



(51) International Patent Classification:

CI2N 15/86 (2006.01) C07K 14/475 (2006.01)

CI2N 15/113 (2010.01) A61K 48/00 (2006.01)

CI2N 15/65 (2006.01) A61P 27/16 (2006.01)

(21) International Application Number:

PCT/US2019/060324

(22) International Filing Date:

07 November 2019 (07.11.2019)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/756,910 07 November 2018 (07.11.2018) US

62/888,105 16 August 2019 (16.08.2019) US

(71) Applicant: AKOUOS, INC. [US/US]; 7 Stetson Street, Brookline, Massachusetts 02446 (US).

(72) Inventors: SIMONS, Emmanuel J.; 7 Stetson Street, Brookline, Massachusetts 02446 (US). NG, Robert; 47 Lincoln Rd, Unit 1, Newton, Massachusetts 02130 (US). LENZ, Danielle; 47 Lincoln Rd, Unit 1, Brookline, Massachusetts 02446 (US). VALERO, Michelle; 650 Ocean Ave., Unit #612, Revere, Massachusetts 02151 (US).

(74) Agent: LYON, Charles E. et al.; Choate, Hall & Stewart LLP, Two International Place, Boston, Massachusetts 02110 (US).

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

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(54) Title: COMPOSITIONS AND METHODS OF INDUCING DIFFERENTIATION OF A HAIR CELL

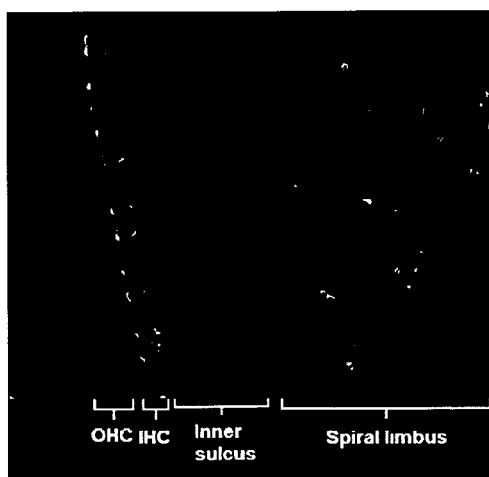


Figure 2B

(57) Abstract: Provided herein are compositions that include one or more adeno-associated virus (AAVs) vectors and methods of inducing differentiation of a hair cell using these vector(s).



Declarations under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
- *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

- *with international search report (Art. 21(3))*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
- *with sequence listing part of description (Rule 5.2(a))*

**COMPOSITIONS AND METHODS OF INDUCING DIFFERENTIATION
OF A HAIR CELL**

CROSS-REFERENCE TO RELATED APPLICATIONS

5 This application claims priority to U.S. Provisional Patent Application Serial No. 62/756,910, filed November 7, 2018 and U.S. Provisional Patent Application Serial No. 62/888,105, filed August 16, 2019; the entire contents of which are herein incorporated by reference.

TECHNICAL FIELD

10 The present disclosure relates to the fields of molecular biology, and more specifically, to the use of nucleic acids for treating hearing loss in a primate.

BACKGROUND OF THE INVENTION

15 Hearing loss can be conductive (arising from the ear canal or middle ear), sensorineural (arising from the inner ear or auditory nerve), or mixed. Most forms of non-syndromic deafness are associated with permanent hearing loss caused by damage to structures in the inner ear (sensorineural deafness), although some forms may involve changes in the middle ear (conductive hearing loss). The great majority of human sensorineural hearing loss is caused by abnormalities in the hair cells of the organ of Corti in the cochlea (poor hair cell function). The
20 hair cells may be abnormal at birth, or may be damaged during the lifetime of an individual (e.g., as a result of noise trauma or infection).

SUMMARY

25 The present invention is based on the discovery that administration of an AAV vector that includes a nucleic acid encoding a gene, to the inner ear of a primate, can result in the successful expression of a protein encoded by the gene in a supporting cell or hair cell in the inner ear of the primate. In view of this discovery, provided here are AAV vector(s) and methods of using these vectors to induce expression and/or activity of a hair cell differentiation protein in a supporting cell or hair cell in the inner ear of a primate or decreasing the expression and/or activity of a hair
30 cell differentiation suppressing gene in a supporting cell or hair cell in the inner ear of a primate.

Provided herein are compositions that include at least two different nucleic acid vectors, where: each of the at least two different adeno-associated virus (AAV) vectors includes a coding sequence that encodes a different portion of a hair cell differentiation protein, each of the encoded portions being at least 30 amino acid residues in length, where the amino acid sequence of each of the encoded portions may optionally partially overlap with the amino acid sequence of a different one of the encoded portions; no single vector of the at least two different vectors encodes the full-length hair cell differentiation protein; at least one of the coding sequences includes a nucleotide sequence spanning two neighboring exons of hair cell differentiation genomic DNA, and lacks an intronic sequence between the two neighboring exons; and when introduced into a primate cell the at least two different vectors undergo concatamerization or homologous recombination with each other, thereby forming a recombined nucleic acid that encodes a full-length hair cell differentiation protein that is expressed in the primate cell.

In some embodiments of any of the compositions described herein, the amino acid sequence of none of the encoded portions overlaps with the amino acid sequence of a different one of the encoded portions. In some embodiments of any of the compositions described herein, the amino acid sequence of each of the encoded portions partially overlaps with the amino acid sequence of a different one of the encoded portions. In some embodiments of any of the compositions described herein, the overlapping amino acid sequence is between 30 amino acid residues to about 390 amino acid residues in length.

In some embodiments of any of the compositions described herein, the vectors include two different vectors, each of which includes a different segment of an intron, where the intron includes the nucleotide sequence of an intron that is present in a hair cell differentiation genomic DNA, and where the two different segments overlap in sequence by at least 100 nucleotides. In some embodiments of any of the compositions described herein, the two different intron segments overlap in sequence by about 100 nucleotides to about 800 nucleotides.

In some embodiments of any of the compositions described herein, the entire nucleotide sequence of each of the at least two different vectors is between about 500 nucleotides to about 10,000 nucleotides in length. In some embodiments of any of the compositions described herein, the entire nucleotide sequence of each of the at least two different vectors is between about 500 nucleotides to about 5,000 nucleotides in length.

In some embodiments of any of the compositions described herein, the number of different vectors in the composition is two. In some embodiments of any of the compositions described herein, a first of the two different vectors includes a coding sequence that encodes an N-terminal portion of the hair cell differentiation protein. In some embodiments of any of the compositions described herein, the N-terminal portion of the hair cell differentiation protein is between about 30 amino acids to about 750 amino acids in length. In some embodiments of any of the compositions described herein, the N-terminal portion of the hair cell differentiation protein is between about 30 amino acids to about 320 amino acids in length.

In some embodiments of any of the compositions described herein, the first vector further includes one or both of a promoter and a Kozak sequence. In some embodiments of any of the compositions described herein, the first vector includes a promoter that is an inducible promoter, a constitutive promoter, or a tissue-specific promoter.

In some embodiments of any of the compositions described herein, the second of the two different vectors includes a coding sequence that encodes a C-terminal portion of the hair cell differentiation protein. In some embodiments of any of the compositions described herein, the C-terminal portion of the hair cell differentiation protein is between about 30 amino acids to about 750 amino acids in length. In some embodiments of any of the compositions described herein, the C-terminal portion of the hair cell differentiation portion is between about 30 amino acids to about 320 amino acids in length.

In some embodiments of any of the compositions described herein, the second vector further includes a poly(dA) sequence. In some embodiments of any of the compositions described herein, the second vector further includes a destabilizing sequence. In some embodiments of any of the compositions described herein, the second vector further includes a FKB12 destabilizing sequence.

Also provided herein are compositions that include two different nucleic acid vectors, where: a first nucleic acid vector of the two different nucleic acid vectors includes a promoter, a first coding sequence that encodes an N-terminal portion of a hair cell differentiation protein positioned 3' of the promoter, and a splicing donor signal sequence positioned at the 3' end of the first coding sequence; and a second nucleic acid vector of the two different nucleic acid vectors includes a splicing acceptor signal sequence, a second coding sequence that encodes a C-terminal portion of a hair cell differentiation protein positioned at the 3' end of the splicing

acceptor signal sequence, and a polyadenylation sequence at the 3' end of the second coding sequence; where each of the encoded portions is at least 30 amino acid residues in length, where the amino acid sequences of the encoded portions do not overlap, where no single vector of the two different vectors encodes the full-length hair cell differentiation protein, and, when the coding sequences are transcribed in a primate cell, to produce RNA transcripts, splicing occurs between the splicing donor signal sequence on one transcript and the splicing acceptor signal sequence on the other transcript, thereby forming a recombined RNA molecule that encodes a full-length hair cell differentiation protein.

In some embodiments of any of the compositions described herein, at least one of the coding sequences includes a nucleotide sequence spanning two neighboring exons of a hair cell differentiation genomic DNA, and lacks an intronic sequence between the two neighboring exons.

Also provided herein are compositions that include: a first nucleic acid vector including a promoter, a first coding sequence that encodes an N-terminal portion of a hair cell differentiation protein positioned 3' of the promoter, a splicing donor signal sequence positioned at the 3' end of the first coding sequence, and a first detectable marker gene positioned 3' of the splicing donor signal sequence; and a second nucleic acid vector, different from the first nucleic acid vector, including a second detectable marker gene, a splicing acceptor signal sequence positioned 3' of the second detectable marker gene, a second coding sequence that encodes a C-terminal portion of a hair cell differentiation protein positioned at the 3' end of the splicing acceptor signal sequence, and a polyadenylation sequence positioned at the 3' end of the second coding sequence; where each of the encoded portions is at least 30 amino acid residues in length, where the respective amino acid sequences of the encoded portions do not overlap with each other, where no single vector of the two different vectors encodes the full-length hair cell differentiation protein, and, when the coding sequences are transcribed in a primate cell to produce RNA transcripts, splicing occurs between the splicing donor signal on one transcript and the splicing acceptor signal on the other transcript, thereby forming a recombined RNA molecule that encodes a full-length hair cell differentiation protein.

In some embodiments of any of the compositions described herein, at least one of the coding sequences includes a nucleotide sequence spanning two neighboring exons of a hair cell differentiation genomic DNA, and lacks an intronic sequence between the neighboring exons. In

some embodiments of any of the compositions described herein, the first or second detectable marker gene is alkaline phosphatase. In some embodiments of any of the compositions described herein, the first and second detectable marker genes are the same.

Also provided herein are compositions that include: a first nucleic acid vector including a promoter, a first coding sequence that encodes an N-terminal portion of a hair cell differentiation protein positioned 3' to the promoter, a splicing donor signal sequence positioned at the 3' end of the first coding sequence, and a F1 phage recombinogenic region positioned 3' to the splicing donor signal sequence; and a second nucleic acid vector, different from the first nucleic acid vector, including a second F1 phage recombinogenic region, a splicing acceptor signal sequence positioned 3' of the second F1 phage recombinogenic region, a second coding sequence that encodes a C-terminal portion of a hair cell differentiation protein positioned at the 3' end of the splicing acceptor signal sequence, and a polyadenylation sequence positioned at the 3' end of the second coding sequence; where each of the encoded portions is at least 30 amino acid residues in length, where the respective amino acid sequences of the encoded portions do not overlap with each other, where no single vector of the two different vectors encodes the full-length hair cell differentiation protein, and, when the coding sequences are transcribed in a primate cell to produce RNA transcripts, splicing occurs between the splicing donor signal on one transcript and the splicing acceptor signal on the other transcript, thereby forming a recombined RNA molecule that encodes a full-length hair cell differentiation protein.

In some embodiments of any of the compositions described herein, at least one of the coding sequences includes a nucleotide sequence spanning two neighboring exons of a hair cell differentiation genomic DNA, and lacks an intronic sequence between the two neighboring exons.

Also provided herein are compositions that include a single adeno-associated virus (AAV) vector, where the single AAV vector includes a nucleic acid sequence that encodes a hair cell differentiation protein; and when introduced into a mammalian cell (e.g., primate cell (e.g., a hair cell or a supporting cell of the inner ear), a nucleic acid encoding the hair cell differentiation protein is generated at the locus of the hair cell differentiation gene and the primate cell expresses the hair cell differentiation protein.

In some embodiments of any of the compositions described herein, the hair cell differentiation gene is selected from the group of: atonal bHLH transcription factor 1 (ATOH1),

POU Class 4 Homeobox 3 (POU4F3), catenin beta 1 (CTNNB1), Noggin (NOG), growth factor independent 1 transcriptional repressor (GFI-1), neurotrophin 3 (NTF3), and brain-derived neurotrophic factor (BDNF).

Also provided herein are compositions including two different nucleic acid vectors, wherein a first nucleic acid vector includes a first nucleic acid sequence that encodes a first hair cell differentiation protein (e.g., any of the hair cell differentiation proteins described herein); and a second nucleic acid vector includes a second nucleic acid sequence that encodes a second hair cell differentiation protein (e.g., any of the hair cell differentiation proteins described herein), and when introduced into a primate cell, the first nucleic acid and the second nucleic acid encoding the first hair cell differentiation protein and the second hair cell differentiation protein are generated at the locus of the hair cell differentiation gene and the primate cell expresses the first hair cell differentiation protein and the second hair cell differentiation protein.

In some embodiments, the first and the second hair cell differentiation proteins are selected from the group consisting of: atonal bHLH transcription factor 1 (ATOH1), POU Class 4 Homeobox 3 (POU4F3), catenin beta 1 (CTNNB1), Noggin (NOG), growth factor independent 1 transcriptional repressor (GFI-1), neurotrophin 3 (NTF3) and brain-derived neurotrophic factor (BDNF).

In some embodiments of any of the compositions described herein, the second nucleic acid vector further includes a destabilizing sequence.

In some embodiments, the second nucleic acid vector further includes a FKB12 destabilizing sequence.

Provided herein are compositions that include at least one adeno-associated virus (AAV) vector that encodes an inhibitory nucleic acid that decreases the expression of a hair cell differentiation-suppressing protein in a primate cell.

In some embodiments of any of the compositions described herein, the inhibitory nucleic acid is a short interfering RNA (siRNA), a short hairpin RNA (shRNA), an antisense oligonucleotide, or a ribozyme.

In some embodiments of any of the compositions described herein, the hair cell differentiation-suppressing gene is HES1, HES5, sex determining region Y-box 2 (SOX2), and p27kip (CDKN1B). In some embodiments of any of the compositions described herein, the composition further includes a pharmaceutically acceptable excipient. Also provided herein are

kits including any of the compositions described herein. In some embodiments of any of the kits described herein, the kit further includes a pre-loaded syringe containing the composition.

Also provided herein are methods of promoting differentiation of a supporting cell of an inner ear of a primate into a hair cell that include: administering to the inner ear of the primate a therapeutically effective amount of any of the compositions described herein, where the administering promotes differentiation of the supporting cell of the inner ear of the primate into a hair cell.

Also provided herein are methods of increasing the expression level of a hair cell differentiation protein in a supporting cell or hair cell of an inner ear of a primate that include: administering to the inner ear of the primate a therapeutically effective amount of any of the compositions described herein, where the administering results in an increase in the expression level of the hair cell differentiation protein in the supporting cell or hair cell of the inner ear of the primate. In some embodiments of any of the methods described herein, the hair cell differentiation protein is selected from the group of: Atoh1, Pou4f3, β -Catenin, Noggin, GFI-1, NTF3, and BDNF. In some embodiments of the methods described herein, the primate has previously been determined to have a defective hair cell differentiation gene.

Also provided herein are methods of decreasing the expression level of a hair cell differentiation-suppressing protein in a supporting cell or hair cell of an inner ear of a primate that include: administering to the inner ear of the primate a therapeutically effective amount of any of the compositions described herein, where the administering results in a decrease in the expression level of the hair cell differentiation-suppressing protein in the supporting cell or hair cell of the inner ear of the primate.

Also provided herein are methods of increasing the number of functional hair cells in a primate in need thereof that include: administering to the inner ear of the primate a therapeutically effective amount of any of the compositions described herein.

Also provided herein are methods of improving hearing in a primate in need thereof that include: administering to the inner ear of the primate a therapeutically effective amount of any of the compositions described herein.

In some embodiments of any of the methods described herein, the method further includes prior to the administering step, determining that the primate has a defective hair cell differentiation gene.

Also provided herein are methods of repairing a hair cell toxicity-inducing mutation in an endogenous hair cell differentiation gene locus in a supporting cell or hair cell of an inner ear of a primate, that include: administering to the inner ear of the primate a therapeutically effective amount of any of the compositions described herein, where the administering results in repair of the hair cell toxicity-inducing mutation in the endogenous hair cell differentiation gene locus in the supporting cell or hair cell of the inner ear of the primate.

Also provided herein are methods of decreasing the risk of hearing loss due to hair cell loss or dysfunction in a primate in need thereof that include: administering to the inner ear of the primate a therapeutically effective amount of any of the compositions described herein.

In some embodiments of any of the methods described herein, the primate has been previously identified as having a defective hair cell differentiation gene.

The term “a” and “an” refers to one or to more than one (i.e., at least one) of the grammatical object of the article.

The term “conservative mutation” refers to a mutation that does not change the amino acid encoded at the site of the mutation (due to codon degeneracy).

Modifications can be introduced into a nucleotide sequence by standard techniques known in the art, such as site-directed mutagenesis and PCR-mediated mutagenesis.

Conservative amino acid substitutions are ones in which the amino acid residue in a protein is replaced with an amino acid residue having a chemically-similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, and histidine), acidic side chains (e.g., aspartic acid and glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine, and tryptophan), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, and methionine), beta-branched side chains (e.g., threonine, valine, and isoleucine), and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, and histidine).

Unless otherwise specified, a “nucleotide sequence encoding an amino acid sequence” includes all nucleotide sequences that are degenerate versions of each other and thus encode the same amino acid sequence.

The term “endogenous” refers to any material originating from within an organism, cell, or tissue.

The term “exogenous” refers to any material introduced from or originating from outside an organism, cell, or tissue that is not produced or does not originate from the same organism, cell, or tissue in which it is being introduced.

The term “isolated” means altered or removed from the natural state. For example, a nucleic acid or a peptide naturally present in a living animal is not “isolated,” but the same nucleic acid or peptide partially or completely separated from the coexisting materials of its natural state is “isolated.” An isolated nucleic acid or protein can exist in substantially purified form, or can exist in a non-native environment such as, for example, a host cell.

The term “transfected,” “transformed,” or “transduced” refers to a process by which exogenous nucleic acid is transferred or introduced into a cell. A “transfected,” “transformed,” or “transduced” primate cell is one that has been transfected, transformed, or transduced with exogenous nucleic acid.

The term “expression” refers to the transcription and/or translation of a particular nucleotide sequence encoding a protein.

The term “transient expression” refers to the expression of a non-integrated coding sequence for a short period of time (e.g., hours or days). The coding sequence that is transiently expressed in a cell (e.g., a primate cell) is lost upon multiple rounds of cell division.

The term “primate” is intended to include any primate (e.g., a human, a non-human primate (e.g., simian (e.g., a monkey (e.g., a marmoset, a baboon, a macaque), or an ape (e.g., a gorilla, a gibbon, an orangutan, or a chimpanzee). In some embodiments, the primate has or is at risk of having hearing loss. In some embodiments, the primate has been previously identified as having a mutation in a hair cell differentiation gene and/or a hair cell differentiation-suppressing gene. In some embodiments, the primate has been previously identified as having a mutation in a hair cell differentiation gene. In some embodiments, the primate has been previously identified as having a mutation in a hair cell differentiation-suppressing gene. In some embodiments, the primate has been identified as having a mutation in hair cell differentiation gene and/or a hair cell differentiation-suppressing gene and has been diagnosed with hearing loss. In some embodiments, the primate has been identified as having hearing loss.

A treatment is “therapeutically effective” when it results in a reduction in one or more of the number, severity, and frequency of one or more symptoms of a disease state (e.g., non-syndromic sensorineural hearing loss or syndromic sensorineural hearing loss) in a primate. In

some embodiments, a therapeutically effective amount of a composition can result in an increase in the expression level of an active hair cell differentiation protein (e.g., a wildtype, full-length hair cell differentiation protein, or an active variant of a hair cell differentiation protein,) (e.g., as compared to the expression level prior to treatment with the composition). In some

5 embodiments, a therapeutically effective amount of a composition can result in an increase in the expression level of an active hair cell differentiation protein (e.g., a wildtype, full-length hair cell differentiation protein or active variant) in a target cell (e.g., a supporting cell of the inner ear or a hair cell (e.g., an outer hair cell or an inner hair cell) of the inner ear). In some embodiments, a therapeutically effective amount of a composition can result in an increase in the expression level
10 of an active hair cell differentiation protein (e.g., a wildtype, full-length hair cell differentiation protein or active variant), and/or an increase in one or more activities of a hair cell differentiation protein in a target cell (e.g., as compared to a reference level, such as the level(s) in a primate cell prior to treatment, the level(s) in a primate cell having a mutation in a hair cell differentiation gene, or the level(s) in a primate cell or a population of primate cells from a
15 subject having non-syndromic sensorineural hearing loss, or the level(s) in a primate cell or a population of primate cells from a subject having syndromic sensorineural hearing loss).

The term “nucleic acid” or “polynucleotide” refers to deoxyribonucleic acid (DNA) or ribonucleic acid (RNA), or a combination thereof, in either single- or double-stranded form. Unless specifically limited, the term encompasses nucleic acids containing known analogues of
20 natural nucleotides that have similar binding properties as the reference nucleotides. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses complementary sequences as well as the sequence explicitly indicated. In some embodiments of any of the nucleic acids described herein, the nucleic acid is DNA. In some embodiments of any of the nucleic acids described herein, the nucleic acid is RNA.

25 The term “hair cell toxicity-inducing mutation” refers to a mutation in a hair cell differentiation gene that encodes a protein that when expressed (e.g., by a supporting cell or a hair cell) induces toxicity in a hair cell (e.g., in a primate).

The term “active hair cell differentiation protein” means a protein encoded by DNA that, if substituted for both wildtype alleles encoding full-length hair cell differentiation protein in
30 supporting cells of the inner ear of what is otherwise a wildtype primate, and if expressed in the supporting cells of that primate, results in that primate’s having a level of hearing approximating

the normal level of hearing of a similar primate that is entirely wildtype. Non-limiting examples of active hair cell differentiation proteins are full-length hair cell differentiation proteins (e.g., any of the full-length hair cell differentiation proteins described herein).

The term “inhibitory nucleic acid” refers to a nucleic acid sequence that hybridizes specifically to a target gene or a target mRNA (e.g., a hair cell differentiation-suppressing gene or a hair cell differentiation-suppressing mRNA) and thereby inhibits the expression and/or activity of the target gene or the target mRNA (e.g., a hair cell differentiation-suppressing gene or a hair cell differentiation-suppressing mRNA). In some embodiments, the inhibitory nucleic acid is a short interfering RNA (siRNA), a short hairpin RNA (shRNA), an antisense oligonucleotide, or a ribozyme. In some embodiments, the inhibitory nucleic acid is between about 10 nucleotides to about 30 nucleotides in length (e.g., about 10 nucleotides to about 28 nucleotides, about 10 nucleotides to about 26 nucleotides, about 10 nucleotides to about 24 nucleotides, about 10 nucleotides to about 22 nucleotides, about 10 nucleotides to about 20 nucleotides, about 10 nucleotides to about 18 nucleotides, about 10 nucleotides to about 16 nucleotides, about 10 nucleotides to about 14 nucleotides, about 10 nucleotides to about 12 nucleotides, about 12 nucleotides to about 30 nucleotides, about 12 nucleotides to about 28 nucleotides, about 12 nucleotides to about 26 nucleotides, about 12 nucleotides to about 24 nucleotides, about 12 nucleotides to about 22 nucleotides, about 12 nucleotides to about 20 nucleotides, about 12 nucleotides to about 18 nucleotides, about 12 nucleotides to about 16 nucleotides, about 12 nucleotides to about 14 nucleotides, about 16 nucleotides to about 30 nucleotides, about 16 nucleotides to about 28 nucleotides, about 16 nucleotides to about 26 nucleotides, about 16 nucleotides to about 24 nucleotides, about 16 nucleotides to about 22 nucleotides, about 16 nucleotides to about 20 nucleotides, about 16 nucleotides to about 18 nucleotides, about 18 nucleotides to about 30 nucleotides, about 18 nucleotides to about 28 nucleotides, about 18 nucleotides to about 26 nucleotides, about 18 nucleotides to about 24 nucleotides, about 18 nucleotides to about 22 nucleotides, about 18 nucleotides to about 20 nucleotides, about 20 nucleotides to about 30 nucleotides, about 20 nucleotides to about 28 nucleotides, about 20 nucleotides to about 26 nucleotides, about 20 nucleotides to about 24 nucleotides, about 20 nucleotides to about 22 nucleotides, about 22 nucleotides to about 30 nucleotides, about 22 nucleotides to about 28 nucleotides, about 22 nucleotides to about 26 nucleotides, about 22 nucleotides to about 24 nucleotides, about 24 nucleotides to about 30

nucleotides, about 24 nucleotides to about 28 nucleotides, about 24 nucleotides to about 26 nucleotides, about 26 nucleotides to about 30 nucleotides, about 26 nucleotides to about 28 nucleotides, about 28 nucleotides to about 30 nucleotides, or 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 or 30 nucleotides).

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Methods and materials are described herein for use in the present invention; other suitable methods and materials known in the art can also be used. The materials, methods, and examples are illustrative only and not intended to be limiting. All publications, patent applications, patents, sequences, database entries, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will control.

BRIEF DESCRIPTION OF DRAWINGS

Figure 1A is a representative image of Myo7a/Iba-1 immunofluorescent staining of cochlear tissue of a cynomolgus macaque (non-human primate) following administration of a single Anc80-GFP AAV vector directly into the inner ear through the round window.

Figure 1B is a representative image of Anc80-GFP immunofluorescent staining of the same cochlear tissue of the cynomolgus macaque as in Figure 1A.

Figure 1C is a representative image of a merged immunofluorescent staining of Myo7a/Iba-1 and Anc80-GFP of the same cochlear tissue of the cynomolgus macaque as in Figure 1A.

Figure 2A is a representative image of Anc80-GFP immunofluorescent staining of a NHP cochlear tissue showing the stria vascularis, the spiral ligament and the lateral wall.

Figure 2B is a representative image of Anc80-GFP immunofluorescent staining of the same NHP cochlear tissue as in Figure 2A showing the spiral limbus, the inner sulcus, inner hair cells (IHC) and outer hair cells (OHC).

Figure 3 is a simplified schematic diagram showing the proteins that play a role during the development of supporting cells and hair cells in the cochlea.

Figure 4A is an exemplary nucleic acid vector (SEQ ID NO: 66), that includes an ITR sequence (SEQ ID NO: 51), a CMV enhancer sequence (SEQ ID NO: 52), a CMV promoter

sequence (SEQ ID NO: 53), a human ATOH1 gene sequence (SEQ ID NO: 67), a 3xFlag sequence (SEQ ID NO: 62), a T2A sequence (SEQ ID NO: 63), a SV40-NLS sequence (SEQ ID NO: 54), a mScarlet gene sequence (SEQ ID NO: 55), a destabilizing domain (DD) sequence (SEQ ID NO: 59), a bGHpA sequence (SEQ ID NO: 56), and an ITR sequence (SEQ ID NO: 57).

Figure 4B is an exemplary nucleic acid vector (SEQ ID NO: 64), that includes an ITR sequence (SEQ ID NO: 51), a CMV enhancer sequence (SEQ ID NO: 52), a CMV promoter sequence (SEQ ID NO: 53), a human GFI1 gene sequence (SEQ ID NO: 65), a 3xFlag sequence (SEQ ID NO: 62), a T2A sequence (SEQ ID NO: 63), a SV40-NLS sequence (SEQ ID NO: 54), a mScarlet sequence (SEQ ID NO: 55), a destabilizing domain (DD) sequence (SEQ ID NO: 59), a bGHpA sequence (SEQ ID NO: 56), and an ITR sequence (SEQ ID NO: 57).

Figure 4C is an exemplary nucleic acid vector (SEQ ID NO: 60), that includes an ITR sequence (SEQ ID NO: 51), a CMV enhancer sequence (SEQ ID NO: 52), a CMV promoter sequence (SEQ ID NO: 53), a human POU4F3 gene sequence (SEQ ID NO: 61), a 3xFlag sequence (SEQ ID NO: 62), a T2A sequence (SEQ ID NO: 63), a SV40-NLS sequence (SEQ ID NO: 54), a mScarlet sequence (SEQ ID NO: 55), a destabilizing domain (DD) sequence (SEQ ID NO: 59), a bGHpA sequence (SEQ ID NO: 56), and an ITR sequence (SEQ ID NO: 57).

Figure 4D is an exemplary nucleic acid vector (SEQ ID NO: 68), that includes an ITR sequence (SEQ ID NO: 51), a CMV enhancer sequence (SEQ ID NO: 52), a CMV promoter sequence (SEQ ID NO: 53), a luciferase (Fluc) gene sequence (SEQ ID NO: 69), a T2A sequence (SEQ ID NO: 63), an mScarlet gene sequence (SEQ ID NO: 55), a SV40pA sequence (SEQ ID NO: 70), a U6 sequence (SEQ ID NO: 71), a short hairpin RNA (shRNA) sequence (SEQ ID NO: 72), and an ITR sequence (SEQ ID NO: 57).

Figure 5A is bar graph showing the relative quantification of Hes-1 RNA in HEK293FT cells transfected with combinations of dual and triple shRNA constructs (S3 (GAAAGTCATCAAAGCCTAT; SEQ ID NO: 73), S5 (ACTGCATGACCCAGATCAA; SEQ ID NO: 74), Kop (ACTGCATGACCCAGATCAA; SEQ ID NO: 75), S3 plus S5, S3 plus Kop, and S5 plus Kop) as determined by real time quantitative polymerase chain reaction (RTqPCR).

Figure 5B is bar graph showing the relative quantification of Hes-1 protein in HEK293FT cells transfected with combinations of dual and triple shRNA constructs (S3, S5, Kop, S3 plus S5, S3 plus Kop, and S5 plus Kop) as determined by Western blotting.

Figure 6A is a bar graph showing the relative quantification of ATOH1, POU4F3, and GFI1 (APG) RN in HEK293FT cells transfected with the individual plasmids of Figures 4A-C.

Figure 6B is an image of a Western blot showing the relative quantification of ATOH1, POU4F3 and GFI1 protein expression in HEK293FT cells transfected with the individual
5 plasmids of Figures 4A-C.

Figure 7A is an exemplary nucleic acid vector (SEQ ID NO: 76), that includes an ITR sequence (SEQ ID NO: 51), a CMV promoter sequence (SEQ ID NO: 53), a mScarlet sequence (SEQ ID NO: 55), a bGHpA sequence (SEQ ID NO: 56) and an ITR sequence (SEQ ID NO: 57).

Figure 7B is an exemplary nucleic acid vector (SEQ ID NO: 77), that includes an ITR
10 sequence (SEQ ID NO: 51), a CMV promoter sequence (SEQ ID NO: 53), a mScarlet sequence (SEQ ID NO: 55), a destabilizing domain (DD) sequence (SEQ ID NO: 59), a bGHpA sequence (SEQ ID NO: 56) and an ITR sequence (SEQ ID NO: 57).

Figure 8A is a dose response curve showing the functionality and reversibility of the destabilizing domain (DD) using fluorescence microscopy. Serial dilutions of TMP (0.1 μ M, 1
15 μ M, 10 μ M, 20 μ M and 100 μ M) were tested in the mScarlet and mScarlet-DD transfected HEK293FT cells.

Figure 8B is a graph showing the functionality and reversibility of the destabilizing domain (DD) by flow cytometry (Attune flow cytometer).

Figure 9A is an image showing mScarlet positive cells in a P1-P3 mouse cochlea explant
20 transfected with AAVanc80 vector at various MOIs. 10 μ M TMP was added at a later time point.

Figure 9B is an image showing mScarlet positive HEK293FT cells transfected with AAVanc80 vector at various MOIs. 10 μ M TMP was added at a later time point.

Figure 10 is an image showing mScarlet positive hair cells and supporting cells in a
25 cochlear explants infected with AAVanc80 with and without 10uM TMP that was added at a later time point.

Figure 11A is an exemplary nucleic acid vector (SEQ ID NO: 83), that includes an ITR sequence (SEQ ID NO: 51), a U6 sequence (SEQ ID NO: 84), a short hairpin HES1 RNA (shHES1) sequence (SEQ ID NO: 85), a CMV enhancer sequence (SEQ ID NO: 52), a CMV
30 promoter sequence (SEQ ID NO: 53), a 3xFlag sequence (SEQ ID NO: 86), a human ATOH1 gene sequence (SEQ ID NO: 87), a destabilizing domain (DD) sequence (SEQ ID NO: 88), a

T2A sequence (SEQ ID NO: 89), a human POU4F3 gene sequence (SEQ ID NO: 61), a bGHpA sequence (SEQ ID NO: 90), a U6 sequence (SEQ ID NO: 91), a short hairpin HES1 RNA (shHES1-2) sequence (SEQ ID NO: 92) and an ITR sequence (SEQ ID NO: 57).

Figure 11B is an exemplary nucleic acid vector (SEQ ID NO: 93), that includes an ITR sequence (SEQ ID NO: 51), a U6 sequence (SEQ ID NO: 84), a short hairpin HES1 RNA (shHES-1) sequence (SEQ ID NO: 85), an ATOH1 enhancer-promoter sequence (SEQ ID NO: 94), a 3xFlag sequence (SEQ ID NO: 86), a human ATOH1 gene sequence (SEQ ID NO: 67), a T2A sequence (SEQ ID NO: 63), a human POU4F3 gene sequence (SEQ ID NO: 95), a bGHpA sequence (SEQ ID NO: 90), a U6 sequence (SEQ ID NO: 84), a short hairpin HES1 RNA (shHES1-2) sequence (SEQ ID NO: 92) and an ITR sequence (SEQ ID NO: 57).

Figure 12A is a bar graph showing the relative quantification of ATOH1, POU4F3, and HES1 in HEK293FT cells transfected with the combined plasmids of Figures 11A-B.

Figure 12B is an image of a Western blot showing the relative quantification of 3xFlag-ATOH1 and HES1 protein expression in HEK293FT cells transfected with the combined plasmids of Figures 11A-B.

DETAILED DESCRIPTION

Provided herein are compositions including at least two different nucleic acid vectors, where: each of the at least two different adeno-associated virus (AAV) vectors comprises a coding sequence that encodes a different portion of a hair cell differentiation protein, each of the encoded portions being at least 30 amino acid residues in length, where the amino acid sequence of each of the encoded portions may optionally partially overlap with the amino acid sequence of a different one of the encoded portions; no single vector of the at least two different vectors encodes the full-length hair cell differentiation protein; at least one of the coding sequences includes a nucleotide sequence spanning two neighboring exons of hair cell differentiation genomic DNA, and lacks an intronic sequence between the two neighboring exons; and when introduced into a primate cell (e.g., a hair cell or a supporting cell of the inner ear) the at least two different vectors undergo concatamerization or homologous recombination with each other, thereby forming a recombined nucleic acid that encodes a full-length hair cell differentiation protein that is expressed in the primate cell.

Also provided herein are compositions including two different nucleic acid vectors, where: a first nucleic acid vector of the two different nucleic acid vectors includes a promoter, a

first coding sequence that encodes an N-terminal portion of a hair cell differentiation protein positioned 3' of the promoter, and a splicing donor signal sequence positioned at the 3' end of the first coding sequence; and a second nucleic acid vector of the two different nucleic acid vectors includes a splicing acceptor signal sequence, a second coding sequence that encodes a C-terminal portion of a hair cell differentiation protein positioned at the 3' end of the splicing acceptor signal sequence, and a polyadenylation sequence at the 3' end of the second coding sequence; where each of the encoded portions is at least 30 amino acid residues in length, where the amino acid sequences of the encoded portions do not overlap, where no single vector of the two different vectors encodes the full-length hair cell differentiation protein, and, when the coding sequences are transcribed in a primate cell (e.g., a hair cell or a supporting cell of the inner ear), to produce RNA transcripts, splicing occurs between the splicing donor signal sequence on one transcript and the splicing acceptor signal sequence on the other transcript, thereby forming a recombined RNA molecule that encodes a full-length hair cell differentiation protein.

Also provided herein are compositions including: a first nucleic acid vector including a promoter, a first coding sequence that encodes an N-terminal portion of a hair cell differentiation protein positioned 3' of the promoter, a splicing donor signal sequence positioned at the 3' end of the first coding sequence, and a first detectable marker gene positioned 3' of the splicing donor signal sequence; and a second nucleic acid vector, different from the first nucleic acid vector, including a second detectable marker gene, a splicing acceptor signal sequence positioned 3' of the second detectable marker gene, a second coding sequence that encodes a C-terminal portion of a hair cell differentiation protein positioned at the 3' end of the splicing acceptor signal sequence, and a polyadenylation sequence positioned at the 3' end of the second coding sequence; where each of the encoded portions is at least 30 amino acid residues in length, where the respective amino acid sequences of the encoded portions do not overlap with each other, where no single vector of the two different vectors encodes the full-length hair cell differentiation protein, and, when the coding sequences are transcribed in a primate cell (e.g., a hair cell or a supporting cell of the inner ear) to produce RNA transcripts, splicing occurs between the splicing donor signal on one transcript and the splicing acceptor signal on the other transcript, thereby forming a recombined RNA molecule that encodes a full-length hair cell differentiation protein.

Also provided herein are compositions including: a first nucleic acid vector including a promoter, a first coding sequence that encodes an N-terminal portion of a hair cell differentiation protein positioned 3' to the promoter, a splicing donor signal sequence positioned at the 3' end of the first coding sequence, and a F1 phage recombinogenic region positioned 3' to the splicing donor signal sequence; and a second nucleic acid vector, different from the first nucleic acid vector, including a second F1 phage recombinogenic region, a splicing acceptor signal sequence positioned 3' of the second F1 phage recombinogenic region, a second coding sequence that encodes a C-terminal portion of a hair cell differentiation protein positioned at the 3' end of the splicing acceptor signal sequence, and a polyadenylation sequence positioned at the 3' end of the second coding sequence; where each of the encoded portions is at least 30 amino acid residues in length, where the respective amino acid sequences of the encoded portions do not overlap with each other, where no single vector of the two different vectors encodes the full-length hair cell differentiation protein, and, when the coding sequences are transcribed in a primate cell (e.g., a hair cell or a supporting cell of the inner ear) to produce RNA transcripts, splicing occurs between the splicing donor signal on one transcript and the splicing acceptor signal on the other transcript, thereby forming a recombined RNA molecule that encodes a full-length hair cell differentiation protein.

Also provided herein are compositions including a single adeno-associated virus (AAV) vector, where the single AAV vector comprises a nucleic acid sequence that encodes a hair cell differentiation protein; and when introduced into a primate cell (e.g., a hair cell or a supporting cell of the inner ear), a nucleic acid encoding the hair cell differentiation protein is generated at the locus of the hair cell differentiation gene and the primate cell expresses the hair cell differentiation protein. Also provided herein are compositions including a single adeno-associated virus (AAV) vector that encodes an inhibitory nucleic acid that decreases the expression of a hair cell differentiation-suppressing protein in a primate cell (e.g., a hair cell or a supporting cell of the inner ear).

Also provided herein are methods of promoting differentiation of a supporting cell of an inner ear of a primate into a hair cell that include: administering to the inner ear of the primate a therapeutically effective amount of any of the compositions described herein, where the administering promotes differentiation of the supporting cell of the inner ear of the primate into a hair cell. Also provided herein are methods of increasing the expression level of a hair cell

differentiation protein in a supporting cell of an inner ear of a primate that include: administering to the inner ear of the primate a therapeutically effective amount of any of the compositions described herein, where the administering results in an increase in the expression level of the hair cell differentiation protein in the supporting cell of the inner ear of the primate.

5 Also provided herein are methods of decreasing the expression level of a hair cell differentiation-suppressing protein in a supporting cell or a hair cell of an inner ear of a primate that include: administering to the inner ear of the primate a therapeutically effective amount of any of the compositions described herein, where the administering results in a decrease in the expression level of the hair cell differentiation-suppressing protein in the supporting cell or the
10 hair cell of the inner ear of the primate.

 Also provided herein are methods of increasing the number of functional hair cells in a primate in need thereof that include: administering to the inner ear of the primate a therapeutically effective amount of any of the compositions described herein. Also provided herein are methods of improving hearing in a primate in need thereof, the method comprising
15 administering to the inner ear of the primate a therapeutically effective amount of any of the compositions described herein.

 Also provided herein are methods of repairing a hair cell toxicity-inducing mutation in an endogenous hair cell differentiation gene locus in a supporting cell or a hair cell of an inner ear of a primate that include: administering to the inner ear of the primate a therapeutically effective
20 amount of any of the compositions described herein, where the administering results in repair of the hair cell toxicity-inducing mutation in the endogenous hair cell differentiation gene locus in the supporting cell or the hair cell of the inner ear of the primate.

 Also provided herein are methods of decreasing the risk of hearing loss due to hair cell loss or dysfunction in a primate in need thereof that include: administering to the inner ear of the
25 primate a therapeutically effective amount of any of the compositions described herein. Also provided herein are methods that include introducing into a cochlea of a mammal a therapeutically effective amount of any of the compositions described herein.

 Also provided are kits that include any of the compositions described herein.

 Additional non-limiting aspects of the compositions, kits, and methods are described
30 herein and can be used in any combination without limitation.

Hair Cell Differentiation Genes

The term “hair cell differentiation gene” refers to a gene encoding a protein (e.g., a transcription factor) that positively contributes, either directly or indirectly, to hair cell differentiation and viability in a primate (e.g., a human). Non-limiting examples of hair cell differentiation genes include: ATOH1, POU4F3, CTNNB1, NOG, GFI-1, NTF3, and BDNF.

The term “mutation in a hair cell differentiation gene” refers to a modification in a wildtype hair cell differentiation gene that results in the production of a hair cell differentiation protein having one or more of: a deletion in one or more amino acids, one or more amino acid substitutions, and one or more amino acid insertions as compared to the wildtype hair cell differentiation protein, and/or results in a decrease in the expressed level of the encoded hair cell differentiation protein in a primate cell as compared to the expressed level of the encoded hair cell differentiation protein in a primate cell not having a mutation. In some embodiments, a mutation can result in the production of a hair cell differentiation protein having a deletion in one or more amino acids (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 amino acids). In some embodiments, the mutation can result in a frameshift in the hair cell differentiation gene. The term “frameshift” is known in the art to encompass any mutation in a coding sequence that results in a shift in the reading frame of the coding sequence. In some embodiments, a frameshift can result in a nonfunctional protein. In some embodiments, a point mutation can be a nonsense mutation (i.e., results in a premature stop codon in an exon of the gene). A nonsense mutation can result in the production of a truncated protein (as compared to a corresponding wildtype protein) that may or may not be functional. In some embodiments, the mutation can result in the loss (or a decrease in the level) of expression of hair cell differentiation mRNA or hair cell differentiation protein, or both the mRNA and protein. In some embodiments, the mutation can result in the production of an altered hair cell differentiation protein having a loss or decrease in one or more biological activities (functions) as compared to a wildtype hair cell differentiation protein.

In some embodiments, the mutation is an insertion of one or more nucleotides into a hair cell differentiation gene. In some embodiments, the mutation is in a regulatory sequence of the hair cell differentiation gene, i.e., a portion of the gene that is not coding sequence. In some embodiments, a mutation in a regulatory sequence may be in a promoter or enhancer region and prevent or reduce the proper transcription of the hair cell differentiation gene.

For example, an active hair cell differentiation protein can include a sequence of a wildtype, full-length hair cell differentiation protein (e.g., a wildtype, human, full-length hair cell differentiation protein) including 1 amino acid substitution to about 160 amino acid substitutions, 1 amino acid substitution to about 155 amino acid substitutions, 1 amino acid substitution to about 150 amino acid substitutions, 1 amino acid substitution to about 145 amino acid substitutions, 1 amino acid substitution to about 140 amino acid substitutions, 1 amino acid substitution to about 135 amino acid substitutions, 1 amino acid substitution to about 130 amino acid substitutions, 1 amino acid substitution to about 125 amino acid substitutions, 1 amino acid substitution to about 120 amino acid substitutions, 1 amino acid substitution to about 115 amino acid substitutions, 1 amino acid substitution to about 110 amino acid substitutions, 1 amino acid substitution to about 105 amino acid substitutions, 1 amino acid substitution to about 100 amino acid substitutions, 1 amino acid substitution to about 95 amino acid substitutions, 1 amino acid substitution to about 90 amino acid substitutions, 1 amino acid substitution to about 85 amino acid substitutions, 1 amino acid substitution to about 80 amino acid substitutions, 1 amino acid substitution to about 75 amino acid substitutions, 1 amino acid substitution to about 70 amino acid substitutions, 1 amino acid substitution to about 65 amino acid substitutions, 1 amino acid substitution to about 60 amino acid substitutions, 1 amino acid substitution to about 55 amino acid substitutions, 1 amino acid substitution to about 50 amino acid substitutions, 1 amino acid substitution to about 45 amino acid substitutions, 1 amino acid substitution to about 40 amino acid substitutions, 1 amino acid substitution to about 35 amino acid substitutions, 1 amino acid substitution to about 30 amino acid substitutions, 1 amino acid substitution to about 25 amino acid substitutions, 1 amino acid substitution to about 20 amino acid substitutions, 1 amino acid substitution to about 15 amino acid substitutions, 1 amino acid substitution to about 10 amino acid substitutions, 1 amino acid substitution to about 9 amino acid substitutions, 1 amino acid substitution to about 8 amino acid substitutions, 1 amino acid substitution to about 7 amino acid substitutions, 1 amino acid substitution to about 6 amino acid substitutions, 1 amino acid substitution to about 5 amino acid substitutions, 1 amino acid substitution to about 4 amino acid substitutions, 1 amino acid substitution to about 3 amino acid substitutions, between about 2 amino acid substitutions to about 160 amino acid substitutions, about 2 amino acid substitutions to about 155 amino acid substitutions, about 2 amino acid substitutions to about 150 amino acid substitutions, about 2 amino acid substitutions to about 145 amino acid substitutions, about 2

21

22

5
10
15
20
25
30

24

5

10

15

20

25

30

[illegible]

5
10
15
20
25
30

5

10

15

20

25

30

5
10
15
20
25
30

5

10

15

20

25

30

5
10
15
20
25
30

[illegible]

5
10
15
20
25
30

5
10
15
20
25
30

substitutions to about 135 amino acid substitutions, between about 140 amino acid substitutions to about 160 amino acid substitutions, about 140 amino acid substitutions to about 155 amino acid substitutions, about 140 amino acid substitutions to about 150 amino acid substitutions, about 140 amino acid substitutions to about 145 amino acid substitutions, between about 150 amino acid substitutions to about 160 amino acid substitutions, or about 150 amino acid substitutions to about 155 amino acid substitutions. One skilled in the art would appreciate that amino acids that are not conserved between wildtype hair cell differentiation proteins from different species can be mutated without losing activity, while those amino acids that are conserved between wildtype hair cell differentiation proteins from different species should not be mutated as they are more likely (than amino acids that are not conserved between different species) to be involved in activity.

An active hair cell differentiation protein can include, e.g., a sequence of a wildtype, full-length hair cell differentiation protein (e.g., a wildtype, human, full-length hair cell differentiation protein) that has 1 amino acid to about 50 amino acids, 1 amino acid to about 45 amino acids, 1 amino acid to about 40 amino acids, 1 amino acid to about 35 amino acids, 1 amino acid to about 30 amino acids, 1 amino acid to about 25 amino acids, 1 amino acid to about 20 amino acids, 1 amino acid to about 15 amino acids, 1 amino acid to about 10 amino acids, 1 amino acid to about 9 amino acids, 1 amino acid to about 8 amino acids, 1 amino acid to about 7 amino acids, 1 amino acid to about 6 amino acids, 1 amino acid to about 5 amino acids, 1 amino acid to about 4 amino acids, 1 amino acid to about 3 amino acids, about 2 amino acids to about 50 amino acids, about 2 amino acids to about 45 amino acids, about 2 amino acids to about 40 amino acids, about 2 amino acids to about 35 amino acids, about 2 amino acids to about 30 amino acids, about 2 amino acids to about 25 amino acids, about 2 amino acids to about 20 amino acids, about 2 amino acids to about 15 amino acids, about 2 amino acids to about 10 amino acids, about 2 amino acids to about 9 amino acids, about 2 amino acids to about 8 amino acids, about 2 amino acids to about 7 amino acids, about 2 amino acids to about 6 amino acids, about 2 amino acids to about 5 amino acids, about 2 amino acids to about 4 amino acids, about 3 amino acids to about 50 amino acids, about 3 amino acids to about 45 amino acids, about 3 amino acids to about 40 amino acids, about 3 amino acids to about 35 amino acids, about 3 amino acids to about 30 amino acids, about 3 amino acids to about 25 amino acids, about 3 amino acids to about 20 amino acids, about 3 amino acids to about 15 amino acids, about 3

36

amino acids to about 25 amino acids, about 10 amino acids to about 20 amino acids, about 10 amino acids to about 15 amino acids, about 15 amino acids to about 50 amino acids, about 15 amino acids to about 45 amino acids, about 15 amino acids to about 40 amino acids, about 15 amino acids to about 35 amino acids, about 15 amino acids to about 30 amino acids, about 15 amino acids to about 25 amino acids, about 15 amino acids to about 20 amino acids, about 20 amino acids to about 50 amino acids, about 20 amino acids to about 45 amino acids, about 20 amino acids to about 40 amino acids, about 20 amino acids to about 35 amino acids, about 20 amino acids to about 30 amino acids, about 20 amino acids to about 25 amino acids, about 25 amino acids to about 50 amino acids, about 25 amino acids to about 45 amino acids, about 25 amino acids to about 40 amino acids, about 25 amino acids to about 35 amino acids, about 25 amino acids to about 30 amino acids, about 30 amino acids to about 50 amino acids, about 30 amino acids to about 45 amino acids, about 30 amino acids to about 40 amino acids, about 30 amino acids to about 35 amino acids, about 35 amino acids to about 50 amino acids, about 35 amino acids to about 45 amino acids, about 35 amino acids to about 40 amino acids, about 40 amino acids to about 50 amino acids, about 40 amino acids to about 45 amino acids, about 45 amino acids to about 50 amino acids, deleted. In some embodiments where two or more amino acids are deleted from the sequence of a wildtype, full-length hair cell differentiation protein, the two or more deleted amino acids can be contiguous in the sequence of the wildtype, full-length protein. In other examples where two or more amino acids are deleted from the sequence of a wildtype, full-length hair cell differentiation protein, the two or more deleted amino acids are not contiguous in the sequence of the wildtype, full-length protein. One skilled in the art would appreciate that amino acids that are not conserved between wildtype, full-length hair cell differentiation proteins from different species can be deleted without losing activity, while those amino acids that are conserved between wildtype, full-length hair cell differentiation proteins from different species should not be deleted as they are more likely (than amino acids that are not conserved between different species) to be involved in activity.

In some examples, an active hair cell differentiation protein can, e.g., include a sequence of a wildtype, full-length hair cell differentiation protein that has between 1 amino acid to about 100 amino acids, 1 amino acid to about 95 amino acids, 1 amino acid to about 90 amino acids, 1 amino acid to about 85 amino acids, 1 amino acid to about 80 amino acids, 1 amino acid to about 75 amino acids, 1 amino acid to about 70 amino acids, 1 amino acid to about 65 amino acids, 1

5
10
15
20
25
30

[illegible]

40

41

amino acids to about 75 amino acids, about 80 amino acids to about 100 amino acids, about 80 amino acid to about 95 amino acids, about 80 amino acids to about 90 amino acids, about 80 amino acids to about 85 amino acids, about 90 amino acids to about 100 amino acids, about 90 amino acids to about 95 amino acids, or about 95 amino acids to about 100 amino acids,

5 removed from its N-terminus and/or 1 amino acid to 100 amino acids (or any of the subranges of this range described herein) removed from its C-terminus.

In some embodiments, an active hair cell differentiation protein can, e.g., include the sequence of a wildtype, full-length hair cell differentiation protein where 1 amino acid to 50 amino acids, 1 amino acid to 45 amino acids, 1 amino acid to 40 amino acids, 1 amino acid to 35 amino acids, 1 amino acid to 30 amino acids, 1 amino acid to 25 amino acids, 1 amino acid to 20 amino acids, 1 amino acid to 15 amino acids, 1 amino acid to 10 amino acids, 1 amino acid to 9 amino acids, 1 amino acid to 8 amino acids, 1 amino acid to 7 amino acids, 1 amino acid to 6 amino acids, 1 amino acid to 5 amino acids, 1 amino acid to 4 amino acids, 1 amino acid to 3 amino acids, about 2 amino acids to 50 amino acids, about 2 amino acids to 45 amino acids, about 2 amino acids to 40 amino acids, about 2 amino acids to 35 amino acids, about 2 amino acids to 30 amino acids, about 2 amino acids to 25 amino acids, about 2 amino acids to 20 amino acids, about 2 amino acids to 15 amino acids, about 2 amino acids to 10 amino acids, about 2 amino acids to 9 amino acids, about 2 amino acids to 8 amino acids, about 2 amino acids to 7 amino acids, about 2 amino acids to 6 amino acids, about 2 amino acids to 5 amino acids, about 2 amino acids to 4 amino acids, about 3 amino acids to 50 amino acids, about 3 amino acids to 45 amino acids, about 3 amino acids to 40 amino acids, about 3 amino acids to 35 amino acids, about 3 amino acids to 30 amino acids, about 3 amino acids to 25 amino acids, about 3 amino acids to 20 amino acids, about 3 amino acids to 15 amino acids, about 3 amino acids to 10 amino acids, about 3 amino acids to 9 amino acids, about 3 amino acids to 8 amino acids, about 3 amino acids to 7 amino acids, about 3 amino acids to 6 amino acids, about 3 amino acids to 5 amino acids, about 4 amino acids to 50 amino acids, about 4 amino acids to 45 amino acids, about 4 amino acids to 40 amino acids, about 4 amino acids to 35 amino acids, about 4 amino acids to 30 amino acids, about 4 amino acids to 25 amino acids, about 4 amino acids to 20 amino acids, about 4 amino acids to 15 amino acids, about 4 amino acids to 10 amino acids, about 4 amino acids to 9 amino acids, about 4 amino acids to 8 amino acids, about 4 amino acids to 7 amino acids, about 4 amino acids to 6 amino acids, about 5 amino acids to 50 amino acids, about 5

amino acids to 45 amino acids, about 5 amino acids to 40 amino acids, about 5 amino acids to
35 amino acids, about 5 amino acids to 30 amino acids, about 5 amino acids to 25 amino acids,
about 5 amino acids to 20 amino acids, about 5 amino acids to 15 amino acids, about 5 amino
acids to 10 amino acids, about 5 amino acids to 9 amino acids, about 5 amino acids to 8 amino
5 acids, about 5 amino acids to 7 amino acids, about 6 amino acids to 50 amino acids, about 6
amino acids to 45 amino acids, about 6 amino acids to 40 amino acids, about 6 amino acids to
35 amino acids, about 6 amino acids to 30 amino acids, about 6 amino acids to 25 amino acids,
about 6 amino acids to 20 amino acids, about 6 amino acids to 15 amino acids, about 6 amino
acids to 10 amino acids, about 6 amino acids to 9 amino acids, about 6 amino acids to 8 amino
10 acids; about 7 amino acids to 50 amino acids, about 7 amino acids to 45 amino acids, about 7
amino acids to 40 amino acids, about 7 amino acids to 35 amino acids, about 7 amino acids to 30
amino acids, about 7 amino acids to 25 amino acids, about 7 amino acids to 20 amino acids,
about 7 amino acids to 15 amino acids, about 7 amino acids to 10 amino acids, about 7 amino
acids to 9 amino acids, about 8 amino acids to 50 amino acids, about 8 amino acids to 45 amino
15 acids, about 8 amino acids to 40 amino acids, about 8 amino acids to 35 amino acids, about 8
amino acids to 30 amino acids, about 8 amino acids to 25 amino acids, about 8 amino acids to 20
amino acids, about 8 amino acids to 15 amino acids, about 8 amino acids to 10 amino acids,
about 10 amino acids to 50 amino acids, about 10 amino acids to 45 amino acids, about 10
amino acids to 40 amino acids, about 10 amino acids to 35 amino acids, about 10 amino acids to
20 amino acids, about 10 amino acids to 25 amino acids, about 10 amino acids to 20 amino
20 acids, about 10 amino acids to 15 amino acids, about 15 amino acids to 50 amino acids, about 15
amino acids to 45 amino acids, about 15 amino acids to 40 amino acids, about 15 amino acids to
35 amino acids, about 15 amino acids to 30 amino acids, about 15 amino acids to 25 amino
acids, about 15 amino acids to 20 amino acids, about 20 amino acids to 50 amino acids, about 20
25 amino acids to 45 amino acids, about 20 amino acids to 40 amino acids, about 20 amino acids to
35 amino acids, about 20 amino acids to 30 amino acids, about 20 amino acids to 25 amino
acids, about 25 amino acids to 50 amino acids, about 25 amino acids to 45 amino acids, about 25
amino acids to 40 amino acids, about 25 amino acids to 35 amino acids, about 25 amino acids to
30 amino acids, about 30 amino acids to 50 amino acids, about 30 amino acids to 45 amino
30 acids, about 30 amino acids to 40 amino acids, about 30 amino acids to 35 amino acids, about 35
amino acids to 50 amino acids, about 35 amino acids to 45 amino acids, about 35 amino acids to

40 amino acids, about 40 amino acids to 50 amino acids, about 40 amino acids to 45 amino acids, or about 45 amino acids to about 50 amino acids, are inserted. In some examples, the 1 amino acid to 50 amino acids (or any subrange thereof) can be inserted as a contiguous sequence into the sequence of a wildtype, full-length protein. In some examples, the 1 amino acid to 50 amino acids (or any subrange thereof) are not inserted as a contiguous sequence into the sequence of a wildtype, full-length protein. As can be appreciated in the art, the 1 amino acid to 50 amino acids can be inserted into a portion of the sequence of a wildtype, full-length protein that is not well-conserved between species.

Atonal basic Helix-Loop-Helix transcription factor 1 (Atoh1)

The ATOH1 gene encodes atonal basic helix-loop-helix (bHLH) transcription factor 1. ATOH1 is a primary regulator of hair cell differentiation (Kawamoto et al., *J. Neurosci.* (2003) 23(11):4395-4400; Izumikawa et al. (2005) *Nat. Med.* 11(3): 271-276; Minoda et al. (2007) *Hear Res.* 232(1-2): 44-51; Atkinson et al. (2014) *PLoS One* 9(7): e102077; Kuo et al. (2015) *J. Neurosci* 35(30): 10786-10798; Walters et al. (2017) *Cell Rep* 19(2): 307-320).

The human ATOH1 gene is located on chromosome 4q22. It contains 1 exon encompassing ~2 kilobases (kb) (NCBI Accession No. NM_005172.1). The full-length wildtype ATOH1 protein expressed from the human ATOH1 gene is 354 amino acids in length.

Non-limiting examples of detecting techniques include: real-time polymerase chain reaction (RT-PCR), PCR, sequencing, Southern blotting, and Northern blotting.

An exemplary human wildtype ATOH1 protein is or includes the sequence of SEQ ID NO: 1. Non-limiting examples of a nucleic acid encoding a wildtype ATOH1 protein is or includes SEQ ID NO: 4. As can be appreciated in the art, at least some or all of the codons in SEQ ID NO: 4 can be codon-optimized to allow for optimal expression in a non-human primate.

Human Full-length Wildtype ATOH1 Protein (SEQ ID NO: 1)

MSRLLHAEWEAEVKELGDHHRQPQPHHLPQPPPPQPPATLQAREHPVYPPELSLLDSTDPRAWLAPTLOGICTARA
AQYLLHSPELGASEAAAPRDEV DGRGELVRRSSGGASSSKSPGPVKVREQLCKLKGGVVVDELGCSRQRAPSSKQVN
GVQKQRRLLAANARERRRMHGLNHAFDQLRNVI PSFNNDKKLSKYETLQMAQIYINALSELLQTPSGGEQPPPPPASC
KSDHHLRLTAASYEGGAGNATAAGAAQASGGGSRPTPPGSCRTRFSAPASAGGYSVQLDALHFSTFEDSALTAMMAQ
KNLSPSLPGSILQPVQEENSKTSPRSHRSDGEFSPHSHYSDSDEAS

Mouse Full-Length Wildtype ATOH1 Protein (SEQ ID NO: 2)

MSRLLHAEWEAEVKELGDHHRHPQPHVPLTPQPPATLQARDLPVYPAELSLLDSTDPRAWLTPTLQGLCTARAAQ
 YLLHSPELGASEAAAPRDEADSQGELVRRSGCGGLSKSPGPVKVREQLCKLKGGVVVDELGCSRQRAPSSKQVNGVQ
 KQRRLAANARERRRMHGLNHAFDQLRNVIPSFNNDKKLSKYETLQMAQIYINALSELLQTPNVGEQPPPTASCKND
 5 HHHLR TASSYEGGAGASAVAGAPAPGGGPRPTPPGPCRTRFSGPASSGGYSVQLDALHFPAFEDRAL TAMMAQKDL
 SPSLPGGILQPVEDNSKTSRSHRSDGEFSPHSHYSDSDEAS

Rat Full-Length Wildtype ATOH1 Protein (SEQ ID NO: 3)

MSRLLHAEWEAEVKELGDHHRHPQPHHIPQLTPQPPATLQARDHPVYPAELSLLDSTDPRAWLTPTLQGLCTARAAQ
 10 YLLHSPELGASEAAAPGDEADGQGELVRRSGCGGLSKSPGPVKVREQLCKLKGGVVVDELGCSRQRAPSSKQVNGVQ
 KQRRLAANARERRRMHGLNHAFDQLRNVIPSFNNDKKLSKYETLQMAQIYINALSELLQTPSVGEQPPPPAASCKND
 HHHLR AAAASYEGGAGASAVAGAPAPGGGPRPTPPGACRTRFSTPASSGGYSVQLDALHFPAFEDRAL TAMMAQKDL
 SPSLPGGILQVPEDSSKTSRSHRSDGEFSPHSHYSDSDEAS

Human Wildtype ATOH1 cDNA (SEQ ID NO: 4)

atgtcccgctgctgcatgcagaagagtgggctgaagtgaaggagttgggagaccaccatcgccagccccagccgca
 tcatctcccgcaaccgcccgcgcgcgcgcagccaccttgcaggcgagagagcatcccgctctaccgcctg
 agctgtccctcctggacagcaccgaccacgcgcctggctggctcccactttgcaggcatctgcacggcacgcgcc
 gccagtatcttgctacattccccggagctgggtgcctcagaggccgctgcgccccgggacgaggtggacggccggg
 20 ggagctggtaaggaggagcagcggcggtgccagcagcagcaagagccccgggcccgggtgaaagtgcgggaacagctgt
 gcaagctgaaaggcgggggtggtgtagacgagctgggctgcagccgccaacgggccccttcagcaaacaggtgaat
 ggggtgcagaagcagagacggctagcagccaacgccaggagcggcgagcaggtgcatgggtgaaccacgccttcga
 ccagctgcgcaatgttatcccgctcgttcaacaacgacaagaagctgtccaaatatgagaccctgcagatggccaaa
 tctacatcaacgccttgtccgagctgctacaaacgcccagcgagggggaacagccaccgcccgcctccagcctcctgc
 25 aaaagcgaccaccaccaccttcgaccgcggcctcctatgaagggggcgcgggcaacgcgaccgcagctggggctca
 gcaggcttcggaggaggagccagcggccgacccccgcccgggagttgccggactcgcttctcagccccagcttctgcgg
 gaggggtactcgggtgcagctggacgctctgcacttctcgactttcgaggacagcgccctgacagcgatgatggcgcaa
 aagaatttgtctccttctctccccgggagcatcttgagccagtgagggagaaaacagcaaaacttcgcctcggtc
 ccacagaagcgacggggaattttccccccattcccattacgtgactcggtatgaggcaagtttag

A non-limiting example of a human wildtype ATOH1 genomic DNA sequence is SEQ ID NO: 5. The exon in SEQ ID NO: 5 is: nucleotide positions 1-1065 (exon 1).

Human Wildtype ATOH1 Gene (SEQ ID NO: 5)

1 atgtcccgcc tgctgcatgc agaagagtgg gctgaagtga aggagttggg agaccacat
 61 cgccagcccc agccgcatca tctcccgcaa ccgcccgcgc cgccgcagcc acctgcaact

121 ttgcaggcga gagagcatcc cgtctaccgc cctgagctgt ccctcctgga cagcaccgac
 181 ccacgcgcct ggctggctcc cactttgcag ggcatctgca cggcacgcgc cggccagtat
 241 ttgctacatt ccccgagct ggggtgcctca gagggcgctg cggcccgga cgaggtggac
 301 ggccgggggg agctggtaag gaggagcagc ggcggtgcca gcagcagcaa gagccccggg
 5 361 ccggtgaaag tgcgggaaca gctgtgcaag ctgaaaggcg ggggtggtgt agacgagctg
 421 ggctgcagcc gccaacgggc cccttcacgc aaacagggtga atgggggtgca gaagcagaga
 481 cggctagcag ccaacgccag ggagcggcgc aggatgcatg ggctgaacca cgccttcgac
 541 cagctgcgca atgttatccc gtcgttcaac aacgacaaga agctgtccaa atatgagacc
 601 ctgcagatgg cccaaatcta catcaacgcc ttgtccgagc tgctacaaac gccacgcca
 10 661 ggggaacagc caccgccgcc tccagcctcc tgcaaaagcg accaccacca ccttcgcacc
 721 gcggcctcct atgaaggggg cgcgggcaac gcgaccgag ctgggggtca gcaggcttcc
 781 ggagggagcc agcggccgac cccgcccggg agttgccgga ctgccttctc agccccagct
 841 tctgcgggag ggtactcggg gcagctggac gctctgcact tctcgacttt cgaggacagc
 901 gccctgacag cgatgatggc gcaaaagaat ttgtctcctt ctctccccgg gagcatcttg
 15 961 cagccagtgc aggaggaaaa cagcaaaact tcgcctcggg ccacagaaag cgacggggaa
 1021 ttttcccccc attccatta cagtgactcg gatgaggcaa gtttag

POU Class 4 Homeobox 3 (Pou4f3)

The POU4F3 gene encodes POU class 4 homeobox 3, and acts as a transcriptional
 20 activator. POU4F3 activates ATOH1 transcription in early development and is later further
 activated by ATOH1 and required for hair cell survival after birth. POU4F3 activates NT3 and
 BDNF. Mutations in POU4F3 have been associated with hearing loss (Lee et al. (2010) *Biochem
 Biophys Res Commun* 396(3):626-630; Clough et al. (2004) *Biochem Biophys Res Commun*
 324(1):372-381; Costa et al. (2015) *Development* 142(11):1948-1959; and Walters et al. (2017)
 25 *Cell Rep* 19(2):307-320).

The human POU4F3 gene is located on chromosome 5q32. It contains 2 exons
 encompassing ~15 kilobases (kb) (NCBI Accession No. NG_011885.1). The full-length
 wildtype POU4F3 protein expressed from the human POU4F3 gene is 338 amino acids in length.

Various mutations in the POU4F3 gene have been associated with hearing loss due to
 30 hair cell degeneration. For example, a nonsense mutation c.337C>T in POU4F3 was identified
 to cause autosomal dominant hearing loss (Zhang et al. (2016) *Neural Plast*
 doi:10.1155/2016/1512831).

Methods of detecting mutations in a gene are well-known in the art. Non-limiting examples of such techniques include: real-time polymerase chain reaction (RT-PCR), PCR, sequencing, Southern blotting, and Northern blotting.

An exemplary human wildtype POU4F3 protein is or includes the sequence of SEQ ID NO: 6. Non-limiting examples of nucleic acid encoding a wildtype POU4F3 protein are or include SEQ ID NO: 9. As can be appreciated in the art, at least some or all of the codons in SEQ ID NO: 9 can be codon-optimized to allow for optimal expression in a non-human primate.

Human Full-length Wildtype POU4F3 Protein (SEQ ID NO: 6)

MMAMNSKQPFGMHPVLQEPKFSSLSHSGSEAMRRVCLPAPQLQGNI FG SFDESL LARAEALAAVDIVSHGKNHPFKPD
 ATYHTMSSVPCTSTSTVPISHPAALTSHPHHAVHQGLEGDLEHISPTLSVSGLGAP EHSVMPAQIHPHHLGAMGH
 LHQAMGMSHPHTVAPHSAMPACLSDVESDPRELEAFAERFKQRRIKLGV TQADVGAALANLKIPGVGSLSQSTICRF
 ESLTSLSHNNMIALKPVLQAWLEEAEEAAYREKNSKPELFNGSERKRKR TSIAAPEKRSLEAYFAIQPRPSSEKIAAIA
 EKLDLKKNVVRVWFCNQRQKQKRMKYS AVH

Mouse Full-Length Wildtype POU4F3 Protein (SEQ ID NO: 7)

MMAMNAKQPFGMHPVLQEPKFSSLSHSGSEAMRRVCLPAPQLQGNI FG SFDESL LARAEALAAVDIVSHGKNHPFKPD
 ATYHTMSSVPCTSTSTPTVPISHPAALTSHPHHAVHQGLEGDLEHISPTLSVSGLGAP EHSVMPAQIHPHHLGAMGH
 LHQAMGMSHPHAVAPHSAMPACLSDVESDPRELEAFAERFKQRRIKLGV TQADVGAALANLKIPGVGSLSQSTICRF
 ESLTSLSHNNMIALKPVLQAWLEEAEEAAYREKNSKPELFNGSERKRKR TSIAAPEKRSLEAYFAIQPRPSSEKIAAIA
 EKLDLKKNVVRVWFCNQRQKQKRMKYS AVH

Rat Full-Length Wildtype POU4F3 Protein (SEQ ID NO: 8)

MMAMNAKQPFGMHPVLQEPKFSSLSHSGSEAMRRVCLPAPQLQGNI FG SFDESL LARAEALAAVDIVSHGKNHPFKPD
 ATYHTMSSVPCTSTSTPTVPISHPAALTSHPHHPVHQGLEGDLEHISPTLSVSGLGAP EHSVMPAQIHPHHLGAMGH
 LHQAMGMSHPHAVAPHSAMPACLSDVESDPRELEAFAERFKQRRIKLGV TQADVGAALANLKIPGVGSLSQSTICRF
 ESLTSLSHNNMIALKPVLQAWLEEAEEAAYREKNSKPELFNGSERKRKR TSIAAPEKRSLEAYFAIQPRPSSEKIAAIA
 EKLDLKKNVVRVWFCNQRQKQKRMKYS AVH

Human Wildtype POU4F3 cDNA (SEQ ID NO: 9)

atgatggccatgaactccaagcagcctttcggcatgcacccggtgctgcaagaacccaaattctccagtctgcactc
 tggctccgaggccatgcgccgagctctgtctcccagccccgcagctgcagggtaatatatttggaagctttgatgaga
 gcctgctggcagcgccgaagctctggcgccggtggatatcgtctcccacggcaagaaccatccgttcaagcccgac
 gccacctaccataccatgagcagcgtgcctgcacgtccacttcgtccaccgtgcccacatctcccaccagctgcgct
 cacctcacaccctcaccacgcccgtgcaccagggcctcgaaggcgacctgctggagcacatctcgccacgctgagtg

tgagcggcctgggcgctccggaacactcgggtgatgcccgcacagatccatccacaccacctgggcgccatggggccac
 ctgcaccaggccatgggcatgagtcacccgcacaccgtggccccctcatagcgccatgcctgcatgcctcagcgacgt
 ggagtcagacccgcgcgagctggaagccttcgccgagcgcttcaagcagcgggcgcatcaagctgggggtgacccagg
 cggacgtgggcgcggtcttggttaatctcaagatccccggcggtgggtcgtgagccaaagcaccatctgcaggttc
 5 gagtctctcactctctcgcacaacaacatgatcgtctctcaagccggtgctccaggcctggttgaggaggccgagggc
 cgcctaccgagagaag aacagcaagccagagctcttcaacggcagcgaacggaagcgcaaacgcacgtccatcgcg
 gcgcccggagaagcggttcactcgaggcctatcttgcgtatccagccacgtccttcatctgagaaga†cgcggccatcgc
 tgagaaactggaccttaaaaagaacgtgggtgagagctctggttctgcaaccagagacagaaacagaaacgaatgaagt
 attcggctgtccactga

A non-limiting example of a human wildtype POU4F3 genomic DNA sequence is SEQ
 ID NO: 10. The exons in SEQ ID NO: 10 are: nucleotide positions 1-209 (exon 1) and nucleotide
 positions 525-1497 (exon 2). The intron in SEQ ID NO: 10 is: nucleotide positions 210-524
 (intron 1).

Human Wildtype POU4F3 Gene (SEQ ID NO: 10)

1 cgctgagcag cgctcacttg gagagcggca agcaagctag acaagcctga ttccatgtca
 61 cccgctgccca ccctgccagg agcgcgaaga tgatggccat gaactccaag cagcctttcg
 121 gcatgcaccc ggtgctgcaa gaacccaaat tctccagtct gactctggc tccgaggcca
 20 181 tgcgccgagt ctgtctccca gccccgcagg tacgtagtgg agcataatta ccgctctaag
 241 gcacatTTTT tgacaggcac tagcttcatg ttttttcat gtcgcccaga acaatcgccg
 301 ctgtctgaac ccctctcctt gtctcccccg cgttctctcc cggcgcgctc tctctctcat
 361 tcatgtctct gatccacacg tctgttccag cagagccgct gcctccgtat taatttttat
 421 gacctgggct ttgaggagag gcatctcggg tgcttgaaaa tgtgttttaa tctgtgttg
 25 481 acagtattcc ctactgaccg tgctgtgctc cttctcgctt gcagctgcag ggtaatatat
 541 ttggaagctt tgatgagagc ctgctggcac gcgccgaagc tctggcggcg gtggatatcg
 601 tctcccacgg caagaaccat ccgttcaagc ccgacgccac ctaccatacc atgagcagcg
 661 tgccctgcac gtccacttcg tccaccgtgc ccattctcca ccagctgcg ctcacctcac
 721 accctcacca cgcggtgcac cagggcctcg aaggcgacct gctggagcac atctcgccca
 30 781 cgctgagtgt gagcggcctg ggcgctccgg aacactcggg gatgcccgca cagatccatc
 841 cacaccacct gggcgccatg ggccacctgc accaggccat gggcatgagt caccgcaca
 901 ccgtggcccc tcatagcgcc atgcctgcat gcctcagcga cgtggagtca gacccgcgcg
 961 agctggaagc cttcgccgag cgcttcaagc agcggcgcat caagctgggg gtgacccagg
 1021 cggacgtggg cgcggctctg gctaattctca agatccccgg cgtgggctcg ctgagccaaa
 35 1081 gcaccatctg caggttcgag tctctcactc tctcgcacaa caacatgatc gctctcaagc
 1141 cgggtgtcca ggcttggttg gaggaggccg aggccgcta ccgagagaag aacagcaagc
 1201 cagagctctt caacggcagc gaacggaagc gcaaacgcac gtccatcgcg gcgcccggaga

1261 agcggttcaact cgaggcctat ttcgctatcc agccacgtcc ttcacatctgag aagatcgcg
 1321 ccacgctga gaaactggac cttaaaaaga acgtggtgag agtctgggtc tgcaaccaga
 1381 gacagaaaca gaaacgaatg aagtattcgg ctgtccactg attgcggcag ggcgcagcgt
 1441 cgggagccgg gagagcctag tgctcatccc tcccgggttc gggggatggt tatcggg

5

Catenin beta 1 (CTNNB1)

The CTNNB1 gene encodes catenin beta 1 (β -Catenin), a protein involved both in transcriptional activation and in adherens junctions. CTNNB1 is required for hair cell development and differentiation. β -Catenin activates ATOH1 through binding to its enhancer.

10 Overexpression or stabilization of CTNNB1 results in supporting cell proliferation and differentiation into hair cells (Shi et al. (2013) *Proc Natl Acad Sci U S A.* 110(34):13851-13856; Kuo et al. (2015) *J. Neurosci.* 35(30):10786-10798). Knock-out of CTNNB1 in early development prevents hair cell differentiation (Shi et al. (2013) *J. Neurosci.* 34(19):6470-6479. Overexpression of CTNNB1 induces ectopic hair cells.

15 The human CTNNB1 gene is located on chromosome 3p22. It contains 15 exons encompassing ~41 kilobases (kb) (NCBI Accession No. NG_013302.2). The full-length wildtype CTNNB1 protein expressed from the human CTNNB1 gene is 781 amino acids in length.

20 Methods of detecting mutations in a gene are well-known in the art. Non-limiting examples of such techniques include: real-time polymerase chain reaction (RT-PCR), PCR, sequencing, Southern blotting, and Northern blotting.

An exemplary human wildtype CTNNB1 protein is or includes the sequence of SEQ ID NO: 11. Non-limiting examples of a nucleic acid encoding a wildtype POU4F3 protein is or includes SEQ ID NO: 14. As can be appreciated in the art, at least some or all of the codons in
 25 SEQ ID NO: 14 can be codon-optimized to allow for optimal expression in a non-human primate.

Human Full-length Wildtype CTNNB1 Protein (SEQ ID NO: 11)

30 MATQADLMELDMAMEPDRKAAVSHWQQQSYLDSGIHSGATTTAPSLSGKGNPEEEDVDTSQVLYEWEQGFSQSFTQE
 QVADIDGQYAMTRAQRVRAAMFPETLDEGMQIPSTQFDAAHPTNVQRLAEPQMLKHAVVNLINYQDDAELATRAIP
 ELTKLLNDEQVVVNKAAMVHQLSKKEASRHAIMRSPQMVSAIVRTMQNTNDVETARCTAGTLHNLSHHREGLLAI
 FKSGGIPALVKMLGSPVDSVLFYAITTLHNLLLHQEGAKMAVRLAGGLQKMVALLNKTNVKFLAITTDCLQILAYGN
 QESKLIILASGGPQALVNIMRTYTYEKLWTTSRVLKVLVSCSSNKPAIVEAGGMQALGLHLTDPSQRLVQNCLWTL

RNLSDAATKQEGMEGLLGLTLVQLLGSDDINVVTCAAGILSNLTCNNYKNKMMVCQVGGIEALVRTVLRAGDREDITE
 PAICALRHILTSRHQEAEMAQNAVRLHYGLPVVVKLLHPPSHWPLIKATVGLIRNLALCPANHAPLREQGAI PRLVQL
 LVRAHQDTQRRTSMGGTQQQFVEGVRMEEIVEGCTGALHILARDVHNRI VIRGLNTIPLFVQLLYSPIENIQRVAAG
 VLCELAQDKEAAEAIEAEGATAPLTELLHSRNEGVATYAAAVLFRMSEDKPQDYKKRLSVELTSSLFRTEPMAWNET
 5 ADLGLDIGAQGEPLGYRQDDPSYRSFHSGGYQDALGMDPMMEHEMGGHHPGADYPVDGLPDLGHAQDLMDGLPPGD
 SNQLAWFDTDL

Mouse Full-length Wildtype CTNNB1 Protein (SEQ ID NO: 12)

MATQADLMELDMAMEPDRKAAVSHWQQQSYLDSGIHSGATTAPSLSGKGNPEEEDVDTSQVLYEWEQGFSQSFTQE
 10 QVADIDGQYAMTRAQRVRAAMFPETLDEGMQIPSTQFDAAHPTNVQRLAEP S QMLKHAVVNLIN YQDDAELATRAIP
 ELTKLLNDEDDQVVVNKAAMVHQLSKKEASRHAIMRSPQMVS AIVRTMQNTNDVETARCTAGTLHNLSSHREGLLAI
 FKSGGIPALVKMLGSPVDSVLFYAITTLHNLLLHQEGAKMAVRLAGGLQKMVALLNKTNVKFLAITTDCQLIAYGN
 QESKLIILASGGPQALVNIMRTYTYEKLWTTSRVLKVL SVCSSNKPAIVEAGGMQALGLHLTDPSQRLVQNCLWTL
 RNLSDAATKQEGMEGLLGLTLVQLLGSDDINVVTCAAGILSNLTCNNYKNKMMVCQVGGIEALVRTVLRAGDREDITE
 15 PAICALRHILTSRHQEAEMAQNAVRLHYGLPVVVKLLHPPSHWPLIKATVGLIRNLALCPANHAPLREQGAI PRLVQL
 LVRAHQDTQRRTSMGGTQQQFVEGVRMEEIVEGCTGALHILARDVHNRI VIRGLNTIPLFVQLLYSPIENIQRVAAG
 VLCELAQDKEAAEAIEAEGATAPLTELLHSRNEGVATYAAAVLFRMSEDKPQDYKKRLSVELTSSLFRTEPMAWNET
 ADLGLDIGAQGEALGYRQDDPSYRSFHSGGYQDALGMDPMMEHEMGGHHPGADYPVDGLPDLGHAQDLMDGLPPGD
 SNQLAWFDTDL

Rat Full-length Wildtype CTNNB1 Protein (SEQ ID NO: 13)

MATQADLMELDMAMEPDRKAAVSHWQQQSYLDSGIHSGATTAPSLSGKGNPEEEDVDTSQVLYEWEQGFSQSFTQE
 QVADIDGQYAMTRAQRVRAAMFPETLDEGMQIPSTQFDAAHPTNVQRLAEP S QMLKHAVVNLIN YQDDAELATRAIP
 ELTKLLNDEDDQVVVNKAAMVHQLSKKEASRHAIMRSPQMVS AIVRTMQNTNDVETARCTAGTLHNLSSHREGLLAI
 25 FKSGGIPALVKMLGSPVDSVLFYAITTLHNLLLHQEGAKMAVRLAGGLQKMVALLNKTNVKFLAITTDCQLIAYGN
 QESKLIILASGGPQALVNIMRTYTYEKLWTTSRVLKVL SVCSSNKPAIVEAGGMQALGPHLTDPSQRLVQNCLWTL
 RNLSDAATKQEGMEGLLGLTLVQLLGSDDINVVTCAAGILSNLTCNNYKNKMMVCQVGGIEALVRTVLRAGDREDITE
 PAICALRHILTSRHQEAEMAQNAVRLHYGLPVVVKLLHPPSHWPLIKATVGLIRNLALCPANHAPLREQGAI PRLVQL
 LVRAHQDTQRRTSMGGTQQQFVEGVRMEEIVEGCTGALHILARDVHNRI VIRGLNTIPLFVQLLYSPIENIQRVAAG
 30 VLCELAQDKEAAEAIEAEGATAPLTELLHSRNEGVATYAAAVLFRMSEDKPQDYKKRLSVELTSSLFRTEPMAWNET
 ADLGLDIGAQGEALGYRQDDPSYRSFHSGGYQDALGMDPMMEHEMGGHHPGADYPVDGLPDLGHAQDLMDGLPPGD
 SNQLAWFDTDL

Human Wildtype CTNNB1 cDNA (SEQ ID NO: 14)

35 atggctactcaagctgatttgatggagttggacatggccatggaaccagacagaaaagcggctgttagtcactggca
 gcaacagtccttacctggactctggaatccattctggtgccactaccacagctccttctctgagtggttaaaggcaatc
 ctgaggaagaggatgtggatacctcccaagtcctgtatgagtggaacagggattttctcagtccttactcaagaa

caagtagctgatattgatggacagtatgcaatgactcgagctcagagggtagcagctgctatgttccctgagacatt
 agatgagggcatgcagatcccattctacacagtttgatgctgctcatcccactaatgtccagcgtttggctgaaccat
 cacagatgctgaaacatgcagttgttaaacttgattaactatcaagatgatgcagaacttgccacacgtgcaatccct
 gaactgacaaaactgctaaatgacgaggaccaggtggtggttaataaggctgcagttatggtccatcagctttctaa
 5 aaaggaagcttccagacacgctatcatgcgttctcctcagatgggtgtctgctattgtacgtaccatgcagaatacaa
 atgatgtagaaacagctcgttgtaccgctgggaccttgcataaacctttcccatcatcgtgagggcttactggccatc
 ttttaagtctggaggcattcctgcccctgggtgaaaatgcttgggtcaccagtggttctgtgtgttttatgccattac
 aactctccacaaccttttattacatcaagaaggagctaaaatggcagtgctttagctgggtggctgcagaaaatgg
 ttgcttgcctcaacaaaaacaaatgttaaattcttggctattacgacagactgccttcaaatttttagcttatggcaac
 10 caagaaagcaagctcatcactactggctagtgggtggaccccaagcttttagtaaatataatgaggacctatacttacga
 aaaactactgtggaccacaagcagagtgctgaagggtgctatctgtctgctctagtaataagccggctattgtagaag
 ctggtggaatgcaagctttaggacttcacctgacagatccaagtcaacgtcttgttcagaactgtctttggactctc
 aggaatctttcagatgctgcaactaaacaggaagggtggaagggtctccttgggactcttgttcagcttctgggttc
 agatgatataaatgtggtcacctgtgcagctggaattctttctaacctcacttgcaataattataagaacaagatga
 15 tggctctgccaagtgggtggtatagaggctcttgtgcgtactgtccttcgggctgggtgacaggggaagacatcactgag
 cctgccatctgtgctcttcgtcatctgaccagccgacaccaagaagcagagatggccagaatgcagttcgcttca
 ctatggactaccagttgtggttaagctcttacaccaccatcccactggcctctgataaaggctactgttggattga
 ttcgaaatcttgccctttgtcccgcaaatcatgcacctttgcgtgagcaggggtgccattccacgactagttcagttg
 cttgttcgtgcacatcaggatacccagcgccgtacgtccatgggtgggacacagcagcaatttgtggaggggggtccg
 20 catggaagaaatagttgaagggtgtaccggagcccttcacatcctagctcgggatgttcacaaccgaattgttatca
 gaggactaaataccattccattgtttgtgcagctgctttattctccattgaaaacatccaaagagtagctgcaggg
 gtcctctgtgaacttgctcaggacaaggaagctgcagaagctattgaagctgaggagccacagctcctctgacaga
 gttacttcactctaggaatgaagggtgtggcgacatatgcagctgctgttttgttccgaatgtctgaggacaagccac
 aagattacaagaaacggctttcagttgagctgaccagctctctcttcagaacagagccaatggcttggatgagact
 25 gctgatcttggacttgatattggtgcccaggaggagaacccttggatatcgccaggatgatcctagctatcgttcttt
 tcactctggtggatatggccaggatgccttgggtatggaccccatgatggaacatgagatgggtggccaccaccctg
 gtgctgactatccagttgatgggtgccagatctggggcatgccaggacctcatggatgggtgcctccaggtgac
 agcaatcagctggcctggtttgatactgacctgtaa

30 A non-limiting example of a human wildtype CTNNB1 genomic DNA sequence is SEQ
 ID NO: 15. The exons in SEQ ID NO: 15 are: nucleotide positions 1-220 (exon 1), nucleotide
 positions 24571-24631 (exon 2), nucleotide positions 25076-25303 (exon 3), nucleotide
 positions 25504-25757 (exon 4), nucleotide positions 25884-26122 (exon 5), nucleotide
 positions 26210-26411 (exon 6), nucleotide positions 27758-27902 (exon 7), nucleotide
 35 positions 33891-33994 (exon 8), nucleotide positions 34079-34417 (exon 9), nucleotide
 positions 34689-34847 (exon 10), nucleotide positions 36274-36393 (exon 11), nucleotide

positions 36899-37049 (exon 12), nucleotide positions 37138-37259 (exon 13), nucleotide positions 38566-38626 (exon 14), and nucleotide positions 39684-40998 (exon 15). The introns in SEQ ID NO: 15 are: nucleotide positions 221-24570 (intron 1), nucleotide positions 24632-25075 (intron 2), nucleotide positions 25304-25503 (intron 3), nucleotide positions 25758-24883 (intron 4), nucleotide positions 26123-26209 (intron 5), nucleotide positions 26412-27757 (intron 6), nucleotide positions 27903-33890 (intron 7), nucleotide positions 33995-34078 (intron 8), nucleotide positions 34418-34688 (intron 9), nucleotide positions 34848-36273 (intron 10), nucleotide positions 36394-36898 (intron 11), nucleotide positions 37050-37137 (intron 12), nucleotide position 37260-38565 (intron 13), and nucleotide position 38627-39683 (intron 14).

