTGGCTCAGAGATGCCCTTCACCTCCCAT GATCAATCTGATCTCGGTTGGGGGACAACATCAAGGTGTTTTTGG ACTCCCTCGATGCCCAGGAGAGAGCTCTCACATCTGTGACTTCAT CCGAAAAACACTGAATGCTGGGGCGTACTCCAAAGTTGTTTCAGG AACGCCTCGTGCAAGCCGAATACTGGCATGACCCCATAAAGGAG GATGTGTATCGCAACCACAGCATCTTCTTGGCAGATATAAATCA GGAGCGGGGTATCAATGAGTCCTACAAGAAAAACCTGATGGCCC TGAAGAAGTTTGTGATGGTGAAATTCCTCAATGATTCCATTGTGG ACCCTGTAGATTTCGGAGTGGTTTGGATTTTACAGAAGTGGCCAA GCCAAGGAAACCATTCCCTTACAGGAGACCTCCCTGTACACACA GGACCGCCTGGGGCTAAAGGAAATGGACAATGCAGGACAGCTA GTGTTTCTGGCTACAGAAGGGGACCATCTTCAGTTGTCTGAAGA ATGGTTTTATGCCACATCATACCATTCCTTGAAGACCTAGAGC AGTGCCTACGCAGGGAGGGAGTGGGAGTGGATCCACTTCATCCT	230

	CTAGAACACTGTGCGGAGGGGAGCTTGTAGACACTCTTCAGTTC GTGTGTGGAGATCGCGGGTTCCTCTTCTCTCGCGGAGGTGGAGGT TCTAGGGGTATACTGGAGGAGTGTTGTTTCAGGGAGTGTGACTT GGCGCTCCTCGAGACCTATTGCGCGACGCCAGCCAGGTCCGAAT GA	
PPT1-33 (wt-PPT1- vIGF2-8Q; native human sequence)	ATGGCGTCGCCCCGGCTGCCTGTGGCTCTTGGCTGTGGCTCTCCTG CCATGGACCTGCGCTTCTCGGGCGCTGCAGCATCTGGACCCGCC GGCGCCGCTGCCGTTGGTGATCTGGCATGGGATGGGAGACAGCT GTTGCAATCCCTTAAGCATGGGTGCTATTAATAAATGGTGGAG AAGAAAATACCTGGAATTTACGTCTTATCTTTAGAGATTGGGAA GACCTGATGGAGGACGTGGAGAACAGCTTCTTCTTGAATGTCA ATTCCAAGTAACAACAGTGTGTCAGGCACTTGCTAAGGATCCT AAATTGCAGCAAGGCTACAATGCTATGGGATTCTCCAGGGAGG CCAATTTCTGAGGGCAGTGGCTCAGAGATGCCCTTACCTCCCAT GATCAATCTGATCTCGGTTGGGGGACAACATCAAGGTGTTTTTGG ACTCCCTCGATGCCCAGGAGAGAGCTCTCACATCTGTGACTTCAT CCGAAAAACACTGAATGCTGGGGCGTACTCCAAAGTTGTTTCAGG AACGCCTCGTGCAAGCCGAATACTGGCATGACCCCATAAAGGAG GATGTGTATCGCAACCACAGCATCTTCTTGGCAGATATAAATCA GGAGCGGGGTATCAATGAGTCCTACAAGAAAAACCTGATGGCCC TGAAGAAGTTTGTGATGGTGAAATTCCTCAATGATTCCATTGTGG ACCCTGTAGATTTCGGAGTGTTTGGATTTTACAGAAGTGGCCAA GCCAAGGAAACCATTCCCTTACAGGAGACCTCCCTGTACACACA GGACCGCCTGGGGCTAAAGGAAATGGACAATGCAGGACAGCTA GTGTTTCTGGCTACAGAAGGGGACCATCTTCAGTTGTCTGAAGA ATGGTTTTATGCCCACATCATACCATTCCCTTGGAAGACCTAGAGC AGTGCCTACGCAGGGAGGGAGTGGGAGTGGATCCACTTCATCCT CTAGAACACTGTGCGGAGGGGAGCTTGTAGACACTCTTCAGTTC GTGTGTGGAGATCGCGGGTTCCTCTTCTCTCGCCCCGCTTCCAGA GTTTCACGGAGGTCTAGGGGTATAGTAGAGGAGTGTGTTTCAG GGAGTGTGACTTGGCGCTCCTCGAGACCTATTGCGCGACGCCAG CCAGGTCCGAATGA	231

PPT1-34 (wt-PPT1- vIGF2-8Q; Codon optimized GENEius)	ATGGCCTCCCCAGGCTGCTTATGGTTGCTGGCCGTAGCACTTTTA CCATGGACATGTGCTAGTCGAGCTTTACAACACTTAGACCCGCC AGCGCCTCTTCCTTTAGTTATCTGGCACGGCATGGGCGACTCGTG TTGTAACCCGCTCAGTATGGGTGCCATAAAGAAGATGGTGGAGA AGAAAATTCCCGGAATCTATGTGCTTAGCCTCGAAATCGGCAAA ACACTTATGGAGGACGTAGAGAACTCATTCTTCCTGAATGTAAA TAGCCAAGTCACCACGGTATGTCAAGCTCTAGCGAAGGACCCTA AACTCCAGCAGGGGTATAACGCAATGGGATTTTCTCAGGGCGGC CAGTTTCTGCGTGCTGTCGCACAGCGTTGCCCTTCTCCGCCTATG ATAAACTTAATTTCCGTAGGAGGGCAACACCAAGGGGTATTCGG CTTACCGAGGTGTCCAGGCGAATCTTCACATATATGCGACTTCAT CCGAAAGACCCTTAATGCCGGGGCCTATTCCAAGGTGGTACAGG AACGGTTGGTGCAAGCTGAGTATTGGCACGACCCTATAAAGGAA GATGTGTATCGGAATCACTCAATCTTTCTTGCGGATATAAATCAA GAGCGCGGCATTAACGAGAGCTACAAGAAGAACCCTCATGGCTCT TAAGAAATTCGTCATGGTCAAATTCCTCAACGACAGTATAGTTG ATCCCGTCGATTTCGGAGTGGTTTGGATTCTACCGCTCTGGGCAAG CCAAAGAGACCATAACCACTACAGGAAACATCGCTATATACCCAA GATCGCTTGGGTTTGAAAGAAATGGATAACGCCGGTCAGCTTGT GTTCTTAGCGACAGAGGGTGATCATCTCCAGCTGTGCGAAGAAT GGTTCTATGCCCACATAATACCTTTCCTTGGACGACCCCGTGCGG TCCAACGCAGGGTGGATCAGGTAGCGGCTCAACTAGTTCCAGC CGTACGTTGTGCGGCGGAGAACTAGTAGACACTCTTCAATTCGTT TGTGGGGATCGGGGCTTCCTCTTCAGCAGGCCAGCGTCACGCGT GTCGCGTCGGAGCCGAGGTATAGTGGAAGAATGCTGCTTCCGCG AATGTGATCTAGCACTCCTTGAAACCTACTGCGCGACGCCTGCCC GAAGTGAATGA	232
PPT1-35 (wt-PPT1- vIGF2-8Q; Codon optimized)	ATGGCTTCCCCTGGCTGCCTGTGGCTGCTCGCTGTGGCCCTCCTG CCCTGGACCTGTGCTTCTCGGGCCCTTCAGCATCTGGACCCTCCA GCCCCCTCCCCTTGGTCATCTGGCACGGCATGGGCGACAGCTGC TGCAACCCTCTGTCCATGGGGGCCATCAAGAAAATGGTTGAGAA GAAGATCCCAGGCATCTACGTGCTGAGCCTGGAAATTGGCAAGA	233

COOL)	CACTGATGGAGGATGTGGAAAACAGCTTCTTCCTGAATGTGAAC TCCCAGGTGACCACCGTGTGCCAGGCTCTGGCCAAAGATCCCAA GCTGCAGCAGGGCTACAATGCCATGGGATTCAGCCAGGGGGGCC AGTTTCTGCGGGCTGTTGCCCAGAGGTGCCCCAGCCCCCATGA TCAATCTCATCTCTGTGGGCGGGCAGCACCAGGGTGTGTTTGGCC TGCCTCGCTGCCCTGGAGAAAAGCAGCCACATTTGTGATTTTCATCA GGAAGACCTTAAATGCTGGAGCCTACAGCAAGGTGGTCCAGGAA AGGCTGGTGCAGGCAGAGTACTGGCATGACCCCATCAAAGAGGA CGTGTACAGAAACCACAGCATCTTCCTGGCTGACATCAACCAGG AGAGAGGAATTAATGAGAGCTACAAGAAGAACCTCATGGCCTTG AAAAAGTTTGTGATGGTGAAGTTCTTGAATGACTCCATCGTGGAT CCTGTGGACAGTGAATGGTTTGGGTTCTACCGCTCTGGACAGGCC AAGGAAACCATCCCCCTGCAAGAAACATCCCTGTACACCCAGGA CCGCCTGGGGCTGAAGGAGATGGACAACGCCGGCCAACTGGTCT TCCTTGCCACAGAAGGAGACCACCTGCAGCTGTCTGAGGAGTGG TTCTATGCCCACATCATCCCCTTCCTGGGCGGGCCAGGGCCGTG CCCACACAGGGAGGCAGTGGCAGCGGCTCCACCAGCTCCAGCAG GACCTGTGTGGCGGCGAGCTGGTTGACACCCTCCAGTTCGTGTG TGGGGACAGAGGCTTCCTCTTCTCCAGGCCCGCCAGCCGGGTGA GCCGCCGCTCCCGGGGCATTGTGGAGGAATGTTGCTTCCGGGAG TGTGACCTGGCCCTGCTGGAGACCTACTGTGCCACCCCTGCCCGG AGTGAGTGA	
PPT1-101 (sequence encoding signal peptide underlined)	<u>ATGAAGCTCTCCCTGGTGGCCGCGATGCTGCTGCTGCTCTGGGTG</u> <u>GCACTGCTGCTGCTCAGCGCGGCGAGGGCCGCGCGTCTAGAAC</u> ACTGTGCGGAGGGGAGCTTGTAGACACTCTTCAGTTCGTGTGTG GAGATCGCGGGTTCCTCTTCTCTCGCGGAGGTGGAGGTTCTAGG GGTATACTGGAGGAGTGTGTTTTCAGGGACTGTGACTTGGCGCTC CTCGAGACCTATTGCGCGACGCCAGCCAGGTCCGAAGGAGGTGG TGGCAGTGGAGGAGGAGGGAGTCGGCCTAGGGCAGTCCCAACC CAGGACCCGCGGCGCGGCTGCCGTTGGTGATCTGGCATGGGAT GGGAGACAGCTGTTGCAATCCCTTAAGCATGGGTGCTATTAAAA AAATGGTGGAGAAGAAAATACCTGGAATTTACGTCTTATCTTTA	234

	GAGATTGGGAAGACCCTGATGGAGGACGTGGAGAACAGCTTCTT CTTGAATGTCAATTCCCAAGTAACAACAGTGTGTCAGGCACTTGC TAAGGATCCTAAATTGCAGCAAGGCTACAATGCTATGGGATTCT CCCAGGGAGGCCAATTTCTGAGGGCAGTGGCTCAGAGATGCCCT TCACCTCCCATGATCAATCTGATCTCGGTTGGGGGACAACATCAA GGTGTTTTTGGACTCCCTCGATGCCCAGGAGAGAGCTCTCACATC TGTGACTTCATCCGAAAAACACTGAATGCTGGGGCGTACTCCAA AGTTGTTTCAGGAACGCCTCGTGCAAGCCGAATACTGGCATGACC CCATAAAGGAGGATGTGTATCGCAACCACAGCATCTTCTTGGCA GATATAAATCAGGAGCGGGGTATCAATGAGTCCTACAAGAAAAA CCTGATGGCCCTGAAGAAGTTTGTGATGGTGAAATTCTCAATG ATTCCATTGTGGACCCTGTAGATTTCGGAGTGGTTTGGATTTTACA GAAGTGGCCAAGCCAAGGAAACCATTCCTTACAGGAGACCTCC CTGTACACACAGGACCGCCTGGGGCTAAAGGAAATGGACAATGC AGGACAGCTAGTGTCTTCTGGCTACAGAAGGGGACCATCTTCAGT TGTCTGAAGAATGGTTTTATGCCACATCATACCATTCTTGGAT GA	
PPT1-104 (sequence encoding signal peptide underlined)	<u>ATGGCGTCGCCCCGGCAGCCTGTGGCTCTTGGCTGTGGCTCTCCTG</u> <u>CCATGGACCTGCGCTTCTCGGGCGCTGCAGCATCTGGACCCGCC</u> GGCGCCGCTGCCGTTGGTGATCTGGCATGGGATGGGAGACAGCT GTTGCAATCCCTTAAGCATGGGTGCTATTAAAAAAATGGTGGAG AAGAAAATACCTGGAATTTACGTCTTATCTTTAGAGATTGGGAA GACCTGATGGAGGACGTGGAGAACAGCTTCTTCTTGAATGTCA ATTCCCAAGTAACAACAGTGTGTCAGGCACTTGCTAAGGATCCT AAATTGCAGCAAGGCTACAATGCTATGGGATTCTCCAGGGAGG CCAATTTCTGAGGGCAGTGGCTCAGAGATGCCCTTCACCTCCCAT GATCAATCTGATCTCGGTTGGGGGACAACATCAAGGTGTTTTTGG ACTCCCTCGATGCCCAGGAGAGAGCTCTCACATCTGTGACTTCAT CCGAAAAACACTGAATGCTGGGGCGTACTCCAAAGTTGTTTCAGG AACGCCTCGTGCAAGCCGAATACTGGCATGACCCCATAAAGGAG GATGTGTATCGCAACCACAGCATCTTCTTGGCAGATATAAATCA	235

	GGAGCGGGGTATCAATGAGTCCTACAAGAAAAACCTGATGGCCC TGAAGAAGTTTGTGATGGTGAAATTCCTCAATGATTCCATTGTGG ACCCTGTAGATTTCGGAGTGGTTTGGATTTTACAGAAGTGGCCAA GCCAAGGAAACCATTCCCTTACAGGAGACCTCCCTGTACACACA GGACCGCCTGGGGCTAAAGGAAATGGACAATGCAGGACAGCTA GTGTTTCTGGCTACAGAAGGGGACCATCTTCAGTTGTCTGAAGA ATGGTTTTATGCCACATCATACCATTCCCTTGAAGACCTAGAGC AGTGCCTACGCAGGGAGGGAGTGGGAGTGGATCCACTTCATCCT CTAGAACACTGTGCGGAGGGGAGCTTGTAGAACTCTTCAGTTC GTGTGTGGAGATCGCGGGTTCCTCTTCTCTCGCGGAGGTGGAGGT TCTAGGGGTATACTGGAGGAGTGTGTTTCAGGGAGTGTGACTT GGCGCTCCTCGAGACCTATTGCGCGACGCCAGCCAGGTCCGAAT GA	
PPT-112 (sequence encoding signal peptide underlined)	<u>ATGGCGTCGCCCCGGCAGCCTGTGGCTCTTGGCTGTGGCTCTCCTG</u> <u>CCATGGACCTGCGCTTCTCGGGCGCTGCAGCATCTGGCCGCGTCT</u> AGAACACTGTGCGGAGGGGAGCTTGTAGAACTCTTCAGTTCGT GTGTGGAGATCGCGGGTTCCTCTTCTCTCGCGGAGGTGGAGGTTC TAGGGGTATACTGGAGGAGTGTGTTTCAGGGACTGTGACTTGG CGCTCCTCGAGACCTATTGCGCGACGCCAGCCAGGTCCGAAGGA GGTGGTGGCAGTGGAGGAGGAGGGAGTCGGCCTAGGGCAGTCC CAACCCAGGACCCGCCGCGCCGCTGCCGTTGGTGATCTGGCAT GGGATGGGAGACAGCTGTTGCAATCCCTTAAGCATGGGTGCTAT TAAAAAATGGTGGAGAAGAAAATACCTGGAATTTACGTCTTAT CTTTAGAGATTGGGAAGACCCTGATGGAGGACGTGGAGAACAGC TTCTTCTTGAATGTCAATTCCCAAGTAACAACAGTGTGTCAGGCA CTTGCTAAGGATCCTAAATTGCAGCAAGGCTACAATGCTATGGG ATTCTCCCAGGGAGGCCAATTTCTGAGGGCAGTGGCTCAGAGAT GCCCTTCACCTCCCATGATCAATCTGATCTCGGTTGGGGGACAAC ATCAAGGTGTTTTTGGACTCCCTCGATGCCCAGGAGAGAGCTCTC ACATCTGTGACTTCATCCGAAAAACACTGAATGCTGGGGCGTAC TCCAAAGTTGTTTCAGGAACGCCTCGTGCAAGCCGAATACTGGCA	236

	TGACCCCATAAAGGAGGATGTGTATCGCAACCACAGCATCTTCT TGGCAGATATAAATCAGGAGCGGGGTATCAATGAGTCCTACAAG AAAAACCTGATGGCCCTGAAGAAGTTTGTGATGGTGAAATTCCT CAATGATTCCATTGTGGACCCTGTAGATTTCGGAGTGGTTTGGATT TTACAGAAGTGGCCAAGCCAAGGAAACCATTCCTTACAGGAGA CCTCCCTGTACACACAGGACCGCCTGGGGCTAAAGGAAATGGAC AATGCAGGACAGCTAGTGTTTCTGGCTACAGAAGGGGACCATCT TCAGTTGTCTGAAGAATGGTTTTATGCCACATCATACCATTCCT TGGATGA	
PPT-114 (sequence encoding signal peptide underlined)	<u>ATGGCGTCGCCCCGGCAGCCTGTGGCTCTTGGCTGTGGCTCTCCTG</u> <u>CCATGGACCTGCGCTTCTCGGGCGCTGCAGCATCTGGCCGCGTCT</u> AGAACTGTGCGGAGGGGAGCTTGTAGAACTCTTCAGTTCGT GTGTGGAGATCGCGGGTTCCTCTTCTCTCGCGGAGGTGGAGGTTC TAGGGGTATACTGGAGGAGTGTGTTTCAGGGAGTGTGACTTGG CGCTCCTCGAGACCTATTGCGCGACGCCAGCCAGGTCCGAAGGA GGTGGTGGCAGTGGAGGAGGAGGGAGTCGGCCTAGGGCAGTCC CAACCCAGGACCCGCCGGCGCCGCTGCCGTTGGTGATCTGGCAT GGGATGGGAGACAGCTGTTGCAATCCCTTAAGCATGGGTGCTAT TAAAAAATGGTGGAGAAGAAAATACCTGGAATTTACGTCTTAT CTTTAGAGATTGGGAAGACCCTGATGGAGGACGTGGAGAACAGC TTCTTCTTGAATGTCAATTCCCAAGTAACAACAGTGTGTCAGGCA CTTGCTAAGGATCCTAAATTGCAGCAAGGCTACAATGCTATGGG ATTCTCCCAGGGAGGCCAATTTCTGAGGGCAGTGGCTCAGAGAT GCCCTTCACCTCCCATGATCAATCTGATCTCGGTTGGGGGACAAC ATCAAGGTGTTTTTGGACTCCCTCGATGCCCAGGAGAGAGCTCTC ACATCTGTGACTTCATCCGAAAAACACTGAATGCTGGGGCGTAC TCCAAAGTTGTTCAAGAACGCCTCGTGCAAGCCGAATACTGGCA TGACCCCATAAAGGAGGATGTGTATCGCAACCACAGCATCTTCT TGGCAGATATAAATCAGGAGCGGGGTATCAATGAGTCCTACAAG AAAAACCTGATGGCCCTGAAGAAGTTTGTGATGGTGAAATTCCT CAATGATTCCATTGTGGACCCTGTAGATTTCGGAGTGGTTTGGATT	237

	TTACAGAAGTGGCCAAGCCAAGGAAACCATTCCTTACAGGAGA CCTCCCTGTACACACAGGACCGCCTGGGGCTAAAGGAAATGGAC AATGCAGGACAGCTAGTGTTTCTGGCTACAGAAGGGGACCATCT TCAGTTGTCTGAAGAATGGTTTTATGCCCACATCATACCATTCCT TGGATGA	
PPT1-115 (sequence encoding signal peptide underlined)	<u>ATGGCGTCGCCCCGGCAGCCTGTGGCTCTTGGCTGTGGCTCTCCTG</u> <u>CCATGGACCTGCGCTTCTCGGGCGCTGCAGCATCTGGCTGCCGAC</u> CCGCCGGCGCCGCTGCCGTTGGTGATCTGGCATGGGATGGGAGA CAGCTGTTGCAATCCCTTAAGCATGGGTGCTATTAaaaaaATGG TGGAGAAGAAAATACCTGGAATTTACGTCTTATCTTTAGAGATTG GGAAGACCCTGATGGAGGACGTGGAGAACAGCTTCTTCTTGAAT GTCAATTCCCAAGTAACAACAGTGTGTCAGGCACTTGCTAAGGA TCCTAAATTGCAGCAAGGCTACAATGCTATGGGATTCTCCCAGG GAGGCCAATTTCTGAGGGCAGTGGCTCAGAGATGCCCTTCACCT CCCATGATCAATCTGATCTCGGTTGGGGGACAACATCAAGGTGT TTTTGGACTCCCTCGATGCCCAGGAGAGAGCTCTCACATCTGTGA CTTCATCCGAAAAACACTGAATGCTGGGGCGTACTCCAAAGTTG TTCAGGAACGCCTCGTGCAAGCCGAATACTGGCATGACCCCAT AAGGAGGATGTGTATCGCAACCACAGCATCTTCTTGGCAGATAT AAATCAGGAGCGGGGTATCAATGAGTCCTACAAGAAAAACCTGA TGGCCCTGAAGAAGTTTGTGATGGTGAAATTCCTCAATGATTCCA TTGTGGACCCTGTAGATTCCGAGTGGTTTGGATTTTACAGAAGTG GCCAAGCCAAGGAAACCATTCCCTTACAGGAGACCTCCCTGTAC ACACAGGACCGCCTGGGGCTAAAGGAAATGGACAATGCAGGAC AGCTAGTGTTTCTGGCTACAGAAGGGGACCATCTTCAGTTGTCTG AAGAATGGTTTTATGCCCACATCATACCATTCCTTGAAGACCTA GAGCAGTGCCTACGCAGGGAGGGAGTGGGAGTGGATCCACTTCA TCCTCTAGAACACTGTGCGGAGGGGAGCTTGTAGACACTCTTCA GTTTCGTGTGTGGAGATCGCGGGTTCCTCTTCTCTCGCGGAGGTGG AGGTTCTAGGGGTATACTGGAGGAGTGTGTTTCAGGGAGTGTG ACTTGGCGCTCCTCGAGACCTATTGCGCGACGCCAGCCAGGTCC	238

	GAATGA	
PPT-116 (sequence encoding signal peptide underlined)	<u>ATGGCGTCGCCCCGGCAGCCTGTGGCTCTTGGCTGTGGCTCTCCTG</u> <u>CCATGGACCTGCGCTTCTCGGGCGCTGCAGCATCTGGCTGCCGAC</u> CCGCCGGCGCCGCTGCCGTTGGTGATCTGGCATGGGATGGGAGA CAGCTGTTGCAATCCCTTAAGCATGGGTGCTATTAaaaaaaATGG TGGAGAAGAAAATACCTGGAATTTACGTCTTATCTTTAGAGATTG GGAAGACCCTGATGGAGGACGTGGAGAACAGCTTCTTCTTGAAT GTCAATTCCCAAGTAACAACAGTGTGTCAGGCACTTGCTAAGGA TCCTAAATTGCAGCAAGGCTACAATGCTATGGGATTCTCCCAGG GAGGCCAATTTCTGAGGGCAGTGGCTCAGAGATGCCCTTCACCT CCCATGATCAATCTGATCTCGGTTGGGGGACAACATCAAGGTGT TTTTGGACTCCCTCGATGCCCAGGAGAGAGCTCTCACATCTGTGA CTTCATCCGAAAAACACTGAATGCTGGGGCGTACTCCAAAGTTG TTCAGGAACGCCTCGTGCAAGCCGAATACTGGCATGACCCATA AAGGAGGATGTGTATCGCAACCACAGCATCTTCTTGGCAGATAT AAATCAGGAGCGGGGTATCAATGAGTCCTACAAGAAAAACCTGA TGGCCCTGAAGAAGTTTGTGATGGTGAAATTCCTCAATGATTCCA TTGTGGACCCTGTAGATTCCGAGTGGTTTGGATTTTACAGAAGTG GCCAAGCCAAGGAAACCATTCCCTTACAGGAGACCTCCCTGTAC ACACAGGACCGCCTGGGGCTAAAGGAAATGGACAATGCAGGAC AGCTAGTGTTTCTGGCTACAGAAGGGGACCATCTTCAGTTGTCTG AAGAATGGTTTTATGCCACATCATAACCATTCCTTGGAAGACCTA GAGCAGTGCCTACGCAGGGAGGGGGTGGCAGTGGCAGTGGAGG CGGCGGTTCCCTCTAGAACACTGTGCGGAGGGGAGCTTGTAGACA CTCTTCAGTTCGTGTGTGGAGATCGCGGGTTCCTCTTCTCTCGCG GAGGTGGAGGTTCTAGGGGTATACTGGAGGAGTGTGTTTCAGG GAGTGTGACTTGGCGCTCCTCGAGACCTATTGCGCGACGCCAGC CAGGTCCGAATGA	239
PPT1-117 (sequence)	<u>ATGGCGTCGCCCCGGCAGCCTGTGGCTCTTGGCTGTGGCTCTCCTG</u> <u>CCATGGACCTGCGCTTCTCGGGCGCTGCAGCATCTGGACCCGCC</u>	240

encoding signal peptide underlined)	GGCGCCGCTGCCGTTGGTGATCTGGCATGGGATGGGAGACAGCT GTTGCAATCCCTTAAGCATGGGTGCTATTAAAAAATGGTGGAG AAGAAAATACCTGGAATTTACGTCTTATCTTTAGAGATTGGGAA GACCCTGATGGAGGACGTGGAGAACAGCTTCTTCTTGAATGTCA ATTCCCAAGTAACAACAGTGTGTCAGGCACTTGCTAAGGATCCT AAATTGCAGCAAGGCTACAATGCTATGGGATTCTCCCAGGGAGG CCAATTTCTGAGGGCAGTGGCTCAGAGATGCCCTTCACCTCCCAT GATCAATCTGATCTCGGTTGGGGGACAACATCAAGGTGTTTTTG ACTCCCTCGATGCCCAGGAGAGAGCTCTCACATCTGTGACTTCAT CCGAAAAACACTGAATGCTGGGGCGTACTCCAAAGTTGTTTCAGG AACGCCTCGTGCAAGCCGAATACTGGCATGACCCCATAAAGGAG GATGTGTATCGCAACCACAGCATCTTCTTGGCAGATATAAATCA GGAGCGGGGTATCAATGAGTCCTACAAGAAAAACCTGATGGCCC TGAAGAAGTTTGTGATGGTGAAATTCCTCAATGATTCCATTGTGG ACCCTGTAGATTTCGGAGTGGTTTGGATTTTACAGAAGTGGCCAA GCCAAGGAAACCATTCCCTTACAGGAGACCTCCCTGTACACACA GGACCGCCTGGGGCTAAAGGAAATGGACAATGCAGGACAGCTA GTGTTTCTGGCTACAGAAGGGGACCATCTTCAGTTGTCTGAAGA ATGGTTTTATGCCACATCATACCATTCCTTGGAAGACCTAGAGC AGTGCCTACGCAGGGAGGGGGTGGCAGTGGCAGTGGAGGCGGC GGTTCCTCTAGAACACTGTGCGGAGGGGAGCTTGTAGACACTCT TCAGTTCGTGTGTGGAGATCGCGGGTTCCTCTTCTCTCGCGGAGG TGGAGGTTCTAGGGGTATACTGGAGGAGTGTGTTTCAGGGAGT GTGACTTGGCGCTCCTCGAGACCTATTGCGCGACGCCAGCCAGG TCCGAATGA	
PPT-118 (sequence encoding signal peptide underlined)	<u>ATGGCGTCGCCCCGGCAGCCTGTGGCTCTTGGCTGTGGCTCTCCTG</u> <u>CCATGGACCTGCGCTTCTCGGGCGCTGCAGCATCTGGCTGCCGAC</u> CCGCCGGCGCCGCTGCCGTTGGTGATCTGGCATGGGATGGGAGA CAGCTGTTGCAATCCCTTAAGCATGGGTGCTATTAAAAAATGG TGGAGAAGAAAATACCTGGAATTTACGTCTTATCTTTAGAGATTG GGAAGACCCTGATGGAGGACGTGGAGAACAGCTTCTTCTTGAAT	241