Human Wildtype CTNNB1 Gene (SEQ ID NO: 15)

	1	aggatacagc	ggcttctgcg	cgacttataa	gagctccttg	tgcggcgcca	ttttaagcct
15	61	ctcgggtctgt	ggcagcagcg	ttggcccggc	ccccgggagcg	gagagcgagg	ggaggcgagg
	121	acggaggaag	gtctgaggag	cagcttcagt	ccccgccgag	ccgccaccgc	aggtcgagga
	181	cggtcggact	cccgcggcgg	gaggagcctg	ttcccctgag	gtgcttgggc	gctcctttcc
	241	ttatccttcc	ggggctgctc	ccgcttcctc	tcggagccaa	acttcgtagc	aggcgcgcg
	301	tccgggcggc	gggctgggcg	cagccgggag	gcctgggggt	gggagcgggg	agctcagggtg
	361	ggggacgggtg	aggggtgggcc	gcgcccgggg	cgcgaggggc	ggcgccgggg	cccgggttcc
20	421	ggtcgcgctg	cctctctggg	gccctggggg	catcgcttgc	ggggaggggg	cgccgcgggg
	481	gcgcgtacag	gagcccggat	ggcaggcggg	gtgggggttg	gggtgggggt	ctgtggtttc
	541	cgtccggggc	tctggccttg	gccgagtttg	ggggagggac	ccggtgcctc	gggatgcgcc
	601	gggccctggg	tggggggcgg	ggtggggacg	gggggctccg	ccttctcagc	tcttgccggc
	661	agttgggggtt	cgggcgctga	ggcagagacg	ccaccctaag	tcccatcagt	cctggggatc
25	721	ggaccagtgg	actttctctt	aagatttcct	ctttcattct	taagaataga	agtgttatta
	781	tttttttttaa	tgccctggct	atgtgagttt	gaatcgaagc	aactttaaac	cttagagcaa
	841	ctaaactcta	agtgcagcgg	gtgcgatgcg	tcagtagggg	gagcacataa	aaaatccatg
	901	tcttgacac	gtatttttagc	gtactatgca	ggtgagtga	agcagtggtg	aatgtactgg
	961	gagtccttatg	gattttatggt	agtgggtatg	agaccctggt	gaaataaggg	ggtggaggaa
30	1021	ggcgaagggtg	atggcttact	gtttcttacc	aagtgaactg	caggattcag	cctctgactc
	1081	agaccgcttc	gagaattttt	ttcgtagaaa	taatttaaat	ttattcaaat	agtttgatgg
	1141	cagctaaaat	tgaattatag	agcacgtttt	cttttcagcg	gagtgaattt	ttccttcgct
	1201	ccaaagctgg	ccaaatggaa	ttcaagcatt	gcaacttctt	tcagtgtttt	gtctggagag
	1261	aggactttga	accgagactt	ttcgaagtta	agttcctata	gcctgcttct	gaatctgcca
35	1321	agcttgaaag	ctttggcagt	tgggtgtatg	tagttgttgc	cttcgttctc	ttcccctttg
	1381	gagggagcgt	tgtctcctac	tttgtatctt	ccagacatct	gtggtcttcc	ccccaccctt
	1441	cgagtttgtg	agtgggtgaat	gaagaaagac	taggctgctg	gtatgcagag	gtcggcaaaa
	1501	ggaaatcgag	gagtggtttt	agtgaatga	gagctttgta	tcataaataa	tgggtggctta
	1561	ggctagacat	caacttgaag	agacggcagc	atttcctttc	ataaagtcta	ggctaattgt
40	1621	tttcagatcg	ctaagttgta	gtttgtctgg	aatttaggaa	gccatttcag	tatttgtcac
	1681	ttgggtgaacg	aacattcaat	accttcagat	gtcttcgtgt	tgacttgtat	tcatacctaag
	1741	aaatagtaaa	tatagtctca	agtgttattt	atgttatact	gctggtttat	tctctgctta
	1801	aattattgac	ataaatttct	actttggagg	cttttcggtt	gaactaaggc	tgtgcggaat
	1861	ttatttttact	tttatattta	aatctttgaa	aaatctctga	ttaaaaaaa	agtaccctta
45	1921	aaggtttgag	gatgtccttt	cacaccagac	aaaatttggt	taatttgcgc	ccaatattca
	1981	ttactttgac	ctaacccttg	ttctgaaggc	cgtgtacaag	gacaaggccc	tgagattatt
	2041	gcaacagtaa	cttgaaaaac	tttcagaagt	ctattctgta	ggattaaagg	aatgctgaga

	2101	ctattcaagt	ttgaagtcct	gggggtgggg	aaaaataaaa	aacctgtgct	agaaagctta
	2161	gtatagcatg	taactttaga	gtcctgtgga	gtcctgagtc	tcccacagac	cagaacagtc
	2221	atttaaaagt	tttcaggaaa	aaccaactta	aaaaaaaata	aggtggctaa	ttaaaaaaaa
5	2281	atgaagcatt	taacagtggt	cagggttcag	agtatggaag	aggggttttt	taaactgtta
	2341	tctgattatt	tctttttacca	acatgatata	gaaaagtgtg	tttccagtat	taaaatttat
	2401	cagactgagc	ttactgttcc	tgtaaatgac	tggaataaaa	attggcataa	atgagggctc
	2461	gtatgcttgt	tttaataaca	ccaccaccaa	gatagaaaac	gaggaggcaa	gtttctccaa
	2521	gggtattttg	aaatgtgtta	gcaaaactat	tcagataact	cgtttttgtt	ataggggtgag
10	2581	gtggggagag	gcgcagtgta	agtattgttg	aaactaggga	tgtagagaat	taaaagtttg
	2641	aatataatta	ttttgtagtt	ataagtagca	gtgaaattaa	atctcctgca	atagactata
	2701	gaagtatat	tagccaaatg	aaacttcagt	gttattgaaa	tgaaataata	catctgtcct
	2761	gttacaagat	tattttttatt	tctcctgtgg	tttcctagct	tctgataatc	aataattgta
	2821	gatgagtagg	tggttaagttt	taagtttgta	ctttgagctt	agtcggaagc	atgcttgact
	2881	gccaaaccgg	ggcacaaggg	atgaaggctt	ttagaactgg	acaaacttct	aacaaaaggt
15	2941	atttgcaact	cttttgtagt	gtgtcatggt	gatttgtgac	attgtttttg	aaaatatgtg
	3001	ttaacttagt	tttcttgtag	ccctcttttt	attggaactg	tggtatctat	tggtgaaact
	3061	gcttgactga	gaacattttt	ataccataaa	agtaaatagt	aaacatagcc	caggagcggc
	3121	ttctggtttg	tccatcgtat	gtagccattg	cctccttgta	ctctcattga	gaagatactg
	3181	atttgcagat	tcagttgtcc	ttctctaaca	gactatttat	gtaatatgtc	agttgtgatt
20	3241	gtgataggta	agtggaaccg	tcgggttaaaa	taaataactca	ggtttcacaa	aaggaaaata
	3301	atatgatttg	tggtgatcta	aatgagtata	ggagttaact	cctatagtgt	ttcatcactt
	3361	aaactcaggg	gaaagtctct	tatttcctct	gtttacttaa	gaatgctgct	tttgtgtttc
	3421	atgcaagact	gagcttgact	cagtttgaaa	cctaggctca	tctgttgagg	cctgaaccct
	3481	gctgtccttg	aagtatgcat	ataatttgct	tccttcctaa	ggaaaaataa	gctcttgaaa
25	3541	gataaagtca	atcacattag	gaacccattt	ttagggttta	gccacttttt	tttttttttt
	3601	tttttaactc	atgggcatct	cttctgttaa	gagacattcc	ccactctcca	agtttccttc
	3661	aagcctgaag	cagcagagtg	agtagtggtg	gagcatgttt	tcattgcatg	cttgggtcat
	3721	gttgagtgcc	ctccagtggg	tatagtataa	tgcttgatga	tttttttttt	ttaattccaa
30	3781	acaagtttat	gtgggatata	tttaggaata	gttctgatga	gggagaatca	actaagaaac
	3841	cttttgatttc	taaaataatt	aatatcatta	ctgctaatta	aaatacaggc	ttgagaaaat
	3901	gtcttctcag	ccaatatattg	cagtagaaaa	gtcggggagg	tttttaaggt	cactttgagt
	3961	aggcagttct	gcttaaatat	atcataatga	taaaccagaa	tctcagtata	gtactttagg
	4021	aggtaaaaga	tcataaatatt	cagttatatt	gatgaattac	agcaactgaa	attctcagaa
	4081	aaaaattaat	gaaaatgtga	attgtcaatt	tgtctaaaaa	cattcacaga	gtaaaacata
35	4141	agtgtcaca	ttgattatat	taggaaaatg	atagaaataa	aggttaattg	gccagtgtat
	4201	gtgacctaaa	atataatgcc	cttagtgacc	ataggggttg	tctcatttgt	acatagtggg
	4261	gggccatgat	gaactgtgtt	ttgccctttg	aatttttcct	taaaaagctt	tctctaggct
	4321	cctatgttca	tggtttttct	gttagtaata	ttattttctg	aaaatccatg	tttcaattca
	4381	gaatcctaatt	agcaacagga	atgaagctta	ttctaaatta	gtttttggaa	gttaaaccgt
40	4441	cagcatatgg	aaatttttca	gggttttagat	ttttaaaaaa	ttgtttttca	gaatatgttg
	4501	ctggaatgaa	aacgttagcg	tagggacgga	aaatgacact	taccagtgat	tgctttactt
	4561	tgctgtgga	attcagtgtg	atttttgtgga	aacattggta	tatgattttt	tactacttaa
	4621	gaaatgtatt	gctatagtta	gggttttttt	ttttttaaa	gcaagaatgc	ctcaagtgtc
	4681	ttatgtgaat	gattatttca	ggatggatta	aatattcctc	catcaaggac	cataacttga
45	4741	aatcagtgat	ttccaagttg	gtgcttagta	tttacagcat	ttactgtcta	taagcttctg
	4801	ttctgatttt	tcaagagttt	tctgagaaat	gagagtaggc	ttaaaagttc	tttgaaaaat
	4861	tatgtacata	caacttactg	aaaaaaattg	ctaccgggga	cttaatttgt	ctcttgaaat
	4921	gggctacttg	ccttcattaa	tgtagcatat	tacaatttga	tgttcaagat	atgttactaa
	4981	gaataagatc	gctttcagaa	gccttatata	ggattgggtc	tactacattg	tagtgggaat
50	5041	ggctactcaa	atgtctccag	ggccagttag	gtattgggta	aatgggacca	tgacagactat
	5101	taaaaattga	agtgcacatg	aagcagccag	tcataagcag	ctccagccac	tggtgtggaa
	5161	tatagtttat	gttgccagat	catctgattt	ctttccccta	agtgggaaat	ccagatcaat
	5221	gtacatctct	tgatttgcaa	gtgttggtga	acaaaattca	tattttaaga	tgctgtattc
	5281	agcacaatt	aaatacactt	atttgctgaa	tactgccagt	ttgtccctct	gcagtagtac
55	5341	catttgaggt	acagtgtttt	cataatgatt	ctgtgaaatg	actggttctg	tgaatgtaca
	5401	taatttagca	gataacattg	ttaaattatt	aggtttgtat	ttatttaggc	acttgggaaa
	5461	tgccctgtgt	caattgatta	tagattagga	gcttaaaagc	aagatttata	ttatcaactt

	5521	atttgtgaag	actgggaaac	ccacattttt	aaagttagga	attaagatgg	ccaggttcaa
	5581	ggaaaagggg	gagaagtaac	tttcttatta	ctcaaccatc	ttaaataagag	ttctttaagt
	5641	gtatttttaa	gaggtctcaa	aacttaatct	gaagggacgt	caaatgctgg	acaaattctg
5	5701	tgtatacaac	tcaagtcagc	ccccaatfff	actgggtctt	aaatcatgtc	ctttttacca
	5761	gaagtttgca	tttctaagct	aaactattac	tgtagacta	gatccaaaac	ttaaaaacag
	5821	tttaggtaat	taaaaattaa	ttgaatataa	acgttttact	taaattaatg	gcaaattggct
	5881	ttttggccaa	tttaagttta	tgtaggcagt	taaatcgatt	ttggttaaat	cttttgctgc
	5941	taacaaggta	tttccagatt	ttgaaaagtg	gggtggcctg	gtgcctgtag	taccagcact
10	6001	ttgggaggct	ggggagggtg	gatcacctga	ggtcaggagt	tcgaactag	cctggccgac
	6061	gtgggtgaata	caaaaatttag	ccaggcatgg	tggcagggtg	ctgtgatccc	agctgcttgg
	6121	aagtctgaag	catgagaatt	gcttgaacct	gggaagcgga	ggttgacgtg	agctgagatc
	6181	acgccactgc	actccagctg	gggcaacaga	gcgagactcc	atctcaagaa	agaaaagtgg
	6241	ggtgttttagt	cttcaaactc	cgtgtttaag	tgactggagt	gaaaatgtaa	atcataggcc
15	6301	ggtgttggtt	taaaaagcat	catctgaaaa	taatgctgta	gtctgcaatt	atttttatta
	6361	cgatacgatg	gtgtaaaaata	caagcagatc	agtgaaccat	tcatgaaaca	ttaatcctaa
	6421	aggcgtctca	ccccaaagtct	atcccacaat	ctccatgaga	cttcgtggaa	ccactgtaaa
	6481	gtttcttgtg	taatatccca	gaagtttctt	acctctggta	tcttttgaac	ttgttgaaaa
	6541	ggcttttcca	ccccctcttt	atgatggttt	gaagagtgtg	aacatctgaa	tgatgctggg
20	6601	gtgaaaactgc	ttcataaacac	ttccattttc	tcccctattt	atttccatat	ttttattttt
	6661	tcactaatat	ccccacggtt	ttacttctgt	tttagtaatt	cacatgttgc	tggactaatt
	6721	cttttttaact	gacttgtaac	agatatgtta	aaccgtttta	aacttggggg	gtatttttaa
	6781	cctacttttaa	gttagttcaa	gttaatcagt	ctacatggca	tataaacctt	atgattaata
	6841	aatctttaat	gctggtagct	gagttggaag	ccaaagacgt	acaaaaaagc	tgaagtgtta
25	6901	ggttttagtgt	gataagcttc	tcttactaac	aggggtttgt	aatagcagaa	atagatatat
	6961	gcataatat	gtgcatatat	atagcatacc	ttattggatg	tccatataaa	aatgtgtgaag
	7021	aagttaaatt	tactgcaaaa	tttcttggga	gtgcaatttg	aagatgatct	taagtgggtga
	7081	tagtagtttg	ctacactggg	ggatagttgt	tgcaaaactgc	tcctaatttt	cctttactgt
	7141	gaagtaaact	gaacagctgt	aatagggatt	aggaactgta	ctccctctct	ctctttttta
30	7201	agtataatta	agtgttttg	gggtaagggg	gtaggagagt	agtgtctttg	aagttttgca
	7261	tatactagat	gaatgccaca	tgtataaggg	aggaacaagg	gattcttgga	aatatttttc
	7321	aatccaagta	actttggagg	cttccaagtg	gagttcattc	ccctgtgtag	gaaagtgtctg
	7381	gggtagaccc	ttaaattcct	ttctgagcca	ttgaaagaat	gtcctcaaac	ttcgcttata
	7441	ctttatagtt	catttagata	caaaagttac	aaactgaatg	ctatttagga	aacgtaatac
35	7501	actgacatac	cgctctttta	atagattata	aatttagtat	atcaattttc	tggcattttg
	7561	ctgaattttta	ttgttttagtt	ttcaagccca	actatcttgt	tactttgtat	atcgtagttg
	7621	tcccccggtg	atcactgttt	cctgcttaat	tgtgctgtcg	tttttctgg	gtcctgattc
	7681	agagtgtcag	cattctgttc	cccatagaat	aagaagaggc	tagaaagttt	acagatgaga
	7741	tatctaggaa	tgccagaaga	tcagggggtca	ccgttgaggc	agagtaatta	attatggtta
40	7801	aaatgggtgt	gctgataagt	gggtgctggg	aaataattaa	aatttgattt	tttagaagaa
	7861	tacttctcat	gcttgaagag	cgccctcatt	atatgctaaa	gggcctcagg	tttttcttta
	7921	ttgccattat	gctgcagatt	ctattacatt	tgtctgaaaa	gatctaagac	agaagggctg
	7981	tttaataacct	tcccttttct	cctgaacttc	ccctctcctc	tccccatca	ggagctaagt
	8041	aggaacccct	tcaccttggt	accatcagat	ttcatcaatg	gtctgtcttt	acaatgaagg
45	8101	aagtagtact	gcattctggg	cagaggccag	tcctgaggca	tgccttttca	aggacattgt
	8161	tacttttagtt	acactggctc	ttctgtttta	actcttatcc	cccagactct	aatcctgttg
	8221	cttttttttg	tccccatctc	ccacettica	tcacttgaaa	tccattcatt	gtaacttctg
	8281	gaactcagtc	gttagaaaaat	cctttatatt	ctcaatcttg	tgaatgttcc	tttctttctt
	8341	attccagctg	taacctagcc	ttctcccaa	gaatgctact	tcccttgacg	ctctctcaag
50	8401	tgggtgaattt	ttcccttctt	gcacacctta	taacactgaa	ctaggagggtg	tgtggactaa
	8461	atgtctgctt	ttgttcctta	ttgtcacttc	ttgaccttta	ttttccaaaa	cttcaagctf
	8521	tgactttcat	gtgatcaaat	tataccaccc	actgcctgtc	tttatttcaa	gcacctgcaa
	8581	accttcctgg	gtcattcaca	tccttctttg	ttcacttcac	tagctcttgg	ctcattgtca
	8641	ctgtctctta	tttctgtcat	aattcttggg	gacatcagta	tctatgtaga	gcaatactag
	8701	tgaagatgtg	gtctggtaac	tgttacctgt	atgaattaag	ataaggagt	atgccagaat
55	8761	ataagtcacc	tgtgtcacta	agtttactgt	ttagcttact	ttttttgtag	caagattttg
	8821	atgaaggacg	caatatgttg	atttacagtc	tggtacaaat	tttgatgtag	aagatgcttc
	8881	caatatcctg	gtctcttagt	tccttgattt	cttctccagt	gatcttattt	tctaccctaa

8941 ctcaactaca tattcccatt gtcatatcct agaataatattt gtctttttatc tgtaactctg
 9001 ctctctttccc ccaatctcat ttcaagcatc ccacttttcta attcctctag taaatacgtc
 9061 agttccaaca gcccatcaat cccattggga cctacagttt atctatccaa gctttttccct
 9121 gttcctcacc ctcactttcta tacagctgaa gtttcatact gaattataat cacttttctcg
 5 9181 tatacacggt taacaatctt gtccctccct ggcttcatgc ccagtgatct cttgtatcta
 9241 tgaccatgtc ctttatcttc tcctctgtca ctggatgaac tgtagccttc caagataagg
 9301 ccaactcagtt catttgtaca gcagattcca tcccctcttg ctctcaagaa tattactgtg
 9361 gtatctctct tttcttgtct ctactggctc tttccatgag caaacatggg attatcccat
 10 9421 tacaaaaaaa attttttctc cgtctctcct tccactcacc acctcagttc ctgcttctct
 9481 tttccgcaaa ataacctga aaaattgctt tatgtactcc cgttttcttt tgaacccctg
 9541 ccagtgacca ccacgttata aatttgtagt tgtcatctca cttaatctgt tagtagtatt
 9601 tggcaccatt gctacagttg cttgaaatgc cttttcattg gtttccaggc caccatgtct
 9661 gttagcagct tttcctctta cttcactagc atttccctct tttgtttttt tggtatcttt
 15 9721 ctgacctctg ttggagtggc tgaaggttta gtccttgaat ctttttttgt tggtcatatt
 9781 tactccagta tcatagcttt atacagatgg tatttacatc tgtttgctaa cgattttccaa
 9841 attggtatcc ttaaactggg atccagctat ttttttgtca gcatttttga tgctaaagaa
 9901 gcttctcaaa ctaaactgac ctcccggttt tccccaaagc tgcatcttag tcttttccga
 9961 aatgcaattc tgtctttcca gttacctagc ttaaaagctt gcagttcttg actcatcttt
 20 10021 ctctcatacc acgtatctga attctctctg caaaaaattg tctgttctcc cttcagaata
 10081 aagtcacgtg tcattttatg atggggatac attcagaaat gcgtcattag gagataatca
 10141 tgggtgtgtg aacatcagag tatacataga caaacctaga tggatatagc tactacacat
 10201 ctaggctata tgggtgtggc aattactatg atgaatactg taggtaattg taacataaag
 10261 gtaggatatt ttatctaaac gtattgaaac atagaaaaag tacagtaaaa aatattggtat
 25 10321 caaaaaataaa aaatggtaca actgtataag gcagttgtga tgaatggagc ttgcaggata
 10381 tggtgctctg ggtgagtcag cgagtgcga ttgaggggaa gtgaaagatg tgggacatca
 10441 ctgtacacta ctgtagactt tataaacact gtacacttgg gctacactac atttttgtaa
 10501 ggttttataaa gacttttttc tataataaac cttaaattac tgcactttt ttactttatg
 10561 aattcttaaa tttttaaacg ttttctactc tgtaataaca cgtagcttaa aacatacatt
 30 10621 gtacagctgt acaaaaattt tctttatctc tttataagct tttttatatt tttaaaatta
 10681 cttttttacct ttttagcttt ttgttgaaaa actaagacat gggccaggcg cgggtggctca
 10741 cgcctgtaat cccagcactt tgggaggctg aggcaggcg atcacgaggt caggagataa
 10801 gagaccatcc tggctaacat ggtgaaaccc cgtctctact aaaaaataca aaaattagcc
 10861 gggcgtggtg gcgggcacct gtagtccgag ctacttggga ggctgaggca ggagaatggc
 35 10921 gtgaaccag gaggcggagt ttgcagttag ccgagatagc gccactgcac tccagtctgg
 10981 gcgacagagc ggaaactccg tctcaaaaaa aaacaaacaa aaaactaaga catgaacaca
 11041 tttagcctagg cctacagagg gtcaggatca tcagtatcac tgtatttcca tctccacatc
 11101 ttgtccttct ggaatgtctt cagaggcagt aaacataaat ggagctgcca cctcctgtga
 11161 taacagtgcc ttctggaata cctcttgaag gacctacctg tggctgtttt atcgttaact
 40 11221 tttttttttt aagaagtaac agaaggagta cactctaattg ataaaaagta tagtaagtac
 11281 ataaacctgt aacaatcatt atcattatca agtgtcatgt actggacata actgtatatg
 11341 ctatactttt tttttttgag atggcatctc actctgtcac ccaggctgga gtgcagtggg
 11401 gcgaggatag ctcactgtaa cctcagactc ctgggctcaa gtgatcctcc tacctcagcc
 11461 tcccaagtag ctgggactac accaggcacc ccacatgcc tggctaatta aaaaaaattt
 45 11521 tttgtagaga cagggtctca ctctgttgcc agggctggcc ttgaattcct ggcatacagt
 11581 aatcctccca ctttggcctc acaaagtgcg aggattacag gtaagagcca ccatgtctgg
 11641 cccactgtac ttttatacaa ctgaagcaca gtaaacctac tgtggtttcg ttacaccag
 11701 catcaccaca aacaccatga gtagaacatt gtgctgcgac gttaacgatg gctacaacat
 11761 cactaggtga taggaatttt tcagctccat tataatctta tgagaccact gttgtatgtg
 50 11821 cagttcatca tccactgaaa tgccttatg tgatgcatgt cttcatatcc aaaaatatta
 11881 atcattttctc actgaagcca tgccatgcca tgccatcttt tgctgtatt attatttttc
 11941 agcttttatt ttagattcag ggtgtacatg tgcaggtttg ttagaaagag tatatcgat
 12001 gatgctgaag tttgggatac agttgaacca gtcaccagc tagtgagcat agtactcaat
 12061 agataacgtt ctaacattac tcctccttcc ctccctgttc ttgtctctgt ctattgtatc
 55 12121 tttatgtcca tgtgtaccaa atgttttagc cattcttgtg agaacatgtg gcatttgatt
 12181 ttgtttctgt gttaatttgc ttacaataaa tagtctccag ctgcatccac attgctacaa
 12241 aggacatgat tttgttcttt tttataggct gcacatatt ccattggtga taggtaccac
 12301 attttcttga tccagtctac cgttcatggg catttgggtt gattgtatct ttgctattat

	12361	ggatggccttt	tgcctatatatt	attggaaag	ccttctaact	ggtgtccctg	cttacaccgt
	12421	tttccccctt	aaatgtgttt	tcaacatggt	agccagagta	acccttttta	taacaataaa
	12481	tcgtgtaact	tttttgttca	gaaacttaca	gggcttacca	tttcattcag	taaaagctca
5	12541	agctcctgta	tagtcagacc	atatccttca	tcacctgtta	cttttctcct	ctgactcttc
	12601	agcctttttg	tttttcctca	aactgatgaa	gccttcatgg	ctgatgtcag	atgttttgcc
	12661	cattgagatc	ttccttggtg	actcagttgc	acttggtcat	atgattttca	tttatttggg
	12721	gtatctaata	ataatctgaa	agttggctac	ttattttttac	ccctttgagg	gtccttgccc
	12781	tgtttttgta	tccctgatag	cgggacagcc	agatatctgg	aacttacagg	tggtcaataa
10	12841	agtttttggtg	aatgaatatt	ctggaatcac	ccaacctttt	ttttccccctc	cacttatttt
	12901	tcttctccct	ttcacggcct	gaaagatgtc	ctatgtatat	ggttccactt	atcactctca
	12961	tcccagtttg	tgatatacta	ttccattata	ttactattat	taatacaatt	ccattgaact
	13021	tgctcttgct	gacttcacca	ctggacctac	atgttggcca	aatggatact	ttataatttt
	13081	agtcttgacc	cctgcctttg	gcacatttct	tacctctagc	acagcactgt	ccagtaatcc
15	13141	acactttctg	agacagtggg	aatgttcagt	atctgtgctg	ttcagttggt	agcaaccagc
	13201	tacccatgcc	tattaaacat	ttgaaatgtg	gctgtgtgac	tagtggcaat	tatgttggag
	13261	agtacagttt	tagaaaactcc	tgtttttctt	acatggcact	acatttagta	tcacaatcta
	13321	attgtgcaag	ccagataggt	aggagtcac	tttattcctg	ttatttaatt	tttctcatct
	13381	actatatcca	gttcacacac	tcaacagcgc	ctgttggttc	tacctcctaa	atatttcttt
20	13441	agtctaacta	ctacttgttc	ctagtgccac	caccatctat	cagctggaat	attgctatag
	13501	ctgccttaca	ggtttccctt	ctttcctggt	ctcttctagt	tttttgaatt	ttagtcagca
	13561	cgagatttta	aaaactcaaa	taagattgtg	ttattcacct	gcttaaaacc	tttcatgact
	13621	ttcagtgta	cgtagaacag	aaaacacttt	tcttaccaaa	ggctagagag	ctctacgtga
	13681	tctggctatt	tttaacgttt	cattgcactc	acccttttcc	tctataatca	aactactctg
25	13741	atctcaaggg	ttagtctctg	aaagatgatc	atgttcttta	atgactttag	gtttttgtgt
	13801	gttattttct	atttctggga	tgtttattct	ctgttcctta	catgctggcc	cttttgcac
	13861	cttcttcagg	tctcagctta	catgttacct	tcaagaagcc	tttgaccact	ctaagtgggc
	13921	ccttccttcc	acttctgctg	tgtaatccca	ctcccttctc	ccacttgta	attagttaca
	13981	tacttttttg	taattgttta	tttggttgct	gtctccctct	caagaatgca	gggaccatgt
30	14041	ctgcattctg	cagtaatcac	tactgcacac	ccagaatcta	ttacagatcc	tggcatgtag
	14101	ctgatgcata	aattttgtt	gaatgaaagt	ctgtacattg	tatttatgct	attgggtattg
	14161	ctatgacctg	aaactaaaag	gagttgtgga	aaagatttct	tatggaacag	aaatatccct
	14221	tttgattaat	atcacaatct	cgtaaattga	gaaaacaaaa	aaatatatac	tactggagca
	14281	ttcatgtata	gttggagatt	atgactcatt	tattgggtgtg	tttttggact	cagaacaaaag
35	14341	atgaggggaat	attccttaaa	gctctgtatt	gaaataacga	aaagcagtca	cattttaata
	14401	atagaagcct	cctagcttac	tctttctgta	atcttctttt	cctaaatgta	agagagcctc
	14461	ataattatga	ggcttattac	tagagtaagg	ctgtcaaagg	cagcaaaatg	tctttctggt
	14521	tggagaata	acataaaact	gacatgtatg	gtgggggaca	gaaggtttca	aaagtttaag
	14581	aatctgtggt	gtcttaacaa	atagatgctt	ctcaaggagc	ttacgctagt	gggtactctg
40	14641	tccagtcagg	gttttttctt	gtttaacttg	ggttcatttc	ctgatggcac	acatgaagtt
	14701	tggatcatat	ggtttgactt	tagctatggt	ccttagctat	ggggagcagc	atcagcgacc
	14761	tgtgacatgt	aaattaaaaa	tacaatgcca	gggcccttcc	ccagccctc	tgatagagaa
	14821	cctcttggtc	atctgtattt	ttagatgttc	caggttagtc	tgattaacac	ccttggttaa
	14881	gaaccattgg	gaggatctga	ttgccagttt	aaggggacct	tcaagcctgt	aggtctttat
45	14941	agttaaaaaa	aaaaaaagat	tttaaaaaatc	atgcatatgt	tgtggctgaa	ttctggttta
	15001	gcacatactg	cttttaaatg	cctgaaatgt	ttttcccaaa	taaattgtct	tgttatagct
	15061	ttcatgtgtg	atgttggtcca	gcttcttggt	ttgaagatac	ttacgggggg	gaacactttg
	15121	tgatttctct	tagtaacata	ttaaccact	taaaaaccct	ttctattaca	ggtcttcaca
	15181	tttaggctta	atgtgcttaa	ttcaaagtga	aaaatacacc	tgcccttggt	ctcagtga
50	15241	gtatgtaata	aataaatgag	gggttggtgaa	actactgcc	accatctgtt	tttttatggc
	15301	ctatgaacta	agaatcggtt	tggatagcta	aaaaaaaaaa	tcaaaaggat	aattattttg
	15361	tgacgtgaaa	attatatgaa	attcaaattt	cagtttctgt	gaatgaagtt	ttaatggaac
	15421	acagccatcc	atgcttatgt	aagtgtgcat	attctctggc	tgttttcact	gcaatagcag
	15481	agttgagtag	ttgtgacaaa	gagtttatgg	cccacaaaac	ctaaaatatt	tactttctga
55	15541	tgctttacag	aaaaagtttc	ctgaacctta	ttctagctat	atgttggtca	taaatgaatc
	15601	tttctgtggt	ctgaaggcat	ttaagaatct	cctaggttat	aaattggctg	ggcgaggtgg
	15661	ctcacgcctg	taatcccagc	actttgggag	gccgaggctg	gtggatcacg	agggcaggag
	15721	ttcaagatca	gcctagccaa	gatggtgaaa	ccctgtctcc	attaaaaaaa	aaaaaaaaaa

	15781	aaaaaaaaata	gctgggggttg	gtgggtgggca	gtaatcccag	ctactcggga	ggctgaggca
	15841	gagaattgct	taaacccagg	aggcggagga	tgcagtgagc	caagatcgcg	ccactgcact
	15901	ccagcctggg	caacagagtg	aaacaccatc	tcaaaaaaaaa	aaaaaaaaaa	aaaacactct
5	15961	taggttataa	ataattgttg	ttagctctcc	aagcctccat	attacatttt	gtgtgttctc
	16021	ctgttcacat	tttgagcatt	ttatttttta	ttagcacatt	cagttcatca	ggtattttaag
	16081	agcttaatat	atgccaaaagc	atatattaag	cgagaagctg	tttctaaatg	tactgtctca
	16141	gccctcacag	agttcacttc	attaggtctc	ttaaaatttc	tttcttttaa	aggtcagcgt
	16201	gctggtatag	tggggaaggg	aaactcttac	aacacgtcga	gtagaggaag	gttatcatta
10	16261	tgggatataa	tttggaagtc	attgagtacc	tgccattaat	tctgcctgta	gtctgaatgt
	16321	agagattaac	atgtagaaac	ttttttgaaa	taaaatcttc	aatttctttg	gcatacttag
	16381	tactgtctag	ctaggcatat	agtcaaagta	tgggtgtatat	ttcaagtatt	aaaagttttt
	16441	ttgggctgta	gtcactgttg	aaaggatata	gttctttact	attacatgtg	atacctttat
	16501	ataaaattgg	ctaaccctcg	tctttcattt	atctgcaaca	ctgactgtta	ccagttgtct
15	16561	ctaactttgg	tatggggggg	ggaaatatga	ttagattgaa	agggtacatg	actgagccac
	16621	aagcagacct	ggatttgaat	tttaactgaa	cgggtttatta	gctattctta	cattaatact
	16681	gctaatacag	tttcttgtga	tatgaggaat	gatgtcttct	ttatgaggtt	gctaggaaga
	16741	ttcaatgaga	taacatacta	ggctcagaac	tgaagttgct	aggaatttaa	ttatgtctacc
	16801	ttgttaaagt	atgtcaaagg	cagaattcag	tgtttagctg	ataccacaag	gcagtatcct
20	16861	aaaattatgc	tgtaaaagat	ataaagatgc	tgtaaagtgc	tcagaaacct	tgtagactttg
	16921	taatgcagtt	gattcttaga	atactgtcac	tttaacagaa	taggagctag	gaatgaagaa
	16981	atagtattta	aattactaaa	atagaaaatt	tattgacaca	tgtaaagtga	catttgctta
	17041	aatattgaaa	aatttgtagt	actatttcct	tgcttttagaa	aacattgggt	accacttttt
	17101	ttatttatag	cagtttggtt	ttgccttgag	gcaagatggg	tgactgagta	gttgccacat
25	17161	ttcttttgta	caaagtccat	ttcataggcc	atctagcttt	tatgcttaga	aacatttcct
	17221	taacgttata	tttcagtatt	tggctaacct	atataggggt	aaattatata	ggctaacttc
	17281	tcggacagat	atttctaata	atttatgtat	ttggttctgc	aaatgtatgc	aaaaatatat
	17341	gtacaaaagg	atgcagatgc	cttgcatact	tgatatatgt	taaatttttt	ttaatgtaga
	17401	cctttttcgt	tctctttaat	gactatatgg	tattccacca	tccccgctc	acctggacaa
30	17461	ctacagtaac	ctcctaaatg	gtgtttctac	tttgctattg	ccccttattg	tcttttttcc
	17521	cctttatagc	tgctggagtg	aatttttagaa	agcctaagtc	atacatcaca	ttgcttcatg
	17581	ggcatcccag	tacacttttg	attttatttt	acatccctac	tgatctgatt	ctcatctctg
	17641	tctcttcatg	gttctctgcc	ttctagttac	actggtgacc	tttcaaaacc	tttaccacat
	17701	tgagttcatt	ccttactttt	cactctttct	ctgcctggag	tgttctgccc	catctttacg
35	17761	tggccagctg	ctcctcctct	gatgaaatgt	ctcttcctca	caggccttcc	ctgaccaccc
	17821	actagagtag	cacatcttct	acctcataaa	cttgtttatt	agtatttctt	actctaaatt
	17881	ttctttttaa	ttgcttaatt	ccctaacagt	agaatataag	cttcactgta	tgatgatctt
	17941	tgttgactct	cttactcatt	gttattgtaa	taccagtaac	aaagggtggt	taaaatttgt
	18001	tcagtgggtg	aatatatggt	ccatttaatg	gataaattat	ttttattcca	gtctcgtgtt
40	18061	gatggacatt	tgaataattt	ccatcttttt	ctctatgaat	gcctcacttg	gcagtcttct
	18121	gacagtattg	ccacagaata	catttctggt	ataaaaaattg	aatttttaag	tcaaagggtta
	18181	gttacacttt	aatggatagt	ggcagcttac	tatcaaaagt	ttctgctagt	ttcaccatat
	18241	ccttatttagc	agtagatatt	atcaatcttt	tcaatctttg	ccaatctgat	aagcaaaaag
	18301	taaatgggtt	taaacatcct	ttgtatatat	tcattgctca	ctttatgttt	ttcctttgaa
45	18361	atgttatttc	ttgttctttc	cctgcagtat	gattctttct	ttttttgact	tgttcccagt
	18421	tttttggtga	ctatggatat	tagcctttta	ttatgtttacg	gatgttctag	tatgttattt
	18481	tttgaattac	ttcaaatgtg	atttgttgct	cagattttta	aaactacata	cacaaattat
	18541	ctcatgtttc	ccttttttgt	ttcaatttcg	actcatgctt	aatcagttca	tcgattgggc
	18601	atggttttat	tcttaatata	taccggtatt	ttatctcatt	ttattttttt	acgtgtaaat
	18661	atttgggtgaa	tataggttta	atttttaatt	aaaataagga	tgaaaaatga	tagttggaat
50	18721	tacaagccca	tttctcctaa	tacttttaat	caagtaatcc	actaattgaa	atattacctt
	18781	cttcattttat	gaaattgcca	cattatatct	gggtgttttt	ctgcctacta	cagtctctta
	18841	cccatttctt	tcctaataat	acaatacttg	aattgctgtg	gttggtgatt	tataatgtta
	18901	tcttaatgat	aacattataa	atgtgatgga	actgggtcct	ccttatagtt	cttcttaaat
	18961	caagaacaag	acatatcttc	ccatttactc	tcgtatgtat	ctcattttac	tgttatgaat
55	19021	gaaatctgtc	ctatttgtgt	ataggaaaat	agtttttgta	tgtaattgtg	atatggccag
	19081	ttttattaaa	aatttgggtta	aactaagagt	tgttttctgt	tcagccttat	catactataa
	19141	aatccacata	aaatgggtat	aaaagtgtcg	caggacactg	ggctcagatg	attctccac

	19201	ctcagcttcc	caagtagctg	ggactacagc	ggcatatgcc	accacaccca	gccaattttt
	19261	aaataagttt	taaaaatagt	atTTTTtagt	gagacaggg	ttcaccatgt	tgcccaggct
	19321	ggtcttgaac	tcctggactc	agacaatcca	cctgccttgg	cttcccaaag	tgttgggatt
5	19381	acaggtgtga	gccaccacac	cttgccgaat	tgcagccata	tttaatactt	ttttccatcc
	19441	tattcccttt	gctgccccca	ggcctcctgt	attgatagcc	cgctattaag	aagctagtgt
	19501	atattctttg	catactttta	cttcataaac	tatatgaagc	attgttctgt	tttttaactt
	19561	aattggtata	aaatttatatt	ttggaaattc	agtatatctt	gtgaaaatta	tttagaaaat
	19621	gtgcctctga	gataaagcct	attcaggatg	tatcttaaag	gagatagctg	tgctttaaca
10	19681	ttatcagtct	ttttggctgc	ttatgttaat	ataagttgga	gaaaaacagt	ctgctttttg
	19741	tgataatatg	ttcttggaga	tggagtgaag	gattgtttta	aaacattgtc	tttttttcc
	19801	cctgaagtac	cagtattttat	tttaggatta	tgttactgat	caaagatgct	gtgtggagtt
	19861	actcattggt	gagactaaca	ataaatcaca	catgcaaagg	atgttaccat	aatctaatta
	19921	ttttaaacag	taaaattata	ttctaagaca	tccagttggc	ctatatgtgc	tatatcaatg
15	19981	actatcaagg	ggctttttat	gtatactgta	tacatgtact	tcacaaaaat	ataaaaggat
	20041	gacatcaaaa	atctggcaag	ccaaaagcct	acattacatg	tagcaataaa	ataagcatat
	20101	gaacttattg	gaatttaaaa	ccctgtagga	tgggcgggtg	atggtatgta	tgtagatgt
	20161	gtggacatat	ctattaaaaag	ttgtgtcaga	taacagctgg	tgctgacaag	cccttggtta
	20221	gatggcagca	tgttcaatat	gttctgtgaa	aattatctca	gtttcagtatt	gtttcagtatt
20	20281	gtggagctat	gcatgaaagg	acttaaaaatt	cttaccctta	aactcagtaa	cagtgtttct
	20341	agaacttctg	gtgatattgg	aaattaagag	aattattttat	atgcaaagg	gtttattgca
	20401	gcattgttgg	aataatagac	aaaatgggga	agaacaagct	cagaatggag	gaggtagctt
	20461	atagtataga	catacgatac	aatccagatg	ataatatatt	ataatagtct	tcacaaggaa
	20521	ttttatatatt	ttatttttta	aaatacatag	cagttagttt	aatataccaa	acataccaaa
25	20581	atgtcatcat	ttactgtgtg	gtggactcat	atgatggaga	tgataaataa	aaatattaat
	20641	ttatttgagg	catatatatta	tggctgagga	aggaagacag	ttatgaagaa	cagctcattc
	20701	tggaaacata	ctaatttttc	ccagccataa	agagatttcc	tatttctttt	ttttttccat
	20761	ttaccttctg	tttcctacct	gagaagattt	catacttcta	ataaccattt	gtgtacctat
	20821	ttaaagacag	taccaaaggc	atacatttta	gtgtttggag	gaccaagggt	catttgatgt
30	20881	ttgatgctta	ttgactattc	gaggatgaca	agacaccttg	agaacacaca	caccacaccc
	20941	cacacccaca	ccctcaccca	ccccccccc	ccccctcccc	gaagaaaagct	gtgaaggaag
	21001	aaagcagaaa	agaacctgga	gtgagttgta	acttaaaaatg	ttagtgttgc	atgaagtgtg
	21061	ttaaaacagg	aagattttgag	gaaattgcat	acatttttcta	gatggcaaag	tattactggt
	21121	gacagttaat	gaaaatgcat	atgcatgtgt	tttttagattt	acaaatttta	ctaagaactt
35	21181	tttaaaaatc	cctgaagggtg	tatcaaaaagt	ttatcatgct	tatgaaatag	agtagcactt
	21241	tctaacttta	aaacggggaa	taattctttg	gatcttgatt	attggaaaag	tgaattatga
	21301	attgctagta	taaaactgtg	gttttaaaaat	atgtctgctt	tatatatttt	tgtagcagat
	21361	ttactcctag	ttataataac	tcaaaacttac	tgaaaactaa	ggtaattaag	ataattctgt
	21421	cctgatggga	agaggaaaaa	taacttcagt	gtgaaatcta	ttatatatta	ttgtgtggca
40	21481	gatttctccc	attgactttg	actggagaca	tttatagggt	taaaatcgga	aatagcacgg
	21541	tgaattttga	agtatccttg	tagttggaaa	gagtattatg	ttcatattgc	caaaaaaaag
	21601	atgcatggat	gcattagact	ggatggaaaa	tacatgagaa	gttggctagc	ccctctttg
	21661	tcaaaacatc	acttggtggt	gataaagctg	ttggaaaaca	cagcattcta	atgtagtctg
	21721	tagtttaaat	ataatctgtg	tcttgaaaca	tttagcgtag	tacttataca	aacctagatg
45	21781	gcatagtgtg	ctgcatgcct	agcctatata	gtatagcctg	ttgcttctag	ggtgtaaagc
	21841	tgtatagcgt	gttactatag	gcagttgaaa	cagtgggtatt	tatgtatcct	tttttttttt
	21901	tttaaatctt	tttaagagac	agggtcttgc	tctgttgccc	aggctggatg	cattggtgtg
	21961	atcatagctc	actataacct	tgaactccta	agtgatcctc	tttgccctcag	cctccccagt
	22021	ggctaggact	acaggcacat	actaccacac	ctggctaatt	tttaacattt	ttttgtagag
50	22081	atggaatttc	gctgtgttgt	ccaggctggg	cttggaaactc	ttgtgctgca	gcaatccacc
	22141	cgcctcccaa	agtgttagaa	ttacaagcca	cttcgcctgg	cttgtttacc	taaacataga
	22201	aaagatccag	taaaaatata	gaattaaaat	cttgtggggc	cactgtagca	tatgtagtcc
	22261	atcttgactg	aaatgtcctt	atgcagtgca	tgattgtact	tcataatttt	taagcactcc
	22321	tcctcttga	ttgggtactta	gtggatttta	tcatttttgt	ttcttcataa	ttctttctga
55	22381	aatgtctact	ggttggtgac	ttgatctcct	gaattgatcg	tgatttcttc	tggtgtattt
	22441	tttgtctttg	tcattttttt	gtactctagg	cagttttctc	aatttttagtt	tctattcaac
	22501	tttttgtttt	tattttattct	ctccagtatt	tatggagata	ctaaattgaa	gtgttctggt
	22561	tctctctcca	ccctatccct	agtttcaagt	tttatctcag	tttctatgga	gtcagttttt

	22621	tcgtttgcttt	aaaaaaaaaat	tttcctgaag	tgatttggtaa	gttttggtcta	attgggagca
	22681	ctagaatttgg	gcccttaatg	gttggcaggg	tgtggtggag	gagagacagc	ccttagtcca
	22741	aaggctcagg	ccagaaaaag	aaagaggaag	gcttttccttt	tcctttccgg	agcagggttc
5	22801	tgccctagggt	cttgcttggc	agtctatttg	atcttcttag	cagttaatgc	tcagtttttt
	22861	ggcatatgtg	gatctgcctc	cagagcaggt	acaaggtgag	tgagtctatg	ctgttaccta
	22921	attagatccc	catttctacc	ctttgttttt	acttctctat	ctactgatag	gtttttaccc
	22981	tccttcacct	catagggttg	cagtgaagag	caagatgaat	ttttatttat	gttgcataaa
	23041	ttttaaaagc	taaaaaatat	atatgtaatg	ttgggaagtc	ccagtgtaca	aatggctatt
10	23101	gtaaatttgg	aacatgaact	tgcttttttc	cattgtaaaa	atgaaatcat	tataaattgc
	23161	ggtcaagtta	ctaggctcagc	ccacacagag	tttaccaggt	aatatgcgta	aatgttttgc
	23221	ctttgcatca	acaacaagga	aaaacagtac	tataaaaaaa	tgttcctgga	agccggatgt
	23281	atcaaagcac	ttctgaaata	gctatatagc	ctatagacat	gaccagttgg	tttctgagtc
	23341	tgttgacatt	ggccaaagga	gaagctcagt	gtagaacatg	tttgaggtct	ccttttgcag
	23401	aaatacattg	gaggctggag	tggggaacca	atcttttcaga	aagggtggtga	agtagttaca
15	23461	tagccactct	tttaaagaca	gtcaaaaagat	agaaaactaag	gccaggtggt	ggctcacatc
	23521	tgagatagga	aaatcacttg	aacctgggag	gctggaggtg	cagtgaagccc	agtatgcacc
	23581	tctgcactcc	agcctggttt	ggcaagagac	caaaaactctg	tctcaaaaaa	aaacaaaaa
	23641	tagttcacac	ttaaatatgt	tattccatat	ctttacatac	ccaatatggt	aattttatagt
	23701	tcaagatgaa	cttggttggg	acagattttg	taataaaagga	aatcgtgtta	ttagaaatat
20	23761	ctagaggcca	tgagccctta	aactgttcta	atctgcaagt	agttccctgt	gtgatgcagt
	23821	ttttttcaat	attgcacaat	aaaggcaaaa	tacggacaaa	ttagatgata	agatttatat
	23881	aaatttttta	aatattgatc	aaaatatgta	tccatatttg	taatatattg	atttataata
	23941	aatcattgct	gtaaatttga	acttagaaaa	atcttactaa	taaaggtgct	tttggtgtgc
	24001	aaactttcat	ttgaaaagta	atctttcttt	gtaccaaaaa	atctaaaatt	cgctattcta
25	24061	gtcaccaaaa	tttgctttat	gaaaaataat	ttttgatggc	actatatcag	aaaacaactt
	24121	gttaaagaaa	atgtggagtt	tttaaaatcc	cactgtacct	ctgttatcca	aaggggatct
	24181	gtgaattttt	ctgtgaaaag	ttaaaaaagg	agagaccttt	aggaattcag	agagcagctg
	24241	atcttttgaat	agtgttttcc	cctccctggc	ttttattatt	acaactctgt	gctttttcat
	24301	caccatcctg	aatatctata	attaatatgt	atactattaa	taaaaagaca	tttttggtaa
30	24361	ggaggagttt	tcactgaagt	tcagcagtgta	tgaggctgtg	gttgaggtgt	ctggaggaga
	24421	ccatgaggtc	tgctgtttcac	taacctggta	aaagaggata	tggtgttttt	ttgtgggtgt
	24481	aatagtgcac	tttaacaggt	atcccagtgta	cttaggagta	ttaatcaagc	taaatttaaa
	24541	tcctaattgac	ttttgattaa	cttttttttag	ggtatttgaa	gtataccata	caactgtttt
	24601	gaaaatccag	cgtggacaat	ggctactcaa	ggtttgtgtc	attaaatctt	tagttactga
35	24661	attggggctc	tgcttcgttg	ccattaagcc	agtctggctg	agatccccct	gctttcctct
	24721	ctccctgctt	acttgtcagg	ctaccttttg	ctccattttc	tgctcactcc	tcctaattggc
	24781	ttggtgaaa	agcaaaacag	ccaccagcag	gaatctagtc	tggtgactg	ttctggagc
	24841	ctggatgcag	taccattctt	ccactgattc	agttagtgag	tgtaggtgg	ttccctaagg
	24901	gattagttat	ttcatcactg	agctaaccct	ggctatcatt	ctgcttttct	tggtgtcttt
40	24961	tcagatttga	ctttatttct	aaaaatatgt	caatgggtca	tatcacagat	tctttttttt
	25021	taaatttaaag	taacatttcc	aatctactaa	tgctaatact	gtttcgtatt	tatagctgat
	25081	ttgatggagt	tgacatggc	catggaacca	gacagaaaag	cggctgttag	tcactggcag
	25141	caacagtcct	acctggactc	tggaatccat	tctggtgcca	ctaccacagc	tccttctctg
	25201	agtggtaaag	gcaatcctga	ggaagaggat	gtggatacct	cccaagtcct	gtatgagtgg
45	25261	gaacagggat	tttctcagtc	cttcaactca	gaacaagtag	ctggtaagag	tattattttt
	25321	cattgcctta	ctgaaagtca	gaatgcagtt	ttgagaacta	aaaagttagt	gtataatagt
	25381	ttaaataaaa	tggtgtgggt	aagaaaagag	agtaatagca	atgtcacttt	taccatttag
	25441	gatagcaaat	acttaggtta	atgctgaact	gtggatagt	agtgttgaat	taaccttttc
	25501	cagatattga	tgacagttat	gcaatgactc	gagctcagag	ggtacgagct	gctatgttcc
50	25561	ctgagacatt	agatgagggc	atgcagatcc	catctacaca	gtttgatgct	gctcatccca
	25621	ctaattgtcca	gcgtttggct	gaaccatcac	agatgctgaa	acatgcagtt	gtaaaactga
	25681	ttaaactatca	agatgatgca	gaacttgcca	cacgtgcaat	ccctgaactg	acaaaactgc
	25741	taaatgacga	ggaccaggta	agcaatgaca	tagctagctt	tttagtctgc	tttgaagtaa
	25801	atgctcaagg	ggagtgtttt	cagaatgtct	acccaatacc	agtacttgaa	aactaacgat
55	25861	gtttctgaat	tcctgtatta	cagggtgggtg	ttaataaggc	tgcatgttat	gtccatcagc
	25921	tttctaaaaa	ggaagcttcc	agacacgcta	tcatgcgttc	tcctcagatg	gtgtctgcta
	25981	ttgtacgtac	catgcagaat	acaaatgatg	tagaaaacagc	tcgttggtacc	gctgggacct

	26041	tgcataacct	ttcccatcat	cgtgagggct	tactggccat	ctttaagtct	ggaggcattc
	26101	ctgccctggt	gaaaatgctt	gggtaagaaa	acatgtcaga	atgcttgaag	ctaaaaagta
	26161	gaagagtata	ctcacaatat	ttctgatgag	gcttttttct	tcttcccagt	tcaccagtgg
5	26221	attctgtgtt	gttttatgcc	attacaactc	tccacaacct	tttattacat	caagaaggag
	26281	ctaaaatggc	agtgcgttta	gctgggtggc	tgcagaaaaat	ggttgccctg	ctcaacaaaa
	26341	caaatgttaa	attcttggct	attacgacag	actgccttca	aatttttagct	tatggcaacc
	26401	aagaaagcaa	ggtaagagaa	ttattcttta	tgtgggtttc	atggagcatt	ggacacctcc
	26461	agtgtcatgt	cattccatgc	agtgttctta	accttttttg	caccagggac	cagtttcgtg
10	26521	gaaaacagtt	tttccatgaa	tgggttgtgg	gaatggtttc	tggatgacac	cattccacct
	26581	cagataatca	ggcattagat	tctcataggg	agcgtgcagc	ctagatccct	cgcagtgtga
	26641	gtccacacta	gggtttctac	tcctatgaga	ctctcatggt	gcagttgatc	tgacaggagg
	26701	tagagctcaa	gccaggtaat	gctcgctcac	ctgccactta	cctcctgctg	tgacagcccag
	26761	ttcattttctg	ttcttttaaa	tttttgagtt	tccatatgta	aagcactatg	cgaagtagta
15	26821	gggatatggt	aggcaagctt	ctcttcacac	ttttgttctt	agggtgggatg	tagatggttg
	26881	gaataataac	ctaataattta	atthgtgtag	tgggaagaag	tggggctatg	agggcacata
	26941	acacaagttg	aaactgactc	tttttgaggg	ttcaaggaga	cctcttgagg	gaagtgatag
	27001	ttgagttcag	tgttcaagga	tgagaaggga	ttcactaggt	gaaggttagg	tgagaaaaca
	27061	acatctttga	aacgaaggaa	ggagatggaa	agtttttgga	atttaagaaa	tactaatagt
20	27121	aaggagggaag	aaaggtttga	ggtgaggcta	ttgagataga	cttagcagat	ctcatagggc
	27181	tttgtagagc	atgttttaaa	gcacaatggg	aaatttcagc	agaagcctga	aatgatgaaa
	27241	tttgttttta	gaaaattggg	gcagtgttga	aagggaagat	atacagggaa	tgaaaggaca
	27301	agcatgaatg	atcattttat	ggtatctggt	tttaagggtg	atataattag	gaaaattaaa
	27361	gggccaaatg	atgaggagtt	aagtgccagt	tctgggtcaa	attttcagtg	aatcagtttt
25	27421	gatataactt	tcactcttagg	gcattactct	tgccataccaa	catagtttct	aaattttttt
	27481	cttttggtgt	gatcactgtg	ggaagaagga	aattgggccc	aaactgatac	attgttttga
	27541	ggactgggat	gtctgaattt	gagtggaaatg	ctttaaaagg	acaagttgga	tagggcccca
	27601	gtatgggggt	ctgagtgatg	gggtccagga	atacatttag	gtccaatggc	aagctggctg
	27661	aaattcttgt	ataataaaa	aggttggtaa	tatggctctt	ctcagacatg	tgatcaagat
30	27721	tccttgacta	acaagatata	tatatataatc	tttctagctc	atcatactgg	ctagtgggtg
	27781	accccaagct	ttagtaaata	taatgaggac	ctataacttac	gaaaaactac	tgtggaccac
	27841	aagcagagtg	ctgaagggtgc	tatctgtctg	ctctagtaat	aagccggcta	ttgtagaagc
	27901	tggtaaagtat	atgtatctat	tctgagtctt	gtgtatagca	tctgcagttc	taattagatt
	27961	acttttctta	ggaaaagggtg	gtagaacttt	aactactgaa	aataaatggt	cctattcagt
35	28021	ttgcagccaa	gatttacatt	cagagtacct	gtcatctgga	ttgtagctaa	atatttaagg
	28081	ctagtttagg	tagagttcct	attatccatc	aaaaatgatg	gcataatggt	tgcttaataa
	28141	aatttgtttg	taatttcagt	tttgagtaaa	cctaagattt	gtaacagag	ctgtgaattt
	28201	ataggagaaa	agacaaattc	taatatagta	cagttttatg	taaagttattg	gctttattag
	28261	tagatgtcta	tgagcagttt	ttgttttgtt	tttaactttta	ggttccgggt	aatgtgcagg
40	28321	cttggttatat	aggtaaattg	catgtcacag	gggtttcgtg	tgagatttat	ttgtcacccc
	28381	aggcagtaag	tattgtaccc	aataggtagt	ttttcagttc	tttacctccc	acccgtaagt
	28441	aggccccagt	gtctgttgtt	cccttctttg	tgcccgtgtg	tactcagttg	ttacctccca
	28501	cttataaagt	agaacatgtg	gtatttggtt	ttctattcct	atgttagttt	gcttaggata
	28561	atggcctcca	gctccatcca	tgttgctgag	gaagacatct	tggatatttt	ttatggctgc
45	28621	ttagtattcc	atagtatata	tgtaccacat	tttctttatc	tagtctacca	ttgatgggca
	28681	tttaggttaa	ttccatatct	ttgctattgt	gaataatgct	gcagtgaaca	tatgcatgca
	28741	tgtgtcttta	tggtaaaaag	atttcttttt	ctttgggcat	atacctaata	ataggattgc
	28801	tggattgaat	ggtaattctg	tcaggttttt	tgagaaatca	ccaaattgct	ttccacaatg
	28861	gctgaactaa	tttactttcc	caccagcagt	gtataagcat	tctcttttct	cagcaacctc
50	28921	accagcatct	gtcatttttt	gactttttat	tagtagccat	tctaactggg	gtgagacggg
	28981	atctcattgt	ggttttgatt	tgcattttctc	taatgatcag	tgatgtcgag	cttttcttca
	29041	tatgtttctt	ggccacttgt	atgtcttctt	ttgaaaagtg	tctgttcag	tcctttgccc
	29101	actttttaat	gggttggttc	ttttttgctt	gttaatttaa	gtttattgta	aactctggat
	29161	attagacctt	tgtcagatgc	atagtttgcc	agtactttct	cccatgccag	tactttctcc
55	29221	cattctgtag	gttgtctgtt	tactctgttg	atthcttttg	ctgcgcagaa	gctctttata
	29281	ctgtcccatt	tgtcagtttt	tgtttttgtt	gcaacttctc	ttggcatctt	cgtcatgaaa
	29341	tcttttgccag	gtcttatgtc	cagaatggta	tttcttaggt	tatcttgag	agtttttaca
	29401	gttttaagtt	ttatatttaa	gtctttaatc	cattctgagt	tgatttttgt	acatcatgta

	29461	aggatggggt	gcagtttcaa	tcttggatgt	ggctagccag	ttatcccagc	accattttatt
	29521	gaataggagg	tcctttcccc	attgcttggt	tttgtttact	tgtaggtgt	gcgccctaac
	29581	ttctgggctt	tcttttctgt	tccattgggtc	tctgtgtctg	tttgataacc	agtaccatgc
5	29641	tgtgattgta	accttggtatt	aacagtatag	cttgaagttg	ggtaaagtga	ttcctccagt
	29701	tttgttcttt	ttgcttagga	ttgccttggc	tattcaggct	cttttttggg	ttcatatgaa
	29761	tttttaaata	gttttttttt	aattatgtga	agaatgccat	tgtagtttg	gtaggaatag
	29821	cattgaatct	gtgaattgct	ttgggctgta	tggccatttt	aacaatattg	attcttcctg
	29881	ccatgaaata	gaatgttttt	tcatttggtg	gtgtcatctc	tgatttcttt	gagcagtgtt
10	29941	ttttgtaatt	ctcattgtag	agatctttca	cctccctggg	tagttgtatt	cctagggtatt
	30001	ttattctttt	tgtggctttg	gtaaatggga	ttgcattctt	gatttggctt	gcagcttggg
	30061	tgttgttggg	gtctagaaat	gcttctgact	tttgtacatt	gatttttata	tcctgaaact
	30121	ttgctgaagt	ttattggatc	aaggagcttt	tgggcagaga	ttatgggggt	ttctaggtat
	30181	agaatcatat	tgtttgcaaa	cagacttcct	atttggatgc	attttctttc	tcttgcttga
15	30241	ttatgagcag	tgttttgccc	tgatattctg	tattctcagt	gaatagatgt	cgtctaagta
	30301	tgagaaacaa	tttttttcta	ttctgagtat	ttttaagaag	gcaacttata	tgtggtactt
	30361	tgtatatatt	gtatgttggc	aattggggaa	aagaatagat	ggtttgtact	agggcctctt
	30421	gggttctgtg	tgtgtgtgtg	tgtgtgtgtg	tgtgtgtcat	gaaaacagtt	acttttttagc
	30481	taccaagcat	tttttctcct	ttcagtaacc	cacctaacaa	cattttactca	gaatttcaaa
20	30541	gcaagcttca	aatcagtatt	gaaagaagga	aaaataataa	ggcatttaata	ggaagaaaat
	30601	gttggaata	aagtataggg	ctggcaaacac	ttacttttct	cacttattga	gagtaatttt
	30661	acttgggaat	ttatgagaga	gaaagacatt	atgattgctc	caggtaacta	ctggcagagg
	30721	aaccatagtc	ttggggatag	acaaatgtgg	ctgagttcat	atagaatgag	gggatgggat
	30781	gtaaatctctg	tcagctgttc	cagcagtaac	ctgtaatgta	ggctaaaaat	acagattttg
25	30841	agatttattt	aatcagaatc	cctggagtgt	taatttttat	atcaagatct	catagtgttt
	30901	tatttgaaagt	gacagggagg	tctgtagata	gctggacatg	tatgggactg	gaagcttagg
	30961	aatctttaag	ttcttccagg	ttattcttat	gttcatttgt	ttattctgaa	aatagcatct
	31021	aatgtatttt	aagaaatgga	ataggcacat	agtatacat	gggtaacaca	acagataggg
	31081	tccccgtgct	taattcttag	tcttgtgaag	gtgacaaaaa	tacttaaaaa	tatgtgatcc
30	31141	taaattagaa	tgagtgttat	gggagaaatg	acagcaaata	gtgatgagaa	ttaatgggga
	31201	ggggaattgt	ctagatgaga	gggaaaaggt	ctccttgaat	aggggatgtt	aagtgggact
	31261	gcaggatgag	agggaaccgt	ctcttgtcta	tatgagaagt	gaggggttaa	cgttttccag
	31321	gtagagaaaa	ggaacaccat	gtgctatgtc	ttagaaccag	ggatatccag	tcttttggct
	31381	tccttgggcc	acattggaag	aagaataatt	gtcttgggct	acacacaaaa	tacactaatg
35	31441	atagctgatg	agctaaaaca	aaaaaaaaatt	gcaaaaagaat	ctcataatgt	ttaagaaagt
	31501	ttacgaattt	gtgttgggct	acattcagag	ctgtcctagg	ccatgtggcc	catgggctgc
	31561	aggttggaca	agcttgcctt	agaaggaaaag	agatttgtca	ggcacgggtg	ctcacgcctg
	31621	taattccagc	actttgggag	gctgaggtgg	gcggatcatg	aggtcaggag	atcgagacca
	31681	gcctggctaa	cacagtgaat	ccccatctct	actaaaaata	caaaaaggta	gccgggctgt
40	31741	gtggcaggcg	cctgtagtcc	cagctacttg	ggaggctgag	gcaggagaat	ggtgtgaacc
	31801	cgggagcgcg	agcttgcagt	gagctgagat	agcgccactg	cacttcagcc	tgggcgacag
	31861	agtgagactc	tatctcaaaa	aaaaaaaaaaa	gggaaagaga	ttgtggagat	ccagggtgctg
	31921	aagagaaggt	ctgcataaac	agaacttagt	aatgaggtgg	atggcctggg	atgaggttga
	31981	ggttaggtaa	gcagagccat	aacatgcagg	actttctagg	ttcctataag	atagttacta
45	32041	ctcatggagt	ttattcatgc	tttattccag	ctttggagcc	atagatacag	aatactttgg
	32101	tcagtttggg	aggctaggtg	ggatccaaat	tctaaacggt	tcctcagggt	tatactaaag
	32161	tatttctatt	atcttaaaag	gatgctgaga	cactttcgtg	ggttgtttat	caatagcaaa
	32221	gcatcacagt	ggtgtgttta	aaatattaat	aatagcattg	tatagattaa	cagtttgaat
	32281	gacaaaagc	tagaagacca	gactactgag	atgttacagg	cttttaggaa	tgaatatagt
50	32341	tgctttttag	actcaatagc	aaaggcgaga	tgtctgagat	gcctgaaaga	atcatagaat
	32401	gtaataatat	aggagctaag	ggagcaacca	aaaacggttt	gtggagggga	caacattggt
	32461	accatgaaga	taaatggaac	cctcagaagg	catccttaat	ttttgaacat	aataatttaa
	32521	gaagctgact	taaagtgact	taaaagggtca	gtaggtagct	ggaaatgtat	gatactagaa
	32581	tgcaagagag	gcaggctaga	gatttgggaag	tttccctctt	agtatatagg	ggtaagggca
	32641	gcagggaagg	ggaggttagag	gtgccacaga	gtcatctgta	tgggactttt	ttttttaccc
55	32701	tagaactgct	gaatcagaat	gtgtgtgttt	taaagtctct	gtaggccatt	ctgatggaca
	32761	tctgggggtta	aaatccattc	tcttagagtt	aatagttatg	taaagggagg	gaatgaagtc
	32821	ttaaagaggg	gaaagaaggt	agtcatttca	caaatactga	gcatcctgat	catcagtcct

	32881	acgcagatca	ttctattagt	agctggagct	actatgaaaa	aggaacccaa	cagagggtgat
	32941	ctttgtcttg	tagggaaagt	ggagtaactt	acactatgaa	ggagaagtgc	aggggtaccat
	33001	aagaattaca	gcagatagac	ctcatctgag	gaaataaaac	agacccgaaa	gatgaaggag
	33061	acaaggaaaa	gtatctctta	ctgcattcag	aagtgtattt	agttgaagat	ggatgagcga
5	33121	agttaatcta	ctatgtgggc	attgggcttc	catttatact	cctttgccag	agtaaagtgc
	33181	ccccatttaa	gggtcctaaa	ggatggaaga	ttgtaaacct	tggaacacat	gttttgtagt
	33241	cagtgaattg	tataaagtcc	ctgacagtaa	gtgttttcat	gccgtctttc	tggattgttc
	33301	ttaccccagg	aatttaccta	gcttctttag	gtcttttagtc	agatgtcacc	ttcacagtga
	33361	ggtgacctaa	ttatctattt	aaaatcgag	ccccactcca	ttatttttct	ccatagccct
10	33421	ttaatatcat	ctgacatact	gtatggtttt	agttttattgt	atatttttct	gcctcttcca
	33481	actagatcat	aaattctgag	ggtaggaact	tctgaatatt	tttgttcact	ggtctatctg
	33541	cagctcagaa	caggacctgg	tactgaataa	atatttttga	aatgattgaa	tggatgaaaa
	33601	gaaatgagta	ataagaatat	tacctaaagg	ggacagtgga	gataacaaaag	gctttttcgg
	33661	cttaggaaaag	gaacagtagc	tatttgagag	tttgtcacta	gtgagggtgaa	ctggc aaaagt
15	33721	gaaggaaaact	gagcaacatt	ctagaaaatg	agaggaaaatc	aaatacttag	gtgaaaggaa
	33781	gtaaaactctg	gaaatacaga	aggacacctc	ctaaggctag	aacagatatt	taggattgat
	33841	aggcacttct	agctaattgac	tagggcctta	tatccttttt	aatttttctag	gtggaatgca
	33901	agcttttagga	cttcacctga	cagatccaag	tcaacgtctt	gttcagaact	gtctttggac
	33961	tctcaggaaat	ctttcagatg	ctgcaactaa	acaggtaaat	tctgagttaa	ctggtgccat
20	34021	gggaatagag	tcaagatgag	tatgtgcttg	tactgaccat	ctgtttttat	ctccatagga
	34081	agggatggaa	ggtctccttg	ggactccttg	tcagcttctg	ggttcagatg	atataaatgt
	34141	ggtcacctgt	gcagctggaa	ttcttttctaa	cctcacttgc	aataattata	agaacaagat
	34201	gatggtctgc	caagtgggtg	gtatagaggc	tcttgtgctg	actgtccttc	gggctggtga
	34261	cagggaagac	atcactgagc	ctgccatctg	tgctcttcgt	catctgacca	gccgacacca
25	34321	agaagcagag	atggcccaga	atgcagttcg	ccttcactat	ggactaccag	ttgtgggttaa
	34381	gctcttacac	ccaccatccc	actggcctct	gataaaagta	aattgtcaaa	gtagaattta
	34441	cctttgttgc	agaattgaaa	atgaagcatc	tctagctggt	ggatggctgt	ctaagcatag
	34501	tgatcaataa	gtaggaattg	tattccttag	taagtaggaa	gtatggctgc	gatagggtga
	34561	agatttctgaa	atgtttgtgt	agtcagaact	acttttagtt	gataccaata	gatttagtgt
30	34621	ggtgggaatt	ttagggtaaag	aaaatgattt	tgttgagttg	tatgccagtt	cttccttctg
	34681	tttttcaggc	tactgttggg	ttgattcgaa	atcttgccct	ttgtcccgc	aatcatgcac
	34741	ctttgcgtga	gcagggtgcc	attccacgac	tagttcagtt	gcttgttcgt	gcacatcagg
	34801	atacccagcg	ccgtacgtcc	atgggtggga	cacagcagca	atttgtggta	ggtaaattct
	34861	tacagtgata	cctggctatc	taaaagggaat	gcataaatcc	aaaggatcct	gaacttcttt
35	34921	ctttgggtcat	tggttccccc	catccgtctt	cctgaagagc	taatgacaaa	gtaaataaat
	34981	aaataattac	acatttctat	ggctgcagag	aaaataaggc	atagtgtggc	cccagtgata
	35041	tttccttgga	cacgtccttc	acatggctag	tcttacaag	gttgggttag	gtgtttcata
	35101	aagtggtctc	atttaattta	cacaaaggcc	cacttcctta	ggaagaggtg	gagtcataat
	35161	ttgagatcaa	atctgtgtaa	tttcagagcc	tcttaccctt	gcctcatcat	gcattttgac
40	35221	tataaatatt	tagcagtcgg	ttttattatc	ttttctgtga	gttaaacttt	tttcatggac
	35281	ctaagaatat	tcagaaataa	gtagtagcat	ttctgtactc	ttaaccacaa	aaatctcaac
	35341	ctgaagcttt	gatacaaaag	ttgtgtctta	aaagtagctt	cattaaaagt	atagtctaata
	35401	gacatttctg	atttctcaga	ctttaagacc	ttattaggtt	agtttagaaa	acaaagatgg
	35461	agcctaccag	aacagatggt	aggaatctca	ttttgctggt	tgctttgtgt	atgtactcat
45	35521	attggggcct	tggctttctt	catttattac	tgttggtatt	ggcccatctc	catgagggtga
	35581	cttaatagaa	cgttgagggc	accttttatt	ttaaatctct	tttctaggaa	gaagagagtt
	35641	tttgtgtcct	tgtaaagaatc	aagttattta	taaaagctgc	taaatgtagc	agaataataa
	35701	ccccttttaa	aactcaaatac	cagaaacagg	agaaacagat	ggtacttaca	tattgcaaaa
	35761	gctatcttcc	ttctatacat	gaggctgtca	gctgaatagt	cttgggaagag	tgaggagtga
50	35821	atttttctgc	tggcaactcg	gttagtttta	gcagttgggtg	ctaaaacttg	gcaaagtttt
	35881	caccaaatac	atggaagata	tacaaaaata	gagggggcat	gtaaaagaaa	aacgttgaca
	35941	tagtctgagc	attactttct	catcttctct	ttttatatac	cttttaccba	gaatgattgg
	36001	tgcccttact	gtaggaaaagt	tgtctttggg	attcagcgct	gtatggaagc	tctgttgcac
	36061	tgtgtatggg	ggaggggtgc	tgctttgaat	tagtgctgcc	aggaggcctc	ttttcagtga
55	36121	cattcaagtt	aatggaatcc	ttcttcttcc	ctgaactaat	tgcaagttac	ggggaacttc
	36181	gggtatataa	tgtaaataat	tacagtctaa	taattgttcc	tcaaacttta	cagaggagaa
	36241	tgccctgttt	gttaacctag	tttcttttgg	caggaggggg	tccgcatgga	agaaatagtt

5 36301 gaagggttgta cccggagccct tcacatccta gctcgggatg ttcacaaccg aattgttatac
 36361 agaggactaa ataccattcc attgtttgtg caggtatggt ttaagtgaag tgttctaggt
 36421 tttatgtcca taaaatttcc agattgtaat gactaataac atttcagaaa attagggacc
 36481 ataatagggg taccacacatt taattttatg aaaattccct acattttttg gtcagtaaga
 36541 gaaacattga gacttgagaa gagggaggag atttcacatt tcaactttat ggggtgcctag
 36601 aggggagagc tgacctgggc tgccagaggc agggcataga cccccaacca attctgggtt
 36661 ttccaaatct tagatcagtt agagctgcct ctgaagaaag ggtttatagc taaaaaatat
 36721 tatggaaatc cagtgtcca gagcattaaa caccccaaga cataaaattc agagaatatt
 10 36781 atttactaca gtgtgaatgc ctcttgact ctgaattggg aatgtttgca ccacagtggg
 36841 gggcttgcca tgttttagct ttagatttaa ttaggttttg tttgtgtttt ctcttagct
 36901 gctttattct cccattgaaa acatccaaag agtagctgca ggggtcctct gtgaacttgc
 36961 tcaggacaag gaagctgcag aagctattga agctgaggga gccacagctc ctctgacaga
 37021 gttacttcac tctaggaatg aaggtgtggg taagtaaaaa ggaaccaaaag cctttagcag
 37081 atgtgtacat tgaagtctca gtttttctct aagggccttt ttctccttgt ctcttagcga
 15 37141 catatgcagc tgctgttttg ttccgaatgt ctgaggacaa gccacaagat tacaagaaac
 37201 ggctttcagt tgagctgacc agctctctct tcagaacaga gccaatggct tggaatgagg
 37261 tagggaaatg tgagcagtta tttatctggt agtttcctag agcaggtatg gcagcttgtt
 37321 ctttcctctc aaaacactta gtacacattc atttgcatg atgtttccct ggcttgagta
 37381 tttcttcttt atgctgtcta gcaactgctc tgaggaaaga ctataatata agctttaaag
 20 37441 agtctgttca gaatcattac aaataagttg tgttatttaa aattataatt cataagggag
 37501 aaagatgaaa aatgtttacca gattaaagaa gatttttcaa aaggatgtaa ggaaagaggc
 37561 agtgttaaac actgttaaga ggacagttaa tcagtatttt ttactaaact ttaataaaac
 37621 ttttctattt gaatttctgc tatgaatttt tcttcagcat ttgtcctcag tacaggtggg
 37681 tccttgaaac attgtttcta ataaaactag aacatcctga tattttatcc attctataga
 25 37741 gatcattgat ggtacacaga catacagtgg attatgtttg ttgagtgaat ggaaagagag
 37801 attgttaggt ttacaacgat gcagctcttg agaccggagt ttaagatcag cctgggcaac
 37861 atagtgaac cccatcttta gctgggcatg gagatggatg cctatagtcc tagctactgg
 37921 ggagacgggg gcaggaggat tgcttgaacc caggagttaa cagactgcac tcagtgcag
 37981 agccagactc caacacaaaa aaaaaaaaagca aattaccagt gagtagtgtg
 30 38041 ttacttgggt ttttaatagg catcttatta acatgttcca acttgagccc ttaactttct
 38101 ccacctaccc ccttccacaa acctgttttc actgtcttct ctgtcttagt taatgtcagc
 38161 tttgtctgtc cagctgtcga ggctaaaact tttctttcat ataacacatc ctatcagcag
 38221 ctctgttttg tgggtaggca ttttgccctt tttttttttt ttttttttaa actgctatat
 38281 ctctagcatg tagaacagtg cctggcagca cataataggg gcttaatata atatttgttg
 35 38341 aaagaacaag tcagtgcagta tttttaatgt gaggtgcaaa gagaaaaaaa aatgtatctt
 38401 tgagggtgtgg agttttgaag aacttccatt ttctaagcat ttgtgtaatg ttggagttac
 38461 ttgttccttt tgtaatctga aagtatgctt taaaaaaaat tagtgtactt ttgagaattt
 38521 tcattttgct ttctattctt ccttgctttg tgcatgttta tctagactgc tgatcttggg
 38581 cttgatattg gtgcccaggg agaacccctt ggataatcgcc aggatggatg ggtctcata
 40 38641 tttctcgatt aactccagat caagctaaag ttctaaaact tttatcagaa gagccgggtt
 38701 gctcatctgg gaaaccagtg ttggcagaaa agtagtggct tcaattaaaa gcagttctta
 38761 aattccagtc agcaacagta tctttaatgg agcacaggga attcagagcc acacaatgag
 38821 tagcagtagg attacaccac caacaaatac atgctactgc taggcctctg cagtgcagga
 38881 tggtacaatt tacctggctt tttattctct ttttgccag aggactcata atacctttgt
 45 38941 ctacaagcta cccaagggaag ataggaaaac tcctgtttct aggctcagat ctcggtggg
 39001 tttttacata gttgcattat catcagggtt ttcttgaaaa gctaatttaa atctgggtaa
 39061 tgaacatgga ggatggcata gaccactaac aattataact gtcttacatt tataaccgca
 39121 tctgcttcta cctaattatg aaaccactaa agcgcagatt cttactgtga gaaataacat
 39181 gtcaacccta agataaaaata tgttgagggt tcatggaaat agtgcccttc cttagtactt
 50 39241 ttgtgggtgt cacttggcct ttttgtcaag atagattaca cctgccagac ctcatattg
 39301 tcttaatcct ccttcccatg acttctcact gcctaggtgg tcacacagta gattcctgct
 39361 tcttctctc gggaacccca agtctcttga caggggtaaa tgcagagtgt tcagggttag
 39421 actaatgatg tgactaggcc ctgctggtgt gcctgtctga tggaaataga tgttatttgt
 39481 gtagtctcat ggggtggcctg gcaactgagta attacttggc taaagaaagc tggaggttga
 55 39541 agaggctaga aagcgttgtt ttctgacaag tttgtgctg aactttggat gccctaacct
 39601 cagtgttaac gtctatgtct gcttctctcc tctctctttt gccttccttc ttgcctattt
 39661 tgttgacacc ctgactcttc tagatcctag ctatcgttct tttcactctg gtggatatgg

39721 ccaggatgcc ttgggtatgg accccatgat ggaacatgag atgggtggcc accaccctgg
 39781 tgctgactat ccagttgatg ggctgccaga tctggggcat gccaggacc tcatggatgg
 39841 gctgcctcca ggtgacagca atcagctggc ctggtttgat actgacctgt aaatcatcct
 39901 ttaggtaaga agttttaaaa agccagtttg ggtaaaatac ttttactctg cctacagaac
 5 39961 ttcagaaaga cttggttggg aggggtgggag tgggttaggc tatttgtaaa tctgccacaa
 40021 aaacagggtat atactttgaa aggagatgtc ttggaacatt ggaatgttct cagatttctg
 40081 gttgttatgt gatcatgtgt ggaagttatt aactttaatg ttttttgcca cagcttttgc
 40141 aacttaatac tcaaataagt aacatttgct gttttaaaca ttaatagcag cctttctctc
 40201 tttatacagc tgtattgtct gaacttgcac tgtgattggc ctgtagagtt gctgagaggg
 10 40261 ctcgaggggt gggctggat ctcagaaagt gcctgacaca ctaaccaagc tgagtttctc
 40321 atgggaacaa ttgaagtaaa ctttttgttc tggctccttt tggctcgagga gtaacaatac
 40381 aaatggattt tgggagtgc tcaagaagtg aagaatgcac aagaatggat cacaagatgg
 40441 aatttatcaa accctagcct tgcttgtaaa attttttttt tttttttttt aagaatatct
 40501 gtaatggtac tgactttgct tgctttgaag tagctccttt tttttttttt tttttttttt
 15 40561 tgcagtaact gttttttaag tctctcgtag tgttaagtta tagtgaatac tgctacagca
 40621 atttctaatt ttaagaatt gagtaatggg gtagaacact aattcataat cactctaatt
 40681 aattgtaatc tgaataaagt gtaacaattg tgtagccttt ttgtataaaa tagacaaata
 40741 gaaaatggtc caattagttt cttttttaat atgcttaaaa taagcagggt gatctatttc
 40801 atgtttttga tcaaaaacta tttgggatat gtatgggtag ggtaaatcag taagaggtgt
 20 40861 tatttgaac cttgttttgg acagtttacc agttgccttt tatcccaaag ttgttgtaac
 40921 ctgctgtgat acgagtcttc aagagaaaat gcggttataa aaaatggttc agaattaaac
 40981 ttttaattca ttcgattg

Noggin (Nog)

25 The NOG gene encodes the noggin protein, and is a bone morphogenetic protein 4 (BMP4) inhibitor. Activation of NOG in supporting cells inhibits BMP4 and induces hair cell regeneration (Lewis et al. (2018) *Hear Res.* 364:1-11).

The human NOG gene is located on chromosome 17q22. It contains 1 exon encompassing ~2 kilobases (kb) (NCBI Accession No. NG_011958.1). The full-length wildtype
 30 NOG protein expressed from the human NOG gene is 232 amino acids in length.

Methods of detecting mutations in a gene are well-known in the art. Non-limiting examples of such techniques include: real-time polymerase chain reaction (RT-PCR), PCR, sequencing, Southern blotting, and Northern blotting.

An exemplary human wildtype NOG protein is or includes the sequence of SEQ ID NO:
 35 16. Non-limiting examples of a nucleic acid encoding a wildtype NOG protein is or includes SEQ ID NO: 19. As can be appreciated in the art, at least some or all of the codons in SEQ ID NO: 19 can be codon-optimized to allow for optimal expression in a non-human primate.

Human Full-length Wildtype NOG Protein (SEQ ID NO: 16)

40 MERCPSLGVTLYALVVVLGLRATPAGGQHYLHIRPAPSDNLPLVDLIEHPDPIFDPKEKDLNETLLRSLLGGHYDPG
 FMATSPPEDRPGGGGAAGGAEDLAELDQLLRQPSGAMPSEIKGLEFSEGLAQGKKQRLSKKLRRKLQMWLWSQTF

CPVLYAWNDLGSRFWPRYVKVGSFCFSKRSCSVPEGMVCKPSKSVHLTVLRWRCQRRGGQRCGWIPYPII SECKCS
C

Mouse Full-length Wildtype NOG Protein (SEQ ID NO: 17)

5 MERCPSLGVTLYALVVVLGLRAAPAGGQHYLHIRPAPSDNLPLVDLIEHPDPI FDPKEKDLNETLLRSLLGGHYDPG
FMATSPPEDRPGGGGGPAGGAEDLAELDQLLRQPSGAMPSEIKGLEFSEGLAQGKKQRLSKKLRRKLQMWLWSQTF
CPVLYAWNDLGSRFWPRYVKVGSFCFSKRSCSVPEGMVCKPSKSVHLTVLRWRCQRRGGQRCGWIPYPII SECKCS
C

10 Rat Full-length Wildtype NOG Protein (SEQ ID NO: 18)

MERCPSLGVTLYALVVVLGLRAAPAGGQHYLHIRPAPSDNLPLVDLIEHPDPI FDPKEKDLNETLLRSLLGGHYDPG
FMATSPPEDRPGGGGGPAGGAEDLAELDQLLRQPSGAMPSEIKGLEFSEGLAQGKKQRLSKKLRRKLQMWLWSQTF
CPVLYAWNDLGSRFWPRYVKVGSFCFSKRSCSVPEGMVCKPSKSVHLTVLRWRCQRRGGQRCGWIPYPII SECKCS
C

15

Human Wildtype NOG cDNA (SEQ ID NO: 19)

atggagcgcgtgccccagcctaggggtcaccctctacgccctggtggtggtcctggggctgcgggacacccggccgg
cggccagcactatctccacatccgcccggcaccagcgacaacctgcccctggtggacctcatcgaacacccagacc
ctatctttgaccccaaggaaaggatctgaacgagacgctgctgcgctcgctgctcgggggccactacgacccaggc
20 ttcattggccacctcgcccccgaggaccggcccggcgggggctgcagctggggcgcgaggacctggcgga
gctggaccagctgctgcggcagcgccgtcgggggccatgccgagcgagatcaaagggttagagttctccgagggt
tgcccgagggaagaagcagcgccctaagcaagaagctgcggaggaagttacagatgtggctgtggtcgagacattc
tgccccgtgctgtacgcgtggaacgacctgggcagccgcttttgccgcgctacgtgaagggtgggcagctgcttcag
taagcgctcgtgctccgtgcccagggcatggtgtgcaagccgtccaagtcctgacacctcacggtgctgcgggtggc
25 gctgtcagcgcgcgggggccagcgctgcggctggattcccatccagtaccccatcatttccgagtgcaagtgctcg
tgctag

A non-limiting example of a human wildtype NOG genomic DNA sequence is SEQ ID
NO: 20. The exons in SEQ ID NO: 20 are: nucleotide positions 1-1892 (exon 1).

30

Human Wildtype NOG Gene (SEQ ID NO: 20)

1 aaaccggtgc caacgtgcgc ggacgccgcc gccgccgcc cgcctggagt ccgccgggca
61 gagccggccg cggagcccgg agcaggcgga gggaagtgcc ctagaacca gctcagccag
121 cggcgcttgc acagagcggc cggacgaaga gcagcgagag gaggagggga gagcggtcgc
35 181 tccacgcgcc ctgcgccgcc gccggcccgg gaaggcagcg aggagccggc gcctcccgcg
241 ccccgcggtc gccctggagt aatttcggat gccagccgc ggccgccttc ccagtagac

301 ccgggagagg agttgcggcc aacttgtgtg cttttcttcc gccccggtgg gagccggcgc
 361 tgcgcgaagg gctctcccgg cggctcatgc tgccggccct gcgcctgccc agcctcgggt
 421 gagccgcctc cggagagacg ggggagcgcg gcggcgccgc gggctcggcg tgctctcctc
 481 cggggacgcg ggacgaagca gcagccccgg gcgcgcgcca gaggcattgga gcgctgcccc
 5 541 agcctagggg tcaccctcta cgccttgggt gtggtcctgg ggctgcgggc gacaccggcc
 601 ggcggccagc actatctcca catccgcccg gcaccagcg acaacctgcc cctggtggac
 661 ctcatcgaac acccagaccc tatctttgac cccaaggaaa aggatctgaa cgagacgctg
 721 ctgcgctcgc tgctcggggg ccactacgac ccaggcttca tggccacctc gcccccgag
 781 gaccggccccg gcggggggcg ggggtgcagct gggggcgcgaggaggacctggc ggagctggac
 10 841 cagctgctgc ggcagcggcc gtcgggggccc atgccgagcg agatcaaagg gctagagtgc
 901 tccgaggggt tggcccaggg caagaagcag cgcctaagca agaagctgcg gaggaagtta
 961 cagatgtggc tgtggtcgca gacattctgc cccgtgctgt acgctggaa cgacctgggc
 1021 agccgctttt ggccgcgcta cgtgaagggt ggcagctgct tcagtaagcg ctctgctcc
 1081 gtgcccaggg gcatggtgtg caagccgtcc aagtccgtgc acctcacggt gctgcgggtg
 15 1141 cgctgtcagc ggcgcggggg ccagcgctgc ggctggattc ccatccagta cccatcatt
 1201 tccgagtga agtgctcgtg ctagaactcg ggggccccct gccgcaccc ggacacttga
 1261 tcgatcccca ccgacgcccc ctgcaccgcc tccaaccagt tccaccaccc tctagcgagg
 1321 gttttcaatg aacttttttt tttttttttt tttttttttc tgggctacag agacctagct
 1381 ttctggttcc tgtaatgcac tgtttaactg ttaggaatg tatatgtgtg tgtatatacg
 20 1441 gtccagttt taatttactt attaaaagg cagtattata cgttaaaagt taccggcttc
 1501 tactgtattt ttaaaaaaaaa gtaagcaaaa gaaaaaaaaa agaacagaga aaagagagac
 1561 ttattctggt tgttgctaataatgttaacc tgctatttat attccagtgc ccttcgcatg
 1621 gcgaagcagg ggggaaaagt tttttttttc ttgaagtaca aagagacggg ggaacttttg
 1681 tagaggactt tttaaaagct attttccatt cttcggaag tgttttggtt ttccttggac
 25 1741 ctcgaagaag ctatagagtt caatgttatt ttacagttat tgtaaataata gagaacaaat
 1801 ggaatgacta atcattgtaa attaagagta tctgctattt attctttata atatcccggtg
 1861 tagtaaatga gaaagaagtg cagagcagga tt

Growth factor independent 1 transcriptional repressor (GFI-1)

30 The GFI-1 gene encodes a nuclear zinc finger protein, and acts as a transcriptional repressor. GFI-1 is activated by Atoh1 and Pou4f3 in early development and is required for hair cell survival after birth (Hertzano et al. (2004) *Hum. Mol. Genet.* 13(18):2143-2153; Costa et al. (2015) *Genom Data* 6:77-80).

35 The human GFI-1 gene is located on chromosome 1p22. It contains 7 exons encompassing ~12 kilobases (kb) (NCBI Accession No. NG_007874.1). The full-length wildtype GFI-1 protein expressed from the human GFI-1 gene is 422 amino acids in length.

Methods of detecting mutations in a gene are well-known in the art. Non-limiting examples of such techniques include: real-time polymerase chain reaction (RT-PCR), PCR, sequencing, Southern blotting, and Northern blotting.

An exemplary human wildtype GFI-1 protein is or includes the sequence of SEQ ID NO:

- 5 21. Non-limiting examples of a nucleic acid encoding a wildtype GFI-1 protein is or includes SEQ ID NO: 24. As can be appreciated in the art, at least some or all of the codons in SEQ ID NO: 24 can be codon-optimized to allow for optimal expression in a non-human primate.

Human Full-length Wildtype GFI-1 Protein (SEQ ID NO: 21)

10 MPRSFLVKSKKAHSYHQPRSPGPDYSLRLENVPAPSRADSTSNAGGAKAEPRDRLSPESQLTEAPDRASASPDSC
SVCERSSEFEDFWRPPSPSPASEKSMCPSLDEAQPFPLPKPYSWGLAGSDLRHLVQSYRPCGALERGAGLGLF
CEPAPEPGHPAALYGPKRAAGGAGAGAPGSCSAGAGATAGPGLGLYGDFGSAAAGLYERPTAAAGLLYPERGHGLHA
DKGAGVKVESELLCTRLLLGGGSYKCIKCSKVFSTPHGLEVHVRRSHSGTRPFACEMCGKTFGHAVSLEQHKAVHSQ
ERSFDCKICGKSFKRSSTLSTHLLIHS DTRPYPCQYCGKR FHQKSDMKKHTFIHTGEKPHKCQVCGKAFSQSSNLIT
15 HSRKHTGFKPFGCDLCGKGFQRKVDLRRHRETQHGLK

Mouse Full-length Wildtype GFI-1 Protein (SEQ ID NO: 22)

MPRSFLVKSKKAHSYHQPRSPGPDYSLRLETVPAPGRAEGGAVSAGESKMEPRERLSPDSQLTEAPDRASASPNSCE
GSVCDPCSEFEDFWRPPSPSPASEKSLCRSLDEAQPYTLFPKPYAWSGLAGSDLRHLVQSYRQCSALERSAGLSL
20 FCERGSEPGRPAARYGPEQAAGGAGAGQPGRCGVAGGATSAAGLGLYGDFAPAAAGLYERPSTAAGRLYQDHGHELH
ADKSVGKVESELLCTRLLLGGGSYKCIKCSKVFSTPHGLEVHVRRSHSGTRPFACEMCGKTFGHAVSLEQHKAVHS
QERSFDCKICGKSFKRSSTLSTHLLIHS DTRPYPCQYCGKR FHQKSDMKKHTFIHTGEKPHKCQVCGKAFSQSSNLI
THSRKHTGFKPFGCDLCGKGFQRKVDLRRHRETQHGLK

25 Rat Full-length Wildtype GFI-1 Protein (SEQ ID NO: 23)

MPRSFLVKSKKAHSYHQPRSPGPDYSLRLETVPVPGRADGGAVSAGESKMEPRERLSPESQLTEAPDRASASPNSCE
GSVCDPSSEFEDYWRPPSPSPASEKSLCRSLDEAQPYTLFPKPYAWSGLAGSDLRHLVQSYRQCSALERSAGLSL
FCERGAESGRPAARYGSEQAAGGAGAGQPGSCGAASGATSAGGLGLYGDFAPAAAGL FERPSTAAGRLYQDRGHELH
ADKSVGKVESELLCTRLLLGGGSYKCIKCSKVFSTPHGLEVHVRRSHSGTRPFACEMCGKTFGHAVSLEQHKAVHS
30 QERSFDCKICGKSFKRSSTLSTHLLIHS DTRPYPCQYCGKR FHQKSDMKKHTFIHTGEKPHKCQVCGKAFSQSSNLI
THSRKHTGFKPFGCDLCGKGFQRKVDLRRHRETQHGLK

Human Wildtype GFI-1 cDNA (SEQ ID NO: 24)

atgccgcgctcatttctcgtcaaaagcaagaaggctcacagctaccaccagccgcgctccccaggaccagactattc
35 cctccgtttagagaatgtaccggcgcctagccgagcagacagcacttcaa atgcaggcggggcgaaggcggagcccc

5 gggaccgtttgtccccgaatcgcagctgaccgaagccccagacagagcctccgcatccccagacagctgcgaaggc
 agcgtctgcgaacggagctcggagtttgaggacttctggaggcccccgtcaccctccgctctccagcctcggagaa
 gtcaatgtgccatcgctggacgaagcccagcccttccccctgcctttcaaaccgtactcatggagcggcctggcgg
 gttctgacctgcggcacctggtgcagagctaccgaccgtgtggggccctggagcgtggcgtggcctgggcctcttc
 10 tgcgaacccgccccggagcctggccacccggccgcgtgtacggccccgaagcgggctgccggcggcggggggccgg
 ggcgccaggagctgcagcgcaggggcccgtgccaccgctggccctggcctagggctctacggcgacttcgggtctg
 cggcagccgggctgtatgagaggcccacggcagcggcgggcttgcgtgtaccccgagcgtggccacgggctgcacgca
 gacaaggcgctggcgtcaaggtggagctcgtgtgcaccgcctgctgctgggcggcggtcctacaagtgc
 catcaagtgcagcaaggtgttctccacgccgcagggctcgaggtgcacgtgcgcaggtcccacagcgggtaccagac
 15 cctttgcctgcgagatgtgcggcaagaccttcgggcacgcggtgagcctggagcagcacaagccgtgcactcgcag
 gaacggagcctttgactgtaagatctgtgggaagagcttcaagaggtcatccacactgtccacacacactgcttatcca
 ctacagacactcggccctacccctgtcagtactgtggcaagaggttcaccagaagtcagacatgaagaaacacactt
 tcatccacactggtgagaagcctcacaagtgcaggtgtgcggcaaggcattcagccagagctccaacctcatcacc
 cacagccgcaaacacacaggcttcaagcccttcggctgcgacctctgtgggaagggtttccagaggaagggtggacct
 ccgaaggcaccgggagacgcagcatgggctcaaatga

A non-limiting example of a human wildtype GFI-1 genomic DNA sequence is SEQ ID
 NO: 25. The exons in SEQ ID NO: 25 are: nucleotide positions 1-151 (exon 1), nucleotide
 positions 3291-3504 (exon 2), nucleotide positions 3831-4013 (exon 3), nucleotide positions
 5789-6276 (exon 4), nucleotide positions 6392-6529 (exon 5), nucleotide positions 8124-8289
 20 (exon 6), and nucleotide positions 10670-12116 (exon 7). The introns in SEQ ID NO: 25 are:
 nucleotide positions 152-3290 (intron 1), nucleotide positions 3505-3830 (intron 2), nucleotide
 positions 4014-5788 (intron 3), nucleotide positions 6277-6391 (intron 4), nucleotide positions
 6530-8123 (intron 5), and nucleotide position 8290-10669 (intron 6).