	GTCAATTCCCAAGTAACAACAGTGTGTCAGGCACTTGCTAAGGA TCCTAAATTGCAGCAAGGCTACAATGCTATGGGATTCTCCCAGG GAGGCCAATTTCTGAGGGCAGTGGCTCAGAGATGCCCTTCACCT CCCATGATCAATCTGATCTCGGTTGGGGGACAACATCAAGGTGT TTTTGGACTCCCTCGATGCCCAGGAGAGAGCTCTCACATCTGTGA CTTTCATCCGAAAAACACTGAATGCTGGGGCGTACTCCAAAGTTG TTCAGGAACGCCTCGTGCAAGCCGAATACTGGCATGACCCCAT AAGGAGGATGTGTATCGCAACCACAGCATCTTCTTGGCAGATAT AAATCAGGAGCGGGGTATCAATGAGTCCTACAAGAAAAACCTGA TGGCCCTGAAGAAGTTTGTGATGGTGAAATTCCTCAATGATTCCA TTGTGGACCCTGTAGATTCCGGAGTGGTTTGGATTTTACAGAAGTG GCCAAGCCAAGGAAACCATTCCCTTACAGGAGACCTCCCTGTAC ACACAGGACCGCCTGGGGCTAAAGGAAATGGACAATGCAGGAC AGCTAGTGTTTCTGGCTACAGAAGGGGACCATCTTCAGTTGTCTG AAGAATGGTTTTATGCCCACATCATACCATTCCCTTGGAAAGACCTA GAGCAGTGCCTACGCAGGGAGGGGGTGGCAGTGGAGGCGGCGG TTCCTCTAGAACACTGTGCGGAGGGGAGCTTGTAGACACTCTTCA GTTCGTGTGTGGAGATCGCGGGTTCCTCTTCTCTCGCGGAGGTGG AGGTTCTAGGGGTATACTGGAGGAGTGTGTTTCAGGGAGTGTG ACTTGGCGCTCCTCGAGACCTATTGCGCGACGCCAGCCAGGTCC GAATGA	
BiP2AA	ATGAAGCTCTCCCTGGTGGCCGCGATGCTGCTGCTGCTCTGGGTG GCACTGCTGCTGCTCAGCGCGGCGAGGGCCGCGCG	242
eSP C6S	ATGGCGTCGCCCCGGCAGCCTGTGGCTCTTGGCTGTGGCTCTCCTG CCATGGACCTGCGCTTCTCGGGCGCTGCAGCATCTG	243
eSP C6S AA (used in PPT1-112 and PPT1-	ATGGCGTCGCCCCGGCAGCCTGTGGCTCTTGGCTGTGGCTCTCCTG CCATGGACCTGCGCTTCTCGGGCGCTGCAGCATCTGGCCGCG	244

114)		
eSP C6S AA (used in PPT1-115, PPT1-116, and PPT1- 118) – different codon usage for AA portion	ATGGCGTCGCCCCGGCAGCCTGTGGCTCTTGGCTGTGGCTCTCCTG CCATGGACCTGCGCTTCTCGGGCGCTGCAGCATCTGGCTGCC	245
WT NAGLU- HPC4 (sequence encoding HPC4 capitalized)	ATGgaggccgtggccgtggccgccgccgtggcgctgctgctggccggcgccggcgccgcc ggcgacgaggccgggaggccgccgccgtggcgccctggtggccggctgctgggccccggcccc ggcgccgacttcagcgttagcgtggagcgggcccgtggccccaagccggcctggacacctacagcctg ggcgggcgccggcgccggcggtgcgggtgcggggcagcaccggcggtggccgccggcgccggcctg caccggatatctgcgggacttctgcggctgccacgtggcctggagcggcagccagctgcggctgccccgg cccctgcccgccgtggcgggcagctgaccgaggccacccccaaaccggtatcggtactaccagaacgtg tgcaccagagctacagcttcgtgtgtgggactgggccccgtgggagcgggagatcgactggatggcc ctgaacggcatcaacctggccctggcctggagcggccaggaggccatctggcagcgggtgtacctggcc ctgggctgacctaggccgagatcaacgagttcttcaccggccccgccttctggcctggggccggatgg gcaacctgcacacctgggacggccccctgctccaagctggcacatcaagcagctgtacctgcagaccg ggtgctggaccagatgcggagcttcggcatgacccccgtgctgcccgccttcgccggccacgtgcccgag gccgtgacccgggtgttccccaaagtaacctgaccaagatgggcagctggggccacttcaactgcagcta cagctgcagcttctgctggccccgaggaccccatcttcccatcatcggcagcctgttctgcgggagct gatcaaggagttcggcaccgaccacatctacggcgccgacaccttcaacgagatgcagcctccaagcagc gagcccagctacctggccgccgccaccaccgccgtgtacgaggccatgaccgccgtggacaccgaggc cgtgtggctgctgcagggtggctgttccagcaccagccccagttctggggacctgcccagatccgggcc gtgctggcgccgtgcctagaggacggctgctggtgctggacctgttcgccgagagccagcccgtgtaca cccgaggccagcttccaggggccgcccttcatctggtgcatgctgcacaacttcggcggaaccacgg cctgttcggcgccctggaggccgtgaacggcgggccccgaggccgccggctgttccccaaagcaccat ggtgggcaccggcatggccccgagggcacccaggaacgaggtggtgtacagcctgatggccgagc	246

	tgggctggcggaaggaccccggtgcccgaacctggccgcctgggtgaccagcttcgccgcccggcggtac ggcgtagccaccccgacccggcgccgcctggcggtgctgctgctgaggagcgtgtacaactgcagcgg cgaggcctgccggggccacaaccggagccccctggtgcgggcgccagcctgcagatgaacaccagc atctggtacaaccggagcgacgtgttcgaggcctggcggtgctgctgaccagcgcggccagcctggcca ccagccccgccttcagatacgacctgctggacctgacctggcaggccgtgcaggagctggtgagcctgta ctacgaggaggccccggagcgctacctgagcaaggagctggccagcctgctgcggggcgccggcggtgc tggcctacgagctgctgcccgcctggacgaggtgctggccagcgacagccggttcctgctgggcagctg gctggagcaggccccggcgccgccgtgagcgaggccgaggccgactttacgagcagaacagccgg tatcagctgacctgtggggacctgagggcaacatcctggactacgccaacaagcagctggccggcctgg tggccaactactacccccaaagtgggcggtgttcctggaggccctggtggacagcgtggcccagggcat ccccttcagcagcaccagttcgacaagaacgtgtccagctggagcaggccttcgtgctgagcaagcagc ggtatcccagccagcctagaggagacaccgtggacctggccaagaagatcttcctgaagtactacccccg gtgggtggccggcagctggggaCTTGAGGTACTGTTCCAAGGGCCCGAGGAC CAGGTAGACCCACGACTCATTGATGGAAAATAG	
vIGF2- NAGLU- HPC4	ATGGAGGCTGTGGCTGTGGCAGCTGCGGTGGGGGTCCTTCTCCT GGCCGGGGCCGGGGGCGCGGCAGGCGACGcTTCTAGGACGTTGT GTGGTGGGGAACCTTGTCGACACACTGCAGTTTGTCTGCGGCGAC CGAGGATTTCTTTTTTCCAGGCCTGCCTCAAGAGTATCTAGGAGG TCCCGCGGTATTGTTGAAGAGTGCTGTTTTAGGTCATGCGACCTT GCGTTGTTGGAGACATATTGTGCTACCCCTGCACGCTCTGAAGGT GGAGGTGGTTCAGGTGGTGGAGGTTCCAGGCCAAGGGCGGTCCC TACTCAGGCCgaggccccggaggccgcccggtgcgggcccctggtggccccggtgctggggc ccggccccgccgacttcagcgttagcgtggagcgggcccctggccgccaagccccggcctggacacct acagcctgggcggcgggcgccgccccgggtgcgggtgcggggcagcaccggcgtggccgccgccc ccggcctgcaccggtatctgcgggacttctgcggctgccacgtggcctggagcggcagccagctgcggct gccccggccccctgcccgccgtgccccggcgagctgaccgaggccacccccaaaccggtatcggtactacca gaacgtgtgacccagagctacagcttcgtgtggtgggactggggccggtgggagcgggagatcgactg gatggccctgaacggcatcaacctggccctggcctggagcggccaggaggccatctggcagcgggtgta cctggccctgggctgaccagccgagatcaacgagttcttcaccggccccgccttcctggcctggggc cggatgggcaacctgcacacctgggacggccccctgcctccaagctggacatcaagcagctgtacctgc agcaccgggtgctggaccagatgcgggagcttcggcatgacccccgtgctgccccgccttcggcgccacgt	247

	gcccagaggccgtgacccgggtgttcccccaagttaacgtgaccaagatgggcagctggggccacttcaac tgcagctacagctgcagcttctctgtgtgccccgaggaccccatcttccccatcatcggcagcctgttctctg cgggagctgatcaaggagttcggcaccgaccacatctacggcgccgacaccttcaacgagatgcagcctc caagcagcgagcccagctacctggccgcccaccaccgccgtgtacgaggccatgaccgccgtggac accgaggccgtgtggctgtgcagggtgtgtgttccagcaccagccccagttctggggacctgcccaga tccggggccgtgtgtggcgccgtgcctagaggacggctgtgtgtgtgtggacctgttcgccgagagccagc ccgtgtacaccggaccgccagcttccagggccagcccttcatctgtgtcatgtgtgcacaacttcggcggc aaccacggcctgttcggcgccctggaggccgtgaacggcgggccccgaggccgcccggtgttccccaac agcaccatgtgtggcaccggcatggccccgagggcacagccagaacgaggtggtgtacagcctgatg gccgagctgggctggcggaaggacccgtgcccgacctggccgctgggtgaccagcttcgccggccg gcggtacggcgtagccaccccgacggcgccgcccgtggcggtgtgtgtgcggagcgtgtacaact gcagcgggcgaggccgtgccggggccacaaccggagccccctgtgtgcggcgggccagcctgcagatgaa caccagcatctgtgtacaaccggagcgacgtgttcgaggcctggcggtgtgtgtaccagcgccccag cctggccaccagccccgcttcagatacgacctgtgtggacctgacctggcagggcgtgcaggagctggt gagcctgtactacgaggaggcccgagcgccctacctgagcaaggagctggccagcctgtgtcgggccg gcgcgctgtgtggctacgagctgtgtccccccctggacgaggtgtgtggcagcgacagccggttcctgt gggagctgggtggagcaggccccggcgccgccgtgagcgaggccgaggccgacttctacgagcag aacagccggtatcagctgacctgtgtgggacctgagggcaacatcctggactacgccaacaagcagctgg ccggcctggtggccaactactacaccccaaggtggcggtgttcttgaggccctggtggacagcgtggc ccagggcaccccccttcagcagcaccagttcgacaagaacgtgttcagctggagcaggccttcgtgtga gcaagcagcggtatcccagccagccctagaggagacaccgtggacctggccaagaagatcttctgaagta ctacccccggtgggtggccggcagctggggaCTTGAGGTACTGTTCCAAGGGCCC GAGGACCAGGTAGACCCACGACTCATTGATGGAAAATAG	
vIGF2-17- NAGLU- HPC4	ATGGAGGCTGTGGCTGTGGCAGCTGCGGTGGGGGTCTTCTCTCT GGCCGGGGCCGGGGGGCGCGGCAGGCGACGcTagcagaacactttgtggcgga gagctggtggacacctgcagttgtgtgtggcgacagaggcttctgttcagcagacctgcacccagagt agcaggcggtccagaggaatcgtggaagagtgtgtcttcagaGAAtgcgatctggccctgtgtgaaac ctactgtgccacaccagccagatctgaaGGTGGAGGTGGTTTCAGGTGGTGGAGGT TCCAGGCCAAGGGCGGTCCCTACTCAGGCCgaggccccggaggccgccgcc gtgcgggccctgtgtggccggctgtgtggccccggccccggccgaccttcagcgttagcgtggagcgg gccctggccgccaagccccggcctggacacctacagcctggcgggcgccggcgccccgggtgtcgggg	248

	tgcggggcagcaccggcgtggccgccgccggcctgcaccggtatctgcgggacttctcggtgcc acgtggcctggagcggcagccagctcggtgccccggccccctgcccgccgtccccggcagctgacc gaggccaccccaaccggtatcggtactaccagaacgtgtgacccagagctacagcttcgtgtgtggg actgggcccgggtgggagcgggagatcgactggatggccctgaacggcatcaacctggccctggcctgga gggccaggaggccatctggcagcgggtgtacctggccctggccctgacccaggccgagatcaacgagt tcttcaccggccccgccttctggcctggggccggatgggcaacctgcacacctgggacggccccctgcc tccaagctggcacatcaagcagctgtacctgcagcaccgggtgctggaccagatcgaggacttcggcatg acccccgtgctgcccgccttcgcggccacgtgcccggaggccgtgacccgggtgttccccaaagttaacg tgaccaagatgggcagctggggccacttcaactgcagctacagctgcagcttctgctggccccgagga ccccatcttccccatcatcggcagcctgttctgctggggagctgatcaaggagtccggcaccgaccacatcta cggcgccgacaccttaacgagatgcagcctccaagcagcggagcccagctacctggccggccaccac cgccgtgtacgaggccatgaccgccgtggacaccgaggccgtgtggctgctgcagggtggtgttccag caccagccccagttctggggacctgcccagatccggccgtgctggggccgtgcttagaggacggctg ctggtgctggacctgttcgccgagagccagcccgtgtacaccggaccgccagcttccaggggccagccct tcatctggtgcatgctgcacaacttcggcggaaccacggcctgttcggcgccctggaggccgtgaacgg cgccccgaggccgcccggctgttccccaacagcaccatggtgggcaccggcatggccccgagggca tcagccagaacgaggtggtgtacagcctgatggccgagctgggctggcgggaaggaccccgtgcccgacc tggccgcctgggtgaccagcttcgccccccggcggtacggcgtgagccaccccagccggcgccgcc tggcggtgctgctgcggagcgtgtacaactgcagcggcgaggcctgccggggccacaaccggagccc cctggtgcggcgccagcctgcagatgaacaccagcatctggtacaaccggagcgacgtgttcgaggc ctggcggtgctgctgaccagcgccccagcctggccaccagccccgccttcagatacgacctgctggac ctgacccggcaggccgtgcaggagctggtgagcctgtactacgaggaggcccgagcgctacctgagc aaggagctggccagcctgctgcggggccggcggtgctggcctacgagctgctgcccgcctggacga ggtgctggccagcgacagccggttctgctgggcagctggctggagcaggccggggccggccgtga gcgaggccgaggccgacttctacgagcagaacagccggtatcagctgacctgtggggacctgagggca acatcctggactacgccaacaagcagctggccggcctggtggccaactactacacccaaggtggcggt gttcctggaggccctggtggacagcgtggcccaggcatcccccttcagcagcaccagttcgacaagaac gtgttccagctggagcaggccttcgtgctgagcaagcagcggtatcccagccagcctagaggagacaccg tggacctggccaagaagatcttctgaagtactacccccggtgggtggccggcagctggggaCTTGA GGTACTGTTCCAAGGGCCCGAGGACCAGGTAGACCCACGACTCA TTGATGGAAAATAG	
--	--	--

vIGF2-31- NAGLU- HPC4	ATGGAGGCTGTGGCTGTGGCAGCTGCGGTGGGGGTCCTTCTCCT GGCCGGGGCCGGGGGCGCGGCAGGCGACGcTagcagaacactttgtggcgga gagctggtggacacctgcagtttgtgtggcgacagaggcttctgttcagcagaGGTGGAGGT GGAtctagaggaatcCTGgaagagtgtctgttcagaGATtgcgatctggccctgctgaaacctact gtgccacaccagccagatctgaaGGTGGAGGTGGTTTCAGGTGGTGGAGGTTC CAGGCCAAGGGCGGTCCCTACTCAGGCCgaggcccgaggccgccgctg cgggcccctgtgtggcccggtgtgtggccccggccccgccgacttcagcgttagcgtggagcgggc cctggccccaagcccggcctggacacctacagcctggcgggcgccggcgccggcggtgctgggtg cggggcagcaccggcgtggcccgcccgccggcctgcaccggtatctgcgggacttctgcggtgccac gtggcctggagcggcagccagctgcggctgccccggccccctgcccggcgagctgaccga ggccaccccaaccggatcgggtactaccagaacgtgtgcaccagagctacagcttcgtgtgtgggact gggcccgggtgggagcgggagatcgactggatggccctgaacggcatcaacctggccctggcctggagc ggccaggaggccatctggcagcgggtgtacctggccctgggcctgaccaggccgagatcaacgagttc ttcaccggccccgccttctgtgctggggccggatgggcaacctgcacacctgggacggccccctgcctc caagctggcacatcaagcagctgtacctgcagcaccgggtgctggaccagatcgcgagcttcggcatgac ccccgtgctgcccgccttcgcccggccacgtgcccaggccgtgaccgggtgttccccaagftaacgtg accaagatgggcagctggggccacttcaactgcagctacagctgcagcttctgtgcccccgaggacc ccatcttccccatcatcgccagcctgttctgcgggagctgatcaaggagtccggcaccgaccacatctacg gcgccgacaccttcaacgagatgcagcctccaagcgcgagcccagctacctggccgccgcccaccaccg ccgtgtacgaggccatgaccgccgtggacaccgaggccgtgtggctgctgcagggtggtgtgtccagca ccagccccagttctggggacctgcccagatccgggccgtgctggcgccgtgcctagaggacggctgct ggtgtgtgacctgttcgcccagagccagcccgtgtacacccggaccgccagcttcagggccagcccttc atctggtgcatgctgcacaacttcggcggcaaccacggcctgttcggcgccctggaggccgtgaacggcg gccccgaggccgccggctgttcccaaacgaccatggtgggcaccggcatggccccgagggcac agccagaacgaggtggtgtacagcctgatggccgagctgggctggcgaaggaccccggtcccgaact ggccgcctgggtgaccagcttcgcccccggcggtacggcgtgagccaccccgacggcgccgcccct ggcggtgctgctgcggagcgtgtacaactgcagcggcgaggcctgcccggggccacaaccggagcccc ctggtgcggcgggccagcctgcagatgaacaccagcatctggtacaaccggagcgacgtgttcgaggcct ggcggtgctgctgaccagcggccccagcctggccaccagccccgccttcagatacgacctgtggacct gaccggcaggccgtgcaggagctggtgagcctgtactacaggaggcccgagcgccctacctgagca aggagctggccagcctgctgcggcgccggcggtgctggcctacgagctgctgcccgcctggacgag gtgctggccagcgacagccggttctgctgggcagctggctggagcagcccgggcgccgcccgtgag	249
-----------------------------	---	-----

	cgaggccgaggccgacttctacgagcagaacagccggtatcagctgaccctgtggggacctgagggcaa catcctggactacgccaacaagcagctggccggcctgggtggccaactactacacccaaggtggcggtg ttcctggaggccctgggtgacagcgtggccagggcatcccttccagcagcaccagttcgacaagaacg tgttccagctggagcaggccttcgtgctgagcaagcagcggatcccagccagcctagaggagacaccgt ggacctggccaagaagatcttctgaagtactacccccgggtgggtggccggcagctggggaCTTGA GGTACTGTTCCAAGGGCCCGAGGACCAGGTAGACCCACGACTCA TTGATGGAAAATAG	
vIGF2-32- NAGLU- HPC4	ATGGAGGCTGTGGCTGTGGCAGCTGCGGTGGGGGTCTTCTCCT GGCCGGGGCCGGGGGCGCGGCAGGCGACGcTagcagaacactttgtggcgga gagctgggtgacaccctgcagtttgtgtgtggcgacagaggcttctgttcagcagaGGTGGAGGT GGAtctagaggaatcCTGgaagagtgtgttcagagaGAAtgcgatctggccctgctggaaacctac tgtgccacaccagccagatctgaaGGTGGAGGTGGTTCAGGTGGTGGAGGTTTC CAGGCCAAGGGCGGTCCCTACTCAGGCCgaggcccgaggccgcccgcctg cgggccctgggtggcccggtgtgtggccccggccccgcccgcacttcagcgttagcgtggagcgggc cctggccccaagcccggtgacacctacagcctgggcggcggcgccgcccgggtgcgggtg cggggcagcaccggcgtggccgcccggcggcctgcaccggtatctgcgggacttctgcggctgccac gtggcctggagcggcagccagctgcggctgccccggccctgcccggcgtggccggcagctgaccga ggccacccccaaaccggtatcggtactaccagaacgtgtgcaccagagctacagcttcgtgtgtgggact ggggccgggtgggagcgggagatcgactggatggcctgaacggcatcaacctggccctggcctggagc ggccaggaggccatctggcagcgggtgtacctggccctgggcctgaccagggcgagatcaacaggttc ttcaccggccccgccttctggcctggggccgatgggcaacctgcacacctgggacggccccctgcctc caagctggcacatcaagcagctgtacctgcagcaccgggtgtggaccagatgcggagcttcggcatgac ccccgtgtgcccgccttcgcccggcacgtgcccaggccgtgaccgggtgttccccaaagtaacgtg accaagatgggcagctggggccacttcaactgcagctacagctgcagcttctgtgcccccgaggacc ccatcttccccatcatcgccagcctgttctgcgggagctgatcaaggagtgcggcaccgaccacatctacg gcgccgacaccttaacgagatgcagcctccaagcagcagcccagctacctggccgcccgcaccaccg ccgtgtacgaggccatgaccggcgtggacaccgaggccgtgtggctgtgcagggtgggtgttccagca ccagccccagttctggggacctgcccagatccggccgtgtgtggcggcgtgcctagaggacgggtgt gggtgtggacctgttcggcgagagccagcccgtgtacaccggaccgccagcttccaggccagcccttc atctgtgtcatgtgcacaacttcggcggaaccacggcctgttcggcgccctggaggccgtgaacggcg gccccgaggccgcccggctgttccccaaacgaccatggtgggcaccggcatggccccgagggcac	250

	agccagaacgaggtggtgtacagcctgatggccgagctgggctggcgggaaggaccccggtccccgacct ggccgcctgggtgaccagcttcgcccccggcggtacggcgtgagccaccccgacggcgccgcct ggcggtgctgctcggagcgtgtacaactgcagcggcgaggcctgccggggccacaaccggagcccc ctggtgcggcgccccagcctgcagatgaacaccagcatctggtacaaccggagcgacgtgttcgaggcct ggcggtgctgctgaccagcgccccagcctggccaccagccccgccttcagatacgacctgctggacct gacccggcaggccgtgcaggagctggtgagcctgtactacgaggaggcccgagcgcctacctgagca aggagctggccagcctgctcggggccggcgcggtgctggcctacgagctgctccccccctggacgag gtgctggccagcgacagccgggtcctgctgggcagctggctggagcaggccggcgccggcgctgag cgaggccgaggccgacttacgagcagaacagccggatcagctgacctgtggggacctgagggcaa catcctggactacgccaacaagcagctggccggcctggtggccaactactacacccaagggtggcggtg ttcctggaggccctggtggacagcgtggccagggcatccctccagcagcaccagttcgacaagaacg tgttcagctggagcaggccttcgtgctgagcaagcagcggtatccagccagcctagaggagacaccgt ggacctggccaagaagatcttcctgaagtactacccccggtgggtggccggcagctggggaCTTGA GGTACTGTTCCAAGGGCCCGAGGACCAGGTAGACCCACGACTCA TTGATGGAAAATAG	
--	---	--

[00159] In some embodiments, the vector comprising the nucleic acid encoding the desired therapeutic fusion protein, such as a vIGF2 fusion or a signal peptide fusion, optionally having an internal ribosomal entry sequence, provided herein is an adeno-associated viral vector (A5/35).

[00160] In some embodiments, the nucleic acid encoding the therapeutic fusion protein, such as a vIGF2 fusion, optionally has an internal ribosomal entry sequence and can be cloned into various types of vectors. For example, in some embodiments, the nucleic acid is cloned into a vector including, but not limited to a plasmid, a phagemid, a phage derivative, an animal virus, and a cosmid. Vectors of interest include expression vectors, replication vectors, probe generation vectors, and sequencing vectors.

[00161] Further, the expression vector encoding the therapeutic fusion protein, such as a vIGF2 fusion or a signal peptide fusion, optionally having an internal ribosomal entry sequence, in some embodiments, is provided to a cell in the form of a viral vector. Viral vector technology is described, e.g., in Sambrook et al., 2012, Molecular Cloning: A Laboratory Manual, volumes 1-4, Cold Spring Harbor Press, NY), and in other virology and molecular

biology manuals. Viruses, which are useful as vectors include, but are not limited to, retroviruses, adenoviruses, adeno-associated viruses, herpes viruses, and lentiviruses. In general, a suitable vector contains an origin of replication functional in at least one organism, a promoter sequence, convenient restriction endonuclease sites, and one or more selectable markers (e.g., WO 01/96584; WO 01/29058; and U.S. Pat. No. 6,326,193).

[00162] Also provided herein are compositions and systems for gene transfer. A number of virally based systems have been developed for gene transfer into mammalian cells. For example, retroviruses provide a convenient platform for gene delivery systems. A selected gene, in some embodiments, is inserted into a vector and packaged in retroviral particles using suitable techniques. The recombinant virus is then isolated and delivered to cells of the subject either in vivo or ex vivo. A number of retroviral systems are suitable for gene therapy. In some embodiments, adenovirus vectors are used. A number of adenovirus vectors are suitable for gene therapy. In some embodiments, adeno-associated virus vectors are used. A number of adeno-associated viruses are suitable for gene therapy. In one embodiment, lentivirus vectors are used.

[00163] Gene therapy constructs provided herein comprise a vector (or gene therapy expression vector) into which the gene of interest is cloned or otherwise which includes the gene of interest in a manner such that the nucleotide sequences of the vector allow for the expression (constitutive or otherwise regulated in some manner) of the gene of interest. The vector constructs provided herein include any suitable gene expression vector that is capable of being delivered to a tissue of interest and which will provide for the expression of the gene of interest in the selected tissue of interest.

[00164] In some embodiments, the vector is an adeno-associated virus (AAV) vector because of the capacity of AAV vectors to cross the blood-brain barrier and transduction of neuronal tissue. In methods provided herein, AAV of any serotype is contemplated to be used. The serotype of the viral vector used in certain embodiments is selected from the group consisting of an AAV1 vector, an AAV2 vector, an AAV3 vector, an AAV4 vector, an AAV5 vector, an AAV6 vector, an AAV7 vector, an AAV8 vector, an AAV9 vector, an AAVrhS vector, an AAVrh10 vector, an AAVrh33 vector, an AAVrh34 vector, an AAVrh74 vector, an AAV Anc80 vector, an AAVPHP.B vector, an AAVhu68 vector, an AAV-DJ vector, and others suitable for gene therapy.

[00165] AAV vectors are DNA parvoviruses that are nonpathogenic for mammals. Briefly, AAV-based vectors have the rep and cap viral genes that account for 96% of the viral genome removed, leaving the two flanking 145 base pair inverted terminal repeats (ITRs) which are used to initiate viral DNA replication, packaging, and integration.

5 **[00166]** Further embodiments include use of other serotype capsids to create an AAV1 vector, an AAV2 vector, an AAV3 vector, an AAV4 vector, an AAV5 vector, an AAV6 vector, an AAV7 vector, an AAV8 vector, an AAV9 vector, an AAVrhS vector, an AAVrh10 vector, an AAVrh33 vector, an AAVrh34 vector, an AAVrh74 vector, an AAV Anc80 vector, an AAVPHP.B vector, an AAV-DJ vector, and others suitable for gene therapy. Optionally,
10 the AAV viral capsid is AAV2/9, AAV9, AAVrhS, AAVrh10, AAVAnc80, or AAV PHP.B.

[00167] Additional promoter elements, e.g., enhancers, regulate the frequency of transcriptional initiation. Typically, these are located in the region 30-110 bp upstream of the start site, although a number of promoters have been shown to contain functional elements downstream of the start site as well. The spacing between promoter elements frequently is
15 flexible, so that promoter function is preserved when elements are inverted or moved relative to one another. In the thymidine kinase (tk) promoter, the spacing between promoter elements is often increased to 50 bp apart before activity begins to decline. Depending on the promoter, it appears that individual elements function either cooperatively or independently to activate transcription.

20 **[00168]** An example of a promoter that is capable of expressing a therapeutic fusion protein, such as a vIGF2 fusion or a signal peptide fusion, optionally having an internal ribosomal entry sequence, transgene in a mammalian T-cell is the EF1a promoter. The native EF1a promoter drives expression of the alpha subunit of the elongation factor-1 complex, which is responsible for the enzymatic delivery of aminoacyl tRNAs to the ribosome. The
25 EF1a promoter has been extensively used in mammalian expression plasmids and has been shown to be effective in driving expression from transgenes cloned into a lentiviral vector (see, e.g., Milone et al., Mol. Ther. 17(8): 1453-1464 (2009)). Another example of a promoter is the immediate early cytomegalovirus (CMV) promoter sequence. This promoter sequence is a strong constitutive promoter sequence capable of driving high levels of expression of any
30 polynucleotide sequence operatively linked thereto. However, other constitutive promoter sequences are sometimes also used, including, but not limited to the chicken β actin promoter, the P546 promoter, the simian virus 40 (SV40) early promoter, mouse mammary tumor virus

(MMTV), human immunodeficiency virus (HIV) long terminal repeat (LTR) promoter, MoMuLV promoter, an avian leukemia virus promoter, an Epstein-Barr virus immediate early promoter, a Rous sarcoma virus promoter, as well as human gene promoters such as, but not limited to, the actin promoter, the myosin promoter, the elongation factor-1a promoter, the hemoglobin promoter, and the creatine kinase promoter. Further, gene therapy vectors are not contemplated to be limited to the use of constitutive promoters. Inducible promoters are also contemplated here. An inducible promoter provides a molecular switch capable of turning on expression of the polynucleotide sequence to which it is operatively linked when such expression is desired, and turning off expression when expression is not desired. Examples of inducible promoters include, but are not limited to a metallothionine promoter, a glucocorticoid promoter, a progesterone promoter, and a tetracycline-regulated promoter.