25

Human Wildtype GFI-1 Gene (SEQ ID NO: 25)

1 gaggggtgccc ccaccgggtcc cgcggggcgc ccgcgggacg cgcgccagg gccctctccg
 61 cgggggggtc ggcgctcgcc cacctcttcc aaatttaacc attacctaaa tccgaaggga
 121 aatgagcaaa cctctcgat tgggtgtcaa ggtattttca gcctcgttgg gcgtatttat
 181 ccccaagtgt ttccacaaca agctatttcg gggcctgcgg ggcaggtttc gctctgcgga
 241 cgcctgtggc actcgccggg ctccaggccg gcggcaccgc gggccggtga ttcacgggtcc
 301 cgacccgggg gtggtgcagc cctaggaggc ggcggggtcg ggggtggggg gggcggtga
 361 ccgaggcctg aggggcgggg agggctctcg gagcggggcg cccccaccc ctctctcgcc
 421 agtcaatctg tgcctcaat ctgtggcttc tctcgtcgcg gaagtctccc tggagccaag
 481 aatagttcat tttctttcaa gtcatttcta gtgcctaagt gtccggacct ccaatttccc
 541 ccatcccctg ccgaccaca gggagagaac tgggaggact actaaggggc gcgcgggacg

5 601 ggctggaaaag gccaggcccc ccaccgcctg gccacttgcg caaaggagcg cgcccggccg
 661 ccgacggggg gttgggagca ggtctgggag ggctatgcga gcgactcagt aacgctcagg
 721 aagtgaagct tgtggttttg ggggctgagc tcggaaggag aatttttttt ttttttaagt
 781 cagagagata gagcggctct tcccgaaaagc aagattccgt ttgaaaactc tcctagcgcg
 841 gtgcccgcgc cgtgactccg caggtaggtc cgccgagcct gttctgcgcc tcctgccctg
 901 gtggggggcg ccgcggggac tcgcagagca ctggcactgc gggggcgatc agagggcgcg
 961 ggcggtttcc cacctgctgc ggaccgccgt gcgggagccc agagagctcc ggcagctggg
 10 1021 ttagggcgcg acccgcgagc tgtgactgga aatctggagc tggggcgcg cagcaggcg
 1081 tctggtggtt cggcagggga gccaaatcca ccagggaagg aaacatctgg tggggaggcg
 1141 gcggcagctg cgctcgggag gacgccccct tagcgccctc ggctccctcc ttcctgggcc
 1201 cggacggtga ggagaggcct gagcgcgcg aggggcccgc ccacctccc cgccagctgc
 1261 agcgcggggt tgccctccca cgcgctcggc ccaggccccg gggccctat ctcctccaaa
 1321 ctctgtcgct cccacagga accagcaaag cgggcccggg tgcgagagag gcgtgggctt
 1381 acagcccggg tggggaggcg gcctccgcgt ccgcctggtc tctggtggcg ccagcccaga
 15 1441 ccagctccg gcgctgacag ttaccccgcc ccatctgtc ccgctcccag ccaacgtggg
 1501 tccaagctgc agcgggaccc tcgggacaac gccactccgt ttttctttt ttcgagttt
 1561 cgtggctggt taaagaattg ggtttgggt ttgtggcgtc taattgtacg gacgagaagt
 1621 gcaggaagcg acaaagctct agccctagga gccaccccgg agggaggcg atggcactct
 1681 caccctagg ggcatctctg cgcttgggta gcgggaaact tcgcgggagc cccgcgacac
 20 1741 gtcccaggcg tcttttctcc caggtctatt cccattcctc cgagagaagg gacacaatgg
 1801 ggctggggat ctggagcagg gggcctgcac cctacaggga ccaaggcctg taggactcgt
 1861 ttgagctgag agcgccaacg gacagacgta gactgtgtgg cctgcatctt gcctaggaag
 1921 ccgaggggct ctagtccg cagtggaaac agcgcgaaag cgaggactg caggtcctgc
 1981 cccggcccag agttcccagc accctcgttt ctgaaccagc cgaggccacg gagaactgct
 25 2041 gtactgcagc tcacgtgtga acccggtcac catcgcttc accccgggag gaaggcagat
 2101 tcgtttactc cagaccacct cgactgtggg gtaccgcccc cgagaccggc tggagcctag
 2161 cggcaggcgc agccacgccc tcccgtcg ctcagatttc gacctggtat taggtgaact
 2221 gattgggggt taatgagagc gacgccccg gcagctagtt ccctcccggg cccgggcccg
 2281 acccccgctc tctgctaatt cagcctgcgc gctctggcgt cctgtctttt ttgtctgcta
 30 2341 aattgtcggg gactaccga ctcgggacac ctagcatttc ccagtcaacg ttcgtggatc
 2401 gggctccacc tccctaggac aagatttttt ggtgagcaga acggaaagtg cttttttccc
 2461 gggacctgat tcccaggtt aggtctccat ggtctgggat ggctcgccgc agcctcgacc
 2521 ggtgcccgcc gcagccggga gtccaagggc aaagtcttc ctacgtggg cactagttag
 2581 gggcgagtgg gatcaccag atgcgagttt ctctggggc .gggggttgg cgtctgttgt
 35 2641 tcccctcact ttctgtccc tttgtctcc cactccttt ctctggcctc tgcgtcccc
 2701 aatccctctg ctgctgtcct ccgcgcgcc cacagtttta tcttgtgttc tgttcccccc
 2761 tccccccggt ctttctactc cagttggtag ctggctcttg gaggtcttg ctccttgccc
 2821 cttccgggtc ctgcaccact gggcatcccc ggcccctaaa ccgatccgcg tgtccccgcc
 2881 ctccctcgcc agccgtaaag cacagccagg caggcgatga gtagctgagt tggggtaacc
 40 2941 caccgatgg gaactacagc tctccaggga gtttgattgc cgagcgagc ttcgctagga
 3001 aaggggagga gctggggggc gtgggcaggg aggaggaaa gggcctgaga caggggcccc

3061 ggacaggttt taccgctgag ctgtgtcagt ggcggcggcg gcaacgacgg cgggttcgcg
 3121 ccacctgtcc aagtgccacc tggtaagcgc ggcgagcag ggtcaagccc ctccctcccgt
 3181 gggccctctg cgcgcctccc tggcccgcgc tctccctccg cctgggtgcc cagtccgccg
 3241 caccggagag tactgaccca cgtctccacc tggttttctt tccctctcag gtctcctccg
 5 3301 ggctggggct gagcaagccc tcggagtgc cgtgggtgac agcggctcca gggactcttg
 3361 gggcgagtg gggaaaagtgc cggaccacca tgccgcgctc atttctcgtc aaaagcaaga
 3421 aggtctcacag ctaccaccag ccgcgctccc caggaccaga ctattccctc cgtttagaga
 3481 atgtaccggc gcctagccga gcaggtgcga ggcgcgcgcg ggccaggcgg ggctgctccc
 3541 ccgatgcct actgcacctc ggacacacct tagtccggag ctgggagggg ctgccccaac
 10 3601 gtcccttttg ctgctgtttt tgtttcctac tgtcctgggt cctccgggtt tgtctcctag
 3661 gtgccatggc ctctctgcgc ctgccctcgg atccgagagg gttcccggcc ggggtctggg
 3721 tggagagggg aagacgctcg gctgccctgg tcgggggatt gggggagcct tcagcacctc
 3781 cagactcaac cgggtcccagc ctgagcccct cacctgcctc ctctttgcag acagcacttc
 3841 aaatgcaggc ggggcgaagg cggagccccg ggaccgtttg tccccgaat cgcagctgac
 15 3901 cgaagcccca gacagagcct ccgcatcccc agacagctgc gaaggcagcg tctgcgaacg
 3961 gagctcggag tttgaggact tctggaggcc cccgtcacc tccgcgtctc caggtaggaa
 4021 cccactggga acctcttggg cgggagctgc agggaccgg cagtgcctgg ggggaattgg
 4081 cgcgaccttg ggcgtagaaa tgctaacggg gagttggaga gtctttccgg gagaaggagg
 4141 ctgattcgta ggggaaggag gcatccggct tctctgggac ttggacagct tgcccgtggg
 20 4201 ggctgctgcc tccatcccag gcggcaggac cctagctgct tgctgcttag attcgtttgc
 4261 gcggagctgg ccagtgcagg aaaacaaacc agtcgtttcg actggcggca acgctgacct
 4321 ttcattttct gaccagacct gactgtttta taaagtgcta ggatcctgca atctagacct
 4381 ccaaacctca aacagagaac agggcagaac gggccaggca gaggagctag gcgctgggcg
 4441 gcagggaggg ggcaggacga aaatctcagc ccgcggcttg gtcttcacag gcgcagattg
 25 4501 ggggcctgtt tcatttttctg ttttgccggg ttaacctagc ctccaggggcc tgctctctgg
 4561 gtttcatttc cagcgagcaa tccagcttca ggcaaaactaa gtgaccacac gttgggtggg
 4621 ggcgtctcga gtcccggccg ggggaaggaa tgagcagacc agccggattc tgtcaagggc
 4681 cggttatata cagaatatgt ttgctagttt tagaagatac caccaccgt cccacaatca
 4741 gtgagttgac ttggcgaaaa ccatagctcc agcaagtgtg tctgggagcc ggcggcggga
 30 4801 ggattcttcc tgccagggcg tcaagtggcc agacaaggat tgggcgcgcc ccgaaccctt
 4861 ccgaacgaaa ctccgggtac agcctctcac tgaagtggcc agcctgaact ggagtgttgt
 4921 gcgcacacac acacacacac atttgtaaatt gccgtatgca ctacatgctg ttgggggtcac
 4981 tagtttttagc aaaattcacg tgggtggggg cgtagcaggc cgagaattca gagctgtctc
 5041 cttgcagggt gtggctaaac cttatgagtt atagttattc tctgagaaat tcaggttccc
 35 5101 cgctccatc aaactgtaac aggaatggg agtatttggc tgtcaattta agcccaaaag
 5161 cccctttcct gctgctcctt tgctacgtac actgggcact taacttcgtg aaatcttaat
 5221 accttcgggt ttattcagac agcagccttt cgggtagttc ggggccgcat ttatggacct
 5281 tctccctcct tcctcttggg ttctgggaag aaaaagaatt gaatgggaac atgtaggggc
 5341 tgggagagtg cctgcgctgg tggtggacc cttccgccct tgagtgtgtg gaggggccga
 40 5401 acggccgcca cttctcctt cttaacagct caactacggg catttataga tgcgcccttc
 5461 cctgtaggat ctccaggtgc gcgggtccag ccagaaaaga tcctcggaac gccgagcgcc

5521 tccgctgcac tcgcacagaa ttacgacct cctctcccga ggtcttttca atgatctgtt
 5581 tactgttctg cctcctatag tggcctgcga ggccccaggg cccgggccac gttttaccct
 5641 ggggcgagcc tggcacctgg cgcacgcagt gttctacaag cgctgggtgc cccgcagtcc
 5701 gcgaacacgc cacgctcgca gccgcagccc ggcggcctcc gctctgccgt ctgaagcctg
 5 5761 accggacgct ccccttgtgc ctccacagcc tcggagaagt caatgtgccc atcgctggac
 5821 gaagcccagc ccttccccct gcctttcaaa ccgtactcat ggagcggcct ggcggttct
 5881 gacctgcggc acctggtgca gagctaccga ccgtgtgggg ccctggagcg tggcgctggc
 5941 ctgggcctct tctgcgaacc cgccccggag cctggccacc cggccgcgct gtacggcccg
 6001 aagcgggctg ccggcggcgc gggggccggg gcgccaggga gctgcagcgc aggggcccgt
 10 6061 gccaccgctg gccctggcct agggctctac ggcgacttcg ggtctgcggc agccgggctg
 6121 tatgagaggc ccacggcagc ggcgggcttg ctgtaccccg agcgtggcca cgggctgcac
 6181 gcagacaagg gcgctggcgt caagggtggag tcggagctgc tgtgcacccg cctgctgctg
 6241 ggcggcggct cctacaagtg catcaagtgc agcaagggtga ggctcccag ctcaccacct
 6301 cgcttgcgt gcgcccgtt cccctacccg cgctcgcct gcgccccgcg gccctctca
 15 6361 gcggccttct ctctggcccc acccgctta ggtgttctcc acgccgcacg ggctcgaggt
 6421 gcacgtgcgc aggtcccaca gcggtaccag accctttgcc tgcgagatgt gcggcaagac
 6481 cttcgggcac gcggtgagcc tggagcagca caaagccgtg cactcgcagg taagcgcggg
 6541 gcgcaccgcc gcgcgcggcc ctgctcgggg atcttctgca tctcctcgt gcagcaccag
 6601 ccaactctctg cctggaagtt ttctcctcga ctccccccag ttctctcccc caagccctcc
 20 6661 gctgcgtccc cttgccctgg tgcaggtgtg tagggaaagg aggattgtgg ccggctcagg
 6721 ccttgaggca gccctggatt ttggtgtcac accactgtga gcctcgagag tgtgatcctc
 6781 attgttactt tgggcttgag gtaggtttgt atgcaactgat tcgtgctgct gatatatcag
 6841 acttactagc tctgtttctt tgtgcctatt cttttcacca aatgggtgtc acttaatttg
 6901 cattgacccc tctcgactga aaaggcagga atctcagctc atttagagca tctagtagca
 25 6961 tattcacccc gctattcatt ctttccttcc ttcttttctt ttcttttctt ttttcttttc
 7021 ttttctttca gagtctggct ctgttgccca ggctggaggg cagttgcaca atctcagctc
 7081 actgcaacct cccctgtca ggttaaagtg attctcgtgc ctcagcctcc taagtagctg
 7141 ggattacagg cgcatgccac tacagcacag ctaatttttg tgttcttagt agagacggga
 7201 tttcgccacg ttagccaggc tggtcacgaa ctctggcct caagtgatcc accagcctgg
 30 7261 gcctcccaaa gtgctgggat tacaggcgtg agccaccatg cccaccgcc gctatttatt
 7321 cattcattca ttaataaata tttgttggt aacttccagg tgccaagtac ttaagaatct
 7381 tataacacat caggtccttg acagcatgcc cacatgaaga ttatagttta gctgagagat
 7441 ggagagtaga tgagcaagta aatatgccaa tagctatctc aggagaatgc ctacttacga
 7501 aggctaaaaa gagtattagc ccatctcccc cagcaccac actggctggg gggagggtggc
 35 7561 atctcaagtg actgaggtct aagcctcctg ttgaggaggg tggagaagtg tgtgctaagt
 7621 ggtgtcaaaa aaagcagggt gtggatatgt atttgccatg ggggtgtggaa ggttgtgggt
 7681 gaagaatgtt ttggtagaaa aagtgttgaa gggccaggca cgggtggctca cgcctgtaat
 7741 ccgaatactt tgagaggccg aggtgggcca atcacttgag gccagggtgtt tgagaccagc
 7801 ctgaccaaca tggtgaaacc ccatctctac taaaaataca aaaattagtc aggtgtgatg
 40 7861 gcgtgtgcct gtagtccctg ctacttggga ggctgagaca cgagaattgc ttgaacctgg
 7921 gaggtggagg gtgcagtga ctgagatcgt tccactgcac tcaagcctgg gcgacagagg

	7981	agactgtctc	aaaaaaagaa	agaaaaagtg	ttcaagggat	tttaggggtca	gctgaggggt
	8041	gaggagagca	gcagtctagt	tgactgcagt	aggagttctg	catctctctc	tctctctctc
	8101	tctctctctc	tctctctctg	caggaacgga	gctttgactg	taagatctgt	gggaagagct
	8161	tcaagaggtc	atccacactg	tccacacacc	tgcttatcca	ctcagacact	cggccctacc
5	8221	cctgtcagta	ctgtggcaag	aggttccacc	agaagtcaga	catgaagaaa	cacactttca
	8281	tccacactgg	tgagctaaaa	aggcccttgg	cttgtaggaa	acaccctgag	gccaacatta
	8341	ctcatcttct	ctgatttctg	gccccagtga	gtggtggatg	aggcctttct	gatggagtta
	8401	ttctctgctc	tgtgttaaag	aaaacaaagg	ggtgggttct	ttggttcatt	taccggcata
	8461	attctcccca	gagccacctt	gatttggggg	tgtgtctgaa	aggccactca	gcaggtcagc
10	8521	tcacaggtac	tctatacttg	gaaagaacat	tttcctttag	gtagcagct	gcttcccctg
	8581	ctgcctgctc	tgggtgaaat	atgaagctcc	agggtcctct	tagagagttg	ctctaaagct
	8641	tacctagaga	ttgaggactt	tccctaacca	cctggccttt	tgtgggaggg	actcgtgtgg
	8701	actctccggc	tgcattttca	ggagtctgag	agcttattct	gattgaagag	gaacaaataa
	8761	tggcaaatat	gattaaactc	tctgctaagc	atTTTTTaaa	tgcattattt	cattttatgc
15	8821	tcacaacaac	tctgagaggt	agcgactact	ccttctcccc	atTTTtagaga	tgaaaaaaat
	8881	gaggctaggt	aatctgcccc	gggtcacact	gctagcaa	gacagggcca	gagctcaa
	8941	tcaggtctga	cctctcaaat	gttactctt	gaccactgtt	tattgtattt	tatgttcaga
	9001	gtcatgaggt	tggtagacag	aaagcttctg	ttcacttatt	gcccttttca	aaatatctgc
	9061	aagttaaatgc	cataataatg	ataattcctt	acctattata	atgctttata	atttacaag
20	9121	tacttttcaa	tctatcattt	catttgattc	ttattgccac	tcaagaaagt	agaaggagct
	9181	gctcttacca	tcctgaaact	cagaaagagt	gaatgattta	tcagaagtaa	gactgaatga
	9241	ataatgtagc	catgtaatgc	tctggctttt	aatcctggac	tgtttgtcta	acactttatg
	9301	tgcgggtggg	agttttaatg	ccaagaacac	tctaatagtc	aaaagacatt	tacatgagac
	9361	ccagaatttc	tgaaaatttt	attgcagaat	atgaatactg	atttagaaca	aatcacagtg
25	9421	tattctaaac	accacccctt	gatgtttata	aatatacttg	ggtaatgtat	atatttccat
	9481	tgaaaaccca	gaaaagtatt	ctactttaat	cattccctct	tacctgaaat	ttccatgtaa
	9541	ttcactcctt	ataagtaagg	tattcaggac	acttatcaaa	atgcaactag	gatcttgact
	9601	gaataaaaca	ttaagccctt	atcaaacatt	tacgttatac	ctagaatttg	ttttctcaga
	9661	tttgtttgac	cctaaaggga	tagaatacat	tttgatgggt	ggtttcttat	caaggaaatc
30	9721	tgaagcatga	aaacagaaaa	gagtttttag	caaggaggac	agaggggtcc	tcaaaaacaa
	9781	acttcatcta	ttttataact	tttccaaggc	tgagccctga	ctataatgcc	atgctgggct
	9841	attggaaatt	catgccattt	acccaacaac	acatgagatg	gggaacaaga	caaaccttc
	9901	ttgtgttctc	ttattttatta	atttgtgggt	aagaattgct	ggtatataaa	gaatcatgtg
	9961	attaacccca	taaaatttaag	gaaaaatcaa	gacagtaaag	tatcagctgc	cttaatcctt
35	10021	tgtggcccaa	atgtggattt	ttaaaataag	atattgaaaa	acgtatcctg	cacatgtacc
	10081	ccggaactta	gaaagaaaga	gagagagaga	gagaaataaa	gaaagaaaga	aagtccatgt
	10141	taagatgttt	tttcagatat	aatctgctgt	ccttcaagaa	caagaaagaa	gacgggctca
	10201	ctgatccata	caaactaaca	cccacttgga	aattcagatt	tgaaaacttc	ctctgaatta
	10261	gaacggagtc	acacggtttt	aggacagctt	ccccctcccc	ttcctgttga	acatctgctc
40	10321	tgagtgttca	tggcttataa	agtcagggga	gtcctcccgg	ggtagattca	gctggggagg
	10381	gcacgtggcc	tttgctctgt	ttccgtttag	caggaaaccg	tttgaggcct	ttggctggga

10441 accccccttc agaaagtctc cctttcacct ggtgccccca tgggtgcttc agggactcgc
 10501 attgcaggct gggagtcagt tcagggttgca acacgtcacc ctccaagttg cttgaaggcc
 10561 ttagactgtg gtgcaaccag ctgctgccaa gagcatgtgg gtcacagtgg gtccccctcta
 10621 gctttatcat agactcatac tttctcccct cccctccca tccccacagg tgagaagcct
 5 10681 cacaagtgcc aggtgtgcgg caaggcattc agccagagct ccaacctcat caccacagc
 10741 cgcaaacaca caggcttcaa gcccttcggc tgcgacctct gtgggaaggg tttccagagg
 10801 aagggtggacc tccgaaggca ccgggagacg cagcatgggc tcaaagtggc accctggctg
 10861 gctgcaagca gcagctacac aacactacag agggcagcct ccctgcttgc caccactctg
 10921 ctccctgctt gcctccactc ccttctgact ttccagaccc cagggtccagt ctgcagatcc
 10 10981 taccagggttg ctccctcctc gccttacctc ctggagctgc cagaagaaat gaggtacctt
 11041 ttcaaagtgc agccgagagt gagaaccaag tgactctcta ggcttcggac acaaataaggc
 11101 tcctctacac ctgaagacaa aggcaaagtc aaatggggac cagaataaat cttagacccc
 11161 acagtccttc ccatttccag ccctaattcta cagacaggaa tgcccttcag gtttcttccc
 11221 tccccctct tgacctaccc cagatatctg tgtggaagag gaggaatcac cattacaag
 15 11281 gtggacaaat gctaataatt ttatctagaa agaagagtga gtgttaactt ttatTTTTTT
 11341 ccttctgggg ggtctgttga ctcccttctt ttgggtgctg cctataaatc ttggaggaat
 11401 catttctcct cctcaaaaac tgattcagaa actgacttgg ggaaggaatt taatactttg
 11461 aagtcattgag atgcaccatc gaggtaccc ccaagaagaa gcagaagaga agttggtaat
 11521 gagaggggat tagaggctct cccttcagta gggctgtgaa aacctcatca ctggaggtaa
 20 11581 aagcacaagc aatgcctgtg gacaagatgt cattcattca ctacagaaat gttcatggat
 11641 caccggctac caaggtacca ggaccatgc taggtattgg ggaagagaga ctgaagtcac
 11701 aacccttgac tgctcctcaa aagctaacgg ttgcacctcc aagtggctgg gtctgttctt
 11761 actcttgagg ggaattctga gaagacagca cagaattgta aacctccct tttgacctt
 11821 ttggatttta tcagggtgaa acaaaaagct gaacagttac ttcaaagata tgtgtgtata
 25 11881 ttcagttttt tattgttaag ctgatatttt aaagatttct gagctagcag gcatgtggga
 11941 aggaaggctc tgtcttcaac tctttgacct tccatgtgta ccatagaggg gggaaagggtg
 12001 gtattttcac tttgatgagg ttggtaaagt ttttagatc ttctggtaag cattatgttt
 12061 gttaatacat atttattaga gtgatgtttt aagttaataa agtattaaga gtatta

30 *Neurotrophin 3 (NTF3)*

The NTF3 gene encodes the neurotrophin 3 protein, and has homology to sulfate transporters. NTF3 is expressed in inner hair cells and in surrounding supporting cells in the adult cochlea. NTF3 supports connectivity to spiral ganglia-like neurons (SGN). NTF3 induces synapse regeneration and SGN protection after damage (Wan et al. (2014) *Elife* 3; Budenz et al. 35 (2015) *Sci Rep* 5:8619; Suzuki et al. (2016) *Sci Rep* 6:24907).

The human NTF3 gene is located on chromosome 12p13. It contains 2 exons encompassing ~63 kilobases (kb) (NCBI Accession No. NG_050629.1). The full-length wildtype NTF3 protein expressed from the human NTF3 gene is 270 amino acids in length.

Methods of detecting mutations in a gene are well-known in the art. Non-limiting examples of such techniques include: real-time polymerase chain reaction (RT-PCR), PCR, sequencing, Southern blotting, and Northern blotting.

An exemplary human wildtype NTF3 protein is or includes the sequence of SEQ ID NO:

- 5 26. Non-limiting examples of a nucleic acid encoding a wildtype NTF3 protein is or includes SEQ ID NO: 29. As can be appreciated in the art, at least some or all of the codons in SEQ ID NO: 29 can be codon-optimized to allow for optimal expression in a non-human primate.

Human Full-length Wildtype NTF3 Protein (SEQ ID NO: 26)

10 MVTFATILQVNKVMISILFYVIFLAYLRGIQGNMQRSLPEDSLNSLI IKLIQADILKNKLSKQMVDVKENYQSTLP
KAEAPREPERGGPAKSAFQPVIAMDTELLRQRRYNSPRVLLSDSTPLEPPPLYLMEDYVGSPVVANRTSRRKRYAE
HKSHRGEYSVCDSESLWVTDKSSAIDIRGHQVTVLGEIKTGNSPVKQYFYETRCKEARPVKNGCRGIDDKHWSQCK
TSQTYVRALTSENKLVGWRWIRIDTSCVCALSRKIGRT

15 Mouse Full-length Wildtype NTF3 Protein (SEQ ID NO: 27)

MSILFYVIFLAYLRGIQGNMQRSLPEDSLNSLI IKLIQADILKNKLSKQMVDVKENYQSTLPKAEAPREPEQGEA
TRSEFQPMIATDTELLRQRRYNSPRVLLSDSTPLEPPPLYLMEDYVGNPVVANRTSPRRKRYAEHKSHRGEYSVCD
SESLWVTDKSSAIDIRGHQVTVLGEIKTGNSPVKQYFYETRCKEARPVKNGCRGIDDKHWSQCKTSQTYVRALTSE
NNKLVGWRWIRIDTSCVCALSRKIGRT

20

Rat Full-length Wildtype NTF3 Protein (SEQ ID NO: 28)

MSILFYVIFLAYLRGIQGNMQRSLPEDSLNSLI IKLIQADILKNKLSKQMVDVKENYQSTLPKAEAPREPEQGEA
TRSEFQPMIATDTELLRQRRYNSPRVLLSDSTPLEPPPLYLMEDYVGNPVVTNRTSPRRKRYAEHKSHRGEYSVCD
SESLWVTDKSSAIDIRGHQVTVLGEIKTGNSPVKQYFYETRCKEARPVKNGCRGIDDKHWSQCKTSQTYVRALTSE
25 NNKLVGWRWIRIDTSCVCALSRKIGRT

Human Wildtype NTF3 cDNA (SEQ ID NO: 29)

atggttacttttggccacgatcttacaggtgaacaaggtgatgtccatcttggtttatgtgatatttctcgcttatct
ccgtggcatccaaggtgaacaacatggatcaaaggagtttgccagaagactcgctcaattccctcattattaagctga
30 tccaggcagatattttgaaaaacaagctctccaagcagatggtggacgttaaggaaaattaccagagcaccctgccc
aaagctgaggctccccgagagccggagcggggagggcccgccaagtcagcattccagccggtgattgcaatggacac
cgaactgctgcgacaacagagacgctacaactcaccgcggtcctgctgagcgacagcacccttgagcccccgcc
ccttgatatctcatggaggattacgtgggcagccccgtggtggcgaacagaacatcacggcggaacggtacgcggag
cataagagtcaccgaggggagtagtactcggtatgtgacagtgagagtcgtggtgaccgacaagtcacggccatcgac
35 cattcggggacaccaggtcacggtgctgggggagatcaaaacgggcaactctcccgtaacaatatattttatgaaa

cgcgatgtaaggaagccaggccgggtcaaaaaacggttcgaggggtattgatgataaacactggaactctcagtgcaaaa
 acatcccaaacctacgtccgagcactgacttcagagaacaataaactcgtgggctggcggtggatacggatagacac
 gtcctgtgtgtgtgccttgtcgagaaaaatcggaagaacatga

5 A non-limiting example of a human wildtype NTF3 genomic DNA sequence is SEQ ID NO: 30. The exons in SEQ ID NO: 30 are: nucleotide positions 1-229 (exon 1) and nucleotide positions 62081-63186 (exon 2). The intron in SEQ ID NO: 30 is nucleotide positions 230-62080 (intron 1).

10 Human Wildtype NTF3 Gene (SEQ ID NO: 30)

	1	agttgaagct	cctctccctt	ccgaacagct	ccgcgcaccg	ccccgcgacg	cagcccggcg
	61	caactacttt	cttctctctc	ctttctttct	tcctctcctt	tttccctcgc	tgggtagtgg
	121	ctgcggcggg	gtgggggaga	ctttgaatga	ccgagctcgc	gtccaccttt	ctcttcatgt
	181	cgacgtccct	ggaaacggcc	acacggatgc	catggttact	tttgccacgg	taaggggagg
15	241	cggcgggcac	cttgggtggg	caggtttggg	gatgggggtc	cacgtgggga	gggattttcc
	301	agtggactgg	tgcggggggc	cccagatccg	catcccgcgc	caccccatc	gcgccgcgct
	361	cactcacttt	cccgggcttg	tgtcttcccc	aaagtgttgc	ctgggatctg	ctcaggccga
	421	agcgcaaccg	cagccacccc	gctacacaca	cacacacaca	cacacacaca	cacacacaca
	481	cacacacaca	cacagacacg	gacacccttc	tccacctcct	cccctcttgt	ccctcggtcg
20	541	cccaagaagc	ttccctcaat	ctgggaaagt	gatcagggtt	aagggacctg	gattggaaaag
	601	ggtgggggca	gaagagggga	aatggggacg	acgaaagagc	aggaagaga	ttcaacagaa
	661	tcaaccaccc	accactccca	accgaccgc	ctgctcctcc	gagaaagctc	ctagcgcac
	721	ctataacaaa	agggggtggc	agacagaact	ccgggcgggg	aggtgccg	gcagctcccc
	781	tgcacacgcc	ctgcactctg	ccggccgctg	agcctgattc	tcagctcgcc	ccagcaccac
25	841	tctggcccgg	gcgtgggctg	gggggagggg	acgcgcagct	caggaccggg	aacctcgctg
	901	tccagttttg	ggagtgggga	ctcactgcca	cgccgcgctg	acctgcgttg	gagttccccg
	961	aaagggtttt	ttcagaaaag	acctcgcgcc	ccgggctcct	cttggccagc	gcccaccggg
	1021	tggccacccc	accctggggc	tttgcgcaga	tgttgagct	ccgtacgcag	cccgcacatc
	1081	tgggacccct	ccggggagcg	gcgggcaccc	gggcccggcc	atcccagggg	atctccttgc
30	1141	ggtatcgtcc	agcctgttct	cggactttga	gcggtggcgt	gggaggccgg	gagacctggg
	1201	caccgcgcga	gccagccagg	tcggagttta	aaggtccac	gacggaccga	actgtcccat
	1261	tgcaccagag	ctttactcag	tgggtgatgc	tcctgatgaa	atttgggacg	cttgggagtt
	1321	gaaggttagg	gacaggaggg	gcgagggccg	agggcatggg	atgggggagt	aggattctgc
	1381	ttgttgctct	ccgcgggagt	gggtgcgcgt	ccaggaggcg	ctgcttcttt	gcgggagttt
35	1441	ggctgctgcg	ttcattcgct	gtctgcgttt	cagatgcacg	gcactgagac	ccttgcgtcc
	1501	gacggtgtcg	gggctgtgga	ctagaaagga	tcccttttgc	tggaatcgag	gctgggggtg
	1561	gattgccggg	gggggaaaca	ccgaaaagat	cgtctggcct	cggcctctgg	cggcggggcg
	1621	caggttctga	gtccgaatgg	aggttgctcc	cgggagcgcc	gggctcagag	ctagagagct
	1681	cgggagactg	tgcgcctgtg	gacttgttta	tgtgtgtgaa	gaggcggggg	cgagggcctg
40	1741	ctgagagggg	aggggagcct	ggaaggggtg	ggtgtttctc	ctggagcctg	atgtttgtaa
	1801	ctcagctgat	tatggagtgc	actgagcgac	ctgcttttta	aataaagagg	tgcccgtccc
	1861	taccccgcaa	aacagcgaac	gaggagaaca	tggaaagcgt	ctgtcctaaa	cgtcaggatg
	1921	ggagaaagtt	gtaacataga	ggagactttt	ccagagggtcc	tgttttcaca	acactcagaa
	1981	gttctccagc	gtactcagcc	tgccctccgc	ctgccctcaa	ttcctttttg	acatgtcaca
45	2041	caaagaaagc	tgaaaggtag	aatgtggagg	ataaactcca	acccctctg	ccttggggcg
	2101	aaacacacag	acttaggtat	cgtgttagga	ggtaaggggg	ttggaaaata	atgcaggctc
	2161	cggtagacag	tgttgaaggg	agatagaaa	tctgggggat	ttcccttagg	gagagtgtgt
	2221	gggattttgt	gttgggtgag	actggagtgt	agctggactt	agagggtttg	gttgtgtgtg
	2281	aatgggatat	gcttatctat	ggagtggagt	tgtgaccatt	gagtggagtg	tcaggagggg
50	2341	atggggatgt	ttttccaaag	tgtgtgtgtg	tgtgtgtgtg	tgtgtgtgtg	tgtaggggaa

	2401	ggtggcaaaa	gagctggagt	cctatccctc	tttgggtgct	gggaggtgtg	tgtttctggg
	2461	gagtctgtcg	gatgggcac	tgtgtgctta	gcgggagaca	agagcttgca	ggaacatagg
	2521	tgtagctaata	gcgaaatctg	tgccatgaa	gggtgtcaga	gcaggaggag	aggtgcagcc
5	2581	tgttgaagac	agatggcttg	tgcatgcctg	cacaaacaca	tggagtgggt	gtggtgtagg
	2641	agggcggctg	tgtgtgtgag	ggaaagggtg	atgcatgttg	aaggttgggc	gcctgaggat
	2701	taccaaagg	atgtgtgtg	gggggtgggg	aagattatct	gaggagcacg	gggtttgaac
	2761	ttcatgagg	tgcaactgga	aattttggct	gaatggcact	ggggcatgta	gggggtagaa
	2821	aagtgtgagc	tcagatccct	tttatggaaa	gacaaagcat	ttgcagggaa	ggacaccggg
10	2881	ttgggtgttg	tggttgccct	ccccctcctt	cctccccaga	gagaagagaa	ttggaggagg
	2941	cttagaatga	gggatgtaca	tttaaggggg	aaagtccttg	tgcatgggca	ctgggtggga
	3001	acaggaagag	atgtgtgcat	aagactgaag	taagtagagg	aatgcactat	ctgtgcctgg
	3061	tgtcagggat	gcatgttgta	taatgttttg	taatgggttt	cttatgtaat	aggagaattc
	3121	actatcacca	gttcataatt	ataatcaatt	aaagctaata	gagtgcctac	tggatgctca
	3181	gccatttgct	gggtactggg	gcatggcaca	ctgccctgga	aaaattcctt	ctccaggctc
15	3241	agtaaaactg	atgcaatttt	gccagaaaata	gacagagctc	gacgcttggt	gaggtttctt
	3301	caagggtctat	aattcagttt	ctagagatgt	ttatggatct	tgggggcca	agggaagctg
	3361	acttgaaata	tatctctggt	tttgtaaaaa	gtggggcata	aagagaggga	tgatgtagac
	3421	ttttgtgtat	tgggaaggga	gctggacggt	gttgtgtgtt	cttgtgtggg	gtgtgggctg
20	3481	ggaattggag	ggtgctgcac	ttcaggtcag	aatgcaggct	cagatccctc	taaggattga
	3541	gtttctaaga	gtacacagag	cctccttcat	agctggaatg	agacctgtta	tgtggagctg
	3601	gagaatggtg	aagtcacgat	caaggaagag	atggactggg	agtgtttggt	catctttgat
	3661	gttttctggg	ctacgagtgt	aaagctaacc	tgagtgtgtg	ttgggggggt	gggtatgcat
	3721	gcagggtcgt	gcacatgcac	tttggcctgg	agtaacgtaa	ctgggagaaa	actagcgagg
	3781	tggataggag	gggtggatat	tctggctcct	gtgaaggggg	tgccagtatg	tgtttttgca
25	3841	ggaattctgt	gtgtctgggg	aaggataaaa	attgtcttgg	atactcagat	gtgcatcttt
	3901	agcgaaaaaa	ggtgtacaca	aaagaagggc	cgtagacttt	ggggaatgag	acaaagggtg
	3961	cgttgatcgg	gtccataccc	tgagatggct	ggtgaggcaa	ctgggtagga	acagagtga
	4021	ggctgtgctt	ttgtgatgca	cacatctgcc	tgtctggagc	agacgatggg	gatccccatt
30	4081	tgattaggtt	ctgaccaccg	ttgggcatat	gtttgggagg	actttatttg	actcatgaca
	4141	tttttgaatt	catgaagtta	cagctgtgcc	attttaaggg	tctaccttat	ttttgagagg
	4201	gaaactcatc	tgtgtaagag	atagtttata	tagggcttga	ctatttgtca	gaagatccgg
	4261	ggcagatatt	tctgagagca	gggtggtgaa	gtgcaagact	ggattgggaa	tgtcttagtt
	4321	ctagtccagt	tctaccactt	gggaaaagtg	tgactttggg	gacattttta	atctctataa
	4381	ctcttctgtt	tctctttttg	taaaggggca	atataatgac	tcaaggagaa	ggaaagttca
35	4441	ctcctctttt	cccccatcaa	aacttgatgt	gtttagttgt	atacagccat	acaaagatga
	4501	aatgtaatta	aatacagggt	ctagcacgta	tgtagggtcat	tgtgaaaata	ggatctggac
	4561	ccctttctct	ttacaatgca	acttaatcag	taatcatgat	gcccacagca	tctctactgc
	4621	tcactcattg	tgtaataaaa	actttttata	actgcctcag	tttactcatc	cagaagctgg
40	4681	aggcagccaa	acctagatta	tagatgtagg	aaatctttta	gatctttgag	ctattcagag
	4741	gaaaggtagt	ctaagaatca	atgttattaa	tgatcagaga	cagacatttc	tttatttttt
	4801	aacatcagct	atgcagacca	cagcaccaaa	agaaatgtga	cgtggagaac	agcaaaaata
	4861	ataattgccc	ttgacatagg	caccatttga	ataaaaagaag	agcaaagtct	agcctgcctg
	4921	gtaagaggaa	agcctgggct	ttgggtcttt	gaaattcttt	gtgccagtg	cagcaagttg
	4981	ggtacagcca	ctttccccag	agatgaagat	ggagtcgtag	aggggcatgg	aaacacactg
45	5041	aaacctccga	ttcctcccag	aagtcagtac	atgctgtcaa	ctcctggctc	agagtcaggc
	5101	tcccatcctt	gagttcagtg	gccataaata	gtcaggagct	aagctatttc	caagaggcac
	5161	taaactaatt	ctgttgcttt	tcttggcgag	attggggcaa	ttaagttgat	tggaaagggt
	5221	cactgctgat	gggtctcaca	ttccccaggc	gggacaggcc	agcagagttt	aattacataa
	5281	catttcccat	gtcttccact	ccactcccag	atgtatcaga	tgttcaggcc	tctcctcttt
50	5341	ccctacattc	aggatgcctg	acggagcagg	cggcttctgc	tacaagcttc	tccagacctt
	5401	ttcctgagca	gttacagcac	tcccacccca	actggaggga	atgcaagtct	gtgcattgcc
	5461	tggaaagtga	agaggggatg	caggctgaca	cagaagatct	ttagagccca	ccaccctgcc
	5521	tcaagcctca	cccgcggctg	ttttactctt	taacgaggca	ggcagtttct	ggcgttgggg
	5581	atgttgaact	aatgaccggt	gctaggggtt	tagatttagt	tttattttaa	aatgagtttg
55	5641	ttggaaaagg	accaaattct	ctactggaga	caagcctaag	ccccgtgtgg	tgctgctgat
	5701	aaggctctgt	gatgtacttt	gtgcatggac	aggcatgtgt	ggagtaagag	gagagcacag
	5761	gatgtgcgtg	cgtgcgtgct	gggggacacg	tgagcgtagg	agatgccttg	ggtagatggg

	5821	gcattttgtag	agcttatgtt	atgggtggca	ggatatctgga	tgcttggtc	agattggatt
	5881	gtcattggat	tgcttcttac	aggtgggaca	gcctgagaaa	aaagttgggg	acaccttct
	5941	tcagatcatt	tttatatggg	catgttgaac	cttgaagtat	gagccagtg	tagatattac
5	6001	ctcattcttg	cctggatgga	tttctaacaa	gattgaggct	ggacaaaaaa	aaaaaaagtc
	6061	tttcagagga	ctcatttttag	aacctgtgag	aacacatgta	gatggcccca	gagcgttaaa
	6121	gtattttttg	ttgttggttaa	tagttttcca	taactcagtc	ttttgtttta	cctaaaaata
	6181	ctgagaggct	agaggttgca	cacaccccca	catggtcaga	atgcatttgc	tgttcatgtc
	6241	cccattctgtg	ccatggtcct	tctcaaatag	gattaattgg	gaactacgga	gagcttgccg
10	6301	cagcttttgg	tgctctcaca	ggacaaacct	atgactgacc	gtgatgtcat	tgaaacagaa
	6361	cagatttttga	cgtgaaat	cgtgtggtct	gttgtgattg	ctcacttttt	agtaacatct
	6421	ccagagaaga	catttggtga	gccccatttt	tttttggtgg	tttggtcagt	cagagatttg
	6481	ttggaagatt	taaggttctc	caacattcaa	ggccagagat	tttgaggaa	acagtagcta
	6541	acaagagccc	agccagggt	gaccttcagt	agcctgaggc	agcaggtagc	tgctgactt
15	6601	agaagtgaag	gaccagggtc	agtgactggg	ctctgtcact	aattagctgc	tactgtgtga
	6661	ccctggacat	tctctgtagg	acatgacttt	gccatctgta	aatgaagggc	ttgaattaga
	6721	agatctccca	gtcccatcc	agccctgtca	ctgtaggatt	ccacagtagc	cacatcctcc
	6781	tacctaatca	gtgctaatta	cctcacttgg	agtgcaccct	ggcacgccct	taacgagctc
	6841	ctgtttttat	ctttggggaa	tggctgagtc	aaaggcagaa	aagcgtgcaa	aaaatttgcc
20	6901	agagtgtttc	cttttgcca	agaatgctct	aaacattggc	ttagctaca	cagggtgtgc
	6961	agagctgaaa	gggtgggat	ttggctttct	ggtaggtctg	atgctaatac	ttcagcttcc
	7021	acgtgaatgt	ctttcagtgg	caatgtctct	gactaaggta	agcatccatc	tttaagatga
	7081	tggatgatga	tgatgatgac	agttaatat	agagtactta	ctgggtgcca	ggcattttta
	7141	cagctattac	atggcattat	ctcagtactc	aaaataatct	catttacaga	tggtgttact
25	7201	gaggttatgt	gaggttaagg	ggtatgctga	tgcttcttgc	tttccagcct	gaagtctgtg
	7261	gcatcccaag	cccattcacct	gctagtcaga	ttctgttttt	tatgtctgtg	atgtcatcca
	7321	gatgggagct	gggtccatct	tcttctgcaa	agttggtcac	cccattaggt	aacctgtctc
	7381	ctgcccccg	aagcctcctt	gtcatgtgca	agacaaatgt	ttaggttttc	tagccacctg
	7441	ctggcaagca	tgtgtcagat	ctgtctcagc	ctccattggc	cctctcctt	tacttctgt
30	7501	tctgttgaca	agggaaacctg	cccttacaga	atacggatat	tgatttaagc	cacaattgcc
	7561	tccttgtaaa	ctctataaaa	ttggagctgg	acatctggtt	ttgatccaca	cgcccacgag
	7621	aggtggtagg	tcccagcatt	cttgagtcct	ggaattgcac	agaatcaaag	tgaaaaagaa
	7681	tatgaatgga	gagcatgagg	tatctttaga	atttgaaga	tggtttctt	tgataggcta
	7741	gggaactcac	accatttaga	ctggggccaa	agttaaggta	ggtttcttgc	tggtgaggtg
35	7801	gggggcagg	gtcactgcta	aagactgagg	tccccaggct	gtgtttgatg	cagagcagag
	7861	cagctgtaga	gccctccagc	tgaggatttg	ccactgcaca	ccattctctg	ggcagtacca
	7921	gaaatagaga	agccaaacag	gttcttgggg	tggtggttgc	cctgtttgaa	gattgtgact
	7981	aaaggctcat	tccttcttcc	tccactgatt	tcctcaccat	cactaccaca	tatacatcat
	8041	caagacagtt	aaggacaggt	gactttgttc	tggtgcttaa	acataattgc	atgtacaaat
40	8101	aaatggtgaa	cactaaccta	atgcctatta	gaccataggt	actcaaaaat	tatgtattta
	8161	ctttgcttaa	aataaccttt	catgtgtagg	gtgataat	ttttcttcat	atggagtata
	8221	tagcctttta	aaaactgata	gccagggtct	tgcaatttac	ttcccccttc	cctctgtttc
	8281	tagatctaag	accccagcat	ttaaaaatat	ccttctctgg	ctctaagtga	tccctcagg
	8341	agtttggttc	aagttctcaa	tctgagatta	gtttcatccc	acatctagag	gacctgttct
45	8401	taatggttca	taaggtactt	ccactgggtt	caagaggaag	acacagctct	agaaatcacc
	8461	tgctttccca	acaggtgtgt	cccattggtg	ggaagaagcc	cagtcctcct	aacaccacc
	8521	tggctacctt	aaatagcttg	ccacagtctc	cctgcagaga	gcccagacta	gctgtgtgac
	8581	tttaagcaaa	tcactttccc	tccgtgggcc	tcctttcctg	gaaaaggagg	gagttggact
	8641	cactagagg	tcacaaggac	cctaccagtg	tgacagatca	catcttccat	gactccgctg
50	8701	ctccctcgcc	cccacactgt	ggcccatgcc	tggtattttc	aggtaacctt	gctggcctag
	8761	aggacttaca	tagagtgtat	tggagggggt	ggaggaggaa	attaagttat	gtgatgccag
	8821	ataagtcatt	gaacatacgg	atgagtgcac	gtttatgaag	cactgaagaa	acaagcttgc
	8881	cagattcgtg	gcggatcttg	gattggaagt	tggtgagtat	cgagaggttt	atgtgggatt
	8941	aaagttgtct	gtgggttagct	gaattcaata	agctaaatca	gagagatccc	aaactctgca
55	9001	gagagatcag	accccgccctc	taactccaac	agaggcta	tacaattttg	ggaagtcctc
	9061	gaaccacact	aggtctcact	tcctcatcta	ttaagagcta	ctgactttga	ttatctctaa
	9121	ggtctttctt	gacttatcat	taattaaatt	gtgagaattg	ggagctagta	gaaaaatctg
	9181	gacaaatgcg	tggatggaga	aggggatacc	tcaaatggcc	taaatataga	gggaaatggt

	9241	ctgtacaagt	gatacgttta	ttcttgaatc	ggagatagca	ttggaggata	tattctgtac
	9301	aaatattgtg	cttcaaattgc	ttacccctcc	atcatcatca	atgatattaa	aatcattgct
	9361	taacctggga	tatttgctta	acctgggaag	cctccctgca	tttcccaga	gagcctaagg
5	9421	ctcccttccc	agcctcctga	tgccacttag	cagcctgtgt	ttaatttgca	ttcctagaat
	9481	gtgacttcac	agagaaaagg	atctgttggt	ttcatttttg	tgtcccccac	cttgaatagc
	9541	gacagccaca	ggaatgtggt	gagtgaatga	atgctgtacc	agccattgtg	agctcagaat
	9601	gtgcagcagc	gatttgacag	tgccaggaac	ttggctcaga	gaggcagtgc	accttgccag
	9661	ggcttcaaaa	cagtgttacc	tttatccccc	ttttggcctg	tcccaggct	aaggctcatga
10	9721	atgctcagcg	cttggtggcc	agaaaaggct	gttcttactc	ccttttagag	acctttcttc
	9781	cccacatctt	agcctgctta	gttctcctgc	aaattggggc	tcttggtcac	agggaaatcg
	9841	gtgtcattgc	atacgcatca	ggataactcc	tcgttagaca	ctcgaaagc	aactcaaaac
	9901	acacgtgtac	agattattca	ttaagcagtc	tggtgtggt	cacgtgctaa	atgttgggga
	9961	agatgcagag	atgaatgtag	tgctgctggc	cttgggtct	ggaaactggt	gaaggaggca
15	10021	gtcaggttac	aatcatagt	ctccaggag	aacaaggga	gcaaggaga	tggaatacag
	10081	ggttcccact	tcaggtgttg	tttttttagta	caaaccatga	ttgagaatcc	tcttggttca
	10141	ggagcaagct	ggaggggtcca	ggaatggtct	ccgcttccac	atgacagtgg	ttttcttgtc
	10201	cctgaattcc	cagggtcagg	gtgatggatg	ctcagagacc	attgtgatgg	gagcccttgg
	10261	tccttgtggg	gagagtcggg	ggaacacgca	tggtgtacag	gtgttgggg	ttggggaggt
20	10321	gtaggtgtgc	atgctcacca	gggtacttct	cagaagtatc	cctgctcgtg	gcagtgtggg
	10381	ggtaggtcca	cctcaggccc	ctaagctgtg	cttctacttc	ctttctaatt	atacagcatg
	10441	tattgtctct	gtgaccactt	ctgacacagg	aacagactgc	catctactgt	ccacttctgt
	10501	cctgagtccc	tcctgtctca	ggaggcattt	attttttctc	atcctgtttt	taatagatcc
	10561	ggacaccagg	agaaacagac	gtggagtcct	atttagtaat	ttttttagaa	tttagaagg
	10621	agccttcaaa	aaacatagca	catctgtcag	atggtttccg	tatctgttta	caaagtatta
25	10681	ttgtaagggg	ccactctgct	gccactgagt	cttgctagcc	gctgtgcgag	gtgccgtgtg
	10741	ggataccaag	agatacaaag	caaggcctta	ccttcaacag	ccttgcaagc	tacttgcaaa
	10801	tacagtgtag	atttgacag	aagcagacta	caggatgtgg	cagtagatat	ttggtaaag
	10861	gcatctgggt	ggtagggact	acagaggttt	tgagggtttg	gagcccttat	gagcccttat
30	10921	aggctggggg	atttgggaaa	gaccaatatg	ggtctccttt	tagcaatccc	agtgttttag
	10981	aagaggtctt	taagagctgc	gcgatgttgt	gggacaggct	atgaactgga	aactcaggaa
	11041	cctggagggg	ctgaggtcta	gtccagcttc	tggcagtcct	gccctgtgac	ccagggcctt
	11101	gggttccgca	tcctcaaagc	aaggctgttg	gtcttcttgt	catgtaagg	ccccccatc
	11161	cccgaccctg	catcatgtac	cactcttgga	tgcatgtgga	ggtggtttcc	tgggttccag
35	11221	cttgtcaaga	gcaatagagt	caattggatc	ccatgcagga	aggatacctg	gatgcagggc
	11281	cctgtgctga	tgaccccaca	tgacagcagga	gagagagaa	ttccccagg	gagttttggg
	11341	tgctgttaaa	atacaaagg	gggaaatgtt	ccatgcaacc	tctgcatgcc	actcctctcc
	11401	agagctttcc	tggactttcc	ttccaaaaat	atatgtgcat	ggtgctttgc	gtccaaagata
	11461	gctaccccaa	aagaaattgt	attagtattc	tagcattgtt	catagctcaa	actaaagatt
40	11521	ccttcacatt	agaattcatt	ccatgattgt	ttttatggag	gttttctttg	agtattttag
	11581	actttttgta	gctgtggatt	aaaatgagga	aaagattgtt	ttctatagct	cttttactcc
	11641	ttgtctcttt	ttctctgtct	ttcttagcat	ttaaaccacc	accaaggaaa	cctccagggt
	11701	gtttattgca	tattcttgcc	ttttgatgtg	tggtgtgtga	tgagagagag	agaaaaaaa
	11761	agaaaggag	agaggagagg	cgagagagag	tacaagagag	aaagatattt	tagactgtgg
45	11821	tctattagca	tgtctaggaa	atcaagttga	atttgattg	ggttacaggt	tgagagacct
	11881	cagctcgtca	ctgtaatcct	ctgacatttc	ccttttactt	tttggtctct	gtctgaatca
	11941	agaatacatc	ttcccgttcc	ctcttccatg	tttacctctt	cttttggggc	agctgtataa
	12001	agttactgtc	tgtctacagc	aagtctcctt	aattcttttg	gggactgctc	gacagaggca
	12061	cagtcaagga	tagaaccatt	agagacggtc	agcttttgga	caaagacaaa	aaaatatgtt
50	12121	gaggaatagg	agaaatgttc	caattgtcga	ttatgtgttt	tcctgtggta	tgctgatgtc
	12181	tgtgactgtt	gtccacagaa	agagagcgtg	agcaggggct	ggggcagaga	taaaaggagg
	12241	attctcccca	aaccaacaaa	tgtggagaga	gagcaggagt	ccccagcaac	agatatggag
	12301	ccaaacaacc	ctgggaagta	actaggaaac	ttcctggtga	cctgagatgt	gttttgtaat
	12361	cactggtgaa	agtaggagga	gatggggtag	aggagagtta	caggtagaga	agttccgtgg
55	12421	gaggagaccc	caaggtgaga	gaaagcagg	acttaaagaa	ctgaaagaag	cccaggagg
	12481	ggatgaacag	agaacatccc	atatatccca	tatccagtat	ggtggagagg	tgggcgggag
	12541	ccaactggga	gcttactttt	tttttcccag	gggcatggga	actatgaaa	actctgagg
	12601	cacaattaag	cttttgtttt	taaaaagatc	cctccggctg	cagtgtggct	ctgaataaag

12661 agttgttcaa tgaagaaagg aatggctgat gggttgattg atggaagaat taatgaatga
 12721 gcaaacgaat gcatgagatg cagagggaac gcagcaagcc tgatttgatc tctggttact
 12781 ttgaccagta ctggctccag gggcgattat ctcagcctgg gaggccaggg agtaatgcat
 12841 tgattagaat gtctctggac acatggatta aaatatctga tattttaggg tggttgatag
 5 12901 tggggaggac ttctgaaact ttttccctct tctatgcatt tccatcctaa tttgatttca
 12961 ttcaggatca aaaaagaaga tggcttttga aattacatcc cagaaaatct gaaactgtgg
 13021 cattgacttg ctccagagag ggctgcttgc atggaagacc ttttcatagg ctcatcgtgg
 13081 aatagggaca gatgataaag tttcttgggc atatgaaggg gtcccagatt tctggacgtc
 10 13141 agatccaccc atagatgatt ccttggatta aatgatgtgt gtgtgtatgt gtgtgtgtgt
 13201 gtgtgtgtgt gtgtgtgtgt gtgtttaaag ttttaaagtc tctgaaaat taccaagtgc
 13261 tactgaacat tttttttgtc agtagtggat atctggataa tttctttcaa ggccacattg
 13321 cttagcatgt ataaggaaag tgtgtgcggc agagaccag atggacagcc ggcatgccat
 13381 ccagttgctt ggggagtgtg aatactcctt ggcaaagcca aaggagacca aaggagacct
 13441 ctagtgtctt tgctctccca tgtcctgact attccaatct cactttgcat tttgagatct
 15 13501 ttttaactttt gctaacagtg attgctcagg atgtcattca ggccaattaa atttgaatgt
 13561 ctagggttgg gacagaagct cacgagggtga atgcaatgtg cagccacgtt gagaaccact
 13621 agctgacact gctaactctac tctgctctcc tccctgcctt ttggcggctt tgccatgata
 13681 tctgtcccat cctcacctca gtttgcctg aactcctcac ttagttaagg aagttcttct
 20 13741 aaggatgttc aactaatatg ggctaaggcc tcctatcccc taaaaatcca gcaattgcct
 13801 gagaaattgg acgctaggag gataagacag gttcttagca gattctgaag cactcacctg
 13861 ctctcatctg ctgaaggggtg tattgaggat gaatgatggg gaacagcagg gagaaagacc
 13921 agcatccata gggcagtaac aggggtgcaga cacctattta tatgtgtcat cgtagtctcc
 13981 ctgacagctc ccagaaccag atattatacc aggcagatat accaggctta caggcaccta
 14041 tttatatgtg tcatctcctt agttcccca acaactcaca gaggcagata ttataccagg
 25 14101 cttgagggca caacaactca cagaggcaga tattatacca gacttgaggg cacagtttgc
 14161 ataatgcatt cataatgcaa agtggtttgt gatcttgaa tgatcagact tcctacactg
 14221 gttctcccca gtaatttgca tatcagccag ataccctcct ctagggttag ctatgatcag
 14281 ggctgctgat ttctggcact gatggctcaa tgggaaggaa agcgctatc cctctgttc
 30 14341 tgtgggattc agcaatcttg ggctgtcct tccaggaggc ccaacctgag gacatgcttt
 14401 aaccaacagg cctttatatt gagaaaatag tagttctctt ccttgccatt tccctgttag
 14461 gaagccatgg cttgcaggac agccagggtg tagagggtca gagaaggagg ctaaaaaaat
 14521 caagacctga aaaagttaag tttccaacca tatattcca aattgtagag tggaccatat
 14581 ggaatcattg caatccctgg gtttctcatt tacaactttt gcaaaagata ttactatgta
 35 14641 tgctgtcagg catgctggag ccggatgtct agcaccctt gctaagtga tgcttcggta
 14701 agagctttgg acttggtggt ctccagactg ctgagtgaac tacgtgtagt ttagcaacat
 14761 aatgcttgga atgagccacg ccgacctgat gcgagaaccg gatggtagcc gaagagagca
 14821 ggccaggtag tgagacaact atgatctgca gaggtgggca gcgggtgga gacaggcttc
 14881 ttctgcacca agtactgcta cctgaactg acatcacaga gccgagagca
 40 14941 gtggaagcag gaaaacacag gcctatctct gtgtctgtca tgtaagtaag tctctcttgg
 15001 ctgtgtgtac ctgccaccat caccatctcc caccctaaat atacagcatc accttcagct
 15061 atagcttttg gtttttgatt aaaggacagt aatatttcca ggaggagaa gagtgatgaa
 15121 cagaagcatt tatgtatgga cactgggcaa tttattgttt ttcaaagtct aatctagaac
 45 15181 atgtttgcaa gaaagcgtga attgagtaaa aagtttacta tcgtattaag gactgttagg
 15241 tacaatgact gaaggaggag ttaagagtga tccttcaata attcccctgt gggattatgc
 15301 acatttaagg aaaaatgttc ttcaggctat ttctgcctta gagctaggca tcattacagc
 15361 gaaatagaga actaaccttt aagcaagagg aaccaagttc tagttctatt tctgtcaatc
 15421 actgtacgca caacctctca aagccttagt ttctctttg caaagtggga taataaacc
 15481 tacctactta acacaatgtg gggtttcagg tgagatgat cataggatcg tgcttgcaa
 50 15541 gctgtaaatc tgtaaatatc aaatatatat tatggtttca actggtacat tcctaagcga
 15601 atagcacatt gctctgttgg gaagacggct cttctccaag tcaggctggg ataagtgtcc
 15661 ctgacaagac actgccatac ctagggtgtc cccaaacatt gtctctggga accttgagga
 15721 agcaccataa gacatgggaa gaaatgttac agcgtggct tgaaagaata acaatgtatc
 15781 agtctactta tttctgataa tgtcatcttg ggataagaga ctcagggtag cttagttagg
 15841 gacatgggca tgcaactgcac agtaaaaaatg gtgtccagga aacctgggtt tatttcagta
 55 15901 tgggttgccc acacttctgc caaccagtc ccctacttcg tccccagctg ctcttgatga
 15961 actctctgca cacacttgca cctgtatctg tgaaacagag ctctctctct tacatgagaa
 16021 tggatctggt tgcaaatcta atagattccg ctaccacaat gtcccctgcc ttttttgtt

5 16081 acttcatttta tgaaaataacc cttgaaacat ccatagtccc atttttaga catggtgctt
 16141 tatgtctttg agattattaa atactcatgc tcctttctga ttgctgtttt cacctcttct
 16201 ttaggcttgg gcttttctat tgggtggaatt tgctgttcct tttcatggtg ctggctttcc
 16261 tgagttgggt ggttgctaatt tcttattgcc agtttgtctt ctgtgacaga ttctccgaca
 16321 tgcctcggat gtggcttccg tgccttgctt tagcttttta tctgggctct cgtgtcctga
 16381 atattttacct tctttcagga atttctcaca gcttctagcc aatgaaaagtc ccccttctta
 16441 ttgtcaaacc cagacagtta taattttatt ttaaagtac tttttgttac attgttttgg
 16501 gagtagggct ggaaggactt gctaattgat gagtggtgga aagttagaga aagaaagata
 16561 tgatggaaga aggagcaatc aaagccccag aaggaagcta tcgattgtt ctggattcc
 10 16621 taagcctgcc agaaaagact gacttacttt acagtttctg agagaactat gtgtgctatt
 16681 actagaagca caagagaaag aaaaatagga tgttcagcat gcttcattta tctaattgta
 16741 aaaatgaact ctgcccagtg acttaattgg cataaactct gtttctaaaa aagccactca
 16801 ttcggcaacg cacttctgag ttcttgatat ggtaaggat tgtgttctgt gctggacagg
 16861 aatacaaaaa tgcacggttc ctacactcca agaacttata gtacatgtag ggaaataaga
 15 16921 caacccttat tgaatatcac tcaagatgga aatgactagg gccatggaat gtaacacaca
 16981 gaggttacct ggagttccta agacttctga tggatgattg aggagagccc tggatcaggt
 17041 aagcaatttg aagggatagg gatagcacag acagcacagg ggtggaaca agtgaggat
 17101 gtcgagagct tgaccagtag gcctgaaagg gagggagtgt acacagagcg ttaattaggag
 17161 ctctgtctcg aggcagcttc cctcagcccc tcccaggaca tcgaggtttt tggagaaaga
 20 17221 gcctattgct cactctcacg gctcttctcc ttttctctg ctttcagttt gttctttgaa
 17281 ctttttggaa acttccccctg ttctttcttt aacactgtgc ttcactctct ggggttctac
 17341 gttttgcagg ttgtagtgtc tgagatccag ccttcccaa tgatttctct gaatttagta
 17401 tttggtatgg gttttgctat tttgtgcca tcccagccct agcaaagaaa cgacttacc
 17461 ggagtatgga cagggcttca gagaaaacc ctaacattcc tgactcccga ctttacagag
 25 17521 ctctgccaaa ccttgccctg cgggagtaag aaaagcgcta acaagccatc ctcttgggtg
 17581 tcaagtgcag acaaatcact tagccctctc gaggtcctcc aacagtaagc tactggttg
 17641 tgaaaccca ggataatcca tctgatttca gtctgcatt tagtactta gaacttctc
 17701 gcacatgcat ttgctggctc atgtacatac gaatatacac atatccctct ttatgcccct
 17761 tcttagcctc tgatgatttc ttctcctcca ggaggcagga atccaaggct tataaacat
 30 17821 gacttctggg aagttttttc ttctgcttaa ccagggtatc attgtttttt ctgccttccc
 17881 ctggagaatc actggccact gccctagtgg ttggggcaag gatcagaggt agcttgcat
 17941 ctgggggttg tcccaaaagc ctcggtggga ctctgcattg gggctgttag cctggatcca
 18001 ctccagtacc ttaactaatc tcttgactcc cagatggtcc aaaatatgtg gatttagaag
 18061 agcaacagac agctgttccct ctgggcctct ccaagaacac ggtttggtgt ctagaccacc
 35 18121 ttagagaaac atggcagagg aaatcatggt ggagcagcat ggaaacaggt gaaaccaga
 18181 cttagtacct tgtaaattc catcctggag tggagatacc agaggagcag atattacctt
 18241 tattaactga tagaaatgtt tggggatttc tctgacttct ttaggggtg gataagcccc
 18301 aaagtgaaga gaattttgct ccttgtttta gccattagga aactcaagac cctgtacag
 18361 tgctattggt ttaatttttc ttaatcacat tgcctctgca acttctgaat ggttgcagcc
 40 18421 atttcttaaa atttccctgc attgtcactc agacaacaag aatagatttg gccttcttca
 18481 tctcaaaata atggtcatga ttaatagtta ttggactggg aacagtgtc agccctctgt
 18541 acgtgatctc aggaatcctc acagtactca atgaaatagc aattttatta tctcattttt
 18601 gcagacaaag caacggaaac ttccacacat tttctacatt gcacctaaga tcatctgaga
 18661 aactatgctg tacttgtttt tctaattgat gatctgattt ttctattata atgttaattc
 45 18721 tatgaggaca gggctttctg tggccttgct tcattgctgt atctccagca cctggactac
 18781 tgcattggcac ctggtagtta cttagttaaag gtttttcaa tgactgagta actcatcaa
 18841 gattaaatgt ctaggaagtg gtggcaccac gcttaggacg actcttttct gattccagag
 18901 tccagacagc cctaaccact atcccacact accttcttgt ttatttttaa atcattttcc
 18961 ttcccttcaa tcccttccca gtgccttaca ccttcttgct gtaatttgaa gcatggccac
 50 19021 agtaagctac ctcaagtttc tcatctgtaa aatggggata atataatgaa ctacctatg
 19081 ggattgtacc cctctgcatg gtagcctcat cctactgtgc ctctaacca cggcctttaa
 19141 atcagcaggt atagttaata tatttagttc ttttaatcta atctgaaaca caaagcattt
 19201 gcttccctaa ttcaagattt ttggctttgc ctgactaag cttaaaacca aagaagtact
 19261 gcagaactga ctgaggctgc cagaagtacc acactcttgc acccagccag tgggaagtgg
 55 19321 aaagataaca gctaagcctt tggggatcct tccagaagta gtgatgacgt acagcattct
 19381 ttctgattat gaagtaaata tctgttctaa tgtatgttca acatagagag ttaagaaat
 19441 ggggaaagaa taaagagtaa aacaatgacc agaaatacct tcaataccct ttgacattct

19501 ttctctgtgt gtgcatgtgt ttgtgtgtct ttgtttctgt gtctgtatat gtgtatttcc
 19561 tttatttttg tttttttact ttaatgtaat ttttagagac aaggctttac tttgtcttgt
 19621 agactcgcgt gtggtggcgt ggcactcatg gttcactgca acctccaatt cctgggctca
 19681 agcgatcctc ttacctcagc ctctctgtga gttaggactg caggcatgca tcaccattcc
 5 19741 tggctaattt ttcaattttt gtgaagacgg gctctcacta tgttactcag gatggctcga
 19801 aactcctggc ccgaagcaat cctatcacct tggcctccca aagtgcctggg attacaggcg
 19861 tgagcgacca tgcccagctc ccttttataa ataaggggct caccatacaa tataaccagt
 19921 ttttacctgg cattttccag tcattattgc attgtacgta tctccccatg tctttttctt
 19981 tttttttttt tttttttttt gttgacggag tctcactctc ttgcccgggt ggaaggcagt
 10 20041 ggcgcaatct cagctcactg caacctccgc ctcccgggtt caagtgttcc tctgacctca
 20101 gcctcccagag tagctgggac tacaggcgcc cgccaccacg cccagctaatt ttttgtattt
 20161 ttagtagaga cgggggtttca ccatgtcagc ccaggaaggc ctcgatctct tgacctcgtg
 20221 atccacccgc ctccggttcc caaagtgtcg gaattacagg cgtgagccac cgcgctggc
 20281 cctctccatg tctttaatta ttcttgcata agatgacttt tctactgata atattccatc
 15 20341 acataccact ctttaaccat ttgtctctg gggcacattt tccttttggg cacacttttt
 20401 atactacagt tgccatcctt ttacatacat ttgaatacac atatctgggt attctctcag
 20461 aatagatttc cagacgttac ctttccaagc ttgaagatgt taacatttta aagaaagatg
 20521 gttattcttg aaagccctga cagctctgag tggggagccg gggctgatgg ttaccacagg
 20581 atagcggaaa ggcacactgg ctggcctgtg tgtactcacg catccccca cctagggcag
 20 20641 ccttgggaag agcactcagg attatgagaa agactgtcgc ctcccctttg cttcattagc
 20701 tgatcctcta agcatatgtg ctttcttggg ctaattttcg gattggctct ctcctatatt
 20761 ctcttcctac tccccacccc gaccttacag ctaagtgcac atctcatgta gtgcagtggg
 20821 aaagaaccgt aaggcagaag ccgggctgac ttggctgtga atcccagctc catcacttgc
 20881 tggccagggtg actgagtaag atcgtttaca catccatcat cctcaagttt ctcactgtga
 25 20941 aaatggggat aataatgtaa ctgccttatg gcattatata aggattgtat gactgaacac
 21001 atgtagaatg cttagaacaa tgcctggcat atatgaagca ttaatacat ggtgtattaa
 21061 attagttttg aaaagaataa attaataaca atgatgaaca ttttggatac ctattttcct
 21121 attgttttga ctctcaaagc cagttgcaag catatttagc actgtgatgt atgtgtgact
 21181 tactgcaaa gcttttttcc agtccctgat accagctctc tcttcacctt cagtgtttcc
 30 21241 taccctcctt gcctcccctt ccctaagaat attgctgttt cacagagtgt aggctttcct
 21301 ctggcttcca gatctgccc catatgcaca cttctctttc ccacccctgt tggactcttt
 21361 ctcccttatca gtttatttgt tccagttggg aagaactgga acctggctcg cagcttttcc
 21421 agttggcttt atctgtgcgc tgcattgtaa aactgttctc tcttgcttag aaatctcttt
 21481 gatccatgtt tagctgtatt tattcttcca acagatgttt tgggtagtga gaggattttc
 35 21541 ttctcgcatt tgcctagtct catgctcctt catgcttccc acttgttcgg gatctttttg
 21601 ccagctgacc acagacaggg gccatctgtc gtgaaggctc cctgggcca gcagaccagg
 21661 aatggcccag caaccaagac tttctgaagg gcttagtgaa ggggaggagg gaggaagatg
 21721 ttggagaact gtgtagggtg gagtttgagt ttccagaga cattccagga gctcttttga
 21781 tccaagggtat acatgatttg gcttggtctc tgtggcagg taacaaaaac acaaccttcc
 40 21841 attgtctcct gtagacaaca gagtggggc cttgggcatg gcaggtagcc taagactacc
 21901 cctgagagtt gggaagtgtg tgagtctcct ggggctgccg taacaaagca ccacaaattg
 21961 ggtggcttag cacaacagaa atgtattgcc tcacagttct ggaggccaga agtccaagat
 22021 caaggttgcg gtcagggccg tgtttcatct gaaggcccag ggaagcagct gccccacgcc
 22081 ttctcctagc ctctggtagc ctctggcatt tcttggtta tagatgcac tgtcaaattc
 45 22141 tgtgtcttca tatggctttc tcctttgtct cacactgtct ttcctctgtg catgtctgtg
 22201 ttcagatgtc cttttttata aggatgtcaa cccaattgga ttaagttcta ccctaagat
 22261 ctatttttaa cttggttacc tctgtaaagt ccttatttcc aaataaggtc atgtgtcaa
 22321 gtactaaagg gttaggactc cagcatactt tgggtgtaga cacaattcaa ccataatgg
 22381 gaaggaaaga tgttgggcac ctgtaactcc tccaaacacc cacagagtgc aggggtgagct
 50 22441 gtgtgctaac acatagtcag ttctcttttg ggtgaggagg cctaggggca gggcccccatt
 22501 gtgggtctc tgtccacacc agcaacaata acaaccaggg aggaagcat ctatttttcc
 22561 ttggctcagt tcagcttttt atgtttttag cacaatgcct gctttgctct tccaacaatt
 22621 tgggaatctc tgggagctgt gcatggaaag caaggaggac agcggcgaga aaaaggggga
 22681 gtagatggag ggtcttgga agcagagggc ctaggcaggc agagaggaca ggaaagtata
 55 22741 gcgagcagag cggcaaattg gtggggagg gcagaaggct gcttggcagc caggagttct
 22801 tgccctggcc ctgccatgag gctgcagtgc tgtggcctag gtattttacct tctccaggcc
 22861 tcagtttctc tgtaggcaag attgggaggg ggatgggtgc tctctaggat cccttctggt

	22921	ccagaataaac	attctcagca	ggagcctaac	gtgtggagca	aatgggagca	ctgggctccg
	22981	gcctcctgca	gtgagcacag	cccctgttct	tgtggaaaca	tcttccaata	gggctgccct
	23041	gcctacagg	tcatgcggca	tgcattctgt	gcctgcctgc	gctcttgaaa	cagcctccac
5	23101	tgctcccctc	ccagctcctg	tctctctgca	cacgcaagcg	tgctactcct	tttcatgata
	23161	cccattagta	ttcttttgacg	atggcataca	tctgtcttcg	atcgttgtca	gctctgggag
	23221	gcttatgcca	agcttcttga	gcgtaaccca	tgactgcctg	ggtttagtgt	tgtgagctgt
	23281	ccaggaggca	ggaggacgat	gcatgcaagt	cagggccttag	ggcagaagtg	cctgggcctg
	23341	gcctcccctt	ggactccagg	agtcctgtcc	taacagagcc	cacagcccc	tatccatctg
10	23401	gcctctgtaa	cccctcccca	acacacacac	acacacacac	acacacacac	acacacacac
	23461	acacacacac	atagcccctg	tgattgaggg	ggccccaatt	cctgttcata	tcctccagga
	23521	tagcccacct	gcacccctga	cagtgaagga	caaagtctta	ttccctgttt	agatgggtgc
	23581	tggggacaat	ggaaaggagg	tgtggctctg	agaagtctat	gtcttgctca	gggcacacag
	23641	cagctgatcg	ggaacatgtt	gctgactcca	agatgctgcc	ttgcaagaag	ctggctctat
15	23701	ccttcttttg	gctgaagtgc	ctttcatgga	tggtgagggg	tgtgagggg	gaagtgtcag
	23761	gagtgaaggt	cagtgggttag	aatcaggcag	tccacagagt	ctgagaaaag	aagacattct
	23821	ctggcagctc	gggggtcatg	atcgcccacc	ccagcccaga	taaccctcac	agctgtgcgg
	23881	gccactagag	aaaaaggagg	gcatgtttgg	ggcaggagag	gcaaatgttt	gcttatctgt
	23941	gacttcttcc	tccaagcatg	tccggacctc	cagtcaatgg	tgggctgtca	tggtcagct
20	24001	gaggttgagc	tttccttagc	aggagcactg	gtcacttggg	ctgggtgggt	tcttagtggt
	24061	acaggatgca	ctgcaagctt	taaatgcgag	tggtcatcatc	cccttccggt	caccatggca
	24121	accagaaaca	ccttgacaca	tttccaaatg	ccctttagta	gggcagtgac	agcccttttg
	24181	agaatcacat	agaatcgcac	tgattgatga	gtgaaaaata	aatggatggg	agcctccttt
	24241	tgtgattttt	gcagcggcct	ttagcttcct	ttactcacc	cagaaatcag	tgggaccctg
25	24301	ggagctgtgt	acccctcaga	cccagttgga	acccagccaa	gagtacttaa	tccatcccca
	24361	cttgtggggc	caacggcacc	taaccacctc	aggcacgggt	gacctgggtc	ctcagagagc
	24421	tctagggaca	gaggagagaa	agggtctgca	ttctgtttgc	agccctgata	gtgagctctg
	24481	ggggctctct	tccaccccca	ccccacccc	cagccctgg	agcaggtact	cggggtcaga
	24541	gctctgtga	gggtctggct	ctgggagggg	aggtttgtgt	aagattccct	cccacgggtc
30	24601	agcacagatg	ggatgacaag	gaccaaattc	tgtttctggg	ctctgatatt	tgccaagatt
	24661	tttaccaggc	ttcctggaat	agacagggaa	gcagagcaag	ctcccgtagg	tcaagtgatt
	24721	tggggccgag	ttgaccacga	gtcccataat	gactgctgtg	tagctacat	gagtgtgctg
	24781	agtggcccat	aggggcaggt	atgagagagg	tgctgagggg	ggcagggggc	ccgcagaacg
	24841	gcctcccatc	tccactgccc	gtcccaggt	ccacaggctc	acagagcaga	cacggtctgt
35	24901	gcctgggttt	gctcaccac	aagaggaaga	acataacatc	tcgctccttt	tgctgcacag
	24961	gataaaacga	gaacagagag	gaaacaggaa	gtgctttgca	ttccagaaaag	agcagcaact
	25021	gtataaaagtc	atgcatatta	ggatttgagg	tatgcatggt	cagaagttag	aaactaacgg
	25081	aatcttgtca	ttgccaggaa	gtttcggggg	tctgtgactg	gtggccactg	atgttctgtg
	25141	gttctctcat	tccagctcct	accttgactg	tgctctctc	ttcacacctc	acttctttag
40	25201	tgaagctcc	tttctctcat	ttcctgttca	atgcttaaac	cccttgcaat	ctggcttcta
	25261	ccatgccttt	atcacagacc	cttctctggc	tttgcttgc	ctggcccttc	catgatgtct
	25321	tctttcttga	aacctctttt	ccttggtagg	ataccacggc	atcctgggtt	ttgtcctacc
	25381	tctgtggctg	tttctgctg	ctttccttag	ctgacttttg	ctcctttatc	tgacctgggc
	25441	tctctctct	ccatacactc	tccatagcct	attctaagtg	tcccagggtc	cttatatctt
45	25501	atctctcaaa	tgacacatta	ctttgtgtta	gctacagacc	catatatcca	gcttccctat
	25561	agatacccca	aatgtctttg	taggctccta	aactcagtg	atgctaagct	gaacaggggc
	25621	tccttttgtg	ccccaaactt	tcccaactcc	agtgagctt	ctccattgtc	tttcacctta
	25681	ataaatggaa	ccacctgcga	ctctagtgtg	tggtccagag	acttgggaag	cacctcagct
	25741	tgtctctctg	tcattccgag	gatcgggcag	cctccaagtc	ctcatcatc	taactctcat
50	25801	aatgcctctg	gagtttgtcc	agatctctcc	gtcaccactg	ccgctacgct	aatcaaaacc
	25861	accattgtct	cttgccatcc	tccatacttt	gcaaagttaa	ttgggtcatt	tttctacttg
	25921	aaatcttata	atggctctcc	agtgcctctg	agtccttggt	ttttcaacac	tgttcacact
	25981	ctccccacct	ctctctcact	catacccat	gcaccagcca	tcctgggttt	tgctgttttt
	26041	gtttcccgag	atgcacaatg	caccttctgg	cctctgagcc	acagcacctg	ggtatttgct
55	26101	cacactactg	cagctctcct	cttccctgcc	accacgcctt	tcttgcttga	ctgttagtat
	26161	gcagcagtg	ccacttgagc	atgaccgcct	ctgggaggct	gtccctagtc	ctctgtcgca
	26221	ttctggggct	ccctatcaca	cactcccat	gcatgctgca	gcattctcag	caccagcat
	26281	tccttattgt	agttcctgat	tcaacacctt	tctcaggaga	ctctggactt	cttgaaggca

	26341	ggaacaattc	ccacttggtc	ctagtagcat	tccaaaccac	acgtgacagt	gtctggctta
	26401	taataagcag	ccaataaaaa	gttgatgaat	gaatgaataa	gtgaaaacag	aaggtgtttg
	26461	cctgcagaaa	tctggaataa	gatcaaagat	cagagctggg	attaaggaaa	aaacttcctt
5	26521	ggggtggcac	tatgaattcc	cagaacaggt	gactaaccct	cattcacttt	ggcaaatgtt
	26581	tatcccatgc	cacgcaacca	ggcaaaaagt	tgaatgaggt	ttaatccctt	cccacacgga
	26641	gcttattcct	tctttggaag	tcctttaaac	aagctctgaa	atgatttttg	caggtagaca
	26701	aactggtcct	catttctctg	tgaccagtaa	gtagggaaa	caagcacaca	tacacacaca
	26761	cacacacacg	tgacacacga	cactgacaga	caaccttgct	cactcacatg	ggcatgcca
10	26821	aaccttcttt	ctattttata	ggatggtaac	tcactcttta	gttttagactc	ttgacgtgcc
	26881	atggaaaatc	ccactcgccc	tagaactggg	ggccggggcag	gtttgactgt	aacaacgaag
	26941	cctggagcct	actctttgct	gattggcctt	cctttctgtc	tccatttttc	ccctggtgag
	27001	cactgcagtt	gtgttcttcc	tcccaaaggt	aatgcctggt	ttggctcact	aaaacctgtt
	27061	ctttctgtac	cgagagctca	tcttctcttc	ctcttctgga	ttctcaaagt	agatgacgtc
15	27121	agaggatgga	ggccaaccac	accttccttc	cttgaccctg	ataaagtttc	ttggaaaccc
	27181	tatactcaga	ggcagccaat	tcttgccagt	ggaagagtga	aaagagggct	tgggaagctc
	27241	aaggctcagt	gtctgtcccc	aggtccccc	gttaaagaca	catctgtcct	tcactctcaa
	27301	agatgttgcc	attgtcctcc	tgctagagt	acacactgca	ctccctcctt	cccttcacac
	27361	cccagcaaga	ggctatttcc	caggggtcct	ataagcagat	ttcatcttct	cttgtgtgt
20	27421	tttcttattt	caattatctt	cagggaagg	cggtcatatt	gcgtcattgc	ctggctgtga
	27481	aatttcat	ccatttcttt	acacctgcag	ttgcaatacg	agagagaaaa	ggccagagct
	27541	tagcggatgt	cctagacgca	ggttatcaag	gtgctgtggc	tgtggtttcc	cggaaaaggc
	27601	cttggtccca	gagcacattt	tatcagcagg	accttcgagg	ggctgcgttc	cttcaattgt
	27661	tttctctttg	gggtctctgg	tctccagttc	tttcttctct	agcatgtgag	atctgtgctt
25	27721	ttgattcatg	cctttaagtc	tgacattgaa	aaaatatctg	atttgccatt	ccagatgctc
	27781	gtcctcattt	gcaaattttc	ctaaagggcc	agattgtcct	ctggcctttt	cccttttctt
	27841	ggtcccacct	caccacccct	ccactggggc	ttcacagagg	cagagctagt	ctcctttcat
	27901	tttttaaaaa	taatagtctt	caatttttag	aacagtttta	agttcacaga	aaaattaac
	27961	agctattaca	gagttaccgt	ataactcctc	cccctcactc	cccagttttc	tccattatta
30	28021	gcatgttgca	ttagtgcagt	aaatttggtg	cagttaataa	gcaaataatta	gcccattcatt
	28081	attaactcaa	gtctatagtt	tacattaagg	tgtattcttt	gtgttttaca	gttttatggg
	28141	ttatgacaaa	tgcataatgt	tatgtatcca	ccattatagc	atacagaata	gtttcactgc
	28201	cctaaaactc	tcctgtgctc	cacctgcca	tcctctctcc	ctcctctgcc	accaatccct
	28261	ggcagccacc	agtcttttga	ctgcctagag	tttcgccttt	ttcagaatat	catagtagtt
35	28321	ggaataatac	agtgcgtagt	cttttcagac	tggtcccttt	cacttagcaa	tatgcatttt
	28381	aagtttcttc	catggaaact	ttgctttcat	ccttttatca	ccacaaggcc	agtcattcaa
	28441	ggaatttctc	catctctgtc	tgttcctttc	tagttctatg	tgtgccactg	cttggcatag
	28501	aataggtatc	catttaatga	acattccctt	tcaccacctg	ggacaccttc	ccagggataa
	28561	caaaaataaa	accagctagg	tcaatgacag	agccccatc	ccagttttta	cctcattctc
40	28621	ccctcttttc	acaataaact	ggatcagaac	cagcagctct	gtaagactgc	atttcttttc
	28681	cttaatacca	ggccccagag	agcatttgat	tccttggcag	agaggtgtag	gcttaattaa
	28741	tttttctcct	tttttctttg	aacatcttgg	aacacacaca	cacacattcg	catttatgca
	28801	caattgggtg	tatgagaatt	ttaatggcag	gtgggtgttag	cagttctttt	cctcctgata
	28861	cagatcaggg	tttttccatc	tgggcctttt	agcagggcct	atgaatattg	actttctaac
45	28921	cacttggaatt	tgggtggagt	gtgcagagt	ggattggggg	ggaaggttca	agggagaact
	28981	atacttatgt	ataaatcaca	tgtgaaggga	gttttgaagt	cattattgct	tcaggatgtg
	29041	cgaaccataa	ttattttttt	aggtcttgat	ttgcccagg	agcatttccc	agggttgctg
	29101	ctccaagcat	gacgtctgtg	ctgtcaggag	gtgcagcata	gtctgattcg	agtttaaatcg
	29161	ctttaaagga	ggccctgggt	aggatctggg	ctctagggtt	tcagctgtgg	tcagtcctcc
50	29221	atgcagcaaa	acatccagat	gacttagatg	attaagacag	cagacttaaa	gtgaagaaga
	29281	gatttttttc	cttattcttt	ccttttatta	ttagttttta	aatggttggc	tacatgggct
	29341	gttggtcatt	ctccatgttc	tctgtgctct	cctcagctct	ctgctcaaaa	caggctgcac
	29401	cggcctgcct	aaacctgaa	agcaacttct	cagctgccta	ctttctgcct	tttgaccccc
	29461	aagccaatcc	ccatctcctt	accacctcc	cgccatgtcc	tcatacacct	gcctctcctt
55	29521	gacttcattc	ttcatgctcc	atcagcaaca	gccctctgtc	aataatgatt	gtcccaggga
	29581	agtgtattca	agggtcacat	aaaatgtgcc	ctctctatgt	gttgagaagg	ttttctgtcc
	29641	ccaaaggagc	tctctggata	atgaggaagg	ttgaactggg	gcagcctaca	ggaagaagcc
	29701	cttagaaggg	aaacctgtgg	cataaacat	gctgatccac	gactcttatt	ttggaatagc

	29761	tattttaaaaa	gaaatatgaa	gaactcgtaa	gacttggaaa	agaaaactag	agaatgttga
	29821	aaatgtccaa	gggttatgtg	tatgatgtgt	atggggaaat	tttaaaaagaa	tggtgtagaa
	29881	aactgaattt	gtggtaaaa	gtgtgcacag	gacggcctgt	tctttcattg	aattatgtct
5	29941	tagtgcaggg	atgacaaaata	aatacagcaa	ctgtgctgcc	attctcacat	cttttcctac
	30001	agtaggcac	actaatcaat	tacaacattc	ttttccactg	tcatgccttt	gcctacagca
	30061	gacatcacta	atcaatcaca	gcactctctc	tcattgagtt	tcaaagggtt	tttaatcctc
	30121	cacataggtc	tctaagtagc	tatataccaa	ttactttgat	ttagaacttg	gattaccaat
	30181	ttgccatctc	tagtttaatg	ggaagaacat	taggattaga	gtcagaatac	ctgaattcaa
10	30241	gttactgacc	tactacttaa	taataattaa	ttttaataat	tctctaacct	cagctttctc
	30301	ctctataata	agaaaataatg	cttaccacag	aggctcgagg	taaggattaa	gtaagataat
	30361	atataacatg	taagataaatt	tgtgtgatgg	ttcctagaat	agtacctggt	gcattgtaag
	30421	cacccagcaa	gtgatagcca	agtatgaatt	tgcagcggca	tggacacatc	atgccacacc
	30481	ccagatatgg	aacaatagtt	aggattccat	tcacctttgc	ttcttttaac	atcctcagtg
15	30541	aaggcagatg	gagaacagct	ggaccctcta	attctacctg	actttaactt	cacattctta
	30601	agatgcttat	taaatctctc	tttcctgaac	taatataatg	cttccttggt	accttggaag
	30661	gacagagtta	acacccatgt	attatgacat	gatggatggt	ctttttgctc	acactgcatg
	30721	catagccatg	tgccaaggcc	aggcctgcct	tagctttttg	gttctccaga	gcagcttcac
	30781	tgtgactgaa	gagagccagg	gaagacatct	tggtagagct	cttttatga	ctcttttcag
20	30841	aatgtttctc	actgatggag	atgacagaaa	cttaggatga	tttgaggcca	ggaggacagg
	30901	cttcttttga	aaaggtttca	ccatagatta	cctcaatcca	gggttggaag	actggaaagt
	30961	ggtctgcagg	gcatcccagg	ggatgccttc	ttcaagacaa	tattagacta	gcattggacc
	31021	ctgcctccag	actagtaaa	gggtattttc	agagcagcaa	caccagagaa	catgttttga
	31081	gcaaatcaaa	tccttaaaat	catgatttca	tacatcctga	gatactagtg	acttcaaaat
25	31141	gccttttctc	aattatacat	atagtgcctt	ctgaatcaac	tgtctttttt	ctgtccaaca
	31201	gtataaacat	aatctctgct	ctcacacagc	ttagaaaactg	tcatgagagg	taaacaaata
	31261	cacaaatgac	aaggtaaaat	agaagagcta	caaaaagagat	catgaaaata	tccaggaggat
	31321	tcgaagggtg	cagataacaa	aaggcttcgt	ggaggaggta	gcctttgtgg	catttgagat
	31381	cagtcctgca	gcagggttat	aggtggacag	gcagagatgg	ttgggggtgg	gcggtaggag
30	31441	ggacagacag	aaggaaccac	aagaatgagg	atatggagaa	aaagaccact	gaagggtgca
	31501	aattattaa	gatggattta	tgtaaaggat	gtggcctaag	aagaaggact	cttctttatg
	31561	atgaaattag	gttttagagg	cagaagtctg	gtgcttccc	gtactaaaat	gacccagta
	31621	tggtgtctgc	agaatactta	gataactgca	tagatggttc	agccttcttt	ccattatacc
	31681	acactactgg	gttaccaact	tgctgtgtga	ccttagataa	gcaactacct	ctctctgggc
35	31741	ctcagttttc	atatctatga	aattaggagg	ttgaattatg	tcttggcata	ataacgaaat
	31801	aaataagtga	tttattttta	tttctttgtc	attcctcttt	gaaatgggag	tgggaataaa
	31861	gtggtatttg	tttccacatg	aaaattaaag	ccagggggcc	aattctggac	ttggcttgag
	31921	ctgtgctttg	tctgtttttc	tgactggcc	cagtaccctt	tacactgggt	taactaactc
	31981	cctggaaagg	gatgcaggta	gacagcgtga	ctgttccttc	ctatctgaga	ggcccaaca
40	32041	gattctatat	tgacagacca	ccaacttggg	atttcagctc	agttcagcaa	acacttttga
	32101	ctgggttcgtg	tgctagacat	tgtgctgggc	ttatgtgggg	tagagatgcc	tgctaactta
	32161	aggctaacag	gaagacaggg	aagtggtttc	cagacctgat	gcttatcaga	gtcactgtgg
	32221	aagaactttt	tagaatgcag	acagcccca	acttaacagt	ggttggaact	tcaatttttc
	32281	accttataat	ggtgcaaaag	tgatgggcat	tcagtatgct	tctcaacata	caaagagggt
45	32341	gtgtgaaata	aacctatcat	taattgaaga	tactgtcaat	caaaaatgta	cttttgactt
	32401	acagtgtttt	caacttagga	ggggtcatag	gttgaggagc	atctccacag	attcctgagc
	32461	cccatccctt	gatgatgatg	atccagttag	cctgggtggg	gcctgagaat	ctgcatacta
	32521	actgctgaag	gtgagtgtga	tgacgggcca	agcttcagaa	cctctgatct	agagaagatt
	32581	ctgtgtcatc	acagctcagg	gtgacctagt	tctctcttac	ttgcacttac	atgaattcat
50	32641	atgaattcag	tattacatga	gtaaattatg	attatatgaa	tttattgagt	atcctatgtg
	32701	cttgatatag	gtgtttacat	cccagggttg	gggggtagag	atgaagaata	taagccggac
	32761	tatgttaaga	tcttcacttc	tcagcacctg	aacaactggc	agcctcatta	ggaaggcaag
	32821	tcattcacag	gtgaaatgac	atcatggctg	tcttcttcat	tcttctcttc	agcatccatt
	32881	taagactcac	attttatcac	cagagactga	aaagagccac	ctaaggcagg	caggtcaggt
	32941	ggtgttatct	ctttattttc	agatgtggag	gctgaggctc	agagagggtg	atccatatgt
55	33001	ccaagctcac	atctcccgcc	ctcagtcctg	ggcttctccc	cacttcatgg	gagaagcatc
	33061	ctcctcccca	gagcagtagg	ttctggagct	gggagaggcc	actgtgggct	ggattgttgg
	33121	ggacagcttc	aggagagacc	cgattcaagg	caatagagaa	ctttggctgc	aggccgttgc

	33181	ctagaatagg	gcagctgaca	cacctttgat	ctggaatgat	tcctgctgct	gagaatgagg
	33241	ttttttatat	ctggattctc	aggtagtaac	accacgacaa	cggtgtgtttg	tgttattttca
	33301	tctgcacagc	attcacgtgt	agcagagaga	aggtagttta	ttccccagaa	gtttaccagt
5	33361	gggaaattga	ggtcaaaggt	aggactttcc	taaatgaaga	gaactactat	ttattggatg
	33421	cctccccatct	gccaaagcagt	gtgcacagca	ggcatatatc	atgaaatgcc	tcattttattc
	33481	ctcataactg	tagcctgatt	ttacattcgg	caaacctgag	gcttacagaa	atcatgtggc
	33541	tcgcactcag	actgctgatg	gccacgcaag	gctttcagct	tctctgattt	ccaggccccc
	33601	atccccacca	cactgtgctg	tccagctcac	ctgggtggaac	tggtatccttt	gagttccagg
10	33661	ccagggatcc	tactgagctc	ttccatcagg	gaaaccatag	cagtagcagc	tccaccaagg
	33721	acttggcatc	tatcttactg	cagcatccgt	gcctgtctaa	tggaaccatg	taccaggagt
	33781	agctacccaa	gaaacatttc	caccagaaac	ttctctttat	agctcccat	cgagcctcag
	33841	agagctgata	ggaattgctc	aagaccaccc	aacttgtaag	tggtagagct	gggactaaaa
	33901	ggcaagtccc	ccaactttca	gcctcactcg	tgcccaatgt	gtctcagcct	ccctgaagaa
15	33961	tatcagacca	agatggccag	aaaaggaacc	tggaagggac	gtgtgggct	catgcagccc
	34021	ttggcacagt	cttcaggctg	agctgcctcc	accttgtcat	ctcatcagag	ctttacatcc
	34081	atccttggga	gaggccagct	ccatgacctc	tcagtgtcat	ttaggatctt	cttcctaaac
	34141	ggctaaccce	agtctgtccg	ctgcatccta	cagtgaagtg	cactgtggcg	agagcagta
	34201	ttggcattct	gcttacgtgc	tgtttcagaa	ggtaaaactca	gtataaatgg	atctacagcc
20	34261	tgtcccattt	tgtacagctc	agttctaaaa	ccagcctgca	aggataatg	tataaaaatg
	34321	tcctgccagc	cccaaggggtg	tcttctacaa	caagttcttg	ttccctcata	attcttctga
	34381	caaattcttc	ttttgatcca	gactttccta	gccatcattt	cattccagaa	gtggcatgtg
	34441	tgtgcaaaca	ctcatgcttg	agggcgggac	aagcaaagg	ataagagggg	aaatgggaac
	34501	tcgaaatctg	ctcaaatgtg	gtaaagaaat	atatccagaa	agtactgcct	actcaccaaa
25	34561	atactatttg	ctttttatat	tctttcctga	gtagaatttc	ctgttcaaac	ttgaaaatga
	34621	aaattcctcc	acttcaaaat	gaacaggcaa	gaaatgctgt	aggctgggtt	tcccggggag
	34681	gggggactga	cagccatctg	ccctgagact	gactgtcaaa	tctgaactct	gtgtacttgt
	34741	tagtgttgtt	tatgggaggg	gtgagggag	ggaggaagca	acagggacct	gctaacccta
	34801	tgaattctcc	ctcataccct	taaaaagtcg	ggtgcttgcc	cgggtgcagt	ggctcacgcc
30	34861	tgaattcca	acacttggga	ggccaggggc	ggttgatcac	aaggctcagga	gttcaagacc
	34921	agcctggcca	agatggtgaa	accccgctct	tactaaaaat	acaaaaatta	gccaggcgta
	34981	ctggcaggcg	ccggtaatcc	cagctactcg	ggaggctgag	gcagagaatt	gcttgaaccc
	35041	aggaagtga	ggttgcaatg	agccgagatc	atgccactgc	actccaacct	gtgcaacaga
	35101	gcgagactcc	gtctcaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	gtcgggggct
35	35161	tgccaagctc	gctttttggg	gcaatgggag	ggaaatttta	agagctgatt	tctgtgatct
	35221	ttatcagacc	tcttcttctg	tcctgccac	aggctcagtga	tgggaaatgat	agtattgcat
	35281	taaaagagaa	gtgattcccc	agatgtggac	atctcagtgc	ctgggaggca	gacgtcttta
	35341	gagtgtttgg	tgttcctccc	catactaaac	ttacctgttc	ttaggacata	ttcccttgag
	35401	tgtcttttat	tatttcgttg	attgattaag	ggttacaaat	ttgttaaaaa	agcagcagaa
40	35461	attttgaaag	ataaaaacag	tagaaaactgt	tatttaattg	gaggcacaaa	agccagttcc
	35521	agcttgagca	ttaactggct	gtgtggcttg	ggcagctccc	ctgcactttt	gggaccatgg
	35581	tttcttcagc	tctaaagtga	aagctgagag	ttcctctgaa	tgctaaaatc	ttgcttctct
	35641	gtgattattg	ttttttaaat	gccgcccgtca	tttcaaaaca	catacatggg	gcagccctcc
	35701	tcagctgcct	gtcctgttct	cttccctcct	tccctttgtg	ccatcagcgt	ctcccctgga
45	35761	tttatcgtgc	tgtactcctt	ggacaccctt	gaaagtggag	agagaataat	tctcattccc
	35821	tttccccaaa	ctctgtcctg	aagcaccact	ccccacctcc	ctaccatgcc	tgcccctcca
	35881	ccccccacat	tttccctctg	atgagatttc	ctctctgagg	gaatattttg	gtcctgttct
	35941	ctgttcccaa	acaaattccc	caggtccttt	aaatctgggt	tcaatcaaca	taatcacagt
	36001	ctcttcgatg	ctatcaactc	tccttgatt	tggttggtg	attcatgctg	tgtcctccac
50	36061	ctcctcaggc	tgttttatat	atacacactt	aagtgtcat	atatatgtgt	atatattaca
	36121	caggttgtat	gtatatgatg	tattgaaaga	gagagatgtt	ttccaccgtg	cctgtaatta
	36181	ctttttatct	tcctccaaat	cagtatctta	atgcatatca	cctaaactgg	attcattttg
	36241	agaacactag	ggcctttttt	ttacttttca	agaacagcta	tctaattctg	ggccaaaaac
	36301	tagttatcaa	atgagggcag	gtaatggagt	tgtatccagc	gagggggcat	ttgcttctcc
	36361	atatgatata	ctcctcacca	gtaccaaggc	gttttctctt	cagtcttcat	gcctatgctg
55	36421	ttacgagtcc	tttctcctat	ttgaaagaaa	agatatgagg	cagcccaggg	aaagctctga
	36481	tggaggctat	aaagacaata	tggtatgtaa	ctaaaaaatg	gtagcttaga	gcttaatggt
	36541	agcctctcag	gacactcacc	acacctgaat	tctaccagcc	tgacgccaat	gctgctccgc

	36601	ttatgctggg	atgtagtgac	acaggacact	ctgctggggg	atagggatga	tctcccccca
	36661	ccaagtgggt	catcaacact	catcactcag	aggggaggtg	ggataccacc	ttgaacagag
	36721	aaagcggctg	ggcatgggtg	ctcacacctg	tagtcccagc	accttgggag	cccagggcag
5	36781	gtggatcgct	ttagcccagg	agtaagagac	cagcctgggc	aacatggcaa	aaccccatct
	36841	tcattaaaaa	tagaaacatt	agcctggcct	ggggttgcac	acctatagtc	ccagctactg
	36901	gggaggctga	ggtgggagga	tcgcttgagc	ctagggggtc	gaggctgcag	tgagccatga
	36961	ttgtgccatt	gaactcctgc	ctaggagaca	gagtgcagtc	ctgtctcaaa	aaaaaaaaaga
	37021	ggaagaagtt	gctgaatccg	tcttataaat	ctgtaacaga	aaattatggc	aagtgtggc
10	37081	ttttaattgc	cagagctggg	cacttaagag	agaaaagggt	ttttgtgaat	tccaaaaatg
	37141	taattgtatt	cattgaacac	tgaagtcagg	agaagtcaga	ccataatgac	atgggagcca
	37201	ttttcgcaac	cagcagacta	gaaggggagg	gtttggaagg	gtggcggcag	agcttagaca
	37261	cctgctgcga	agggagagaa	gtggggacag	aaagatggcc	tgtgtagagt	cccatgagaa
	37321	agacagaact	gcactggcag	gcactcttag	gggcccaggt	cacaatcatg	gggccggtgg
15	37381	acagtctcca	gggcactgtg	attgtcacag	tgcacagcct	aatggggaaa	attgcacagc
	37441	ttcacttaaa	tataggtgac	atatggacgt	aagaattcat	gattaatcta	gaactaacc
	37501	tgaccagcaa	ggccgaatga	agaaaaaatg	tagaaacact	gaccaaacct	tcctcaaagg
	37561	tcacagatct	tagggaatgt	gtctctcttt	cacatttcaa	aataacaaca	atttttgaaa
	37621	tatgtatttt	ttaacattta	attgttttat	ttggggataa	ttttgactc	atggcatcca
20	37681	ctgttttttt	agttaccac	atatttgatt	cagtaagatt	ccctgtattg	atacaaaaaa
	37741	agaagaggga	ccctgtttcc	ctttgacact	gttacttttt	cctgagctca	ttgtttgttt
	37801	gtatttttgt	ttgttttgtt	ttggtttttg	agatgtagtc	ttgctctgtc	acccaggcta
	37861	gagtgcgaat	gcgcgatctc	agctcactgc	aagctccgcc	tcccgggttc	actccattct
	37921	cctacctcag	cctcctgagt	agctgggact	acaggtgccc	gccaccatgc	ctggctaatt
25	37981	ttttgtattt	ttagtagaga	tggggattca	cggtgttagc	catgatggtc	ttgatctcct
	38041	gacctcgtga	tccaccggcc	tcggcctccc	aaagtgcctg	gattacaggc	gtgagccact
	38101	gcgctcggcc	tcctgagctc	attttaagag	agacttctgg	cctagagggt	tgaatgagaa
	38161	gaagtataac	agctgggatt	cttccctttc	tctgatatga	ggacaggagt	tctctctcat
	38221	ctccgccaag	agcaggaagc	tggagtaact	gccacaagct	ccaggaggga	gtgtctagaa
30	38281	catccacgtt	ttgcagcagg	aaaacacccc	ctcacgctga	agtttgattc	ctgaatcctg
	38341	tgtcgcagtc	taaatgctga	ggcagaaggg	gacatccgtg	ttcctggggc	attccacttg
	38401	cagtcctggc	tgtaaccoga	gtgagccatc	cggtgtagttc	ctgttgctaa	gtctcccctg
	38461	ccacctcttc	ttcccatggc	tgacgggcag	ggggccatgc	cctcctcttc	atttctgtc
	38521	cctgggtgag	cggtccccct	gccttctccc	agatctctgc	tgtggcagct	tcacgtggga
35	38581	ttcagcactg	tgtctccttc	ccctctgctc	ggcctgcccc	tacctgtcca	gcagagctgt
	38641	aagaccagaa	gacagagcat	tccccttatc	tatgaagtca	aatgcatgtg	tggaaatgc
	38701	caccagcct	gcagtctctc	tactataaaa	tactgtcat	aagacaaatg	tgtggcccag
	38761	atgatttctg	ataaagtcta	ttattttgaa	atacatatgt	atgtctctca	gccactgata
	38821	cacgcagaag	ctgcacatgt	tgtcaacacc	tgctttgagc	tcctttcctt	cccacccttc
40	38881	cttttggcaa	tgcaagtttc	cattctattt	ctgcattact	ggtctcctct	cttctcccct
	38941	actactagat	cttacaataa	acatttgaaa	tagtttattt	gtccacagtg	tgatttctgt
	39001	tgagaaacat	gggctcaccg	acttttgggt	ctcttctaac	actagaaatt	cctctggttt
	39061	tttagactat	ttcaagggtc	attaatgtgg	aacagacggc	ctttagaaac	agcaatccac
	39121	agggggcccg	gagacctgga	ttcatatttc	tcctagcgtc	aactagtggg	tgaccttaga
45	39181	aatgtcattt	tccttgctcg	gcttttagttc	ccttatctgt	cacacagaag	cactgtgtga
	39241	gtttgggaaa	ccaatactat	gttgaaatgt	acaaaataat	cttaaagcac	agatattctg
	39301	ttccctccaa	gaatacatca	aacaaaggaa	ctgacattgc	aagaagattt	tgaggagatg
	39361	gctggatgca	ctggagcagg	gattgctgag	ggaagccagg	ccctcacctg	gagcgtctca
	39421	ggagaggcag	cttcggtgct	ggctgtttat	tgacggcatc	tctttctgtg	tctgtgcgtt
50	39481	taggggctct	tctttggaga	taagaaaagg	gttctggatg	gagggcagtg	aagaacagtg
	39541	agaacttaac	atgaggatgt	ttgtatagag	gggaagactt	ctggacagtg	gcttgacttt
	39601	gctcactggg	catttctctc	tggatctctg	tagaagtcag	ggacagatct	cctcgtgcat
	39661	atctgtctcc	cagagacaga	tctcttcta	gcagaaagta	gaaagtgggc	ttcaggcatc
	39721	ctggaagttt	tctttcttgg	tgggtgataa	aagggcttgc	agagagagga	gaatcaaatc
55	39781	tcccacatgt	gcacatgccc	tgcgagtctc	atgcagagat	gtcttatgtt	caacatagaa
	39841	agcaagcctg	gcagcccaa	gaccttctc	tgacaccgt	ccatttttac	ttggtttcat
	39901	tttgataaact	gtgcggtctg	aggtcctggc	caagaaagca	tcacctggca	agaagtgtat
	39961	ttggccaatg	gtaaggttac	catctctgtg	taattaggct	ccgtaaagct	ttgtttttaa

	40021	at tt t a t t a a t	g g g a a t g a t t	t g a c a t t c c t	a c a c a c t g a c	a t t a c c c t c a	t g g a a t g g a t
	40081	a a g a a t c t c a	a g g c t t g t t g	g g t g a a a g a a	g g g c a g t g t t	t g g t g t g a c g	g g a a g g g a a a
	40141	g t a t a a g c a g	g c a g c t c g t g	c g c a t g a g c a	t t t g g g a a a c	a g a a c a g a a a	t c a t a g a a t g
	40201	g c a g g c t t a a	t t c t a g c t c t	g t c a c c t a c t	g g c t g t t t g t	c a t t a g a a a a	a t t a t t t a c c
5	40261	c t t c g t g a g a	t t c a g t t t c c	t t a c a t t t a a	a a t a a a g a a a	a t a t t c g t c c	t c a t a t t g a a
	40321	a t g a a t t g g g	c t a t c a t g a t	c a a c t t t a a a	a t a c a a c g a a	c a g t a t a a a t	g t c a g g a a t t
	40381	a t a t g a c a t t	c g g g a c c t c c	a c t g c c a c c c	t c a c t t t t c t c	c c t c c a g t g g	t c a c t t a c t c
	40441	t c t g t c c c t c	t t t c t g g g t c	a g a g c t t c t g	t t a c t c c a g c	c t g g g c c t g c	c t t a a g t g g g
	40501	g a c a t g t g c t	g a t c c c t c a c	a a t g c c g g g t	g a c a a g g a g g	g t t t t c a a g g	c t g g c t t g a c
10	40561	t g c c a c t g c t	g g t c t c t c t t	c t c a t t t g c a	a c t g t c t t c t	c c c c g c t g g g	c t c a g t g t t c
	40621	c t g g g a g g g t	g a t g c t g a g g	g a g a a a g c t g	t g g c a g a g g g	a c g t g g c a g g	g t c a g a g a c c
	40681	a c t g a t t c g c	a g g a g c t g g c	c t c a g a g c t a	g c c t t t t t g c	a t t g a t c t a g	g g a a c c a g t g
	40741	a t c a t a g a t a	t c t a t g t t g a	c g c c t g t g t c	a a t t t a t c c c	t a g a g c c a t t	a t t c a g t g a a
	40801	t t t c c t a a g g	g g a a a a c a a t	t c t c c a g c a t	t a t t t t t c t t	t g g g c c a g g a	g a g c a c t t t c
15	40861	t c t g a g t t t t	a c t g g c a a g c	t a g a t a t a t t	c t t g a a a g g c	t c c a g c a g c a	g a g t t c c c g t
	40921	c t t g t t t a g t	t c a a a a a c a g	c t c c t g g c c c	g t c t c t a a a t	g g t c t t g c t a	a a a a c a t c t c
	40981	c t c c c a c c c a	t a g a c c t g a a	c t t a a g c c t c	a g a c t g c t a t	c c t c t c t t t c	t g c c a c t g t g
	41041	a g a g a c c t a t	g c c t c t t t t c	t c a t t g g c t t	t t g c c t g c c c	a g c c c t c c t c	a g a t c c t t g t
	41101	a c a t c c c t a g	g a a g t a c a t t	c t t t c c a c t c	c a c a c a c a t a	a t t g t g c a a c	t t g t g c t a g g
20	41161	a c c g c a t g a g	g c a c t g g a g c	t g c a g a t a g c	a a g g a a a t g t	g g g c c c t g c c	c t c a g g a a c t
	41221	t g a a a t c c g g	t a g t t a c c t a	g a g c t a c a c t	g a g t t c c c t g	a c g t g g t a g g	a a g c c c t c c c
	41281	a g a g c c t t g t	c c g t g c t t a g	a c g t t g c c t t	c a c a g a g g t g	g c t a a g g g g c	a t t t t g t c c c
	41341	t g c c c t a g t t	t t t a c a a g t c	c c c t g g a t g t	t a a c t c c t a c	t t g c t t t t a t	t t g c a g g t t c
	41401	t c c a g t c t t a	t g c a t t c t c t	c t t a t t c c t a	a a a a t t t c c a	a t c c a g t g c t	g a t a g t a t g t
25	41461	t a g a g c t g t a	g g g c c a g g a a	a c c c t g c t g g	g g g a a t c a t c	g t c t a g t a g g	t g g a g t g t g a
	41521	g a g a g a g g a g	a c t c a g g c c a	g a g g g g c t t c	t g a g t t c t g g	g c a a g t c c c t	t a c c a t c c t a
	41581	g t g c a g c t a t	c c t t c t c c a t	g t c c t a c t g a	g c t g c c t t c c	t t c t t g c t t c	t c a t t c c c a a
	41641	a g a g a g g c a a	c c a t g c c a t t	c t g g g c a g t g	g g c a a g g g c t	g c t g g a a t t g	a g a t t t t t a t
	41701	t t t t c t c t t t	g g c a t c t g a a	t c c t t c c t t t	g c a g t t g c t g	a c a t g c a g t t	t t g t g t g a g a
30	41761	t c a c c c a t g t	c a t c c a t c a t	c t t c a g g a c a	c c c a g a a a c t	c c t c c t a c c t	c t c t c a g c c t
	41821	c a c a t g c g c a	t c t t c a g t c c	c c t t c t g a t t	c a c c c t a g t t	g g t c t a t t g c	t c t t g g c t c t
	41881	g a a t a a a t a c	c t t t a a a g g g	t a c a g g t a t c	c a t a t g g a a g	g c c t t c a g a g	a a g a g a g a g c
	41941	a t g g a a t t t a	t t t t c c a a a g	t g g g a c a c t c	g a g a g t g a a a	g g g t g a g c t g	a c a a t a a t g a
	42001	t g g c a g g a c a	a a a g c a t a t t	t c a g g c t t c t	t a a g a g c t t c	a t c c c t t t g a	t c c a g t g a c t
35	42061	t t a t t t c t a g	g a a t c t t t t a	t g g a g t a c c t	a c c a t g a g c c	a g g c a c c a t t	g c a g g t g c t a
	42121	a g g a t a t c a a	a a c a c a t c a a	a a g a g a g c a t	a c a g t t a a g t	g g a g c a g a c a	g g c a a g a a a a
	42181	a g g a t g g t t g	t a a a a g a a t g	c g g t a a g t g c	t g a g a a g t c t	t t c t c a g a g a	c g t g a t g t a t
	42241	g a g c a g g c t c	a g a c a g t c a a	a t a g t a g t t a	g c c a g g g a a g	g a a a g g a g a	a a t t g c c t t c
	42301	c t g g c a c a g c	g a g t t g t c a g	g g g a g c c c t g	g c a g a g g t g a	g a g t a t a c a t	g g t g t g t t t g
40	42361	g a a a a c c c a g	g t c c c t c g c t	g a g g c t g c t g	t g c a c a g g g a	g a g g a g g g c t	g g c a a g a g a t
	42421	g a g g c t g g c a	a g t g a g g c a c	c c a g g a g c c a	g a c g a t g a a g	g g c a g a c a a t	g c t g a g g a a t
	42481	t t t a g c t t g a	t c t t g g a g g t	t a t a g a g a a g	a a c a g a a a t t	t c t c c a g c t g	a g g a a t g a c c
	42541	c a a t t c c a t t	c c a g t g c g g g	g a a g a a a t g a	t a g g g a a g g c	t g g g a g a g g g	c a a a a t a g g g
	42601	t c t c g g a a a c	c a t t t a a g a g	t c t a t t g a a g	t a a t c c a g g c	g a a a g a g g t a	g c a t g g a c t g
45	42661	t a g a g a a a a g	g a g g g t g g a t	t t t g a g a a a t	a g g a g g a g g c	a g a g t c a a c t	g g a a t t a t t a
	42721	a a g a t c a g t t	c a a t g t g g g g	g c a g g g a a g a	a a g g a a g a a a	g a g g a c t c a a	a c a t g a t t t c
	42781	t g g t g c t g c c	t t t t c c c a a g	a g a g a a a a t c	c g a g a g g c t g	g a g g a g g a g c	a g g g t t t g g g
	42841	g g a g g a t g g t	g a c t t c a g c t	t t g g a t a t a g	c a c c c t t g a g	a t t c c t c t g c	a g t t c a g g t g
	42901	g a g a t g t c t a	g t a g g c c a t g	g a t g t c c c a t	g g a t g t c t a t	g g g t t t g g a g	t a t a g g a c g g
50	42961	a g g t c t g t g c	c t g g a g a t g t	g a g c a t a g a c	g t a t c t a c a c	a t c a g c a g g g	t t g a t g a g a t
	43021	t g c a c a g t c a	t a t c c t c a g a	t g a g a a g g g a	a a g c a g a a t g	t t g g a g a a g a	a t a t g g a a a g
	43081	a a t g c a a g g g	a a g t g c t g g t	g g t g g t a g g a	g g a a g a c c a a	g a g a g g t g c c	g t c t g g a g g a
	43141	g g g a t c g g a g	t g t c c a g t g a	t t c a g a g a g c	a t g a g g g a g g	a g g g g c t g t g	g c a c a c c t g c
	43201	c t g a c t t t g a	a a t g a a g a a t	g g t a t t c t a g	g t g g g g c t g t	t t t t t t a g g g	a c c t g g t g c a
55	43261	g g t g g g g a t g	a t g g a c a t a g	g g a a t t c t a c	t c t c t g g g a g	a c a g c a t g t g	g c t g g g a t c a
	43321	g a t a g t t a c a	a a g t t a t a t g	a c a a c t a t t t	t a t t t c a t t t	c t c t a a t a c t	g g c a a c a t c t
	43381	g a a c a t t a t c	t a c a g a g g a a	g g a a a c c t t	t a t g g a g g g g	t a g g t t g a a a	a t a t g t c t g g

43441 gccaggcatg gtggctcaca cctgtaatcc cagcactttg ggaggctatg gtgggaggat
 43501 tgcttgagcc caggagtttg agaccagcct gtgcaacatg gcaaaacccc atctctacaa
 43561 aaaatacaaaa aaaatagctg ggcttggtgt gcattcctgt agtcctagct acttggaagg
 43621 ctgagggtggg aggcattcacc tgagcccaag aggtcggggc tgcagtgaagc tgtgattatg
 5 43681 cccctgcact ctagactgag tgacagaatg agactctgcc tcaaaaaaga aaaaagaaga
 43741 aagagagaga gagagagaga gagaaggaaa gaaagaaaaga aaaagagaaa gaaagaaaaga
 43801 aagagagaaa gagagaaaaa gaaagagaaa agagaagaaa agacaagaca agaaaaagaaa
 43861 agagaaaaata cacctgggat ggaggccctg atggaagaag gtttctgagg aaatgggtgt
 43921 ggggtcccaga gtgtcagtat agggcctgtt ggagagcgag ,tatgataagg tcagccctgc
 10 43981 cactcagaag aaattctggt tatactcaca gtaatcacac ttctgtggtt acccagggca
 44041 gtcctaccgt gcaaccatta cccaacaaa attattgctt acaccccttt tcaacttttga
 44101 aagtttaggt gataaagtat ctggtctcct cagagtatac atctaaggag attcgtaggt
 44161 ggagtttagac agctgaggga gctcctacct gaaggactct gtgcagcagg aggcaggatt
 44221 ctctgtatga tgttccaaga ggcttgagga gataggagca gctgtgaggg gaggggaagac
 15 44281 ctgaaatgag aaaaagagca aattgttgat ctacagtaag ggctgagcag agcagatcct
 44341 tgagaacatc aatttgctgg gcaccaagta gttcataggc cgcgtttccc tgtcttttagc
 44401 agcctgcgtg tctgctggga gaatagatgt ataaactgat ctgaggttgg agctgcacag
 44461 ggcagatgtg gccagagggt aaggaagtgg ctggagtggg ggatcatgag atagcacac
 20 44521 ccgtagtaag ctctgcacat ggagataact attgcagcac tgtttatggg agcagaggat
 44581 tcaaaatagt aaatgtccaa caaggggggtg tcatccggtt aattatggca aattaacgtg
 44641 atagaatatt atgtggccat caagattttg tttaggtaga attgtaatat aaggagagaa
 44701 agcctagatt gtaatagtaa gtgaaataag aaagtccag aatgatgtat attaaacat
 44761 cccgcctata atgcatagag atttttttat ttatttttaa aaggttggg acaaatttc
 25 44821 cataataata cctgtgggag ctgatatag atatatttt gtatttttaa atttttcttc
 44881 aataagttca ttatattact tttataataa ggaaaaaaca tctctctagg aaaaattatt
 44941 taaggaaaaa ataaacatga aaaaggggtg ctgtaatcac agtttaattt ggccttctca
 45001 gaaaaccatt gtagggagtt tttattctct agtttcccca ggttggggaa gatgagcagc
 45061 atggccctag tgtttgagaa cgtgggtttt gatatcagac atgcctgggt tgaattccca
 30 45121 gctccgctaa gctgagctat agtattgcca tgggtgtgac tcaagcaggc tatccttatt
 45181 tatgaagtgg aataatacac atccccagtc tcttctttat aattccgaaa ccaaaaaagc
 45241 cctgaaaacc cagagtcggt tcttaaagtg acagcaaatt catttggtta taaaacctaa
 45301 cctgaattga gctagtcttt atttcacttc agtgaatata attttgagc ggaaatatta
 45361 acatgtttta tcattggatg ttgcgccaga tctcacggag ggtatgagat aatagcagg
 35 45421 gttcaggtaa gttgtgaact acaggatctg cgtcataggg ttattgtaga atgaagtcag
 45481 gttgtgattg cagggttctt agcacagttc ttgccagaga gtaagtactg acaaagttag
 45541 ttatcatcac tattgatgct ataattattc gtgagccttc actaagtatg gagtgcata
 45601 tttctgcttg cccctcttcc ctccatttcc ttctttattg ggatgtttga gaagattaat
 45661 gtatcacata taatgagaat ttaggatggc attgtttagg acacaggact gcaaaaaagg
 40 45721 gaagttccag cacagatatt ttctgcctc cagatttct cactttcttg
 45781 ggaattagct gtaggataata atgcacatta acaggacacc caaatacttg aagagcta
 45841 ggagaagtaa cccttactcc ctgtgtgact gctttcgtct caaagcacat gtgttcatta
 45901 taaaatgtta atgtatttac accatagttg ctgtacattt aaaggtttac tgtgctttaa
 45961 gctaaaacta gctggatata aaatgtgctt ttaattgaaa agtggttatt tactgctttc
 45 46021 cataggaatc cacaataaga ttttttcatt gcaaaccctc ctagtatctt taaaatgcaa
 46081 ttcagttcac aaacattgat ttgcattcaa cgtttttcag aagcctaata agagtgc
 46141 atgaggtcca cctgcaagac tacagttatt acttctctcc ttttcttgga agcatccatg
 46201 cattgtaaat tctgtttgtg tggcatcagc caactacatt attaacatca ccaggaggaa
 46261 ttaaaatgta tcatgttata ggatcaaaag tttcacattc tgaatcaatg ccagtataaa
 50 46321 aagaatgtcc agacccttcc gtctgactac agtcggcagg gatccccgcc ccagcagcca
 46381 tgtgactcac ttgcgattgt ggtgagggtc actcatggct catacggggc tgcggccgga
 46441 taggcctccc catggggcag catttagcct tttcaaaatc agcttaggat cacagaccct
 46501 caggtcttga aggaactttt tggagtttcc tctattaaac ccagttttaa tctggtcatt
 46561 tagactgcat gtctagaaac actacagcat ttttagtagt aaactgaaat aaatcttcaa
 46621 ataaacgact ttaacatttt gcacattggg ccagggggcag aagagctcgt ctgttgctt
 55 46681 ttaggagggc tctaattctt ttctgtttca gctcatcggg tcctcattgt cccctactgt
 46741 atgtgccag cacaatgtca gggcacattt ctacacaatg ttggattcca gctgatctc
 46801 cgtgttaatc cattcgttag tgccaaattg cagggttcct tgccaagccc gtcagcacia

	46861	tcagagcttc	tgaagtcagg	tacctcaatt	acattagttc	agttgtctca	ttagccccc
	46921	tgcattcatcc	aaactcatca	tcaaacgtgt	tttccttcat	tattgttgtc	atittcttaa
	46981	taccagtctt	gttgaaaaag	gattgtttta	gatacaattt	tgcctaactt	gtttcatctc
5	47041	cagaatacaa	tcctaaatag	agccaagatt	tgccggagag	aggagatagc	ggttgaggct
	47101	ggtgaagctt	ccctccggtt	agaattagag	agaagtggat	gcagaacttg	ggccagtaat
	47161	ggacccatcc	atctccctag	tcaagaggac	attcttgcag	ccagagggtg	ttccgctt
	47221	gaaatgatct	actttgtgta	aaccgggtgca	tcacagatac	tgtagagta	tctcacagat
	47281	actgatacag	catgggtctgc	attttgtaga	tatgagggtc	ctcccttagg	tttacagccg
10	47341	gataaaggag	ctccacttac	ctgggtgatca	caggcatgtg	gtccaccctg	ctgggctcta
	47401	ggactttgaa	catagagaag	gctatctttt	aggaagactg	caggatcagg	agtcgggaag
	47461	cagggattct	gctttagatt	ttacatctag	cagtaacaac	ttttagccca	tcctcacct
	47521	actctgtgcc	ctggatttct	tggctttaac	atagcaaaa	cctctcttta	tagttttcct
	47581	aactccatct	ctcccaaaaa	agctggagat	gtaagtttta	aattccagtc	cttggcacat
15	47641	agtaggtgct	caatctctca	tccttctgtc	cctcaagata	gccgatggct	cagccatata
	47701	tgttccagac	aggtctctct	gtgtcttttg	ggaaagcctt	gttgtctctc	gactggcctc
	47761	taggcctcat	gttttccac	ttgtcttgc	cttcagtgtt	aggtcatggg	cttgccctcat
	47821	tacctgtttc	aaggaggcct	ctatttgatc	cctggagcat	ccctgggagt	ctgaggcctt
	47881	taaggatccc	tgtgggtcca	aggaaaaaga	actataagta	gttctcaggc	tccttgatc
20	47941	aattcaaa	catcttcagc	tcttccagt	cagccctcc	agagctgttg	aacgccacct
	48001	ttttcctctt	tcgggaaacc	cacaaaacct	tgcatgggt	tgtgtctgt	atggttttcc
	48061	agcatcaaat	tcactcatca	aaagatgcat	caggagggtg	gagggggtg	cgctgggtg
	48121	gagggcaca	caaggaaggc	cagaatgcct	gtttgcattc	gcaacgggaa	cattggaagt
	48181	tcgggtggaa	aaacaatccc	accggccttt	atttaccaca	gtgggtagcc	tcgggaggag
25	48241	ggggagctgg	gtgggaaagg	aagacaatag	aatgctgtgt	tgcttgcgtg	taaaagcttg
	48301	ctctaagcaa	ataaagtggc	aagctcagtc	agagccctgc	tcaattagcc	ccagtgccta
	48361	aaggagggtc	tccttgcagc	tgttggcggc	gggcgggctg	gagcaaaggc	aggcgactc
	48421	tggggcactc	gggaggcgaa	ccggcaggaa	tcttgcattg	gagctgacc	gggagggaga
	48481	ccaaaggacc	ctccaacctg	atcccagtc	ctgcttcttg	aacagaggc	tacagaagg
30	48541	gttgggtggg	ccactctggg	gagaggcagt	gtggggagac	cacggaggag	gtgaccagaa
	48601	aagttagtgc	aaacgtttca	tgccgagagt	gaccgacata	tggaacgtgt	tatccggaag
	48661	ggttacaggc	tgaggagct	cctagagttt	ttcaggggtg	gggggtgggg	agggggaagc
	48721	cttcagaaac	tcaaggaggt	ttctaataca	aaggaatttg	agaagggttg	gagaaataat
	48781	gagaaatgca	aacttgacct	agagtttctc	cttccatctc	acggcttctc	ccctgcctgt
35	48841	cctcacatgt	ttccagttca	tggggaccca	ggaaggccac	tggagccctg	tgctgactc
	48901	cacgtgcacc	tcactggggg	tgtgggtggc	ggggtagggg	ggaccgcag	agctggctcg
	48961	ctgctctgtg	ctgaaaggga	cccagagagc	gagagccctg	cctggcttta	gaccctgtg
	49021	gactagcagg	ctgctagcgc	cgggatggtc	tctggatgat	ttactccgtg	tcctgcctt
	49081	gctggaggag	caatttgatg	ccggactggg	agtgaaaaac	agacttggcc	agggactcac
40	49141	agtggccaaa	ggggaggctg	ggagtagacg	caggctcttg	atcccttctg	ctttttggca
	49201	tctcccttcc	tcagcatccc	ttgccagca	caccacgcac	acatgtacat	acacacgcag
	49261	acacacacac	caaacagatg	cacacagaat	aacaacacca	cacctacatg	cacacataca
	49321	cagacataca	tgacgcacag	acacacacag	aataacacca	cacatacatg	cacaccacag
	49381	atacacagaa	taccacacac	acatgcacac	acacagacac	attcacagaa	tacatatatg
45	49441	cacacacaca	catacatata	tgcaccacac	agaataacac	cacacacaca	tacagacaca
	49501	cagacataca	catacatgca	cacatacaga	tatactcaca	gaatacacac	atgcacacac
	49561	atgtatagac	atacaccaca	caagcagaat	aacaccacac	gcacacacca	cacacacaca
	49621	cagaataaca	ccacacacat	atgcacacca	tacagagaca	cactcacacc	tagacatata
	49681	caccacactc	atacagagac	acatgcacac	acatgcatac	tacacacatg	cacatgcaca
50	49741	cacaaaaaca	tacacagaca	tcacacatgc	atacatatac	acaacacaca	gacacaccca
	49801	cgggcttaca	cagacaccac	acacacgtag	atatgcacac	agcacacaca	cacacagaga
	49861	aacagacaca	ctcgcaggca	tacatgcatg	catacacaca	ctacacacag	acacactcgc
	49921	aggcatacat	gcatgcatac	acacactaca	cacagacaca	tctgcagtca	tacacaggca
	49981	ctacatacat	agagacatac	aatgcacact	cacacacaga	cacacataga	cacatgttgt
55	50041	gtgctcatgc	gtacacgagc	gcgcacacac	acacactgcc	tgacttgttt	cccacaagac
	50101	gggtactggc	ctgtcgttgc	tagccctcct	gtcccagcct	gtgttggcca	gggtgccagg
	50161	cactgccacc	cctcttggga	caaggtacag	gtggccagtg	tgatcacggc	ctgttctgag
	50221	agcctcttcc	tgagccagga	agcgtctgtg	gatactgagt	gcccgctgct	tcgtctttcc

	50281	catgggtgctg	ggctcttgcc	acctgctgca	ttgatagcac	ccgcatgttc	acttccctgg
	50341	cagtagaaaag	aaatgtaggt	tagtgcaggg	aggctactgg	ctttggagat	ggtggcaggg
	50401	tggagcagtc	tctaagtga	atggaagtgc	acatgccctc	tggacctgca	gcaatgctgg
5	50461	cttcagaggg	gccttctctg	gtacctttcc	aattttcccc	aaccaggtag	aagagccact
	50521	gcccagtgct	ttgggctcaa	ctgaaacccc	ctgtacaaga	aagagacccc	cctttccact
	50581	gtgtctctct	cctcctcccc	actcactgtc	tctctgtctc	cctcactccg	tgctctctct
	50641	ctctcttctg	tgtctctctg	cctctccctc	tctctgtttc	tctctgtatc	gctctttttt
	50701	ttcaatctct	gtctctatct	ctcctccttt	ctctttatct	ctctgtctct	cctctgtct
10	50761	ttctctctgt	gtgtgttttt	ttttctgtct	ttctgcatct	ctgtttctgt	ctccttctct
	50821	cggttctgta	tctttgtccc	cctctctccc	tatctctgtt	tctctgtctc	agtccttttc
	50881	tgtctttatg	tctctgttta	tctctctcag	tctctgtatc	tctgtctctt	tctcttttta
	50941	tcctttctct	gtctctttgt	tctctatccc	tccatctctc	tctccctctc	cctcagctct
	51001	tctctgtctc	ttcctttctc	tgtctcagcc	tctctctgtc	tctgtatctc	tatctctttc
	51061	tctctctccc	agtctctctc	tttctatctc	tctctctctc	tctctctctg	gcactcactc
15	51121	accacttac	ttgaagtctc	catgagcagt	gggtggcttc	acctttctgt	ttcgccactt
	51181	tgcagtccga	cccgtggggc	ttgcagaccc	tcttctggcg	cacaccttca	ggagaccaac
	51241	ggtgccaggg	cactcccgtg	ttcttcaggg	ttccagcccc	gagtagttgg	taaacaatcg
	51301	taatcgtcct	agagatccac	tgtagattcc	tcatccaggt	actgaatgag	gcccttctga
	51361	gcaaattta	tggaatgacc	ttggtgacat	tacaagatga	cggctcatct	cctgtagtct
20	51421	attattgttg	ggcatttagg	ttgatactat	gtcttcaacta	ttgtgagtag	tgctgcagtg
	51481	aacataatg	tgcataatgt	tctttatgaa	agaatgattt	atagtccttt	gggtatatac
	51541	ccagtaaatg	gattgctggg	ttgagtggtg	tttctatttt	taggtctttg	aggaattgct
	51601	taggtctttg	aggaatcgca	tttgctgttt	caaacaatcat	ggagaacaca	ggtcttgagg
	51661	atgtgcagga	acatggagac	caaggaaatga	aaaagccatc	aggttggaag	tgaagtcacc
25	51721	cagagtgatt	gcaggtttat	ggacagaaaag	gaaaatgctc	tgattcctgg	tgctcatgag
	51781	gaagggtgat	gccaatagat	ggtgctacac	agacacttgt	gaaaggaaga	agggaaagag
	51841	tctgcctcca	tgggctgagt	agggctgctg	gggcctcagg	cttcacacat	agtgctcagac
	51901	atgctgtgtt	ggtcaggttc	atctgggacc	cctgtgcctc	ccccaggaaa	cctatttggg
	51961	aggtgtgggc	aggcttccct	ccatgtgtcc	ttaaatacac	tgactgctct	gtgtgtgact
30	52021	gtggtgagc	agcgtagggc	tggtcatctg	gagggctggg	cttggttcca	gctctgttac
	52081	tgatggattg	attggttttg	ggcaagtcac	ttccccccat	aacctttcca	ttgtaccaca
	52141	atcctaatag	agttggatgc	cagtttctgg	caccctgggc	ttcacaggga	cacagagaaa
	52201	cagagaggac	cagaaaatta	ggctgataat	aatcatttct	tctctatttg	gtctggagaa
	52261	gaaatgactt	aggttggtga	caggttaata	tgtcaactag	aaagaacctt	tgaatatctt
35	52321	tagttcaaga	attctaaacc	taaaatctgt	gggacactag	gatgattttt	atggactctt
	52381	tgaatcacc	tgaattttt	taaaatattt	tttctttatg	ggcacatgtg	tggtttttta
	52441	gcagaaggat	tccttagcat	tgattacatt	gtaaaagggc	caacaatcat	aaaacaatta
	52501	attccccagc	atcatgttat	gtgattcata	actgaggggt	ctccgcagat	ggcatttggg
	52561	aacgtgcggg	tccttttttt	cagttgtcag	catgactggg	gtcacgatgg	atatttagta
40	52621	ggtgggggcc	tgctgtgcat	gcgtggggga	ttaaacttaa	caataaaaaa	ctgtcctgtt
	52681	tgaatccca	gtagcacc	tctggagaaa	cactgctcag	gaactgagcc	ccaagatgc
	52741	acagtgattc	tcccggtggc	acaaagctca	ttagtggcag	agttgggttt	tattaacaga
	52801	tcaaaacagg	ggatgtgcca	agaacctaca	ttattttgtt	tctcatcatc	agctgcaaat
	52861	gtgctgcaat	ctgtgaaaac	aaaagaatca	gaaatctgtc	ccaccttcaa	cgctccagcc
45	52921	attttttaaaa	atgagtcttt	gggttcttgg	ttgtttttta	ccatcaaagt	agaaatgagg
	52981	aaaggaatat	tttacatttg	gagaaactaa	accataaaca	attgatccca	ctagccccat
	53041	atcacctaga	tgtgttcttc	agctactgtg	aattggtgac	gcaagcctta	gactggaaat
	53101	ttccccatta	tgtttcagag	gatctgtctt	gaatgtctct	tactctacaa	agaaagaaac
	53161	attaatatat	cagcatgctg	cttgctacct	agtttactgc	aatgaagtgg	caggtgcctt
50	53221	agacttttga	gtgaaattga	gagatattcc	gcagtattag	ctaagagtta	ggcctgtggg
	53281	atcacagaga	caggggtttg	ttccctggct	ctgtctctca	attgatgtgt	cttcttaaga
	53341	aaaagtactt	aatccttctg	agcctcagtt	tccttggctg	aaaagtgggg	ataataatcg
	53401	tatccacaaa	gattaattga	gatcatccat	gtgaagtgtt	ccgcacagtc	tagcatatgg
	53461	ggaggctcaa	taaatgtgag	ctgttattac	aagcgattat	tatgacttgt	gtctgtgatt
55	53521	aaagacagcc	tgaggttcag	agtgtttgtg	ccaaaattgc	tttcagagga	caagctctgg
	53581	gttggttttt	cagtcttgcc	cagctgccc	gagtcactag	agtgtttggg	gctggagtca
	53641	cagtgtttat	taagtgccta	ccacacactc	cagttttgat	ggaagattgg	ttttcttctt

	53701	gtgtgcacat	ccccgcatca	ctacatgagt	gggtgagaag	agactcagaa	aagcaccaga
	53761	cacatgtttc	ttctgcctca	gttccctagt	ctgtaaaagt	actcaagagc	atgatggcgg
	53821	ccacagggaa	cagaggggag	agacacctgg	aggagcaaaa	gacagtccca	gctgtcagtc
5	53881	ttcagacacc	atggttttct	ggagaaggga	tccagaccag	accgccaatc	aatatagccg
	53941	cagaccgcca	gtcaatatac	ctcatccttg	tgaaaggcgg	ttgtctgtgt	ggcagcaggg
	54001	agacaggagg	ggccatatatt	gacacaagcc	tgctggccaa	gtctaaagga	gtcagcctgc
	54061	catctgacca	cactttctgc	agccaagtcc	tctggggggc	aaatcagtat	ggtttgagtt
	54121	tatccccgac	ttgtctgctat	ggttgatgca	cacaaaaaag	gtttggccat	tgccaagcca
10	54181	gctgtgtgtc	tggtttgaa	gaggcaacca	aagtaagaat	tgaagggaag	gcatgctttg
	54241	ggattttccct	tagccttcaa	ccctctagag	gaagccaact	tctttattct	gccattttga
	54301	gactcatgct	tgtctgggtg	gataagataa	atactttcat	cattatcatc	attattatta
	54361	attataatta	ttgaaaatca	attactcagc	agcccctacc	ttgtaatgg	atcacttaca
	54421	gttgtatagc	attttgcaga	caacgtatac	actagaccat	aaactcagct	cgaacataag
15	54481	cataaattta	aaacaaaaac	attgacttgg	gcagctatta	tctgaggtca	aagtatttct
	54541	ttaattccag	ctgtaccctc	agctgcctat	gtcgtggaag	caacttcccc	ttgtgtgttt
	54601	tcacacttag	agattggagt	aatccagaca	cagaaaatga	tacaaaaatt	gaaatgtttc
	54661	tgaacgttaa	atgattcaat	aatttgaaaa	tagtttcaaa	tttttcaaaa	ttcaaaaaat
	54721	tgaaaatagt	tcgactagag	cccctcctca	cctcaggtag	ttacgtggtg	tgtgtgtgtg
20	54781	tgtgtgtgtg	tgtgtgtgtg	tgtgtgtagg	cacacaaacc	catgcatgca	cactgttatg
	54841	cacagtgtgtg	gatgagtgtg	agttaccaaa	acaataacca	ctaaatgcag	gataacacca
	54901	tgatggctag	tagctggatt	gactcagatt	ttctctgcag	atccttccta	gtccaccctc
	54961	cttccatgac	atgggagttt	gtgatatgca	aaggataaag	taaacctgag	tttctggggg
	55021	cttgtgtcgg	cataccaagc	tcctttaaga	tctctctttg	ccaagatggg	gctaagccag
25	55081	aggcttcctt	ggcctgggga	ataactgtcc	taccttcctc	cctggagcgg	ctggcgacac
	55141	ctgttttgac	gtgtgccatg	ccttccatct	tttacggcct	cggggccaga	ggagggatg
	55201	ctttgaacaa	tgtgaaattc	ctctcctaac	agctgtgcaa	aggaaactca	cagctttcta
	55261	tctccatctc	ggtccactta	gctttctttt	ggggtatagg	tttcttgtct	gttcaggggt
	55321	ccctgttcct	cactttctgg	ccttcagaag	gaccagtttc	agtacttcct	ttcttaggga
30	55381	ggcaagtcat	tcctctgagg	ttctctgttt	tattttattt	tatcctagag	tatgccttca
	55441	ctgttgaccg	tgtcttgcc	catttctgct	aagaagccag	catcttctga	aaattgagtg
	55501	tcttgtcctg	tttcagatct	tgggccgagg	attgaccatg	tatgttagtc	agtttcagag
	55561	ccctgagttt	taagtgctaa	gttttaagga	gaggaacaac	agtagaatta	gatatcttaa
	55621	agaacaggat	aggaatccac	ttccctgccc	cttctctcac	ctggtttgac	actacataga
35	55681	tacagacctg	tcctgagttc	agggattgtc	taatgaggcc	tcaccaaata	ccaacagaga
	55741	gactagctgt	ttctcctcta	gctcactgac	tgcttctagt	tttccgtgac	tttagagcat
	55801	gtgcgtcagc	aggttttggg	gaccatctct	ggcctgcttt	ctgatcttgg	ctctgtccct
	55861	caatttatgt	cttttttgtg	ctccataaga	cagggagttt	catgggtgcc	actaccggag
	55921	aggtgtgcta	aagtgaatga	ggtggaggac	acagagctag	caggggaagaa	gggtccgtgg
40	55981	aaacccacac	agctttattc	atgtgagctc	atttctcatt	tactccttgt	agcaatgcca
	56041	agtcaccccc	tccttccatc	cctttccaag	catttccctc	ccttggttgt	tttggatcca
	56101	ttgtaacccc	agtgaatga	cctacatgtt	ttagtttgaa	tctaactcatg	ctatttccag
	56161	gttgccaacc	cccacgggtg	cttcagccta	attgggtctgt	cctcactggg	gtctgcagca
	56221	ccttcccttt	gagcgtcatc	tgtgcatttt	aattaatctg	ccactttttc	ctccctcctg
45	56281	cttccggatc	attaataaag	agattaacca	aaaccagacc	aatccccagc	ggaggctcca
	56341	ccaggatctc	tctctcccag	ctctcagctt	tgtcatgggt	cagaacatca	tcatgtggcc
	56401	ccattgacta	attctgaggc	ctcaagacct	atgttcagag	caaagccagt	gtgaattaat
	56461	ttttcaatta	agatttgggg	aacaagcata	gccagtcttt	tatatatggc	cttagtttaa
	56521	tatttccacc	acattccttt	cccttttgta	attctacttt	aaatctatca	ctttcccacc
50	56581	tctctccctt	tccagtcctt	gaacaagaca	ttagggttat	ctggcaggat	gtctcctgag
	56641	cctgtagcat	tagtcagaga	tcattgtcata	gccgaagaaa	ttctccaaag	aatcaggccc
	56701	ttctggacac	agctggccag	cgctctcttt	gaccccagcc	agagaatgca	tttgcttcca
	56761	gagaatatgg	ctacaccaga	aataatgaaa	tctcttgaat	ttgcatatta	aaaatttggc
	56821	aaattaaaaat	ttgctttcac	ctctgtctctc	atgactcctc	tctgaggaaa	gctaaaacag
55	56881	agcctttctg	ctttctttgt	aagtcaccca	gatttccagc	gctgggtctca	aaccaggat
	56941	gtgagatgg	gcactccgtg	ttctgtccac	tacaacacag	ggcctctgca	ggcacagggt
	57001	acctgcagat	gggctgcact	ggagagcatc	aggcccagga	gggactaggg	tgagggaatt
	57061	ggattgtgtt	tacttttgtg	gggatgctt	tcatagggaa	tacagggagg	aggacacagg

	57121	attgatggcc	tcaaatctct	gcaaactctg	cttagagctg	aggcttcttg	gggaaggaca
	57181	cagcctgctg	acagcagccg	tcctggctcc	aggcttggag	cttgggcttc	ccctggaagt
	57241	aaactctcta	gaatgtttga	gagcaaggac	agacggagag	gcagccgagt	gtcatcatta
5	57301	ataacatgag	ctctagagcc	agaaggtgca	aatcctgact	ctgttgcttg	ttagctggtt
	57361	gtggaacccc	tcctttgcctt	ggtttcccca	ttcaaaatgg	gaaatataat	agtacgtact
	57421	tcaggatggg	tgttactagg	atttaagatt	tatgtaagcc	agtacctggc	actttgcatt
	57481	tttattcttg	gtgatctcac	ccctaagcag	tctaaaaatt	tacttcacac	atgtgcagca
	57541	agaggcgtct	ttaggtacac	ggtcaggagg	tgctcctgtg	tagggaagct	ggcctggag
10	57601	aggtgaattg	gaacctgaag	gctgcagccc	agtctctgca	cgtgacagg	cagacgcagg
	57661	ttgctggggg	ctcagattct	gtgccactgt	ctacctcagc	ttccttttcc	atgtcagctg
	57721	cttcttgaat	acccccaaac	cctctgtgga	tgctgtaaac	tgagccaaaa	aggaggaatc
	57781	agaaattgct	ctctttttgag	tattttgagg	gatggtgaaa	ccttgtagga	atgttgctcc
	57841	agggactgag	aaaacaccat	atgtgggaaa	gcgtttctaa	gtgaggaaaa	tcgcacatgc
15	57901	acgtcgggtca	ggatcttgga	aaactctgag	caggggaaat	gggcactggg	gagagtgtag
	57961	aaagggggtt	cactaagggtc	tttaacaggt	ctttttctaa	ttatagccaa	gaagagcctg
	58021	gggacctagg	ccgacttcag	agccagggcc	tgtattttct	ttgggaaaag	aaggcagagt
	58081	tgattggcct	ccaaaaccag	gcttggcaaa	agcccatgca	ggctctccag	ccagtaagca
	58141	gtgcctgggg	taggtctcga	tcgtagcagg	cgctcattca	gtaaatagac	gtcgaagtc
20	58201	cgcagcgtgt	ccagccctat	gctgggagtc	aaggatgtgt	ggatggaaaag	gacatgggtcc
	58261	ctgtcgtcaa	agagaaggca	caccaacaat	aatagcacia	tatgacaatt	gcttaaatga
	58321	agctctctgc	agatgacagg	gacccctaga	gaaagagcgt	agacattccc	ctaaatagca
	58381	gaactcgtcg	ctgtgccttg	cctggatgga	gctggcactg	ccttctctgt	tccattcccc
	58441	actccaggga	actggcccac	tgggggactc	ccaccccgga	catggagtga	gaaatgaagg
25	58501	tattggtggc	caaaaagtga	ccagagatca	gaatgcatct	attcttgccc	ccagcaaatt
	58561	cagaaacacc	acctctccct	tgcccgtgtt	tgctttcgtc	atttgtctgt	tctctgggct
	58621	tcgagttcat	aatattccta	tcttctccca	caaccagtgt	aactggctctg	agatccaagc
	58681	agtcgggcac	catggcatct	gggggaaagc	cttgtgggga	ccgtgttggt	aaagattgct
	58741	gtgagccaga	tgtagaggag	ggagctctcc	agactctggg	tcctctcgcc	tgctgcgtca
30	58801	agggcagggt	tgccacctct	cccgtggcca	caggccattc	agcttcatgt	tctctaacat
	58861	ttctagtgtg	ctcactctgt	acctgatgct	atgctagggtg	cttccatgtg	gatttcattg
	58921	tttacccttc	acgataactt	tcatgggtgca	caagccatgt	gattttttgt	gcgcctacga
	58981	accaaagtcc	cccgtcactc	tctgccattg	tttagcagtc	tgaccttaga	cagctcactt
	59041	caccacaact	gcaggcctcc	gttcccttat	atgtataaat	gaggaggttg	tctagggtcc
35	59101	tttccagctg	caaacctctg	tgttttgggg	agttctgtta	gcaacattct	tggacttgct
	59161	ctctacgaga	aaagctagaa	gtcgattatt	aaaatggagt	ccacaggaag	cctgcagggc
	59221	tttaagacag	ccttctaaag	agtttagaaa	cactcaccag	cctgcagatg	acttactccc
	59281	actacttatt	acaaagtatc	atggcaagta	agaaagaaat	ggtcagccac	tgggagagca
	59341	ggctgaggtt	gcaaggaagt	ggcaggtgct	atcggtggga	tcttggggga	gctattgtaga
40	59401	aaggtcagct	ttggaaccag	gcactgaaaa	acgggtgaaa	tttcaacagg	agaggcagtg
	59461	cacatgggtt	gtgatttgag	cagcagtgag	tagtgaatgg	ggttggagca	agaggcctgt
	59521	gagcccagtc	atcacatgga	aaagggaggt	gagggccaga	gtgcagagaa	cccctctcac
	59581	tttctctgtg	gtctctctct	ttcttttttt	tttttttttt	ttttgaattg	gagtcttgct
	59641	ctgtcgccca	ggctggagtg	cagtggcggtg	atgttggctc	actgtaagct	ccgcctcccg
45	59701	ggttcacacc	attctcctgc	ctcagcctcc	cgagttagctg	ggactacagg	ctgccaccgt
	59761	gccctgctaa	tttttgtatt	ttttttttaga	cggggtttca	ccttgtttagc	caggatggtc
	59821	tcgatctcct	gacctcgtga	tccacccacc	tcggcctccc	aaagtgtctg	gattacaggc
	59881	ataagccacc	gtacccggcc	tttctctctc	tcttttaaact	gcatttaggt	tgtcctcatc
	59941	tataagatga	aggaataaga	ttagatggcc	tcttttatatg	gcttatctac	ttctaataaa
50	60001	ctttggttcc	atattccagc	caccacccat	aaagactctt	actttttcct	cctcaatcca
	60061	tcagcatcag	ccagcattct	ccctctcagt	tcactcctgg	cattacctca	ctaactctga
	60121	ccttgactct	tattctacct	gggacctccc	ctcactaacc	atacttgatt	tatttcttta
	60181	agagcggata	tcctgtgtat	tcttagctta	cagtggctga	gcttttggga	ggtttaacaa
	60241	attgtttttc	aatagactct	cagtcctaaa	tgattttatt	gcagctgtat	tattcttttt
	60301	gaggggagta	gagaatgtaa	tcctagttgt	atgccatatt	ttttgtattg	ttcaagttgc
55	60361	attttgttgg	gcctgtaata	ttgaagaaaa	tgtcacttgt	atgcagaata	ggagggagat
	60421	tccttatgct	gtagaggaac	cattttccta	ggatagtaca	attccttaga	gtcatctcta
	60481	agggtgagca	gcaaagctag	aatccctctc	ttggctttga	ccttgaagga	gtcagttttc

5 60541 ccaaagacta gttcccatgg aaagaagatg gtctccttat cacagcagca agaggaggca
 60601 gtatgagcag agtgccctcac tgttggtgac ctctagggac aagtgagccg gcagtatttc
 60661 agacaagctg ggaaagaggc tgtctgtgag tcctggggagc gagtgagcac tggcaggctc
 60721 agacataggt gctgggttagc aggactgctt ttctgtttct tgtgtcggct ttgtttattt
 60781 cctctctatt ttccccctgg acttagtaaa gtctttccga aaataccaaa ggtgaaccag
 60841 gggaagagtt tttattttcc atgtttggac agaactttaa agaggaaatg atgtaccccc
 60901 ctgggagcca gtgaggtggc agcgtatggtg attaaggagt gaatatctca aggaggtgga
 60961 cgaattcggg gatcactagc tcagctgccc ctctccacct ggagcatctc cttccagtgc
 10 61021 taccctcaga acatctgggc tttgctctag tgagggagag actagcaatg aaggtgtctt
 61081 gagatcagca ctgtaattcc accaggacgc caccggagtc cgggtgttaag cttctactat
 61141 ggcaacagaa tgagagcgtg gatgggttga aatgccattt caacaaggaa atagtagaat
 61201 tagaagaagc ataagcacta aaaggaacat tttgtagcag aaatgttaaa aatctaaagg
 61261 aggcaagtga atcaacaatg actcttctat tctgggcaag tccagcccat ttatgtaagg
 61321 tggttattct gcatctctgt cttctgcaag tagtgctgtg gcagagctgc gttttgtgga
 15 61381 gagcgtcccc ggggatggag cagatcagtt ggtgatgctg atgtatcaga aagctcggca
 61441 gagcaccctg gaacgtaggc cctctcggcg agtgggtagt ggccctacat gttcatttcc
 61501 aagggcagga gaatagaccg ttccagctgc ggccctggcca gggatgacct cacatctgac
 61561 actgcaatat gggggcaact gaaccagtcc tcagcctcag tgtgttccag gggctgcagc
 61621 tggggagcag tcgaactctt tcttgagaca attacaaggc caccgctgct gctgctgaag
 20 61681 ggaagttact ccatgtttac aattctcagg tttgaagttt tcatgctttg ccaaggtaga
 61741 gtgaaccatg cgtcttttga ggctcaaggg atgtttaaag aagcggtagg acatcgtcca
 61801 cccacaagca gagaccgcag gataaagcag acatccaatg taaatacaac ccgtgcaaaa
 61861 agcagagtcg gcagacctgg agtgcattcg cagtatctcc cgggggtggg ggaaagaaat
 61921 cacctcttca gaatgtccag aggggagttg ccttgcttac ctggggggcg gtaccctctc
 25 61981 tcgtgccctc acagggctac tcagcctcag gtagctgggt ccagaataac acagactcag
 62041 ctgccagagc ctgctcttaa cacctgtgtt tccttttcag atcttacagg tgaacaaggt
 62101 gatgtccatc ttgttttatg tgatatttct cgcttatctc cgtggcatcc aaggtaaaca
 62161 catggatcaa aggagtttgc cagaagactc gctcaattcc ctcatatta agctgatcca
 62221 ggcagatatt ttgaaaaaca agctctccaa gcagatgggt gacgttaagg aaaattacca
 30 62281 gagcaccctg cccaaaagctg aggtcctccg agagccggag cggggagggg ccgccaagtc
 62341 agcattccag ccggtgattg caatggacac cgaactgctg cgacaacaga gacgctacaa
 62401 ctaccgcggt gtcctgctga gcgacagcac ccccttggtg ccccgccctc tgtatctcat
 62461 ggaggattac gtgggcagcc ccgtggtggc gaacagaaca tcacggcgga aacggtacgc
 62521 ggagcataag agtcaccgag gggagtactc ggtatgtgac agtgagagtc tgtgggtgac
 35 62581 cgacaagtca tcggccatcg acattcgggg acaccaggtc acggtgctgg gggagatcaa
 62641 aacgggcaac tctcccgtca aacaatattt ttatgaaacg cgatgtaagg aagccaggcc
 62701 ggtcaaaaac ggttgacagg gtattgatga taaacactgg aactctcagt gcaaaacatc
 62761 ccaaacttac gtccgagcac tgacttcaga gaacaataaa ctctgtgggt ggcggtggat
 62821 acggatagac acgtcctgtg tgtgtgcctt gtcgagaaaa atcggaagaa catgaattgg
 40 62881 catctctccc catatataaa ttattacttt aaattatatg atatgcatgt agcatataaa
 62941 tgtttatatt gtttttatat attataagtt gacctttatt tattaaactt cagcaaccct
 63001 acagtatata agcttttttc tcaataaaaat cagtgtgctt gccttccctc aggcctctcc
 63061 catctgttaa aacttgtttt gtgatccggc tctcaggagt cactctgtaa aatctgtgta
 63121 caccagtatt ttgcattcag tattgtcaag gccatgactg ttgttttagt aaacttgta
 45 63181 aaatca

Brain-derived neurotrophic factor (BDNF)

The BDNF gene encodes the brain-derived neurotrophic factor protein. BDNF is expressed only in inner hair cells and outer hair cells during the neonatal stage. BDNF supports connectivity to SGN. BDNF induces synapse regeneration and SGN protection after damage (Takada et al. (2014) *Hear Res* 309:124-135; Budenz et al. (2015) *Sci Rep.* 5:8619).

The human BDNF gene is located on chromosome 11p14. It contains 2 exons encompassing ~67 kilobases (kb) (NCBI Accession No. NG_011794.1). The full-length wildtype BDNF protein expressed from the human BDNF gene is 255 amino acids in length.

Methods of detecting mutations in a gene are well-known in the art. Non-limiting examples of such techniques include: real-time polymerase chain reaction (RT-PCR), PCR, sequencing, Southern blotting, and Northern blotting.

An exemplary human wildtype BDNF protein is or includes the sequence of SEQ ID NO: 31. Non-limiting examples of a nucleic acid encoding a wildtype BDNF protein is or includes SEQ ID NO: 34. As can be appreciated in the art, at least some or all of the codons in SEQ ID NO: 34 can be codon-optimized to allow for optimal expression in a non-human primate.

Human Full-length Wildtype BDNF Protein (SEQ ID NO: 31)

MFHQVRRVMTILFLTMVISYFGCMKAAPMKEANIRGQGGLAYPGVRTHGTLESVNGPKAGSRGLTSLADTFEHVIEE
LLDEDQKVRPNEENNKDADLYTSRVMLSSQVPLEPPLLFLLEEYKNYLDAAANMSMRVRRHSDPARRGELSVCD
15 WVTAADKKTAVDMSGGTVTVLEKVPVSKGQLKQYFYETKCNPMGYTKEGCRGIDKRHWNSQCRTTQSYVRALTMDSK
KRIGWRFIRIDTSCVCTLTIKRGR

Mouse Full-length Wildtype BDNF Protein (SEQ ID NO: 32)

MTILFLTMVISYFGCMKAAPMKEVNVHGQGNLAYPGVRTHGTLESVNGPRAGSRGLTTTSLADTFEHVIEELLDEDQ
20 KVRPNEENHKDADLYTSRVMLSSQVPLEPPLLFLLEEYKNYLDAAANMSMRVRRHSDPARRGELSVCD
KKTAVDMSGGTVTVLEKVPVSKGQLKQYFYETKCNPMGYTKEGCRGIDKRHWNSQCRTTQSYVRALTMDSKKRIGW
FIRIDTSCVCTLTIKRGR

Rat Full-length Wildtype BDNF Protein (SEQ ID NO: 33)

MTILFLTMVISYFGCMKAAPMKEANVHGQGNLAYPAVRTHGTLESVNGPRAGSRGLTTTSLADTFEHVIEELLDEDQ
25 KVRPNEENHKDADLYTSRVMLSSQVPLEPPLLFLLEEYKNYLDAAANMSMRVRRHSDPARRGELSVCD
KKTAVDMSGGTVTVLEKVPVSKGQLKQYFYETKCNPMGYTKEGCRGIDKRHWNSQCRTTQSYVRALTMDSKKRIGW
FIRIDTSCVCTLTIKRGR

Human Wildtype BDNF cDNA (SEQ ID NO: 34)

atgaccatccttttcttactatgggttatcttcatacttttggttgcatgaaggctgccccatgaaagaagcaaacat
ccgaggacaaggtggcttggcctaccaggtgtgcggacccatgggactctggagagcgtgaatgggcccgaaggcag
gttcaagaggcttgacatcattggctgacactttcgaacacgtgatagaagagctgttgatgaggaccagaaagt
30 cggcccaatgaagaaaacaataaggacgcagacttgtacacgtccagggtgatgctcagtagtcaagtgcccttgg

gcctcctcttctcttcttctgctggaggaatacaaaaaattacctagatgctgcaaacatgtccatgaggggtccggcgcc
 actctgaccctgcccgcgaggggagctgagcgtgtgtgacagtattagtgagtgggtaacggcggcagacaaaaag
 actgcagtggacatgtcgggcgggacgggtcacagtccttgaaaagggtccctgtatcaaaaggccaactgaagcaata
 cttctacgagaccaagtgcatacccatgggttacacaaaaagaaggctgcaggggcatagacaaaaggcattggaact
 5 cccagtgccgaactaccagtcgtacgtgcggggcccttaccatggatagcaaaaagagaattggctggcgattcata
 aggatagacacttcttgtgtatgtacattgaccattaaaaggggaagatag

A non-limiting example of a human wildtype BDNF genomic DNA sequence is SEQ ID
 NO: 35. The exons in SEQ ID NO: 35 are: nucleotide positions 1-647 (exon 1) and nucleotide
 10 positions 63474-64238 (exon 2). The intron in SEQ ID NO: 35 is: nucleotide positions 648-
 63473 (intron 1).

Human Wildtype BDNF Gene (SEQ ID NO: 35)

15	1	attattaaag	cggtagtctg	ccggcgctga	taagcaacaa	gttccccagc	ggtcttcccg
	61	ccctagcctg	acaaggcgaa	ggttttctta	cctggcgaca	gggaaatctc	ccgagccgaa
	121	ttcagcttcg	ccggagcccc	aggtgtgacc	tgcgtagtgg	gcaagggagc	ggtgtgcagg
	181	ctgagttttt	ttttttacag	gggtaccctg	aaactcctca	ctttctctgg	gaactttcag
	241	tgccaggacc	cagtaacggg	cggttagaag	gcagccctag	gaaacacctg	ctacatagca
	301	gggcagttgg	gcaatcattg	gtaacctcgc	tcattcatta	gaatcacgta	agaactcaaa
20	361	aggaaacgtg	tctctcggag	tgagggcggt	tgcgtaaaatc	tataggtttt	tcgacatcga
	421	tgccagttgc	tttgtcttct	gtagtgcgca	aggtggttga	gagtttaagc	ttgcggatat
	481	tgcaaagggg	tattagattc	ataagtcaca	ccaagtgggtg	ggcgaaccac	tgagcaaagc
	541	cgaacttctc	acatgatgac	ttcaaacaag	acacattacc	ttccagcatc	tggtggggag
	601	acgagattht	aagacacttg	agtctccagg	acagcaaagg	cacaatgggtg	agtagcaata
25	661	aaactctgcat	tataattgaa	aaatcttgac	atgttgctta	acaacgggca	tatcacggct
	721	cttcctagca	cttcacacgc	caaagaacag	cagctactca	ggccagggga	atcgggtttt
	781	tacacagtgc	aactttaatt	ggaatcattt	gagatttgac	acagctatgt	ggaactgcgt
	841	ggaacaaact	tggagctggg	tgggggggtg	tgtgttatat	tgggtgttca	aggctgatgc
	901	ttgtctctca	gcagtcttgc	attctattct	tttccttaat	gtgtatgggtg	tatgatcata
30	961	ttctatgatt	tatatgtggg	catgtaattg	acatttgcaa	gggggttaat	ttccatctaa
	1021	aaacaataat	gctgttagag	ggtgggggtta	gggggtggag	tgggggttaag	ggtggggtaa
	1081	agactgggag	tttaggtgta	gatggggggg	ggggttgggg	ggagagaaat	aagtcagaag
	1141	tgcataatcac	cggtaatggg	taatcctctc	gtagaagaaa	aggttctcat	caacatgtga
	1201	tcaactatta	acaggatggc	tttggcaagg	ccatccgcac	gtgacaaacc	gtaagggaag
35	1261	ggaagaaaacc	gtctagagca	atatcaagta	tcacttaatt	agagattttt	aagccttttc
	1321	ctcctgctgt	gccgggtgtg	taatccgggc	gataggagtc	cattcagcac	cttggacaga
	1381	gccaacggat	ttgtccgagg	tggcgggtacc	cccaggtagt	cttcttggcc	ccgctgtaaa
	1441	gccaaccctg	tgtcgcctt	aaaaagcgtc	ttttctgagg	ttcggctcac	actgagatcg
	1501	gggctggaga	gagagtcaga	ttttggagcg	gagcgttttg	aaagcgagcc	ccagtttggg
40	1561	cccctcattg	agctcgctga	agttggcttc	ctagcgggtg	aggctggaat	agactcttgg
	1621	caagctccgg	gttgggtatac	tgggttaact	ttgggaaatg	caagtgttta	tctccaggat
	1681	ctagccaccg	gggtgggtgta	agccgcaaag	aaggtaagca	ccagggcggg	gaccccttgc
	1741	atccccaatt	cttgagctat	tttgatactg	tcttccggag	aggacgcgtg	gtggagggga
	1801	ggaggtagag	ggagagcatg	agaggggggt	gtttcttggg	atttgcccag	tttgaattgc
45	1861	cctaggtgag	aaccgtggg	caaagggaga	aagaaaaaaa	agaaactcag	tcttctctgc
	1921	gatataatga	gttttagttaa	cttggacctg	caaagtgtctg	attcaaatgt	aagattttatc
	1981	tctctttttc	tcctcttcac	ctccctcttt	tccgttctct	ttgctgggtg	gtgtgtgtgt
	2041	gtgtacagta	gattcattac	taattatgaa	gcttttgcaa	aacattcgaa	ttcctaaaat

	2101	ttgactttgt	agcattttaga	atcaggcggt	ggaggtgggtg	tgcggtgggg	agaggagggtg
	2161	gaggttggga	agagggaagg	aggtaaagct	aaacctccaa	cacaaaaaaa	tgaatcaagg
	2221	taatttcagc	tcttctagtg	agaaggattc	attctctctg	tatccctccc	tccctctctt
5	2281	tccccctccc	tccctccttc	ccgccccct	tcttccaccc	cgccccctcc	tccagcctcc
	2341	atccctccct	cattctatct	cttctctcc	gtcgcctctg	ctcctcgctg	gatgcttctt
	2401	tctgggtttt	cttttttttt	tcccttctgt	cctccctccc	cgcgagtttc	gggcgctggc
	2461	ttagagggtt	cccgttttct	caagggaagg	ggagctgccc	agaccgcgct	ccgctcccca
	2521	gcccggccgg	atgcctcact	gagcccaggt	ccgagtcagt	cggggtaact	cagggaaaagg
10	2581	ggagcctccg	cctgggagta	gaaggtcctt	tccggaccga	agagccagag	agcgggcccgg
	2641	gcgagggggc	ctgggcccgt	ggaggcggtg	gagaagaaca	cttttagctc	cgtgcggcgg
	2701	ctggacagag	ccaccaatca	gctggacgcg	cagaccgccc	tgccaggggc	aggttgcgtc
	2761	cggaggcgcc	ggtggagggc	ggccggctag	tcgctgagcc	gccgcgcca	cccgggtggg
	2821	caggggactg	gcggtgggtg	gaggtgaggg	gcttggcggg	tgagatagaa	gcggcgcgga
	2881	gccgcccaga	cctgtgttct	acctctcccg	cccccgctg	cacccccggg	ggacagcgaa
15	2941	ctgccggaac	gcgcggctgc	gttatcctct	tgccactctt	cagggagctc	agggacttag
	3001	gcgccccctg	gcgggggcca	ccaggctctc	cacactccta	taaccctcac	ccccaccccc
	3061	ttctcaggcc	ttttgttccg	gccacagagc	caagcccggg	ggcagttttc	gccccagggt
	3121	ttggtcgttc	tggactccta	ccccgatcag	ctcttcttaa	ttaaaaggca	agagcttggg
	3181	gtaggggtga	ggggaacgta	ggaaaaatct	gtttccgaaa	ctcaagacca	ctgttttaac
20	3241	gaacgaaaga	aagaatccca	actctgcgca	ggtggattca	taggcgaagc	gaggatattg
	3301	tggaaattca	gaaggaaaaag	ataaaaaaca	ggcgctagga	tcagatgacg	gtgataggct
	3361	gctcggcaca	caaagggagc	gtagggcagg	gtttacggag	caagcctgca	gcgaatgggg
	3421	cacagattgt	tccgagatcc	agtcgttttc	tcagtcagat	ctacgcgaag	ggaggggagg
	3481	ggaggggccc	gcaggggagc	gtggcgggag	gggtgagct	tgggggccc	gggatttctg
25	3541	atcagtctga	tgcaattcca	agcgtgctgc	aaaggaaatc	caaggcgccc	gcaccacat
	3601	cgccacccac	ccttcccaga	tgggtgctgt	ttaaatacgg	atctgcaggg	ctgaacgcag
	3661	aactgggaga	tttattgcaa	aatcccggga	ggggcggggg	ggggtgtgtg	gcggaacggg
	3721	gaatggagga	gcagaattta	aaggtgcaac	gcttgctttt	tccaatcagg	cggcaaccgg
	3781	ccggaattat	tatttttttc	tttctgtctg	cttgtctctg	gatttctaatt	caccaagaaa
30	3841	gaggtgtaaa	tattgtgaca	ttttgaggca	gcttgatgga	tgggaaagaa	atcatctgtc
	3901	actctaaatt	gcagagttcc	ctctccccgc	gccatccctt	gctagcgaat	actcgtctgt
	3961	gcctaataca	gttgctaggg	cttcaaatga	atgcatcggt	aagggaatat	tatcctttta
	4021	gttgacttgc	caatttagtt	gacagttgaa	tcgagaaaat	tgtagatttc	gtgtctctgg
	4081	gaggaanaat	gcttaacagt	ctaagtcttg	taaccttgag	gtctttaaca	acttaataaa
35	4141	acctcaaaaag	tgtcacgtca	tcctctacac	acacacacac	acacacacac	acacacactc
	4201	aacttgtaag	atgacatggt	ttcacctaaa	ctgttggtga	aatgaatagc	actttaaaaa
	4261	tgggtgcacct	gatattcact	gtttatgtgt	atttacaag	agctcttcag	catgaaggca
	4321	agacatttca	attgtcctgt	ttggaatcag	tcagaagact	agaagtgat	gaggaagaa
40	4381	agggaaagaa	gaggaagag	agagaatttt	aacctagatg	ctattaaatt	aacagtaacc
	4441	tagcctactt	ttataccctt	tggctctgca	tattaatatt	tctgtatgtg	agatttttagc
	4501	ttggtctagc	tcccccaatg	gagtatacca	gtattgattc	agatgagaat	atgagcatcc
	4561	tgccagtagc	tttttcagtg	tcattgatag	taagacctac	aacacagcaa	tttttgagg
	4621	atagaagaga	atatatataa	gggctttgca	aactgggaag	caggcactcc	ataaatggga
	4681	ggtatcatta	tgacattctc	tttgacata	tcattttcat	ttactgaac	cagagtacta
45	4741	gttattttta	aacataatct	aatgtatatg	ctcaaggtag	taagtgggga	ttttaaaagc
	4801	aagtgattag	ttggcttata	aaatattatt	tttcaattgt	ctattaatgt	acattggaaa
	4861	gaaggctttt	aaagatctaa	aatcaacata	aataagcttc	ccctttcatt	tgccagactc
	4921	tttccctatc	agatttctga	tctaaattct	taataagaag	agaagctggt	gaatttagtt
	4981	tctttccttt	tcctggcctg	tcctctaggg	gaagctttag	taagaaacaa	cattccaaaa
50	5041	tcaggcagtg	agcgagagag	aaggcaaggg	actggatgac	cacaaaatag	ataatcagcc
	5101	aagaaacaga	aatgagggaa	aaccagcatt	aaagcatgac	ttacaaaggg	tttttatttt
	5161	gtaattctgt	aattttggga	ccaggctcaa	acttgctcaa	gtaacattca	ctcgatcata
	5221	ttgcttacia	tctgtcagtt	aaaatgatgt	ctaactgagc	atatttttta	ttaaatatac
	5281	ttctctcaaa	ggccagtaaa	gctactcttt	ggttttaatt	agacaaacta	gtctaaccac
55	5341	ttaaataact	ctaataaata	tgaactgata	tcatacagatt	taaaagctct	gctgaaaact
	5401	aaattttattc	tgaaaagcac	tgacttgcca	gaaaaaatatc	tatttttgca	gctttctttt
	5461	cactctatgg	ataatttaac	gagttgctta	ttttaatttt	acaactgcta	cctcagaagt

5521 atctcaaaatt atcttttcttt ggctgggtgtc tttctctgtct gatctgctac tgctgtgtgt
 5581 gtgtgtgtgt gtgtgtgtgt gtctgtgtgt gtgtgtgtat gggcggtgtgt gtctctgtgt
 5641 gtgtttttcta gtgggaattt aacaagcagt gagtctctta aatttacatg ccataatcta
 5701 tgtcaagaac attgctact acttagcaat aaaaataaac attagcatct agtgaaagct
 5761 taccatcatt gagtgcctatg gaaatagagg tcttaaagaa agattaaatt tttcaacaaa
 5821 aaaatttttc cccttttttg cttaaagggtg attataattt caaaaatatg acatctttcc
 5881 tcttttactt tggaatgtag agctgctgct ttaacaagtg tcttttgaga aagatacacg
 5941 tgtttcataa agattaatac ccttaaaaaca ctatggtgca gagagggaag gatgaattct
 6001 ttaacccctgc ctctaactctc atttggcaat ttttggagta ttcattctga ctttttaaaa
 6061 attcagggtg attttttttc tgcttttctt ccaacattat aaaacaatcc tataagagat
 6121 ttttctgcta tagtgcagac tttatttgta tttcctagta ataacacttt agattcatat
 6181 agtactttga cagctctcta taggtttcat ttgatttcct tatcagtcac gtaaggtagg
 6241 aatcaccaat caccttttac agatgaggaa agtaagggtc agaattatct aacactacac
 6301 tgccagtaag atgtaaagac taactcagag tctttcttca aattttaagg aaatttgtgt
 6361 ttgttctctc ctagaccatg ctgacctaaa ctccactagg gcatcagagg gagctgtagg
 6421 cattattttc tcctattttg atttatttaa ttaaattaaa aacatttttt acaaatagtt
 6481 ttcaaaaatt caggcctaag ggaaagtgtt aactagtctt cctaagacag tatttcctc
 6541 tcccacagtt agacactcaa agaaagcagg actcttctc tagttgacat accatctaag
 6601 tcatagttgc taattcccca aaaaacaaat acaagaata agaagaacca aagccaagat
 6661 gtacattacc tttacttgtg aatcatagaa tctgggttct gggaagggtc tcagagggtca
 6721 cttagcccag ctcatatctg atacatgaat tcatgctggc ttgttttgtt tcaatcctct
 6781 ggaatttcat ccagctcttg gacactgggt acttggttat cttggtagat gttctgatac
 6841 ttatgtggat catgagcttg gctgatcatt tcccattttg cccaaaagg atactttcta
 6901 tagagactct agcattcata attttatttt gtaagtatta attggcaaca cataatttgc
 6961 cattatgtgc caggaaactgt attacaatct ggagattcaa gaacaagcaa gaccaaatgg
 7021 tccctatctt ggtggatctt gtggcctaga gacaaagact ggtctcttta tttgctccaa
 7081 tccttagaag gggacagacc tctggactca taagattcct ttttcaaga gaacacagt
 7141 aataagatta aaagcctgag tttaggcttt acttccaacc ccttgcagt acccaaaaa
 7201 gtcaatcagt cactcttgat ctctgcttcc tcacacataa aacaaaagag atgttctcag
 7261 aataaactct aaattccact tccgttccaa gtttgagtga tataaaagat atagtctata
 7321 actatttctt gcagtgtagg ggaaattaag gcctaagtta ccctaacctt tgggtgattta
 7381 caattctggg tgggtaccgc aaatttttaa cttgttaaga agtatatcat gaaaaaatc
 7441 aaagtaacat atttgaaccc agaatagaag gaatctaggt gttgaacctg ttcatttatg
 7501 gaatgatggc taggaaagtt ttaatttaga gaatgatctc aaatttctgc ggattatttt
 7561 taaaagcagg tgggtgcgat gaagacctgc tatcaaattt actgcttatt tcttttgtgc
 7621 aagcagaagt tataattttt caggtcattt ccttttaaaa tcaaaaatat tacatcctga
 7681 aattgcctgg gtctcatgaa taatgcatta tatacacatg ataatagata attagatgga
 7741 caagccaaaa gaaacatgaa aggaagcagg tagcccaagg attgcagaag gtgtgtggc
 7801 attttgacat ccaggaaatg ctatagatct gtcccttaact aactcagcct ggtggagata
 7861 attaaagaaa aaaatgtggg tgtagaaaga ctgcaagcca ttccctggga ttggctagat
 7921 tgctgcagta gttcaaaaac aattggcaca gccaccaca ctagacatga ttgcccttg
 7981 atgaggcagc tattgacttt tataaagatg tcatatttaa aataacttct gatgcactag
 8041 gcataacaga catcattctt gaattctatt ttagacaaat ggcaaaatgt attacaaagt
 8101 attaatTTAA aaataaaaaa tctttaaagt ctagtgtcta aaaaccagca gtttagtaac
 8161 atgcaacctc tggatttaag aattcagcct gaagctggga gaaagctgta gcttgatag
 8221 gacattttga tccactctgg gcatttccca gaccactaca ggaagtaaaa tgactttgt
 8281 caaagttttt aacctttgag tgaatgttaa atccactcca aaatcttcgc aacctgggaa
 8341 aggtgatcca acaattttcc taaatagcgg cagaaaatgc tctgagatct ttgttccag
 8401 agtgaatgtt ataattgtat gctatctaga aatttccttg tagcaccatg ctcatcagta
 8461 ccaaaaggag ttagaattga ttccctccgc ttcaaggaaa tatatcaacc acctcctgtc
 8521 tctaagtaac aaggttactg tggggaaaaa atacacaaat taggtgattg cagaaagggtg
 8581 tcacaaacat ccaaagcctt tgggatagg cattgcagtg tgagtgaata gagaaaagaa
 8641 agagaatgtg ggaaaaaatt gagaaataaa aagggaagtc acagtggagt tctaattata
 8701 caggggctct tgaattgact gttctctacc ttccatgctc attgtgttgc tggctacttt
 8761 agtaggaaac aatgatttct tctgctttca ccttctcctc ccgctaagga cttcttactt
 8821 gccataaact tccataatca atgtttaaga attgctctga tgcccagtg ggtggctaac
 8881 gcctgtaatc ccaacacttt gggaggccga ggtgagtga tcacttgagg tcaggagtcc

	8941	cagaccagcc	tggccaacat	ggcaaaaccc	tgtctctact	aaaaatacaa	aaaattagtg
	9001	gttcatgcct	gtaatctcag	ctacttggga	ggctgaggta	ggagactaac	ttgaaccatg
	9061	gagacagagg	ctgtagtaag	ccgagatcat	gccattgcac	tccagcctgg	caaaaaaaaaa
5	9121	aaaaaaaaaa	agaaagaaag	aaagaaagaa	ttgctcactg	taatgacttt	catgccatgg
	9181	actcaactct	cttggcagtc	tggtaaaagct	tatgtaaacc	cttctcataa	aaatgtctaa
	9241	atggggccag	gcgcgacggc	tcacacctgt	aatctcagca	ctttggggagg	ccaaggcggg
	9301	tggatcactt	gaggtcagga	gttccagacc	agcctggcca	acatggcaaa	accctgtctt
	9361	taccaaaaaa	aaaaaattag	ctgggcgtgg	tggcatgtac	ctttaatccc	agctacttgg
	9421	aaggctgagg	cacgagaatc	acttgaatct	gggagggtga	ggttgcagtg	agccgagatt
10	9481	agccactgca	ctccagcctg	gatgacacag	tgagactttg	tctcaaaaaa	aaaaaaaaaat
	9541	tctaagtga	tgaaataaat	gtataagatt	acaaaggaag	ccagtggcat	tgatgtacag
	9601	ttataaaaaa	atttaaaaata	atatattgtg	tgatataagta	atatatatgc	tttttaatac
	9661	attaaataag	atctaacagc	aaggtaaata	ttataatttt	gaaataatga	taagtataca
	9721	tgtattttga	aatatctata	aaactgacgt	gatatgaagg	tgtctgtgat	gtatactggg
15	9781	gagaaaagcat	gcaagtacta	ctgtgtgaaca	tttcccacac	atatttaaca	acagaacact
	9841	tgagaagcac	ttattaacac	agcatagatt	cagaaatatt	aatttagtaa	atgtcaacat
	9901	tagccattgt	tgtttccttc	ctggcaaaag	gaaatcagca	ttggggagaaa	acttttaaaa
	9961	ttcacatttg	ccattagaca	agctgtcaag	tggggaaaagg	accaaatact	gagaaggcca
20	10021	gggtatggta	agcatgtttc	tattgactga	gcttgctatt	actctaactg	ttatctttag
	10081	catcaccagc	acaaccccat	taccctagca	atccatcact	ccattgaaaa	agataaaaaag
	10141	ttcagattct	ggtcattaac	tcagcattgc	ttaagatacc	tgttctgacc	tcactaaccc
	10201	aagagattac	tgaaactctt	cctgtttgtc	attactacac	catgggaaat	tataatgatg
	10261	tgggatgaca	tttactctgc	attcatccag	tgctgttatt	tgttttgtat	ttggcatata
	10321	ttacttaact	cttaaagtaa	ctctcagaga	tagattaaga	aaactagagc	tcagagaatt
25	10381	taagtaactt	gcccuaacta	acacagaaaa	tctgaagtgg	agaagctaaa	cttcaaacc
	10441	aaagtthtct	ggttccaaag	tccattatga	agttgtgcct	ccccatctta	tagctaccac
	10501	ccagatttaa	tctgggtctc	ccattatcag	atggtttaca	tacacatttt	cttacaagat
	10561	cttgaccaca	actctttgag	atggccatga	gtctcacaat	tcatttccag	gagtgtact
	10621	ttagaatcat	tttgatcttt	gctaaccgat	gagagatttt	caaatagcta	attgtcacct
30	10681	accctthtgg	aagcccagtt	ttcataatca	taaaatggaa	acagtattac	aatgtthtgg
	10741	taggatcata	tacattaata	ataaaatcta	actttgttga	gctcactatg	gtgagcattc
	10801	tgcatthtct	tagttcattg	aatcctcaca	acaatctttc	taggctaaga	cgattatttt
	10861	tctthttaaag	ataaggaaac	tgaggcatca	gtaattaatt	aactatctta	aattagcaga
	10921	gccaataagt	ggcaaagctg	ggttcaaacc	taggtctgtc	taatgtcaaa	gccctthtth
35	10981	taatcactaa	tctgcaaatc	actattcaat	cttagcttht	attattataa	ttatcatcac
	11041	acttaaaaca	ctatcaagat	acagaatgat	ccagacataa	gtatatagtc	actgaagaga
	11101	ttagaatctg	aaactthtca	cctgcatgth	cttccttcca	ctthtthtta	tttaaccaat
	11161	ggatgatgtc	tgactcctth	cttaacttgt	ttagggcagt	tcctaagttag	ttgacttctg
	11221	agagttattg	agtaagaat	gttataaatt	gtttggatta	ggatttagta	tgthttagaag
40	11281	ctattthcata	agthtgcctt	tgcgaaactgt	tactggctat	aatactgcag	atgctgtgat
	11341	gaggaacacc	ctctccaaag	acacacagtg	gatgacaaac	ctccaaagct	aacatgttgt
	11401	ttacagatat	ggagaagaag	aggatggaca	agcacagtct	aaaacgtaat	tacaaggctt
	11461	atagtccttg	ttggggacta	gaatgttht	tggctthtct	tgtgcaattc	aatgctcttc
	11521	ctccaaagga	tccactccaa	acttggaact	ttcctgaaaa	tagcatctca	tttgggagca
45	11581	tgccaggaat	tgggtgtctgg	gtcctthtgtg	tctthtgcacc	aactcagaac	tctggatact
	11641	agctctagaa	actaagctgg	gatataattct	gggtaaggga	gtagcatatc	tacttgggca
	11701	tcttctgat	acattthtth	catccatctt	cctcctagag	agcacctcct	agaaagatgt
	11761	ggtthttaa	gaggatttg	atgcatactg	gtatgtctta	gcacacaagt	cagtggctct
	11821	tgagagctg	ccaaaggcat	ataagtaatc	aaagatgcgg	aagtctatga	agagacttca
50	11881	tcccacctcc	actctgattt	attcagggaa	ggaccccatg	aacacataat	ggatttgata
	11941	cgtcccagag	ctctgaaagc	agcctagcaa	aaaaggataa	tcttgaaagga	cattthttagt
	12001	tatgaaaaag	tccaccta	gctthtgcag	agataactaa	gtaatatgat	ggctggtaga
	12061	ctgtaagtcc	ttaccttggc	tcaggaactg	tatatcattt	ggtaactaa	acttgtcgtt
	12121	caaattthtga	tagaaaaagt	accttacaaa	tgatctagtt	cactgattcc	cttcatgcat
55	12181	tgaatcacc	taaatcatct	cttctthtctg	agataaggct	tgaatgtgtt	gccagcttht
	12241	gcaaactcag	ttttagctcc	actgacctca	tttgattgat	tgggcaactg	aggtgcacag
	12301	tggtagatct	ctcaatttht	tcaataaaca	attatatggc	ccttacgata	tctatctgaa

	12361	caatcttggg	ctagtgaagt	tgcttgccca	ggttacatgg	ccagaaaactg	acagtttttaa
	12421	attaggacca	aagttctttt	gactactatc	tgggccttaa	aataatatca	tatgacaaag
	12481	atatttcttc	tgtttcctaa	tagtcacatc	aaaaggaaac	aatggacagt	ttgtgcaaga
5	12541	ttttagttac	tttaatgttc	aaaataaaat	taaaaacaga	ttattactaa	aacataagca
	12601	taacaacact	ttaatagcat	tctaatacaga	tattattaat	ttcaaaatgg	taggacaaaa
	12661	ctaattatac	tttatacttc	ttaaataatcc	tatagttact	ttatgactat	tgagacacta
	12721	gctaaaaactt	gaaacttcaa	gttttcattg	attcctatat	tattacttat	ttcagagtta
	12781	cttcatttgg	ttctttttatc	tgagattgga	caacagcttt	atttgatttt	cagcgacaaa
10	12841	attctttttca	ctcctgatcc	tccaccccaa	gaaaacaaca	gctactaata	tattttccct
	12901	aaagtgtatca	agaaataaaaa	gaggaattct	agccaggcgc	ggtggctcat	gcctgtaatc
	12961	ccagctcttt	gggaggctga	ggcggttgga	tcacctgagg	tcaggagttc	gagaccagcc
	13021	tggccaacat	ggtgaaactc	cgtctctact	aaaacttaaa	aaatgagcca	agtgtggtgg
	13081	cgcattgcctg	taatcccagc	tacttgggag	gctgaggcag	gagaattgct	tgaaccagcag
15	13141	aggcagaggt	tgcaagtgagc	caagattgcg	ccattgcaact	ccagctctggg	tgacagagtg
	13201	agactctgta	tcaaaaaaaaa	aaaaaaaaaaaa	aaagaggaat	tctaaaaatta	attatatcta
	13261	ttaatatccc	tactcttaaa	acgttagaaa	atgtttgctc	atttaaaatt	tttattttta
	13321	aaaccacctt	atattccaac	taaatactct	ttggagcaat	ttctttgttc	ctcatataat
	13381	atccatacat	ataattctgc	ttttgtgatt	aacttttatt	actactcttc	taaaattgtg
20	13441	ctcttataaa	catcagttaa	ttaagagttaa	atctgatgtt	ttataaaattc	ttctagaaa
	13501	cagagagcaa	aatcatataa	ataacaatat	gaattttcaa	aagtacaata	ataaaaaaaaa
	13561	attagaaaaa	aattaatcta	ggaaatagtc	aagaatatgt	caaacttgta	catacttttg
	13621	agataaattg	gcatcatgta	gatttagcatg	attcttcttt	atggaattca	acttattttt
	13681	actcactttg	ctctaattag	ttttgtgtgtg	cggacaagat	ggaaggtaat	ggaaatttgg
25	13741	cttgcaaagt	agttctaaca	tgatctacat	ccacaatctg	gttataatgc	tataagaata
	13801	ttatgtggga	atagtagttc	aaatcagtat	ttagtatgaa	cataaaggga	caaacaatgc
	13861	aaagctaact	taagttgttt	acacttggaa	cttattttaa	ttaaaaaggc	cagtggatgg
	13921	tcatatgttt	ggctcattct	tctcaaggcc	ttcaggaaaa	catgcctatg	aaataaaaaga
	13981	tcctcaatat	taaacatttt	actgcatttg	ggggacacat	gaaattctgtg	ataaaaggaa
30	14041	gtgttggtct	tcatttttct	aattcagcat	ggaaactatc	ttgaggaaaa	ctgactatgg
	14101	tcttagtttg	tgtctcagaa	atatatttag	tctgaatcat	ggcgtcgaca	tctgacttcc
	14161	aaaattggat	atctagccgt	atagtacctc	acctcccaca	cacaccaccc	cccattccca
	14221	ggtcatgact	actgtccaag	cagcaaaaaa	agaagtaatt	tcccagagta	catacatggc
	14281	agtgacaacc	aaccaaaaca	aaaacaatta	taggggctgg	aatttaaatt	aatggctgta
35	14341	ctctcaccaa	ttcattcccc	attccacccc	atctctctgt	cttcaacttt	tatgaaacat
	14401	tatatttgtc	ctattcttct	gtatcagcat	cagcctttcc	tatatccaac	tagacttata
	14461	acttcttggg	gcctctcact	ggctgactaa	ggtttcagaa	gtacctactt	acagcaaaaca
	14521	cttgacagcag	tctctttttg	gttacaaaag	ccctggacaa	tttctcaagg	cgatattatg
	14581	aagagggaag	aaacattctc	ctctgctact	ccatttcttt	ttagagtgtc	aactttattc
40	14641	tatatctggt	ttaatttctt	cttaggccaa	ttggactgat	ttacagaca	ccatagaata
	14701	tctcctgagt	aatgggaaca	atatttctgc	tgatcccatg	atttggcttc	attgggttgt
	14761	taggccataa	tggagacata	cttgatgaat	ttatgaagac	ttgattctag	gtatcatgta
	14821	ggttagcata	attctctttt	actgaattca	acttagtttt	attcacttta	ttctaactgg
	14881	attttgtgtg	cagaccaa	gaaaagaaat	ggttcaattt	aggtgaaagg	taaagcttca
45	14941	aaagtagtgt	agtatttcat	agaccttacc	tttgagagaa	attatatcag	tatataataa
	15001	gcacctgaga	atatgaaagc	acaaatccaa	tttaaagtgt	aaaggtctac	aacttgggat
	15061	tttaaagtga	gtacagaaaa	gccactgttt	cttaaacaa	tttgttgagg	gggaaaacag
	15121	tgaagctaa	atgttctatt	caagagttgt	ttcttttgaa	aataatgctt	catttaaaag
	15181	ctaaggacag	aagacgtagc	tttgttatga	aggctcatct	ttttattaaa	caaccactac
50	15241	tttgtctcca	agttgcaaag	ggaagatttg	tcaatctgat	tgaatcttcc	ctttagtttt
	15301	tcccaacagc	tgtgtccaga	taattcatga	ctcctgtgtt	tcttgagccc	tggataattt
	15361	cacacacatg	tctggtttgg	ggctccacat	tttcagaaaa	atatagaaat	cttggaggag
	15421	gtccagggtga	gaccaaggga	aatgattaat	gggttgaaag	ttggagttaa	tgaagaaagg
	15481	ttgtgagatc	tgatcttttg	ctacgagaaa	agtctgagtg	gtgacttaat	aacataagga
55	15541	ggttagtaag	cagctgttct	ccatcttcac	taaggttgaa	tgaaatgaaa	taagatataa
	15601	attgcaacag	gaacaaaaat	gcattacaag	tgaggacttc	caagcaccag	cattgctgga
	15661	ttctagatag	ctcccaaaaa	gaaggatgtg	tagtctactt	ccctggctctg	caatgacagg
	15721	cctataaata	gtgagaaaaga	tgagataatc	tcttaagatc	ccttctggac	ctatctttta