[00169] In order to assess the expression of a therapeutic fusion protein, such as a vIGF fusion or a signal peptide fusion, optionally having an internal ribosomal entry sequence, or portions thereof, the expression vector to be introduced into a cell often contains either a selectable marker gene or a reporter gene or both to facilitate identification and selection of expressing cells from the population of cells sought to be transfected or infected through viral vectors. In other aspects, the selectable marker is often carried on a separate piece of DNA and used in a co-transfection procedure. Both selectable markers and reporter genes are sometimes flanked with appropriate regulatory sequences to enable expression in the host cells. Useful selectable markers include, for example, antibiotic-resistance genes, such as neo and the like.

[00170] Methods and compositions for introducing and expressing genes into a cell are suitable for methods herein. In the context of an expression vector, the vector is readily introduced into a host cell, e.g., mammalian, bacterial, yeast, or insect cell by any method in the art. For example, the expression vector is transferred into a host cell by physical, chemical, or biological means.

[00171] Physical methods and compositions for introducing a polynucleotide into a host cell include calcium phosphate precipitation, lipofection, particle bombardment, microinjection, gene gun, electroporation, and the like. Methods for producing cells comprising vectors and/or exogenous nucleic acids are suitable for methods herein (see, e.g., Sambrook et al., 2012, *Molecular Cloning: A Laboratory Manual*, volumes 1-4, Cold Spring Harbor Press, NY). One method for the introduction of a polynucleotide into a host cell is calcium phosphate transfection.

[00172] Chemical means and compositions for introducing a polynucleotide into a host cell include colloidal dispersion systems, such as macromolecule complexes, nanocapsules, microspheres, beads, and lipid-based systems including oil-in-water emulsions, micelles, mixed micelles, nucleic acid-lipid particles, and liposomes. An exemplary colloidal system for use as a delivery vehicle in vitro and in vivo is a liposome (e.g., an artificial membrane vesicle). Other methods of state-of-the-art targeted delivery of nucleic acids are available, such as delivery of polynucleotides with targeted nanoparticles or other suitable sub-micron sized delivery system.

[00173] In the case where a non-viral delivery system is utilized, an exemplary delivery vehicle is a liposome. The use of lipid formulations is contemplated for the introduction of the nucleic acids into a host cell (in vitro, ex vivo or in vivo). In another aspect, the nucleic acid is associated with a lipid. The nucleic acid associated with a lipid, in some embodiments, is encapsulated in the aqueous interior of a liposome, interspersed within the lipid bilayer of a liposome, attached to a liposome via a linking molecule that is associated with both the liposome and the oligonucleotide, entrapped in a liposome, complexed with a liposome, dispersed in a solution containing a lipid, mixed with a lipid, combined with a lipid, contained as a suspension in a lipid, contained or complexed with a micelle, or otherwise associated with a lipid. Lipid, lipid/DNA or lipid/expression vector associated compositions are not limited to any particular structure in solution. For example, in some embodiments, they are present in a bilayer structure, as micelles, or with a "collapsed" structure. Alternately, they are simply interspersed in a solution, possibly forming aggregates that are not uniform in size or shape. Lipids are fatty substances which are, in some embodiments, naturally occurring or synthetic lipids. For example, lipids include the fatty droplets that naturally occur in the cytoplasm as well as the class of compounds which contain long-chain aliphatic hydrocarbons and their derivatives, such as fatty acids, alcohols, amines, amino alcohols, and aldehydes.

[00174] Lipids suitable for use are obtained from commercial sources. For example, in some embodiments, dimyristyl phosphatidylcholine ("DMPC") is obtained from Sigma, St. Louis, Mo.; in some embodiments, dicetyl phosphate ("DCP") is obtained from K & K Laboratories (Plainview, N.Y.); cholesterol ("Choi"), in some embodiments, is obtained from Calbiochem-Behring; dimyristyl phosphatidylglycerol ("DMPG") and other lipids are often obtained from Avanti Polar Lipids, Inc. (Birmingham, Ala.). Stock solutions of lipids in chloroform or chloroform/methanol are often stored at about -20 °C. Chloroform is used as the

only solvent since it is more readily evaporated than methanol. "Liposome" is a generic term encompassing a variety of single and multilamellar lipid vehicles formed by the generation of enclosed lipid bilayers or aggregates. Liposomes are often characterized as having vesicular structures with a phospholipid bilayer membrane and an inner aqueous medium. Multilamellar liposomes have multiple lipid layers separated by aqueous medium. They form spontaneously when phospholipids are suspended in an excess of aqueous solution. The lipid components undergo self-rearrangement before the formation of closed structures and entrap water and dissolved solutes between the lipid bilayers (Ghosh et al., 1991 Glycobiology 5: 505-10). However, compositions that have different structures in solution than the normal vesicular structure are also encompassed. For example, the lipids, in some embodiments, assume a micellar structure or merely exist as non-uniform aggregates of lipid molecules. Also contemplated are lipofectamine-nucleic acid complexes.

[00175] Regardless of the method used to introduce exogenous nucleic acids into a host cell or otherwise expose a cell to the therapeutic fusion protein, such as a vIGF2 fusion or a signal peptide fusion, optionally having an internal ribosomal entry sequence, provided herein, in order to confirm the presence of the recombinant DNA sequence in the host cell, a variety of assays are contemplated to be performed. Such assays include, for example, "molecular biological" assays suitable for methods herein, such as Southern and Northern blotting, RT-PCR and PCR; "biochemical" assays, such as detecting the presence or absence of a particular peptide, e.g., by immunological means (ELISAs and western blots) or by assays described herein to identify agents falling within the scope herein.

[00176] The present disclosure further provides a vector comprising a therapeutic fusion protein, such as a vIGF2 fusion or a signal peptide fusion, optionally having an internal ribosomal entry sequence, encoding nucleic acid molecule. In one aspect, a therapeutic fusion protein vector is capable of being directly transduced into a cell. In one aspect, the vector is a cloning or expression vector, e.g., a vector including, but not limited to, one or more plasmids (e.g., expression plasmids, cloning vectors, minicircles, minivectors, double minute chromosomes), retroviral and lentiviral vector constructs. In one aspect, the vector can be used to express the vIGF2-therapeutic fusion protein construct in mammalian cells. In one aspect, the mammalian cell is a human cell.

Uses and Methods of Treatment

[00177] Also provided herein are methods of treating genetic disorders using gene therapy comprising administering to an individual a nucleic acid encoding a therapeutic fusion protein (such as a vIGF2 fusion or a signal peptide fusion or a signal peptide-vIGF2 fusion), optionally having an internal ribosomal entry sequence, disclosed herein. Genetic disorders
5 suitable for treatment using methods herein comprise disorders in an individual caused by one or more mutations in the genome causing lack of expression or expression of a dysfunctional protein by the mutant gene.

[00178] Further provided herein are pharmaceutical compositions comprising a gene therapy vector, such as a gene therapy vector comprising a nucleic acid encoding a therapeutic
10 fusion protein (such as a vIGF2 fusion or a signal peptide fusion or a signal peptide-vIGF2 fusion), optionally having an internal ribosomal entry sequence, disclosed herein and a pharmaceutically acceptable carrier or excipient for use in preparation of a medicament for treatment of a genetic disorder.

[00179] In some embodiments, genetic disorders suitable for treatment using methods
15 provided herein are lysosomal storage disorder. In some embodiments, lysosomal storage disorders are treated herein using gene therapy to deliver missing or defective enzymes to the patient. In some embodiments, methods herein deliver an enzyme fused to a vIGF2 or fused to a signal peptide to the patient in order to deliver the enzyme to the cell where it is needed. In some embodiments, the lysosomal storage disorder is selected from the group consisting of
20 aspartylglucosaminuria, Batten disease, cystinosis, Fabry disease, Gaucher disease type I, Gaucher disease type II, Gaucher disease type III, Pompe disease, Tay Sachs disease, Sandhoff disease, metachromatic leukodystrophy, mucopolipidosis type I, mucopolipidosis type II, mucopolipidosis type III, mucopolipidosis type IV, Hurler disease, Hunter disease, Sanfilippo
disease type A, Sanfilippo disease type B, Sanfilippo disease type C, Sanfilippo disease type D,
25 Morquio disease type A, Morquio disease type B, Maroteau-Lamy disease, Sly disease, Niemann-Pick disease type A, Niemann-Pick disease type B, Niemann-Pick disease type C1, Niemann-Pick disease type C2, Schindler disease type I, and Schindler disease type II. In some embodiments, the lysosomal storage disorder is selected from the group consisting of activator deficiency, GM2-gangliosidosis; GM2-gangliosidosis, AB variant; alpha-
30 mannosidosis (type 2, moderate form; type 3, neonatal, severe); beta-mannosidosis; aspartylglucosaminuria; lysosomal acid lipase deficiency; cystinosis (late-onset juvenile or adolescent nephropathic type; infantile nephropathic); Chanarin-Dorfman syndrome; neutral

lipid storage disease with myopathy; NLSDM; Danon disease; Fabry disease; Fabry disease type II, late-onset; Farber disease; Farber lipogranulomatosis; fucosidosis; galactosialidosis (combined neuraminidase & beta-galactosidase deficiency); Gaucher disease ; type II Gaucher disease ; type III Gaucher disease; type IIIC Gaucher disease; Gaucher disease, atypical, due to saposin C deficiency; GM1-gangliosidosis (late-infantile/juvenile GM1-gangliosidosis; adult/chronic GM1-gangliosidosis); Globoid cell leukodystrophy, Krabbe disease (Late infantile onset; Juvenile Onset; Adult Onset); Krabbe disease, atypical, due to saposin A deficiency; Metachromatic Leukodystrophy (juvenile; adult); partial cerebroside sulfate deficiency; pseudoarylsulfatase A deficiency; metachromatic leukodystrophy due to saposin B deficiency; Mucopolysaccharidoses disorders: MPS I, Hurler syndrome; MPS I, Hurler-Scheie syndrome; MPS I, Scheie syndrome ; MPS II, Hunter syndrome ; MPS II, Hunter syndrome ; Sanfilippo syndrome Type A / MPS IIIA ; Sanfilippo syndrome Type B / MPS IIIB ; Sanfilippo syndrome Type C / MPS IIIC ; Sanfilippo syndrome Type D / MPS IIID ; Morquio syndrome, type A / MPS IVA ; Morquio syndrome, type B / MPS IVB ; MPS IX

hyaluronidase deficiency ; MPS VI Maroteaux-Lamy syndrome ; MPS VII Sly syndrome; mucopolipidosis I, sialidosis type II; I-cell disease, Leroy disease, mucopolipidosis II; Pseudo-Hurler polydystrophy / mucopolipidosis type III; mucopolipidosis IIIC / ML III GAMMA; mucopolipidosis type IV; multiple sulfatase deficiency; Niemann-Pick disease (type B; type C1 / chronic neuronopathic form; type C2; type D / Nova Scotian type); Neuronal Ceroid

Lipofuscinoses: CLN6 disease - Atypical Late Infantile, Late-Onset variant, Early Juvenile; Batten-Spielmeyer-Vogt/Juvenile NCL/CLN3 disease; Finnish Variant Late Infantile CLN5; Jansky-Bielschowsky disease/Late infantile CLN2/TPP1 Disease; Kufs/Adult-onset NCL/CLN4 disease (type B); Northern Epilepsy/variant late infantile CLN8; Santavuori-Haltia/Infantile CLN1/PPT disease; Pompe disease (glycogen storage disease type II); late-onset Pompe disease; Pycnodysostosis; Sandhoff disease / GM2 gangliosidosis; Sandhoff disease / GM2 gangliosidosis; Sandhoff disease / GM2 Gangliosidosis; Schindler disease (type III / intermediate, variable); Kanzaki disease; Salla disease; infantile free sialic acid storage disease (ISSD); spinal muscular atrophy with progressive myoclonic epilepsy (SMAPME); Tay-Sachs disease / GM2 gangliosidosis; juvenile-onset Tay-Sachs disease; late-onset Tay-Sachs disease; Christianson syndrome; Lowe oculocerebrorenal syndrome; Charcot-Marie-Tooth type 4J, CMT4J; Yunis-Varon syndrome; bilateral temporooccipital polymicrogyria (BTOP); X-linked hypercalciuric nephrolithiasis, Dent-1; and Dent disease 2. In some

embodiments, the therapeutic protein is associated with a lysosomal storage disorder and the therapeutic protein is selected from the group consisting of GM2-activator protein ; α -mannosidase ; MAN2B1 ; lysosomal β -mannosidase ; glycosylasparaginase ; lysosomal acid lipase ; cystinosis ; CTNS ; PNPLA2 ; lysosome-associated membrane protein-2; α -galactosidase A; GLA; acid ceramidase; α -L-fucosidase; protective protein/cathepsin A; acid β -glucosidase; GBA ; PSAP; β -galactosidase-1; GLB1 ; galactosylceramide β -galactosidase ; GALC ; PSAP; arylsulfatase A; ARSA ; α -L-iduronidase; iduronate 2-sulfatase; heparan N-sulfatase; N- α -acetylglucosaminidase; heparan acetyl CoA: α -glucosaminide acetyltransferase; N-acetylglucosamine 6-sulfatase; galactosamine-6-sulfate sulfatase; β -galactosidase; hyaluronidase; arylsulfatase B; β -glucuronidase; neuraminidase; NEU1; gamma subunit of N-acetylglucosamine-1- phosphotransferase; mucolipin-1; sulfatase-modifying factor-1; acid sphingomyelinase; SMPD1; NPC1; and NPC2.

[00180] In some embodiments, treatment via methods herein delivers a gene encoding a therapeutic protein to a cell in need of the therapeutic protein. In some embodiments, the treatment delivers the gene to all somatic cells in the individual. In some embodiments, the treatment replaces the defective gene in the targeted cells. In some embodiments, cells treated ex vivo to express the therapeutic protein are delivered to the individual.

[00181] Gene therapy for disorders disclosed herein provides superior treatment outcomes to conventional treatments, including enzyme replacement therapy, because it does not require long infusion treatments.

Definitions

[00182] As used herein “ex vivo gene therapy” refers to methods where patient cells are genetically modified outside the subject, for example to express a therapeutic gene. Cells with the new genetic information are then returned to the subject from whom they were derived.

[00183] As used herein “in vivo gene therapy” refers to methods where a vector carrying the therapeutic gene(s) is directly administered to the subject.

[00184] As used herein “fusion protein” and “therapeutic fusion protein” are used interchangeably herein and refer to a therapeutic protein having at least one additional protein, peptide, or polypeptide, linked to it. In some instances, fusion proteins are a single protein molecule containing two or more proteins or fragments thereof, covalently linked via peptide bond within their respective peptide chains, without chemical linkers. In some embodiments, the fusion protein comprises a therapeutic protein and a signal peptide, a peptide that increases

endocytosis of the fusion protein, or both. In some embodiments, the peptide that increases endocytosis is a peptide that binds CI-MPR.