	15781	taggtctatc	tatcatat	agaaaaatta	tttgctcaa	acaaaaatta	tctgatttcc
	15841	tccctctcac	cctatccact	ccttctcttt	tgtctacctt	ttgtaaaaca	ctgctaaccg
	15901	aaataactgg	ggactgatta	accgtgggtg	gccctcccc	gcctctaagt	gccactccag
5	15961	ctttgggagc	aagtttcttg	tccatcacta	ccacccctg	gccactaggg	gcatgtttac
	16021	catcatcttt	ctacacacca	aacctacggc	aagggaaaaat	aaaacaaaac	aaaacttcct
	16081	agacttaaca	aatttgcaag	tgtcaccatg	gattaaaaata	caactcttat	gtcctagaat
	16141	atgagcatgt	aaagggctaa	aatgtat	atgcatctgc	ctgtatcagc	ccatagaata
	16201	gcctcctgac	agatagtaga	tactcagcaa	tctttcatca	actgaatgac	tgtaactatg
10	16261	aagtgaaagg	caactaaagt	tgagaaagtc	aggagtttcg	gatgtttcca	aatgattctg
	16321	tatgccagag	taatctaaag	cctaaccocat	tcttcacaac	catgcactat	taaggatttc
	16381	attctcacca	tgctgtgct	atctggaggt	agaaagaggg	ccagttgcac	atcctgctca
	16441	agtccttggg	caaaaagacc	actaaagagt	gctttgtaga	ttcatgtatc	agaatcacat
	16501	gaaagtaggc	caaattctta	gtgtgtgttt	tfaaaataag	actttaggaa	gttcacttat
15	16561	ttttttctaa	attat	catattcttc	tttttcattt	ttttcatgaa	gaatttaaaa
	16621	tttggctgta	gaaaatctct	cactccaaac	atcacacagc	ctaaataggt	gagtctcaaa
	16681	aataagctaa	tgttcatctt	tcatctgatt	caatgtcctg	aaaccctttg	gtttaaat
	16741	gttaattctt	ctcatggctt	ttctcctagc	aaaaccaact	aataccacag	ctatttatta
	16801	ctgtcagctc	taacttatgc	ccacaatctc	acatcccttt	tgaccacgct	tatagaacta
20	16861	ttacaacaag	taaaccaaat	ttattcttca	ttatttaattt	ttaaatgttc	tcagcacaaa
	16921	tctggtaact	tggagggcta	caagttgata	tttctcatat	gtttgggggt	ttagtctcaa
	16981	cagtttctta	atggttttcta	tgccgttttt	cttgatccaa	ctaaatatta	ttcccagatg
	17041	ggatcagctt	ttgaccctct	tgttctactc	tcctagtctt	ggcccttcta	aaagtttctt
	17101	gctgtgggtt	ctttcttttg	tctgccacta	atggctatgc	ctggttacat	aactcctgta
25	17161	acaggtgttg	accaatttga	acacattttg	gtatggtatt	gagctattct	tatggttcat
	17221	aaaaagctta	gtgagaacgt	aacatctcat	gaatagggaa	attacttctc	ccttaagggtt
	17281	tttctcagga	caggcctcat	acaagaat	caaggattgc	gagtgacata	gtttaacatt
	17341	ggaccaggcc	tttcaaatta	tccaggatga	gtttgaaaac	acctgtgcca	ctctgtctaa
	17401	cagcagagtt	ttctgtttac	taagtatttt	ccctatgcta	attacggaaa	gtttcaacag
30	17461	tttttttagg	ccaacttatt	tgatgctaga	ctagacaact	tatttttttt	ttcttgcaag
	17521	gaataactgaa	ggtaggagta	actaggaagc	ttaaataaac	ataaatataa	aatgcttata
	17581	gtgatagaat	tgacctcagc	caattaaaat	tattaataga	aaaaacatgt	caatgtcaag
	17641	cctactacct	ctgttctcac	ttgagtaatg	aggattagtt	tatatttccc	gacaagaata
	17701	gatgggaatt	caaatttctt	cctgaccttt	gttccccctg	gaacattggg	ttaggatcat
35	17761	attagaacat	aaccaaaaag	aaataaagat	tcagacgaat	tcacaattaa	tttttaagcc
	17821	ccacaaaagt	gaaataggtta	gcattat	ttcaagctgt	gaaactttcc	ctcattttag
	17881	taatagagaa	aatgttcaga	ttataaaactt	ggaaactttg	ctcctaacat	atcaattatg
	17941	ccagaggcca	atttttaaga	agaagagaaa	tgcatgctct	atattctcag	catcatcctt
	18001	gccacaata	gggaaataat	tttgtaaaat	gtttgatttt	agacctccaa	aattatctct
40	18061	atatgctacc	tgaatttaagc	aaataaaaaa	taatat	aattccatgc	aaggcactgg
	18121	tacaattttg	tttatcttgg	cttcattgtt	tttgaatgta	agatgtactt	ttaaggcaaa
	18181	taagtacatg	ttttaagctg	gtcgcataca	gtattggcaa	tgctataatc	acaaatcaga
	18241	aagtttggaa	atgcttacaa	gtgttaagag	gtgtgattca	tcatgggtat	ctgaattggc
	18301	atctgatctt	cttttctttc	taaatatccc	tgacatttct	gactcctctg	tcttttcctc
45	18361	agtaaaactg	caccacacac	tggaaagcga	agatacacac	at	ataatgtcaa
	18421	gggagagtag	gaataagaag	attggccata	gacccaccca	atcagagtct	gggaaatgag
	18481	aacacttttt	ccttcagcag	aatgctgac	gtgccaatgt	gaatttagca	gaaaaaagat
	18541	ttgccataac	ttctaagtga	gcagccttca	gaatgctagc	ttagattcct	ggcattaact
	18601	tgccaggtat	tttttcagga	aggaaataaa	ttacaattga	gcttaaaaac	ctgagggtag
50	18661	aactcat	caagcaaatg	tgaagcatca	gtttgaagtt	aacaaagtta	aagtttggag
	18721	tagggttcct	ccagtccttt	ataatgtagt	acaagtattt	tttttaaatg	tataacacta
	18781	gcctttttaa	ttgtattgtg	ctactaaaag	aaattgtgcc	tgcatcctc	ttacaacctg
	18841	ggaaccaacg	cagaggtctc	gtgggtagc	ggtatccagc	ttcatgccct	ctgtccttta
	18901	ttgctttctg	gttagcctgc	gtatttcaca	tacattaaat	attccacaat	aaactctgcc
55	18961	atctgtgctg	tagggtagtt	tgtattggtc	atgtgctctg	tcaagttgac	agaggtgcaa
	19021	agctaaatgt	gtgacactcg	aagaatatgc	atataattga	ataatttgac	tatttagtcc
	19081	aacaatttgc	aaaggcgctc	tgaatgatca	cacattctga	taacacttcc	aaggaacaga
	19141	tagcttcact	taggggtg	gggagatgga	agcagggtta	tttctagcag	gaattcttga

5 19201 gttcactgaa gtcttgtccc tggactttca ctgtgtgaac gtgggtaaat tatttcctgg
 19261 cagaggatcg gattcttctt ttataaaacg ggtaaataat ttctgtcact agtctttaga
 19321 agttctaaaa tagctaattg tagtgaattc attttgctaa ctgtaaaccct ttaggtaaat
 19381 tgaactgagt atgtaataat attatatatt cagttcaaca gcacattctt ggtaaccaca
 19441 agaggggtcca ggaaaggaaa ctgtttataa atctttccct ttagcaaaat taatgtttgga
 19501 gtcttttaggg aaattctttac agcaatagtc ttcgcaatta ttaggtcaaaa cccctttgag
 19561 attacagaaa aacgcacaca cacagaaaagc tgcctgcaga atttgggtgt gggcttggtg
 19621 ggagattcct ctgataccca gtgttgtacc cccaagagag tgtttctcaa agtgtgactt
 19681 cagattgtct gcattcgaat tgcttgtggt atttattaaa attataactc ctggggccctg
 10 19741 cccacccct actaaatcac aatttcagga ggagggacct tcattttaac actcaccag
 19801 gtgattttta tgctccgagg aggtccaggg actccaagtt aagtacggta ctgctgtctt
 19861 attctttatt ctaaaattta aggtctgcac aaattggtt aactaatgag aagaaaattc
 19921 agctttaaag cagaaacaca ggtagacggt tgacagagtt catcaaatgg ataattgaaa
 19981 atgtcctctg gaccctagcc atataagttc tcttcaaggg tcttggctac aggcaaatga
 15 20041 gaaccgaaa ggctatttgc tcttttgctg cgggcagtggt tgggggtgga gggcggggga
 20101 ggattaactg agccagttct gccccaccc tcgaatcacc tacccccact ctggttaaag
 20161 cagaagactt tttattttatc ttggctgccc tggttcgta ttaaaagggt tagcttatac
 20221 gtgtgtttgc tggggctgga agtgaaaaca tctgcaaaag catgcaatgc cctggaacgg
 20281 aactcttcta ataaaagatg tatcatttta aatgcgctga atttggattc tggtaattcg
 20 20341 tgcactagag tgtctatttc gaggcagcgg aggtatcata tgacagcgca cgtcaaggca
 20401 ccgtggagcc ctctcgtgga ctcccaccca ctttccatt caccgcggag agggctgctc
 20461 tcgctgccgc tccccccggc gaactagcat gaaatctccc tgctctgcc gagatcaaat
 20521 ggagcttctc gctgatggg tgcgagtatt acctccgcca tgcaatttcc actatcaata
 20581 atttaacttc tttgctgcag aacagaagga gtacataacc ggcaccaaag actcgcgccc
 25 20641 cctccccct ttaattaagc gaagggaacg tgaaaaaata atagagtgtg ggagtttgg
 20701 ggccgaagtc tttcccggag cagctgcctt gatggttact ttgacaagta gtgactgaaa
 20761 aggtgggttt gttttctttc tttctctttc cgtttttctg tttggtcggc tagaaagcgt
 20821 gtggctttag cgaggtctgt cattgcctgg gcttcttggc tggacaagt aacttgggt
 20881 aacgttatct gggggcggtt atcaataaaa aatgctgtta ttatctgat tgaattccta
 30 20941 ttaggcaaac tctagagagg tcagtgcgcg aactctgtt aagccggcgt gtttaaggca
 21001 gcagagtaaa ccaatagccc ccatgctctg tgcgatttca ttgtgtgctc gcgttcgcaa
 21061 gctccgtagt gcaggaaggt gcgggaaggt gtgtctgtgg cccgggaaac gcacgccctc
 21121 tcccagagaa cttgggtgct gggatgggga ggaaggggag agttgaaagc taggggagcg
 21181 agacctcggg gcgtgcgatt ctactcgtc cctcccgc ccagcgcca cagccgggt
 35 21241 ttctgcagag ggcgcgggac gcgggttcc ccggggctga ggctggggct ggaacacccc
 21301 tcgaagccgc gggcgctcctg tccaaggcgc cccaggaggg cgcaggactc gcagggcgat
 21361 gtcgcggggc cctaggggag gaggtgagga caggccccgg gggagcgggg agttccgggc
 21421 gccctcggg tccccgcgcg aggaagagc gcggcgttcc cttaagcgg cgcctcgaa
 40 21481 cgggtatcgg tagcgcggg gagcggggag cggggggcgg gggggggcgg gggggggcgg
 21541 gcgcggtttg accaatcgaa gctcaaccga agagctaaat aatgtctgac ccgggcgcaa
 21601 ggcgagccct ggagctccgg gtccccgacg ctgccgccgc cgcgccggg cgcacccgcc
 21661 cgctcgtgtt cccgcgcacc ccgtagcgcc tcgggctccc gggccggaca gaggagccag
 21721 cccggtgcgc ccctccacct cctgctcggg gggctttaat gagacacca ccgctgctgt
 45 21781 ggggcccggc gggagcagca ccgcgacggg gaccggggct gggcgctgga gccagaatcg
 21841 gaaccacgat gtgactccgc cgccggggac ccgtgaggtt tgtgtggacc ccgaggtagg
 21901 caagcgtgtg gaatggggct tgggtgcagga gctgccgctc cgcgggagag agttgactgg
 21961 gggatcccc accccaaagt tgtgggacga ggccagtctc cttctttcct cccctccggt
 22021 agaagggacg atttgaggtt actcttggg agttttctcc cccatccac aaccagaag
 50 22081 gtcagccgg accaccagg aaaaagggac ccggggaagt cacgaagtag aggagggaag
 22141 gcctggagga gaccagagc tgcgtgatgg gagcaaagac ggcgaccgg ggatccctcg
 22201 cagccctccc ccagcccagg agtagtcgag agagacttag ggggcccag ctgtcgaggg
 22261 tcctgactga ggggaggggt ctggggctag gctaggaatc cttccagggg gtgggtggct
 22321 cccgcgccga cttgcgggg gagtgggagg gaagcttgcg ccttcagccc gcatcccttc
 22381 cccggagctg cacacggcta cctgctcccc aggaattgag actgaagtgg acttacaagt
 55 22441 ccgaagccaa ttagctttgg aaaacttgg aggcggaatt cctaccgctg ggaactgaaa
 22501 gggctctgca cactctcggg caggccgaac ccacatctct acccatcctg cgccctctt
 22561 ctgaagcgcc ctccagggaa gttaagagtt ttgactttcg gggagtgggt gggatgtacg

	22621	tggggggattc	ttgactcggg	ttagtctctg	gggatgcaga	gccgggaaga	ggaatgggtg
	22681	agtgaagttac	tcctggaaaag	aaatagctga	ggattggggg	ctctgtgcct	gacgggcaag
	22741	aagaaggggga	gattacagac	taggggcatc	cctaagggaag	aagcctcggg	gctgcgaggg
5	22801	tgaactggag	gatgcagtgt	ttgtgtgttg	ggggtagagc	ggggatgagg	gaccgggggtg
	22861	gaggggaggc	gaggaggagg	aggggaccca	gagaacgaag	ctaggggaagg	tagaggggtgc
	22921	cctctgccgg	ccatgctgcc	aagagcagct	actggggggc	ggaggctggg	ggtgggggaag
	22981	tggtaaaagga	aggttttgcg	ggatccctta	gagagctggt	aggagggact	tggtgaatgg
	23041	tgctgctgac	tccagctcgg	tggggcggtc	gactcgtcgt	cggtggattt	tgactcctcg
10	23101	ttcttgtttg	gcttctatgc	aagttttcct	cgcgctgggg	gagctttgat	aagcctcgat
	23161	tggcgggtgtg	tttagggcttc	ttggatctta	ttttagggtc	ctctagtatt	cctgcactta
	23221	ctccttaatg	tcagtagcaa	ccaaagaaca	ttttccgaca	agcacgcagg	aatgttcttg
	23281	gccagaagca	aagaaaggca	tattttctgag	tgtttattaa	tcctcctagt	aatcttttaa
	23341	agcaaagtaa	tatgtaattg	ggaacgttga	ttttctaact	gcataataaa	ggcgacatga
	23401	tattaaatga	gaccctccc	tactgactca	atatcctgca	aaatctctct	ctccccttta
15	23461	ttattatgga	aaaaatctatt	tttatatgag	tttgttgtaa	ggtcaaaaagc	catttttggtc
	23521	ttacaatttg	atatgtcttt	acatttttaac	ttattgaggc	ataattacag	atttaatttg
	23581	tatgaacgtg	tgtgccttca	atgcttatct	catgcaacat	aatttttagg	ttggagattt
	23641	ctgatgttat	ggcatgtagc	gtttcaaggc	attacacata	ataggttaac	tagcatgttg
	23701	aaattacacc	acaaagtttt	gaccctggga	acagcacctt	ttaaaaacaa	tcactaaact
20	23761	cctgttcctg	ttttctgatt	ttgcaaatgc	cttgcttaag	actttttttt	tttttttttt
	23821	tttttttttt	ttgggaaatt	tacctctggg	ttagcaggag	aggtaaaaaa	aaggaaagag
	23881	acacttggtg	aaatgtaacc	ataaccttta	ctggaattta	aaacatgttg	gtcaccatta
	23941	ctggaattcc	agggccataa	agtcgttggtc	ttttttttct	tctacttcat	tttgtaaaat
	24001	gtgataaatg	ttggtaaata	tagaccagta	gtaagtatta	tgacactaaa	agcattatgt
25	24061	atgtggaact	attttaaagt	attacagaac	attttctatt	tataaatgat	ataagcagaa
	24121	agaaatgatt	tccagataaa	caaggcttac	gtacatgttt	tgaagcatta	gaacattgca
	24181	gacactctta	gacatcacat	tttttaaagc	aaaataacag	taatttttca	catacctttg
	24241	gagcctttca	tagcccattc	agagctgagt	tagtagctgg	aagtttcctt	tatttttaagg
	24301	tgatatttta	aaaccattta	acatgtatag	taggtcaaca	ttggtgcatc	cagaaaatga
30	24361	agcattttagg	aaatctgttt	cagtgtcttt	tcaatgtgtg	taacttttac	ttgcaaacca
	24421	atggaaccaa	gaaagtcatc	atttgcctaa	aatgcagtca	tcacctcaaa	tgattcattt
	24481	atactatgtg	agttaattgc	cttcatctca	ttaatggcca	aggaggggaag	ggaggtcctg
	24541	gggtattttct	tgttcatttt	gactcaccag	gagggaaaat	cctgtaaaaa	aagaaatgca
	24601	aatttctaaa	atcctggctc	aaagtccgtg	ggtttcctgt	ttaaaagggg	cgccatgaaa
35	24661	atgtaagcta	ttcccttttt	cctggaatct	ttaagagtcc	cagcttttca	atagtcaaaa
	24721	tgtagatgat	tgatatcatt	tcttatatga	atagcactgg	ttgtagttc	agcacgcaca
	24781	gtgagctggg	cacgcccacc	tgatagtata	gcagagaact	tgtttacatt	ctttttacat
	24841	tcatcttcta	aaacctgggg	tgctctctct	ctctctctct	ctctgtgtgt	ctctgtgtgt
	24901	gtgtgtgtgc	gtgtgtgtgc	acgtgcgcgt	gtgtgtagag	ggggagagag	agagagagag
40	24961	aactgtgaac	tgtgaaatat	aacacagcca	gcagctttgg	gtctcaatcg	tagacttact
	25021	cttaaggaaa	tttacagaat	ggaaaggtca	tgttcaagta	gtttattaac	attttgagat
	25081	gtaggaaatt	aatcccggag	tacagaagaa	caatttcaga	cttcctgaat	aaaaacagac
	25141	agcatagaga	gtggatgata	gctaaactct	gaatatcttt	tgagaagaaa	ggcactccca
	25201	tttcagggtgc	ccataaatatg	gatttgattt	tagtgattaa	aacattaatt	ttcaacttgc
45	25261	atctccctgt	gtggaagagt	tcaatttgtg	tgaggggtct	cgcctatcca	acaaaagtga
	25321	atatgtccct	tttatagggt	aattgctaac	ttgtctcaac	ttgttttcaa	acaattgtta
	25381	tagagcactc	agtttccact	aattgcaaaa	ttgttgctta	attgaaggac	tctcagccat
	25441	ctagtgcagc	cattcagcca	ctggcaggct	ctgtgatctc	aaactgtgaa	ttgcatttta
	25501	aagaggaatc	gaggagagaa	ttctgtggaa	ttctagggtt	taagtgtctg	ctgttgttca
50	25561	atggaagagg	aaatcatttg	aacaagaatc	gcatcaagtt	gtgttgtgat	aaattttctt
	25621	tattaggatg	aataacatgc	acagatgagc	ttcaaaaagt	aatgagcaaa	cttactgggt
	25681	acactctgca	tccatttact	ctgttttagta	tgagtaaatg	ttaggcaata	aatgatgctg
	25741	gcaaataaaa	tccgtatggt	atttgcagtgt	ggtattttaa	cctaggaaac	atagagtggc
	25801	tttggtattt	gtaggcttag	tcatgtgtgt	cctaaacgtc	ctcttaaaact	tctacttaag
55	25861	gcatagaatt	atttaaatcct	aaataatttt	atacttaagt	gcctcactgg	atttccagaa
	25921	tattttacact	gtaaagattt	agaaaggtca	tgaacccaat	tattgactat	atggaatcat
	25981	tattgatggc	agatgcaaaa	tggagctcac	taatgtactg	acattgaaaa	ccttttgcag

	26041	gggagaggag	ggggagtggg	aaatgtgtgt	gttctttaag	tggaacagga	aggatttctc
	26101	ttttctgtag	aaaaatttga	gtatctgggc	agataagtgt	ggaagctttc	atttaaatta
	26161	agtattttaag	ttcaagtaga	agctctaggg	cacttatcct	cttgatgaga	caaactcttat
5	26221	caaataatact	agatgctaag	aagtggctca	ttgccctgat	gtctcattta	tagattgatg
	26281	tttgaggatg	ggttgcatta	agtgaagttag	ggggctgagt	gtgggacagg	agaacgattg
	26341	gaaggaagca	aagtaaat	acaagcttta	gtgacagcca	taataaagta	aaagtttatt
	26401	tccagagagc	ctagagagta	aggaacgtta	tatagttttc	cccaaagggt	cacttgaaag
	26461	aacttttcat	tggttgtcat	ggtagtaatg	tcttgatttt	gaaatctccc	agaacctagt
10	26521	agctcttaaa	catgctttca	tcttggttcc	tttgggtctga	cggaaacttt	atgacgaccc
	26581	tctgtgtttt	tgacatgcct	ctgcattttt	ggagagagga	ggtcaggcaa	gggaggattt
	26641	cttaaaacta	agacagtata	gtaaggaaac	ataaaattat	atgataaaaa	atcactgaac
	26701	ttcaaattga	cttactgaaa	taaaacctag	aaggcaacct	gtcgtttaat	tacaactagc
	26761	ttgtataaaa	ttaaaattta	taaaatggga	attcaaagaa	aataaacggg	cagttccaag
15	26821	taattttaagc	aactcaccaa	aaattgaagt	aatagtgcc	cctagagaac	aaaatcacca
	26881	gctttactag	ccaaatggct	tatttccata	tgaaccattt	ttccaacgct	acagttacta
	26941	ggattttcctt	gttaccatat	tcagatcttg	tgagtgtgta	tgggggtggg	ggttgcattg
	27001	ggaattacag	atgaaat	aaaacaagca	gatccacaat	ttgatataatg	cactaaatcc
	27061	ttttaacggt	gtaatgtagc	caaatgtaga	atagcatgcc	aggaatcaac	gtctagcatc
20	27121	ctttttaaca	tttattttt	tcattggaat	gtaccaaacc	gaaccattga	gtataaagg
	27181	tctgatttta	tttatttgc	acaggcaatt	cattatactt	tctgagatac	aataacacca
	27241	aataatttga	gtagagagac	ctttaagaat	gttttctgatt	tatgatctac	ctttaacttt
	27301	aatgtactca	gaagatgtga	gaataaaata	aagtcaaata	taagcaagat	tttaaacaca
	27361	cacacaaaaa	acaaacaaac	aagaaaaagg	aagaaaatta	taaggattgc	cttaacctta
25	27421	gaatagatga	aggtatacat	ctgagccagc	accaaaaaaa	aaaaaaaaaa	aagttatgga
	27481	accaggaacc	aataattaca	aattgactta	aaattcttgg	atgacaaaaa	tctatattta
	27541	gttcatTTTT	gcatgcgccc	acaacagcat	ccaaaacagt	tctggggagg	cactttgata
	27601	aatggttgctg	aatgcactaa	tagattgatt	aatggctgct	tcagattatc	actagtgtg
	27661	tagacagaaa	cttcatgaaa	atggtttgtc	ttgctggaag	aaaggcagaa	attggaggaa
30	27721	aaggtttaaat	aatatttttc	cccagtacct	attataaaaag	tcatttagtt	ggcttagttc
	27781	tataatttct	tatgtgtaat	ttgattcact	tatgaaattg	tgaatatatg	aaatgttaaa
	27841	gttgatttag	acagcaacta	taagcttggtg	gatttttctt	taaatgtctt	caaattttta
	27901	aatgccagtg	gagatgccag	cgactgtgct	tcagggagta	gaatatagta	tatcttaaat
	27961	ttgtgccaat	ttctggtgaag	cagagaaaaa	attgcatgat	aaccaaaagaa	agtcattattg
35	28021	tttgtgcttt	gtgttattca	tggaagcaat	caggtgcaga	aaactttctt	tttcagaaaa
	28081	aaaaaattac	taaaataaag	gtgctgtgtg	gtgtatgcac	atataatctaa	aggagagag
	28141	ggagaaggaa	acttactaaa	taaaattttt	gccacatggg	atttagtcta	atcagtcttg
	28201	gttttgaggt	tgctatcatc	agtagttcca	ttttgtgatt	ctttctttct	gccttcatgt
	28261	gcctttgaaa	actgaaacta	tgcccaaatt	aaaacaagtt	tttctgtctt	tcctcatggt
40	28321	cacttatttc	ttgaatgtgt	ttttaaacac	agacaaactt	cttttacatc	atgtagaatc
	28381	tgaaggctga	gaaatttgca	gtcattttgc	tggagagaga	tgcttgccgg	agtcacaggc
	28441	cacattccta	ggccaaactc	tcgaagggtat	tcctcttatg	caacattggg	aaaatacatc
	28501	cagcaccgac	atgttggtg	ataatgtgtc	tgaaggcaca	gacgatatgc	ttatcatatg
	28561	aaacataaag	ccagcagata	ttgcagacat	tctgttgaat	gatagaatct	ggatcattta
45	28621	catttactta	aatgtaaaat	actatgatta	agtacaaaaa	aatcaattta	ggagagaata
	28681	gagagttgctg	ggcacggttt	taggggatga	cttatcagca	gattgtagaa	aggaagcttg
	28741	aatgttttaa	attaactgca	agttcagtat	aagccagtgg	tgtgacaaga	ggctgttatc
	28801	atagctactg	aaattttggg	ctgactgtct	agaaatataa	tactgaaatg	gagaagctaa
	28861	taattcttca	ctttttaaat	agactgtatc	tagaatatta	tcactagttc	aaggaaatga
50	28921	aataagttgt	tttaggtaca	tcactgataa	attagtgtac	attcaaatca	ctgtgaccag
	28981	gatgcatagg	gaatttgaaa	gcattgcatg	tgagcaatgg	ttgaggggac	ttggaatgca
	29041	tgacttaggg	acaagaaaac	ttaggctgga	atggcaagtg	gtttttgaat	gttgggttga
	29101	gaagaattct	aaaactgtga	aggattagta	aaaataacat	tcagattgct	aatgcctact
	29161	gtggctggga	gattagagtg	tcaacatgtg	tgatgtat	ttgacatcct	tattttgagg
55	29221	atgggcttca	aagatttgac	gaactgtcat	aagtgtat	tgtgttgctt	cagacagcag
	29281	ttctagaacc	aatgatgtaa	atttagatac	tctacatggt	agttagaaaa	ctttccatta
	29341	atttaattta	gcaaatattg	aatgtctcact	gcatacagag	cactttatta	gaggaatata
	29401	taataaagaa	aaaagaggtc	tggtgtggtg	gtcatgcct	gtaatccag	cactctggga

	29461	ggctgaggtg	ggaggatcac	ttgagcccag	gagatcatta	cgagtctgga	caacatagca
	29521	agaccccatc	tttacaaaag	acaaaaaaa	ttatccaagc	ctggtgacag	gcacttgtag
	29581	tcccagctac	ttgggaggct	gaggcagtag	gatcgtttga	gccaggagg	ttggggctgc
5	29641	agtgagccgt	gattgtccca	ctgctcttca	gcctgggtga	cagagtgaga	ccccgttga
	29701	gggaagggaag	ggaagggaag	aagaaaggta	ggaaggctga	aatgaaaccc	ttttagaaat
	29761	gacactaaaa	tgggagggtg	gagtaaggta	ttttctgaag	tgcctttgta	cttgtttttt
	29821	tctaattgcat	tggccataag	tctgctcctt	atttatagtc	cataaacaat	cctaattgaga
	29881	acagttatat	atttctgcct	ttgaatcatc	tacttgaagt	gtttagcatc	atgaattgag
10	29941	tatcagaaat	ccctcccatt	tccttgcaaa	gcgctgtatt	ttacttttcc	ttatttgat
	30001	acagattctc	aaaattggct	attttctcct	tgggttagac	agaacagaat	gtctggaana
	30061	aaaagttctt	atcaaatcca	ggtgcccana	ttgcttaaga	aattaacttt	tgaggttata
	30121	tttttttagg	gttcagtagc	taaaactaaga	aaacttctca	ccgttcacct	tcacttttgg
	30181	aaaccacaaa	atcttcagat	attacagttt	tccaaagagt	ttctcttttt	aaatataaac
15	30241	taaaagggaat	tgactcctcc	cccaactccc	tcaggcctca	gcattggatag	agttactttt
	30301	tttctttaat	aattttattta	taacttattt	tgctcttctg	tagaacagct	ggagattaag
	30361	caacatggcc	atgacataaa	atgcaagtta	gaccataaga	tgagcagccc	actccaagta
	30421	tgaatgagta	cttattcttt	gtgatctctc	atactgcttt	taggtattaa	tagtgctagt
	30481	cagcaaaagca	aacagtttaa	tattttacatc	tccttttagga	tatcatatag	ttatagtttt
20	30541	gtatgtgttc	ttgcgtgtat	gttttcttct	gtttcaaatt	cttttttctt	aaagtaagaa
	30601	tgttatatgt	agcaaatggg	tccttcatta	attcatttgt	taattcaccg	tgcatthaatt
	30661	gagtgcccag	tgagtgtcag	gcactgggct	acttggtatt	cttttccctg	tattgatcca
	30721	tatatcttca	ggtgctcctt	gatattggctc	tccattgact	ctctcatata	aggctttcat
	30781	tacctattca	catactccct	cccaaagagg	actggtccaa	aagtaaaatc	ttgagcaagt
25	30841	tctctgagtt	atttcagaac	ttcttgcccc	caaactgtat	ttttaattat	tgactggtag
	30901	catttttgaa	taacttacct	ctctttttta	aagattgaag	ttctttatcc	tctctgattt
	30961	tcagggtgtca	gtttgtgtac	aaattgagac	aataaaaaatg	tttggttagac	attttcttaa
	31021	agcattttgc	tatgtgagaa	tctttcatga	agaactcttt	ttaacaatga	ctctatagca
	31081	gaagccacag	tagagggaga	actactgaat	caaagatggg	gtttgagtct	atgattttat
30	31141	gatggatttt	tttttttttt	acatacaaag	atattgtatg	gtcttttagt	attcaggaaat
	31201	ctgttcttca	cttgacagta	tttataaaatt	gtgtgtttcc	ccctaaaaaa	acttaaatg
	31261	tgagaatgct	tccattttact	agaagttggg	taatgattat	gccaaataag	gaaaataaga
	31321	cagaaaaatc	agtttagtga	atactatttt	gcctttaaat	ttagtaattt	agtaacagta
	31381	tctctttggg	gtttactaga	aaccactttt	taatccaata	ggctcttttc	attgtgaagt
35	31441	caggaggtga	ttttgcttaa	atgtgtagta	taggaatcta	tatgtgggtg	tcaaggatca
	31501	tgtaaatatg	ctgatataat	cggagcacag	tttggcatca	ttaactcaga	aatatttaaa
	31561	ctcttgctat	acaacatgga	agcaaaactg	tgcatagttt	gtgtgtgtgt	gtgtgtgtga
	31621	atctcaaaaa	aagaaaaaaa	tcacaggatc	aggaagtcgg	aataggtccc	acttttcttc
40	31681	tagtaccaaa	cctacagcca	tgctcctagc	cttctctttt	actccaagc	aagacagaca
	31741	ggcaaatgac	catcctgtgc	cccatttctg	tgatatttca	cttgcttaga	ttttagttat
	31801	cacctgcttg	ttgagagtat	tcacaaatgg	tacctgataa	agtagatact	tctttaaaca
	31861	tgtgaatttt	tttgctattg	ataatgttta	gaaataatca	tgtataaatg	gttgaatatt
	31921	aatacaggat	tgccttatca	agtattttat	taatcattaa	aatgtgggtg	cattaataca
45	31981	atttattttta	agtgtctttc	ctaaaatacc	agattatttt	tctgattttc	acatccctga
	32041	caatgacttt	cttaaaactg	gtagccagga	acagaaaacc	taacactgca	tggtctcact
	32101	cataagtggg	agctgaacag	agagaacaca	tggactcagg	gaggggaacc	acacacactg
	32161	gggcctggag	cagggggcag	gggaaggagg	agagtgcgtc	aggacaaaca	gacaaatagc
	32221	taatgcatgc	gagccttaat	acctaggtga	tgggttgata	ggtgcagcaa	accaccatga
	32281	cacataattta	cctatgtaac	aaacctgcac	attctgcaca	tgtatcccgg	aacttaaaag
50	32341	aaaaataaaaa	ataataataa	aataaaaataa	acttagtagc	atctattgtt	ccagagcctg
	32401	taattgctct	tcaggcagtc	tcacataaaa	acctaggaga	accttactg	tcactgttcc
	32461	atgaggtgtt	aggaaaactt	gctctactgc	agtgccccag	taggcattgg	tactgagacc
	32521	aaaattcagc	tgggttggtg	ttactacgat	tcctacgtga	tttacttgt	catgtagaca
	32581	agattgcaca	cttcaataat	aatcttgctc	aaatgtgtgg	tattccatac	atttttaaaa
55	32641	tgcattcaca	tatctcattc	catttgatcc	tacaaaataac	tctataaaaa	agattggcag
	32701	acattattttc	tatataacag	aggaggaaac	tggagcttag	agaagctaaa	tagcaatcca
	32761	aatgacacag	ctgtagaacc	agagcaagga	tgatagccca	gtgacttcac	ctaacctagt
	32821	ccccttacca	ccactccagc	tgtctataac	caaaacctgc	agtattcaag	taagaaacca

	32881	tatcttgccc	ttgatgcatt	aatgtgagac	ctggagcagg	aacaggctga	tattgtcacc
	32941	ctggcctact	gtccaccttt	gtctccagca	gagactggta	cccttctgtg	tgccaaggaa
	33001	taaagtggta	atgggaagat	taaaaatgtt	ttttccaagg	agttttttaa	tttaattttt
5	33061	ttaaaaaaga	aaaaactctt	agagggaaaa	atgaatatat	gacttttgat	gtattgttcc
	33121	ttagtaactt	agttataatt	ttacttaaac	ctgagactct	tgctaagtga	atgattagaa
	33181	atattaggtg	gctggccaga	tggcaaatag	gaacagctgc	agtctgcagc	tcccagagag
	33241	atcaatgcag	aaggtgggtg	atttctgcat	ttccaactga	ggtaccgggc	tcatctcatt
	33301	gggactggtt	agacagtggg	tgcagcccat	ggagggtag	ccaaagcagg	gtggagcatt
10	33361	gcctcactca	ggaagcgcaa	ggggtcaggg	gaactccctc	ccctagccaa	gggaagccct
	33421	gagggactgt	gccatgaggg	acagtgcctat	ctggcccaga	tactacacat	ttcctacagt
	33481	ctttgcagcc	ggcagaccag	gagattccct	tgggtgccta	caccaccagg	gccctgggtt
	33541	tcaagtacaa	aactgggtgg	ccattttgggc	agacacccag	ttagctgcag	gagttttttc
	33601	tcatacccca	gtggcacctg	aaattccagt	gagacagaac	cattcactcc	cccggaaaagg
15	33661	ggctgaaggc	caggcagcca	agtgtatctag	ctcagcagat	cccaccccca	tggagcacgg
	33721	caagtaaga	tctgctggtt	tgaattctc	actgccagca	cagctgcctg	aagtcaacct
	33781	gggatgctcc	agcttggctg	ggggaggggc	atccgccatt	actgaggctt	gagtaggctg
	33841	ttttcctctc	acaatgtaaa	caaagccact	gggaagtttg	aactgggtgg	agcctaccac
	33901	agctcagcaa	agcccctgta	gccagattgc	ctctctagat	tctccctctc	tgggcagggc
20	33961	atctgggaaa	gaaaggcagc	agccccagtc	aggggcttat	agataaaact	cccatctcat
	34021	gggacagagc	acctgggaga	gggggtggct	gtgggcccag	cttcagcaga	cttaaatgtt
	34081	ctttgcctgt	tggctgtgaa	gagagcagtg	gatctcccag	cacagcactt	gagctctgct
	34141	aagggacaga	ctgccttctt	aagcaggtcc	ctgaccctcg	tgattcctga	gtgggagaca
	34201	cctcccagca	ggggtcgaca	gacacttcat	acaggagagc	tctggctggc	atctgggtggg
25	34261	tgcccctctg	ggacaaacct	tccagaggaa	ggaacaggca	gcagtctttg	ctgttctgca
	34321	gcttctgctg	gtgataccca	ggcaaacagg	gtctggagtg	gacctccacc	aaattccagc
	34381	agacctgcag	cagagggggc	tgactgttag	aaggaaaact	aacaaacagg	aatagcatca
	34441	acatcaacaa	aaaggatgtc	cacacgaaaa	ccccgtacaa	aggtcgccaa	catcaaagat
	34501	caaacataga	taaatccaca	aggatgagga	aaatccagca	caaaaaggct	gaaaattcca
30	34561	aaaaccagaa	tgctcttctt	cctccaaggg	agcacaactc	tttgccagca	agggaaacaaa
	34621	actggtaggga	gaatgagttt	gatgaattga	cagaagtatg	cttcagaaag	tgggtaaaaa
	34681	cagactcctc	caagctaaag	gagcatgctc	taacccaatg	caaagaagct	aagaaccttg
	34741	aaaaaagggtt	ggaggaattg	ctaactagaa	taactagttt	agagaagaac	ataaatgacc
	34801	taatggagct	gaaaaacaca	gcacgagaac	tctgtgaagc	atacacaagc	ttcaataact
35	34861	gaatcgataa	agcagaggaa	aggatatcag	agattgaaga	tcaatttaat	gaaataaagc
	34921	atgaagacaa	gattagagaa	aaaaagaatg	aaaagggaagg	aacaaagcct	ccaagaaatg
	34981	tgggactatg	tgaaaagacc	aaacctacat	ttcattgggtg	tacctgaaag	tggcgggggac
	35041	aatggaacca	agttggaaaa	cactccttag	gatattatcc	aggagaactt	ccccaaacta
	35101	gcaagacaag	ccaacattca	aattcaggaa	atacagagaa	caccacaaa	ataccctca
40	35161	agaagagcaa	acccaagaca	tgtaattgtc	agattcacca	aggttgaaat	gcaggaaaaa
	35221	aagttaaggg	cagcgagaga	gaaaggtcgg	gttaccctaaa	aagggaagcc	catcagacta
	35281	acagtggatc	tctcagcaga	aaccctacaa	gcctacaagc	cagaagagag	tggggggccaa
	35341	tattcaacat	tcttaaagaa	aagaattttc	aaccacagaat	ttcatatcca	gccaactaag
	35401	cttcagaagt	gaagtagaaa	taaaatcctt	tacagacgag	caaagtctga	gagattttgt
45	35461	caccaccagg	catgccttac	aagagctcct	gaagggaagta	ctaaataagg	aaaggaaaaa
	35521	ccggtaccag	ccactgcaga	aacataccaa	attgtaaaga	ccattgaaac	tatgaagaaa
	35581	ctgcatcaac	taatgggcaa	aataaccagc	taacatcata	atgacaggat	caaattcaca
	35641	cataacaata	ttaaccttaa	atataaatgg	gctaaatgcc	ccaattaaaa	gaccacagac
	35701	tggcaaattg	gataaagagt	caagacccat	cagtgtgctg	tgttctggag	acccatctca
50	35761	catgcaaaga	cacacatagg	ctgaaaaata	agggatggag	gaagatctac	caagcaaatg
	35821	gaaagcaaaa	aaaaagcagg	ggttgcaatc	ctagtctctg	ataaaacaga	ctttaaacca
	35881	acaaagatca	aaagagacaa	agaaggccat	tacataatga	taaagggatc	aattcaacaa
	35941	gaagagctaa	ctatcctaaa	catatatgca	cccaatacag	gagcaccag	attcataaag
	36001	caagttctta	gagaccacaa	aagagaccaa	gactcccaca	caataatagt	gtgagacttt
55	36061	aacaccccaa	tgtcaatatt	aggtcaacga	gacagaaaat	taacaagcat	attcaggatt
	36121	tgaactcagc	tctggaccca	gtggaactaa	tagacatcta	cagaactctc	caccccatat
	36181	caacagaata	tacattcttc	tcagcaccac	atcacactta	ttctaaaatt	gaccacataa
	36241	ttggaagtaa	aacactcctc	agcaaatgca	aaagaatgga	aatcataaca	aacagtctct

	36301	cagaccaaaag	tgcaattaaa	ttagaactca	ggattaagaa	actaactcaa	aaccatacaa
	36361	ctacagtggga	aactgaacaa	cctgctcctg	aatgactact	gagtaaataa	caaaaagaag
	36421	gcagaaataa	atacattatt	tgagaccaat	gagaataaaag	atacaacata	ccagaatctc
5	36481	tgggacacag	ctaaaacagt	gtttagggga	aattcatagc	aataaatgcc	cacaggagaa
	36541	agcaggaaaag	agctaaaatc	aacactctaa	catcacaatt	aaaggaaacta	gagaagcaag
	36601	agcaaacaca	ttcaaaagct	agcagaagac	aagaaataac	taagatcaga	gcagaactga
	36661	aggagattag	agacacaaaa	aacccttcaa	aaaaatcagt	gaatccagaa	gctggttttt
	36721	tgaaaagatt	aacaaaatag	atagaatgct	agccagattg	ataaagaaga	aaagagagaa
	36781	gaatcaaata	gacgcaataa	aagatgataa	agaggatatac	accactgatac	ccacaaaaat
10	36841	acaatctacc	atcagagaac	actataaaca	cctctatgca	aataaactag	aaaatctaga
	36901	agaaatgaat	aaattcctgg	acacatacac	cctcctaaga	ctaaaggaag	aagtcaaatt
	36961	cctgaataga	ccaataataa	gttctgaaat	cgaggcagta	attaacagcc	taccaaccaa
	37021	aaaaagccca	ggaccagacg	gattcacagc	tgaattctac	cagaagtaca	aagaagagct
	37081	ggtaccattc	cttctgaaac	tattccaatc	aatagaaaag	gagggaatcc	tccctaactc
15	37141	attttatgag	tccggcatca	tcctgatata	aaaacctggc	agagacacag	caaaaaata
	37201	aaattgtagg	ccaatatccc	tgatgaacat	tgatgcaaaa	atcttcaata	aaaaactggg
	37261	aaactgaatc	cagcagcaca	tcaaaaagct	tatctaccat	gataatttgg	cttcatccct
	37321	gggtgcaag	gctgggtcaa	catatgcaaa	tcaataaaga	taatccatca	cataaagaga
	37381	accaatgaca	aaaaccacat	gattatttca	atagatgcag	aaaaggcctt	tcataaaatt
20	37441	caacagccct	tcattgctaaa	aactctcaat	aaactaggta	ttgatggaac	atatctcaaa
	37501	ataataagag	ctattttatga	gaaaccacaca	gccaatatca	tactgaatgg	gcaaaagctg
	37561	gaagcattca	tttgaaaacc	ggcacaaaac	aaggatgcc	tctgtcacca	ctcctattca
	37621	acatagtatt	ggacgttcta	gccagggcaa	tcaggcaata	gaaagaaata	aagcatattc
	37681	aaataggaag	agaggaagtc	aaattgtctc	tgtttgcaga	tgacatgatt	gtatatttag
25	37741	aaaaccccat	catctcagcc	caaaatctcc	ttaagctgat	aagcaacttc	agcaaagtct
	37801	caggatacaa	aatcaatgtg	caaaaatcac	gagcattcct	atacaccaat	aatgacaaac
	37861	agccaagtca	tgagtgaact	cccattcaca	attgctacaa	agagaataaa	atgcctagga
	37921	atacaactta	caagggatgt	gaaggacctc	tttaaagaga	actacaaaac	actgctcaat
	37981	gaaataagag	aggacacaaa	caaatgggaag	aacattccat	tctcatggat	aggaagaatc
30	38041	aatatcgtga	aaatggccat	actgcccaca	gtaattttata	gatccaatgc	tatccccatc
	38101	aagctaccat	tgacttttctt	catagaatta	gaaaaaacta	cttttaaattt	catatggaac
	38161	caaaaaacag	cccgtatagc	caagacaatc	ctaagcaaaa	tgaacaagct	ggaggcatca
	38221	tgctacctga	cttcaaaacta	tactacaagg	ctacagtaac	caaaacatca	tggtactggg
	38281	acataaacag	atagatagac	caatggaaca	gaacagaggc	ctcagaaata	acgccacaca
35	38341	tctacaacca	tctgatcttt	gacaaacatg	acaaaaacaa	gcaatgcaga	aaggattccc
	38401	tatttaataa	atggtgtcgg	gaaaactggc	tagccatttg	cagaaaactg	aaactggacc
	38461	ccttccttac	acgttatata	aaaattaact	caagatggat	taaagactta	aacataaaaac
	38521	ataaaaccat	aaaaacccta	gaagaaaacc	taggcaatac	cattcaggac	ataggcatgg
	38581	caaagacttc	atgactaaaa	tacaaaaagc	aatggcaaca	aaagccaaaa	ttgacaaatg
40	38641	ggatctaatt	aaactaaaga	gcttctgcac	agcaaaaagaa	actaacatca	gagtgaacag
	38701	gcaaccgaca	gaatgggtga	aatttttttgc	aacgtatcca	tctgacaaaa	ggctaatatc
	38761	cagaatctac	aaggaaccta	aacaagttta	caagaaaaaa	aacaacccca	tcaaaaagtg
	38821	ggcgaagggt	atgaacagat	gcttctcaaa	agaagaaatt	tatgctgcca	acaaacatac
	38881	gaagaaaagc	tcatcatcac	tggtcattag	agaaatgcaa	atcaaaaacca	cagtgcagata
45	38941	ccatcttatg	ccagttagaa	tggcgatcat	taaaaagtca	ggaaacaaca	gatgcaggag
	39001	aggatgtaga	gaaataggaa	cactttttaca	ctgttggtgg	gagtgtaaat	tagttcaacc
	39061	attgtggaag	acagtgtggg	gattcctcaa	ggatctagaa	ccagaaatat	cttttgacct
	39121	agccatccca	ttactgggta	tatactcaaa	ggattataaa	tcatgctact	ataaagacac
	39181	atgcacatgt	atgttttattg	tggcactatt	cacaatagca	aagacttggg	accaatccga
50	39241	atgcccatca	atgatagact	ggataaagaa	aatgtgacac	acatacacca	tggaatacta
	39301	tgagcccata	aaaaaggatg	agttcatgtc	ctttgcaggg	acatggatga	agctggaaac
	39361	catcattctt	ggcaaggtaa	cacaggaaca	gaaaacccaa	caccacatgt	tctcactcat
	39421	aagtgggagt	tgaacagtga	gaacacatgg	acactgggag	gagaacatca	cacactgggg
	39481	cctgtcaggg	tgtaggaggc	taggggaggg	atagcattag	gagaaatacc	taatgtagat
55	39541	gacaagttga	tgagtgcagc	aaaccacat	ggcatctgta	tacctaggta	acaaacctgc
	39601	acgttctgca	catgtacccc	agaacttaaa	agtattatta	ttattattat	aataataata
	39661	ataaaagaaa	tacaataaaa	tagaatgcag	catacagcag	tgatttctca	acacattcag

	39721	catcagaatt	acccttgaat	ctttaaata	tatatacata	tgagatctta	gtctccaaga
	39781	tttgttaagtt	tggtattggg	tccctgggcc	tatgttgggt	ttagaaactt	ctacagatgg
	39841	tttggatgta	tgggacagtt	taagaatcgc	tgaactaaaa	tcaaataaac	tgaatatcct
5	39901	gtgattttaga	gagacttatc	gtttatttca	ctatccaagt	acttgcatta	gagcgtggct
	39961	agaagggatt	tgcagccttg	taaataatca	gaaattcaga	cattttgaga	tgagagaact
	40021	gctgaagatt	ttattctgac	ttgaaataaa	ttttctaatt	agaaacttcc	aggtgagagc
	40081	aaaggcctgg	aacaatattc	ctgagccaga	ggaggatcga	gtttgactcc	aggcctaaca
	40141	cttactaggt	ctatgacctt	gggtcagtaa	tttaaattct	ctgtatctca	acctctcaac
10	40201	agggtattgg	tagggattaa	atgtgttagt	gtctgtgaag	tgcttagagc	agtgttggc
	40261	atagtaaattg	cttaatgaat	ttcagccact	gtttttat	ttagtacttt	ccagctcccc
	40321	caaaaagata	ctttttttag	acttgtatta	agacaataaa	aagttaatc	agcatgcttc
	40381	atacctaaat	atgcttcact	ttatagcaaa	gtttacaaga	ctaaaactgt	ttgtgtgtaa
	40441	ttctctgagt	ctcatgtgtt	tattaatgat	tttttctgct	gtttattcat	ctgaattcta
15	40501	ctcattcttc	aagacctagc	tggaaatcctg	tttctagaaa	gactcttgcc	cataataata
	40561	aacctgccct	atctgagttc	ctagggtggtc	tgtacctcat	aatttggtaa	ttaattgtat
	40621	atgcacttat	ataacaaaac	attattgtgt	gtctttgctg	tatcagattc	taggctggaa
	40681	gtttagata	tgatgttttt	gtctagaaaa	atgttctaga	atgcctact	caggacagtc
	40741	tggtgacttt	aaagacacat	ttcctaagaa	gacacttcat	gaggcagccc	gagcctgtac
20	40801	ctgtgttcct	ggacctgatg	atcaagtttg	atttaagcct	caccacttac	tagctctgtg
	40861	attttgggca	agttacttga	attctctgtg	tgtagataga	acaatgttga	gggaaatccc
	40921	tttcccccat	ccttgtgttt	ccacaaggga	acttgcttcc	taataagtaa	cactttcagg
	40981	ggaatattct	aggcccttct	cttatcccca	ttacttggtc	tttctgtgaa	aagaggagag
	41041	gttaatctga	tggatgaaat	ccttaatctt	tcatcttctg	gactgtagag	cctgtgaacc
25	41101	aaagcaatgg	accacttgca	ctgaaattga	ggctgaccct	gtattttgat	tcttatttgg
	41161	caacttattt	ctattctgtt	cccaattcaa	aatcccaagg	ggagaaggaa	gataattgat
	41221	taccagaagt	atgtaattgt	ggttaggaagt	tgaataaatg	gtaacttttt	aaaagttgca
	41281	tgagatatag	tccttatccc	agagaagcta	agtttgcttt	tctttcctct	catgtatttt
	41341	agtattattt	ctacaattag	attgtaaacc	ctttaaagc	aagaatattt	ctacattttc
30	41401	ttactcctga	tagcacacag	tagactgctg	ggcacatata	tagtaggttg	ctctgtaggt
	41461	acttgctaaa	tgattcaaca	tgtttttccc	tcatggaaaa	gaaagatttc	agtattgttc
	41521	ttatcagcta	ggaaggcact	ctgaatagga	aatcagttct	aggcaggtat	ccataaatgg
	41581	gttatgattt	ccaacttact	tgccccagag	gctcgcta	gttgaactct	tcatgggtac
	41641	tttgtcttgc	ttcatgagct	atacatgcta	aggggttagc	agatcatata	atcttttgat
35	41701	ctacaaaata	tgatctttat	tgaacaaaaa	cttgggccaa	aggcctttct	cctttgccac
	41761	cttcctccct	cttttcattc	tcttttttgg	gaatgccctt	tgtgcatgtt	agttacagca
	41821	tgtaccacat	tgactgttat	tgttggtttt	tgggtctaac	ccacccttaa	cactgcagtc
	41881	cccaagggca	gaaattcagt	ctcattcatt	ttgatgtcct	cagtgcctgt	gctcagagaa
	41941	tatctattat	ttgaaaaaat	agtgcaaaag	taaaatttag	gagactacat	cacactcatc
40	42001	taaaactgcaa	gtttgacaag	ttgacatcca	aaagaaaggc	tctccta	aaactcgcca
	42061	cagaaatttg	ggtgaccttt	gtagctctgg	agaaagcaga	ggcaaaaatg	aaaccta
	42121	attatttgtg	ggttttttaa	aaatgttttc	tcatggagta	aaggtctaca	gctgagttct
	42181	tttcatatga	gggaatgaca	gaaacacagc	tggttctgac	tttcagcttc	aactgagcga
	42241	ccagagctct	gctggtgaaa	caggaacttg	tattgtgccc	ctgacgtgca	ccttgaagg
45	42301	gtcagctcat	tgtccctttg	ttcacataaa	tagtttttta	agaattgttt	ttgatcttgt
	42361	gagcctctaa	ctaaatgatt	aacctgcaa	agttggccat	ttggggta	actgaagcac
	42421	ttctcttgag	ggctattgac	aggtgggaat	gtgccacct	ccttgggtct	ctggttttca
	42481	tgtcatactt	gcaaatcagt	gacagtttaa	acttggggca	atcacttagc	aagtctattg
	42541	agttaccaag	ttaattattc	ccactttgca	tgaagcaacc	ttgaaaatga	ttttcctaaa
50	42601	gcaaagtaca	tccaaactca	gtaccttctt	aataaccttt	gctgaatgaa	taaatgacta
	42661	attcataaaa	aatgtaacat	atctttta	cttacttacg	ggcagtttaa	gcctcttg
	42721	taagaggagg	cctcggcttg	agataacata	ggatagtaag	cctcctagag	aaatttctat
	42781	atggaaacat	ggtctgctat	gaagctagaa	gtgagaggac	attatatttg	accattatat
	42841	ttggcttcag	agcttctcaa	catggggccc	aaagtcaagg	tcccttggtt	cattaagagg
55	42901	aggtccagga	gtgcatgaca	cccatcagac	tactgagacc	cagctggaac	taggcacctt
	42961	gcacaggggc	cttgccta	caaaatagtt	cttatttttt	ctgagttcca	agtaactagt
	43021	ttcctaacc	agtgtctgga	tagtagtgcc	aagtgggagt	accttcaatg	aacttctca
	43081	tgaggttatt	tctagcctat	tggaaatg	cgtttttagga	gggtgaggaa	gggaagtctt

	43141	gaattttttgt	gcttagttta	atgttgtgat	acagctttga	ccatccgttt	aatgggagat
	43201	ctgtttttcca	gatgactata	catgtggaaa	ggagaagttt	tttgagtgtt	ttttttaacc
	43261	cctttttaaaag	aatgggttttt	catttagtct	ctacatttgg	gggtaaaagg	tcctctaggg
5	43321	agactttttca	aaagtatttg	aagtttgc	ctgatttcag	aggtgagttg	gaggcctatc
	43381	tgtgtatgac	agacacatgt	ctccaacaac	tatatgttca	caaggactaa	gagccatcct
	43441	tttgggtcca	tcattcaaca	ttgatctcac	attcgtgttc	gtatcagtat	ctttacagt
	43501	cgctcccagt	tacatctccc	taatttccct	tagtaggctt	cacagaattt	gcagtgtatg
	43561	caatggcaga	tgaccacatg	tggagtcatt	taaccacatc	ttccactgca	agtcagccccg
10	43621	ctcttgatgt	ctgttttatgt	ttagattcca	tcttttggaa	gatttcattc	ctctgacta
	43681	tctcagtatc	tcagatgctt	ttgagactgg	gtccttttcc	cctcctatgt	ttggccatgg
	43741	ccacccccctc	aggggttgtgt	tgtgtttcac	agctgctgtt	tgtagggttg	acctttacaa
	43801	tgtacaaagc	tctttcccat	atgttgacaa	tccctgggtg	gatgctgtga	gttaggcagg
	43861	gtgtgtatac	gtgtcctcat	catattacag	tggtaaggca	acagggtttt	tgaatttgat
15	43921	cacccatgaa	tttgtcta	ttgttggtaa	aaaatggtca	tgtatcagcc	gtttcacagg
	43981	gtcagcttaa	tagaaagtgg	gagttaggca	ggaccagaat	tcaggacttc	agccccgggt
	44041	cccagggact	attctctata	cccaattgtc	ccaccttgaa	tcagtttctt	ctagggaat
	44101	atctccaaaa	ctgagatggc	acccacagga	cttcttaatt	gtagtcatta	ccaggaaaaa
	44161	caagcaaaag	aactgggtga	aatctctgtt	tttgggtgatt	gggtggagatt	tggagattgt
20	44221	cttgtgtcaa	aagtaaagcc	actagattaa	atgttttgtt	aataaattgg	ttatttttaa
	44281	tttaattatt	tgacagttaa	tttacattat	tcaaaaatca	aaataaaatt	taaaagaagt
	44341	ttacactgaa	aagtcttgcc	ccacttatac	cctgctcacc	tcagtatccc	ccaatacata
	44401	ccatctataa	ggtgatcatt	tgtattagtt	tcttgtgaat	ccttgatagt	gtgttttata
	44461	tagatacagg	taaatatgag	tatgtactat	tatttcccc	ccacccacc	ctgttttttt
25	44521	tttgagacgg	agtctcgctc	tgtcgcccag	gctggagtgc	agtggcacga	tctcggtca
	44581	ccgcaagctc	caccattttt	ccccattttt	aaaacaaaag	gtagtagcca	tatatacact
	44641	attttacacc	ttgttttatc	acttactaat	atataccaga	gagctttcca	tcattttgta
	44701	catatgcacc	tatatctgtc	aattattccc	agaagtggaa	ttgctgggtc	agcaggaaaa
	44761	atcatgtata	attttgatag	gtattgccta	attgtcctgc	acagggttg	aattgtttgt
30	44821	actcccacct	ttagtgtatg	agaagacctg	tttctccata	gcctcatcaa	acagagtgtg
	44881	tgagattaga	tgagaaatag	gaggtgagca	gtcttttacc	ccatccgtag	tttgagtggt
	44941	gaacactgca	cagttgcaag	agctggtgca	ggtatcagat	tagttccagt	ggaaacgctg
	45001	cctcaccatg	gccatgggct	tgcgccagct	ctagtacac	acacggaatg	gaccacggt
	45061	gccacttgca	gaatttcctg	tagcagaaaag	ttgaacatgc	attcattatt	catctaacta
35	45121	gccatgctgg	atctaaagag	cacaacagtg	ttttttagaa	ccaaaaagaa	aattgtttca
	45181	ctacaacaca	ctgtgtataa	ggctttcaat	gctcttttct	cagctattaa	cattattttc
	45241	aggactgagt	tcaagagatg	tatcccaaat	cacagggatg	tcttgctaag	cttggaactt
	45301	tcatactcaa	gggatgcttt	tttgaggaat	gattttacac	ttactcaaca	tttgtaatta
	45361	aataattagt	actttataag	ataaatttaa	actgtccaag	tacaatataa	acattgaact
40	45421	atgatgcatt	attgctagac	tttttcctta	aagttgccaa	gtggtttcct	gcattaggca
	45481	aataggggat	catataaaaa	tgccatgatt	tacggcctag	ataacatctc	caccatttga
	45541	gcagcatata	ttccagggtca	tccccacata	actccttacc	attctcatta	gaaaggttga
	45601	ttcttagtct	tattttttctc	tgaggacagc	aaaaaaaaaa	atccccctca	gttccactgc
	45661	atagaaaagt	gtggtaaatg	gagccgggca	cagtggttat	ttaatttaaa	tggacaatat
45	45721	tttttataga	attttgacag	ggccactgta	taggggaaag	tcactcctct	tcccctttat
	45781	agaagagttg	cacctggaca	gttgcatatg	tgactgtatc	cagtctacac	aagaggtcat
	45841	tcctgggcat	aagaatggac	tgccaaaatc	tagctgaaac	accattgaca	aatagacatt
	45901	ttcttttggt	aataatacct	gtgaaggctt	tcataacaga	catttccagt	tttgttctca
	45961	ggctccttgc	agctgctcct	ctaaaagtgt	gctctcttcc	aagagctgac	aatggccaga
50	46021	agcaaggtgt	tctgtctttt	gtgccatcat	catctaactt	gccacacaca	tttgggatgt
	46081	cagcctaggt	ataggttttt	tatccactca	gtatggcttg	tgggtctggg	tgcttttgtt
	46141	attcatgctg	agggcctctg	ggcatcagtt	tgggtgtgaga	gaacccattc	catgaccctc
	46201	cttccttttg	ctgttttgac	tcgatggctc	ttgttggcac	agtctgtgag	tgtctgatgc
	46261	tctatccatg	ccggaccatc	tgttctgctg	tctctgtggt	ctgaagtcgt	tttctgaact
55	46321	attccttgat	aataaatttg	agatgatctt	gttctacctt	tcttttcaag	tcacatctta
	46381	gcccccttagc	cacattcccc	aagaacatga	caaatggatg	ggtcacaagt	cacgtagcat
	46441	aggggtgtcag	accacgaggc	tttgaaggga	ttctgttggg	tgctaaaaag	aaagattttg
	46501	tgtcaccacg	atttttttta	aaggcatggt	gacacttagg	ccttaattga	aagcgttctt

	46561	actcaagtag	agttgacaga	ggagtatttg	gtagtcgcgg	ttgctggctct	gaagagcatg
	46621	tggttctggt	tcaatgccca	atgagatctt	ctcacgggaa	aatgttctga	catctcaaac
	46681	aaatgacctt	catgcatagt	tttgacaaaa	taccctatta	agtatgcata	tatggttggt
5	46741	accttggtgt	aataattcaa	tactggaaac	agagtagcaa	caaagaaaca	ttagggttat
	46801	atttaacctc	tgtggaatta	gtgtgtaaac	aaactgctta	tcagaaatgc	tcatatgggg
	46861	ctttgtttta	ataaataaga	aactggcata	tagggctctgc	aggatatttc	tgccaagtag
	46921	acctccctca	cattataaga	caccacatct	atgtctgacc	ccatatggaa	agaggcatag
	46981	caagccagca	ctggttcata	ttccctctcc	accacataat	gggtatgtga	tcttagggaa
10	47041	tccaccgaaa	ctctctgggc	ctcagtttcc	tcagctataa	atgggtggata	atcaaattat
	47101	ttacctcacc	attaataaat	gtagctattt	atTTTTtatt	aagtttaata	caaagagaaa
	47161	cattttactt	atTTTTccag	ctatccagag	catcttccaa	aatcctatca	ccaacaaata
	47221	ctgtattgta	tttattatag	caactatgta	aaaaatggag	ccctgtccta	tgcttagatg
	47281	aaatatgttg	gtatttgagt	ttgcatgtct	tctataggaa	tcagtgttta	gtgaaaacgg
	47341	gtggagataa	acagatgttt	tcacagtcct	gttggtcaca	gtaccgccaa	attgaatgtt
15	47401	tccatatagg	tgcatctcaa	tggcttaaat	gatgcagata	ttttctggcc	agccatatgg
	47461	atcttttgtc	atctaagatg	ttaatatatt	ccttatattt	tatagtagtt	ctggagtaca
	47521	gccagtttct	tgaatagggt	ccacatggct	cattatgcac	agggcctgga	aactgcctta
	47581	ctcgtgctgt	tgaatgaac	cgtgacactt	cagaagagct	gggagctggg	gtagagcagt
	47641	ggctaggaga	acataattcaa	ttatatctcc	tcctgcatta	agctacaagt	aatgagcact
20	47701	ttcctgtgct	ttacagttaa	gtaattaaaa	gaaattatag	agtgggatgc	aaaaataacc
	47761	cgaaggacaa	ctggatgtgt	ggagccacca	gttttctcca	tgagtgcaca	aggttaatcc
	47821	ttgttactac	tcagaatgct	gagtttctac	agaaaggggt	gcaggtccac	acatgttttg
	47881	gcgtctaccc	acacgcttct	gtatggcatg	actgtgcatc	ccagaagaag	ggctgtgctg
	47941	tgtacctcca	cgtttcagtg	gaatttaaca	aactgatccc	tgaaaatggg	ttcataaagg
25	48001	tgagtaacag	agagctaata	gccttctctt	gctaatttta	tctttccccc	aagatttctt
	48061	gataatagtt	tgaagaggag	tggtattctt	tggtctctag	aggcaactta	cctttccagt
	48121	ttcttccatc	acctgttttc	atctctcttg	tttttttaaa	tttaatgctg	tatgtatttc
	48181	agaggatagg	atctaactta	gtgcggtccc	ttcatcaggt	gagaattatt	catctcattt
	48241	tcatttttagc	ccttctgaat	taatgacatt	gaagcccggc	agtttgggtc	taagatgggt
30	48301	ttaattatgt	acagatactc	tttctataat	ggaaattgct	cagataacta	attaaccaca
	48361	agaatacact	gtctatggaa	aatttcagga	gcaccgtctg	tggaataact	gggaagggca
	48421	tgctgtcacc	acagctctgg	ggctatttaa	aagtgtgggt	atgcagcact	ggtgtctagt
	48481	gggtgtgttg	ctctcaactg	ccagaattcc	catagcattt	catggcagaa	agtcaagggt
	48541	tccagcaata	ctctgaaagt	gacctgttga	ttaaagtcgt	caattctgaa	gaaagagact
35	48601	gaaataagac	aaatgggtct	taactttttt	tctctttctc	tctcttgtaa	aaatgtgtga
	48661	ttgttctggc	atgttcccaa	tccccacata	atgccaacat	cttttcttaa	agggggattc
	48721	cctttatcct	tggtctgag	aattattgca	tggtctccct	ttagggacaa	tgaatgcagt
	48781	tgctacacc	ttgttttttt	tttttttttt	gtacacagca	tgcttattct	tggtatgcagg
	48841	gacttgaaag	acaaagcccc	acctggcttt	cacaacatct	cctattagta	ggtgtgcctt
40	48901	gtgtgtaatt	tgaaggaggc	ggtcccttag	ctgtgtttac	actgtacttt	taaatgtggg
	48961	gctgaaggta	gaatcaacca	tacttaagat	gccacctggg	aaaatagggt	tctgtgtcat
	49021	ctcagcccca	cccatttgca	aatgacttaa	cagcagcact	attagggttc	ctagtgtgag
	49081	tcatttgcat	ttggactggg	gaacttggtg	acttcttggt	gtttggaaac	aaacaacctt
	49141	tgcagtcttt	cgtaaaaaagc	ctgaacagtg	gaccagtctc	cagttctact	tgcaaagctg
45	49201	cccccatcaa	atccctcata	atgttcaact	taaaaaatgt	tacacttttc	tctggaaatc
	49261	taaccttttt	tccttttttt	aaagccattt	taagtacttc	agtcttgaat	caaatgatcc
	49321	caaataattgg	acaccaacct	agaaattggg	ttacctcctg	ggaactttat	cgaagaagag
	49381	agatttttgt	tgagaggggg	gttttgatgt	ttgatactta	tatttactat	tttaatatatt
	49441	cattgttgtt	gttgctgctg	ctgctgtatt	atTTTtgcgag	tttctgttgt	ttaaattttca
50	49501	tggtatttgg	taggagagag	ctggatctgt	tggtttcagg	acaagtctag	aaataagaaa
	49561	tctgccttga	gtgagtgaag	tggttccctc	tggtgtctatt	tcaccattaa	ggacgaaagg
	49621	aactcacaag	gaccagagac	atctggctga	aagcaatact	agtgtgactg	gacatctact
	49681	acctgccata	gttggtcata	tcgtttccag	tatgattctg	attgagttag	tgatattagg
	49741	ctatgttcag	ggatcagggg	ggctaattat	gcttatattg	ccttgtagca	ttttggtaag
55	49801	aattaatgat	tgtgtagatg	tccagattta	ggtcagcaat	attctaaaag	ttctcattga
	49861	actaatcatg	tttataagta	gcctgtactt	tctatcataa	taacaatagt	ggaaaagcta
	49921	gttgacataa	aaggagccca	gatttttactt	aagtaaaaac	acaaaagcaa	agatattttc

	49981	ccacataaat	tacaaaagca	aagatatattt	cccacataaa	tgtccccata	aaacaagttg
	50041	aaccaaaagag	gaaagatgac	aggtaaccgt	atgacacgct	aagaaagtat	cataataactt
	50101	aagttaactt	caacctttta	tttccttatc	ctaagcagcc	tcttttctct	ttatcattta
5	50161	gtcctgtgct	tctcaacttt	gataagtaaa	aaagtatttg	cactaaataa	atcttattga
	50221	aatgcaggat	ctgattgagt	gggtggggta	ggtggaatga	gggtggggaa	gttgagattc
	50281	tgcatttctt	agaagtttct	actttatgtt	aaaatggcta	atccatctca	acattgagaa
	50341	gtaaggtttc	acttaatttc	agcctgtgta	agtttatccc	atatgtacat	ttcctaaaac
	50401	tctaattctca	ggccccagga	atttctcctt	tagttaaaat	atttttagga	ataaatttga
	50461	attgcattaa	tacacaattt	ataaatttaa	cacaaaaaat	tatttgaagt	ttgagacttt
10	50521	aggttgcatg	aaatcaattt	catacttgaa	aattttctat	aaattcaaaa	gtctgtgat
	50581	ttaaatacaa	tttaaatacc	tgtgttacag	tgacatttgt	ttttctgtct	ctctctccac
	50641	catttccaga	gtcatcatcc	ctgtacagaa	aaatttttcc	cacatgattt	caccataaat
	50701	tcattaaata	tgatgcttac	ttgataattt	ctccagggtc	tttttttttt	taattatact
	50761	ttaaagtctta	gggtacatgt	gcacaacctg	caggtttgtt	acatatgtat	acatgtgccca
15	50821	tgttgggtgtg	ctgcacccat	taactcgtca	tttacattag	gtatatctcc	taatgctatc
	50881	cctccccctt	acccttactc	catgacaggt	cccagtgtgt	gatgttcccc	accctgtgtc
	50941	caagtgttct	cattgttcag	ttcccaccta	tgagtggagaa	catgcggtgt	ttgggtttct
	51001	gtccttgcca	tagtttgctc	agaatgatgt	ccttgctcac	tgatggacat	ttggttggct
	51061	ccaagtattt	gctattgtaa	atagtgccgc	aataaacata	cgtgtgcatg	tgtctttata
20	51121	gtagcatgat	ttataatcct	ctgggtatat	accagtaat	gggatggctg	gctcaaattg
	51181	tatttctagt	tctagatcct	agaggaatcg	ccacactgtc	ttccacaatg	tttgaactag
	51241	tttacagtcc	catcaacagt	gtaaaagtgt	tcctatttct	ctacatcctt	tccagcacct
	51301	gttggttccg	gactttaatg	atcgccattc	taactggtgt	gagatgggat	ctcattgtgg
	51361	ttttgatttg	catttctctg	atggccagtg	atgatgagca	tttttcatg	tgtcttttgg
25	51421	ctacataaat	gtcttctttt	gagaagtgtc	tgttcataatc	cttcaccac	ttttgatgg
	51481	ggtcatttga	tttttctttg	taaatttgtt	taagttcttt	tagattctgg	atattagccc
	51541	tttgtcagat	gggtagattg	taaaaatttt	ctccatttcc	gtaggtttcc	gtattcactt
	51601	gatggtagtt	tcttttgctg	tgcaagaagt	ctttagttaa	attagatccc	atttgtcaat
	51661	tttggctttt	gttgccattg	cttttggtgt	tttagtcatg	aagtccttgt	ccatgcctat
30	51721	gtcctgaatg	gtattgccta	ggttttcttc	taggggtttt	atgggttttag	gtctaactgt
	51781	taagtcttta	attcatcttg	aattaatttt	tgtataaggt	gtaaagaagg	gatccagttt
	51841	cagctttcta	catatggcta	gccagttttc	ccagcaccat	ttattaaata	gagaatcctt
	51901	tccccatttc	ttgtttttgt	caggtttgct	aaagatcaga	tggttgtaga	tgtgtggtat
	51961	tgtttctgag	ggctctgttc	tgttccattg	gtctatatct	ctgttttgg	accagtacca
35	52021	tgctgttttg	gttactgtag	ccttgtaata	tagtttgaag	tcaggtagcg	tgatgcctcc
	52081	agctttgttc	ttttggttta	ggattgtcct	ggcgatgcgg	gctctttttt	ggttccatat
	52141	gaactttaaa	gtagtttttt	tccaattctg	tggagaaaagt	cattggtagc	ttgatgggga
	52201	tgccattgaa	tctataaatt	accttgggta	gtatggccat	tttcatgata	ttgattcttc
	52261	ctacccatga	gaatggaatg	ttcttccatt	tgtttgctgc	ctcttttatt	tccttgagca
40	52321	gtgggtttgta	gttctccttg	aagaggtctt	ccacatccct	tgtaagttgg	attcctaagt
	52381	attttattct	ctttgaaaca	attgtgaatg	ggagttcact	catgatttgg	ctctctgttt
	52441	gtctgttatt	gggtgatagg	aatgcttgtg	atttttgcac	attgattttg	tatcctgaga
	52501	ctttgctgaa	gttgcttatc	agcttaagga	gattttgggc	tgagatgatg	gggttttcta
	52561	aatatacaat	catgtcatct	gcaaacagag	acaatttgac	ttcctctctt	cctatttgaa
45	52621	tatcctttat	ttctttctat	tgccctgattg	ccctggctag	aacgtccaat	actatgttga
	52681	ataggagtgg	tgacagagga	catccttggt	ttgtgccagt	tttcaaagg	aatgcttcca
	52741	gcttttgccc	attcagtatg	acattggctg	tgggtttgtc	gtgaatagct	cttattattt
	52801	tgagatatgt	cccatcaata	cctagtttat	ttagagtttt	tagcacaaag	gctgttgaat
	52861	tttgtcaaag	gccttttctg	catctattga	gataatcatg	gtttttgtct	ttgattctgt
50	52921	ttatatgatg	gattatattt	attgatttgc	atatgttgaa	ccagccttgc	atcccaggga
	52981	tgaagccaac	ttgatcatgg	tggataagct	ttttgatgtt	ctgctggatt	cgggttgcca
	53041	gtattttact	gaggattttt	ccatcgatct	tcatacaggga	tattggcctg	aaattctctt
	53101	tttttggtgt	gtctctgtca	ggctgtggta	tcaggatgat	gctggcctca	taaaatgagt
	53161	tagggaggat	tccctctttt	tctattgatt	agaatagttt	cagaatggta	ccagctcctc
55	53221	cttataacctc	tggtagaatt	cagctgtgaa	tccatctggt	cctgatggat	ttttttgggt
	53281	ggtaggctat	taattattgc	ctcaatttca	gagcctgtta	ttgggtctatt	aagagattca
	53341	acttcttcct	ggtttagtcc	tgggagggtg	tgtgtgtcca	ggaatttata	aatttctttt

	53401	agggttttcta	gtttatttgc	atagaagtgt	ttatagtgtt	ctctgatggg	agtttgtatt
	53461	tctgtgggat	tgggtgggat	atccccctta	tcacctttta	ttgcatctat	ttgattcttt
	53521	tctcttttct	tctttattag	tcttgctagt	gatctatcaa	ttttgttgat	ctttttaaaa
5	53581	aaccagctcc	tgggttcatt	gattttttga	aggagttttt	ctgtctctat	ctccttcagt
	53641	tctactctga	tcttagttat	ttcttgtctt	ctgctagctt	ttgaatgtgt	ttgctcttgc
	53701	ttctctaaat	tgtgatgtta	gggtgtcaat	tttagatctt	tcttgctttc	tcttgtgggc
	53761	atttagtgct	ataaaatttc	ctctacacac	tgctttaaat	gtgtcccaga	gattctggta
	53821	tgttgtgtct	ttgttctcat	tggtttcaaa	gaacatcttt	atttctgcct	tcacttcggt
10	53881	aagtaccag	tagtcactca	ggagcaggtt	gctcagtttc	catgtagtgt	agtgggtctg
	53941	agtgaagttc	ttaatcctga	gttctagtgt	gaaagcactg	tagtctgaga	ggcagtttgt
	54001	tataatttct	gttcttttac	atttgctgag	gagtgcctta	cttccaaacta	tgtagtcaat
	54061	ttttggaata	agtgtgatgt	ggtgccgaga	agaatgtata	ttctgttgat	ttggagtgga
	54121	gagttctgta	gatgtctatt	aggctccgct	ggtgcagagc	tgagttcaat	ttctggatat
	54181	ctttgttaat	tttctgtctt	gttgatctgt	ctaataattga	ccgtgggggtg	ataaagtctc
15	54241	ccattattat	tgtgtgggag	tctaagtctc	ttttaggttc	tctaaggact	tgctttgtga
	54301	atctgggtgct	cctgtattag	gtgcatatat	ttttaggata	gttagctctt	cttgttgaat
	54361	tgatccccctt	atcattatgt	aatggccctt	tttgcctctt	ttgatctttg	ttggtttaaa
	54421	gtctgtttta	tcagagacta	ggattgcaac	tctgctttt	ttttgccttt	catttccttg
20	54481	gtagatcttc	ctccatccct	ttattttgag	cctatgtgctg	tctctgcaca	tgaatgggt
	54541	ctgctgaata	cagcacactg	atgggtcttg	actctttatc	caatttgcca	gtccatgtct
	54601	tttaactgga	gcatttagcc	cattttacatt	taagggttaat	attgttatgt	gtgaatttga
	54661	tcctgtcatt	atgatgttag	ctgggtattt	tgctcgtag	ttgatgcagt	ttcttcctag
	54721	cctcaatgat	ctttacaatt	tggcatgttt	ttgcagtgcc	tggtactggg	tggtccttct
25	54781	catgtttagt	gcttccttca	ggagctcttg	taaggcaggc	ctggtgggtga	caaaatctct
	54841	cagcatttgc	ttgtctgtaa	aggattttat	ttctccttca	cttatgaagc	ttagtttggc
	54901	tggatatgaa	attctggggt	gaaaattctt	ttctttaaga	atgttgaata	ttggccccc
	54961	ctctcttctg	gctttagtag	tttctgccga	aagatgctgt	tagtctgatg	gacttccctt
	55021	tgtgggtaac	ctgcccttct	tctctcgctg	cacttaatgt	ttttccttc	atttcaactt
30	55081	tggtagaatct	gacaattatg	tgtctttgag	ttactcttct	tgaggagtat	ctttgcggca
	55141	ttctctgtat	ttcctgaatt	tgaatgctgg	cctgcctcac	tagattgggg	aagtctcctt
	55201	ggataaatatc	ctgcagagcg	ttttccaact	tggttccatt	ctccccatca	ctttcaggta
	55261	caccaatcag	atgtagattt	ggtcttttca	catagtccca	tatttcttgg	aggctttgtt
	55321	catttctttt	tactcttttt	tctctaaact	tctcttcttg	cttcatttca	ttcatttgat
35	55381	cttcaatccc	tttcttccac	ttgattgaat	cagctactga	agcttgtgca	tgtgtcacat
	55441	agttctcgtg	ccatggtttt	cagctccatc	aggctattta	aggcttcttc	tatgctgttt
	55501	tttctagtta	gccattcgtc	taatgttttt	tcaaggtttt	tagcttcttt	gctaaaagg
	55561	tcaaacatcc	tccttttagct	cggaggaggt	tgttattact	gatcatctga	agccttcttc
	55621	tctcaacttg	tgaaagtcat	tctctgtcca	gctttgttcc	attgtggcgc	aggagctgca
40	55681	ttcctttgga	ggagaagacg	tgtctgtatt	tttagaattt	tcagcttctc	tgtctgtgtt
	55741	tctccccatc	ttattgggtt	tatctacctt	tggtctttga	tgatgggtgac	gtacagatgg
	55801	ggttttgggtg	tggatgttct	ttctctttgt	tagttttcct	tctaacagtc	aggaccctca
	55861	gctgcaggctc	tgttggaggt	tgctggaggt	ccactccaga	ccctgtttgc	ttgggtatca
	55921	ccagcagagg	ctgcagaaca	gcaaataattg	cagaacggca	aatgttgctc	cctgattgtt
45	55981	cctctggaag	cttcgtctca	gaggggcacc	tggccgtatg	agggtgtcagt	cggcccctac
	56041	tgggaggtgc	ctcccagtta	ggctactcag	gggtcaggaa	cccacttgaa	gaggcagact
	56101	gtccattctc	agatatcata	ttccatgctg	ggaggacccc	tactcttttc	aaagctgtca
	56161	gacagggaca	tttaagtctg	cagaagtttc	tgctgtcttt	tggtcagctg	tgccctgccc
	56221	ctagaggtgg	agtctacaga	ggcaggcagg	cctccttgag	ctgcggtggg	ctccaccat
	56281	ttcgagcttc	ctggctgctt	tgtttaccta	ctcaagtctc	agcaatgggtg	gacaccctc
50	56341	ccccagcctc	gctgctgctt	tgcagttcga	tctcagactg	ctgtgctagc	agtgagccag
	56401	gctccgtggg	catgggaccc	tccgagccag	gcctgggaca	taatctcctg	gtgtgccgtt
	56461	tgctaagacc	attggaaaag	cacagtatta	gggtggggag	tgctctgatt	ttccaggtag
	56521	cgtcagtcac	ggcttccctt	ggctaggaaa	gggaattccc	caacccttgc	tgcttccctg
	56581	gtgaggtgat	gccccaccct	gctttggctc	atgctccgtg	ggttgtaccc	actgtctgac
55	56641	aagccccagt	gagatgaacc	cggtagctca	gttggaaatg	cagaaatcac	ccgtcttctg
	56701	catcactcac	gctgggggct	gtagactgga	gctgttcata	tttggccatc	ttggaacctc
	56761	cctttccaag	ttctttatta	cagagtgggt	cactgaaact	tcatggaaca	aattggaaat

56821 tatcttcttta attaattgtca ctgtctacca tgtatgggaa tttggtaaatt attatatggt
 56881 ttcaataaca tagtagatag aacattgtca aatctaaact tcagtgaatt gtaacagatc
 56941 ccacctgaaa ttctaaagaa aacagaattc taattgaaga ggtaaacttt ttacaggga
 57001 tgtcaactgc catttgggtc ctgtaaacaa aaaactgttt tttaaaaaag taaactttta
 57061 aagtattttc agatgacctc atttgctatc caagtggctt gagtatgctt gatgctaaga
 57121 cttctttgtt acagactgga gatgtgtgct actggggcag tgttgctctg tgacaaggag
 57181 gcagaggatg agggcaaggt tcgatgtgac tgtgaattct gggtggtctt ggctatcggg
 57241 agccttcatt gattacagca aaacagttgc tttcctaggg caatagtgtc tctgtcacc
 57301 aggctggagt tcagtggcat gatcaatcgc tcaactgtagc ctcaacttct tagactcaag
 57361 taatcctccc acctcagcct cccaagtagc tgaaactaca ggtgtgcacc accacaccta
 57421 attttttttaa tttttaagtt tttgtagaga catggtctca ctgtgttgcc cagggtgatc
 57481 tcgaattcct gggctctagt gatcctcccg cctcggcctc ccaaagtgtt gggattacaa
 57541 atgtgagcca ctgcacctgg ccttttgcaa ccttcttgac aatgcattcc tttattccct
 57601 aactggaagt aacttctttt tctttataaa attgtatctg taccttttct gggcatttc
 57661 tacctttata ttctagttac gtatgtccta cctccctcct agggaggag gtaagtaaga
 57721 ctggaaaagta gacttcatgt gtgatgaatg aatgaacaaa aggaagtcta acatatggat
 57781 atagtcaact ggatgcaaatt taaaaatttt taaatattga tttgcaagat ttcattaagg
 57841 tcaactctta atagtttgta tcatatatgt taggaaccaa atattaataa cttcttcagc
 57901 attaccatta tctttatagg actgtctaaa atgagcagcc atatctttaa actgtgtttt
 57961 ctctgattac acgctcacag gtaaaaccca aaggggctgg gaacaaacaa gactttttt
 58021 tttttctgta tgctgaatt atctgtactg ttgcttgttt tcccaccttt ggccatagaa
 58081 acttagttct aacatgctac aatttttgca gttctttctc ttagaaaaag accacattgt
 58141 ctgaaatttc atccatttaa gtaatcaagc cttaaagtgt aaggatcttg gtcattgata
 58201 atctagacct acaaagtagt atcttaatgg cactcctttt agaaagttag gttccaggac
 58261 acacatagct gcagtgtcca cattttgtaa gtccttctgt tgtcacagcc actctcttct
 58321 ctgtggctga tattctaaaa ctggcaacac atcctgatgg taaaagcttg gttcaggaga
 58381 caggtagacct actagcttta tggcatttga caggttacct aacctctctg acgcataatt
 58441 gcctcatcta tataatgggg ataataatac ccatcctgtc tcttgtaaa aatcaaatta
 58501 gatgacgcct gtgaatgttc tatagtctct tagacaaatg taagttatga ctacagcaag
 58561 agtaaaaagag catgttgttt ttagcattct ttcagtgaat tgtctaagac ttgtgagtca
 58621 cacttaaaagc taaacttgat atctacttca ttgattttct ttttagttct atgtactata
 58681 ttgaatttcc tgacagtggg gctatgaaag ccttcctagc attttataga tgtggttgaa
 58741 ttaatggctg taagccttaa agcagaatta gacagcatca atgaatttat taagtataaa
 58801 taaatatata atctgcttag caatattaca cagcctcttt atcttatgtg tgataaagag
 58861 tcatccgaag gttgaaaatg aagaattgtc ctggaagctc ttacttaatc ttttattatt
 58921 tcctaataca gtatataaaa ttactcattg aaagcttagc agaataagaa acaagaagtt
 58981 aaaaggctga aaactacaaa ttttgctatt attattgtta ttacttccca agtctcttat
 59041 tgatctgtta gaaatagagc tacacaggaa attgtaggac agttagatg tggtagtggt
 59101 atctgctttt taattattca agtaaggttt tattccatta gaggaactca tgaagttggt
 59161 catggctgat aattgctatc tgtcaaatc cttagagcag ggatccgcaa caccaggcc
 59221 atggattgtt accagtccct ggctgttag gaaccaggct gcacagtagg aagtgagcgg
 59281 cgggtgagca aacattgcca tctgagctct gtctcctttc agatcagcag cagcattaga
 59341 ttctcataga agcatgagcc ctgtgtgtaa ctgcacatgc aaggatcta ggttgtgtgc
 59401 tccttatgag aatccaattc cttatgataa tttaactgat gatctaaggt ggaaccattt
 59461 catcccaaaa ccatccaccc tgctactccc agaccgtgga aaaattgtct tccacaaaac
 59521 tggttcctgc tgccaaaaag gttgggacca ctgccttaga gtttataatt tggggttagc
 59581 acagcctata ttacctgag aatttcaatg gtttactga tctttccaaa tgaaaaggct
 59641 tcttacgaaa attatatcca aactgtcttt tctcttagtt taataaacct atcagtaagt
 59701 ttttactgag tactgtatt acatttttct ctgttaagca ttatggggc tcagacatga
 59761 tccattccct caaagaactt acctttcagc tgaagactga ctagaatgag caaatacgg
 59821 taacaattaa caagttagta ggccagctcg gccaacatgg tgaaacctg tctctactaa
 59881 aaatacaaaa attagccggg catggtgtg ggcgcctgta atcccagcta cctcctgctg
 59941 aggcaggaga attgcttgaa cccaggaggt ggagattgca gtgagccgag attgcacat
 60001 tgcactccag cctgggagac agagcaagac tctgtcttgg gaaaaaaaca aaacaaaaca
 60061 agtgagaagg gaatcaagta ctacgtaaga tgtaattgtg aatttttagg aaggaaggca
 60121 gtgtatgctg gagtaattag agaaagggg atgcatgatt tgatgcttga actggatcat
 60181 gaaggataag caagattttg gcagcaagt agaggagag aggagtttgt cagaggaatg

	60241	gaacaagtca	ggaggcagta	atgtgtacgg	cactccaagg	actctgtcct	gatcggagca
	60301	gaagtgatga	agcattgagt	agtttgagaa	agtttagctaa	gaagggtgag	ggcagatggt
	60361	ggagagagtt	gagtatcaga	gagaagacat	tagatttgag	cagatagaac	aaaaatgcc
	60421	ttgccagttt	ttgtgtaaga	gaatagtatt	aggggtggta	cccaggaagt	ggcatcaga
5	60481	gttgaatgga	gcagaaagag	tgtctgatac	agcatcggga	cgttgtgggc	acaaaatgag
	60541	gtgggtgagg	ctgagccaag	atgggtggcag	tgggatggat	aaaaagggat	ggccaggaca
	60601	aatattttta	aggaaaaatt	aacaggacat	ctttactgac	tggattggaa	ggctatgcag
	60661	gaaatatatt	gtcaaacttg	attccaggat	ttctatccta	tgcttgggtt	gccccaaata
	60721	tcagggaacc	attgttagaa	aaggtaggag	ataccactgt	tccaacaaaa	agtattgagt
10	60781	ttgggtgtgc	accacttaaa	cttcaaggcc	ttacaagtga	gtagacagtt	agttagaatt
	60841	gcagaagtgc	cactcagaga	gcagggtctg	caatgtgggg	ttggactttg	tcaccattgt
	60901	gttaattcct	aattctatgc	agatgctcag	cttgaggaat	acccatgttt	gggcttcaga
	60961	atgaaagcca	agtaatat	actcagatgc	caattttccc	tctgaaatat	ttgctcatgg
15	61021	aactgagaga	acaatatata	aagcattaat	tatttttctc	ataaagttat	taataaaaag
	61081	ataagatcag	tgaaaggcag	agtaaactag	aagccaagta	tagaaaaatg	tatcattcaa
	61141	agactcatta	ctgtagtggt	gaaaacaaaa	caattttcca	acagcttaag	atgcctcagt
	61201	attttggacc	atttttaagt	agttagtgtg	ggcacttagt	aaatatgtat	taaactatag
	61261	ttcattaatt	cttttttttt	tttttttgag	atggagtttc	actctgtgtc	cccaggctgc
	61321	atttttgctt	tcttagtgat	atataaaatg	tcagagtttc	caatgagtgt	atcttagatt
20	61381	tgattaaata	tggtattaaa	aaatagctga	tcacagaaaag	tctctaccag	tgtgatgtag
	61441	atggctaaaag	tattccacat	ttgcaaaact	ttattgacct	aaataagagg	tgccccctgg
	61501	gttgttttta	tttggactgg	gaatattagg	agaaaagctt	ttcattcagt	gtgtaagtac
	61561	aatctaccag	aaatagaaac	ccccatggac	gatctatttc	ttttagtggt	caggactcag
	61621	aacattcaca	aagatttagt	tgtagcggga	atagacatct	gtattttatt	caaaccaatt
25	61681	ttcccttcct	aatctgagaa	cattgtgcaa	tctaagcagt	tctaagcatg	tttgctattc
	61741	gtgcaaaagt	agagtaaatc	taaaagaaat	ttttttgtgt	gtttagggat	ggtataaaag
	61801	tctcttagtg	gttgaaaatg	ttatttctta	caaaaagtga	gaacatttgc	ttttcaatac
	61861	cagagttttc	agccatttct	gcattctgac	ctattgactg	gaggtagggt	gcctttgaat
	61921	tcagtaaaac	ttcatgggca	gaaacacagt	tccttttcct	acttatttgg	atatcatgat
30	61981	ggccattgca	tgtatgtgtc	tttttgtaag	tccatgcctc	agaactgaga	agtaggaata
	62041	aaattagggg	cagggtctgg	gatgctactc	tttgctgctg	agaaaacaca	tgcttcaggt
	62101	aagtgattct	gaagtccttc	accacctgac	ggtaaccttg	ggttggtcca	taggtatggt
	62161	ttcattttgc	ttgttcatcc	attttaattg	gcttcctaga	gcatgcttgt	agatgtagag
	62221	ccaaatttag	agtagagcaa	ccctctggca	aacaggaaga	gattaatttt	gtggtatgct
35	62281	tttaagggac	ttcccaggaa	acttcaaaaag	cagaaaaaga	agcactagct	gcctattcca
	62341	aaatgtgtaa	aacaccactc	agctttttta	aagtaggata	aactcagagc	gcgcgcacac
	62401	gcgcgcgcgc	acacacacac	acacacacag	agagaacatc	tctagtaaaa	agaaaagtgt
	62461	agcttttcta	gctagatgtg	tgtattagcc	agaaaaagcc	aaggagtga	gggttttaga
	62521	gaactggagg	agataaagtg	gagctcgcac	atgggaggca	tttgaatggg	acttaaatgt
40	62581	cttttttaag	ctgacttttt	cagttttctc	cttaccagac	acattgtttt	catgacatta
	62641	gccccaggca	tagacacatc	attaaaatga	acatgtcaaa	aaatgatttc	tgtttagaaa
	62701	taagcaaaac	attttccagt	gtgaccaccc	aggtgtagaa	taaagaacag	tggaattggg
	62761	agccctgagt	tctaacataa	actttcttca	tgacataagg	caagtcttct	atggcctttg
	62821	gtttccttac	ctgtaaaaca	ggatgggtca	atgaaattat	ctttcttctt	tgctataata
45	62881	gagtatctct	gtgggaagag	gaaaaaaaaa	gtcaatttta	aggctcctta	tagttcccca
	62941	actgctgttt	tattgtgcta	ttcatgccta	gacatcacat	agctagaaag	gcccacaga
	63001	cccctcaggc	cactgctgtt	cctgtcacac	attcctgcaa	aggacctagt	tgtaacttg
	63061	aaaaaaatta	ctattaatta	cacttgacgt	tgttgcttag	taacatttat	gattttgtgt
	63121	ttctcgtgac	agcatgagca	gagatcatta	aaaattaaac	ttacaaagct	gctaaagtgg
50	63181	gaagaaggag	aacttgaagc	cacaattttt	gcacttgctt	agaagccatc	taatctcagg
	63241	tttatatgct	agatcttggg	ggaaacactg	catgtctctg	gtttatatta	aaccacatac
	63301	agcacactac	tgacactgat	ttgtgtctgg	tgacagctga	gtttatcacc	aagacataaa
	63361	aaaaccttga	ccctgcagaa	tggcctggaa	ttacaatcag	atgggccaca	tgccatccc
	63421	gtgaaagaaa	gccctaacca	gttttctgtc	tgttttctgc	tttctcccta	cagttccacc
55	63481	aggtgagaag	agtgatgacc	atccttttcc	ttactatggt	tatttcatac	tttggttgca
	63541	tgaaggctgc	ccccatgaaa	gaagcaaaac	tccgaggaca	aggtggcttg	gcctaccag
	63601	gtgtgcggac	ccatgggact	ctggagagcg	tgaatgggcc	caaggcaggt	tcaagaggct

	63661	tgacatcatt	ggctgacact	ttcgaacacg	tgatagaaga	gctgttggat	gaggaccaga
	63721	aagttcggcc	caatgaagaa	aacaataagg	acgcagactt	gtacacgtcc	agggatgatgc
	63781	tcagtagtca	agtgcctttg	gagcctcctc	ttctctttct	gctggaggaa	tacaaaaatt
5	63841	acctagatgc	tgcaaacatg	tccatgaggg	tccggcgcca	ctctgaccct	gcccgccgag
	63901	gggagctgag	cgtgtgtgac	agtattagt	agtgggtaac	ggcggcagac	aaaaagactg
	63961	cagtggacat	gtcggggcggg	acggtcacag	tccttgaaaa	ggccctgtga	tcaaaaggcc
	64021	aactgaagca	atacttctac	gagaccaagt	gcaatcccat	gggttacaca	aaagaaggct
	64081	gcaggggcat	agacaaaagg	cattggaact	cccagtgcg	aactaccag	tcgtacgtgc
	64141	gggcccttac	catggatagc	aaaaagagaa	ttggctggcg	attcataagg	atagacactt
10	64201	cttgtgtatg	tacattgacc	attaaaaagg	gaagatagt	gatttatgtt	gtatagatta
	64261	gattatattg	agacaaaaat	tatctatttg	tatatataca	taacagggtg	aattattcag
	64321	ttaagaaaaa	aataatttta	tgaactgcat	gtataaatga	agtttataca	gtacagtggg
	64381	tctacaatct	atattattgga	catgtccatg	accagaagg	aaacagtc	ttgcgcacaa
	64441	cttaaaaaag	ctgcattaca	ttccttgata	atgttgtgt	ttgttgccgt	tgccaagaac
15	64501	tgaaaaacata	aaaagttaaa	aaaaataata	aattgcatgc	tgctttaatt	gtgaattgat
	64561	aataaaactgt	cctctttcag	aaaacagaaa	aaaacacaca	cacacacaac	aaaaatttga
	64621	acccaaacat	tccgtttaca	tttttagacag	taagtatctt	cgttcttgtt	agtactatat
	64681	ctgttttact	gcttttaact	tctgatagcg	ttggaattaa	aacaatgtca	aggtgctgtt
20	64741	gtcattgctt	tactggctta	ggggatggg	gatgggggt	atatttttgt	ttgttttgtg
	64801	tttttttttc	gtttgtttgt	tttgtttttt	agttcccaca	gggagtagag	atggggaaag
	64861	aattcctaca	atataatatt	tggctgataa	aagatacatt	tgtatgttgt	gaagatgttt
	64921	gcaatatcga	tcagatgact	agaaagtga	taaaaaattaa	ggcaactgaa	caaaaaaatg
	64981	ctcacactcc	acatcccgtg	atgcacctcc	caggccccgc	tcattctttg	ggcgttggct
25	65041	agagtaagct	gcttttgacg	gaaggaccta	tgtttgctca	gaacacattc	ttccccccc
	65101	tccccctctg	gtctcctctt	tgttttgttt	taaggaagaa	aaatcagttg	cggttcttga
	65161	aatattttac	cactgctgtg	aacaagtga	cacattgtgt	cacatcatga	cactcgtata
	65221	agcatggaga	acagtgattt	tttttttagaa	cagaaaacaa	caaaaaataa	ccccaaaatg
	65281	aagattattt	tttatgagga	gtgaacattt	gggtaaatca	tggctaagct	taaaaaaaac
30	65341	tcattggtgag	gcttaacaat	gtcttgtaag	caaaaggtag	agccctgtat	caaccagaa
	65401	acacctagat	cagaacagga	atccacattg	ccagtgcac	gagactgaac	agccaaatgg
	65461	aggctatgtg	gagttggcat	tgcatttacc	ggcagtgcg	gaggaatttc	tgagtggcca
	65521	tcccaaggct	taggtggagg	tggggcatgg	tatttgagac	attccaaaac	gaaggcctct
	65581	gaaggaccct	tcagaggtgg	ctctggaatg	acatgtgtca	agctgcttgg	acctcgtgct
35	65641	ttaagtgcct	acattatcta	actgtgctca	agaggttctc	gactggagga	ccacactcaa
	65701	gccgacttat	gcccaccatc	ccacctctgg	ataattttgc	ataaaattgg	attagcctgg
	65761	agcaggttgg	gagccaaaatg	tggcatttgt	gatcatgaga	ttgatgcaat	gagatagaag
	65821	atgtttgcta	cctgaacact	tattgctttg	aaactagact	tgaggaaacc	agggtttatc
	65881	ttttgagaac	ttttggttaag	ggaaaaggga	acaggaaaag	aaaccccaaa	ctcaggccga
40	65941	atgatcaagc	ggaccatag	gaaatcctgt	ccagagacaa	gacttcggga	aggtgtctgg
	66001	acattcagaa	caccaagact	tgaagggtgcc	ttgctcaatg	gaagaggcca	ggacagagct
	66061	gacaaaattt	tgtcctccag	tgaaggccac	agcaaccttc	tgcccatcct	gtctgttcat
	66121	ggagagggtc	cctgcctcac	ctctgccatt	ttgggttagg	agaagtcaag	ttgggagcct
	66181	gaaatagtgg	ttcttggaag	aatggatccc	cagtgaatac	tagagctcta	agcccattca
45	66241	gcccatctca	cacctgaaaa	tgtagtgat	caccacttgg	accagcatcc	ttaagtatca
	66301	gaaagcccca	agcaattgct	gcattcttagt	agggtgagg	ataagcaaaa	gaggatgttc
	66361	accataaccc	aggaatgaag	ataccatcag	caaagaattt	caatttgttc	agtccttcat
	66421	ttagagctag	tctttcacag	taccatctga	atacctcttt	gaaagaagga	agactttacg
	66481	tagttagat	ttgttttggt	ttgtttgaaa	atattatctt	tgtaattatt	tttaatatgt
50	66541	aaggaaatgct	tggaatatct	gctatatgtc	aactttatgc	agcttccttt	tgagggacaa
	66601	attttaaaca	aacaaccccc	catcacaac	ttaaaggatt	gcaagggcca	gatctgttaa
	66661	gtggtttcat	aggagacaca	tccagcaatt	gtgtggctcag	tggtcttttt	acccaataag
	66721	atacatcaca	gtcacatgct	tgatgggtta	tggtgacct	agatttattt	tgtaaaatc
	66781	tctctctgtt	gtgttcgttc	ttgttctgtt	ttgttttgtt	ttttaaagtc	ttgctgtggg
55	66841	ctctttgtgg	cagaagtgtt	tcattgcatg	cagcaggcct	gttgcttttt	tatggcgatt
	66901	cccattgaaa	atgtaagtaa	atgtctgtgg	ccttggtctc	tctatggtaa	agatattatt
	66961	caccatgtaa	aacaaaaaac	aatattttatt	gtatttttagt	atatttatat	aattatgtta
	67021	ttgaaaaaaa	ttggcattaa	aacttaaccg	catcagaacc	tattgtaaat	acaagttcta

67081 ttttaagtgtta ctaattaaca tataatatat gttttaaata tagaattttt aatgttttta
 67141 aatataatttt caaagtacat aaaa

Hair Cell Differentiation-Suppressing Gene

5 The term “hair cell differentiation-suppressing gene” refers to a gene encoding a protein (e.g., a transcription factor) that positively contributes (directly or indirectly) to the suppression of hair cell differentiation from supporting cells in a primate (e.g., a human). Non-limiting examples of hair cell differentiation-suppressing genes include: HES1, HES5, CDKN1B, and SOX2.

10 The term “mutation in a hair cell differentiation-suppressing gene” refers to a modification in a hair cell differentiation-suppressing gene that results in the production of a hair cell differentiation-suppressing protein having one or more of: one or more amino acid substitutions, and one or more amino acid insertions as compared to the wildtype hair cell differentiation-suppressing protein, and/or results in an increase in the expressed level of the
 15 encoded hair cell differentiation-suppressing protein in a primate cell as compared to the expressed level of the encoded hair cell differentiation-suppressing protein in a primate cell not having a mutation. In some embodiments, the mutation can result in the gain (or an increase in the level) of expression of a hair cell differentiation-suppressing mRNA or a hair cell differentiation-suppressing protein, or both the mRNA and protein. In some embodiments, the
 20 mutation can result in the production of an altered hair cell differentiation-suppressing protein having a gain or increase in one or more biological activities (functions) as compared to a wildtype hair cell differentiation-suppressing protein.

In some embodiments, the mutation is an insertion of one or more nucleotides into a hair cell differentiation-suppressing gene. In some embodiments, the mutation is in a regulatory
 25 sequence of the hair cell differentiation-suppressing gene, i.e., a portion of the gene that is not coding sequence. In some embodiments, a mutation in a regulatory sequence may be in a promoter or enhancer region and prevent or reduce the proper transcription of the hair cell differentiation-suppressing gene (e.g., a mutation in a regulatory sequence that increases the transcription of the hair cell differentiation-suppressing gene).

Hes family basic Helix-Loop-Helix (bHLH) transcription factor 1 (HES1)

The HES1 gene encodes hes family bHLH transcription factor 1, and acts as a transcriptional repressor. HES1 binds to the ATOH1 promoter to inhibit transcription in supporting cells and drives lateral inhibition (Abdolazimi et al. (2016) *Development* 143:841-850). Loss of HES1 results in supernumerary inner hair cells in early development. HES1 inhibition after damage induces hair cell regeneration (Du et al. (2018) *Mol. Ther.* 26(5):1313-1326).

The human HES1 gene is located on chromosome 3q29. It contains 4 exons encompassing ~15 kilobases (kb) (NCBI Accession No. NM_005524). The full-length wildtype HES1 protein expressed from the human HES1 gene is 280 amino acids in length.

Methods of detecting mutations in a gene are well-known in the art. Non-limiting examples of such techniques include: real-time polymerase chain reaction (RT-PCR), PCR, sequencing, Southern blotting, and Northern blotting.

An exemplary human wildtype HES1 protein is or includes the sequence of SEQ ID NO: 36. Non-limiting examples of a nucleic acid encoding a wildtype HES1 protein is or includes SEQ ID NO: 37.

Human Full-length Wildtype HES1 Protein (SEQ ID NO: 36)

MPADIMEKNSSSPVAATPASVNTTPDKPKTASEHRKSSKPIMEKRRRRARINESLSQLKTLILDALKKSSRHSKLEK
ADILEMTVKHLRLNLQRAQMTAALSTDPVLGKYRAGFSECMNEVTRFLSTCEGVNTEVTRRLGLHLANCMTQINAMT
YPGQHPALQAPPPPPGPGGPQHAPFAPPPPLVPIPGGAAPPPGGAPCKLGSQAGEAAKVFGGFQVVPAPDGQFAF
LIPNGAFAHSGPVI PVYTSNSGTSVGPNAVSPSSGPSLTADSMWRPWRN

Human Wildtype HES1 cDNA (SEQ ID NO: 37)

atgccagctgatataatggagaaaaattcctcgtccccgggtggctgctaccccagccagtggtcaacacgacaccgga
taaaccaaagacagcatctgagcacagaaagtcataaagcctattatggagaaaagacgaagagcaagaataaatg
aaagtctgagccagctgaaaacactgattttggatgctctgaagaaagatagctcgcggcattccaagctggagaag
gcgacattctggaaatgacagtgaagcacctccggaacctgcagcgggcgagatgacggctgcgctgagcacaga
cccaagtgtgctggggaagtaccgagccggcttcagcgagtgcatgaacgaggtgacccgcttcctgtccacgtgcg
agggcggttaataccgaggtgcgcactcggctgctcgccacactggccaactgcatgaccagatcaatgccatgacc
taccggggcagccgcaccccgcttgcaggcgccgccaccgccccaccgggacccggcgggccccagcacgcgcc
gttcgcgcgcgcgcgcactcgtgccatccccggggcgcgggcgccccctccggggcgccccctgcaagctgg
gcagccaggctggagaggcggttaaggtgtttggaggcttcagggtgtaccggctcccgatggccagtttgcttcc

ctcattcccaacggggccttcgcgcacagcggccctgtcatccccgtctacaccagcaacagcggcacctccgtggg
 cccaacgcagtgacaccttcagcggccctcgcttacggcggaactccatgtggaggccgtggcggaactga

A non-limiting example of a human wildtype HES1 genomic DNA sequence is SEQ ID
 5 NO: 38. The exons in SEQ ID NO: 38 are: nucleotide positions 1-347 (exon 1), nucleotide
 positions 348-443 (exon 2), nucleotide positions 444-531 (exon 3), and nucleotide positions 532-
 1461 (exon 4).

Human Wildtype HES1 Gene (SEQ ID NO: 38)

10 1 gggatcacac aggatccgga gctggtgctg ataacagcgg aatcccccggt ctacctctct
 61 ccttggtcct ggaacagcgc tactgatcac caagtagcca caaaatataa taaaccctca
 121 gcacttgctc agtagttttg tgaaagtctc aagtaaaaga gacacaaaca aaaaatttctt
 181 tttcgtgaag aactccaaaa ataaaattct ctagagataa aaaaaaaaaa aaaaggaaaa
 241 tgccagctga tataatggag aaaaattcct cgtccccggt ggctgctacc ccagccagtg
 15 301 tcaacacgac accggataaa ccaaagacag catctgagca cagaaagtca tcaaagccta
 361 ttatggagaa aagacgaaga gcaagaataa atgaaagtct gagccagctg aaaacactga
 421 ttttgatgc tctgaagaaa gatagctcgc ggcatccaa gctggagaag gcggacattc
 481 tggaaatgac agtgaagcac ctccggaacc tgcagcgggc gcagatgacg gctgcgctga
 541 gcacagaccc aagtgtgctg gggaagtacc gagccggctt cagcgagtgc atgaacgagg
 20 601 tgaccgcgtt cctgtccacg tgcgagggcg ttaataccga ggtgcgcact cggctgctcg
 661 gccacctggc caactgcatg acccagatca atgccatgac ctacccccggg cagccgcacc
 721 ccgccttgca ggcgccgcca ccgccccac cgggaccggg cggccccag cagcgccgt
 781 tcgcgccgcc gccgccactc gtgccatcc ccgggggcgc ggcgccccct cccggcggcg
 841 cccctgcaa gctgggcagc caggctggag aggcggctaa ggtgtttgga ggcttccagg
 25 901 tggtaaccggc tcccgatggc cagtttgctt tcctcattcc caacggggcc ttcgcgcaca
 961 gcggccctgt catccccgtc tacaccagca acagcggcac ctccgtgggc cccaacgcag
 1021 tgtcaccttc cagcggcccc tcgcttacgg cggactccat gtggaggccg tggcggaact
 1081 gagggggctc aggccacccc tcctcctaaa ctcccaacc cacctctctt ccctccggac
 1141 tctaaacagg aacttgaata ctgggagaga agaggacttt tttgattaag tggttacttt
 30 1201 gtgttttttt aatttctaag aagttacttt ttgtagagag agctgtatta agtgactgac
 1261 catgcactat atttgtatat attttatatg ttcattattg attgcgctt tgtattataa
 1321 aagctcagat gacatttcgt tttttacacg agatttcttt tttatgtgat gccaaagatg
 1381 tttgaaaatg ctcttaaaat atcttccttt ggggaagttt atttgagaaa atataataaa
 1441 agaaaaaagt aaaggctttt aaaaaaaaaa aaaaa

35

Non-limiting examples of siRNA targeting HES1 are described in, e.g., Zhang et al., *World J. Gastroenterol.* 24(29):3260-3272, 2018; Du et al., *Mol. Ther.* 26(5):1313-1326, 2018; Li et al., *Oncol. Lett.* 14(4):3989-3996, 2017; and Du et al., *Hear Res.* 304:91-110, 2013. Non-limiting examples of shRNA targeting HES1 are described in, e.g., Cenciarelli et al., *Oncotarget* 8(11):17873-17886, 2017, and Wang et al., *Oncotarget* 6(34):36713-36730, 2015.

Hes family bHLH transcription factor 5 (HES5)

The HES5 gene encodes hes family bHLH transcription 5, and acts as a transcriptional repressor. HES5 is a Notch-pathway activator, and binds the ATOH1 promoter to inhibit transcription in supporting cells. Loss of HES5 results in supernumerary outer hair cells in early development. HES5 inhibition in adult mouse utricle results in increased regeneration after aminoglycoside damage (Jung et al. (2013) *Mol. Ther.* 21(4):834-841; Abdolazimi et al. (2016) *Development* 143:841-850).

The human HES5 gene is located on chromosome 1p36. It contains 3 exons encompassing ~18 kilobases (kb) (NCBI Accession No. NM_001010926.3). The full-length wildtype HES5 protein expressed from the human HES5 gene is 166 amino acids in length.

Methods of detecting mutations in a gene are well-known in the art. Non-limiting examples of such techniques include: real-time polymerase chain reaction (RT-PCR), PCR, sequencing, Southern blotting, and Northern blotting.

An exemplary human wildtype HES5 protein is or includes the sequence of SEQ ID NO: 39. Non-limiting examples of a nucleic acid encoding a wildtype HES5 protein is or includes SEQ ID NO: 40.

Human Full-length Wildtype HES5 Protein (SEQ ID NO: 39)

MAPSTVAVELLSPKEKNRLRKPVVEKMRRDRINSSIEQLKLLLEQEFARHQPNKLEKADILEMAVSYLKHSKAFVA
AAGPKSLHQDYSEGYSWCLQEAVQFLTLHAASDTQMKLLYHFQRPPAAPAPAKEPKAPGAAPPPALSAKATAAAAA
AHQPACGLWRPW

Human Wildtype HES5 cDNA (SEQ ID NO: 40)

atggccccccagcactgtggccgtggagctgctcagccccaaagagaaaaaccgactgcggaagccggtggtggagaa
gatgcgcccgcgaccgcatcaacagcagcatcgagcagctgaagctgctgctggagcaggagttcgcgcgccaccagc
ccaactccaagctggagaaggccgacatcctggagatggctgtcagctacctgaagcacagcaaagccttcgtcgcc
gccgcccggccccaagagcctgcaccaggactacagcgaaggctactcgtggtgcctgcaggaggccgtgcagttcct

gacgctccacgcccagccagcgacacgcagatgaagctgctgtaccacttccagcgcccccgccgccccgcccgcgc
ccgccaaggagcccaaggcgccggcgccgccccgccccgcgctctccgccaaggccaccgcccgcgcccgcgc
gcgcaccagccccgctgcgccctctggcgccctggtga

- 5 A non-limiting example of a human wildtype HES5 genomic DNA sequence is SEQ ID NO: 41. The exons in SEQ ID NO: 41 are: nucleotide positions 1-135 (exon 1), nucleotide positions 136-301 (exon 2), and nucleotide positions 302-1306 (exon 3).

Human Wildtype HES5 Gene (SEQ ID NO: 41)

10 1 cgcgcttgcc cttgcccgcg cccgctcgcc tcgtctcgcc cggcctcccc gcgtcgccctc
61 gtcgcctgtt ccgcgccagg catggccccc agcactgtgg ccgtggagct gtcagcccc
121 aaagagaaaa accgactgcg gaagccggtg gtggagaaga tgcgcccga ccgcatcaac
181 agcagcatcg agcagctgaa gctgctgctg gagcaggagt tcgcgcgga ccagcccaac
241 tccaagctgg agaaggccga catcctggag atggctgtca gctacctgaa gcacagcaaa
15 301 gccttcgctc ccgcccgcg ccccaagagc ctgcaccagg actacagcga aggctactcg
361 tgggtgctgc aggaggccgt gcagttcctg acgctccacg ccgcccagcga cacgcagatg
421 aagctgctgt accacttcca gcggcccccg gccgcgcccg ccgcgcccgc caaggagccc
481 aaggcgccgg gcgcccgcgc cccgcccgcg ctctccgcca aggccaccgc cgccgcccgc
541 gccgcgcacc agcccgcctg cggcctctgg cggccctggt gaccgggcg gacctgcggg
20 601 cgcgcggccc gacgaccaga gggcgagcct gtcctctcg cctgtaggga agcgcccttc
661 cgccgtcgtc cgccccgggc ttggacgcgc ccttctccgg aaggctctgg cccaagctg
721 gccggcccgc aggagcccca ttctcagaga atgtgtgtgc agagtccctg ccgttttagg
781 acaatcaggg cccatcttct gccaagtgtc tgaccccatg gggttgttct gtgtttgcat
841 ttaagcaagt gacttctggg aagtccccgg ccgcccgggg ttctatgata tttgtagtgc
25 901 cggggctcgc aactgctgc cccagcctg tagaggactt tcttcagggc ccgtagctgc
961 tgggctgacc cctggcaggc gggctgtgcc gcgggcacat ttgccttttg tgaaggccga
1021 actcgagctg taccctcata ggaaacagt atcaccgccg acgggctcc aggaccctga
1081 gggccatggc caaaaggctc ctgagtgtgc ctggtggtct ggctggggct cacggtgggc
1141 tgtctgggga ggggtgggtc ctccactatg atccttaaag gattcctctg tgtgggtgga
30 1201 tgcgtgtggg cagactttg tactcagaaa ttgaactctc agtcacgtgg aagccacggg
1261 actgctccga agccgccata ataaaatctg attgttcagc ccccaaaaaa aaaaaaaaaa

- Non-limiting examples of siRNA targeting HES5 are described in, e.g., Gu et al., *Oncol. Rep.* 37(1):474-482, 2017; Zhu et al., *Exp. Mol. Pathol.* 99(3):474-484, 2015; Du et al., *Hear Res.* 304:91-110, 2013; Jung et al., *Mol. Ther.* 21(4):834-841, 2013; and Liu et al., *Int. J. Gynecol. Cancer* 20(7):1109-1116, 2010. Non-limiting examples of shRNA targeting HES5 are

described in, e.g., Lee et al., *J. Neurochem.* 100(6):1531-1542, 2007; and Osario et al., *Development* 140:1-12, 2013.

Cyclin dependent kinase inhibitor 1B (Cdkn1b) (p27^{kip1})

5 The CDKN1B gene encodes a cyclin-dependent kinase inhibitor (p27^{kip1}). CDKN1B is a cell cycle regulator and controls the cell cycle exit of supporting cells. For example, p27^{kip1} binds to and prevents activation of cyclin E (CDK2) and cyclin D (CDK4). Inhibition of CDKN1B promotes supporting cell proliferation and regeneration induction through its canonical pathway and a non-canonical pathway that involves Gata3 (Minoda et al. (2007) *Hear Res.* 232(1-2):44-51; Walters et al. (2014) *J. Neurosci* 34(47):15751-15763; Walters et al. (2017) 10 *Cell Rep* 19(2):307-320).

The human CDKN1B gene is located on chromosome 12p13. It contains 3 exons encompassing ~5 kilobases (kb) (NCBI Accession No. NG_016341.1). The full-length wildtype CDKN1B protein expressed from the human CDKN1B gene is 198 amino acids in length.

15 Methods of detecting mutations in a gene are well-known in the art. Non-limiting examples of such techniques include: real-time polymerase chain reaction (RT-PCR), PCR, sequencing, Southern blotting, and Northern blotting.

An exemplary human wildtype CDKN1B (p27^{kip1}) protein is or includes the sequence of SEQ ID NO: 42. Non-limiting examples of a nucleic acid encoding a wildtype CDKN1B (p27) 20 protein is or includes SEQ ID NO: 43.

Human Full-length Wildtype CDKN1B (p27^{kip1}) Protein (SEQ ID NO: 42)

MSNVRVSNNGSPSLERMDARQAEHPKPSACRNLFPGVDHEELTRDLEKHCRDMEEASQRKWNFDFQNHKPLEGKYEWQ
EVEKGSLEPEFYRPPRPPKGACKVPAQESQDVSGSRPAAPLIGAPANSEDTLHVDPKTDPSDSQTGLAEQCAGIRKR
25 PATDDSSSTQNKRANRTEENVSDGSPNAGSVEQTPKKPGLRRRQT

Human Wildtype CDKN1B (p27^{kip1}) cDNA (SEQ ID NO: 43)

atgtcaaactgtgcgagtggtctaacgggagccctagcctggagcggatggacgccaggcaggcggagcaccccaagcc
ctcggcctgcaggaacctcttcggcccggtggaccacgaagagttaaccgggacttggaagaagcactgcagagaca
30 tgggaagaggcgagccagcgcaagtggaaatttcgattttcagaatcacaaccctagagggaagtagcagtggtgcaa
gaggtggagaagggcagcttgcccaggttctactacagacccccgcggccccccaaaggtgcctgcaaggtgccggc
gcaggagagccaggatgtcagcgggagccgcccggcgccgctttaattggggctccggctaactctgaggacacgc
atttggtggacccaaagactgatccgtcggacagccagacggggttagcggagcaatgcgcaggaataaggaagcga

cctgcaaccgacgattcttctactcaaaacaaaagagccaacagaaacagaagaaaatgtttcagacggttccccaaa
tgccggttctgtggagcagacgccaagaagcctggcctcagaagacgtcaaacgtaa

A non-limiting example of a human wildtype CDKN1B (p27^{kip1}) genomic DNA sequence
5 is SEQ ID NO: 44. The exons in SEQ ID NO: 44 are: nucleotide positions 1-1045 (exon 1),
nucleotide positions 1556-1685 (exon 2), and nucleotide positions 3767-5114 (exon 3). The
introns in SEQ ID NO: 44 are: nucleotide positions 1046-1555 (intron 1) and nucleotide
positions 1686-3766 (intron 2).

10 Human Wildtype CDKN1B (p27^{kip1}) Gene (SEQ ID NO: 44)

1 ttaaggccgc gctcgccagc ctcggcgggg cggctccgc cgccgcaacc aatggatctc 61
ctcctctgtt taaatagact cgccgtgtca atcattttct tcttcgtcag cctcccttcc 121
accgccatat tgggccacta aaaaaagggg gctcgtcttt tcggggtgtt tttctcccc 181
tcccctgtcc ccgcttgctc acggctctgc gactccgacg ccggcaagggt ttggagagcg 241
15 gctgggttcg cgggaccgcg gggcttgac ccgccagac tcggacgggc tttgccaccc 301
tctccgcttg cctggtcccc tctcctctcc gccctccgc tcgccagtcc atttgatcag 361
cggagactcg gcggccgggc cggggcttcc ccgcagcccc tgcgcgctcc tagagctcgg 421
gccgtggctc gtccgggtct gtgtcttttg gctccgaggg cagtcgctgg gcttccgaga 481
ggggttcggg ctgcgtaggg gcgctttgtt ttgttcgggt ttgttttttt gagagtgcga 541
20 gagaggcggg cgtgcagacc cgggagaaa atgtcaaacg tgcgagtgtc taacgggagc 601
cctagcctgg agcggatgga cgccaggcag gcggagcacc ccaagccctc ggctgcagg 661
aacctcttcg gcccggtgga ccacgaagag ttaaccggg acttgagaa gcactgcaga 721
gacatggaag aggcgagcca gcgcaagtgg aatttcgatt ttcagaatca caaaccctca 781
gagggcaagt acgagtggca agaggtggag aagggcagct tgcccagatt ctactacaga 841
25 ccccgcggc ccccaaagg tgctgcaag gtgccggcg aggagagcca ggatgtcagc 901
gggagccgcc cggcgcgcc ttaattggg gctccggcta actctgagga cacgcatttg 961
gtggacccaa agactgatcc gtcggacagc cagacgggt tagcggagca atgcgcagga 1021
ataaggaagc gacctgcaac cgacggtaat gacccttcc caaccataga atgtgtttgg 1081
ggccccgctt tgctgtctgg aggggtgtta ccttagcttg cttttcggcg tattctgatt 1141
30 tagctttggg agagctaact ttattggtct taggtgttca gtgctacctg gccactgct 1201
tgtctgtttg tgacttttaa gtcagaaact ggagatggta agatccgata atttcctaa 1261
cttaatacat cgcggtccct ctactagca actcctaggt atgtgacaaa gttgggatgt 1321
ttatcaacgg tccgcctcct ggctagggaa agagctctgg ggcgagaaat gcactttctg 1381
ttttttgaaa acaacctcat tttgtgccct taaaagccac tggggatgac ggatccagga 1441
35 ttgtgggtgg aggtagtggg tttttcatcc cctgactatg gggccaactt ctgccagcca 1501
ttgttttttc taataaagat tgtgtgttct ttttaaaat tcccctgcg cttagattct 1561

tctactcaaa acaaaaagagc caacagaaca gaagaaaatg tttcagacgg ttccccaat 1621
 gccggttctg tggagcagac gccaagaag cctggcctca gaagacgtca aacgtaaaca 1681
 gctcggtagg ttgatcacta aaggagcacg cactggaacc cggggccttc agacctcacg 1741
 atacctgatc ttactggttg ctggcaaatt aaaagcttat ggggttttgt tttgtttata 1801
 5 cttcgtgagg tcaaaaaagt agcaatgggg aaggctgggg atacggtaat tcctcagagt 1861
 ttctatgcc agagatactt tctcttcaaa ctgttgacca gagcagctac ttgtaacca 1921
 ggcccatcg gtaggaag tcgtttccct gtgagtgcca ctaaaacgtg ttgggagcaa 1981
 taggttcttt gcccatccga acaagaacta gggactccc tcagtccgaa ttaatgagaa 2041
 ttaatttcct agaggttcag cttgagtcgg taacagattt tgagccatac atggaaaaat 2101
 10 ggcaaataca tgattaagtt tcaattttga gggggaatgt ttggtagaaa ttgctcatct 2161
 ttggttatgc aagggttag agatgtgaat aggatggtat gttgtgttct ttgacatttt 2221
 aataaactgt cactttccct gttgtctcct aagtttgag agagaaggaa ccagtatttg 2281
 caaaaaccaa atggaaagat aaaaaagtta ctaaagtttc tacagaattt ctggtaacac 2341
 tgaagttgca aagcagaagt taaattaact cttgtcagta agcaatccag gaacacgtca 2401
 15 gccagtgtat gctaattgtg ccgtaacagg gtgatttgga tattttagg ggaaatgggt 2461
 agtaaatatc aagactggtg accgtaggtc agcccagcac aaagggaagtg gagatttttc 2521
 catgcacaag aatctgatca ctgtaaatac ctaatttgaa taattcagtc ccagataac 2581
 caacatgggt tgggtattca taataaacta catattttaa tagtttatta gcttccttta 2641
 gaccaagact gtgacctctt tattttctaa agcacacacg tagtttagca tatgaggcga 2701
 20 taaaatattg atgttaactt tttaaatccc cagttataaa aattttaaaa taacagggat 2761
 taaggtaga ttcaggtttg ttgtgtcttt aaattgtata tgtgacttca catatctttt 2821
 tcagcgctta tacaaaacgg cactatagaa cctccatttt acagcaccat atgaagtggg 2881
 aaaattaggt gaaaattttc ctgaagcaac cttacatgc gcagcccttg ttggtttgtg 2941
 acttgtggcc tagctcatca gatgagccac gagaatcaga cctggatttt gatctggccc 3001
 25 tgttctgaca tgcaatgagg cattttagg atttagtaat attgctagtt caaagaatac 3061
 tagaaatatt agtaagaacc tattcaaaag tattcatgag tattttctgc atatgaatca 3121
 ggaattagaa tattttgaaa atgatgttaa taaaattttc ctctggaagg cttttataat 3181
 ttttattccc aatcattttt caaatttaga aagtttaatc tgtcacagga gaaaaaaat 3241
 taaaaatttt caaaaattta gaaaattttt acccgtaagt attacagttt cctaattatc 3301
 30 ctattttattt ccacttgcc tttgacctag attatttaga gtagggtttc tcagcctctg 3361
 cactaatgac attttgggcc gaataattct gttgtaggag gctgtcctgt gtgttttaga 3421
 ttgtttgaa ttatccctgg cctctcacac tagatgccag cagtatcctc ctccccagt 3481
 gtgacaacct aaaatgtctc cagacattgc caaatgttcc ctgtggggcg ggggcaacat 3541
 tgcctactgt taagaactac tgctctagac caaagaacac agcacagagg aaaggaaaaa 3601
 35 aaaatccagt taagagaatg ttaggtggag atgactatag tcatcaaact tttttcccca 3661
 tcaagtattt ccaagctaac atagtgacaa aataattcct gtactctact ggtaacgtta 3721
 atctagtgt cttcctttta ttcttccgtt ttgttttctt ttgcagaatt aagaatatgt 3781
 ttccttgttt atcagataca tctgtcttg atgaagcaag gaagatatac atgaaaattt 3841

taaaaataca tatcgctgac ttcattggaat ggacatcctg tataagcact gaaaaacaac 3901
 aacacaataa cactaaaatt ttaggcactc ttaaatgac tgcctctaaa agcgttggat 3961
 gtagcattat gcaattaggt ttttccttat ttgcttcatt gtactacctg tgtatatagt 4021
 ttttaccttt tatgtagcac ataaactttg gggaagggag ggcaggggtg ggctgaggaa 4081
 5 ctgacgtgga gcgggggtatg aagagcttgc tttgatttac agcaagtaga taaatatttg 4141
 acttgcataga agagaagcaa ttttggggaa gggtttgaat tgttttcttt aaagatgtaa 4201
 tgtccctttc agagacagct gatacttcat ttaaaaaaat cacaaaaatt tgaacactgg 4261
 ctaaagataa ttgctattta tttttacaag aagtttattc tcatttggga gatctggtga 4321
 tctcccaagc tatctaaagt ttgtagata gctgcatgtg gcttttttaa aaaagcaaca 4381
 10 gaaacctatc ctactgccc tccccagtct ctcttaaagt tggaatttac cagttaatta 4441
 ctgagcagaa tggtagcac tccaggtagt ttggggcaaa aatccgaggt gcttgggagt 4501
 tttgaatgtt aagaattgac catctgcttt tattaatttt gttgacaaaa ttttctcatt 4561
 ttcttttcac ttcgggctgt gtaaacaacag tcaaaataat tctaaatccc tcgatatttt 4621
 taaagatctg taagtaactt cacattaaaa aatgaaatat tttttaattt aaagcttact 4681
 15 ctgtccattt atccacagga aagtgttatt tttcaaggaa gggtcatgta gagaaaagca 4741
 cactttagg ataagtgaat tggatactac atctttaaac agtatttcat tgcctgtgta 4801
 tggaaaaacc atttgaagt tacctgtgta cataactctg taaaaacact gaaaaattat 4861
 actaacttat ttatgttaaa agattttttt taatctagac aatatacaag ccaaagtggc 4921
 atgttttgtg catttgtaaa tgctgtgttg ggtagaatag gttttcccct cttttgttaa 4981
 20 ataatatggc tatgcttaaa aggttgcata ctgagccaag tataattttt tgtaatgtgt 5041
 gaaaaagatg ccaattattg ttacacatta agtaatcaat aaagaaaact tccatagcta 5101
 ttcattgagt caaa

Non-limiting examples of siRNA targeting CDKN1B (p27^{kip1}) are described in, e.g.,
 25 Galardi et al., *J. Biol. Chem.* 282:23716-23724, 2007; Liang et al., *Nature Cell Biol.* 9:218-224,
 2007; Tamamori-Adachi et al., *J. Biol. Chem.* 279:50429-50436, 2004; Akashiba et al., *Cell.*
Mol. Life Sci. 63:2397-2404, 2006; and Lee et al., *J. Mol. Med.* 83(4):296-307, 2005. Non-
 limiting examples of shRNA targeting CDKN1B (p27^{kip1}) are described in, e.g., Lin et al.,
Nature 464:374-379, 2010.

30

Sex determining region Y-box 2 (SOX2)

The SOX2 gene encodes the sex determining region Y- box 2 protein. SOX2 is a
 transcription factor that binds the ATOH1 3'- enhancer and activates initial hair cell
 differentiation. Low SOX2 expression levels are required for proper hair cell maturation.
 35 Haploinsufficiency of SOX2 results in a few extra inner hair cells. SOX2 also increases the

susceptibility to induce transdifferentiation in the presence of other contributing components, e.g., beta-catenin (Kempfle et al. (2016) *Sci Rep* 6:23293; Atkinson et al. (2018) *J Clin Invest* 128(4):1641-1656).

The human SOX2 gene is located on chromosome 3q26. It contains 1 exon encompassing ~3 kilobases (kb) (NCBI Accession No. NG_009080.1). The full-length wildtype SOX2 protein expressed from the human SOX2 gene is 317 amino acids in length.

Methods of detecting mutations in a gene are well-known in the art. Non-limiting examples of such techniques include: real-time polymerase chain reaction (RT-PCR), PCR, sequencing, Southern blotting, and Northern blotting.

An exemplary human wildtype SOX2 protein is or includes the sequence of SEQ ID NO: 45. Non-limiting examples of a nucleic acid encoding a wildtype SOX2 protein is or includes SEQ ID NO: 46. As can be appreciated in the art, at least some or all of the codons in SEQ ID NO: 46 can be codon-optimized to allow for optimal expression in a non-human primate.

Human Full-length Wildtype SOX2 Protein (SEQ ID NO: 45)

MYNMMETELKPPGPQQTSGGGGNGNSTAAAAGGNQKNSPDRVKRPMNAFMVWSRGQRRKMAQENPKMHNSEISKRLGA
 EWKLLSETEKRPFI DEAKRLRALHMKEHPDYKYRPRRKTKTLMKKDKYTLPGGLLAPGGNSMASGVGVGAGLGAGVN
 QRMDSYAHMNGWSNGSYSMQDQLGY PQHPGLNAHGAAQM QPMHRYDVSALQYNSMTSSQTYMNGSPTYSMSYSQQG
 TPGMALGSMGSVVKSEASSSPVVTSSSHSRAPCQAGDLRDMISMYLPGA EVPEPAAPSRLHMSQHYQSGPVPGTAI
 NGTLPLSHM

Human Wildtype SOX2 cDNA (SEQ ID NO: 46)

atgtacaacatgatggagacggagctgaagccgcccggcccgagcaaaacttcggggggcgggcgggcgaactccac
 cgcgggcgggcgccggcggaaccagaaaaacagcccgaccgctcaagcgggcccatgaatgccttcattggtgtggt
 cccgcgggcagcgcgcaagatggcccaggagaaccccaagatgcacaactcggagatcagcaagcgctggcgcc
 gaggtgaaacttttgcggagacggagaagcgccgttcattcgacgaggctaagcggtgagcgctgcacatgaa
 ggagcaccggtattataaataaccggccccggcggaacccaagacgctcatgaagaaggataagtacacgctgcccc
 gcgggctgctggcccccgggcaatagcatggcgagcggggtcggggtggcgccggcctggcgcgggcggtgaac
 cagcgcatggacagttacgcgcacatgaacggctggagcaacggcagctacagcatgatgcaggaccagctgggcta
 cccgcagcaccgggctcaatgcgcacggcgagcgagatgcagcccatgcaccgctacgacgtgagcgccctgc
 agtacaactccatgaccagctcgacagctacatgaacggctcgccacctacagcatgtcctactcgacagggc
 acccctggcatggctcttggtccatgggttcgggtggtcaagtccgaggccagctccagccccctgtggttacctc
 ttctccactccaggcgccctgccaggccggggacctccgggacatgatcagcatgtatctccccggcgccgagg

tgccggaacccgccgccccagcagacttcacatgtcccagcactaccagagcggccccgtgcccggcacggccatt
aacggcacactgcccctctcacacatgtga

A non-limiting example of a human wildtype SOX2 genomic DNA sequence is SEQ ID

5 NO: 47. The exon in SEQ ID NO: 47 is nucleotide positions 1-2520 (exon 1).

Human Wildtype SOX2 Gene (SEQ ID NO: 47)

1 ggatggttgt ctattaactt gttcaaaaaa gtatcaggag ttgtcaaggc agagaagaga
10 61 gtgtttgcaa aagggggaaa gtagtttgct gcctctttaa gactaggact gagagaaaga
121 agaggagaga gaaagaaagg gagagaagtt tgagccccag gcttaagcct ttccaaaaaa
181 taataataac aatcatcggc ggccggcagga tcggccagag gaggaggga ggcgtttttt
241 tgatcctgat tccagtttgc ctctctcttt tttcccca aattattctt cgctgattt
301 tcctcgcgga gccctgcgct cccgacaccc ccgcccgcct cccctcctcc tctccccccg
15 361 ccgcggggcc ccccaaagtc ccggccgggc cgagggtcgg cggccgcccg cgggcccggc
421 ccgcgcacag cgcccgcag tacaacatga tggagacgga gctgaagccg ccgggcccgc
481 agcaaacttc ggggggccc ggccgcaact ccaccgccc ggccgcccgc ggcaaccaga
541 aaaacagccc ggaccgcgtc aagcggcca tgaatgcctt catggtgtgg tcccgccggc
601 agcggcgcaa gatggcccag gagaaccca agatgcacaa ctccgagatc agcaagcgcc
20 661 tgggcgcccga gtgaaactt ttgtcggaga cggagaagcg gccgttcac gacgaggcta
721 agcggctgcg agcgtgcac atgaaggagc acccgatta taaataccg ccccgccgga
781 aaaccaagac gtcataag aaggataagt acacgctgcc cggcgggctg ctggcccccg
841 gcggcaatag catggcgagc ggggtcggg tggcgccgg cctgggcgcg ggcgtgaacc
901 agcgcagtgga cagttacgc cacaatgaac gctggagcaa cggcagctac agcatgatgc
25 961 aggaccagct gggctacccg cagcaccg gcctcaatgc gcacggcgca gcgcagatgc
1021 agcccatgca ccgctacgac gtgagcgccc tgcatgaca ctccatgacc agtcgcgaga
1081 cctacatgaa cggctcgccc acctacagca tgcctactc gcagcagggc acccctggca
1141 tggctcttgg ctccatgggt tcggtgtgca agtccgaggc cagctccagc cccctgtgg
1201 ttacctcttc ctccactcc agggcgccct gccaggccgg ggacctccg gacatgatca
30 1261 gcatgtatct ccccgccgccc gaggtgccg aaccgcccgc cccagcaga cttcacatgt
1321 ccagcacta ccagagcggc ccggtgccc gcacggccat taacggcaca ctgcccctct
1381 cacacatgtg agggccggac agcgaactgg aggggggaga aattttcaaa gaaaaacgag
1441 ggaaatggga ggggtgcaaa agaggagagt aagaaacagc atggagaaaa cccggtacgc
1501 tcaaaaagaa aaaggaaaaa aaaaaatccc atcaccaca gcaaatgaca gctgcaaaag
35 1561 agaacaccaa tcccatccac actcacgcaa aaaccgcat gccgacaaga aaacttttat
1621 gagagagatc ctggacttct ttttggggga ctatttttgt acagagaaaa cctggggagg
1681 gtggggaggg cgggggaatg gaccttgat agatctggag gaaagaaagc tacgaaaaac
1741 tttttaaaag ttctagtgg acggtaggag ctttgacgga agtttgcaa agtctttacc

1801 aataatat ttt agagctagtc tccaagcgac gaaaaaaatg ttttaatat tgcaagcaac
 1861 ttttgtacag tatttatcga gataaacatg gcaatcaaaa tgtccattgt ttataagctg
 1921 agaatttgcc aatat ttttc aaggagaggc ttcttgctga attttgattc tgcagctgaa
 1981 atttaggaca gttgcaaacg tgaaaagaag aaaattattc aaatttggaac attttaattg
 5 2041 tttaaaaatt gtacaaaagg aaaaaattag aataagtact ggcgaacat ctctgtggtc
 2101 ttgttttaaa agggcaaaag ttttagactg tactaaattt tataacttac tgttaaaagc
 2161 aaaaatggcc atgcagggtg acaccgttg taatttataa tagcttttgc tcgatcccaa
 2221 ctttccattt tgttcagata aaaaaacca tgaaattact gtgtttgaaa tattttctta
 2281 tggtttgtaa tattttctga aatttattgt gatattttta ggttttcccc cttttatttt
 10 2341 ccgtagttgt attttaaaag attcggctct gtattatttg aatcagtctg ccgagaatcc
 2401 atgtatataat ttgaactaat atcatcctta taacagggtac attttcaact taagttttta
 2461 ctccattatg cacagtttga gataaataaa tttttgaaat atggacactg aaa

15 Non-limiting examples of siRNA targeting SOX2 are described in, e.g., Kondo et al.,
Genes Develop. 18:2963-2972, 2004; Tani et al., *J. Cancer Res. Clin. Oncol.* 133(4):263-269,
 2007; Chen et al., *J. Biol. Chem.* 283:17969-17978, 2008; and Card et al., *Mol. Cell. Biol.*
 28(20):6426-6438, 2008. Non-limiting examples of shRNA targeting SOX2 are described in,
 e.g., Rudin et al., *Nature Genetics* 44:1111-1116, 2012; Basu-Roy et al., *Oncogene* 31:2270-
 20 2282, 2012; and Marques-Torrejon et al., *Cell Stem Cell* 12(1):88-100, 2013.

Vectors

Some of the compositions provided herein can include at least two (e.g., two, three, four,
 five, or six) AAV vectors, where: each of the at least two different AAV vectors includes a
 25 coding sequence that encodes a different portion of a hair cell differentiation protein, each of the
 encoded portions being at least 30 amino acids (e.g., about 30 amino acids to about 800 amino
 acids, about 30 amino acids to about 780 amino acids, about 30 amino acids to about 760 amino
 acids, about 30 amino acids to about 750 amino acids, about 30 amino acids to about 740 amino
 acids, about 30 amino acids to about 720 amino acids, about 30 amino acids to about 710 amino
 30 acids, about 30 amino acids to about 700 amino acids, about 30 amino acids to about 690 amino
 acids, about 30 amino acids to about 680 amino acids, about 30 amino acids to about 670 amino
 acids, about 30 amino acids to about 660 amino acids, about 30 amino acids to about 650 amino
 acids, about 30 amino acids to about 640 amino acids, about 30 amino acids to about 630 amino
 acids, about 30 amino acids to about 620 amino acids, about 30 amino acids to about 610 amino

127

5
10
15
20
25
30

5
10
15
20
25
30

[illegible]

131

132

133

5
10
15
20
25
30

135

[illegible]

about 710 amino acids, about 650 amino acids to about 700 amino acids, about 700 amino acids to about 800 amino acids, about 700 amino acids to about 790 amino acids, about 700 amino acids to about 780 amino acids, about 700 amino acids to about 770 amino acids, about 700 amino acids to about 760 amino acids, about 700 amino acids to about 750 amino acids, or about 750 amino acids to about 800 amino acids), where the amino acid sequence of each of the encoded portions may optionally partially overlap with the amino acid sequence of a different one of the encoded portions; no single vector of the at least two different vectors encodes the hair cell differentiation protein (e.g., a full-length hair cell differentiation protein (e.g., a full-length wildtype hair cell differentiation protein)); and, when introduced into a primate cell (e.g., a hair cell or a supporting cell of the inner ear), the at least two different AAV vectors undergo homologous recombination with each other, where the recombined nucleic acid encodes a hair cell differentiation protein (e.g., a full-length hair cell differentiation protein).

In some embodiments of the compositions that include at least two AAV vectors, at least one of the coding sequences includes a nucleotide sequence spanning two neighboring exons of hair cell differentiation genomic DNA, and lacks the intronic sequence that naturally occurs between the two neighboring exons.

In some embodiments of the compositions that include at least two AAV vectors, the amino acid sequence of none of the encoded portions overlaps even in part with the amino acid sequence of a different one of the encoded portions. In some embodiments of the compositions that include at least two AAV vectors, the amino acid sequence of one or more of the encoded portions partially overlaps with the amino acid sequence of a different one of the encoded portions. In some embodiments of the compositions that include at least AAV vectors, the amino acid sequence of each of the encoded portions partially overlaps with the amino acid sequence of a different one of the encoded portions.

In some embodiments of the compositions that include at least two AAV vectors, the overlapping amino acid sequence is between about 30 amino acid residues to about 800 amino acids (e.g., or any of the subranges of this range described herein) in length.

In some examples, the compositions include two different AAV vectors, each of which comprises a different segment of an intron, where the intron includes the nucleotide sequence of an intron that is present in a hair cell differentiation genomic DNA, and where the two different segments overlap in sequence by at least 100 nucleotides (e.g., about 100 nucleotides to about

3,000 nucleotides, about 100 nucleotides to about 2,500 nucleotides, about 100 nucleotides to about 2,000 nucleotides, about 100 nucleotides to about 1,500 nucleotides, about 100 nucleotides to about 1,000 nucleotides, about 100 nucleotides to about 800 nucleotides, about 100 nucleotides to about 600 nucleotides, about 100 nucleotides to about 400 nucleotides, about 100 nucleotides to about 200 nucleotides, about 200 nucleotides to about 3,000 nucleotides, about 200 nucleotides to about 2,500 nucleotides, about 200 nucleotides to about 2,000 nucleotides, about 200 nucleotides to about 1,500 nucleotides, about 200 nucleotides to about 1,000 nucleotides, about 200 nucleotides to about 800 nucleotides, about 200 nucleotides to about 600 nucleotides, about 200 nucleotides to about 400 nucleotides about 400 nucleotides to about 3,000 nucleotides, about 400 nucleotides to about 2,500 nucleotides, about 400 nucleotides to about 2,000 nucleotides, about 400 nucleotides to about 1,500 nucleotides, about 400 nucleotides to about 1,000 nucleotides, about 400 nucleotides to about 800 nucleotides, about 400 nucleotides to about 600 nucleotides, about 600 nucleotides to about 3,000 nucleotides, about 600 nucleotides to about 2,500 nucleotides, about 600 nucleotides to about 2,000 nucleotides, about 600 nucleotides to about 1,500 nucleotides, about 600 nucleotides to about 1,000 nucleotides, about 600 nucleotides to about 800 nucleotides, about 800 nucleotides to about 3,000 nucleotides, about 800 nucleotides to about 2,500 nucleotides, about 800 nucleotides to about 2,000 nucleotides, about 800 nucleotides to about 1,500 nucleotides, about 800 nucleotides to about 1,000 nucleotides, about 1,000 nucleotides to about 3,000 nucleotides, about 1,000 nucleotides to about 2,500 nucleotides, about 1,000 nucleotides to about 2,000 nucleotides, about 1,000 nucleotides to about 1,500 nucleotides, about 1,500 nucleotides to about 3,000 nucleotides, about 1,500 nucleotides to about 2,500 nucleotides, about 1,500 nucleotides to about 2,000 nucleotides, about 2,000 nucleotides to about 3,000 nucleotides, about 2,000 nucleotides to about 2,500 nucleotides, or about 2,500 nucleotides to about 3,000 nucleotides), in length.

The overlapping nucleotide sequence in any two of the different AAV vectors can include part or all of one or more exons of a hair cell differentiation gene.

In some embodiments, the number of different AAV vectors in the composition is two, three, four, or five. In compositions where the number of different AAV vectors in the composition is two, the first of the two different vectors can include a coding sequence that encodes an N-terminal portion of the hair cell differentiation protein. In some embodiments, the N-terminal portion can include a portion having about 30 amino acids to about 800 amino acids

(or any of the subranges of this range described herein). In some examples, the N-terminal portion encoded by one of the two vectors can include a portion comprising amino acid position 1 to about amino acid position 800, about amino acid position 790, about amino acid position 780, about amino acid position 770, about amino acid position 760, about amino acid position 750, about amino acid position 740, about amino acid position 730, about amino acid position 720, about amino acid position 710, about amino acid position 700, about amino acid position 690, about amino acid position 680, about amino acid position 670, about amino acid position 660, about amino acid position 650, about amino acid position 640, about amino acid position 630, about amino acid position 620, about amino acid position 610, about amino acid position 600, about amino acid position 590, about amino acid position 580, about amino acid position 570, about amino acid position 560, about amino acid position 550, about amino acid position 540, about amino acid position 530, about amino acid position 520, about amino acid position 510, about amino acid position 500, about amino acid position 490, about amino acid position 480, about amino acid position 470, about amino acid position 460, about amino acid position 450, about amino acid position 440, about amino acid position 430, about amino acid position 420, about amino acid position 410, about amino acid position 400, about amino acid position 390, about amino acid position 380, about amino acid position 370, about amino acid position 360, about amino acid position 350, about amino acid position 340, about amino acid position 330, about amino acid position 320, about amino acid position 310, about amino acid position 300, about amino acid position 290, about amino acid position 280, about amino acid position 270, about amino acid position 260, about amino acid position 250, about amino acid position 240, about amino acid position 230, about amino acid position 220, about amino acid position 210, about amino acid position 200, about amino acid position 190, about amino acid position 180, about amino acid position 170, about amino acid position 160, about amino acid position 150, about amino acid position 140, about amino acid position 130, about amino acid position 120, about amino acid position 110, about amino acid position 100, about amino acid position 90, about amino acid position 80, about amino acid position 70, about amino acid position 60, about amino acid position 50, or about amino acid position 40 of a wildtype hair cell differentiation protein.

In compositions where the number of different AAV vectors in the composition is two, the second of the two different vectors can include a coding sequence that encodes a C-terminal

portion of the hair cell differentiation protein. In some embodiments, the C-terminal portion can include a portion having about 30 amino acids to about 800 amino acids (or any of the subranges of this range described herein).

As used herein, the term “vector” means a composition including a polynucleotide capable of carrying at least one exogenous nucleic acid fragment, e.g., an adeno-associated virus (AAV) vector. A vector can, e.g., include sufficient cis-acting elements for expression; other elements for expression can be supplied by the host primate cell or in an *in vitro* expression system. The term “vector” includes any genetic element (e.g., a plasmid, a transposon, a cosmid, an artificial chromosome, or a viral vector, etc.) that is capable of replicating when associated with the proper control elements.

“Recombinant AAV vectors” or “rAAVs” are typically composed of, at a minimum, a transgene or a portion thereof and a regulatory sequence, and optionally 5' and 3' AAV inverted terminal repeats (ITRs). Such a recombinant AAV vector is packaged into a capsid and delivered to a selected target cell (e.g., an inner or outer hair cell, or a supporting cell of the inner ear).

The AAV sequences of the vector typically comprise the cis-acting 5' and 3' ITR sequences (See, e.g., B. J. Carter, in "Handbook of Parvoviruses", ed., P. Tijsser, CRC Press, pp. 155-168, 1990). Typical AAV ITR sequences are about 145 nucleotides in length. In some embodiments, at least 75% of a typical ITR sequence (e.g., at least 80%, at least 85%, at least 90%, or at least 95%) is incorporated into the AAV vector. The ability to modify these ITR sequences is within the skill of the art. (See, e.g., texts such as Sambrook et al., "Molecular Cloning. A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory, New York, 1989; and K. Fisher et al., *J Virol.* 70:520-532, 1996). In some embodiments, any of the coding sequences described herein are flanked by 5' and 3' AAV ITR sequences in the AAV vectors. The AAV ITR sequences may be obtained from any known AAV, including presently identified AAV types. In some examples of any of the vectors described herein, the vector includes a 5' ITR sequence

CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCCGGGCGTCGGGCGAC
CTTTGGTTCGCCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACT
CCATCACTAGGGGTTTCCT (SEQ ID NO: 51) and/or a 3' ITR sequence
AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTG

AGGCCGGGCGACCAAAGGTCGCCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTG
AGCGAGCGAGCGCGCAGCTGCCTGCAGG (SEQ ID NO: 57).

AAV vectors as described herein may include any of the regulatory elements described herein (e.g., one or more of a promoter, a polyA sequence, and an IRES).

5 In some embodiments, the AAV vector 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 AAV2.7m8 vector, an AAV8BP2 vector, and an AAV293 vector. Additional exemplary AAV vectors that can be used herein are known in the art. See, e.g., Kanaan et al., *Mol. Ther. Nucleic Acids* 8:184-197, 2017; 10 Li et al., *Mol. Ther.* 16(7): 1252-1260; Adachi et al., *Nat. Commun.* 5: 3075, 2014; Isgrig et al., *Nat. Commun.* 10(1): 427, 2019; and Gao et al., *J. Virol.* 78(12): 6381-6388.

In some embodiments, an AAV vector provided herein includes or consists of a sequence that is at least 80% identical (e.g., at least 82%, at least 84%, at least 85%, at least 86%, at least 88%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 98%, at least 15 99%, or 100% identical) to SEQ ID NO: 50, 58, 60, 64, 66, 68, 78, 79, 81, 82, 83 or 94.

The AAV vectors provided herein can be of different sizes. In some embodiments, the AAV vector(s) can include a total number of nucleotides of up to 5 kb. In some embodiments, the AAV vector(s) can include a total number of nucleotides in the range of about 1 kb to about 2 kb, about 1 kb to about 3 kb, about 1 kb to about 4 kb, about 1 kb to about 5 kb, about 2 kb to 20 about 3 kb, about 2 kb to about 4 kb, about 2 kb to about 5kb, about 3 kb to about 4 kb, about 3 kb to about 5 kb, or about 4 kb to about 5 kb.

In some embodiments of any of the compositions, kits, and methods provided herein, the at least two different AAV vectors can be substantially the same type of vector and may differ in size. In some embodiments, the at least two different AAV vectors can be different types of 25 AAV vector, and may have substantially the same size or have different sizes.

In some embodiments, any of the at least two AAV vectors can have a total number of nucleotides in the range of about 500 nucleotides to about 10,000 nucleotides, about 500 nucleotides to about 9,500 nucleotides, about 500 nucleotides to about 9,000 nucleotides, about 500 nucleotides to about 8,500 nucleotides, about 500 nucleotides to about 8,000 nucleotides, 30 about 500 nucleotides to about 7,800 nucleotides, about 500 nucleotides to about 7,600 nucleotides, about 500 nucleotides to about 7,400 nucleotides, about 500 nucleotides to about

[illegible]

5
10
15
20
25
30

5
10
15
20
25
30

5
10
15
20
25
30

[illegible]

2,400 nucleotides, about 2,000 nucleotides to about 2,200 nucleotides, about 2,200 nucleotides to about 10,000 nucleotides, about 9,500 nucleotides, about 9,000 nucleotides, about 8,500 nucleotides, about 8,000 nucleotides, about 7,800 nucleotides, about 7,600 nucleotides, about 7,400 nucleotides, about 7,200 nucleotides, about 7,000 nucleotides, about 6,800 nucleotides, about 6,600 nucleotides, about 6,400 nucleotides, about 6,200 nucleotides, about 6,000 nucleotides, about 5,800 nucleotides, about 5,600 nucleotides, about 5,400 nucleotides, about 5,200 nucleotides, about 5,000 nucleotides, about 4,800 nucleotides, about 4,600 nucleotides, about 4,400 nucleotides, about 4,200 nucleotides, about 4,000 nucleotides, about 3,800 nucleotides, about 3,600 nucleotides, about 3,400 nucleotides, about 3,200 nucleotides, about 3,000 nucleotides, about 2,800 nucleotides, about 2,600 nucleotides, about 2,400 nucleotides, about 2,400 nucleotides to about 10,000 nucleotides, about 2,400 nucleotides to about 9,500 nucleotides, about 2,400 nucleotides to about 9,000 nucleotides, about 2,400 nucleotides to about 8,500 nucleotides, about 2,400 nucleotides to about 8,000 nucleotides, about 2,400 nucleotides to about 7,800 nucleotides, about 2,400 nucleotides to about 7,600 nucleotides, about 2,400 nucleotides to about 7,400 nucleotides, about 2,400 nucleotides to about 7,200 nucleotides, about 2,400 nucleotides to about 7,000 nucleotides, about 2,400 nucleotides to about 6,800 nucleotides, about 2,400 nucleotides to about 6,600 nucleotides, about 2,400 nucleotides to about 6,400 nucleotides, about 2,400 nucleotides to about 6,200 nucleotides, about 2,400 nucleotides to about 6,000 nucleotides, about 2,400 nucleotides to about 5,800 nucleotides, about 2,400 nucleotides to about 5,600 nucleotides, about 2,400 nucleotides to about 5,400 nucleotides, about 2,400 nucleotides to about 5,200 nucleotides, about 2,400 nucleotides to about 5,000 nucleotides, about 2,400 nucleotides to about 4,800 nucleotides, about 2,400 nucleotides to about 4,600 nucleotides, about 2,400 nucleotides to about 4,400 nucleotides, about 2,400 nucleotides to about 4,200 nucleotides, about 2,400 nucleotides to about 4,000 nucleotides, about 2,400 nucleotides to about 3,800 nucleotides, about 2,400 nucleotides to about 3,600 nucleotides, about 2,400 nucleotides to about 3,400 nucleotides, about 2,400 nucleotides to about 3,200 nucleotides, about 2,400 nucleotides to about 3,000 nucleotides, about 2,400 nucleotides to about 2,800 nucleotides, about 2,400 nucleotides to about 2,600 nucleotides, about 2,600 nucleotides to about 10,000 nucleotides, about 2,600 nucleotides to about 9,500 nucleotides, about 2,600 nucleotides to about 9,000 nucleotides, about 2,600 nucleotides to about 8,500 nucleotides, about 2,600 nucleotides to about 8,000 nucleotides, about 2,600 nucleotides to about 7,800 nucleotides, about 2,600

[illegible]

5
10
15
20
25
30

[illegible]

5
10
15
20
25
30

[illegible]

[illegible]

[illegible]

5
10
15
20
25
30

nucleotides, about 7,200 nucleotides to about 9,500 nucleotides, about 7,200 nucleotides to about 9,000 nucleotides, about 7,200 nucleotides to about 8,500 nucleotides, about 7,200 nucleotides to about 8,000 nucleotides, about 7,200 nucleotides to about 7,800 nucleotides, about 7,200 nucleotides to about 7,600 nucleotides, about 7,200 nucleotides to about 7,400 nucleotides, about 7,400 nucleotides to about 10,000 nucleotides, about 7,400 nucleotides to about 9,500 nucleotides, about 7,400 nucleotides to about 9,000 nucleotides, about 7,400 nucleotides to about 8,500 nucleotides, about 7,400 nucleotides to about 8,000 nucleotides, about 7,400 nucleotides to about 7,800 nucleotides, about 7,400 nucleotides to about 7,600 nucleotides, about 7,600 nucleotides to about 10,000 nucleotides, about 7,600 nucleotides to about 9,500 nucleotides, about 7,600 nucleotides to about 9,000 nucleotides, about 7,600 nucleotides to about 8,500 nucleotides, about 7,600 nucleotides to about 8,000 nucleotides, about 7,600 nucleotides to about 7,800 nucleotides, about 7,800 nucleotides to about 10,000 nucleotides, about 7,800 nucleotides to about 9,500 nucleotides, about 7,800 nucleotides to about 9,000 nucleotides, about 7,800 nucleotides to about 8,500 nucleotides, about 7,800 nucleotides to about 8,000 nucleotides, about 8,000 nucleotides to about 10,000 nucleotides, about 8,000 nucleotides to about 9,500 nucleotides, about 8,000 nucleotides to about 9,000 nucleotides, about 8,000 nucleotides to about 8,500 nucleotides, about 8,500 nucleotides to about 10,000 nucleotides, about 8,500 nucleotides to about 9,500 nucleotides, about 8,500 nucleotides to about 9,000 nucleotides, about 9,000 nucleotides to about 10,000 nucleotides, about 9,000 nucleotides to about 9,500 nucleotides, or about 9,500 nucleotides to about 10,000 nucleotides (inclusive).

Figures 4A-D, Figures 7A-B and Figures 11A-B provide schematic representations of exemplary nucleic acid vectors that can be included in any of the compositions and methods described herein.

In some embodiments of any of the compositions described herein, the vector comprises or consists of pITR-CMV-mScarlet (SEQ ID NO: 50). In some embodiments of any of the compositions described herein, the vector comprises a sequence that has at least 75% (e.g., at least 80%, at least 82%, at least 84%, at least 85%, at least 86%, at least 88%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 98%, at least 99%) sequence identity to SEQ ID NO: 50.

pITR-CMV-mScarlet (SEQ ID NO: 50)

cctgcaggcagctgcgcgctcgctcgctcactgaggccgcccggcgctcgggacaccttggctgcccggcctcagt
 gagcgagcgagcgcgagagagggagtgcccaactccatcactaggggttcctctagatcccatatatggagttccg
 cgttacataacttacggtaaatggcccgctggctgaccgcccacgacccccgcccattgacgtcaataatgacgt
 5 atgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggagtatttacggtaaaactgccacttg
 gcagtacatcaagtgtatcatatgccaaagtacgccccctattgacgtcaatgacggtaaatggcccgctggcatta
 tggccagttacatgaccttatgggactttcctacttggcagttacatctacgtatttagtcatcgctattaccatgggtga
 tgccgttttggcagttacatcaatggcgctggatagcgggttgactcacggggatttccaagtctccacccccattgac
 gtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaaactccgccccattgacgca
 10 aatgggcggttaggcgtgtacgggtgggaggtctatataagcagagctcgtttagtgaaccgtcagatcgctggagac
 gcatgcctaagaagaagcggaaagtccggctccggcgctgagcaagggcgaggcagtgatcaaggagttcatgcggttc
 aaggtgcacatggagggctccatgaacggccacgagttcgagatcgaggcgaggcgaggcgccccctacgaggg
 caccagacccgccaagctgaaggtgaccaaggtggccccctgcccttctcctgggacatcctgtccccctcagttca
 tgtacggctccagggccttcatcaagcaccgcccgcacatccccgactactataagcagtccttccccgagggcttc
 15 aagtgggagcgctgatgaacttcgaggacggcgcgccgtgaccgtgacccaggacacctccctggaggacggcac
 cctgatctacaaggtgaagctccgcggcaccaacttccctcctgacggccccgtaatgcagaagaagacaatgggct
 gggaagcgctccaccgagcggttgtaccccgaggacggcgctgtaagggcgacattaagatggccctgcgcctgaag
 gacggcgcccgctacctggcggaacttcaagaccacctacaaggccaagaagcccgctgcagatgcccggcgccctaca
 cgctcgaccgcaagttggacatcacctcccacaacgaggactacaccgtggtggaacagtacgaacgctccgagggcc
 20 gccactccaccggcgcatggacgagctgtacaagtaagctgatcagcctcgactgtgccttctagttgccagccat
 ctgttgtttgccccctcccccgctgccttccctgaccctggaaggtgccactcccactgtcctttcctaataaaatgag
 gaaattgcatcgcatgtctgagtaggtgtcattctattctgggggtgggggtggggcaggacagcaagggggagga
 ttgggaagacaatagcaggcatgctggggatgcggtgggctctatggaggaacccctagtgatggagttggccactc
 cctctctgcgcgctcgctcgctcactgaggccggggcgaccaaaggtcgcccgacgcccgggctttgcccggcgggcc
 25 tcagtgcgagcgagcgcgagctgcctgcagg

5' ITR cDNA sequence (SEQ ID NO: 51)

cctgcaggcagctgcgcgctcgctcgctcactgaggccgcccggcgctcgggacaccttggctgcccggcctcagt
 gagcgagcgagcgcgagagagggagtgcccaactccatcactaggggttcct
 30

CMV_enhancer cDNA sequence (SEQ ID NO: 52)

ctagatcccatatatggagttccgcgttacataacttacggtaaatggcccgctggctgaccgcccacgaccccc
 gccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggag
 tatttacggtaaaactgccacttggcagttacatcaagtgtatcatatgccaaagtacgccccctattgacgtcaatga
 35 cggtaaatggcccgctggcattatgccagttacatgaccttatgggactttcctacttggcagttacatctacgtat
 tagtcatcgctattaccatg

CMV_promoter cDNA sequence (SEQ ID NO: 53)

gtgatgCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTC AAGTCTCCACCCCAT
 TGACGTCAATGGGAGTTTGTGTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACCTCGCCCCATTGA
 CGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGG
 5 agacgc

SV40-NLS cDNA sequence (SEQ ID NO: 54)

atgcctaagaagaagcggaaagtcggctccggc

10 mScarlet cDNA sequence (SEQ ID NO: 55)

gtgagcaagggcgaggcagtgatcaaggagttcatgCGGTTCAAGGTGCACATGGAGGGCTCCATGAACGGCCA
 GTTCGAGATCGAGGGCGAGGGCGAGGGCGGCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTG
 GCCCCCTGCCCTTCTCCTGGGACATCCTGTCCCTCAGTTTCATGTACGGCTCCAGGGCCTTCATCAAGCACCCCGCC
 GACATCCCCGACTACTATAAGCAGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGTGAACCTCGAGGACGGCGG
 15 CGCCGTGACCGTGACCCAGGACACCTCCCTGGAGGACGGCACCCCTGATCTACAAGGTGAAGCTCCGCGGCACCAACT
 TCCCTCCTGACGGCCCCGTAATGCAGAAGAAGACAATGGGCTGGGAAGCGTCCACCGAGCGGTTGTACCCCGAGGAC
 GGCCTGCTGAAGGGCGACATTAAGATGGCCCTGCGCCTGAAGGACGGCGGCGCTACCTGGCGGACTTCAAGACCAC
 CTACAAGGCCAAGAAGCCCGTGCAGATGCCCGGCGCCTACAACGTCGACCGCAAGTTGGACATCACCTCCCACAACG
 AGGACTACACCGTGGTGGAACAGTACGAACGCTCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAG
 20 taa

BGHpA cDNA sequence (SEQ ID NO: 56)

gctgatcagcctcgactgtgccttctagttgccagccatctgttgTTTGCCCTCCCCCGTGCCTTCTTGACCCGTG
 GAAGGTGCCACTCCCCTGTCTTTCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTAT
 25 TCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGG
 GCTCTATGG

3' ITR cDNA sequence (SEQ ID NO: 57)

aggaacccttagtgatggagttggccactccctctctgcgcgctcgctcgctcactgaggccgggCGACCAAAGGTC
 30 gcccgacgcccgggctttgccgggCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG

In some embodiments of any of the compositions described herein, the vector comprises or consists of pITR-CMV-mScarlet-DD (SEQ ID NO: 58). In some embodiments of any of the compositions described herein, the vector comprises a sequence that has at least 75% (e.g., at
 35 least 80%, at least 82%, at least 84%, at least 85%, at least 86%, at least 88%, at least 90%, at

least 92%, at least 94%, at least 95%, at least 96%, at least 98%, at least 99%) sequence identity to SEQ ID NO: 58.

pITR-CMV-mScarlet-DD (SEQ ID NO: 58)

5 cctgcaggcagctgcgcgctcgctcgctcactgaggccgcccggcgctcgggcgacctttggctcgcccggcctcagt
gagcgagcgagcgcgagagagggagtgcccaactccatcactaggggttcctctagatcccatatatggagttccg
cgttacataacttacggtaaatggcccgctggctgaccgcccacgacccccgccattgacgtcaataatgacgt
atgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggagtatctacggtaaactgccacttg
gcagtacatcaagtgtatcatatgccaaagtacgccccctattgacgtcaatgacggtaaattggcccgctggcatta
10 tggccagtagacgttatgggactttcctacttggcagtagatctacgtatttagtcatcgctattaccatgggtga
tgccgttttggcagtagatcaatggcggtgtagcggttgactcacggggatttccaagtctccacccccattgac
gtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgca
aatggcggttaggcgtgtacgggtgggaggtctatataagcagagctcgtttagtgaaccgctcagatcgccctggagac
gcattgcctaagaagaagcggaaagtccggtccggcggtgagcaaggcgaggcagtgatcaaggagttcatgcggttc
15 aaggtgcacatggagggctccatgaacggccacgagttcgagatcgagggcgagggcgagggcgccctacgaggg
caccagaccgccaagctgaaggtgaccaagggtggccccctgcccttctcctgggacatcctgtcccctcagttca
tgtacggctccagggccttcatcaagcaccgcccgcacatccccgactactataagcagtccttccccgagggcttc
aagtgggagcgcggtgatgaacttcgaggacggcgggcgccgtgaccgtgacccaggacacctccctggaggacggcac
cctgatctacaaggtgaagctccgcgccaccaacttccctcctgacggccccgtaatgcagaagaagacaatgggct
20 gggaagcggtccaccgagcggttgtagcccgaggacggcggtgctgaaggcgacattaagatggccctgcgcctgaag
gacggcgcccgctacctggcggaacttcaagaccacctacaaggccaagaagcccgtagcagatgcccggcgcctacaa
cgctcgaccgcaagttggacatcacctcccacaacgaggactacaccgtggtggaacagtagcaacgctccgagggcc
gccactccaccggcgcatggacgagctgtacaagggtaccatcagtcctgattgcggcggttagcggtagattacgtt
atcggcgatggaaaacgcatgccgtggaacctgcctgccgatctcgccctggtttaaacgcaacaccttaataaaacc
25 cgtgattatgggcccgcatacctgggaatcaatcggtcggtccgttgccaggacgcaaaaatattatcctcagcagtc
aaccgagtagcgacgatcgcgtaacgtgggtgaagtcggtggatgaagccatcgcgcggtgtggtgacgtaccagaa
atcatggtgattggcgcggtcgcggtattgaacagttcttgccaaaagcgcaaaaactgtatctgacgcataatcga
cgcaagaagtgaaggcgacacccatttcccggattacgagccggatgactgggaatcggtattcagcgaattccacg
atgctgatgcgcagaactctcacagctattgctttgagattctggagcggcgataagctgatcagcctcgactgtgc
30 cttctagttgccagccatctgttgtttgcccctcccccgtagccttcttgaccctggaagggtgccactcccactgtc
ctttcctaataaaatgaggaaattgcatcgcatgtgtctgagtaggtgtcattctattctgggggggtgggggtggggca
ggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggaggaacccttag
tgatggagttggccactccctctctgcgcgctcgctcgctcactgaggccggcgaccaaaggtcgcccgacgcccg
ggctttgcccggcgccctcagtgagcgagcgagcgcgagctgcctgcagg

35

DHFR-DD cDNA sequence (SEQ ID NO: 59)

ggtaccatcagtcctgattgcggttagcggtagattacgttatcggcatggaaaacgccatgccgtggaacctgcc
 tgccgatctcgctgggtttaaacgcaacaccttaataaaccctgattatgggcccgcataacctgggaatcaatcg
 gtcgtccgttgccaggacgcaaaaatattatcctcagcagtcacccgagtagcgacgatcgcgtaacgtgggtgaag
 5 tccgtggatgaagccatcgcggtgtggtgacgtaccagaaatcatggtgattggcggtcggttattgaaca
 gttcttgccaaaagcgcaaaaactgtatctgacgcataatcgacgcagaagtgaaggcgacacccatttcccgatt
 acgagccgatgactgggaatcggtattcagcgaattccacgatgctgatgcgagaactctcacagctattgcttt
 gagattctggagcggcgataa

- 10 In some embodiments of any of the compositions described herein, the vector comprises
 or consists of pITR-CMV-hPou4f3-T2A-mScarlet-DD (SEQ ID NO: 60). In some embodiments
 of any of the compositions described herein, the vector comprises a sequence that has at least
 75% (e.g., at least 80%, at least 82%, at least 84%, at least 85%, at least 86%, at least 88%, at
 least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 98%, at least 99%)
 15 sequence identity to SEQ ID NO: 60.

pITR-CMV-hPou4f3-T2A-mScarlet-DD 60

- cctgcaggcagctgcgcgctcgctcgctcactgaggccgcccggcgctcgggcgacctttgggtcgcccggcctcagt
 gagcgagcgagcgcgagagagggagtgccaaactccatcactaggggttcctctagatcccatatatggagttccg
 20 cgttacataacttacggtaaatggcccgctggctgaccgcccacgacccccgccattgacgtcaataatgacgt
 atgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggagtatttacggtaaactgcccacttg
 gcagtacatcaagtgtatcatatgccaaagtacgccccctattgacgtcaatgacggtaaatggcccgctggcatta
 tgcccagtagacgttatgggactttcctacttggcagtacatctacgtatttagtcacgtattaccatgggtga
 tgccgttttgccagtagcatcaatggcggtggatagcggtttgactcacggggatttccaagtctccacccattgac
 25 gtcaatgggagtttgttttgccacaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgca
 aatggcggttaggcgtgtacgggtgggaggtctatataagcagagctcggttagtgaaacctcagatcgccgtggagac
 gcatgatggccatgaactccaagcagcctttcgccatgcacccgggtgctgcaagaacccaaattctccagtctgcac
 tctggctccgaggccatgcgccgagctgtctctccagccccgcagctgcagggtaatatatttggaaagctttgatga
 gagcctgctggcacgcgccgaagctctggcggtggatctctccacggcaagaacctccgttcaagccccg
 30 acgccacctaccataccatgagcagcgtgcctgcacgtccacttcgtccaccgtgcccattctccacccagctgcg
 ctccacctcacacctcaccacgcccgtgcaccagggcctcgaaggcgacctgctggagcacatctcgccacgctgag
 tgtgagcggcctggcgctccggaacactcggtgatgccgcacagatccatccacaccacctggcgccatgggccc
 acctgcaccaggccatgggcatgagtcacccgcacaccgtggccctcatagcgccatgctgcatgcctcagcgac
 gtggagtcagacccgcgcgagctggaagccttcgccgagcgttcaagcagcggcgcatcaagctgggggtgaccca
 35 ggcggacgtggcgcggtcttgctaattctcaagatccccggcggtgggctcgctgagccaaagcaccatctgcaggt
 tcgagctctctcactctctcgcaacaacatgatcgctctcaagccggtgctccaggcctggttgaggaggccgag
 gccgcctaccgagagaagaacagcaagccagagctcttcaacggcagcgaacggaagcgcaaacgcacgtccatcg

ggcgcgggagaagcggttcactcgaggcctatttcgctatccagccacgtccttcatctgagaagatcgcgggccatcg
 ctgagaaactggaccttaaaaagaacgtggtgagagtctggttctgcaaccagagacagaaacagaaacgaatgaag
 tattcggctgtccacgttaacgattacaaggatgacgacgataaggactataaggacgatgatgacaaggactacaa
 agatgatgacgataaaggatccggcgagggcagaggaagtctgctaacatgcggtgacgtcgaggagaatcctggcc
 5 caatgcctaagaagaagcggaaagtcggctccggcgtgagcaagggcgaggcagtgatcaaggagttcatgcggttc
 aaggtgcacatggagggctccatgaacggccacgagttcgagatcgagggcgagggcgagggccgcccctacgaggg
 caccagaccgccaagctgaaggtgaccaaggtggccccctgcccttctcctgggacatcctgtcccctcagttca
 tgtacggctccagggccttcatcaagcaccgcccacatccccgactactataagcagtccttccccgagggcttc
 aagtgggagcgctgatgaacttcgaggacggcgggcgccgtgaccgtgaccaggacacctccctggaggacggcac
 10 cctgatctacaagggtgaagctccggcgaccaacttccctcctgacggccccgtaatgcagaagaagacaatgggct
 gggaagcgctccaccgagcggttgtaacccgaggacggcggtgctgaagggcgacattaagatggccctgcgcctgaag
 gacggcgggcgctacctggcggaacttcaagaccacctacaaggccaagaagcccgatgcagatgcccggcgcctacaa
 cgctcgaccgcaagttggacatcacctcccacaacgaggactacaccgtggtggaacagtagcaacgctccgagggcc
 gccactccaccggcgccatggacgagctgtacaaggtaccatcagtcgtgattgcgggcgttagcggttagattacgtt
 15 atcggcatggaaaacgccatgccgtggaacctgcctgccgatctcgccctgggtttaacgcaacaccttaataaacc
 cgtgattatgggcccatacctgggaatcaatcggtcgctccgttgccaggacgcaaaaatattatcctcagcagtc
 aaccgagtagcgacgatcgcgtaacgtgggtgaagtcggtggatgaagccatcgcgggcgtgtggtgacgtaccagaa
 atcatggtgattggcgggcggtcgcggttattgaacagttcttgccaaaagcgcaaaaactgtatctgacgcatatcga
 cgcagaagtggaaaggcgacaccatttcccggattacgagccggatgactgggaatcggtattcagcgaattccacg
 20 atgctgatgcgcagaactctcacagctattgctttgagattctggagcgcgataagctgatcagcctcgactgtgc
 cttctagttgccagccatctgttgtttgcccctcccccgctgccttcccttgaccctggaaggtgccactcccactgtc
 ctttcctaataaaaatgaggaaattgcatcgcatgtctgagtaggtgtcattctattctggggggtgggggtggggca
 ggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggaggaacccctag
 tgatggagttggccactccctctctgcgcgtcgctcgctcactgaggccggcgaccaaaggctgcccgcagcccg
 25 ggctttgcccggcgccctcagtgagcgagcgagcgcgagctgcctgcagg

hPou4f3 cDNA sequence (SEQ ID NO: 61)

atgatggccatgaactccaagcagcctttcggcatgcacccggtgctgcaagaacccaaattctccagtcctgcactc
 tggctccgaggccatgcgccgagtcgtctcccagccccgcagctgcagggtaatatatttggaaagctttgatgaga
 30 gcctgctggcacgcgccgaagctctggcgcggtggatatcgtctcccacggcaagaaccatccgttcaagcccgac
 gccacctaccataccatgagcagcgtgccctgcacgtccacttcgtccaccgtgcccatctcccaccagctgcgct
 cacctcacaccctcaccacgcccgtgcaccagggcctcgaaggcgacctgctggagcacatctcgcccacgctgagtg
 tgagcggcctggcgctccggaactcggtgatgcccgcacagatccatccacaccacctggcgccatggggcac
 ctgcaccaggccatgggcatgagtcacccgcacacccgtggcccctcatagcgccatgcctgcatgcctcagcgacgt
 35 ggagtcagaccgcgcgagctggaagccttcgccgagcgcttcaagcagcgcgcatcaagctgggggtgaccacgg
 cggacgtggcgcggtctggctaattctcaagatccccggcggtgggctcgctgagccaaagcaccatctgcaggttc
 gagtctctcactctctcgacaacaacatgatcgctctcaagccggtgctccaggcctggttgaggaggccgaggg
 cgccctaccgagagaagaacagcaagccagagctcttcaacggcagcgaacggaagcgcaaacgcacgtccatcgcg

cgccggagaagcggttcactcgaggcctatttcgctatccagccacgtccttcatctgagaagatcgcgggccatcgct
gagaaactggaccttaaaaagaacgtggtgagagtctggttctgcaaccagagacagaaacagaaacgaatgaagta
ttcggctgtccacgttaac

5 3x FLAG cDNA sequence (SEQ ID NO: 62)

gattacaaggatgacgacgataaggactataaggacgatgatgacaaggactacaaagatgatgacgataaaggatc
cggc

T2A cDNA sequence (SEQ ID NO: 63)

10 gagggcagaggaagtctgctaacatgcggtgacgtcgaggagaatcctggccca

T2A cDNA sequence (SEQ ID NO: 89)

GCGAGGGCAGAGGAAGTCTGCTAACATGCGGTGACGTCGAGGAGAATCCTGGCCCA

15 In some embodiments of any of the compositions described herein, the vector comprises
or consists of pITR-CMV-hGFI1-T2A-mScarlet-DD (SEQ ID NO: 64). In some embodiments
of any of the compositions described herein, the vector comprises a sequence that has at least
75% (e.g., at least 80%, at least 82%, at least 84%, at least 85%, at least 86%, at least 88%, at
least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 98%, at least 99%)
20 sequence identity to SEQ ID NO: 64.

pITR-CMV-hGFI1-T2A-mScarlet-DD (SEQ ID NO: 64)

cctgcaggcagctgcgcgctcgctcgctcactgagggccgcccggcgctcgggcgacctttggtcgcccgccctcagt
gagcgagcgagcgcgagagagggagtggccaaactccatcactaggggttcctctagatcccatatatggagttccg
25 cgttacataacttacggtaaatggcccgctggctgaccgcccacgacccccgcccattgacgtcaataatgacgt
atgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggagtatttacggtaaactgcccaacttg
gcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgctggcatta
tgcccagtagatgaccttatgggacttttcctacttggcagtagatctacgtatttagtcatcgctattaccatgggtga
tgcggttttggcagtagatcaatggcggtgtagcggtttgactcacggggatttccaagtctccacccattgac
30 gtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgca
aatgggcggtaggcggtgtacgggtgggaggtctatataagcagagctcggttagtgaaaccgtcagatcgcttgagac
gcatgccgcgctcatttctcgtaaaaagcaagaaggctcacagctaccaccagccgcgctccccaggaccagactat
tcctccggttagagaatgtaccggcgcttagccgagcagacagcacttcaaatgcaggcgggggaaggcgagacc
ccgggaccgtttgtccccgaatcgagctgaccgaagccccagacagagcctccgcatccccagacagctgcgaag
35 gcagcgctctgcgaacggagctcgagtttgaggacttctggaggcccccgctaccctccgctctccagcctcgag
aagtcaatgtgccatcgctggacgaagcccagcccttccccctgcctttcaaacgtactcatggagcggtggc

gggttctgacctgcgccacctggtgcagagctaccgacctgtggggccctggagcgtggcgctggcctgggcctct
 tctgcgaaccgccccggagcctggccaccgcccgcgctgtacggcccgaagcgggctgcccggcgcgggggcc
 gggcgccagggagctgcagcgcaggggcccgtgccaccgctggccctggcctagggctctacggcgacttcgggtc
 tgccgcagccgggctgtatgagaggccacggcagcggcggttgcgtgtaccccagcgtggccacgggctgcacg
 5 cagacaagggcgctggcgtcaaggtggagtcggagctgctgtgcaccgcctgctgctggcgcgcggtcctacaag
 tgcataaagtgcagcaaggtgttctccacgcgcacgggctcgaggtgcacgtgcccaggtcccacagcggcaccag
 accctttgcctgagagatgtgcggcaagaccttcgggcacgcgggtgagcctggagcagcacaagccgtgcactcgc
 aggaacggagctttgactgtaagatctgtgggaagagcttcaagaggtcatccacactgtccacacacctgcttacc
 cactcagacactcggccctaccctgtcagtagctgtggcaagaggttccaccagaagtgcagacatgaagaaacacac
 10 tttcatccacactggtgagaagcctcacaagtgccaggtgtgcggcaaggcattcagccagagctccaacctcatca
 cccacagccgcaaacacacaggttcaagcccttcggctgcgacctctgtgggaagggtttccagaggaaggtggac
 ctccgaaggcaccgggagacgcagcatgggctcaaagttaacgattacaaggatgacgacgataaggactataagga
 cgatgatgacaaggactacaaagatgatgacgataaaggatccggcgagggcagaggaagtctgctaacatgcggtg
 acgtcgaggagaatcctggcccaatgcctaagaagaagcggaaagtgcggtccggcggtgagcaagggcgaggcagt
 15 atcaaggagtcatgcggttcaaggtgcacatggagggctccatgaacggccacagagttcgagatcgagggcgaggg
 cgagggccgccccctacgagggcaccagaccgccaagctgaaggtgaccaagggtggccccctgcccttctcctggg
 acatcctgtccccctcagttcatgtacggctccagggccttcatcaagcaccgccgacatccccgactactataag
 cagtccttccccgagggcttcaagtgaggcgcgtgatgaacttcgaggacggcgcgccgtgacctgacctcagga
 cacctccctggaggacggcaccctgatctacaaggtgaagctccgcggcaccaacttccctcctgacggccccgtaa
 20 tgcagaagaagacaatgggctgggaagcgtccaccgagcgggtgtaccccagggacggcggtgctgaaggcgacatt
 aagatggccctgcgctgaaggacggcgccgctacctggcggacttcaagaccacctacaaggccaagaagccggt
 gcagatgcccggcgccctacaacgtcgaccgcaagttggacatcacctcccacaacgaggactacacctggtggaac
 agtacgaacgctccgagggccgcccactccaccggcggtatggacgagctgtacaagggtaccatcagttctgattgcg
 gcgttagcggtagattacgttatcgccatggaaaacgccatgccgtggaacctgcctgccgatctcgcttggtttaa
 25 acgcaacaccttaataaaccggtgattatgggcccacatacctgggaatcaatcggtcgtccgttgccaggacgca
 aaaatattatcctcagcagtcacaccgagtagcgacgacatcgcgtaacgtgggtgaagtcgggtggatgaagccatcgcg
 gcgtgtggtgacgtaccagaaatcatggtgattggcgccggtcgcggttattgaacagttcttgccaaaagcgcaaaa
 actgtatctgacgcatatcgacgcagaagtgggaaggcgacaccatttcccggattacgagccggatgactgggaat
 cggatttcagcgaattccacgatgctgatgcgcagaactctcacagctattgctttgagattctggagcggcgataa
 30 gctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgccctccccctgccttcttgacctg
 gaaggtgccactcccactgtcctttcctaataaaaatgaggaaattgcatcgcatgtctgagtaggtgtcattctat
 tctggggggtggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcggtgg
 gctctatggaggaacccttagtgatggagttggccactccctctctgcgcgctcgctcgctcactgaggccggggcga
 ccaaaggtcgcccgacgcccgggctttgccggggcgccctcagtgagcgagcgagcgcgagctgcctgcagg

35 **hGFI1 cDNA sequence (SEQ ID NO: 65)**
 ATGCCGCGCTCATTTCTCGTCAAAAGCAAGAAGGCTCACAGCTACCACCAGCCGCGCTCCCCAGGACCAGACTATTC
 CCTCCGTTTAGAGAATGTACCGGCGCCTAGCCGAGCAGACAGCACTTCAAATGCAGGCGGGGCGAAGGCGGAGCCCC

GGGACCGTTTGTCCCCGAATCGCAGCTGACCGAAGCCCCAGACAGAGCCTCCGCATCCCCAGACAGCTGCGAAGGC
 AGCGTCTGCGAACGGAGCTCGGAGTTTGAGGACTTCTGGAGGCCCCCGTCACCCTCCGCGTCTCCAGCCTCGGAGAA
 GTCAATGTGCCCATCGCTGGACGAAGCCCAGCCCTTCCCCCTGCCTTTCAAACCGTACTCATGGAGCGGCCTGGCGG
 GTTCTGACCTGCGGCACCTGGTGCAGAGCTACCGACCGTGTGGGGCCCTGGAGCGTGGCGCTGGCCTGGGCCTCTTC
 5 TCGGAACCCGCCCCGGAGCCTGGCCACCCGGCCGCGCTGTACGGCCCCGAAGCGGGCTGCCGGCGGCGCGGGGGCCGG
 GCGGCCAGGGAGCTGCAGCGCAGGGGCCGGTGCCACCGCTGGCCCTGGCCTAGGGCTCTACGGCGACTTCGGGTCTG
 CGGCAGCCGGGCTGTATGAGAGGCCACGGCAGCGGGGCTTGTGTACCCCGAGCGTGGCCACGGGCTGCACGCA
 GACAAGGGCGCTGGCGTCAAGGTGGAGTCGGAGCTGCTGTGCACCCGCTGCTGCTGGGCGGCGGCTCCTACAAGTG
 CATCAAGTGCAGCAAGGTGTTCTCCACGCCGCACGGGCTCGAGGTGCACGTGCGCAGGTCCACAGCGGCACACAGAC
 10 CCTTTGCCTGCGAGATGTGCGGCAAGACCTTCGGGCACGCGGTGAGCCTGGAGCAGCACAAAGCCGTGCACTCGCAG
 GAACGGAGCTTTGACTGTAAGATCTGTGGGAAGAGCTTCAAGAGGTATCCACACTGTCCACACACCTGCTTATCCA
 CTCAGACACTCGGCCCTACCCCTGTCTAGTACTGTGGCAAGAGGTTCCACCAGAAGTCAGACATGAAGAAACACACTT
 TCATCCACACTGGTGAGAAGCCTCACAAGTGCCAGGTGTGCGGCAAGGCATTTCAGCCAGAGCTCCAACCTCATCACC
 CACAGCCGCAAACACACAGGCTTCAAGCCCTTCGGCTGCGACCTCTGTGGGAAGGGTTTCCAGAGGAAGGTGGACCT
 15 CCGAAGGCACCGGGAGACGCAGCATGGGCTCAAAGTTAAC

In some embodiments of any of the compositions described herein, the vector comprises
 or consists of pITR-CMV-hATOH1-T2A-mScarlet-DD (SEQ ID NO: 66). In some
 embodiments of any of the compositions described herein, the vector comprises a sequence that
 20 has at least 75% (e.g., at least 80%, at least 82%, at least 84%, at least 85%, at least 86%, at least
 88%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 98%, at least
 99%) sequence identity to SEQ ID NO: 66.

pITR-CMV-hATOH1-T2A-mScarlet-DD (SEQ ID NO: 66)
 25 cctgcaggcagctgcgcgctcgctcgctcactgaggccgcccggcgctcgggcgacctttggtcgcccgccctcagt
 gagcgagcgagcgcgagagagggagtgcccaactccatcactaggggttcctctagatcccatatatggagttccg
 cgttacataacttacggtaaatggcccgctggctgaccgccccacgacccccgcccattgacgtcaataatgacgt
 atgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggagattttacggtaaactgcccacttg
 gcagtacatcaagtgtatcatatgccaaagtacgccccctattgacgtcaatgacggtaaatggcccgccctggcatta
 30 tgcccagtagacgttatgggactttcctacttgaggtagcatctacgtattagtcacgctattaccatgggtga
 tgcggtttttggcagtagcatcaatggcggtgtagcggtttgactcacggggatttccaagtctccacccattgac
 gtcaatgggagttttgttttgccacaaaatcaacgggactttccaaaatgtcgtacaaactccgccccattgacgca
 aatggcggttaggcgtgtacggtgggaggtctatataagcagagctcgtttagtgaaccgtcagatcgctggagac
 gcatgtcccgccctgctgcatgcagaagagtggtgaagtgaaggagttgggagaccaccatcgccagccccagccg
 35 catcatctcccgcaaccgcccgcgcccgcgagccacctgcaactttgcaggcgagagagcatcccgctctaccgccc
 tgagctgtccctcctggacagcaccgaccacgcgctggctggctccactttgcaggcgatctgcacggcacgcg
 ccgcccagtagtttgctacattccccggagctgggtgcctcagaggccgctgcgccccgggacgaggtggacggccgg

ggggagctggtaaggaggagcagcggcggtgccagcagcagcaagagccccgggcccgggtgaaagtgcgggaacagct
 gtgcaagctgaaaggcgggggtggtggttagacgagctgggctgcagccgccaacggggccccttccagcaaacaggtga
 atgggggtgcagaagcagagacggctagcagccaacgccaggaggagcggcgaggatgcatgggctgaaccacgccttc
 gaccagctgcgcaatgttatcccgtcgttcaacaacgacaagaagctgtccaaatatgagaccctgcagatggcca
 5 aatctacatcaacgccttgtccgagctgctacaaacgcccagcggagggggaacagccaccgcgcctccagcctcct
 gcaaaagcgaccaccaccaccttcgcaccgcggcctcctatgaagggggcgcgggcaacgcgaccgcagctggggct
 cagcaggcttccggaggagaccagcggccgaccccgccgggagttgcccggactcgttctcagccccagcttctgc
 gggagggtactcgggtgcagctggacgctctgcacttctgcacttctcagggacagcgccctgacagcgatgatggcg
 aaaagaatttgtctccttctctccccgggagcatcttgcagccagtgcaggaggaaaacagcaaaacttcgcctcgg
 10 tcccacagaagcgacggggaattttccccccattcccattacagtgactcggatgaggcaagtgttaacgattacaa
 ggatgacgacgataaggactataaggacgatgatgacaaggactacaaagatgatgacgataaaggatccggcgagg
 gcagaggaagtctgtaacatgcggtgacgtcgaggagaatcctggcccaatgcctaagaagaagcggaaagtccggc
 tccggcgtgagcaagggcgaggcagtgatcaaggagttcatgcggttcaaggtgcacatggagggtccatgaacgg
 ccacgagttcgagatcgaggggcgaggggcgaggggccgcccctacgagggcacccagaccgccaagctgaagggtacca
 15 aggggtggccccctgccttctcctgggacatcctgtcccctcagttcatgtacggctccagggccttcatcaagcac
 cccgcgcacatccccgactactataagcagtccttccccgagggttcaagtgggagcgcgatgaacttcgagga
 cggcgggcgccgtgaccgtgacccaggacacctccctggaggacggcaccctgatctacaaggtgaagctccgcggca
 ccaacttccctcctgacggccccgtaatgcagaagaagacaatgggctgggaagcgtccaccgagcgggtgtacccc
 gaggacggcgtgctgaaggggcgacattaagatggccctgcgcctgaaggacggcgccgctacctggcggacttcaa
 20 gaccacctacaaggccaagaagcccgtgcagatgcccggcgccctacaacgtcgaccgcaagttggacatcacctccc
 acaacgaggactacaccgtggtggaacagtagcaacgctccgaggggccgccaactccaccggcgcatggacgagctg
 tacaagggtaccatcagctctgattgcggcggttagcggtagattacgttatcggcatggaaaacgccatgccgtggaa
 cctgcctgccgatctcgctgggtttaaacgcaacaccttaaataaaacccgtgattatgggcccgcatacctgggaat
 caatcggtcgtccgttgccaggacgcaaaaatattatcctcagcagtagcaaccgagtagcgacgatcgcgtaacgtgg
 25 gtgaagtcgggtggatgaagccatcgcgcggtgtggtgacgtaccagaaatcatggtgattggcgcggtcgcggttat
 tgaacagttcttgccaaaagcgcaaaaactgtatctgacgcatatcgacgcagaagtggaaaggcgacacccatttcc
 cggattacgagccggatgactgggaatcggtattcagcgaattccacgatgctgatgcgcagaactctcacagctat
 tgctttgagattctggagcggcgataagctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgc
 ccctcccccgctgccttcccttgaccctggaaggtgccactcccactgtcctttcctaataaaaatgaggaaattgcac
 30 gcattgtctgagtaggtgtcattctattctgggggtgggtggggcaggacagcaagggggaggattgggaagaca
 atagcaggcatgctggggatgcggtgggctctatggaggaacccctagtgatggagttggccactccctctctgcgc
 gctcgctcgctcactgaggccggggcgaccaaaggtcgcccgacgcccgggctttgcccggggcgccctcagtgagcga
 gcgagcgcgcagctgcctgcagg

35 hATOH1 cDNA sequence (SEQ ID NO: 67)

atgtccgcctgtgcatgcagaagagtggtgtaagtgaaggagttgggagaccaccatcgccagccccagccgca
 tcatctcccgcaaccgcgcgcgcgcgcgcagccacctgcaactttgcaggcgagagagcatcccgctctaccgcctg
 agctgtccctcctggacagcaccgaccacgcgcctggctggctcccactttgcagggcacgtgcacggcacgcgc

gccagtagtttgcacattccccggagctgggtgcctcagagggcgctgcgccccgggacgaggtggacggccgggg
 ggagctggtaaggaggagcagcggcggtgccagcagcagcaagagccccgggcccgggtgaaagtgcgggaacagctgt
 gcaagctgaaaggcgggggtggtggttagacgagctgggtgcagccgccaacggggccccttcagcaaacaggtgaat
 ggggtgcagaagcagagacggctagcagccaacgccaggaggagcggcgaggtgcatgggtgaaccacgccttcga
 5 ccagctgcgcaatgttatcccgtcgttcaacaacgacaagaagctgtccaaatatgagaccctgcagatggcccaa
 tctacatcaacgccttgtccgagctgctacaaacgcccagcggagggggaacagccaccgcgcctccagcctcctgc
 aaaagcgaccaccaccaccttcgcaccgcggcctcctatgaagggggcggggcaacgcgacgcagctggggctca
 gcaggcttcggaggaggagccagcggcgacccccggggagttgcgggactcgccttcagccccagcttctgcgg
 gagggtagctcgggtgcagctggacgctctgcacttctcgcacttctcagggacagcgccctgacagcgatgatggcgcaa
 10 aagaatttgcctccttctcctccccgggagcatcttgcagccagtgaggaggaaaacagcaaaacttcgcctcggtc
 ccacagaagcgacggggaattttccccccattcccattacagtgactcggatgaggcaagtgttaac

In some embodiments of any of the compositions described herein, the vector comprises
 or consists of pITR-CMV-Luc-T2A-mScarlet-U6-Hes1 –S3 (SEQ ID NO: 68). In some
 15 embodiments of any of the compositions described herein, the vector comprises a sequence that
 has at least 75% (e.g., at least 80%, at least 82%, at least 84%, at least 85%, at least 86%, at least
 88%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 98%, at least
 99%) sequence identity to SEQ ID NO: 68.