[00185] As used herein “vector”, or “gene therapy vector”, used interchangeably herein, refers to gene therapy delivery vehicles, or carriers, that deliver therapeutic genes to cells. A
5 gene therapy vector is any vector suitable for use in gene therapy, e.g., any vector suitable for the therapeutic delivery of nucleic acid polymers (encoding a polypeptide or a variant thereof) into target cells (e.g., sensory neurons) of a patient. In some embodiments, the gene therapy vector delivers the nucleic acid encoding a therapeutic protein or therapeutic fusion protein to a cell where the therapeutic protein or fusion is expressed and secreted from the cell. The vector
10 may be of any type, for example it may be a plasmid vector or a minicircle DNA. Typically, the vector is a viral vector. These include both genetically disabled viruses such as adenovirus and nonviral vectors such as liposomes. The viral vector may for example be derived from an adeno-associated virus (AAV), a retrovirus, a lentivirus, a herpes simplex virus, or an adenovirus. AAV derived vectors. The vector may comprise an AAV genome or a derivative
15 thereof.

[00186] “Construct” as used herein refers to a nucleic acid molecule or sequence that encodes a therapeutic protein or fusion protein and optionally comprises additional sequences such as a translation initiation sequence or IRES sequence.

[00187] As used herein “plasmid” refers to circular, double-stranded unit of DNA that
20 replicates within a cell independently of the chromosomal DNA.

[00188] As used herein “promoter” refers to a site on DNA to which the enzyme RNA polymerase binds and initiates the transcription of DNA into RNA.

[00189] As used herein “somatic therapy” refers to methods where the manipulation of gene expression in cells that will be corrective to the patient but not inherited by the next
25 generation. Somatic cells include all the non-reproductive cells in the human body

[00190] As used herein “somatic cells” refers to all body cells except the reproductive cells.

[00191] As used herein “tropism” refers to preference of a vector, such as a virus for a certain cell or tissue type. Various factors determine the ability of a vector to infect a particular
30 cell. Viruses, for example, must bind to specific cell surface receptors to enter a cell. Viruses are typically unable to infect a cell if it does not express the necessary receptors.

[00192] The term “transduction” is used to refer to the administration/delivery of the nucleic acid encoding the therapeutic protein to a target cell either *in vivo* or *in vitro*, via a replication-deficient rAAV of the disclosure resulting in expression of a functional polypeptide by the recipient cell. Transduction of cells with a gene therapy vector such as a rAAV of the disclosure results in sustained expression of polypeptide or RNA encoded by the rAAV. The present disclosure thus provides methods of administering/delivering to a subject a gene therapy vector such as an rAAV encoding a therapeutic protein by an intrathecal, intraretinal, intraocular, intravitreous, intracerebroventricular, intraparechymal, or intravenous route, or any combination thereof. “Intrathecal” delivery refers to delivery into the space under the arachnoid membrane of the brain or spinal cord. In some embodiments, intrathecal administration is via intracisternal administration. The present disclosure also provides methods of administering/delivering cells that have been transduced *ex vivo* with a gene therapy vector such as an rAAV vector encoding a therapeutic protein by an intrathecal, intraretinal, intraocular, intravitreous, intracerebroventricular, intraparechymal, or intravenous route, or any combination thereof.

[00193] The terms “recipient”, “individual”, “subject”, “host”, and “patient”, are used interchangeably herein and in some cases, refer to any mammalian subject for whom diagnosis, treatment, or therapy is desired, particularly humans. “Mammal” for purposes of treatment refers to any animal classified as a mammal, including humans, domestic and farm animals, and laboratory, zoo, sports, or pet animals, such as dogs, horses, cats, cows, sheep, goats, pigs, mice, rats, rabbits, guinea pigs, monkeys etc. In some embodiments, the mammal is human.

[00194] As used herein, the terms “treatment,” “treating,” “ameliorating a symptom,” and the like, in some cases, refer to administering an agent, or carrying out a procedure, for the purposes of obtaining a therapeutic effect, including inhibiting, attenuating, reducing, preventing or altering at least one aspect or marker of a disorder, in a statistically significant manner or in a clinically significant manner. The term “ameliorate” or “treat” does not state or imply a cure for the underlying condition. “Treatment,” or “to ameliorate” (and like) as used herein, may include treating a mammal, particularly in a human, and includes: (a) preventing the disorder or a symptom of a disorder from occurring in a subject which may be predisposed to the disorder but has not yet been diagnosed as having it (e.g., including disorders that may be associated with or caused by a primary disorder; (b) inhibiting the disorder, i.e., arresting its development; (c) relieving the disorder, i.e., causing regression of the disorder; and (d)

improving at least one symptom of the disorder. Treating may refer to any indicia of success in the treatment or amelioration or prevention of a disorder, including any objective or subjective parameter such as abatement; remission; diminishing of symptoms or making the disorder condition more tolerable to the patient; slowing in the rate of degeneration or decline; or making the final point of degeneration less debilitating. The treatment or amelioration of symptoms is based on one or more objective or subjective parameters; including the results of an examination by a physician. Accordingly, the term "treating" includes the administration of the compounds or agents of the present invention to prevent or delay, to alleviate, or to arrest or inhibit development of the symptoms or conditions associated with the disorder. The term "therapeutic effect" refers to the reduction, elimination, or prevention of the disorder, symptoms of the disorder, or side effects of the disorder in the subject.

[00195] The term "affinity" refers to the strength of binding between a molecule and its binding partner or receptor.

[00196] As used herein, the phrase "high affinity" refers to, for example, a therapeutic fusion containing such a peptide that binds CI-MPR which has an affinity to CI-MPR that is about 100 to 1,000 times or 500 to 1,000 times higher than that of the therapeutic protein without the peptide. In some embodiments, the affinity is at least 100, at least 500, or at least 1000 times higher than without the peptide. For example, where the therapeutic protein and CI-MPR are combined in relatively equal concentration, the peptide of high affinity will bind to the available CI-MPR so as to shift the equilibrium toward high concentration of the resulting complex.

[00197] "Secretion" as used herein refers to the release of a protein from a cell into, for example, the bloodstream to be carried to a tissue of interest or a site of action of the therapeutic protein. When a gene therapy product is secreted into the interstitial space of an organ, secretion can allow for cross-correction of neighboring cells.

[00198] "Delivery" as used herein means drug delivery. In some embodiments, the process of delivery means transporting a drug substance (e.g., therapeutic protein or fusion protein produced from a cell transduced with a gene therapy vector) from outside of a cell (e.g., blood, tissue, or interstitial space) into a target cell for therapeutic activity of the drug substance.

[00199] "Engineering" or "protein engineering" as used here in refers to the manipulation of the structures of a protein by providing appropriate a nucleic acid sequence

that encodes for the protein as to produce desired properties, or the synthesis of the protein with particular structures.

[00200] A "therapeutically effective amount" in some cases means the amount that, when administered to a subject for treating a disorder, is sufficient to effect treatment for that disorder.

[00201] As used herein, the term "about" a number refers to a range spanning that from 10% less than that number through 10% more than that number, and including values within the range such as the number itself.

[00202] As used herein, the term "comprising" an element or elements of a claim refers to those elements but does not preclude the inclusion of an additional element or elements.

EXAMPLES

[00203] The following examples are given for the purpose of illustrating various embodiments of the invention and are not meant to limit the present invention in any fashion. The present examples, along with the methods described herein are presently representative of preferred embodiments, are exemplary, and are not intended as limitations on the scope of the invention. Changes therein and other uses which are encompassed within the spirit of the invention as defined by the scope of the claims will occur to those skilled in the art.

Example 1: Binding of variant IGF2 peptide to CI-MPR Receptor

[00204] Surface plasmon resonance (SPR) experiments were conducted using Biacore to measure binding of wildtype and variant IGF2 (vIGF2) to the CI-MPR receptor. The wildtype, human mature IGF2 peptide (wt IGF2) has the sequence set forth in SEQ ID NO: 68. The vIGF2 sequence differs from wt IGF2 in that it lacks residues 1-4 and contains the following mutations: E6R, Y27L, and K65R. It has the amino acid sequence:

SRTLCCGGELVDTLQFVCGDRGFLFSRPASRVSRRSRGIVEECCFRSCDLALLETYCATPARSE (SEQ ID NO: 80). vIGF2 also has an N-terminal linker with the sequence

GGGGSGGGG (SEQ ID NO: 181). The combined sequence is

GGGGSGGGGSRTLCCGGELVDTLQFVCGDRGFLFSRPASRVSRRSRGIVEECCFRSCDLALLETYCATPARSE. **FIG. 4** shows that as expected, the wildtype IGF2 peptide binds to the CI-MPR receptor with high affinity (0.2 nM). **FIG. 5** shows that the variant IGF2 peptide

(vIGF2) also binds to the CI-MPR receptor with high affinity (0.5 nM). These data indicate that vIGF2 peptide has high affinity for the intended CI-MPR receptor for targeting therapeutics to lysosomes.

5 **[00205]** SPR was utilized to measure peptide binding to the Insulin Receptor to assess potential side effects. Insulin binds the Insulin Receptor with high affinity (~8 nM; data not shown). Wildtype IGF2 and a vIGF2 were tested, where the vIGF2 had the sequence SRTLGGELVDLTLQFVCGDRGFLFSRPASRVSRRSRGIVEECCFRSCDLALLETYCATP
10 ID NO: 181). **FIG. 8** shows that wildtype IGF2 also binds the Insulin Receptor with relatively high affinity (~100 nM). IGF2 peptide from Biomarin/Zystor IGF2-GAA fusion protein (BMN-701) also binds the Insulin Receptor with high affinity and was shown to cause hypoglycemia in clinical trials. **FIG. 9** shows no measurable binding of vIGF2 peptide to the insulin receptor. These data show that vIGF2 peptide confers a superior safety profile
15 compared with wt IGF2 peptide fusions.

[00206] The same SPR binding analysis was utilized to characterize vIGF2 peptide interaction with the IGF1 Receptor. **FIG. 10** shows that the wildtype IGF2 peptide binds IGF1 receptor with relatively high affinity (~100 nM). **FIG. 11** shows no measurable binding of
20 vIGF2 peptide to the IGF1 Receptor, showing an improved safety profile compared to wt IGF2.

Table 8: SPR Affinity Results		
Receptor	wt IGF2 Kd (nM)	vIGF2 Kd (nM)
CI-MPR	0.2	0.5
Insulin Receptor	100	No Binding Detected
IGF1 Receptor	100	No Binding Detected

Example 2: vIGF2 Converts Low Affinity Ligand to High Affinity ERT for CI-MPR

25 **[00207]** The vIGF2 peptide (SEQ ID NO: 80) with an N-terminal linker (SEQ ID NO: 181) was chemically coupled to alglucosidase-alfa, designated here as vIGF2- alglucosidase-alfa, to determine whether the vIGF2 peptide could improve affinity for CI-MPR. As shown in

FIG. 6, binding affinities of alglucosidase-alfa and vIGF2- alglucosidase-alfa were directly compared using CI-MPR plate binding assays in 96-well plates coated with CI-MPR.

Unbound enzyme was washed away prior to measuring bound enzyme activity. Varying concentrations of both enzyme preparations were used with or without free WT IGF2 peptide.

- 5 vIGF2 substantially improved the affinity for CI-MPR. Further, binding of vIGF2-
alglucosidase-alfa was blocked by free WT IGF2 indicating that binding was IGF2-dependent.
(Data not shown.) Coupling of vIGF2 peptide did not impair GAA enzyme activity.

- [00208]** The vIGF2 was coupled to recombinant human N-acetyl- α -D-glucosaminidase
10 (rhNAGLU). RhNAGLU, a lysosomal enzyme lacking M6P, to determine whether peptide
can convert a non-ligand to high affinity ligand for CI-MPR. In this experiment, rhNAGLU
and vIGF2-rhNAGLU were directly compared using CI-MPR plate binding assays, utilizing
CI-MPR-coated plates. Unbound enzyme was washed away prior to measuring bound enzyme
activity. Varying concentrations of both enzyme preparations were used with or without free
15 vIGF2 peptide. As shown in **FIG. 7**, vIGF2-rhNAGLU has significantly higher affinity for CI-
MPR than rhNAGLU lacking vIGF2. Further, vIGF2-rhNAGLU binding was blocked by free
vIGF2 peptide indicating that receptor binding was specific for IGF2 peptide. These results
show that vIGF2 peptide can be utilized to improve drug targeting to lysosomes.

20 Example 3: Myoblast Uptake of vIGF2-GAA Fusion Proteins

[00209] vIGF2-GAA fusion proteins (same sequences as in Examples 1-2) were
administered and L6 myoblast uptake of the enzyme was measured. **FIG. 6** shows superior
uptake of the vIGF2-rhGAA compared to rhGAA and M6P-GAA. Therefore, vIGF2 is
effective at targeting GAA to the cells.

25

Example 4: Constructs for ERT Delivered by Gene Therapy

- [00210]** Two different constructs are illustrated in **FIG. 12**. In the top panel is a
construct which contains a Kozak sequence and a nucleic acid encoding a recombinant human
GAA with the native signal peptide encoding “natural hGAA” (SEQ ID NO: 189). In the
30 middle panel is the construct Kozak-BiP-vIGF2-2GS-GAA, encoding “engineered hGAA”
(SEQ ID NO: 190). This construct is characterized by a Kozak sequence, a nucleic acid
encoding BiP signal peptide, a nucleic acid encoding the vIGF2 peptide having the sequence

set forth in SEQ ID NO: 80, and a nucleic acid encoding a 2GS linker (SEQ ID NO:181) followed by a nucleic acid encoding a recombinant human GAA (SEQ ID NO:1) with the N-terminal 60 amino acids removed to prevent premature processing and removal of the vIGF2. The amino acid sequence of “engineered hGAA” is set forth in SEQ ID NO:2.

5

Example 5: Enhanced Secretion of Gene Therapy Constructs

[00211] *Engineered hGAA has greater secretion and is able to interact with a cell surface receptor appropriate for cellular uptake and lysosomal targeting*

CHO expressing engineered hGAA, described in more detail below, or natural hGAA were cultured and conditioned media was collected for measurement of GAA activity. FIG. 15 shows the relative activity of engineered and natural hGAA showing that engineered hGAA has increased activity compared to natural hGAA, indicative of more efficient secretion of engineered hGAA.

Example 6: Analysis of PPT1 in conditioned media

15 **[00212]** *Cloning of PPT1 constructs*

PPT1 constructs were cloned into the pcDNA3.1 expression vector (ThermoFisher cat# V79020), which contains a CMV promoter. The tested constructs included PPT1-1 (WT-PPT1) (SEQ ID NO: 4) ; PPT1-2 (WT-vIGF2-PPT1) (SEQ ID NO: 5); PPT1-29 (BiP2aa-vIGF2-PPT1) (SEQ ID NO: 6).

20 *PPT1 secretion & binding*

[00213] The PPT1 constructs were transiently expressed in HEK293T cells for 3 days and the PPT1 secreted into the media. Secreted PPT1 was quantified by Western Blotting, and assayed for CI-MPR binding using established methods. Secreted PPT1 is shown in FIG. 13. CI-MPR binding is shown in FIG. 14.