20 **pITR-CMV-Luc2-T2A-mScarlet-U6-Hes1-S3 (SEQ ID NO: 68)**
 cctgcaggcagctgcgcgctcgcctcactgaggccgcccggcgctcgggcgacctttgggtcgcccgccctcagt
 gagcgagcgagcgcgcagagagggagtgcccaactccatcactaggggttcctctagatcccatatatggagttccg
 cgttacataacttacggtaaatggcccgccctggctgaccgccaacgacccccgccattgacgtcaataatgacgt
 atgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggagtatattacggtaaactgccacttg
 25 gcagtagcatcaagtgtatcatatgccaaagtacgccccctattgacgtcaatgacggtaaatggcccgccctggcatta
 tgcagtagcatgaccttatgggactttcctacttggcagtagcatctacgtattagtcacgctattaccatgggtga
 tgccggttttggcagtagcatcaatggcggtgtagcggtttgactcacggggatttccaagtctccacccattgac
 gtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgca
 aatgggcggtaggcgtgtacggtgggaggtctatataagcagagctcgttttagtgaaccgtcagatcgccctggagac
 30 gcatggaagatgcaaaaaacattaagaaggggccagcgccattctaccactcgaagacgggacggcgagcag
 ctgcacaaagccatgaagcgtacgcccctgggtgccggcaccatcgcccttaccgacgcacatatcgaggtggacat
 tactacgcccagtagtacttcgagatgagcgttcgggtggcagaagctatgaagcgtatgggtgaatacaaacctac
 ggatcgtggtgtgcagcgagaatagcttgagttcttcatgcccgtgttgggtgccctgttcacggtgtggtgtg
 gccccagctaacgacatctacaacgagcgcgagctgctgaacagcatgggcatcagccagcccaccgtcgtattcgt
 35 gagcaagaaagggtgcaaaagatcctcaacgtgcaaaagaagctaccgatcataaaaagatcatcatcatggata
 gcaagaccgactaccagggttccaaagcatgtacaccttcgtgacttccatttgccacccggcttcaacgagtag
 gacttcgtgcccagagcttcgaccgggacaaaaccatcgccctgatcatgaacagtagtggcagtagcggattgcc

caagggcgtagccctaccgcaccgcaccgcttggtggtcgattcagtcattgcccgcgaccccatcttcggcaaccaga
 tcatccccgacaccgctatcctcagcgtggtgccatttcaccacggcttcggcatggttcaccacgctgggctacttg
 atctgcggcttttcgggtcgtgctcatgtaccgcttcgaggaggagctattcttgcgagcttgcaagactataagat
 tcaatctgccctgctggtgccacactatcttagcttcttcgctaagagcactctcatcgacaagtacgacctaaagca
 5 acttgacagagatcgccagcggcgggcgccgctcagcaaggaggtaggtgaggccgtggccaaacgcttccacct
 ccaggcatccgcccagggtacggcctgacagaaaacaaccagcgccattctgatcacccccgaaggggacgacaagcc
 tggcgagtaggcaagggtggtgcccttcttcgaggctaagggtggtggacttggacacaggttaagacactgggtgtga
 accagcgcgggagctgtgcgtccgtggccccatgatcatgagcggctacgttaacaaccccgaggctacaaacgct
 ctcatcgacaaggacggctgggtgcacagcggcgacatcgccactactgggacgaggacgagcacttcttcacgtgga
 10 ccggctgaagagcctgatcaaatacaagggtaccaggtagccccagccgaactggagagcatcctgctgcaacacc
 ccaacatcttcgacgcccgggtgcgcggcctgcccgcgacgatgccggcgagctgccggccgagctcgtcgtgctg
 gaacacggtaaaaccatgaccgagaaggagatcgtggactatgtggccagccagggttacaaccgccaagaagctgcg
 cgggtggtgtgtgttcgtggacgaggtgcctaaaggactgaccggcaagttggacgcccgaagatccgcgagattc
 tcattaaggccaagaagggcggaagatcgccgtgggctccggagagggcgaggaagctcgtacatgaggtgcacat
 15 gtcgaggagaatcctggcccaatggtgagcaaggcgaggcagtgatcaaggagttcatgagggttcaagggtgcacat
 ggagggtccatgaacggccacgagttcgagatcgaggcgaggcgaggcgcccccctacgagggcacccagaccg
 ccaagctgaagggtgaccaagggtggccccctgcccttctcctgggacatcctgtccccctcagttcatgtacggctcc
 agggccttcatcaagcaccgccgacatccccgactactataagcagtccttccccgagggttcaagtgggagcg
 cgtgatgaacttcgaggacggcgccgctgaccgtgaccagggacacctccctggaggacggcaccctgatctaca
 20 aggtgaagctccgcggcaccaacttccctcctgacggccccgtaatgcagaagaagacaatgggctgggaagcgtcc
 accgagcgggtgtaccccgaggacggcgctgtaaggcgacattaagatggccctgcgcctgaaggacggcgccg
 ctacctggcgacttcaagaccacctacaaggccaagaagcccgtgcagatgccggcgccctacaacgtcgaccgca
 agttggacatcacctcccacaacgaggactacaccgtggtggaacagtagcaacgctccgaggggccgcccactccacc
 ggcgccatggacgagctgtacaagtaagctgatcagcctcgataagatacattgatgagtttggaacaaccacaact
 25 agaatgcagtgaaaaaatgctttatgtgaaatgtgtgatgctattgctttatgttaaccattataagctgcaa
 taaacaagtttaaggctgggcaggaagaggccctatttcccatgattccttcatatttgcatatacgatacaaggctg
 ttagagagataattagaattaatttgactgtaaacacaaagatattagtagcaaaatacgtgacgtagaaagtaataa
 tttcttgggtagtttgacgttttaaaattatgttttaaaatggactatcatatgcttacggtaacttgaaagtattt
 cgatttcttggctttatatatcttgtggaaaggacgaaacaccgaaagtcatcaaagcctatcgaaataggctttga
 30 tgactttcttttttaggaaccctagtgtgaggttgccactccctctctgcgcgctcgtcgtcactgaggccg
 ggcgaccaaaggctcgccgcgacgcccgggctttgccggggcgccctcagtgagcgagcgagcgcgagctgc

In some embodiments of any of the compositions described herein, the vector comprises
 or consists of pITR-CMV-Luc-T2A-GFP-U6-Hes1 –S5 (SEQ ID NO: 78). In some
 35 embodiments of any of the compositions described herein, the vector comprises a sequence that
 has at least 75% (e.g., at least 80%, at least 82%, at least 84%, at least 85%, at least 86%, at least

88%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 98%, at least 99%) sequence identity to SEQ ID NO: 78.

pITR-CMV-Luc2-T2A-GFP-U6-Hes1 –S5 (SEQ ID NO: 78)

5 CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCGTCGGGCGACCTTTGGTCGCCCCGGCCTCAGT
GAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCTCTAGATCCCATATATGGAGTTCCG
CGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCATTGACGTCAATAATGACGT
ATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAACTGCCCCACTTG
10 GCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATT
TGCCCAGTACATGACCTTATGGGACTTTCTACTTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGA
TGCGGTTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGAC
GTCAATGGGAGTTTGTGGTGGCACAAAATCAACGGGACTTTCCAAAATGTGTAACAACCTCCGCCCCATTGACGCA
AATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAGAC
GCATGGAAGATGCCAAAAACATTAAGAAGGGCCAGCGCCATTCTACCCACTCGAAGACGGGACCGCCGGCGAGCAG
15 CTGCACAAAGCCATGAAGCGCTACGCCCTGGTGCCCGGCACCATCGCCTTTACCGACGCACATATCGAGGTGGACAT
TACCTACGCCGAGTACTTCGAGATGAGCGTTCGGCTGGCAGAAGCTATGAAGCGCTATGGGCTGAATACAAACCATC
GGATCGTGGTGTGCAGCGAGAATAGCTTGCGATTCTTCATGCCCCGTGTTGGGTGCCCTGTTTCATCGGTGTGGCTGTG
GCCCCAGCTAACGACATCTACAACGAGCGCGAGCTGCTGAACAGCATGGGCATCAGCCAGCCCACCGTCGTATTCTG
GAGCAAGAAAGGGCTGCAAAAGATCCTCAACGTGCAAAAGAAGCTACCGATCATACAAAGATCATCATCATGGATA
20 GCAAGACCGACTACCAGGGCTTCCAAAGCATGTACACCTTCGTGACTTCCCATTGTCACCCGGCTTCAACGAGTAC
GACTTCGTGCCCGAGAGCTTCGACCGGGACAAAACCATCGCCCTGATCATGAACAGTAGTGGCAGTACCGGATTGCC
CAAGGGCGTAGCCCTACCGCACCGCACCGCTTGTGTCCGATTTCAGTCATGCCCGCGACCCCATCTTCGGCAACCAGA
TCATCCCCGACACCGCTATCCTCAGCGTGGTGCCATTTACACACGGCTTCGGCATGTTACACACGCTGGGCTACTTG
ATCTGCGGCTTTTCGGGTGCTGCTCATGTACCGCTTCGAGGAGGAGCTATTCTTGCGCAGCTTGCAAGACTATAAGAT
25 TCAATCTGCCCTGCTGGTGCCCACTATTTAGCTTCTTCGCTAAGAGCACTCTCATCGACAAGTACGACCTAAGCA
ACTTGACGAGATCGCCAGCGGCGGGCGCCGCTCAGCAAGGAGGTAGGTGAGGCCGTGGCCAAACGCTTCCACCTA
CCAGGCATCCGCCAGGGCTACGGCCTGACAGAAACAACAGCGCCATTCTGATCACCCCCGAAGGGGACGACAAGCC
TGGCGCAGTAGGCAAGGTGGTGCCCTTCTTCGAGGCTAAGGTGGTGGACTTGGACACAGGTAAGACACTGGGTGTGA
ACCAGCGCGGCGAGCTGTGCGTCCGTGGCCCCATGATCATGAGCGGCTACGTTAACAACCCCGAGGCTACAAACGCT
30 CTCATCGACAAGGACGGCTGGCTGCACAGCGGCGACATCGCCTACTGGGACGAGGACGAGCACTTCTTCATCGTGGA
CCGGCTGAAGAGCCTGATCAAATACAAGGGCTACCAGGTAGCCCCAGCCGAAGTGGAGAGCATCCTGCTGCAACACC
CCAACATCTTCGACGCCGGGGTCGCCGGCCTGCCGACGACGATGCCGGCGAGCTGCCCGCCGAGTCGTGCTGCTG
GAACACGGTAAAACCATGACCGAGAAGGAGATCGTGGACTATGTGGCCAGCCAGGTTACAACCGCCAAGAAGCTGCG
CGGTGGTGTGTGTTTCGTGGACGAGGTGCCTAAAGGACTGACCGGCAAGTTGGACGCCCCGAAGATCCGCGAGATTC
35 TCATTAAGGCCAAGAAGGGCGGCAAGATCGCCGTGGGCTCCGAGAGGGCAGAGGAAGTCTGCTAACATGCGGTGAC
GTCGAGGAGAATCCTGGCCCAATGGTGAGCAAGGGCGAGGCAGTGATCAAGGAGTTCATGCGGTTCAAGGTGCACAT
GGAGGGCTCCATGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCGCCCCCTACGAGGGCACCCAGACCG
CCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCTCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCC

AGGGCCTTCATCAAGCACCCCGCCGACATCCCCGACTACTATAAGCAGTCCTTCCCCGAGGGCTTCAAGTGGGAGCG
 CGTGATGAACTTCGAGGACGGCGGCCCGTGACCGTGACCCAGGACACCTCCCTGGAGGACGGCACCCCTGATCTACA
 AGGTGAAGCTCCGCGGCACCAACTTCCCTCCTGACGGCCCCGTAATGCAGAAGAAGACAATGGGCTGGGAAGCGTCC
 ACCGAGCGGTTGTACCCCGAGGACGGCGTGCTGAAGGGCGACATTAAGATGGCCCTGCGCCTGAAGGACGGCGGCCG
 5 CTACCTGGCGGACTTCAAGACCACCTACAAGGCCAAGAAGCCCCGTGCAGATGCCCCGGCGCCTACAACGTCGACCGCA
 AGTTGGACATCACCTCCCACAACGAGGACTACACCGTGGTGGAACAGTACGAACGCTCCGAGGGCCGCCACTCCACC
 GGCGGCATGGACGAGCTGTACAAGTAAGCTGATCAGCCTCGATAAGATACATTGATGAGTTTGGACAAACCACAAC
 AGAATGCAGTGAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAA
 TAAACAAGTTAAGGTGCGGCAGGAAGAGGGCCTATTTCCCATGATTCCCTTCATATTTGCATATACGATACAAGGCTG
 10 TTAGAGAGATAATTAGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAA
 TTTCTTGGGTAGTTTGCAGTTTTAAATTATGTTTTAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTT
 CGATTTCTTGGCTTTATATATCTTGTGGAAGGACGAAACACCACTGCATGACCCAGATCAAcgaaTTGATCTGGGT
 CATGCAGTTTTTTTAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCG
 GGCGACCAAAGGTCGCCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGC

15 In some embodiments of any of the compositions described herein, the vector comprises
 or consists of pITR-CMV-Luc-T2A-GFP-U6-Hes1 –KOP (SEQ ID NO: 79). In some
 embodiments of any of the compositions described herein, the vector comprises a sequence that
 has at least 75% (e.g., at least 80%, at least 82%, at least 84%, at least 85%, at least 86%, at least
 20 88%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 98%, at least
 99%) sequence identity to SEQ ID NO: 79.

pITR-CMV-Luc2-T2A-GFP-U6-Hes1 –KOP (SEQ ID NO: 79)

cctgcaggcagctgcgcgctcgcctcactgagggccgcccggcgctcgggcgacctttggctcgcccgccctcagt
 25 gagcgagcgagcgcgagagaggagtgccaaactccatcactaggggttcctctagatcccatatatggagttccg
 cggtacataacttacggtaaatggcccgctggctgaccgcccacgacccccgcccattgacgtcaataatgacgt
 atgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggagtatttacggtaaactgcccacttg
 gcagtacatcaagtgtatcatatgccaaagtacgccccctattgacgtcaatgacggtaaatggcccgccctggcatta
 tgcccagttacatgaccttatgggactttcctacttggcagttacatctacgtatttagtcatcgctattaccatgggtga
 30 tgcggttttggcagttacatcaatggcgctggatagcggtttgactcacggggatttccaagtctccacccattgac
 gtcaatgggagttttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgca
 aatgggcggtaggcgtgtacggtgggaggtctatataagcagagctcgtttagtgaaccgtcagatcgccctggagac
 gcatggaagatgcaaaaaacattaagaaggggccagcgccattctaccactcgaagacgggaccgcccggcgagcag
 ctgcacaaagccatgaagcgctacgcccgtggtgccggcaccatcgccctttaccgacgcacatatcgaggtggacat
 35 tacctacgcccagttacttcgagatgagcggttcggctggcagaagctatgaagcgctatgggctgaatacaaacatc
 ggatcggtgtgtgcagcgagaatagcttgagttcttcatgcccgtgttggtgcccgtgttcacggtgtggctgtg
 gccccagctaacgacatctacaacgagcgcgagctgctgaacagcatgggcatcagccagcccaccgtcgatttcgt

gagcaagaaagggctgcaaaagatcctcaacgtgcaaaagaagctaccgatcatataaaaagatcatcatcatggata
 gcaagaccgactaccagggcttccaaagcatgtacaccttcgtgacttcccatttgccacccggcttcaacgagtac
 gacttcgtgcccagagagcttcgacccgggacaaaaccatcgccctgatcatgaacagtagtggcagtagccgattgcc
 caagggcgtagccctaccgcaccgcaccgcttggtggtccgattcagtcattgcccgcgaccccatcttcggcaaccaga
 5 tcatccccgacaccgctatcctcagcgtggtgccatttcaccacggcttcggcatggtcaccacgctgggctacttg
 atctgcggcttttcgggtcgtgctcatgtaccgcttcgaggaggagctattcttgccgagcttgcaagactataagat
 tcaatctgcccgtgctggtgcccacactatttagcttcttcgctaagagcactctcatcgacaagtagcacctaagca
 acttgacagagatcgccagcggcgggcgccgctcagcaaggaggtaggtgaggccgtggccaaacgcttccacct
 ccaggcatccgcccagggtacggcctgacagaaacaaccagcgccattctgatccccccgaaggggacgacaagcc
 10 tggcgcagtaggcaaggtggtgcccttcttcgaggctaaggtggtggacttggacacaggtaagacactgggtgtga
 accagcgcggcgagctgtgctgctccgtggccccatgatcatgagcggctacgttaacaaccccgaggctacaaacgct
 ctcatcgacaaggacggctggctgcacagcggcgacatcgccctactgggacgaggacgagcacttcttcatcgtgga
 ccggctgaagagcctgatcaaatacaagggtaccaggtagccccagccgaactggagagcatcctgctgcaacacc
 ccaacatcttcgacgcgggggtgcggcgccctgcccgcgacgatgcccgcgagctgcccgcgcgagctcgtcgtgctg
 15 gaacacggtaaaaccatgaccgagaaggagatcgtggactatgtggccagccaggttacaacccccaagaagctgcg
 cggtggtgttgtgttcgtggacgaggtgcctaaaggactgaccggcaagttggacgcccgaagatccgcgagattc
 tcattaaggccaagaagggcggaagatcgccgtgggctccggagagggcagaggaagtctgctaacatgcggtgac
 gtcgaggagaatcctggcccaatggtgagcaaggcgaggcagtgatcaaggagttcatgcggttcaagggtgcacat
 ggagggctccatgaacggccacgagttcgagatcgagggcgagggcgagggccgcccctacgagggcaccacagaccg
 20 ccaagctgaagggtgaccaaggggtggccccctgcccttctcctgggacatcctgtcccctcagttcatgtacggctcc
 agggccttcatcaagcaccgccgacatccccgactactataagcagtccttccccgagggttcaagtgggagcg
 cgtgatgaacttcgaggacggcgccgctgaccgtgaccacggacacctccctggaggacggcacccctgatctaca
 aggtgaagctccgcggcaccaacttccctcctgacggccccgtaatgcagaagaagacaatgggctgggaagcgtcc
 accgagcgggtgtaccccgaggacggcgctgtaaggggcgacattaagatggccctgcgcctgaaggacggcgggccg
 25 ctacctggcggacttcaagaccacctacaaggccaagaagcccgtgcagatgcccggcgccctacaacgtcgaccgca
 agttggacatcacctcccacaacgaggactacaccgtggtggaacagtacgaacgctccgagggccgcccactccacc
 ggcgcatggacgagctgtacaagtaagctgatcagcctcgataagatacattgatgagtttggaacaaaccacaact
 agaatgcagtgaaaaaaatgctttatttgtgaaatttgtgatgctattgctttatttgaaccattataagctgcaa
 30 taaacaagttaaggctcgggcaggaagagggcctatttcccatgattccttcatatttgcataacgatacaaggctg
 ttagagagataattagaattaatttgactgtaaacacaaagatatttagtataaaaatacgtgacgtagaaagtaataa
 tttcttgggtagtttgagttttaaaattatgttttaaaatggactatcatatgcttaccgtaacttgaaagtattt
 cgatttcttggctttatatatcttgtggaaaggacgaaacaccgcagctgatataatggagaacgaattctccatta
 tatcagctgttttttaggaaccctagtgtgaggtggccactccctctctgcgcgctcgtcgtcactgagggc
 gggcgaccaaaaggtcgcccgacggcggttggccggcgccctcagtgagcgagcgagcgcgagctgc
 35

Hes1 –KOP (SEQ ID NO: 80)
 gcagctgatataatggagaa

SV40pA cDNA sequence (SEQ ID NO: 70)

gctgatcagcctcgataagatacattgatgagtttggacaaaccacaactagaatgcagtgaaaaaatgctttatt
tgtgaaatttgatgctattgctttatttgtaaccattataagctgcaataaacaagtt

5 U6 cDNA sequence (SEQ ID NO: 71)

aaggctcgggcaggaagagggcctatttcccatgattccttcattttgcatatacgatacaaggctgtagagagat
aattagaattaatttgactgtaaacacaaagatattagtagacaaaatacgtgacgtagaaagtaataatttcttggt
agtttgcagttttaaaattatgttttaaaatggactatcatatgcttaccgtaacttgaaagtatttcgatttcttg
gctttatatacttctgtggaaaggacgaaacacc

10

U6 cDNA sequence (SEQ ID NO: 84)

CGGTGTTTCGTCCTTTCCACAAGATATATAAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAG
TCCATTTTAAAACATAATTTTAAAACACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACGTATTTGTACTAA
TATCTTTGTGTTTACAGTCAAATTAATTCTAATTATCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAA
15 TCATGGGAAATAGGCCCTCTTCCTGCCCGACC

siRNA cDNA sequence (SEQ ID NO: 72)

(N)₂₀CGAA(N)₂₀TTTTTT

20 In some embodiments of any of the compositions described herein, the vector comprises
or consists of pITR-CMV-mScarlet-bGHpA (SEQ ID NO: 76). In some embodiments of any of
the compositions described herein, the vector comprises a sequence that has at least 75% (e.g., at
least 80%, at least 82%, at least 84%, at least 85%, at least 86%, at least 88%, at least 90%, at
least 92%, at least 94%, at least 95%, at least 96%, at least 98%, at least 99%) sequence identity
25 to SEQ ID NO: 76.

pITR-CMV-mScarlet (SEQ ID NO: 76)

cctgcaggcagctgcgcgctcgcctcactgaggccgcccggcgctcgggcgacctttggtcgccggcctcagt
gagcgagcgagcgcgagagagggagtggtgccaactccatcactaggggttcctgtgatgcggttttggcagtagatc
30 aatgggctgtgatagcggtttgactcacggggatttccaagtctccacccattgacgtcaatgggagtttgttttg
gcacaaaaatcaacgggactttccaaaatgtcgtgaacaaactccgccccattgacgcaaatgggcggtaggcgtgtac
ggtgggaggtctatataagcagagctcgttttagtgaaccgtcagatcgcttgagacgcgtgagcaagggcgaggca
gtgatcaaggagttcatgcggttcaaggtgcacatggagggctccatgaacggccacgagttcgagatcgagggcga
gggagggggccgcccctacgagggcaccagaccgccaagctgaaggtgaccaaggggtggccccctgcccttctcct
35 gggacatcctgtcccctcagttcatgtacgggtccagggccttcatcaagcaccgcccgcacatccccgactactat
aagcagtccttccccgagggcttcaagtggtgagcgctgatgaacttcgaggacggcgccgctgaccgtgaccca
ggacacctccctggaggacggcaccctgatctacaaggtgaagctccgcgccaccaacttccctcctgacggccccg
taatgcagaagaagacaatgggctgggaagcgtccaccgagcggttgtaacccgaggacggcgctgtaagggcgac

attaagatggccctgcgctgaaggacggcgccgctacctggcggacttcaagaccacctacaaggccaagaagcc
 cgtgcagatgccccggcgctacaacgtcgaccgcaagttggacatcacctcccacaacgaggactacaccgtggtgg
 aacagtacgaacgctccgagggccgcccactccaccggcgcatggacgagctgtacaagtaagctgatcagcctcga
 ctgtgccttctagttgccagccatctgttgtttgcccctccccctgcttcttgaccctggaaggtgccactccc
 5 actgtcctttcctaataaaaatgaggaaattgcattgcattgtctgagtaggtgtcattctattctggtgggtgggt
 ggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggaggaac
 ccctagtgtgaggttgccactccctctctgcgctcgctcgctcactgaggccggcgaccaaaggtcgcccgga
 cgcccggtttgcccggcgccctcagtgcgagcgagcgcgagctgcctgcagg

10 In some embodiments of any of the compositions described herein, the vector comprises
 or consists of pITR-CMV-mScarlet-DD-bGHpA (SEQ ID NO: 77). In some embodiments of
 any of the compositions described herein, the vector comprises a sequence that has at least 75%
 (e.g., at least 80%, at least 82%, at least 84%, at least 85%, at least 86%, at least 88%, at least
 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 98%, at least 99%) sequence
 15 identity to SEQ ID NO: 77.

pITR-CMV-mScarlet (SEQ ID NO: 77)

cctgcaggcagctgcgcgctcgctcgctcactgaggccgccccggcgctcgggcgacctttggtcgcccgccctcagt
 gagcgagcgagcgcgagagaggagtggccaaactccatcactaggggttcctgtgatgcgggttttggcagtacatc
 20 aatgggcgtggatagcggtttgactcacgggatttccaagtctccacccattgacgtcaatgggagtttgttttg
 gcacaaaaatcaacgggactttccaaaatgtcgtaaaactccgccccattgacgcaaatggcggttaggcgtgtac
 ggtgggaggtctatataagcagagctcgtttagtgaaccgtcagatcgctggagacgcgtgagcaagggcgaggca
 gtgatcaaggagttcatgcggttcaaggtgcacatggagggctccatgaacggccacgagttcgagatcgagggcga
 gggcgagggccgcccctacgagggcaccagaccgccaagctgaaggtgaccaaggtggccccctgcccttctcct
 25 gggacatcctgtcccctcagttcatgtacggctccagggccttcatcaagcaccgcccagacatccccgactactat
 aagcagtccttccccgagggcttcaagtgggagcgctgatgaacttcgaggacggcgccgctgaccgtgaccca
 ggacacctccctggaggacggcaccctgatctacaaggtgaagctccgcgccaccaacttccctcctgacggccccg
 taatgcagaagaagacaatgggctgggaagcgctccaccgagcggttgtaacccgaggacggcggtgctgaagggcgac
 attaagatggccctgcgctgaaggacggcgccgctacctggcggacttcaagaccacctacaaggccaagaagcc
 30 cgtgcagatgccccggcgctacaacgtcgaccgcaagttggacatcacctcccacaacgaggactacaccgtggtgg
 aacagtacgaacgctccgagggccgcccactccaccggcgcatggacgagctgtacaagtaaatcagctctgattgcg
 gcgttagcggttagattacgttatcgcatggaaaacgcatgccgtggaacctgcctgccgatctcgcttggtttaa
 acgcaacaccttaataaaacccgtgattatgggcccgcatacctgggaatcaatcggtcgtccgttgccaggacgca
 aaaatattatcctcagcagtaaccgagtagcgacgatcgcgtaacgtgggtgaagtcggtggatgaagccatcgcg
 35 gcgtgtggtgacgtaccagaaatcatggtgattggcgccggtcgcggttattgaacagttcttgccaaaagcgcaaaa
 actgtatctgacgcataatcgacgcagaagtgaaggcgacacccatttcccgattacgagccggatgactgggaat

cggatttcagcgaattccacgatgctgatgcgacagaactctcacagctattgctttgagattctggagcggcgagct
 gatcagcctcgactgtgccttctagttgccagccatctgttggtttgcccctcccccgctgccttccttgaccctggaa
 ggtgccactcccactgtcctttcctaataaaaatgaggaaattgcatcgcatgtgtctgagtaggtgtcattctattct
 ggggggtgggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcggtgggct
 5 ctatggaggaaccctagtgatggagttggccactccctctctgcgcgctcgctcgctcactgaggccggggcgacca
 aaggtcgcccgcacgcccgggctttgcccggggcgccctcagtgagcgagcgagcgcgagctgcctgcagg

In some embodiments of any of the compositions described herein, the vector comprises
 or consists of pITR-CMV-mScarlet (SEQ ID NO: 81). In some embodiments of any of the
 10 compositions described herein, the vector comprises a sequence that has at least 75% (e.g., at
 least 80%, at least 82%, at least 84%, at least 85%, at least 86%, at least 88%, at least 90%, at
 least 92%, at least 94%, at least 95%, at least 96%, at least 98%, at least 99%) sequence identity
 to SEQ ID NO: 81.

15 pITR-CMV-mScarlet (SEQ ID NO: 81)

cctgcaggcagctgcgcgctcgctcgctcactgaggccgcccggggcgctcgggcgacctttggtcgcccggccctcagt
 gagcgagcgagcgcgagagaggagtgcccaactccatcactaggggttcctgcgggccgcacgcgtctagatccca
 tatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccacgacccccgcccattgac
 gtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggagtatattacggt
 20 aaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatgg
 cccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcacgc
 tattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacgggggatttccaagtc
 tccacccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactcc
 gccccattgacgcaaattggcggttaggcgtgtacgggtgggaggtctatataagcagagctcggttagtgaaccgtca
 25 gatcgcttgagagcgacccggtgccaccatgcctaagaagaagcggaagtcgggtccggcggtgagcaagggcgagg
 cagtgatcaaggagttcatgcggttcaagggtgcacatggagggtccatgaacggccacgagttcgagatcgagggc
 gagggcgagggccgcccctacgagggcacccagaccgccaagctgaaggtgaccaaggggtggccccctgcccttctc
 ctgggacatcctgtcccctcagttcatgtacgggtccagggccttcatcaagcaccgcccgacatccccgactact
 ataagcagtccttccccgagggcttcaagtgggagcgcggtgatgaacttcgaggacggcgcgccgtgaccgtgacc
 30 caggacacctccctggaggacggcacccctgatctacaaggtgaagctccgcggcaccaacttccctcctgacggccc
 cgtaatgcagaagaagacaatgggctgggaagcgtccaccgagcggttgtagccccgaggacggcggtgctgaagggcg
 acattaagatggccctgcgccctgaaggacggcgccgctacctggcggaacttcaagaccacctacaaggccaagaag
 cccgtgcagatgcccggcgctacaacgtcgaccgcaagttggacatcacctcccacaacaggactacaccgtggt
 ggaacagtacgaacgctccgagggccgcccactccaccggcgcatggacgagctgtacaagtaataagagctcgctg
 35 atcagcctcgactgtgccttctagttgccagccatctgttggtttgcccctcccccgctgccttccttgaccctggaag
 gtgccactcccactgtcctttcctaataaaaatgaggaaattgcatcgcatgtgtctgagtaggtgtcattctattctg

gggggtggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctc
 tatggaagcttgaattcagctgacgtgcctcggaccgctaggaacccctagtgatggagttggccactccctctctg
 cgcgctcgcctcgcctcactgaggccggcgaccaaaggctgccccgacgcccgggctttgccccggcgccctcagtgag
 cgagcgagcgcgagctgcctgcagggcgccctgatgcggtatcttctccttacgcatctgtgcggtatttcacacc
 5 gcatacgtcaaagcaaccatagtagcgccctgtagcggcgcatataagcgcgggcggtgtggtggttacgcgcagcg
 tgaccgctacacttgccagcgccctagcgcgcgctcctttcgctttcttcccttcctttctcgccacgttcgcgggc
 tttccccgtcaagctctaatacgggggctccctttagggttccgatttagtgctttacggcacctcgacccccaaaa
 acttgatttgggtgatggttcacgtagtgggccatcgccctgatagacgggttttgcgcctttgacgttggagtcca
 cggttctttaatagtggactcttggttccaaactggaacaacactcaaccctatctcgggctattcttttgatttataa
 10 gggattttgcccatttcggcctattgggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaaat
 attaacgtttacaattttatgggtgcaactctcagtacaatctgctctgatgccgcatagttaagccagccccgacacc
 cgccaacacccgctgacgcgcctgacgggcttgctctgctccggcatccgcttacagacaagctgtgacggctctcc
 gggagctgcatgtgtcagaggttttcaccgctcatcaccgaaacgcgcgagacgaaagggcctcgtgatacgcctatt
 tttataggttaatgtcatgaacaataaaactgtctgcttacataaacagtaatacaaggggtgttatgagccatatt
 15 caacgggaaacgtcgaggccgcgattaaattccaacatggatgctgatttatatgggtataaatgggctcgcgataa
 tgtcgggcaatcaggtgcgacaatctatcgcttgatgggaagcccgatgcgccagagttgtttctgaaacatggca
 aaggtagcgttgccaatgatgttacagatgagatgggtcagactaaactggctgacggaatttatgcctcttccgacc
 atcaagcattttatccgtactcctgatgatgcatggttactcaccactgcgatccccgaaaaacagcattccaggt
 attagaagaatatcctgattcaggtgaaaatattgttgatgcgctggcagtggtcctgcgccggttgcatcgcattc
 20 ctgtttgtaatgtccttttaacagcgatcgcgattttcgtctcgcctcagggcgcaatcacgaatgaataacggtttg
 gttgatgcgagtgattttgatgacgagcgtaatggctggcctgttgaaacagtcctggaaagaaatgcataaactttt
 gccattctcaccggattcagtcgctcactcatggtgattttctcacttgataaccttatttttgacgaggggaaattaa
 taggttgattgatgttgacgagtcggaatcgacagaccgataaccaggatcttgccatcctatggaactgcctcgggt
 gagttttctccttcattacagaaacggctttttcaaaaatatgggtattgataatcctgatatgaataaattgcagtt
 25 tcatattgatgctcgatgagtttttctaactctcatgacaaaatcccttaacgtgagttttcgttccactgagcgtca
 gaccccgtagaaaagatcaaaggatcttcttgagatccttttttctgcgcgtaatctgctgcttgcaaaacaaaaaa
 accaccgctaccagcggtgggtttgtttgccggatcaagagctaccaactcttttccgaaggtaactggcttcagca
 gagcgagataccaaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcct
 acatacctcgctctgctaatacctgttaccagtggtgctgctgacagtgagcagataagtcgtgtcttaccgggttgactc
 30 aagacgatagttaccggataagggcgacggtcggttgacggggggttcgtgcacacagcccagcttgagcgaa
 cgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaaggagaaaaggcggac
 aggtatccggtaagcggcagggctcggaacaggagagcgacagggagcttccagggggaaacgccttggtatcttta
 tagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcagggggcgagcctatgga
 aaaacgcccagcaacgcggcctttttacggttccctggccttttgccttttgcctcacatgt

35

In some embodiments of any of the compositions described herein, the vector comprises or consists of pITR-CMV-mScarlet-DD (SEQ ID NO: 82). In some embodiments of any of the

compositions described herein, the vector comprises a sequence that has at least 75% (e.g., at least 80%, at least 82%, at least 84%, at least 85%, at least 86%, at least 88%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 98%, at least 99%) sequence identity to SEQ ID NO: 82.

5

pITR-CMV-mScarlet-DD (SEQ ID NO: 82)

cctgcaggcagctgcgcgctcgcctcactgaggccgcccggcgctcgggcgacctttggctcgcccgccctcagt
gagcgagcgagcgcgagagaggagtgcccaactccatcactaggggttcctgcggccgcacgcgtctagatccca
tatatggagttccgcgttacataacttacggtaaattggccgcctggctgacgcgccaacgacccccgcccattgac
10 gtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggagtatttacggg
aaactgcccacttggcagtacatcaagtgtatcatatgccaaagtacgccccctattgacgtcaatgacggtaaatgg
ccgcctggcattatgccagtacatgaccttatgggactttcctacttggcagtacatctacgtatttagtcatcgc
tattaccatgggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtc
tccacccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactcc
15 gccccattgacgcaaattgggcggtaggcgtgtacgggtgggaggtctatataagcagagctcggttagtgaaccgtca
gatcgctggagacgcaccggtgccaccatgcctaagaagaagcgaaagtcggctccggcgtgagcaaggggcgagg
cagtgtcaaggagttcatgcggttcaagggtgcacatggagggctccatgaacggccacgagttcgagatcgagggc
gagggcgagggccgcccctacgagggcaccacagaccgccaagctgaagggtgaccaagggtggccccctgcccttctc
ctgggacatcctgtcccctcagttcatgtacggctccagggccttcatcaagcaccgcccgcacatccccgactact
20 ataagcagtccttccccgagggttcaagtgggagcgcggtgatgaacttcgaggacggcgccgctgaccgtgacc
caggacacctccctggaggacggcaccctgatctacaagggtgaagctccgcggcaccaacttccctcctgacggccc
cgtaatgcagaagaagacaatgggctgggaagcggtccaccgagcggttgtaacccgaggacggcggtgctgaaggcg
acattaagatggccctgcgcctgaaggacggcgccgctacctggcggaacttcaagaccacctacaaggccaagaag
ccggtgcagatggccggcgctacaacgtcgaccgcaagttggacatcacctcccacaacgaggactacaccgtggt
25 ggaacagtacgaacgctccgagggccgcccactccaccggcgcatggacgagctgtacaagggtaccatcagtcgtga
ttgcggcggttagcggttagattacgttatcggcgtggaacacgcatgcccgtggaacctgcctgccgatctgcctgg
tttaaacgcaacaccttaataaaacccgtgattatgggcccatacctgggaatcaatcggtcgtccgttgccagg
acgcaaaaatattatcctcagcagtcacccgagtagcgacgatcgcgtaacgtgggtgaagtcgggtggatgaagcca
tcgcggcggtgtggtgacgtaccagaaatcatggtgattggcggcggtcgcggttattgaacagttcttgccaaaagcg
30 caaaaactgtatctgacgcataatcgacgcagaagtgaaggcgacaccatttcccggattacgagccggatgactg
ggaatcggtattcagcgaattccacgatgctgatgcgagaactctcacagctattgctttgagattctggagcggc
gataataagagctcgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctccccctgtgc
cttccttgaccctggaagggtgccactcccactgtcctttcctaataaaaatgaggaaattgcatcgcatgtctgagt
aggtgtcattctattctgggggggtgggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgc
35 tggggatgcggtgggctctatggaagcttgaattcagctgacgtgcctcgaccgctaggaacccctagtgtgag
ttggccactccctctctgcgcgctcgcctcactgaggccggcgaccaaaggtcgcccgacgcccgggcttttgc
ccgggcccctcagtgcgagcgagcgcgagctgcctgcagggggcgctgatgcggtattttctccttacgcac

tgtgcggtatttcacaccgcatacgtcaaagcaaccatagtagcgccctgtagcggcgcatataagcgcggcggtg
 tgggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttccttt
 ctgcgcacggttcgcccgttttccccgtcaagctctaaatcgggggctccctttaggggtccgatttagtgctttacg
 gcacctcgacccccaaaaaacttgatttgggtgatggttcacgtagtgggccatcgccctgatagacgggtttttcgcc
 5 ctttgacggttgaggtccacggttctttaatagtggtactcttggttccaaactggaacaacactcaaccctatctcgggc
 tattcttttgatttataagggattttgccgatttcggcctattgggttaaaaaatgagctgatttaacaaaaatttaa
 cgcgaaattttaacaaaatattaacggtttacaattttatgggtgactctcagtacaatctgctctgatgccgcatagt
 taagccagccccgacaccgccaacaccgctgacgcgcctgacgggcttgctctgctcccggcatccgcttacaga
 caagctgtgaccgtctccgggagctgcatgtgtcagagggttttcaccgctcatcaccgaaacgcgcgagacgaaaggg
 10 cctcgtgatacgcctatttttataggttaatgtcatgaacaataaaactgtctgcttacataaacagtaatacaagg
 ggtgttatgagccatattcaacgggaaacgtcgaggccgcgattaaattccaacatggatgctgatttatatgggtta
 taaatgggctcgcgataatgtcgggcaatcaggtgcgacaatctatcgcttgatgggaagcccgatgcgccagagt
 tgtttctgaaacatggcaaaggtagcgttgccaatgatgttacagatgagatgggtcagactaaactggctgacggaa
 tttatgcctcttccgaccatcaagcattttatccgtactcctgatgatgcatgggttactcaccactgcgatccccgg
 15 aaaaacagcattccaggtattagaagaatatcctgattcaggtgaaaatattgttgatgcgctggcagtggttctctgc
 gccggttgcatcgcattcctgtttgtaattgtccttttaacagcgatcgcgatatttcgtctcgctcaggcgcaatca
 cgaatgaataacgggttggttgatgcgagtgattttgatgacgagcgtaatggctggcctgttgaaacaagtctggaa
 agaaatgcataaacttttgccattctcaccggattcagtcgctcactcatgggtgattttctcacttgataaccttattt
 ttgacgaggggaaattaataggttgattgatgttgacgagtcggaatcgagaccgataaccaggatcttgccatc
 20 ctatggaaactgcctcggtgagttttctccttcattacagaaacggctttttcaaaaatatggtattgataatcctga
 tatgaataaattgcagttttcatttgatgctcgatgagtttttctaattctcatgacccaaaatcccttaacgtgagttt
 tcggtccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatccttttttctgcgcgtaatctg
 ctgcttgcaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccga
 aggttaactggcttcagcagagcgcagataccaaataactgtccttctagtgtagccgtagtttaggccaccacttcaag
 25 aactctgtagcaccgcctacatacctcgctctgctaactcctgttaccagtggtgctgccagtgggcgataagtcgtg
 tcttaccgggttgactcaagacgatagttaccggataaggcgcagcggctcgggctgaacggggggttcgtgcacac
 agcccagcttgagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcggccacgcttccc
 gaagggagaaaaggcggacaggtatccggtaagcggcagggctcggaacaggagagcgcacgaggagcttccaggggg
 aaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcag
 30 gggggcggagcctatggaaaaacgccagcaacgcggcctttttacgggttcttgcccttttgctggccttttgctcac
 atgt

In some embodiments of any of the compositions described herein, the vector comprises
 or consists of pITR-U6-shHES1-S5-CMV-3xFLAG-hATOH1-DD-T2A-hPOU4F3-U6-shHES1-
 35 S3 (SEQ ID NO: 83). In some embodiments of any of the compositions described herein, the
 vector comprises a sequence that has at least 75% (e.g., at least 80%, at least 82%, at least 84%,

at least 85%, at least 86%, at least 88%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 98%, at least 99%) sequence identity to SEQ ID NO: 83.

pITR-U6-shHES1-S5-CMV-3xFLAG-hATOH1-DD-T2A-hPOU4F3-U6-shHES1-S3 (SEQ ID NO: 83)

5 cctgcaggcagctgcgcgctcgcctcactgaggccgcccggcgctcggcgacctttggctcgcccggcctcagt
gagcgagcgagcgcgagagagggagtgccaaactccatcactaggggttcctgcgccgcaaaaaaactgcatgac
ccagatcaattcgttgatctgggtcatgcagtcggtgtttcgtcctttccacaagatatataaagccaagaaatcga
aatactttcaagttacggtaagcatatgatagtcatttttaaaacataattttaaaactgcaaactacccaagaaat
10 tattactttctacgtcacgtattttgtactaatatctttgtgtttacagtcaaattaattctaattatctcttaac
agccttgatatgtatatgcaaatatgaaggaatcatgggaaataggccctcttctgcccagaccacgcgtctagatc
ccatatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccacgacccccgcccatt
gacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggagtatattac
ggtaaaactgcccacttggcagtagatcaagtgtatcatatgccagtagccccctattgacgtcaatgacggtaaa
15 tggcccgctggcattatgccagtagatgaccttatgggactttcctacttggcagtagatctacgtattagtcac
cgctattaccatgggtgatgcggttttggcagtagatcaatgggcgtggatagcggtttgactcacggggattttccaa
gtctccacccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtacaac
tccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctaccggtgccaccat
ggattacaaggatgacgacgataaggactataaggacgatgatgacaaggactacaaagatgatgacgataaagtta
20 actcccgctgctgcatgcagaagagtggtgtaagtgaaggagttgggagaccaccatcgccagccccagccgcat
catctcccgcaaccgcccgcgcgcgcgcagccacctgcaactttgcaggcgagagagcatcccgtctaccgcctga
gctgtccctcctggacagcaccgacccacgcgcctggctggctcccactttgcagggcatctgcacggcacgcgcgcg
cccagtatattgctacattccccggagctgggtgcctcagaggccgctgcgccccgggacgaggtggacggccggggg
gagctggttaaggaggagcagcgcggtgccagcagcagcaagagccccgggcccgtgaaagtgcgggaacagctgtg
25 caagctgaaaggcgggggtggtgtagacgagctgggtgcagccgcaacggggcccttccagcaaacaggtgaatg
gggtgcagaagcagagacggctagcagccaacgcccaggagcggcgcaggatgcatgggtgaaccacgccttcgac
cagctgcgcaatgttatcccgctgttcaacaacgacaagaagctgtccaaatatgagaccctgcagatggcccaaat
ctacatcaacgccttgtccgagctgctacaaacgcccagcggaggggaacagccaccgcccctccagcctcctgca
aaagcgaccaccaccaccttcgacccgcgccctcctatgaagggggcgcgggcaacgcgaccgcagctggggctcag
30 caggcttcggagggagccagcgccgacccccgcccggagttgccggactcgttctcagccccagcttctgcggg
agggtactcgggtgcagctggacgtctgcacttctcgaactttcgaggacagcgccctgacagcgatgatggcgcaaa
agaatttgtctccttctctccccgggagcatcttgcagccagtgcaggaggaaaacagcaaaacttcgcctcgggtcc
cacagaagcgacggggaattttccccccattcccattacagtgactcggatgaggcaagtgttaacatcagttctgat
tgcggcgttagcggtagattacgttatcggcatggaaaacgccatgccgtggaacctgcctgccgatctcgcttggt
35 ttaaacgcaacaccttaataaaaccggtgattatgggccgccatacctgggaatcaatcggtcgtccgttgccagga
cgcaaaaatattatcctcagcagtaaccgagtagcgacgatcgcgtaacgtgggtgaagtgcgttgatgaagccat
cgcggcgtgtggtgacgtaccagaaatcatggtgattggcgggcggtcgcggtattgaacagttcttgccaaaagcgc

aaaaactgtatctgacgcataatcgacgcagaagtggaaaggcgacacccatttcccgattaccagccggatgactgg
gaatcgggtatttcagcgaattccacgatgctgatgcgcagaactctcacagctattgctttgagattctggagcggcg
aggatccggcgagggcgagaggaagtctgctaacatgcggtgacgtcgaggagaatcctggcccaatgatggccatga
actccaagcagcctttcggcatgcaccgggtgctgcaagaacccaaattctccagttctgcactctggctccgaggcc
5 atgcgccgagttctgtctcccagccccgcagctgcagggtaatatatttggaaagctttgatgagagcctgctggcacg
cgccgaagctctggcgggcggtggatatcgtctcccacggcaagaacctccgttcaagcccgcgcccacctaccata
ccatgagcagcgtgccctgcacgtccacttcgtccaccgtgcccatctcccaccagctgcgctcacctcacacct
caccacgccgtgcaccagggcctcgaaggcgacctgctggagcacatctcgccacgctgagtgtagcgggcctggg
cgctccggaacactcgggtgatgccgcacagatccatccacaccacctgggcgccatggggccacctgcaccaggcca
10 tgggcatgagtcacccgcacaccgtggccctcatagcgccatgcctgcatgcctcagcgacgtggagtcagaccg
cgcgagctggaagccttcgcccagcgttcaagcagcggcgcatcaagctgggggtgaccaggcgagcgtgggcgc
ggctctggctaattctcaagatccccggcggtgggctcgctgagccaaagcaccatctgcaggttcgagttctctactc
tctcgcaacaacatgatcgctctcaagccggtgctccaggcctggttggaggaggccgagggccctaccgagag
aagaacagcaagccagagctcttcaacggcagcgaacggaagcgcaaacgcacgtccatcgcgggcgccggagaagcg
15 ttcactcgaggcctatttgcgtatccagccacgtccttcattctgagaagatcgcgcccatcgctgagaaactggacc
ttaaaaagaacgtggtgagagttctggttctgcaaccagagacagaaacagaaacgaatgaagtattcggtgtccac
taaataataaaaatatctttattttcattacatctgtgtgttggttttttgtgtgttaataaaaaaaagaaagtcac
caaagcctatttcgataggctttgatgactttcggtgtttcgtcctttccacaagatatataaaggccaagaaatcga
aatactttcaagttacggtaagcatatgatagtcatttttaaacataatttttaaaactgcaaaactacccaagaaat
20 tattactttctacgtcacgtattttgtactaatatctttgtgtttacagtcaaattaatttctaattatctctctaac
agccttgatatcgtatatgcaaatatgaaggaatcatgggaaataggccctcttctgcccgaacccggaccgctagga
accctagtgtatggagttggccactccctctctgcgcgtcgctcgctcactgaggccggggcgaccaaaggtcgccc
gacgcccgggctttgcccggggcgccctcagtgagcgagcgagcgcgagctgcctgcagg

25 U6 cDNA sequence (SEQ ID NO: 91)

25 **U6 cDNA sequence (SEQ ID NO: 31)**
Ggtgttttcgtccttttcacacaagatatataaagccaagaaatcgaaatactttcaagttacggttaagcatatgatagt
ccatttttaaacaataattttaaaactgcaaaactaccaagaaattattacttttctacgtcacgtattttgtactaat
atctttgtgtttacagtcaaattaattctaattatctctctaacagccttgatcgtatatgcaaatatgaaggaat
catgggaaataggccctcttctcgtcccgacc

shHES1-1 (SEQ ID NO: 85)

cggtgtttcgtcctttccacaagatatataaagccaagaaatcgaaatactttcaagttacggtaagcatatgatag
tccattttaaaacataattttaaaactgcaaaactaccaagaaattattactttctacgtcacgtattttgactaa
tattctttgtgtttacagtcaaattaatttctaattatctctctaacagccttgatatcgatatgcaaatatgaaggaa
tcattgggaaataggccctcttctctgccgacc

atggattacaaggatgacgacgataaggactataaggacgatgatgacaaggactacaaagatgatgacgataaa

5 gttaactcccgcctgctgcatgcagaagagtgggctgaagtgaaggagtgtgggagaccaccatcgccagccccagcc
gcatacatctcccgaaccgccgcgcgcgcagccacctgcaactttgcaggcgagagagcatcccgtctaccgc
ctgagctgtccctcctggacagcacccgacccacgcgcctggctggctcccactttgcagggcatctgcacggcacgc
gccgcccagtatattgctacattccccggagctgggtgcctcagaggccgctgcgccccgggacgaggtggacggccg
gggggagctggtaaggaggagcagcggcggtgccagcagcagcaagagccccgggcccgggtgaaagtgcgggaacagc
10 tgtgcaagctgaaaggcgggggtggtggtagacgagctgggctgcagccgccaacgggccccttcagcaaacaggtg
aatgggggtgcagaagcagagacggctagcagccaacgccaggggagcggcgaggatgcatgggctgaaccacgcctt
cgaccagctgcgcaatgttatcccgtcgttcaacaacgacaagaagctgtccaaatatgagaccctgcagatggccc
aaatctacatcaacgccttgtccgagctgctacaaacgcccagcggagggggaacagccaccgccgcctccagcctcc
tgcaaaagcgaccaccaccaccttcgcaccgcggcctcctatgaagggggcgcggggaacgcgaccgcagctggggc
15 tcagcaggcttccggaggggagccagcggccgaccccgccgggagttgcgggaactcgcttctcagccccagcttctg
cgggagggtaactcgggtgcagctggacgctctgcacttctcgacttctcagagacagcgccctgacagcgatgatggcg
caaaagaatttgtctccttctctccccgggagcatcttgagccagtgaggaggaaaacagcaaaacttcgcctcg
gtccacagaagcgacggggaattttccccccattcccattacagtgactcggatgaggcaagt

ataataaaatatctttattttcattacatctgtgtgttggttttttggtg

25 ggtgtttcgtcctttccacaagatatataaagccaagaaatcgaaatactttcaagttacggtaagcatatgatagt
ccattttaaaacataattttaaaactgcaaactacccaagaaattattactttctacgtcacgtattttgactaat
atctttgtgtttacagtc aaatttaattcttaattatctctctaacagccttgatatcgtaatatgcaaatatgaaggaat
catgggaaataggccctcttctcgccgacc

35

**pITR-U6-shHES1-S5, hATOHessps-3xFLAG-hATOH1-T2A-hPOU4F3-US-shHES1-S3
(SEQ ID NO: 93)**

cctgcaggcagctgcgcgctcgctcgctcactgaggccgcccgggcgctcgggcgacctttggctcgcccggcctcagt
gagcgagcgagcgcgagagagggagtgccaaactccatcactaggggttcctgcgggcgcaaaaaaactgcatgac
5 ccagatcaattcggtgatctgggtcatgcagtcggtgtttcgtcctttccacaagatatataaagccaagaaatcga
aatactttcaagttacggtaagcatatgatagtcattttaaaacataattttaaaactgcaaactaccaagaat
tattactttctacgtcacgtattttgtactaatatctttgtgtttacagtcaaattaattctaattatctctaac
agccttgatatcgatatatgaaatatgaaggaatcatgggaaataggccctcttctgcccagaccacgcgtctatgga
gtttg'cataacaaacggtttggcagctcgctctcttacctccattaacaagctgtaacatatagctgcaggttgcta
10 taatctcattaataattttggaaacttgaatataggatatttctgagtgctcattccccatatgccagccacttctgc
catgctgactgggttcctttctctccattattagcaattagcttcttaccttccaaagtcagatccaaggtatccaag
atactagcaaaggaatcaactatgtgtgcaagttaagcatgcttaatatcacccaaacaaacaaagaggcagcattt
cttaaagtaatgaagatagataaatcggttagtcctttgcgacactgctggtgctttctagagttttatatatttt
aagcagcttgctttatattctgtctttgcctcccacccccaccagcacttttatttggagggttttggctcgccac
15 actttgggaaacttatttgatttcacggagagctgaaggaagatcatttttggcaacagacaagtttaaacacgatt
tctatgggacattgctaactggggcccctaaggagaaaggggaaactgagcggagaatgggttaaactccttggaagc
aggggagaggcaggggaggagagaagtcggaggagtataaagaaaaggacaggaaccaagaagcgtgggggtggttt
gccgtaatgtgagtggttcttaattagagaacggttgacaatagagggtctggcagaggctcctggccgcggtgcgg
agcgtctggagcggagcacgcgtgtcagctggtgagcgcactctcctttcaggcagctccccggggagctgtgcgg
20 ccacatttaacaccatcatcaccctccccggcctcctcaacctcggcctcctcctcgctcgacagccttcttggcc
cccaccagcagagctcacagtagcgagcgtctctcgccgtctcccgcaactcggccggggcctctctctccccagc
tgcgagcggggagccgcaactgccactgcacctcccagcaaccagcccagcacgcaaagaagctgcgcaaagttaa
agccaagcaatgccaaggggaggggaagctggaggcgggctttgagtggttctgggcgctggcgggtccagaatc
gccagagccgcccgcggtcggtgcacatctgacccgagtcagcttgggcaccagccgagagccggctccgcaccgct
25 cccgcaccccagccgcccgggtggtgacacacaccggagtcgaattacagccctgcaattaacatatgaatctgacg
aatttaaaagaaggaaaaaaaaaaaaaaaaacctgagcaggcttgggagtcctctgcacacaagaacttttctcggggt
gtaaaaactctttgattggctgctgcgacgcgctgcccgcgcccctccattggctgagaagacacgcgaccggcgcg
aggagggggttgggagaggagcggggggagactgagtgcgcgctgccgctttttaaggggcgagcgccttcagca
accggagaagcatagttgcacgcgacctggtgtgtgatctccgagtggtgggggagggctgaggagggaaaaaaaa
30 ataagacgttgagaagagacccggaaaggccttttttttgggtgagctggtgtcccagtgctgcctccgatcctg
agcctccgagcctttgagtgcaaccggtgccaccatggattacaaggatgacgacgataaggactataaggacgat
gatgacaaggactacaaagatgatgacgataaagttaacatgtcccgctgctgcatgcagaagagtgggctgaagt
gaaggagttgggagaccaccatcgccagccccagccgatcatctcccgcaaccgcccgcgcccgcagccacctg
caactttgcaggcgagagagcatcccgtctaccgcctgagctgtccctcctggacagcaccgacccacgcgcctgg
35 ctggctcccactttgcagggcacgtgcacggcacgcgcccagctatttgctacattccccggagctgggtgcctc
agaggccgctgcgccccgggacgaggtggacggccggggggagctggtaaggaggagcagcggcggtgccagcagca
gcaagagccccgggcccgtgaaagtgcgggaacagctgtgcaagctgaaaggcgggggtggtggtagacgagctgggc
tgagccgccaacggggccccttcagcaaacaggtgaatggggtgcagaagcagagacggctagcagccaacgccag

ggagcggcgagcagatgcatgggctgaaccacgccttcgaccagctgcgcaatgttatcccgctcgttcaacaacgaca
 agaagctgtccaaatatgagaccctgcagatggcccaaattctacatcaacgccttggtccgagctgctacaaacgccc
 agcggaggggaacagccaccgcccgcctccagcctcctgcaaaagcgaccaccaccaccttcgaccgcgccctccta
 tgaagggggcgcgggcaacgcgaccgcagctggggctcagcaggcttccggaggagccagcgccgaccccgccc
 5 ggagttgccggactcgcttctcagccccagcttctgcgggaggggtactcgggtgcagctggacgctctgcacttctcg
 acttttcgaggacagcgccctgacagcgatgatggcgcaaaagaatttgctctccttctctccccgggagcatcttgca
 gccagtgcaggaggaaaacagcaaaacttcgcctcgggtcccacagaagcgacggggaattttccccccattccatt
 acagtgactcggatgaggcaagtgttaacgagggcagagggaagtctgctaacatgcgggtgacgtcgaggagaatcct
 ggcccaatgatggccatgaactccaagcagcctttcgcatgcacccgggtgctgcaagaacccaaattctccagtct
 10 gcactctggctccgaggccatgcgccgagctctgtctcccagccccgcagctgcagggtaatatatttggaagctttg
 atgagagcctgctggcagcgccgaagctctggcggggtggatatcgtctcccacggcaagaacccatccgttcaag
 cccgacgccacctaccataccatgagcagcgtgccctgcacgtccacttcgtccaccgtgcccatctcccaccacgc
 tgcgctcacctcacaccctcaccacgcggtgcaccagggcctcgaaggcgacctgctggagcacatctcgccacgc
 tgagtgtgagcggcctggcgctccggaacactcgggtgatgcccgacagatccatccacaccacctggcgccatg
 15 ggccacctgcaccaggccatgggcatgagtcacccgcacaccgtggcccctcatagcgccatgcctgcatgcctcag
 cgacgtggagtcagacccgcgcgagctggaagccttcgccgagcgcttcaagcagcgcgcatcaagctgggggtga
 cccaggcggacgtggcgcggtcttggttaattctcaagatccccggcggtgggtcgcgtgagccaaagcaccatctgc
 aggttcgagttctctcactctctcgcaacaacatgatcgtctctcaagccggtgctccaggcctgggtggaggaggc
 cgaggccgcctaccgagagaagaacagcaagccagagctcttcaacggcagcgaacggaagcgcaaacgcacgtcca
 20 tcgcgggcgccggagaagcgttccactcgaggccctatttcgctatccagccacgtccttcatctgagaagatcgcgggc
 atcgcgtgagaaactggaccttaaaaaagaacgtgggtgagagctctgggtctgcaaccagagacagaaacagaaacgaat
 gaagtattcggctgtccactaaataataaaaatatctttattttcattacatctgtgtgttggtttttgtgtgttaa
 ttaaaaaaaagaaagtcataaagcctatttcgataggctttgatgactttcgggtgtttcgtcctttccacaagata
 tataaagccaagaaatcgaaatactttcaagttacggtaagcatatgatagtcattttaaaacataatttttaaaac
 25 tgcaaaactaccaagaaattattactttctacgtcacgtattttgtactaataatctttgtgtttacagtcaaattaa
 ttctaattatctcttaacagccttgatcgtatatgcaaatatgaaggaaatcatgggaaataggccctcttcctgc
 ccgacccggaccgctaggaacccctagtgtgagttggccactccctctctgcgcgctcgcctcactgaggcc
 gggcgaccaaaggctcgcccgacgcccgggctttgcccggcgccctcagtgagcgagcgagcgcgagctgcctgca
 gg

30

Human POU4F3 sequence (SEQ ID NO: 95)

atgatggccatgaactccaagcagcctttcgcatgcacccgggtgctgcaagaacccaaattctccagtctgcactc
 tggctccgaggccatgcgccgagctctgtctcccagccccgcagctgcagggtaatatatttggaagctttgatgaga
 gcctgctggcacgcgcgaagctctggcgggcggtggatatcgtctcccacggcaagaacccatccgttcaagcccgac
 35 gccacctaccataccatgagcagcgtgccttcacgtccacttcgtccaccgtgcccatctcccaccagctgcgct
 cacctcacaccctcaccacgcggtgcaccagggcctcgaaggcgacctgctggagcacatctcgccacgctgagtg
 tgagcggcctggcgctccggaacactcgggtgatgcccgacagatccatccacaccacctggcgccatgggcccac
 ctgcaccaggccatgggcatgagtcacccgcacaccgtggcccctcatagcgccatgcctgcatgcctcagcgacgt

ggagtcagaccgcgcgagctggaagccttcgccgagcgcttcaagcagcgccgcacatcaagctgggggtgaccagc
 cggacgtgggcgcggctctggctaattctcaagatccccggcggtgggtcgctgagccaaagcaccatctgcaggttc
 gagtctctcactctctcgcacaacaacatgatcgctctcaagccggtgctccaggcctgggtggaggaggccgagggc
 cgctaccgagagaagaacagcaagccagagctcttcaacggcagcgaacggaagcgaaacgcacgtccatcgcg
 5 cgccggagaagcggttcactcgaggcctatttcgctatccagccacgtccttcatctgagaagatcgcgccatcgct
 gagaaactggaccttaaaaagaacgtggtgagagtcgtggttctgcaaccagagacagaaacagaaacgaatgaagta
 ttcggtgtccactaa

A variety of different methods known in the art can be used to introduce any of the AAV
 10 vectors disclosed herein into a primate cell (e.g., a supporting cell or a hair cell (e.g., an inner or
 outer cochlear hair cell)). Non-limiting examples of methods for introducing an AAV vector
 into a primate cell include: lipofection, transfection (e.g., calcium phosphate transfection,
 transfection using highly branched organic compounds, transfection using cationic polymers,
 dendrimer-based transfection, optical transfection, particle-based transfection (e.g., nanoparticle
 15 transfection), or transfection using liposomes (e.g., cationic liposomes)), microinjection,
 electroporation, cell squeezing, sonoporation, protoplast fusion, impalefection, hydrodynamic
 delivery, gene gun, magnetofection, viral transfection, and nucleofection.

Skilled practitioners will appreciate that any of the AAV vectors described herein can be
 introduced into a primate cell (e.g., a hair cell or a supporting cell of the inner ear) by, for
 20 example, lipofection.

Various molecular biology techniques that can be used to correct a mutation(s) in an
 endogenous gene are also known in the art. Non-limiting examples of such techniques include
 site-directed mutagenesis, CRISPR (e.g., CRISPR/Cas9-induced knock-in mutations and
 CRISPR/Cas9-induced knock-out mutations), and TALENs. These methods can be used to
 25 correct the sequence of a defective endogenous gene present in a chromosome of a target cell
 (e.g., any of the exemplary cells described herein).

Any of the AAV vectors described herein can further include a control sequence, e.g., a
 control sequence selected from the group of a transcription initiation sequence, a transcription
 termination sequence, a promoter sequence, an enhancer sequence, an RNA splicing sequence, a
 30 polyadenylation (polyA) sequence, a Kozak consensus sequence, and a destabilizing domain
 sequence. Non-limiting examples of these control sequences are described herein. In some
 embodiments, a promoter can be a native promoter, a constitutive promoter, an inducible

promoter, and/or a tissue-specific promoter.

Some embodiments of any of the compositions and kits described herein can include any combination of the AAV vectors described herein. Some embodiments of any of the methods described herein can include the use of any combination of the AAV vectors described herein.

5

Promoters

The term “promoter” means a DNA sequence recognized by enzymes/proteins in a primate cell required to initiate the transcription of a specific gene (e.g., a hair cell differentiation gene). A promoter typically refers to, e.g., a nucleotide sequence to which an RNA polymerase and/or any associated factor binds and at which transcription is initiated. Non-limiting examples of promoters are described herein. Additional examples of promoters are known in the art.

In some embodiments, an AAV vector encoding an N-terminal portion of a hair cell differentiation protein (e.g., a human hair cell differentiation protein) can include a promoter and/or an enhancer. The AAV vector encoding the N-terminal portion of the hair cell differentiation protein can include any of the promoters and/or enhancers described herein or known in the art.

In some embodiments, the promoter is an inducible promoter, a constitutive promoter, a primate cell promoter, a viral promoter, a chimeric promoter, an engineered promoter, a tissue-specific promoter, or any other type of promoter known in the art. In some embodiments, the promoter is a RNA polymerase II promoter, such as a primate RNA polymerase II promoter. In some embodiments, the promoter is a RNA polymerase III promoter, including, but not limited to, a H1 promoter, a human U6 promoter, a mouse U6 promoter, or a swine U6 promoter. The promoter will generally be one that is able to promote transcription in cochlear cells such as hair cells or supporting cells. In some examples, the promoter is a cochlea-specific promoter or a cochlea-oriented promoter.

A variety of promoters are known in the art that can be used herein. Non-limiting examples of promoters that can be used herein include: human EF1a, human cytomegalovirus (CMV)

(GTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGG
ATTTCCAAGTCTCCACCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCA
ACGGGACTTTCCAAAATGTCGTAACAACCTCCGCCCCATTGACGCAAATGGGCGGTA

GGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATC
 GCCTGGAGACGC; SEQ ID NO: 53; US Patent No. 5,168,062), human ubiquitin C (UBC),
 mouse phosphoglycerate kinase 1, polyoma adenovirus, simian virus 40 (SV40), β -globin, β -
 actin, α -fetoprotein, γ -globin, β -interferon, γ -glutamyl transferase, mouse mammary tumor virus
 5 (MMTV), Rous sarcoma virus, rat insulin, glyceraldehyde-3-phosphate dehydrogenase,
 metallothionein II (MT II), amylase, cathepsin, MI muscarinic receptor, retroviral LTR (e.g.
 human T-cell leukemia virus HTLV), AAV ITR, interleukin-2, collagenase, platelet-derived
 growth factor, adenovirus 5 E2, stromelysin, murine MX gene, glucose regulated proteins
 (GRP78 and GRP94), α -2-macroglobulin, vimentin, MHC class I gene H-2k b, HSP70,
 10 proliferin, tumor necrosis factor, thyroid stimulating hormone α gene, immunoglobulin light
 chain, T-cell receptor, HLA DQ α and DQ β , interleukin-2 receptor, MHC class II, MHC class II
 HLA-DR α , muscle creatine kinase, prealbumin (transthyretin), elastase I, albumin gene, c-fos, c-
 HA-ras, neural cell adhesion molecule (NCAM), H2B (TH2B) histone, rat growth hormone,
 human serum amyloid (SAA), troponin I (TN I), duchenne muscular dystrophy, human
 15 immunodeficiency virus, and Gibbon Ape Leukemia Virus (GALV) promoters. Additional
 examples of promoters are known in the art. See, e.g., Lodish, Molecular Cell Biology, Freeman
 and Company, New York 2007. In some embodiments, the promoter is the CMV immediate
 early promoter. In some embodiments, the promoter is a CAG promoter or a CAG/CBA
 promoter.

20 The term “constitutive” promoter refers to a nucleotide sequence that, when operably
 linked with a nucleic acid encoding a protein (e.g., a hair cell differentiation protein), causes
 RNA to be transcribed from the nucleic acid in a primate cell (e.g., a hair cell or a supporting cell
 of the inner ear) under most or all physiological conditions.

Examples of constitutive promoters include, without limitation, the retroviral Rous
 25 sarcoma virus (RSV) LTR promoter, the cytomegalovirus (CMV) promoter (see, e.g., Boshart et
 al, *Cell* 41:521-530, 1985), the SV40 promoter, the dihydrofolate reductase promoter, the beta-
 actin promoter, the phosphoglycerol kinase (PGK) promoter, and the EF1-alpha promoter
 (Invitrogen).

Inducible promoters allow regulation of gene expression and can be regulated by
 30 exogenously supplied compounds, environmental factors such as temperature, or the presence of
 a specific physiological state, e.g., acute phase, a particular differentiation state of the cell, or in

replicating cells only. Inducible promoters and inducible systems are available from a variety of commercial sources, including, without limitation, Invitrogen, Clontech, and Ariad. Additional examples of inducible promoters are known in the art.

Examples of inducible promoters regulated by exogenously supplied compounds include
 5 the zinc-inducible sheep metallothioneine (MT) promoter, the dexamethasone (Dex)-inducible mouse mammary tumor virus (MMTV) promoter, the T7 polymerase promoter system (WO 98/10088); the ecdysone insect promoter (No et al, *Proc. Natl. Acad. Sci. U.S.A.* 93:3346-3351, 1996), the tetracycline-repressible system (Gossen et al, *Proc. Natl. Acad. Sci. U.S.A.* 89:5547-5551, 1992), the tetracycline-inducible system (Gossen et al, *Science* 268:1766-1769, 1995, see
 10 also Harvey et al, *Curr. Opin. Chem. Biol.* 2:512-518, 1998), the RU486-inducible system (Wang et al, *Nat. Biotech.* 15:239-243, 1997) and Wang et al, *Gene Ther.* 4:432-441, 1997), and the rapamycin-inducible system (Magari et al. *J. Clin. Invest.* 100:2865-2872, 1997).

The term “tissue-specific” promoter refers to a promoter that is active only in certain specific cell types and/or tissues (e.g., transcription of a specific gene occurs only within cells
 15 expressing transcription regulatory proteins that bind to the tissue-specific promoter).

In some embodiments, the regulatory sequences impart tissue-specific gene expression capabilities. In some cases, the tissue-specific regulatory sequences bind tissue-specific transcription factors that induce transcription in a tissue-specific manner.

Exemplary tissue-specific promoters include but are not limited to the following: a liver-specific thyroxin binding globulin (TBG) promoter, an insulin promoter, a glucagon promoter, a somatostatin promoter, a pancreatic polypeptide (PPY) promoter, a synapsin-1 (Syn) promoter, a creatine kinase (MCK) promoter, a primate desmin (DES) promoter, an alpha-myosin heavy chain (a-MHC) promoter, and a cardiac Troponin T (cTnT) promoter. Additional exemplary promoters include Beta-actin promoter, hepatitis B virus core promoter (Sandig et al., *Gene*
 20 *Ther.* 3:1002-1009, 1996), alpha-fetoprotein (AFP) promoter (Arbuthnot et al., *Hum. Gene Ther.* 7:1503-1514, 1996), bone osteocalcin promoter (Stein et al., *Mol. Biol. Rep.* 24:185-196, 1997); bone sialoprotein promoter (Chen et al., *J. Bone Miner. Res.* 11:654-664, 1996), CD2 promoter (Hansal et al., *J. Immunol.* 161:1063-1068, 1998); immunoglobulin heavy chain promoter; T cell receptor alpha-chain promoter, neuronal such as neuron-specific enolase (NSE) promoter
 25 (Andersen et al., *Cell. Mol. Neurobiol.* 13:503-515, 1993), neurofilament light-chain gene
 30

promoter (Piccioli et al., *Proc. Natl. Acad. Sci. U.S.A.* 88:5611-5615, 1991), and the neuron-specific vgf gene promoter (Piccioli et al., *Neuron* 15:373-384, 1995).

In some embodiments, the tissue-specific promoter is a cochlea-specific promoter. In some embodiments, the tissue-specific promoter is a cochlear hair cell-specific promoter. Non-limiting examples of cochlear hair cell-specific promoters include but are not limited to: a
 5 ATOH1 promoter, a ATOH1 3'-enhancer, a POU4F3 promoter, a LHX3 promoter, a MYO7A promoter, a MYO6 promoter, a CHRNA9 promoter, and a CHRNA10 promoter. In some embodiments, the promoter is an outer hair cell-specific promoter such as a SLC26A5 promoter or an OCM promoter. See, e.g., Zheng et al., *Nature* 405:149-155, 2000; Tian et al. *Dev. Dyn.*
 10 231:199-203, 2004; and Ryan et al., *Adv. Otorhinolaryngol.* 66: 99-115, 2009.

In some embodiments of any of the AAV vectors described herein, the AAV vector includes a human ATOH1 enhancer-promoter (SEQ ID NO: 94).

Human ATOH1 enhancer-promoter (SEQ ID NO: 94)

15 ctatggaggtttgcataacaaacgtttggcagctcgctctcttactccattaacaagctgtaacatatagctgcag
 gttgctataatctcattaatatatttggaaacttgaatattgagtatttctgagtgtcattccccatagccagcca
 cttctgccatgctgactggttcctttctctccattattagcaattagcttcttaccttccaaagtgcagatccaaggt
 atccaagatactagcaaaggaatcaactatgtgtgcaagttaagcatgcttaatatcacccaaacaaacaaagagggc
 agcatttcttaaagtaataagatagataaatcgggttagtcctttgcgacactgctggtgctttctagagttttat
 20 atattttaagcagcttgctttatattctgtctttgcctcccaccccaccagcacttttatttggagggttttggc
 tcgccacactttgggaaacttatttgatttcacggagagctgaaggaagatcatttttggcaacagacaagttaaa
 cacgatttctatgggacattgctaactggggccctaaaggagaaaaggggaaactgagcggagaatgggttaaatcct
 tggagcaggggagaggcaggggaggagagaagtcggaggagtataaagaaaaggacaggaaccaagaagcgtgggg
 gtggtttgccgtaatgtgagtgtttcttaattagagaacggttgacaatagagggtctggcagaggctcctggccgc
 25 ggtgaggagcgtctggagcggagcacgcgctgtcagctggtgagcgcactctcctttcaggcagctccccggggagc
 tgtgaggccacatttaacaccatcatcacccctccccggcctcctcaacctcggcctcctcctcgtcgacagccttc
 cttggccccaccagcagagctcacagtagcgagcgtctctcgccgtctcccgactcggcggggcctctctcctc
 cccagctgcgagcgggagcggcactgcccactgcacctcccagcaaccagcccagcacgcaaagaagctgcgca
 aagttaaagccaagcaatgccaaggggaggggaagctggaggcgggctttgagtggcttctgggagcctggcggtc
 30 cagaatcgcccagagccgcccgggtcgtgcacatctgacccgagtcagcttgggcaccagccgagagccgggtccg
 caccgctcccgaccccagccgcccgggtggtgacacacaccggagtcgaattacagccctgcaattaacatatgaa
 tctgacgaatttaaaagaaggaaaaaaacactgagcaggcttgggagtcctctgcacacaagaacttttc
 tcggggtgtaaaaactctttgattggctgctcgacgcgcctgccgcgcctccattggctgagaagacacgcgac
 cggcgagaggaggggttgggagaggagcgggggagactgagtgccgctgcccgttttttaaggggagcagcgcc
 35 ttcagcaaccggagaagcatagttgcacgcgacctggtgtgtgatctccgagtggtgggggaggggtcgaggagga

aaaaaaaaataagacgtttgcagaagagacccggaaagggccttttttttggttgagctggtgtcccagtgctgcctcc
gacctgagcctccgagcctttgcagtgcaa

Enhancers and 5' Cap

5 In some instances, an AAV vector can include a promoter sequence and/or an enhancer sequence. The term “enhancer” refers to a nucleotide sequence that can increase the level of transcription of a nucleic acid encoding a protein of interest (e.g., a hair cell differentiation protein). Enhancer sequences (50-1500 basepairs in length) generally increase the level of transcription by providing additional binding sites for transcription-associated proteins (e.g.,
10 transcription factors). In some embodiments, an enhancer sequence is found within an intronic sequence. Unlike promoter sequences, enhancer sequences can act at much larger distance away from the transcription start site (e.g., as compared to a promoter). Non-limiting examples of enhancers include a RSV enhancer, a CMV enhancer
(CTAGATCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGG
15 CTGACCGCCCAACGACCCCCGCCATTGACGTCAATAATGACGTATGTTCCCATAGT
AACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGC
CCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAA
TGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCC
TACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATG; SEQ ID NO: 52), and a
20 SV40 enhancer.

In some embodiments of any of the AAV vectors described herein, the AAV vector includes a CMV enhancer-promoter sequence (SEQ ID NO: 96)

CMV enhancer-promoter sequence (SEQ ID NO: 96)

25 CGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCC
ATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTG
ACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTA
TCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCA
TTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTA
30 GTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAG
CGGTTTGA CTACGGGGATTTC AAGTCTCCACCCCAT TGACGTCAATGGGAGTTTG

TTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACCTCCGCCCCATTG
ACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCT

Poly(A) Sequences

5 In some embodiments, any of the AAV vectors provided herein can include a poly(A) sequence. Most nascent eukaryotic mRNAs possess a poly(A) tail at their 3' end which is added during a complex process that includes cleavage of the primary transcript and a coupled polyadenylation reaction (see, e.g., Proudfoot et al., *Cell* 108:501-512, 2002). The poly(A) tail confers mRNA stability and transferability (Molecular Biology of the Cell, Third Edition by B. Alberts et al., Garland Publishing, 1994). In some embodiments, the poly(A) sequence is positioned 3' to the nucleic acid sequence encoding the C-terminus of the hair cell differentiation protein or a protein of interest (e.g., a Cas9 endonuclease, e.g., a SaCas9 endonuclease (e.g., any of the SaCas9 endonucleases described herein), a reporter protein (e.g., a GFP protein, a mScarlet protein)).

15 As used herein, "polyadenylation" refers to the covalent linkage of a polyadenylyl moiety, or its modified variant, to a messenger RNA molecule. In eukaryotic organisms, most messenger RNA (mRNA) molecules are polyadenylated at the 3' end. The 3' poly(A) tail is a long sequence of adenine nucleotides (e.g., 50, 60, 70, 100, 200, 500, 1000, 2000, 3000, 4000, or 5000) added to the pre-mRNA through the action of an enzyme, polyadenylate polymerase. In higher eukaryotes, the poly(A) tail is added onto transcripts that contain a specific sequence, the polyadenylation signal or "poly(A) sequence." The poly(A) tail and the protein bound to it aid in protecting mRNA from degradation by exonucleases. Polyadenylation is also important for transcription termination, export of the mRNA from the nucleus, and translation.

25 Polyadenylation occurs in the nucleus immediately after transcription of DNA into RNA, but additionally can also occur later in the cytoplasm. After transcription has been terminated, the mRNA chain is cleaved through the action of an endonuclease complex associated with RNA polymerase. The cleavage site is usually characterized by the presence of the base sequence AAUAAA near the cleavage site. After the mRNA has been cleaved, adenosine residues are added to the free 3' end at the cleavage site.

As used herein, a “poly(A) sequence” is a sequence that triggers the endonuclease cleavage of an mRNA and the additional of a series of adenosines to the 3' end of the cleaved mRNA.

There are several poly(A) sequences that can be used, including those derived from
 5 bovine growth hormone (bgh) (Woychik et al., *Proc. Natl. Acad. Sci. U.S.A.* 81(13):3944-3948, 1984; U.S. Patent No. 5,122,458), mouse- β -globin, mouse- α -globin (Orkin et al., *EMBO J.* 4(2):453-456, 1985; Thein et al., *Blood* 71(2):313-319, 1988), human collagen, polyoma virus (Batt et al., *Mol. Cell Biol.* 15(9):4783-4790, 1995), the Herpes simplex virus thymidine kinase gene (HSV TK), IgG heavy-chain gene polyadenylation signal (US 2006/0040354), human
 10 growth hormone (hGH) (Szymanski et al., *Mol. Therapy* 15(7):1340-1347, 2007), the group of SV40 poly(A) sites, such as the SV40 late and early poly(A) site (Schek et al., *Mol. Cell Biol.* 12(12):5386-5393, 1992).

The poly(A) sequence can be a sequence of AATAAA. The AATAAA sequence may be substituted with other hexanucleotide sequences with homology to AATAAA which are capable
 15 of signaling polyadenylation, including ATTAAA, AGTAAA, CATAAA, TATAAA, GATAAA, ACTAAA, AATATA, AAGAAA, AATAAT, AAAAAA, AATGAA, AATCAA, AACAAA, AATCAA, AATAAC, AATAGA, AATTAA, or AATAAG (see, e.g., WO 06/12414).

In some embodiments, the poly(A) sequence can be a synthetic polyadenylation site (see,
 20 e.g., the pCI-neo expression vector of Promega which is based on Levitt et al., *Genes Dev.* 3(7):1019-1025, 1989). In some embodiments, the poly(A) sequence is the polyadenylation signal of soluble neuropilin-1 (sNRP) (AAATAAAATACGAAATG) (see, e.g., WO 05/073384). Additional examples of poly(A) sequences are known in the art.

In some embodiments, the poly(A) sequence is a bGHpA sequence
 25 (GCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCCTCCC
 CCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACCTGTCCTTTCCTAATAAAATG
 AGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGG
 GGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGC
 GGTGGGCTCTATGG; SEQ ID NO: 56).

30

Internal Ribosome Entry Site (IRES)

In some embodiments, an AAV vector encoding the C-terminus of the hair cell differentiation protein can include a polynucleotide internal ribosome entry site (IRES). An IRES sequence is used to produce more than one polypeptide from a single gene transcript. An IRES forms a complex secondary structure that allows translation initiation to occur from any position with an mRNA immediately downstream from where the IRES is located (see, e.g., Pelletier and Sonenberg, *Mol. Cell. Biol.* 8(3):1103-1112, 1988).

There are several IRES sequences known to those skilled in the art, including those from, e.g., foot and mouth disease virus (FMDV), encephalomyocarditis virus (EMCV), human rhinovirus (HRV), cricket paralysis virus, human immunodeficiency virus (HIV), hepatitis A virus (HAV), hepatitis C virus (HCV), and poliovirus (PV). See e.g., Alberts, *Molecular Biology of the Cell*, Garland Science, 2002; and Hellen et al., *Genes Dev.* 15(13):1593-612, 2001.

In some embodiments, the IRES sequence that is incorporated into the vector that encodes the C-terminus of a hair cell differentiation protein is the foot and mouth disease virus (FMDV) 2A sequence. In some embodiments, the IRES sequence that is incorporated into the vector that encodes the C-terminal portion of a protein of interest (e.g., a Cas9 endonuclease, e.g., a SaCas9 endonuclease (e.g., any of the SaCas9 endonucleases described herein)) is the FMDV 2A sequence. The Foot and Mouth Disease Virus 2A sequence is a small peptide (approximately 18 amino acids in length) that has been shown to mediate the cleavage of polyproteins (Ryan, M D et al., *EMBO* 4:928-933, 1994; Mattion et al., *J. Virology* 70:8124-8127, 1996; Furler et al., *Gene Therapy* 8:864-873, 2001; and Halpin et al., *Plant Journal* 4:453-459, 1999). The cleavage activity of the 2A sequence has previously been demonstrated in artificial systems including plasmids and gene therapy vectors (AAV and retroviruses) (Ryan et al., *EMBO* 4:928-933, 1994; Mattion et al., *J. Virology* 70:8124-8127, 1996; Furler et al., *Gene Therapy* 8:864-873, 2001; and Halpin et al., *Plant Journal* 4:453-459, 1999; de Felipe et al., *Gene Therapy* 6:198-208, 1999; de Felipe et al., *Human Gene Therapy* 11:1921-1931, 2000; and Klump et al., *Gene Therapy* 8:811-817, 2001).

Destabilizing Domain (DD)

Any of the AAV vectors provided herein can optionally include a sequence encoding a destabilizing domain (“a destabilizing sequence”) for temporal control of protein expression.

Non-limiting examples of destabilizing sequences include sequences encoding: a FK506

5 sequence, a dihydrofolate reductase (DHFR) sequence. An exemplary DHFR destabilizing sequence is:

MISLIAALAVDYVIGMENAMPWNL PADLAWFKRNTLNKPVIMGRHTWESIGRPLPGRK
NIILSSQPSTDDRVTWVKSVD EAIACGDVPEIMVIGGGRVIEQFLPKAQKLYLTHIDAEV
EGDTHFPDYEPDDWESVFSEFHDADAQNSHSYC FEILERR (SEQ ID NO: 48). An

10 exemplary DHFR destabilizing domain sequence is

GGTACCATCAGTCTGATTGCGGCGTTAGCGGTAGATTACGTTATCGGCATGGAAAAC
GCCATGCCGTGGAACCTGCCTGCCGATCTCGCCTGGTTTAAACGCAACACCTTAAAT
AAACCCGTGATTATGGGCCGCCATACCTGGGAATCAATCGGTTCGTCCTGTCAGGA
CGCAAAAATATTATCCTCAGCAGTCAACCGAGTACGGACGATCGCGTAACGTGGGT
15 GAAGTCGGTGGATGAAGCCATCGCGGCGTGTGGTGACGTACCAGAAATCATGGTGA
TTGGCGGCGGTTCGCGTTATTGAACAGTTCTTGCCAAAAGCGCAAAAACCTGTATCTGA
CGCATATCGACGCAGAAAGTGAAGGCGACACCCATTTCCCGGATTACGAGCCGGAT
GACTGGGAATCGGTATTCAGCGAATTCCACGATGCTGATGCGCAGAACTCTCACAGC
TATTGCTTTGAGATTCTGGAGCGGCGATAA (SEQ ID NO: 59).

20

In some embodiments of any of the AAV vectors described herein, the AAV vector includes a destabilizing domain (SEQ ID NO: 88).

Destabilizing domain (SEQ ID NO: 88)

25 Atcagtctgatttgcggcgtagcggttagattacgttatcggcattggaacacgcatgcccgtggaacctgcctgcca
tctcgcttggtttaaacgcaacaccttaataaaacccgtgattatgggcccgcatacctgggaatcaatcggtcgtc
cgttgccaggacgcaaaaatattatcctcagcagtcacccgagtagcgacgatcgcgtaacgtgggtgaagtcgggtg
gatgaagccatcgcgccgtgtggtgacgtaccagaaatcatggtgattggcgccggtcgcggttattgaacagttctt
gccaaaagcgcaaaaactgtatctgacgcataatcgacgcagaagtggaaggcgacacccatttcccggttacgagc
30 cggatgactgggaatcggtattcagcgaattccacgatgctgatgcgcaactctcacagctattgctttgagatt
ctggagcggcga

Additional examples of destabilizing sequences are known in the art. In some embodiments, the destabilizing sequence is a FK506- and rapamycin-binding protein (FKBP12) sequence, and the stabilizing ligand is Shield-1 (Shld1) (Banaszynski et al. (2012) *Cell* 126(5): 995-1004). An exemplary FKBP12 destabilizing sequence is:

5 MGVEKQVIRPGNGPKPAPGQTVTVHCTGFG
 KDGDLSQKFWSTKDEGQKPFSFQIGKGA VIKGWDEGVIGMQIGEVARLRCSSDYAYGA
 GGFWAWGIQPNVLD FEIEVLSVQ (SEQ ID NO: 49). In some embodiments, the destabilizing sequence is a DHFR sequence, and the stabilizing ligand is trimethoprim (TMP) (Iwamoto et al. (2010) *Chem Biol* 17:981-988).

10 In the absence of a stabilizing ligand, the protein sequence operatively linked to the destabilizing sequence is degraded by ubiquitination. In contrast, in the presence of a stabilizing ligand, protein degradation is inhibited, thereby allowing the protein sequence operatively linked to the destabilizing sequence to be actively expressed. As a positive control for stabilization of protein expression, protein expression can be detected by conventional means, including
 15 enzymatic, radiographic, colorimetric, fluorescence, or other spectrographic assays; fluorescent activating cell sorting (FACS) assays; immunological assays (e.g., enzyme linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and immunohistochemistry).

In some embodiments, the destabilizing sequence is a FKBP12 sequence, and the presence of an AAV vector carrying the FKBP12 gene in a primate cell (e.g., a supporting
 20 cochlear outer hair cell) is detected by western blotting. In some embodiments, the destabilizing sequence can be used to verify the temporally-specific activity of any of the AAV vectors described herein.

In some embodiments of any of the AAV vectors described herein, the AAV vector comprising the C-terminal portion of a hair cell differentiation gene, the vector further includes a
 25 destabilizing sequence 3' of the C-terminal portion of the hair cell differentiation gene. In some embodiments of the AAV vector including a sequence encoding the C-terminal portion of an ATOH1 protein, the vector further comprises a sequence encoding a destabilizing domain (DD) (e.g., any of the destabilizing domain described herein).

Reporter Sequences/Detectable Marker Genes

Any of the AAV vectors provided herein can optionally include a sequence encoding a reporter protein or a detectable marker ("a reporter sequence" or "a detectable marker gene"). Non-limiting examples of reporter sequences or detectable marker genes include DNA sequences encoding: a beta-lactamase, a beta-galactosidase (LacZ), an alkaline phosphatase, a thymidine kinase, a green fluorescent protein (GFP), a red fluorescent protein, an mCherry fluorescent protein, a yellow fluorescent protein, a chloramphenicol acetyltransferase (CAT), and a luciferase. Additional examples of reporter sequences or detectable markers are known in the art. When associated with regulatory elements which drive their expression, the reporter sequence or detectable marker gene can provide signals detectable by conventional means, including enzymatic, radiographic, colorimetric, fluorescence, or other spectrographic assays; fluorescent activating cell sorting (FACS) assays; immunological assays (e.g., enzyme linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and immunohistochemistry).

In some embodiments, the reporter sequence or detectable marker gene is a 3x Flag sequence

(GATTACAAGGATGACGACGATAAGGACTATAAGGACGATGATGACAAGGACTACA AAGATGATGACGATAAAGGATCCGGC; SEQ ID NO: 62). In some embodiments, the reporter sequence or detectable marker gene is a luciferase sequence

(ATGGAAGATGCCAAAACATTAAGAAGGGCCCAGCGCCATTCTACCCACTCGAAG
ACGGGACCGCCGGCGAGCAGCTGCACAAAGCCATGAAGCGCTACGCCCTGGTGCCC
GGCACCATCGCCTTTACCGACGCACATATCGAGGTGGACATTACCTACGCCGAGTAC
TTCGAGATGAGCGTTTCGGCTGGCAGAAGCTATGAAGCGCTATGGGCTGAATACAAA
CCATCGGATCGTGGTGTGCAGCGAGAATAGCTTGCAGTTCTTCATGCCCCGTGTTGGG
TGCCCTGTTTCATCGGTGTGGCTGTGGCCCCAGCTAACGACATCTACAACGAGCGCGA
GCTGCTGAACAGCATGGGCATCAGCCAGCCCACCGTCGTATTCGTGAGCAAGAAAG
GGCTGCAAAAGATCCTCAACGTGCAAAAGAAGCTACCGATCATAAAAAGATCATC
ATCATGGATAGCAAGACCGACTACCAGGGCTTCCAAAGCATGTACACCTTCGTGACT
TCCCATTGTCACCCGGCTTCAACGAGTACGACTTCGTGCCCCGAGAGCTTCGACCGG
GACAAAACCATCGCCCTGATCATGAACAGTAGTGGCAGTACCGGATTGCCCAAGGG
CGTAGCCCTACCGCACCGCACCGCTTGTGTCCGATTCAAGTCATGCCCCGCGACCCCAT
CTTCGGCAACCAGATCATCCCCGACACCGCTATCCTCAGCGTGGTGCCATTTACCA

CGGCTTCGGCATGTTACCCACGCTGGGCTACTTGATCTGCGGCTTTCGGGTCGTGCTC
 ATGTACCGCTTCGAGGAGGAGCTATTCTTGCGCAGCTTGCAAGACTATAAGATTCAA
 TCTGCCCTGCTGGTGCCCACTATTTAGCTTCTTCGCTAAGAGCACTCTCATCGACA
 AGTACGACCTAAGCAACTTGCACGAGATCGCCAGCGGGCGGGGCGCCGCTCAGCAAG
 5 GAGGTAGGTGAGGCCGTGGCCAAACGCTTCCACCTACCAGGCATCCGCCAGGGCTA
 CGGCCTGACAGAAACAACCAGCGCCATTCTGATCACCCCCGAAGGGGACGACAAGC
 CTGGCGCAGTAGGCAAGGTGGTGCCCTTCTTCGAGGCTAAGGTGGTGGACTTGAC
 ACAGGTAAGACACTGGGTGTGAACCAGCGCGGCGAGCTGTGCGTCCGTGGCCCCAT
 GATCATGAGCGGCTACGTTAACAACCCCGAGGCTACAAACGCTCTCATCGACAAGG
 10 ACGGCTGGCTGCACAGCGGCGACATCGCCTACTGGGACGAGGACGAGCACTTCTTC
 ATCGTGGACCGGCTGAAGAGCCTGATCAAATACAAGGGCTACCAGGTAGCCCCAGC
 CGAACTGGAGAGCATCCTGCTGCAACACCCCAACATCTTCGACGCCGGGGTCGCCG
 GCCTGCCCCGACGACGATGCCGGCGAGCTGCCCCGCCGACGTCGTGCTGCTGGAACAC
 GGTAACCATGACCGAGAAGGAGATCGTGGACTATGTGGCCAGCCAGGTTACAAC
 15 CGCCAAGAAGCTGCGCGGTGGTGTGTTGTTTCGTGGACGAGGTGCCTAAAGGACTGA
 CCGGCAAGTTGGACGCCCGCAAGATCCGCGAGATTCTCATTAAGGCCAAGAAGGGC
 GGCAAGATCGCCGTGGGCTCCGGA; SEQ ID NO: 69).

In some embodiments, the reporter sequence or detectable marker gene is the LacZ gene,
 and the presence of a vector carrying the LacZ gene in a primate cell (e.g., a supporting cochlear
 20 outer hair cell) is detected by assays for beta-galactosidase activity. In other embodiments, the
 reporter sequence or detectable marker gene is a fluorescent protein (e.g., green fluorescent
 protein) or luciferase, the presence of a vector carrying the fluorescent protein or luciferase in a
 primate cell (e.g., a supporting cochlear outer hair cell) may be measured by fluorescent
 techniques (e.g., fluorescent microscopy or FACS) or light production in a luminometer (e.g., a
 25 spectrophotometer or an IVIS imaging instrument). In some embodiments, the reporter sequence
 or detectable marker gene can be used to verify the tissue-specific targeting capabilities and
 tissue-specific promoter regulatory activity of any of the vectors described herein.

Flanking Regions Untranslated Regions (UTRs)

30 In some embodiments, any of the AAV vectors described herein (e.g., any of the at least
 two different vectors) can include an untranslated region. In some embodiments, an AAV vector

can includes a 5' UTR or a 3' UTR.

Untranslated regions (UTRs) of a gene are transcribed but not translated. The 5' UTR starts at the transcription start site and continues to the start codon but does not include the start codon. The 3' UTR starts immediately following the stop codon and continues until the transcriptional termination signal. There is growing body of evidence about the regulatory roles played by the UTRs in terms of stability of the nucleic acid molecule and translation. The regulatory features of a UTR can be incorporated into any of the vectors, compositions, kits, or methods as described herein to enhance the stability of a hair cell differentiation protein or of a protein of interest (e.g., a Cas9 endonuclease, e.g., a SaCas9 endonuclease (e.g., any of the SaCas9 endonucleases described herein), a reporter protein (e.g., a GFP protein, a mScarlet protein).

Natural 5' UTRs include a sequence that plays a role in translation initiation. They harbor signatures like Kozak sequences, which are commonly known to be involved in the process by which the ribosome initiates translation of many genes. Kozak sequences have the consensus sequence CCR(A/G)CCAUGG, where R is a purine (A or G) three bases upstream of the start codon (AUG), which is followed by another "G". The 5' UTR have also been known, e.g., to form secondary structures that are involved in elongation factor binding.

For example, in some embodiments, a 5' UTR is included in any of the AAV vectors described herein. Non-limiting examples of 5' UTRs including those from the following genes: albumin, serum amyloid A, Apolipoprotein A/B/E, transferrin, alpha fetoprotein, erythropoietin, and Factor VIII, can be used to enhance expression of a nucleic acid molecule, such as a mRNA.

In some embodiments, a 5' UTR from a mRNA that is transcribed by a cell in the cochlea can be included in any of the vectors, compositions, kits, and methods described herein.

3' UTRs are known to have stretches of adenosines and uridines embedded in them. These AU-rich signatures are particularly prevalent in genes with high rates of turnover. Based on their sequence features and functional properties, the AU-rich elements (AREs) can be separated into three classes (Chen et al., *Mol. Cell. Biol.* 15:5777-5788, 1995; Chen et al., *Mol. Cell Biol.* 15:2010-2018, 1995): Class I AREs contain several dispersed copies of an AUUUA motif within U-rich regions. For example, c-Myc and MyoD mRNAs contain class I AREs. Class II AREs possess two or more overlapping UUAUUUA(U/A) (U/A) nonamers. GM-CSF and TNF-alpha mRNAs are examples that contain class II AREs. Class III AREs are less well

defined. These U-rich regions do not contain an AUUUA motif. Two well-studied examples of this class are c-Jun and myogenin mRNAs.

Most proteins binding to the AREs are known to destabilize the messenger, whereas members of the ELAV family, most notably HuR, have been documented to increase the stability of mRNA. HuR binds to AREs of all the three classes. Engineering the HuR specific binding sites into the 3' UTR of nucleic acid molecules will lead to HuR binding and thus, stabilization of the message *in vivo*.

In some embodiments, the introduction, removal, or modification of 3' UTR AREs can be used to modulate the stability of an mRNA encoding a hair cell differentiation protein. In other embodiments, AREs can be removed or mutated to increase the intracellular stability and thus increase translation and production of a hair cell differentiation protein.

In other embodiments, non-UTR sequences may be incorporated into the 5' or 3' UTRs. In some embodiments, introns or portions of intron sequences may be incorporated into the flanking regions of the polynucleotides in any of the vectors, compositions, kits, and methods provided herein. Incorporation of intronic sequences may increase protein production as well as mRNA levels.

Inhibitory Nucleic Acids

Some embodiments of the compositions provided herein include a single AAV vector that encodes an inhibitory nucleic acid that decreases the expression of a hair cell differentiation-suppressing protein in a primate cell (e.g., a hair cell or a supporting cell of the inner ear). Inhibitory nucleic acids include, e.g., siRNA, shRNA, antisense nucleic acids, and ribozymes.

Non-limiting examples of siRNAs that can decrease the expression of a hair cell differentiation-suppressing protein in a primate cell (e.g., a hair cell or a supporting cell of the inner ear) are described herein. An inhibitory nucleic acid can be, e.g., a chemically-modified siRNAs or a vector-driven expression of short hairpin RNA (shRNA) that are then cleaved to siRNA. In some examples, an inhibitory nucleic acid can be a dsRNA (e.g., siRNA) including 16-30 nucleotides, e.g., 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 nucleotides in each strand, where one of the strands is substantially identical, e.g., at least 80% (or more, e.g., 85%, 90%, 95%, or 100%) identical, e.g., having 3, 2, 1, or 0 mismatched nucleotide(s), to a target region in the hair cell differentiation-suppressing mRNA, and the other strand is

complementary to the first strand. dsRNA molecules can be designed using methods known in the art, e.g., Dharmacon.com (see, siDESIGN CENTER) or “The siRNA User Guide,” available on the Internet at mpibpc.gwdg.de/abteilungen/100/105/sirna.html website.

Several methods for expressing siRNA duplexes within cells from a vector to achieve
5 long-term target gene suppression in cells are known in the art, e.g., including vectors that use a mammalian Pol III promoter system (e.g., H1 or U6/snRNA promoter systems (Tuschl, *Nature Biotechnol.*, 20:440-448, 2002) to express functional double-stranded siRNAs; (Bagella et al., *J. Cell. Physiol.*, 177:206–213, 1998; Lee et al., *Nature Biotechnol.*, 20:500-505, 2002; Paul et al., *Nature Biotechnol.*, 20:505-508, 2002; Yu et al., *Proc. Natl. Acad. Sci. U.S.A.*, 99(9):6047-6052,
10 2002; Sui et al., *Proc. Natl. Acad. Sci. U.S.A.* 99(6):5515-5520, 2002). Transcriptional termination by RNA Pol III occurs at runs of four consecutive T residues in the DNA template, and can be used to provide a mechanism to end the siRNA transcript at a specific sequence. The siRNA is complementary to the sequence of the target gene in 5'-3' and 3'-5' orientations, and the two strands of the siRNA can be expressed in the same construct or in separate constructs.
15 Hairpin siRNAs, driven by H1 or U6 snRNA promoter and expressed in cells, can inhibit target gene expression (Bagella et al., 1998, supra; Lee et al., 2002, supra; Paul et al., 2002, supra; Yu et al., 2002, supra; Sui et al., 2002, supra).

Animal cells express a range of noncoding RNAs of approximately 22 nucleotides termed micro RNA (miRNAs) and can regulate gene expression at the post transcriptional or
20 translational level during animal development. miRNAs are excised from an approximately 70 nucleotide precursor RNA stem-loop. By substituting the stem sequences of the miRNA precursor with miRNA sequence complementary to the target mRNA, a vector construct that expresses the novel miRNA can be used to produce siRNAs to initiate RNAi against specific mRNA targets in mammalian cells (Zeng, *Mol. Cell*, 9:1327-1333, 2002). When expressed by
25 DNA vectors containing polymerase III promoters, micro-RNA designed hairpins can silence gene expression (McManus, *RNA* 8:842-850, 2002).

In some examples, an inhibitory nucleic acid can be an antisense nucleic acid molecules, i.e., nucleic acid molecules whose nucleotide sequence is complementary to all or part of an mRNA encoding a hair cell differentiation-suppressing protein. An antisense nucleic acid
30 molecule can be antisense to all or part of a non-coding region of the coding strand of a nucleotide sequence encoding a hair cell differentiation-suppressing protein. The non-coding

regions ("5' and 3' untranslated regions") are the 5' and 3' sequences that flank the coding region and are not translated into amino acids. Based upon the sequences disclosed herein, one of skill in the art can easily choose and synthesize any of a number of appropriate antisense molecules to target a hair cell differentiation-suppressing gene described herein. For example, a "gene walk" comprising a series of oligonucleotides of 15-30 nucleotides spanning the length of a nucleic acid (e.g., a hair cell differentiation-suppressing mRNA) can be prepared, followed by testing for inhibition of expression of the gene. Optionally, gaps of 5-10 nucleotides can be left between the oligonucleotides to reduce the number of oligonucleotides synthesized and tested.

An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45, or 50 nucleotides or more in length.

In some embodiments, the inhibitory nucleic acid can be a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes (described in Haselhoff and Gerlach, *Nature*, 334:585-591, 1988)) can be used to catalytically cleave mRNA transcripts to thereby inhibit translation of the protein encoded by the mRNA. Methods of designing and producing ribozymes are known in the art (see, e.g., Scanlon, 1999, *Therapeutic Applications of Ribozymes*, Humana Press). A ribozyme having specificity for a hair cell differentiation-suppressing mRNA can be designed based upon the nucleotide sequence of a hair cell differentiation-suppressing cDNA (e.g., any of the exemplary cDNA sequences described herein). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a hair cell differentiation-suppressing mRNA (Cech et al. U.S. Patent No. 4,987,071; and Cech et al., U.S. Patent No. 5,116,742). Alternatively, an mRNA encoding a hair cell differentiation-suppressing protein can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules (See, e.g., Bartel and Szostak, *Science*, 261:1411-1418, 1993).

In some embodiments, the administration of the single AAV vector including a sequence that encodes an inhibitory nucleic acid results in at least a 1% to about 99% decrease (e.g., a 1% decrease to about a 99% decrease, a 1% decrease to about a 95% decrease, a 1% decrease to about a 90% decrease, a 1% decrease to about a 85% decrease, a 1% decrease to about a 80% decrease, a 1% decrease to about a 75% decrease, a 1% decrease to about a 70% decrease, a 1%

decrease to about a 65% decrease, a 1% decrease to about a 60% decrease, a 1% decrease to about a 55% decrease, a 1% decrease to about a 50% decrease, a 1% decrease to about a 45% decrease, a 1% decrease to about a 40% decrease, a 1% decrease to about a 35% decrease, a 1% decrease to about a 30% decrease, a 1% decrease to about a 25% decrease, a 1% decrease to about a 20% decrease, a 1% decrease to about a 15% decrease, a 1% decrease to about a 10% decrease, about a 20% decrease to about a 99% decrease, about a 20% decrease to about a 95% decrease, about a 20% decrease to about a 90% decrease, about a 20% decrease to about a 85% decrease, about a 20% decrease to about a 80% decrease, about a 20% decrease to about a 75% decrease, about a 20% decrease to about a 70% decrease, about a 20% decrease to about a 65% decrease, about a 20% decrease to about a 60% decrease, about a 20% decrease to about a 55% decrease, about a 20% decrease to about a 50% decrease, about a 20% decrease to about a 45% decrease, about a 20% decrease to about a 40% decrease, about a 20% decrease to about a 35% decrease, about a 20% decrease to about a 30% decrease, about a 50% decrease to about a 99% decrease, about a 50% decrease to about a 95% decrease, about a 50% decrease to about a 90% decrease, about a 50% decrease to about a 85% decrease, about a 50% decrease to about a 80% decrease, about a 50% decrease to about a 75% decrease, about a 50% decrease to about a 70% decrease, about a 50% decrease to about a 65% decrease, about a 50% decrease to about a 60% decrease, about a 50% decrease to about a 55% decrease, about a 70% decrease to about a 99% decrease, about a 70% decrease to about a 95% decrease, about a 70% decrease to about a 90% decrease, about a 70% decrease to about a 85% decrease, about a 70% decrease to about a 80% decrease, about a 70% decrease to about a 75% decrease, about a 80% decrease to about a 99% decrease, about a 80% decrease to about a 95% decrease, about a 80% decrease to about a 90% decrease, about a 80% decrease to about a 85% decrease, about a 90% decrease to about a 99% decrease, or about a 90% decrease to about a 95% decrease) in the level of expression of the hair cell differentiation-suppressing mRNA or protein in a primate cell (e.g., as compared to the level of expression before administration of the single AAV vector that encodes the inhibitory nucleic acid that targets the hair cell differentiation-suppressing mRNA).

Primate Cells

Also provided herein is a cell (e.g., a primate cell, e.g., a hair cell or a supporting cell of the inner ear) that includes any of the nucleic acids, vectors (e.g., at least two different vectors

described herein), or compositions described herein. In some embodiments, the primate cell is a human cell (e.g., a human supporting cell or a human hair cell of the inner ear). In other embodiments, the primate is a non-human primate (e.g., simian cell (e.g., a monkey cell (e.g., a marmoset cell, a baboon cell, a macaque cell), or an ape cell (e.g., a gorilla cell, a gibbon cell, an orangutan cell, a chimpanzee cell). Skilled practitioners will appreciate that the AAV vectors described herein can be introduced into any primate cell (e.g., a primate supporting cell or a primate hair cell of the inner ear). Non-limiting examples of AAV vectors and methods for introducing AAV vectors into primate cells are described herein.

In some embodiments, the primate cell can be a supporting hair cell of the inner ear of a mammal. For example, a supporting cell can be Hensen's cells, Deiters' cells, inner pillar cells, outer pillar cells, Claudius cells, inner border cells, inner phalangeal cells, or cells of the stria vascularis.

In some embodiments, the primate cell is a specialized cell of the cochlea. In some embodiments, the primate cell is a hair cell. In some embodiments, the primate cell is a cochlear inner hair cell or a cochlear outer hair cell. In some embodiments, the primate cell is a cochlear inner hair cell. In some embodiments, the primate cell is a cochlear outer hair cell.

In some embodiments, the primate cell is *in vitro*. In some embodiments, the primate cell is present in a primate. In some embodiments, the primate cell is autologous cell obtained from a primate and cultured *ex vivo*.

Methods

Also provided herein are methods of promoting differentiation of a supporting cell of an inner ear of a primate into a hair cell that include: administering to the inner ear of the primate a therapeutically effective amount of any of the compositions described herein, where the administering promotes differentiation of the supporting cell of the inner ear of the primate into a hair cell. Differentiation of a supporting cell of the inner ear into a hair cell can be determined using, e.g., indirect functional assays (e.g., hearing testing, e.g., pure tone audiometry).

Also provided herein are methods of increasing the expression level of a hair cell differentiation protein in a hair cell or a supporting cell of an inner ear of a primate that include: administering to the inner ear of the primate a therapeutically effective amount of any of the compositions described herein, where the administering results in an increase (e.g., a 1% to

500% increase, a 1% to 450% increase, a 1% to 400% increase, a 1% to 350% increase, a 1% to 300% increase, a 1% to 250% increase, a 1% to 200% increase, a 1% to 150% increase, a 1% to 100% increase, a 1% to 50% increase, a 50% to 500% increase, a 50% to 450% increase, a 50% to 400% increase, a 50% to 350% increase, a 50% to 300% increase, a 50% to 250% increase, a 50% to 200% increase, a 50% to 150% increase, or a 50% to 100% increase) in the expression level of the hair cell differentiation protein in the hair cell or the supporting cell of the inner ear of the primate (e.g., as compared to the level of expression of the hair cell differentiation protein in the hair cell or the supporting cell of the inner ear of the primate before administration of the composition).

Also provided herein are methods of decreasing the expression level of a hair cell differentiation-suppressing protein in a hair cell or a supporting cell of an inner ear of a primate that include: administering to the inner ear of the primate a therapeutically effective amount of any of the compositions described herein, where the administering results in a decrease (e.g., a 1% decrease to 99% decrease, or any of the subranges of this range described herein) in the expression level of the hair cell differentiation-suppressing protein in the hair cell or the supporting cell of the inner ear of the primate (e.g., as compared to the level of expression of the hair cell differentiation-suppressing protein in the hair cell or the supporting cell of the inner ear of the primate before administration of the composition).

Also provided herein are methods of increasing (e.g., a 1% to 500% increase, or any of the subranges of this range described herein) the number of functional hair cells in a primate in need thereof (e.g., as compared to the number of functional hair cells in a primate prior to the administration of the composition) that include: administering to the inner ear of the primate a therapeutically effective amount of any of the compositions described herein.

Also provided herein are methods of improving hearing in a primate in need thereof, the method comprising administering to the inner ear of the primate a therapeutically effective amount of any of the compositions described herein. In some embodiments, the administering improves hearing in a primate following environmental damage (e.g., noise, chemotherapeutic treatment (e.g., cisplatin treatment) or aminoglycoside treatment).

Also provided herein are methods of repairing a hair cell toxicity-inducing mutation in an endogenous hair cell differentiation gene locus in a hair cell or a supporting cell of an inner ear of a primate that include: administering to the inner ear of the primate a therapeutically effective

amount of any of the compositions described herein, where the administering results in repair of the hair cell toxicity-inducing mutation in the endogenous hair cell differentiation gene locus in the hair cell or the supporting cell of the inner ear of the primate.

Also provided herein are methods of decreasing the risk of hearing loss due to hair cell loss or dysfunction in a primate in need thereof that include: administering to the inner ear of the primate a therapeutically effective amount of any of the compositions described herein.

In some embodiments of any of these methods, the primate has been previously identified as having a defective hair cell differentiation gene (e.g., a hair cell differentiation gene having a mutation that results in a decrease in the expression and/or activity of a hair cell differentiation protein encoded by the gene). In some embodiments of any of these methods, the primate has been previously identified as having a defective hair cell differentiation-suppressing gene (e.g., a hair cell differentiation-suppressing gene having a mutation that results in an increase in the expression and/or activity of a hair cell differentiation-suppressing protein encoded by the gene). Some embodiments of any of these methods further include, prior to the introducing or administering step, determining that the primate has a defective hair cell differentiation gene and/or a defective hair cell differentiation-suppressing gene. Some embodiments of any of these methods can further include detecting a mutation in a hair cell differentiation gene and/or a hair cell differentiation-suppressing gene in a primate. Some embodiments of any of the methods can further include identifying or diagnosing a primate as having non-syndromic sensorineural hearing loss. Some embodiments of any of the methods can further include identifying or diagnosing a primate as having syndromic sensorineural hearing loss.

In some embodiments of any of these methods, two or more doses of any of the compositions described herein are introduced or administered into the cochlea of the primate. Some embodiments of any of these methods can include introducing or administering a first dose of the composition into the cochlea of the primate, assessing hearing function of the primate following the introducing or the administering of the first dose, and administering an additional dose of the composition into the cochlea of the primate found not to have a hearing function within a normal range (e.g., as determined using any test for hearing known in the art).

In some embodiments of any of the methods described herein, the composition can be formulated for intra-cochlear administration. In some embodiments of any of the methods described herein, the compositions described herein can be administered via intra-cochlear

administration or local administration. In some embodiments of any of the methods described herein, the compositions are administered through the use of a medical device (e.g., any of the exemplary medical devices described herein).

In some embodiments, intra-cochlear administration can be performed using any of the methods described herein or known in the art. For example, a composition can be administered or introduced into the cochlea using the following surgical technique: first using visualization with a 0 degree, 2.5-mm rigid endoscope, the external auditory canal is cleared and a round knife is used to sharply delineate an approximately 5-mm tympanomeatal flap. The tympanomeatal flap is then elevated and the middle ear is entered posteriorly. The chorda tympani nerve is identified and divided, and a curette is used to remove the scutal bone, exposing the round window membrane. To enhance apical distribution of the administered or introduced composition, a surgical laser may be used to make a small 2-mm fenestration in the oval window to allow for perilymph displacement during trans-round window membrane infusion of the composition. The microinfusion device is then primed and brought into the surgical field. The device is maneuvered to the round window, and the tip is seated within the bony round window overhang to allow for penetration of the membrane by the microneedle(s). The footpedal is engaged to allow for a measured, steady infusion of the composition. The device is then withdrawn and the round window and stapes foot plate are sealed with a gelfoam patch.

In some embodiments of any of the methods described herein, the primate has or is at risk of developing non-syndromic sensorineural hearing loss. In some embodiments of any of the methods described herein, the primate has been previously identified as having a mutation in a hair cell differentiation gene and/or a hair cell differentiation-suppressing gene. In some embodiments of any of the methods described herein, the primate has any of the mutations in a hair cell differentiation gene and/or a hair cell differentiation-suppressing gene that are described herein or are known in the art to be associated with non-syndromic sensorineural hearing loss or syndromic sensorineural hearing loss.

In some embodiments of any of the methods described herein, the primate has been identified as being a carrier of a mutation in a hair cell differentiation gene and/or a hair cell differentiation-suppressing gene (e.g., via genetic testing). In some embodiments of any of the methods described herein, the primate has been identified as having a mutation in a hair cell differentiation gene and/or a hair cell differentiation-suppressing gene and has been diagnosed

with non-syndromic sensorineural hearing loss. In some embodiments of any of the methods described herein, the primate has been identified as having a mutation in a hair cell differentiation gene and/or a hair cell differentiation-suppressing gene and has been diagnosed with syndromic sensorineural hearing loss. In some embodiments of any of the methods described herein, the primate has been identified as having non-syndromic sensorineural hearing loss. In some embodiments of any of the methods described herein, the primate has been identified as having syndromic sensorineural hearing loss.

In some embodiments, successful treatment of non-syndromic sensorineural hearing loss, or syndromic sensorineural hearing loss, can be determined in a primate using any of the conventional functional hearing tests known in the art. Non-limiting examples of functional hearing tests are various types of audiometric assays (e.g., pure-tone testing, speech testing, test of the middle ear, auditory brainstem response, and otoacoustic emissions).

In some embodiments of these methods, the primate cell is *in vitro*. In some embodiments of these methods, the primate cell is originally obtained from a primate and is cultured *ex vivo*. In some embodiments, the primate cell has previously been determined to have a defective hair cell differentiation protein and/or a defective hair cell differentiation-suppressing protein.

Methods for introducing any of the compositions described herein into a primate cell are known in the art (e.g., via lipofection or through the use of a viral vector, e.g., any of the viral vectors described herein).

An increase in expression of an active hair cell differentiation protein and/or an active hair cell differentiation-suppressing protein (e.g., a full-length hair cell differentiation protein and/or a full-length hair cell differentiation-suppressing protein) as described herein is, e.g., as compared to a control or to the level of expression of an active hair cell differentiation protein and/or a hair cell differentiation-suppressing protein (e.g., a full-length hair cell differentiation protein and/or a full-length hair cell differentiation-suppressing protein) prior to the introduction of the vector(s).

Methods of detecting expression and/or activity of a hair cell differentiation protein and/or a hair cell differentiation-suppressing protein are known in the art. In some embodiments, the level of expression of a hair cell differentiation protein and/or a hair cell differentiation-suppressing protein can be detected directly (e.g., detecting hair cell differentiation protein

and/or a hair cell differentiation-suppressing protein or detecting hair cell differentiation mRNA and/or a hair cell differentiation-suppressing mRNA). Non-limiting examples of techniques that can be used to detect expression and/or activity of hair cell differentiation proteins and/or hair cell differentiation-suppressing proteins directly include: real-time PCR, Western blotting, immunoprecipitation, immunohistochemistry, or immunofluorescence. In some embodiments, expression of a hair cell differentiation protein and/or a hair cell differentiation-suppressing protein can be detected indirectly (e.g., through functional hearing tests).

Pharmaceutical Compositions and Kits

In some embodiments, any of the compositions described herein can further include one or more agents that promote the entry of any of the AAV vectors described herein into a primate cell (e.g., a liposome or cationic lipid).

In some embodiments, any of the AAV vectors described herein can be formulated using natural and/or synthetic polymers. Non-limiting examples of polymers that may be included in any of the compositions described herein can include, but are not limited to, DYNAMIC POLYCONJUGATE® (Arrowhead Research Corp., Pasadena, Calif.), formulations from Mirus Bio (Madison, Wis.) and Roche Madison (Madison, Wis.), PhaseRX polymer formulations such as, without limitation, SMARTT POLYMER TECHNOLOGY® (PhaseRX, Seattle, Wash.), DMRI/DOPE, poloxamer, VAXFECTIN® adjuvant from Vical (San Diego, Calif.), chitosan, cyclodextrin from Calando Pharmaceuticals (Pasadena, Calif.), dendrimers and poly (lactic-co-glycolic acid) (PLGA) polymers, RONDEL™ (RNAi/Oligonucleotide Nanoparticle Delivery) polymers (Arrowhead Research Corporation, Pasadena, Calif.), and pH responsive co-block polymers, such as, but not limited to, those produced by PhaseRX (Seattle, Wash.). Many of these polymers have demonstrated efficacy in delivering nucleic acid *in vivo* into a primate cell (see, e.g., deFougerolles, *Human Gene Ther.* 19:125-132, 2008; Rozema et al., *Proc. Natl. Acad. Sci. U.S.A.* 104:12982-12887, 2007; Rozema et al., *Proc. Natl. Acad. Sci. U.S.A.* 104:12982-12887, 2007; Hu-Lieskovan et al., *Cancer Res.* 65:8984-8982, 2005; Heidel et al., *Proc. Natl. Acad. Sci. U.S.A.* 104:5715-5721, 2007).

Any of the compositions described herein can be, e.g., a pharmaceutical composition. A pharmaceutical composition can include any of the compositions described herein and one or more pharmaceutically or physiologically acceptable carriers, diluents, or excipients. Such

compositions may comprise one or more buffers, such as neutral-buffered saline, phosphate-buffered saline, and the like; one or more carbohydrates, such as glucose, mannose, sucrose, and dextran; mannitol; one or more proteins, polypeptides, or amino acids, such as glycine; one or more antioxidants; one or more chelating agents, such as EDTA or glutathione; and/or one or more preservatives.

In some embodiments, the composition includes a pharmaceutically acceptable carrier (e.g., phosphate buffered saline, saline, or bacteriostatic water). Upon formulation, solutions will be administered in a manner compatible with the dosage formulation and in such amount as is therapeutically effective. The formulations are easily administered in a variety of dosage forms such as injectable solutions, injectable gels, drug-release capsules, and the like.

As used herein, the term “pharmaceutically acceptable carrier” includes solvents, dispersion media, coatings, antibacterial agents, antifungal agents, and the like that are compatible with pharmaceutical administration. Supplementary active compounds can also be incorporated into any of the compositions described herein.

In some embodiments, a single dose of any of the compositions described herein can include a total amount (e.g., total sum amount of the at least two different AAV vectors, or the total amount of the single AAV vector) of at least 1 ng, at least 2 ng, at least 4 ng, about 6 ng, about 8 ng, at least 10 ng, at least 20 ng, at least 30 ng, at least 40 ng, at least 50 ng, at least 60 ng, at least 70 ng, at least 80 ng, at least 90 ng, at least 100 ng, at least 200 ng, at least 300 ng, at least 400 ng, at least 500 ng, at least 1 μ g, at least 2 μ g, at least 4 μ g, at least 6 μ g, at least 8 μ g, at least 10 μ g, at least 12 μ g, at least 14 μ g, at least 16 μ g, at least 18 μ g, at least 20 μ g, at least 22 μ g, at least 24 μ g, at least 26 μ g, at least 28 μ g, at least 30 μ g, at least 32 μ g, at least 34 μ g, at least 36 μ g, at least 38 μ g, at least 40 μ g, at least 42 μ g, at least 44 μ g, at least 46 μ g, at least 48 μ g, at least 50 μ g, at least 52 μ g, at least 54 μ g, at least 56 μ g, at least 58 μ g, at least 60 μ g, at least 62 μ g, at least 64 μ g, at least 66 μ g, at least 68 μ g, at least 70 μ g, at least 72 μ g, at least 74 μ g, at least 76 μ g, at least 78 μ g, at least 80 μ g, at least 82 μ g, at least 84 μ g, at least 86 μ g, at least 88 μ g, at least 90 μ g, at least 92 μ g, at least 94 μ g, at least 96 μ g, at least 98 μ g, at least 100 μ g, at least 102 μ g, at least 104 μ g, at least 106 μ g, at least 108 μ g, at least 110 μ g, at least 112 μ g, at least 114 μ g, at least 116 μ g, at least 118 μ g, at least 120 μ g, at least 122 μ g, at least 124 μ g, at least 126 μ g, at least 128 μ g, at least 130 μ g, at least 132 μ g, at least 134 μ g, at least 136 μ g, at least 138 μ g, at least 140 μ g, at least 142 μ g, at least 144 μ g, at least 146 μ g, at least 148

μg, at least 150 μg, at least 152 μg, at least 154 μg, at least 156 μg, at least 158 μg, at least 160 μg, at least 162 μg, at least 164 μg, at least 166 μg, at least 168 μg, at least 170 μg, at least 172 μg, at least 174 μg, at least 176 μg, at least 178 μg, at least 180 μg, at least 182 μg, at least 184 μg, at least 186 μg, at least 188 μg, at least 190 μg, at least 192 μg, at least 194 μg, at least 196 μg, at least 198 μg, or at least 200 μg, e.g., in a buffered solution.

The compositions provided herein can be, e.g., formulated to be compatible with their intended route of administration. A non-limiting example of an intended route of administration is local administration (e.g., intra-cochlear administration). In some embodiments, the therapeutic compositions are formulated to include a lipid nanoparticle. In some embodiments, the therapeutic compositions are formulated to include a polymeric nanoparticle. In some embodiments, the therapeutic compositions are formulated to comprise a synthetic perilymph solution. An exemplary synthetic perilymph solution includes 20-200 mM NaCl; 1-5 mM KCl; 0.1-10 mM CaCl₂; 1-10 mM glucose; 2-50 mM HEPES, having a pH of between about 6 and about 9.

Also provided are kits including any of the compositions described herein. In some embodiments, a kit can include a solid composition (e.g., a lyophilized composition including the single AAV vector or the at least two different vectors described herein) and a liquid for solubilizing the lyophilized composition. In some embodiments, a kit can include a pre-loaded syringe including any of the compositions described herein.

In some embodiments, the kit includes a vial comprising any of the compositions described herein (e.g., formulated as an aqueous composition, e.g., an aqueous pharmaceutical composition).

In some embodiments, the kits can include instructions for performing any of the methods described herein.

Devices and Surgical Methods

Provided herein are therapeutic delivery systems for treating non-syndromic sensorineural hearing loss, or syndromic sensorineural hearing loss. In one aspect, the therapeutic delivery systems include i) a medical device capable of creating one or a plurality of incisions in a round window membrane of an inner ear of a primate in need thereof, and ii) an effective dose of a composition (e.g., any of the compositions described herein). In some

embodiments, the medical device includes a plurality of micro-needles.

Also provided herein are surgical methods for treatment of hearing loss (e.g., non-syndromic sensorineural hearing loss, syndromic sensorineural hearing loss). In some embodiments, the methods include the steps of: introducing into a cochlea of a primate first
5 incision at a first incision point; and administering intra-cochlearly a therapeutically effective amount of any of the compositions provided herein. In some embodiments, the composition is administered to the primate at the first incision point. In some embodiments, the composition is administered to the primate into or through the first incision.

In some embodiments of any of the methods described herein, any of the compositions
10 described herein is administered to the primate into or through the cochlea oval window membrane. In some embodiments of any of the methods described herein, any of the compositions described herein is administered to the primate into or through the cochlea round window membrane. In some embodiments of any of the methods described herein, the composition is administered using a medical device capable of creating a plurality of incisions in
15 the round window membrane. In some embodiments, the medical device includes a plurality of micro-needles. In some embodiments, the medical device includes a plurality of micro-needles including a generally circular first aspect, where each micro-needle has a diameter of at least about 10 microns. In some embodiments, the medical device includes a base and/or a reservoir capable of holding the composition. In some embodiments, the medical device includes a
20 plurality of hollow micro-needles individually including a lumen capable of transferring the composition. In some embodiments, the medical device includes a means for generating at least a partial vacuum.

The invention is further described in detail by reference to the following experimental examples. These examples are provided for purposes of illustration only, and are not intended to
25 be limiting unless otherwise specified. Thus, the invention should in no way be construed as being limited to the following examples, but rather should be construed to encompass any and all variations that become evident as a result of the teaching provided herein.

Without further description, it is believed that one of ordinary skill in the art can, using the preceding description and the following illustrative examples, make and utilize the
30 compounds of the present invention and practice the claimed methods. The following working

examples specifically point out various aspects of the present invention, and are not to be construed as limiting in any way the remainder of the disclosure.

EXAMPLES

Example 1. AAV single vector injection into the inner ear

Immunofluorescent staining was performed on cochlear tissue of a cynomolgus macaque (non-human primate) following administration of a single Anc80-GFP AAV vector directly into the inner ear through the round window.

The cochlear tissue from the treated macaque was processed for immunofluorescence analysis using Myo7a as a marker for hair cells and Iba-1 as a marker for macrophages. The middle turn is representative of the entire sensory epithelium. The data in Figures 1A-1C show clear GFP expression in both the hair cells and the supporting cells, including the following supporting cell subtypes: Hensen's cells (HC), Claudius cells (CC), Dieter cells (DC), inner and outer pillar cells (OPC/IPC), inner border cells, and inner phalangeal cells (IPHC/IBC). These data demonstrate successful Anc80-GFP AAV vector transduction into different cell types of the inner ear sensory epithelium, and the resulting expression of the encoded reporter gene (GFP) in these different cell types. These data indicate the present claimed compositions including a single AAV vector or two or more AAV vectors can be used to express a gene in hairs cells and supporting cells, and can be used to repair a mutation in a gene in hair cells and supporting cells. Figures 2A and 2B are representative images of Anc80-GFP immunofluorescent staining of the cochlear tissue. As shown in Figure 2B, expression is detected in inner hair cells.

Example 2. Exemplary vectors for promoting differentiation of a supporting cell of an inner ear of a primate into a hair cell

As shown in Figure 3, progenitor cells differentiate into either supporting cells or hair cells. Expression of Notch 1 and Hes1/5 in progenitor cells leads to the generation of supporting cells, whereas expression of Atoh1 and Wnt in progenitor cells leads to the generation of hair cells. Figures 4A-4C are exemplary vectors that can be used to promote differentiation of a supporting cell. Figure 4D is an exemplary vector that encodes a shRNA that decreases the expression of a hair cell differentiation-suppressing protein in a primate cell. The data in Figure

5A shows the relative mRNA expression levels of Hes1 in HEK293 cells that were transfected with a vector encoding S3 (SEQ ID NO: 68), a vector encoding S5 (SEQ ID NO: XX), a vector encoding Kop (SEQ ID NO: 75), vectors encoding S3 plus S5, vectors encoding S3 plus Kop and vectors encoding S5 and Kop. Relative expression was determined using RTqPCR. Cells
 5 transfected with the dual vectors show increased reduction in Hes1 mRNA levels. The data in Figure 5B shows reduced Hes1 protein levels in these same cells as determined by Western blotting. Taken together, the data in Figures 5A and 5B confirms the ability of vectors to decrease target mRNA and protein levels.

The data in Figures 6A and 6B show overexpression of ATOH1, POU4F3 and GFI-1 in
 10 HEK293FT cells that were transfected with the vectors of Figures 4A-4D. As shown in Figure 6A, overexpression of POU4F3 in HEK293FT cells also led to an increase in ATOH1 and GFI-1 mRNA levels. Figure 6B shows overexpression of ATOH1, GFI-1 and POU4F3 in HEK293FT cells, respectively.

HEK293FT cells were transfected with mScarlet and mScarlet-DD vectors (Figures 7A-
 15 7B). The data in Figures 8A and 8B show the functionality and reversibility of the destabilizing domain (DD) using fluorescence microscopy and flow cytometry, respectively. As shown in Figure 8A, the percentage of mScarlet positive cells increased proportionately with increasing concentration of TMP in mScarlet-DD transfected HEK293FT cells, whereas the percentage of mScarlet positive cells remained constant regardless of TMP concentration in mScarlet
 20 transfected HEK293FT cells. As shown in Figures 9A and 9B, mScarlet expression was seen in all HEK293FT cells transfected with mScarlet, whereas mScarlet expression was primarily seen in mScarlet-DD transfected cells in the presence of TMP. Figure 10 displays the same response in cochlear explants, where transduction and subsequent expression of mScarlet is seen in hair cells and supporting cells, whereas expression of mScarlet-DD is only seen in the presence of
 25 TMP.

Figures 11A and 11B are exemplary combined vectors that can be used to promoter differentiation of a supporting cells. The vectors are combined from vectors of Figures 4A-C.

The data in Figures 12A and 12B show overexpression of ATOH1 and POU4F3 and reduction in HES1 mRNA and protein respectively, after transfection with the vectors of Figures
 30 11A and 11B.

OTHER EMBODIMENTS

It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will control. In addition, section headings, the materials, methods, and examples are illustrative only and not intended to be limiting.

What is claimed is:

1. A composition comprising at least two different nucleic acid vectors, wherein:
 - each of the at least two different adeno-associated virus (AAV) vectors comprises a coding sequence that encodes a different portion of a hair cell differentiation protein, each of the encoded portions being at least 30 amino acid residues in length, wherein the amino acid sequence of each of the encoded portions may optionally partially overlap with the amino acid sequence of a different one of the encoded portions;
 - no single vector of the at least two different vectors encodes the full-length hair cell differentiation protein;
 - at least one of the coding sequences comprises a nucleotide sequence spanning two neighboring exons of hair cell differentiation genomic DNA, and lacks an intronic sequence between the two neighboring exons; and
 - when introduced into a primate cell the at least two different vectors undergo concatamerization or homologous recombination with each other, thereby forming a recombined nucleic acid that encodes a full-length hair cell differentiation protein that is expressed in the primate cell.
2. The composition of claim 1, wherein the amino acid sequence of none of the encoded portions overlaps with the amino acid sequence of a different one of the encoded portions.
3. The composition of claim 1, wherein the amino acid sequence of each of the encoded portions partially overlaps with the amino acid sequence of a different one of the encoded portions.
4. The composition of claim 3, wherein the overlapping amino acid sequence is between 30 amino acid residues to about 390 amino acid residues in length.
5. The composition of any one of claim 1-4, wherein the vectors include two different vectors, each of which comprises a different segment of an intron, wherein the intron comprises the nucleotide sequence of an intron that is present in a hair cell differentiation genomic DNA, and wherein the two different segments overlap in sequence by at least 100 nucleotides.

6. The composition of claim 5, wherein the two different intron segments overlap in sequence by about 100 nucleotides to about 800 nucleotides.
7. The composition of any one of claims 1-6, wherein the entire nucleotide sequence of each of the at least two different vectors is between about 500 nucleotides to about 10,000 nucleotides in length.
8. The composition of claim 7, wherein the entire nucleotide sequence of each of the at least two different vectors is between about 500 nucleotides to about 5,000 nucleotides in length.
9. The composition of any one of claims 1-8, wherein the number of different vectors in the composition is two.
10. The composition of claim 9, wherein a first of the two different vectors comprises a coding sequence that encodes an N-terminal portion of the hair cell differentiation protein.
11. The composition of claim 10, wherein the N-terminal portion of the hair cell differentiation protein is between about 30 amino acids to about 750 amino acids in length.
12. The composition of claim 11, wherein the N-terminal portion of the hair cell differentiation protein is between about 30 amino acids to about 320 amino acids in length.
13. The composition of any one of claims 10-12, wherein the first vector further comprises one or both of a promoter and a Kozak sequence.
14. The composition of claim 13, wherein the first vector comprises a promoter that is an inducible promoter, a constitutive promoter, or a tissue-specific promoter.
15. The composition of any one of claim 10-14, wherein the second of the two different vectors comprises a coding sequence that encodes a C-terminal portion of the hair cell differentiation protein.
16. The composition of claim 15, wherein the C-terminal portion of the hair cell differentiation protein is between about 30 amino acids to about 750 amino acids in length.

17. The composition of claim 16, wherein the C-terminal portion of the hair cell differentiation portion is between about 30 amino acids to about 320 amino acids in length.

18. The composition of any one of claim 15-17, wherein the second vector further comprises a poly(dA) sequence.

19. The composition of any one of claims 15-18, wherein the second vector further comprises a destabilizing sequence.

20. The composition of claim 19, wherein the second vector further comprises a FKB12 destabilizing sequence.

21. A composition comprising two different nucleic acid vectors, wherein:

a first nucleic acid vector of the two different nucleic acid vectors comprises a promoter, a first coding sequence that encodes an N-terminal portion of a hair cell differentiation protein positioned 3' of the promoter, and a splicing donor signal sequence positioned at the 3' end of the first coding sequence; and

a second nucleic acid vector of the two different nucleic acid vectors comprises a splicing acceptor signal sequence, a second coding sequence that encodes a C-terminal portion of a hair cell differentiation protein positioned at the 3' end of the splicing acceptor signal sequence, and a polyadenylation sequence at the 3' end of the second coding sequence;

wherein each of the encoded portions is at least 30 amino acid residues in length, wherein the amino acid sequences of the encoded portions do not overlap, wherein no single vector of the two different vectors encodes the full-length hair cell differentiation protein, and, when the coding sequences are transcribed in a primate cell, to produce RNA transcripts, splicing occurs between the splicing donor signal sequence on one transcript and the splicing acceptor signal sequence on the other transcript, thereby forming a recombined RNA molecule that encodes a full-length hair cell differentiation protein.

22. The composition of claim 21, wherein at least one of the coding sequences comprises a nucleotide sequence spanning two neighboring exons of a hair cell differentiation genomic DNA, and lacks an intronic sequence between the two neighboring exons.

23. A composition comprising:

a first nucleic acid vector comprising a promoter, a first coding sequence that encodes an N-terminal portion of a hair cell differentiation protein positioned 3' of the promoter, a splicing donor signal sequence positioned at the 3' end of the first coding sequence, and a first detectable marker gene positioned 3' of the splicing donor signal sequence; and

a second nucleic acid vector, different from the first nucleic acid vector, comprising a second detectable marker gene, a splicing acceptor signal sequence positioned 3' of the second detectable marker gene, a second coding sequence that encodes a C-terminal portion of a hair cell differentiation protein positioned at the 3' end of the splicing acceptor signal sequence, and a polyadenylation sequence positioned at the 3' end of the second coding sequence;

wherein each of the encoded portions is at least 30 amino acid residues in length, wherein the respective amino acid sequences of the encoded portions do not overlap with each other, wherein no single vector of the two different vectors encodes the full-length hair cell differentiation protein, and, when the coding sequences are transcribed in a primate cell to produce RNA transcripts, splicing occurs between the splicing donor signal on one transcript and the splicing acceptor signal on the other transcript, thereby forming a recombined RNA molecule that encodes a full-length hair cell differentiation protein.

24. The composition of claim 23, wherein at least one of the coding sequences comprises a nucleotide sequence spanning two neighboring exons of a hair cell differentiation genomic DNA, and lacks an intronic sequence between the neighboring exons.

25. The composition of claim 23, wherein the first or second detectable marker gene is alkaline phosphatase.

26. The composition of claim 23 or 25, wherein the first and second detectable marker genes are the same.

27. A composition comprising:

a first nucleic acid vector comprising a promoter, a first coding sequence that encodes an N-terminal portion of a hair cell differentiation protein positioned 3' to the promoter, a splicing

donor signal sequence positioned at the 3' end of the first coding sequence, and a F1 phage recombinogenic region positioned 3' to the splicing donor signal sequence; and

a second nucleic acid vector, different from the first nucleic acid vector, comprising a second F1 phage recombinogenic region, a splicing acceptor signal sequence positioned 3' of the second F1 phage recombinogenic region, a second coding sequence that encodes a C-terminal portion of a hair cell differentiation protein positioned at the 3' end of the splicing acceptor signal sequence, and a polyadenylation sequence positioned at the 3' end of the second coding sequence;

wherein each of the encoded portions is at least 30 amino acid residues in length, wherein the respective amino acid sequences of the encoded portions do not overlap with each other, wherein no single vector of the two different vectors encodes the full-length hair cell differentiation protein, and, when the coding sequences are transcribed in a primate cell to produce RNA transcripts, splicing occurs between the splicing donor signal one transcript and the splicing acceptor signal on the other transcript, thereby forming a recombined RNA molecule that encodes a full-length hair cell differentiation protein.

28. The composition of claim 27, wherein at least one of the coding sequences comprises a nucleotide sequence spanning two neighboring exons of a hair cell differentiation genomic DNA, and lacks an intronic sequence between the two neighboring exons.

29. A composition comprising a single adeno-associated virus (AAV) vector, where the single AAV vector comprises a nucleic acid sequence that encodes a hair cell differentiation protein; and

when introduced into a primate cell, a nucleic acid encoding the hair cell differentiation protein is generated at the locus of the hair cell differentiation gene and the primate cell expresses the hair cell differentiation protein.

30. The composition of any one of claims 1-29, wherein the hair cell differentiation gene is selected from the group consisting of: atonal bHLH transcription factor 1 (ATOH1), POU Class 4 Homeobox 3 (POU4F3), catenin beta 1 (CTNNB1), Noggin (NOG), growth factor

independent 1 transcriptional repressor (GFI-1), neurotrophin 3 (NTF3) and brain-derived neurotrophic factor (BDNF).

31. A composition comprising two different nucleic acid vectors, wherein:

a first nucleic acid vector comprises a first nucleic acid sequence that encodes a first hair cell differentiation protein; and

a second nucleic acid vector comprises a second nucleic acid sequence that encodes a second hair cell differentiation protein, wherein the first hair cell differentiation protein and the second hair cell differentiation protein are different proteins, and

when introduced into a primate cell, the first nucleic acid and the second nucleic acid encoding the first hair cell differentiation protein and the second hair cell differentiation protein are generated at the locus of the hair cell differentiation gene and the primate cell expresses the first hair cell differentiation protein and the second hair cell differentiation protein.

32. The composition of claim 31, wherein the first and the second hair cell differentiation proteins are selected from the group consisting of: atonal bHLH transcription factor 1 (ATOH1), POU Class 4 Homeobox 3 (POU4F3), catenin beta 1 (CTNNB1), Noggin (NOG), growth factor independent 1 transcriptional repressor (GFI-1), neurotrophin 3 (NTF3) and brain-derived neurotrophic factor (BDNF).

33. The composition of any one of claims 21-32, wherein the second nucleic acid vector further comprises a destabilizing sequence.

34. The composition of claim 33, wherein the second nucleic acid vector further comprises a FKB12 destabilizing sequence.

35. A composition comprising at least one adeno-associated virus (AAV) vector that encodes an inhibitory nucleic acid that decreases the expression of a hair cell differentiation-suppressing protein in a primate cell.

36. The composition of claim 35, wherein the inhibitory nucleic acid is a short interfering RNA (siRNA), a short hairpin RNA (shRNA), an antisense oligonucleotide, or a ribozyme.

37. The composition of claim 30 or 32, wherein the hair cell differentiation-suppressing gene is hes family bHLH transcription factor 1 (HES1), hes family bHLH transcription factor 5 (HES5), sex determining region Y-box 2 (SOX2), and p27kip (CDKN1B).

38. The composition of any one of claims 1-37, further comprising a pharmaceutically acceptable excipient.

39. A kit comprising a composition of any one of claims 1-38.

40. A kit of claim 39, further comprising a pre-loaded syringe comprising the composition.

41. A method of promoting differentiation of a supporting cell of an inner ear of a primate into a hair cell, the method comprising:

administering to the inner ear of the primate a therapeutically effective amount of a composition of any one of claims 1-38,

wherein the administering promotes differentiation of the supporting cell of the inner ear of the primate into a hair cell.

42. A method of increasing the expression level of a hair cell differentiation protein in a supporting cell or a hair cell of an inner ear of a primate, the method comprising administering to the inner ear of the primate a therapeutically effective amount of a composition of any one of claims 1-32,

wherein the administering results in an increase in the expression level of the hair cell differentiation protein in the supporting cell or the hair cell of the inner ear of the primate.

43. The method of claim 42, wherein the hair cell differentiation protein is selected from the group consisting of: Atoh1, Pou4f3, β -Catenin, Noggin, GFI-1, NTF3, and BDNF.

44. The method of claim 42, wherein the primate has previously been determined to have a defective hair cell differentiation gene.

45. A method of decreasing the expression level of a hair cell differentiation-suppressing protein in a supporting cell or a hair cell of an inner ear of a primate, the method comprising administering to the inner ear of the primate a therapeutically effective amount of a composition of any one of claims 35-37,

wherein the administering results in a decrease in the expression level of the hair cell differentiation-suppressing protein in the supporting cell or the hair cell of the inner ear of the primate.

46. A method of increasing the number of functional hair cells in a primate in need thereof, the method comprising administering to the inner ear of the primate a therapeutically effective amount of a composition of any one of claims 1-38.

47. A method of improving hearing in a primate in need thereof, the method comprising administering to the inner ear of the primate a therapeutically effective amount of a composition of any one of claims 1-38.

48. The method of claim 47, further comprising, prior to the administering step, determining that the primate has a defective hair cell differentiation gene.

49. A method of repairing a hair cell toxicity-inducing mutation in an endogenous hair cell differentiation gene locus in a supporting cell or a hair cell of an inner ear of a primate, the method comprising administering to the inner ear of the primate a therapeutically effective amount of a composition of any one of claims 1-32,

wherein the administering results in repair of the hair cell toxicity-inducing mutation in the endogenous hair cell differentiation gene locus in the supporting cell or the hair cell of the inner ear of the primate.

50. A method of decreasing the risk of hearing loss due to hair cell loss or dysfunction in a primate in need thereof, the method comprising administering to the inner ear of the primate a therapeutically effective amount of a composition of any one of claims 1-38.

51. The method of claim 50, wherein the primate has been previously identified as having a defective hair cell differentiation gene.

1/10

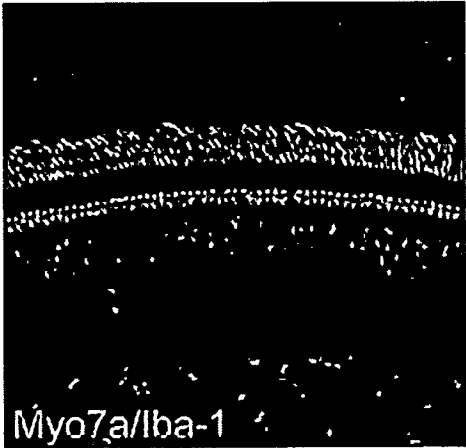


Figure 1A

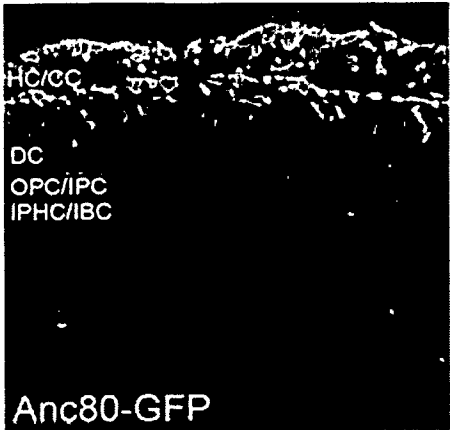


Figure 1B

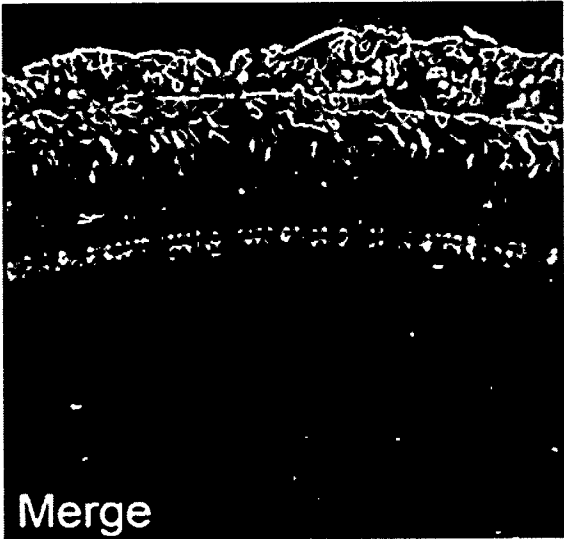


Figure 1C

2/10

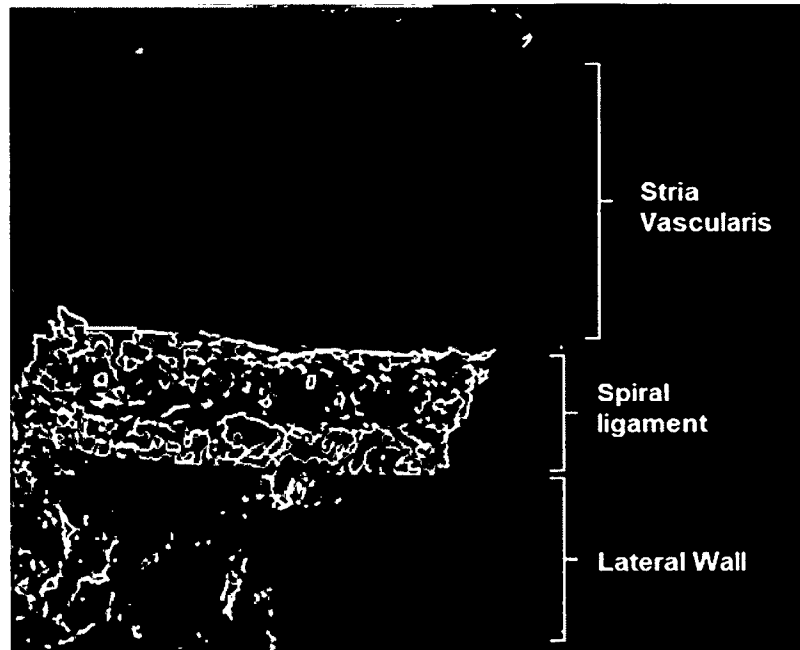


Figure 2A

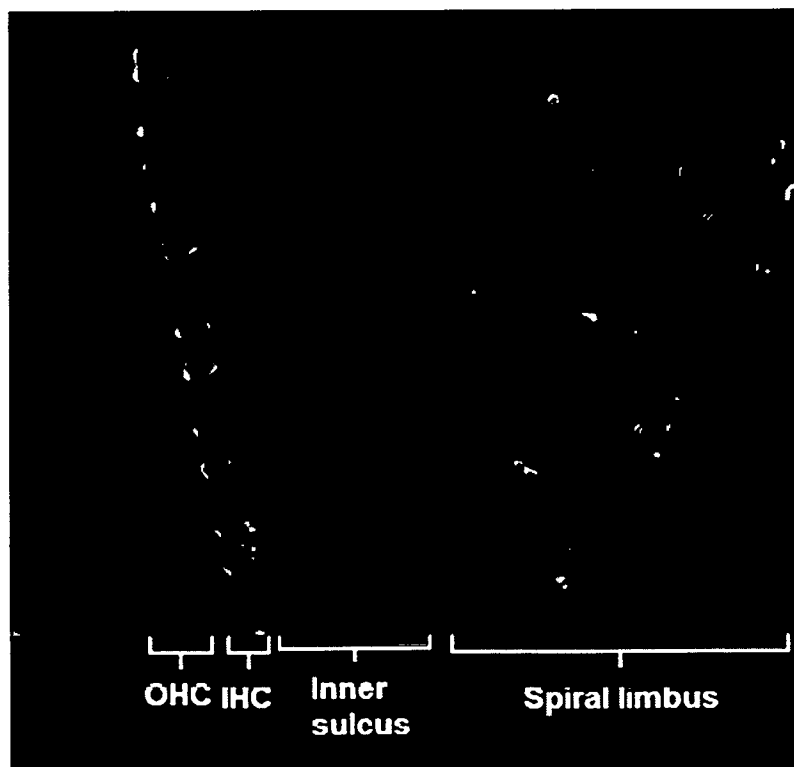


Figure 2B

3/10

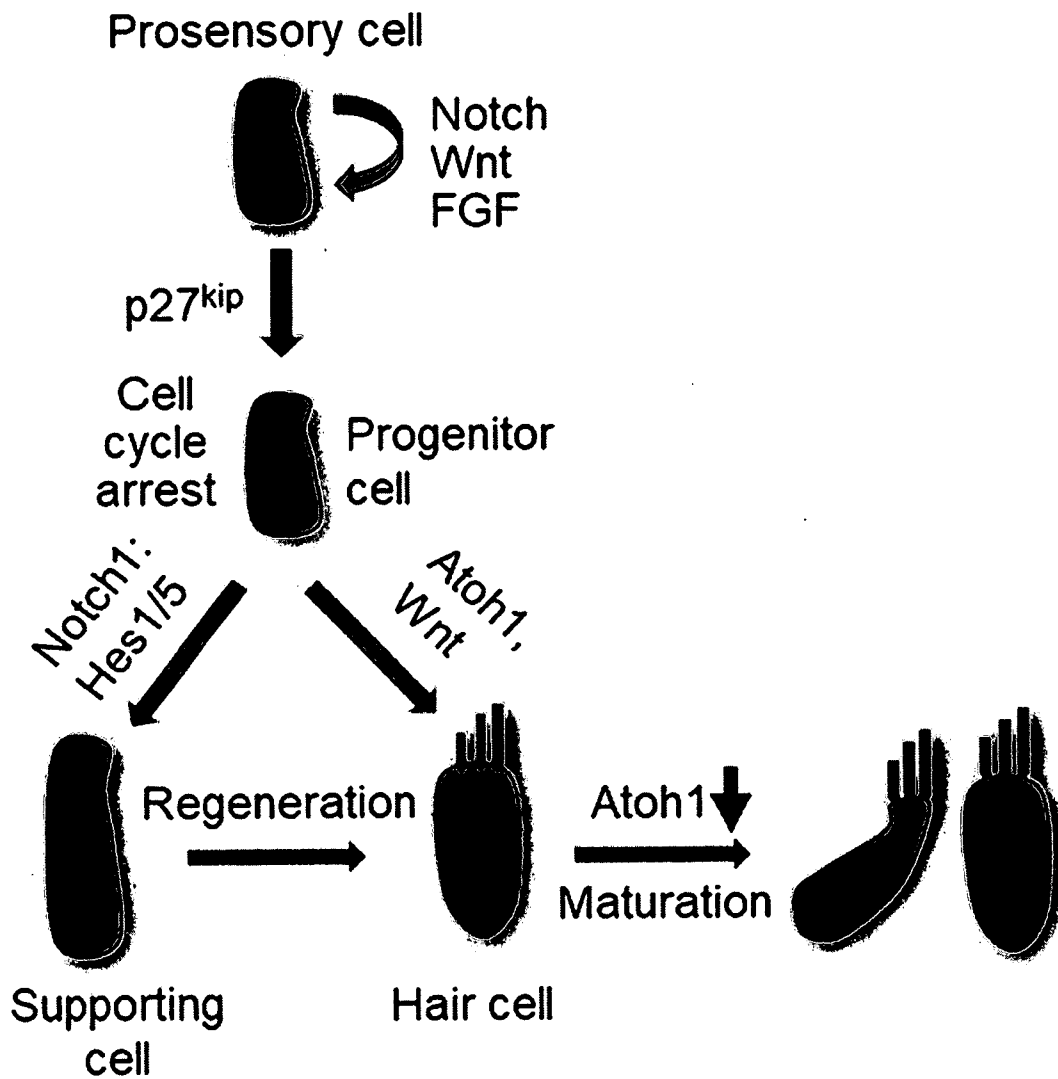


Figure 3

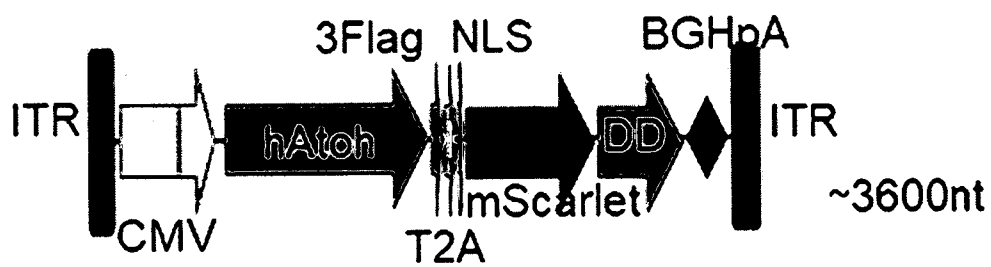


Figure 4A

4/10

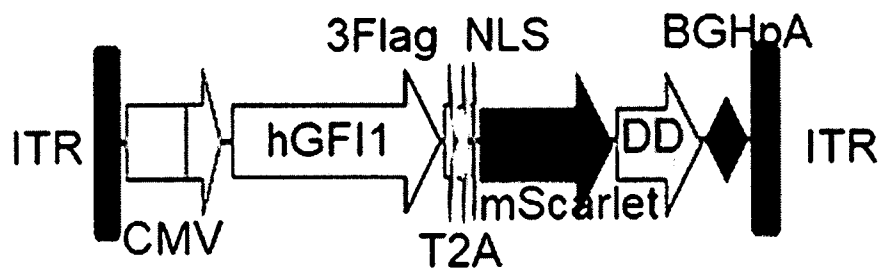


Figure 4B

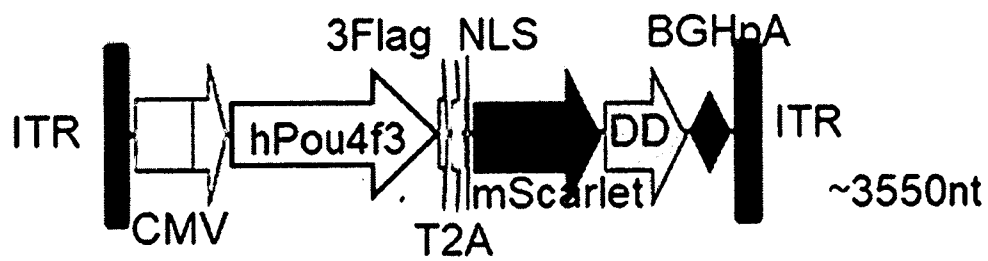


Figure 4C

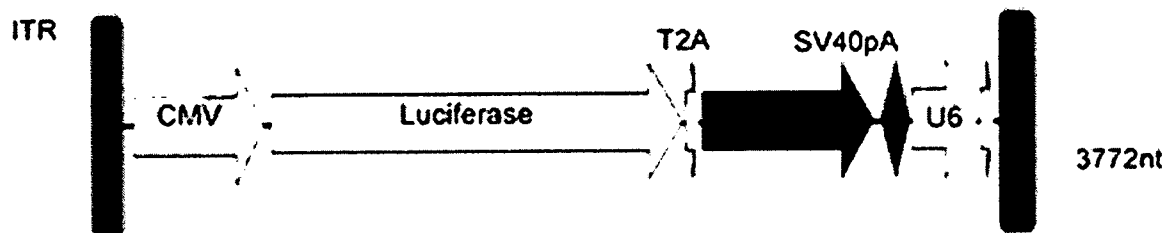


Figure 4D

5/10

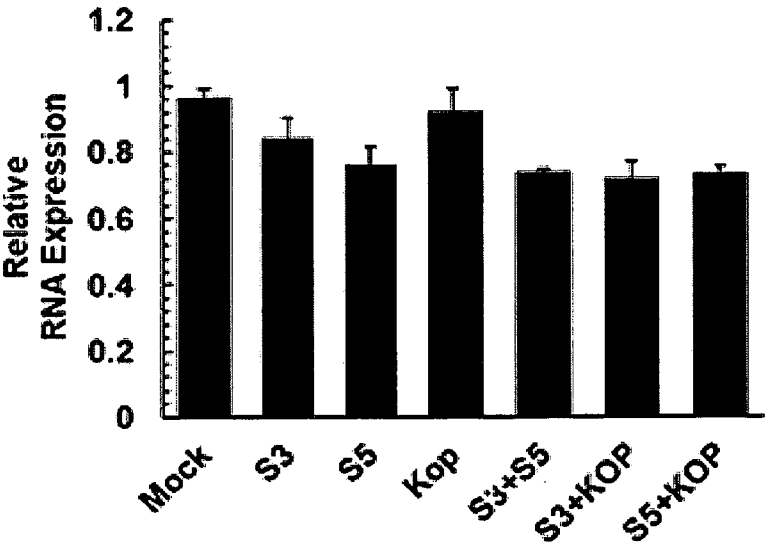


Figure 5A

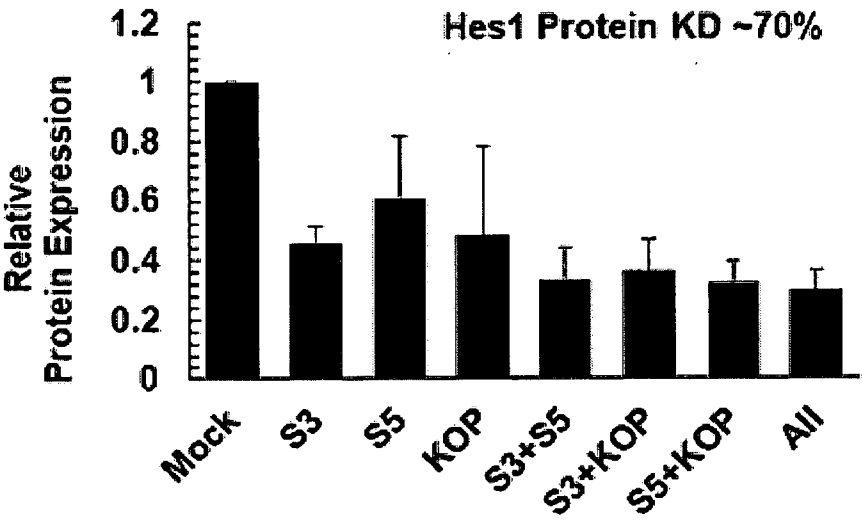


Figure 5B

6/10

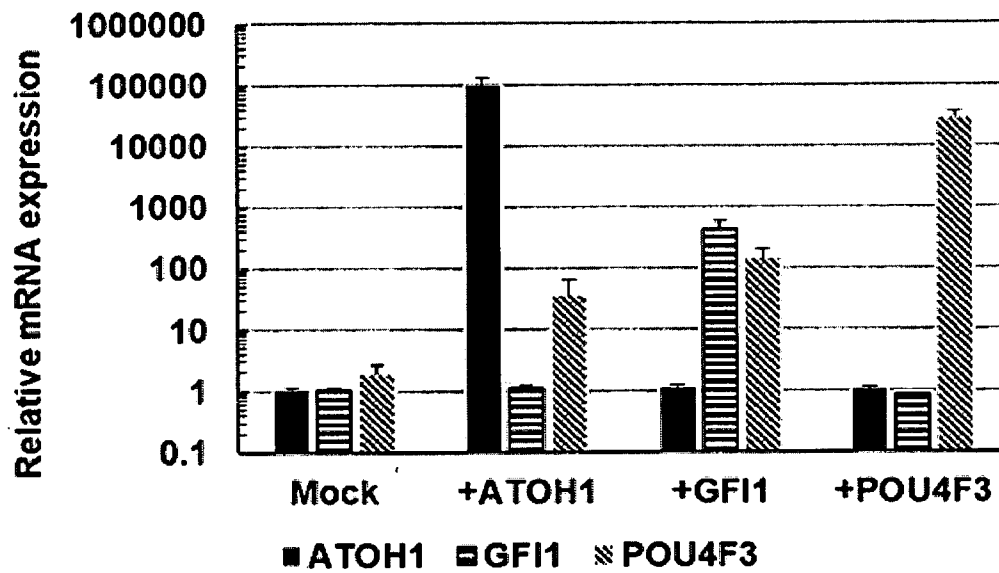
ATOH1, POU4F3 and GFI-1 overexpression in HEK293ft

Figure 6A

Mock +A +P +G +AP +AG +APG

GFI1
ATOH1
POU4F3

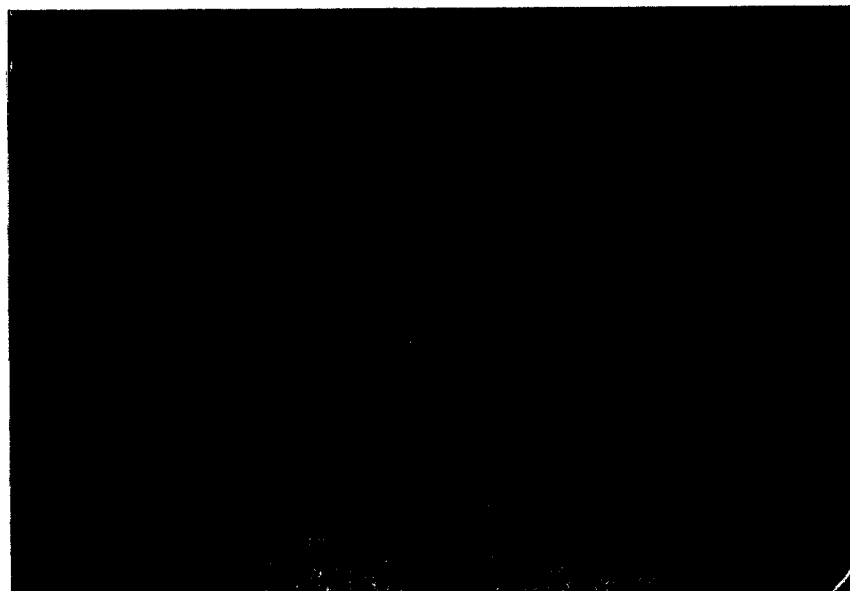


Figure 6B

7/10

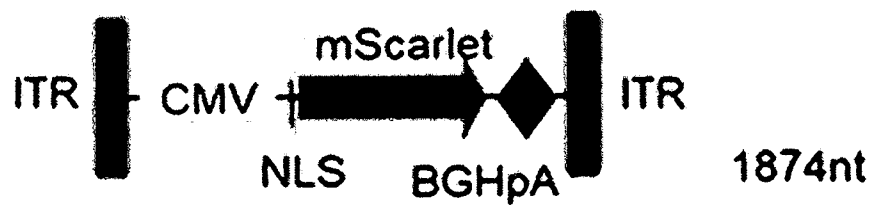


Figure 7A

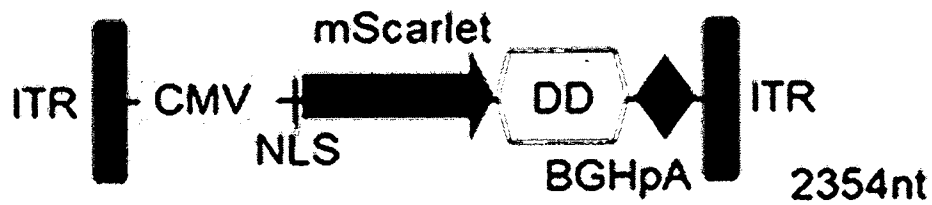


Figure 7B

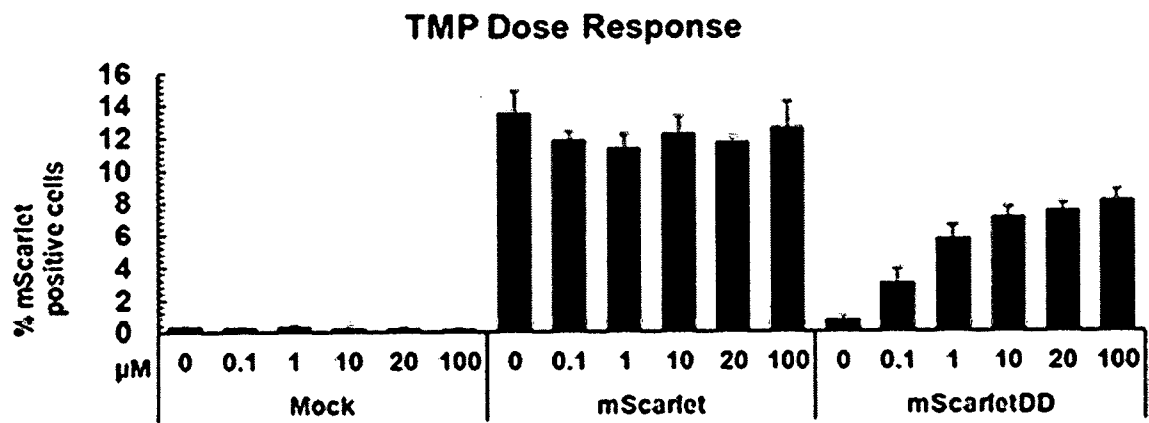


Figure 8A

8/10

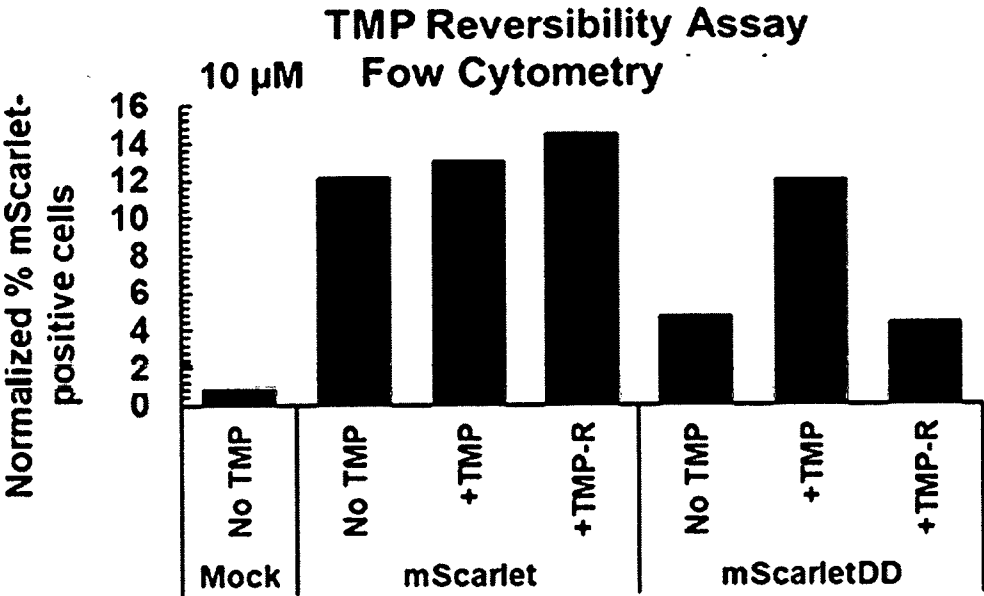


Figure 8B

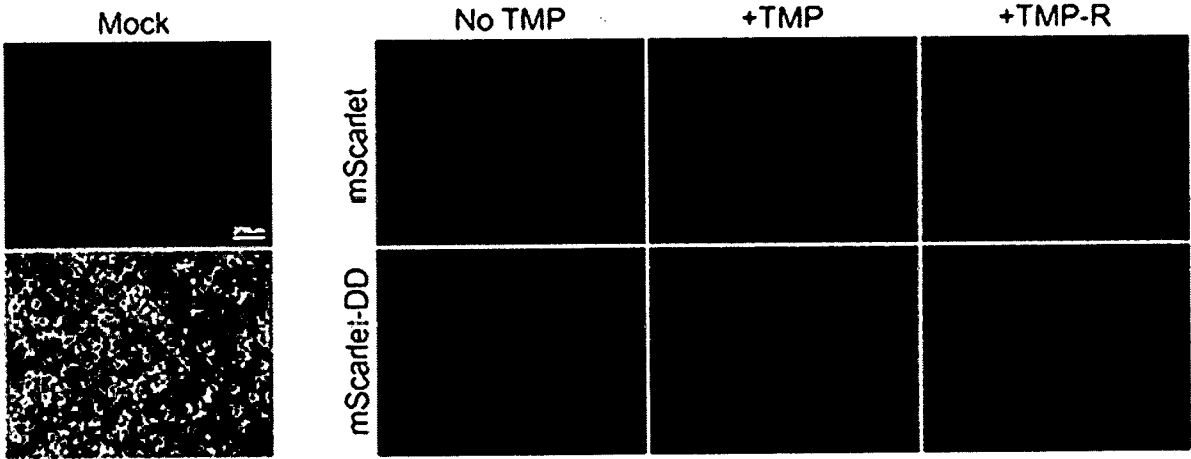


Figure 9A

Figure 9B

9/10

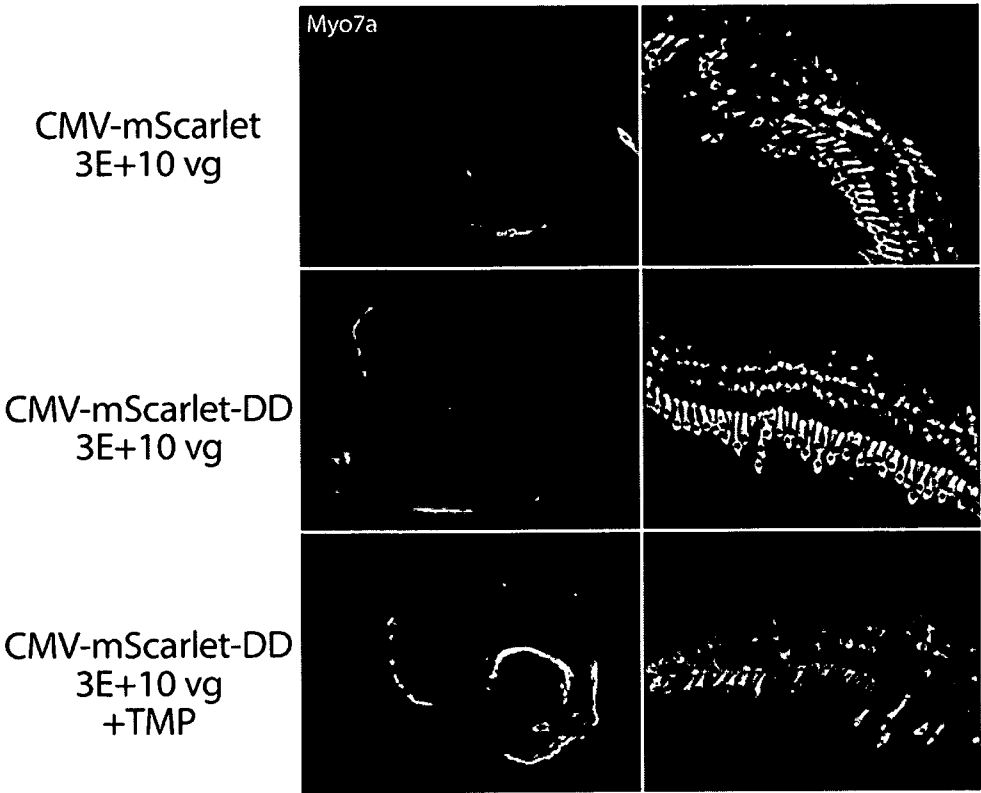


Figure 10

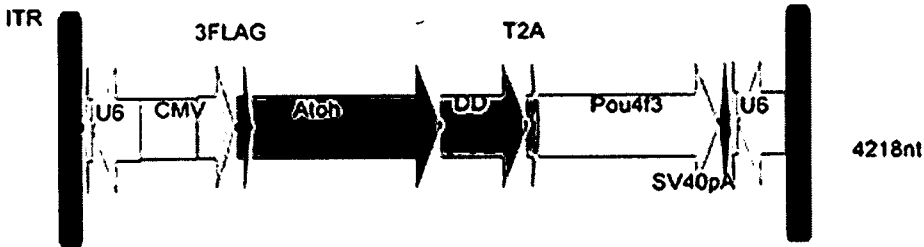


Figure 11A

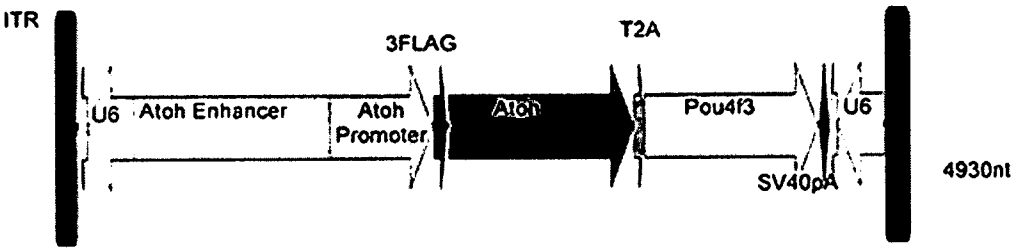


Figure 11B

10/10

Transfection of combined plasmid in HEK

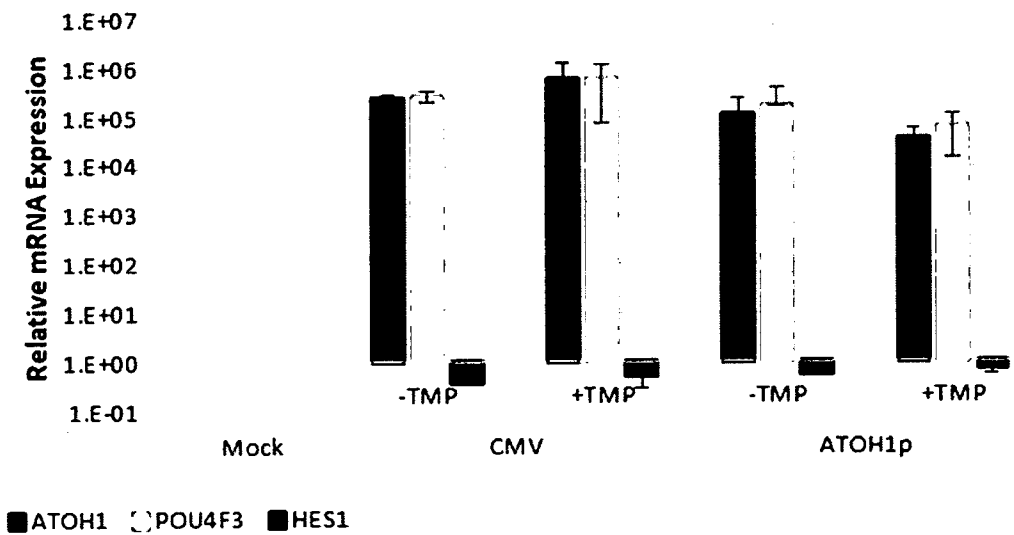


Figure 12A

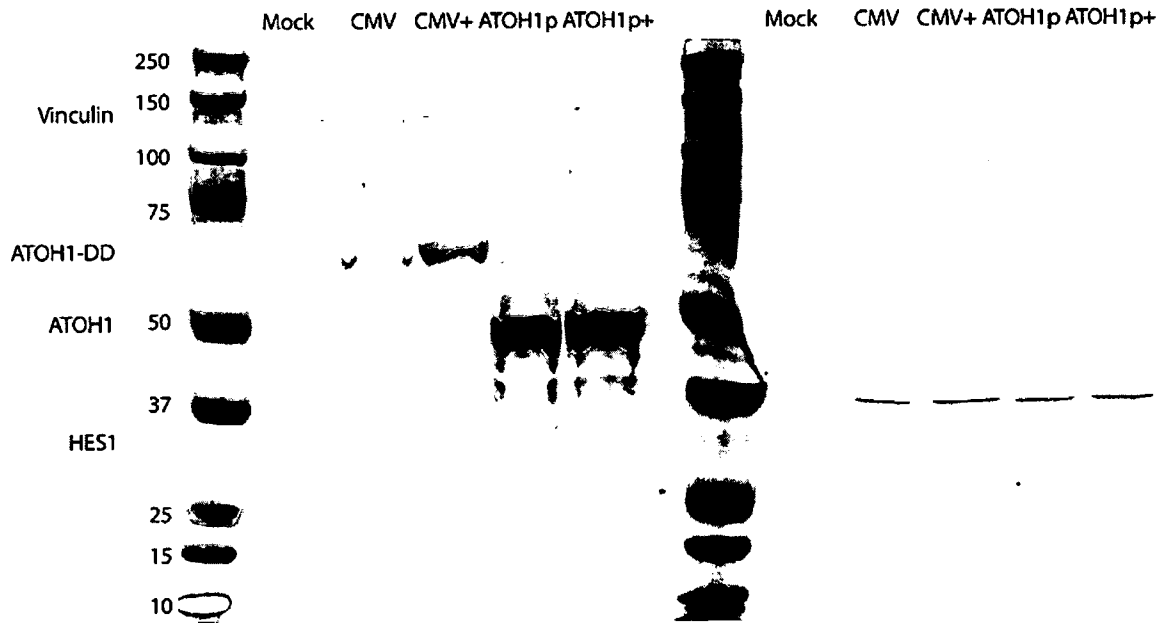


Figure 12B

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2019/060324

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.: 41-51
because they relate to subject matter not required to be searched by this Authority, namely:
Claims 41-51 pertain to methods for treatment of the human body by therapy (PCT Article 17(2)(a)(i) and Rule 39.1(iv))
2. ☒ Claims Nos.: 8, 10-12, 14, 16-17, 20, 34, 40, 43-44, 48, 51
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
Claims 8, 10-12, 14, 16-17, 20, 34, 40, 43-44, 48, 51 are regarded to be unclear because they refer to claims which do not comply with PCT Rule 6.4(a).
3. ☒ Claims Nos.: 7, 9, 13, 15, 18-19, 30, 33, 37-39, 41-42, 45-47, 49-50
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of any additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2019/060324**A. CLASSIFICATION OF SUBJECT MATTER****C12N 15/86(2006.01)i, C12N 15/113(2010.01)i, C12N 15/65(2006.01)i, C07K 14/475(2006.01)i, A61K 48/00(2006.01)i, A61P 27/16(2006.01)i**

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12N 15/86; A61K 48/00; C07K 14/705; C12N 15/85; C12N 15/113; C12N 15/65; C07K 14/475; A61P 27/16

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean utility models and applications for utility models

Japanese utility models and applications for utility models

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS(KIPO internal) & Keywords: dual vector, AAV, hair cell differentiation protein, splicing, overlapping, F1 phage recombinogenic region, alkaline phosphatase

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	AHMED, HENA et al., 'Emerging gene therapies for genetic hearing loss', Journal of the Association for Research in Otolaryngology, 2017, vol. 18, pp. 649-670 See pages 653-654, 657, 659-660.	1-6, 21-22, 29, 31-32, 35-36
Y		23-28
Y	WO 2018-039375 A1 (AKOUOS, INC. et al.) 01 March 2018 See claims 41-46.	23-28
X	BUDENZ, CAMERON L. et al., 'Differential effects of AAV.BDNF and AAV.Ntf3 in the deafened adult guinea pig ear', Scientific Reports, 2015, vol. 5, article no. 8619, pp. 1-8 See pages 1, 5-6.	29, 31-32
X	WO 2017-136764 A1 (THE GENERAL HOSPITAL CORPORATION et al.) 10 August 2017 See claims 1-3, 7, 9-10.	29, 31-32
Y	WO 2017-108931 A1 (INSERM (INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE) et al.) 29 June 2017 See abstract; pages 5-6; claims 1-2.	23-28

☒ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"D" document cited by the applicant in the international application

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

06 March 2020 (06.03.2020)

Date of mailing of the international search report

09 March 2020 (09.03.2020)

Name and mailing address of the ISA/KR

International Application Division

Korean Intellectual Property Office

189 Cheongsa-ro, Seo-gu, Daejeon, 35208, Republic of Korea



Facsimile No. +82-42-481-8578

Authorized officer

HEO, Joo Hyung

Telephone No. +82-42-481-8150



INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2019/060324

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2016-0076054 A1 (FONDAZIONE TELETHON) 17 March 2016 See paragraphs [0015], [0092]–[0093]; claims 1–4; figure 1.	23–28
Y	MCCLEMENTS, MICHELLE E. et al., ‘Adeno-associated virus (AAV) dual vector strategies for gene therapy encoding large transgenes’ , Yale Journal of Biology and Medicine, 2017, vol. 90, pp. 611–623 See pages 615–618.	23–28
Y	TRAPANI, IVANA et al., ‘Effective delivery of large genes to the retina by dual AAV vectors’ , EMBO Molecular Medicine, 2014, vol. 6, no. 2, pp 194–211 See pages 195, 201–202; figure 1.	23–28

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2019/060324

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2018-039375 A1	01/03/2018	CN 110225975 A EP 3510161 A1 JP 2019-530737 A US 2019-0185864 A1	10/09/2019 17/07/2019 24/10/2019 20/06/2019
WO 2017-136764 A1	10/08/2017	US 2019-0038778 A1	07/02/2019
WO 2017-108931 A1	29/06/2017	EP 3393522 A1 JP 2019-500039 A US 2019-0002916 A1	31/10/2018 10/01/2019 03/01/2019
US 2016-0076054 A1	17/03/2016	BR 112015026422 A2 CN 105408352 A EP 2986635 A1 EP 2986635 B1 JP 2016-516424 A JP 6453307 B2 US 10494645 B2 WO 2014-170480 A1	14/11/2017 16/03/2016 24/02/2016 03/10/2018 09/06/2016 16/01/2019 03/12/2019 23/10/2014