25

[00214] Example 7: Testing Gene Therapy Vectors in an Animal Model of Pompe Disease

[00215] *Pompe Gene Therapy: Preclinical Proof of Concept Study Design*

30 **[00216]** A preclinical study was conducted in GAA knockout (GAA KO) mice using a high dose for initial comparison of constructs. The constructs are shown in FIG. 12. Mice were treated with vehicle or one of two constructs, Natural – hGAA or Engineered – hGAA.

Mice were administered 5×10^{11} gc/mouse (approximately 2.5×10^{13} gc/kg). GAA knockout mice were used at age 2 months. Normal (wildtype) mice were used as a control. The study design is outlined in FIG. 16.

[00217] *Pompe Gene Therapy: Plasma*

- 5 **[00218]** Plasma was collected from wild type (normal) mice or GAA KO mice treated with vehicle or a gene therapy vector as indicated and GAA activity and cell surface binding was measured. Data are summarized in FIG. 17, FIG. 27, and FIG. 19. Similar high GAA levels were seen in mice treated with gene therapy vectors (FIG. 17, FIG. 18). However, greater cell targeting receptor binding was observed with the engineered construct (FIG. 19).

10 **[00219]** *Pompe Gene Therapy: Quadriceps*

- [00220]** GAA activity, and glycogen storage/cytoplasmic vacuolization were assessed in normal (wild type) mice and treated GAA KO mice (FIG. 28). GAA activity in the quadriceps was about 20-fold higher than wild type. Glycogen PAS (FIG. 29) and immunohistochemistry (FIG. 30) were also assessed. Immunohistochemistry showed greater lysosomal targeting of engineered hGAA compared to wild type. Glycogen reduction was more consistent for engineered hGAA by PAS staining.

[00221] *Pompe Gene Therapy: Triceps*

- [00222]** GAA activity, and glycogen storage/cytoplasmic vacuolization were assessed in normal (wild type) mice and in treated GAA KO mice (FIG. 31). GAA activity was about 10-15-fold higher than wild type. Immunohistochemistry and glycogen PAS were also assessed (FIG. 32 and FIG. 33). Immunohistochemistry illustrated greater lysosomal targeting of engineered hGAA compared to wildtype GAA. Glycogen reduction was more consistent for engineered hGAA as measured by PAS staining.

[00223] *Pompe Gene Therapy: Tibialis Anterior (TA)*

- 25 **[00224]** GAA activity, and glycogen storage/cytoplasmic vacuolization were assessed in normal (wild type) and treated GAA KO mice (FIG. 20). GAA activity in the TA was about 15-20-fold higher than wild type. Immunohistochemistry and glycogen PAS were also assessed (FIG. 21 and FIG. 22). Immunohistochemistry illustrated greater lysosomal targeting of engineered hGAA compared to wildtype GAA. Glycogen levels were close to wildtype levels. Glycogen reduction was more consistent for engineered hGAA by PAS staining.

[00225] *Pompe Gene Therapy: Brain and Spinal Cord*

[00226] GAA activity, glycogen content, and glycogen storage/cytoplasmic vacuolization were assessed in normal (wild type) mice and treated GAA KO mice (FIG. 23). GAA activity in the brain was about 5-fold lower than wildtype. Immunohistochemistry and glycogen PAS were also assessed (FIG. 24, FIG. 25, FIG. 26, FIG. 27).

- 5 Immunohistochemistry indicated that there may be a direct transduction of some cells. However, little to no glycogen clearance was obtained with the natural construct. Glycogen levels were close to wild type levels for the engineered construct even though activity was only 20% of wild type. PAS staining in the spinal cord shows little to no glycogen clearance with the natural construct. Glycogen levels close to wild type for engineered construct was
- 10 observed in the ventral horn including motor neurons. Immunohistochemistry demonstrated direct transduction in spinal cord neurons. Engineered hGAA produced by the choroid plexus and neuronal cells was able to reduce glycogen by cross correction in the spinal cord while little glycogen reduction was observed for natural hGAA.

[00227] *Conclusions*

- 15 **[00228]** Overall, the data in this example demonstrated that the engineered gene therapy constructs have dramatically better uptake into tissues and glycogen reduction than the wildtype GAA used in conventional treatments, including effects in the brain and spinal cord.

EXAMPLE 8: Animal Study Protocols

- 20 **[00229]** AAVhu68 vectors were produced and titrated by the Penn Vector Core as described. (Lock, Alvira et al. 2010, "Rapid, simple, and versatile manufacturing of recombinant adeno-associated viral vectors at scale." Hum Gene Ther 21(10): 1259-1271).
- [00230]** Mus musculus, Pompe mice Gaa knock-out, in a C57BL/6/129 background founders were purchased at Jackson Labs (stock#004154, also known as 6neo mice).
- 25 **[00231]** Mice received 5×10^{11} GCs (approximately 2.5×10^{13} GC/kg) of AAVhu68.CAG.hGAA (comprising either natural hGAA (SEQ ID NO: 189) or engineered hGAA (SEQ ID NO: 190) in 0.1 mL via the lateral tail vein, were bled on Day 7 and Day 21 post vector dosing for serum isolation, and were terminally bled (for plasma isolation) and euthanized by exsanguination 28 days post injection. Tissues were promptly collected, starting
- 30 with brain.
- GAA Activity*

[00232] Plasma was mixed with 5.6 mM 4-MU- α -glucopyranoside pH 4.0 and incubated for three hours at 37 °C. The reaction was stopped with 0.4 M sodium carbonate, pH 11.5. Relative fluorescence units, RFUs were measured using a Victor3 fluorimeter, ex 355 nm and emission at 460 nm. Activity in units of nmol/mL/hr was calculated by interpolation from a standard curve of 4-MU. Activity in individual tissue samples were further normalized based on total protein content in the homogenate.

GAA Signature Peptide by LC/MS

[00233] Plasma was precipitated in 100% methanol and centrifuged. Supernatants were discarded. The pellet was spiked with a stable isotope-labeled peptide unique to hGAA as an internal standard and resuspended with trypsin and incubated at 37 °C for one hour. The digestion was stopped with 10% formic acid. Tryptic peptides were separated by C-18 reverse phase chromatography and Identified and quantified by ESI-mass spectroscopy. The total GAA concentration in plasma was calculated from the signature peptide concentration.

Cell surface Receptor Binding assay

[00234] A 96-well plate was coated with receptor, washed, and blocked with BSA. 28-day plasma from AAV treated mice was serially diluted to give a series of decreasing concentrations and incubated with coupled receptor. After incubation the plate was washed to remove any unbound hGAA and 4-MU- α -glucopyranoside added for one hour at 37 °C. The reaction was stopped with 1.0 M glycine, pH 10.5 and RFUs were read by a Spectramax fluorimeter; ex 370, emission 460. RFU's for each sample were converted to activity (nmol/mL/hr) by interpolation from a standard curve of 4-MU. Nonlinear regression was done using GraphPad Prism.

Histology

[00235] Tissues were formalin fixed and paraffin embedded. Muscle slides were stained with PAS; CNS slides with luxol fast blue/Periodic Acid-Schiff (PAS). A board-certified veterinary pathologist (JH) blindly reviewed histological slides. A semi-quantitative estimation of the total percentage of cells with glycogen storage and cytoplasmic vacuolization was done on scanned slides. A score from 0 to 4 was attributed as described in table below.

Table 9: Histology Scoring	
	Storage/Vacuolization
0	0
1	1 to 9%

2	10 to 49%
3	50 to 74%
4	75 to 100%

Immuno-histochemistry (IHC)

We studied transgene expression and cellular localization from slides immunostained using an anti-human GAA antibody (Sigma HPA029126).

- 5 **[00236]** EXAMPLE 9: Histology-Tissue Processing- Protocols and Results in an Animal Model of Pompe Disease

[00237] All tissues were fixed in 10% NBF (neutral buffered formalin). The assays (PAS and IHC) are routinely used in the field.

- 10 **[00238]** PAS staining of quadriceps and triceps (FIG. 29 and FIG. 32) - Tissues were fixed in 10% NBF and embedded in paraffin. Sections were post-fixed in 1% periodic acid and stained with Schiff's reagent. Afterwards, sections were counterstained with hematoxylin. Glycogen appears as magenta aggregates (lysosomal bound) or diffused pink (cytosolic); nuclei are blue. Based on the images and assuming each is representative of a group, the
- 15 ranking order in terms of glycogen clearance is: Engineered hGAA > Natural hGAA. The Engineered hGAA construct produced more staining across the entire image compared to the rest, showing an improved endocytosis of GAA protein mediated through the binding of vIGF2 to CI-MPR.

- [00239]** PAS staining of spinal cord (FIG. 26) - Tissues were fixed in 10% NBF. Post-fixation in 1% periodic acid could have been done prior to or after paraffin embedding. Sections were stained with Schiff's reagent and counterstained likely with methylene blue. Glycogen appears as magenta aggregates (lysosomal bound); nerve fibers appear blue. The images focused on the ventral horn of the spinal cord and glycogen accumulation in the motor neurons. Engineered hGAA appeared most effective in glycogen reduction among the
- 25 constructs.

[00240] GAA IHC (FIG. 22, FIG. 25, FIG. 27, FIG. 30, and FIG 35) - Tissues were fixed in 10% NBF and embedded in paraffin. Sections were incubated with an anti-GAA primary antibody, followed by a secondary antibody that recognizes the primary antibody and carries an enzyme tag --- HRP. Subsequently, an enzymatic reaction was carried out and a

brown-colored precipitating product was formed. Sections were then counterstained with hematoxylin. The constructs showed GAA uptake into muscle fibers (FIG. 31). Engineered hGAA > Natural hGAA . The BiP-vIGF2 construct had more diffused staining across the entire image compared to the rest.

5 **[00241]** Compared to other vectors, engineered hGAA produced more GAA IHC signals with a punctum-like appearance inside the muscle fibers, showing a much more efficient lysosomal targeting (FIG. 22).

[00242] In all, engineered hGAA consistently demonstrated superiority in tissue uptake, lysosomal targeting, and glycogen reduction in various tissues among the constructs.

10 **EXAMPLE 10: Binding of Fusion Proteins to CIMPR**

[00243] In this example, therapeutic enzymes were engineered to be targeted to the CI-MPR. Data in this example show that the fusion proteins bind better to CIMPR when they contain a vIGF2 tag. This was shown even for enzymes that are known to be well-phosphorylated, such as PPT1.

15 **[00244]** Each transgene was cloned into a pIRESHyg3 plasmid and the DNA was transfected in suspension HEK 293K cells using PEI transfection reagent. Cells were grown in FreeStyle 293 expression media. The conditioned media was harvested from the cells three to four days post-transfection. The amount of secreted enzyme in the conditioned media was determined by activity assay or by Signature Peptide assay. These concentrations were used to
20 set up CIMPR binding assays.

[00245] In the binding assay, a plate was first coated with CI-MPR. Next, a sample containing the enzyme of interest was incubated on the plate. The plate was washed so that only substances bound to CI-MPR remain on the plate. The amount of the enzyme of interest bound to the plate was determined by enzyme assay or by mass spec. The binding assay was
25 performed at a range of concentrations of the enzyme of interest in order to obtain a binding curve.

[00246] The amount of tagged and untagged enzyme bound to the plate was determined in order to construct binding curves. In the case of AGA and TPP1, enzyme activity assays were performed to make this determination. In other cases, the Signature Peptide assay was
30 performed to determine the amount of enzyme bound.

[00247] TPP1 activity assay is described at
www.rndsystems.com/products/recombinant-human-tripeptidyl-peptidase-i-tppl-protein-
cf_2237-se#product-details.

[00248] AGA activity assay described at YaV, et al. Applications of a new fluorometric
5 enzyme assay for the diagnosis of aspartylglucosaminuria. *J Inherit Metab Disease* 1993 and
Banning, et al. Identification of Small Molecule Compounds for Pharmacological Chaperone
Therapy of Aspartylglucosaminuria. *Sci Rep* 2016.

[00249] FIG. 34 shows increased binding of engineered PPT1 compared to wild type
PPT1. FIG. 35 shows increased binding of engineered TPP1 compared to wild type TPP1.
10 FIG. 36 shows increased binding of engineered AGA compared to wild type AGA. FIG. 37
shows increased binding of engineered GLA compared to wild type GLA.

EXAMPLE 11: Cloning of PPT1 Fusions

[00250] All PPT1 constructs were assembled into the pcDNA3.1 expression vector
15 using the In-Fusion cloning kit from Takara Bio.

[00251] The linearized pcDNA3.1 vector and each PPT1 gene fragments were
recombined via the InFusion reaction to yield the final pcDNA3.1 vector harboring the stated
PPT1 constructs.

EXAMPLE 12: Cloning vIGF2 Mutants

[00252] All of the vIGF2 mutants were swapped into the pcDNA3.1-BiP-vIGF2-2GS-
GAA expression vector using the In-Fusion cloning kit from Takara.

[00253] Recombination of the ordered vIGF2 fragment and the linearized pcDNA3.1-
GAA vector via the InFusion reaction gave the final pcDNA3.1-BiP-vIGF2*-2GS-GAA
25 circular expression vector.

EXAMPLE 13: Characterization of vIGF2-GAA Constructs

[00254] *Transient transfection of HEK293T cells with pcDNA3.1-vIGF2-GAA plasmids*

[00255] HEK293T cells were transiently transfected with 1 µg of DNA using Fugene
30 HD transfection reagent. The cultures were incubated for an additional 2-5 days at 37°C
supplemented with 5% CO₂ before harvesting the conditioned media and cell pellet.

[00256] *Western blot analysis of vIGF2-GAA in conditioned media*

[00257] Western blots were performed using a common standard method using the Licor Odyssey detection system. The primary antibody used for vIGF2-GAA detection was an in-house rabbit Anti-GAA antibody (FL059). The secondary antibodies used for GAA were goat anti-rabbit DyLight 800 (ThermoFisher cat# SA5-35571).

5 [00258] *GAA Activity Assay*

[00259] GAA activity was measured as described above.

[00260] *CI-MPR Binding Assay*

[00261] CI-MPR binding was measured as described above.

[00262] *Cellular Uptake Assay*

10 [00263] Results from the creation of 30+ IGF2-GAA constructs is as follows.

[00264] vIGF2-GAA constructs that exhibited secretion/expression level not less than 80% of the original vIGF2 are vIGF2-4, 5, 10, 11, 14, 16, 17, 31, and 32 (**FIG. 38** and **FIG. 39**).

[00265] vIGF2-GAA constructs that exhibited secretion/expression level not less than
15 50% of the original vIGF2 are vIGF2-4, 5, 6, 9-14, 16-23, 25, 27, and 29-34 (**FIG. 38** and **FIG. 39**).

[00266] All vIGF2-GAA constructs appeared to have processed correctly inside cells where the 70/76KDa mature GAA peptide fragment was observed (**FIG. 38**).

[00267] vIGF2-17 consistently gave a CI-MPR binding Bmax significantly higher than
20 the original vIGF2 (**FIG. 40, FIG. 41, FIG. 44, and FIG. 45**).

[00268] vIGF2-24 has binds CI-MPR significantly better than the original vIGF2 (**FIG. 42** and **FIG. 43**).

vIGF2-GAA constructs that have a comparable or better PM25 cellular uptake properties to the original vIGF2 include vIGF2-7, vIGF2-10, vIGF-17, vIGF2-18, vIGF2-20, vIGF2-22, &
25 vIGF2-23 (**FIG. 46** and **FIG. 47**).

EXAMPLE 14: Testing of PPT1 Constructs

[00269] vIGF2 peptides were designed as discussed elsewhere herein. Variants were selected based on increased selective binding to CI-MPR and improved protein expression.
30 Exemplary peptides and their structure are provided in **FIG. 48**.

[00270] *Transient transfection of HEK293T cells with pcDNA3.1-PPT1 plasmids*

[00271] HEK293T cells grown to about 80% confluence in 1 mL OptiMEM media supplemented with 5% FBS in 12-well culture were transiently transfected with 1 µg of DNA using Fugene HD transfection reagent. The cultures were incubated for an additional 2-5 days at 37°C supplemented with 5% CO₂ before harvesting the conditioned media and cell pellet.

5

[00272] *Western blot analysis of PPT1 in conditioned media*

[00273] Western blots were performed using a common standard method using the Licor Odyssey detection system. The primary antibody used for PPT1 detection was a mouse polyclonal antibody from Abcam (catalog cat# ab89022). The secondary antibodies used for PPT1 were goat anti-mouse DyLight 800 (ThermoFisher cat# SA5-35521).

10

[00274] Western blots of PPT1 expression and a graph showing band intensity are shown in **FIG. 49**. A graph showing PPT1 in conditioned media quantified by Western blot is shown in **FIG. 50**.

15 [00275] *PPT1 activity assay*

[00276] The PPT1 activity assay used was essentially that described by Van Diggelen et al. (Mol Genet Metab. 66:240-244, 1999). Briefly in a typical PPT1 activity assay, 10ul of conditioned media containing secreted PPT1 was mixed with 90ul of reaction buffer containing 75uM MU-6S-Palm-βGlc (4-methylumbelliferyl-6-thio-palmitate-β-D-glucopyranoside, Cayman Chemical; CAS 229644-17-1), 2 U/mL β-glucosidase (Sigma Chemicals; CAS 9001-22-3; G4511), 20 mM citrate pH 4.0, 5 mM DTT, 0.02% Triton X-100, and 50 mM NaCl in a 96-well black, clear bottom plate (Corning cat# 3631). Using an excitation wavelength of 330 nm and emission wavelength of 450 nm, fluorescence was monitored at 30-second intervals over a 1 hr period at 25°C using the SpectraMax M2. The rate of the PPT1 reaction was extracted by fitting the time course fluorescence data with a linear regression.

25

[00277] A graph showing PPT1 in conditioned media quantified by activity is shown in **FIG. 51**. Activity was found to have a strong correlation with the Western blot results. **FIG. 52** shows the correlation between activity and Western blot quantification.

30

[00278] *PPT1 stability assay*

[00279] Briefly, in a typical stability assay, 180 μ L of conditioned media containing PPT1 was diluted with 20 μ L of 10X PBS, pH 7.4 and incubated at 37°C. At different time points, an aliquot of 15 μ L was taken out and flashed frozen in ethanol cooled with dry ice. At the end of the time course experiment, the frozen samples were thawed and PPT1 activity was measured using the PPT1 activity assay.

[00280] *CI-MPR binding assay*

[00281] CI-MPR plate-binding assay performed as previously described, then amount bound was determined by PPT1 activity assay.

10 [00282] Binding of PPT1 constructs to CI-MPR in presence of M6P in the table below. Binding curves are shown in FIG. 53.

[00283] **Table 10:** Binding of PPT1 constructs to CI-MPR in presence of M6P

	Bmax	Relative Kd
PPT1-9		ND
PPT1-27		ND
PPT1-28		ND
PPT1-29	8.98	0.107
PPT1-30	4.88	0.056
PPT1-32	6.05	0.121
PPT1-33	9.19	0.143
PPT1-2		ND

15 [00284] Six PPT1 constructs were selected for further analysis. These six constructs are shown in FIG. 54. PPT1 secretion into the media (FIG. 55), PPT1 processing in-cell (FIG. 56), PPT1 quantification by Western blot (FIG. 57) and activity (FIG. 58) were determined for these six constructs.

20 **EXAMPLE 15 Engineering and Testing of Additional PPT1 IGF2 Fusion Constructs**

[00285] Additional PPT1 constructs were designed and cloned as shown in FIG. 62. These constructs contain either an endogenous signal sequence with a C6S mutation (SEQ ID NO:177), optionally with a two alanine extension to improve cleavage (SEQ ID NO:178), or a modified BiP signal peptide, BiP-2 (SEQ ID: 171), a PPT1 sequence comprising amino acid residues 21-306 or 28-306 of wild-type human PPT1 (SEQ ID NO: 4), a GS linker (SEQ ID

NO:181-187), and a variant IGF2-31 or 32 (SEQ ID NOs:120 or 121), separated by a lysosomal cleavage site, RPRAVPTQA (SEQ ID NO: 188).

All PPT1 constructs (Fig. 62) were transiently expressed in FreeStyle 293 suspension cells.

- 5 Briefly, FreeStyle 293 cells were transfected with each PPT1 construct in a pcDNA3.1 backbone, using polyethylenimine (PEI) as a transfection reagent. After four days of expression in FreeStyle 293 expression medium, the conditioned medium from each transfection was collected and run on western blots, using an anti-PPT1 primary antibody. Relative PPT1 levels in the medium were quantified from the band density on these western
- 10 blots. Fig. 63 shows that several constructs tested have higher levels secreted into the medium than WT PPT1. Higher PPT1 levels in the conditioned medium are reflective of both good expression and efficient secretion from the cell. Although vIGF2-31 (SEQ ID NO:120) and vIGF2-32 (SEQ ID NO:32) were designed to improve CIMPR binding, the surprisingly enhanced the expression and secretion of PPT1 compared to an earlier IGF2 variant (SEQ ID
- 15 NO:80).

Neuronal uptake experiments with purified protein constructs PPT1-101 and PPT1-104 showed successful uptake of both proteins, with approximately twice as much PPT1-104 taken up as PPT1-101 (Fig. 64 A). For this experiment, rat cortical neurons were cultured in NeuroCult

20 medium and plated on poly-L-lysine coated cover slips. The neurons were treated with 5 ug/ml purified PPT1-101 or PPT1-104, which had been labeled with Alexa Fluor 680 fluorescent dye. After a one-hour incubation, the cells were fixed, permeabilized, and imaged using a Leica SP8 confocal microscope.

- 25 Neuronal uptake experiments with conditioned medium were performed using conditioned medium obtained from FreeStyle 293 cell transfections, as described above. The concentration of each PPT1 construct protein in the media was first determined via western blot, using a standard curve generated using a sample of PPT1 of known concentration. Each sample of conditioned media was concentrated before treating the neurons. Rat cortical neurons were
- 30 cultured in Primary Neuron Growth Medium and plated on poly-L-lysine coated cover slips. The neurons were treated with the following concentrations of PPT1 protein in media:

WT	PPT1-101	PPT1-104	PPT1-112	PPT1-114	PPT1-117
----	----------	----------	----------	----------	----------

5.6 ug/ml	6.8 ug/ml	12.8 ug/ml	14.8 ug/ml	15.4 ug/ml	17.8 ug/ml
-----------	-----------	------------	------------	------------	------------

After a one-hour incubation, the cells were fixed, permeabilized, and imaged using a Leica SP8 confocal microscope. Uptake with all PPT1 variants was higher than with WT PPT1; PPT1-104 and PPT1-117 showed the highest levels of uptake (Fig. 64 B).

5

EXAMPLE 16 Analysis of NAGLU Constructs

[00286] Mutant fusion proteins comprising recombinant human NAGLU protein having an N-terminal vIGF2 tag inserted between the signal peptide and the NAGLU protein were designed as shown in **FIG. 65**. Several variants were prepared including fusion proteins comprising vIGF2 (SEQ ID NO:80), vIGF2-17 (SEQ ID NO:106), vIGF2-31 (SEQ ID NO:120) and vIGF2-32 (SEQ ID NO:121). The fusion proteins were expressed in HEK293F cells. The NAGLU content as determined by Western blotting with ab214671 (R&Dsystems) is shown in the lysate and media fractions for each fusion protein tested. (**FIG. 66 A-B**) Enzyme activity in conditioned media for each fusion protein was determined by a 4-MU assay. (Fig 66 C) Protein amounts in conditioned media were not normalized/equalized and activity data represent relative secretion of constructs into conditioned media rather than relative specific activity of equal quantities of proteins. As seen in Fig. 66, the presence of variant IGF2 led to decreased expression and secretion as compared to untagged NAGLU. However, the CIMPR binding of IGF2-tagged NAGLU improved significantly compared to untagged NAGLU. (**FIG. 67**) Notably, ~2.5-fold less IGF2-tagged NAGLU compared to WT was used as input for the binding assay, yet more of the tagged compared to WT bound to the immobilized receptor.

25

EXAMPLE 17 Analysis of TPP1 Constructs

[00287] A series of nucleic acid constructs for expressing TPP1 fusion proteins linked to IGF2 variants were designed and tested for expression, secretion and CIMPR binding. The fusion proteins comprise a signal peptide (SEQ ID NO:179, a variant IGF2 sequence (SEQ ID NOs:80, 106, 111, 133, 119-121), a GS linker (GGGGSGGGGS, SEQ ID NO:186), a

30

lysosomal cleavage site (RPRAVPTQA, SEQ ID NO:188), a TPP1 propeptide (SEQ ID NO:45), and a TPP1 mature peptide (SEQ ID NO:46). Both N-terminally and C-terminally vIGF2 tagged constructs were generated and tested. Examples of PPT1 fusion proteins that were designed and tested are shown in Table 11.

5 [00288] **Table 11. TPP1 Fusion Constructs**

PPT1 Construct	SEQ ID
pSvelte001- Native TPP1 Signal Peptide – vIGF2 – GS linker – Lyso Cleave – TPP1 propeptide – TPP1 mature peptide	47
pSvelte057 - Native TPP1 Signal Peptide – vIGF2v17 – GS linker – Lyso Cleave – TPP1 propeptide – TPP1 mature peptide	48
pSvelte059 - Native TPP1 Signal Peptide – vIGF2v22 – GS linker – Lyso Cleave – TPP1 propeptide – TPP1 mature peptide	49
pSvelte060 - Native TPP1 Signal Peptide – vIGF2v24 – GS linker – Lyso Cleave – TPP1 propeptide – TPP1 mature peptide	50
pSvelte061 - Native TPP1 Signal Peptide – vIGF2v30 – GS linker – Lyso Cleave – TPP1 propeptide – TPP1 mature peptide	51
pSvelte062 - Native TPP1 Signal Peptide – vIGF2v31 – GS linker – Lyso Cleave – TPP1 propeptide – TPP1 mature peptide	52
pSvelte063- Native TPP1 Signal Peptide – vIGF2v32 – GS linker – Lyso Cleave – TPP1 propeptide – TPP1 mature peptide	53

Expression & Secretion

[00289] For each construct, Freestyle 293 cells (3.7million cells in 1.5ml of Freestyle
 10 293 media) were transfected with 9ul of 1mg/ml PEI and 3ug DNA and grown in 24-well deep well plates under shaking conditions (37 deg C, 5% CO₂, 80%RH, 250 RPM). ~24 hrs following transfection, valproic acid (final concentration 2.2mM) and an additional 1.5ml freestyle media was added to the transfection. Cultures were harvested 3 days post transfection and centrifuged to separate cells and conditioned media. Protein in conditioned media was
 15 separated on an SDS-PAGE gel and transferred to a nitrocellulose membrane. The membrane was blocked with 5% milk and probed with anti-TPP1 (abcam EPR16537) and Licor Anti-

rabbit 800CW (926-32213). Blots were imaged and bands were quantified with a Licor Odyssey CLX as show in **FIG. 68**.

CIMPR Binding

- 5 **[00290]** CIMPR binding was measured essentially as described in Example 10. The results are shown in **FIG. 69**. rhTPP1 (R&D system #2237-SE-010, expressed in Mouse myeloma NS0 cells) and WT TPP1 (SEQ ID NO:8) were included as controls. As shown in FIG. 69, the novel TPP1 constructs all showed improved binding compared to rhTPP1.

10 **EXAMPLE 18** *Testing of Novel PPT1 Variants CLN1 Mouse Model*

- [00291]** The PPT1-101 (SEQ ID NO:60) and PPT1-104 (SEQ ID NO:61) constructs were tested in CLN1^{R151X} mouse model. (Miller, 2014, Human Molecular Genetics, 24(1)185–196). Gene therapy constructs comprising the coding sequences of PPT1-101(SEQ ID NO:228) and PPT1-104 (SEQ ID NO:235) were prepared. Postnatal Day 1 (P1) mice were
15 intracerebroventricularly injected with the viral constructs (or PBS control) at doses of 5x10¹⁰, 1x10¹⁰, or 1x10⁹ vg/animal. Wild-type PPT1 (p546) was included as a control. The transgenes were introduced using an AAV9 vector. Outcomes were assessed at 2 months of age.

20 *Transgene Expression*

- [00292]** Human CLN1 transgene expression was detected by RT-qPCR. As seen in **FIG. 70**, brain and spinal cord extracts showed similar gene expression between the various constructs, with higher expression in the cortex.

25 *Reduction in Autofluorescent Storage Material*

- [00293]** FIGs. 71-72 show the effect of each construct on brain autofluorescent storage material (ASM) accumulation, a correlate of lysosomal dysfunction. At the 5x10¹⁰ and 1x10¹⁰ doses in the cortex, and at the 1x10¹⁰ and 1x10⁹ doses in the thalamus, the 101 and 104 constructs trend towards greater reductions in ASM, as compared to the WT p546 construct.

30

Reduction in Glial Fibrillary Acidic Protein (GFAP)

[00294] FIG. 73 shows the effect of each construct on the glial fibrillary acidic protein (GFAP), a correlate of astrogliosis and neuroinflammation. At the 1x10⁹ dose in the cortex, the 104 construct trended towards greater reductions in GFAP. At the 1x10¹⁰ dose in the thalamus, the 101 construct trended towards greater reductions in GFAP. GFAP-positive cells
5 were morphologically consistent with a reactive astrocyte phenotype.

[00295] Thus, the novel PPT1 101 and 104 gene therapy constructs show improved cross-correction compared to wildtype PPT1 in a CLN1 mouse model, leading to greater reduction in both ASM and GFAP in the cortex and thalamus.

10

[00296] While preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. It should be understood
15 that various alternatives to the embodiments described herein may be employed. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.