

LIVER-SPECIFIC EXPRESSION CASSETTES, VECTORS AND USES THEREOF FOR EXPRESSING THERAPEUTIC PROTEINS

RELATED APPLICATIONS

[0001] This application claims priority to U.S. Provisional Application No. 63/245,013, filed on September 16, 2021, the contents of which is hereby incorporated by reference in its entirety.

BACKGROUND

[0002] Gene therapy aims to improve clinical outcomes for patients suffering from either genetic mutations or acquired diseases caused by an aberration in the gene expression profile. Gene therapy includes the treatment or prevention of medical conditions resulting from defective genes or abnormal regulation or expression, *e.g.*, underexpression or overexpression, that can result in a disorder, disease, malignancy, *etc.* For example, a disease or disorder caused by a defective gene might be treated, prevented or ameliorated by delivery of a corrective genetic material to a patient, or might be treated, prevented or ameliorated by altering or silencing a defective gene, *e.g.*, with a corrective genetic material to a patient resulting in the therapeutic expression of the genetic material within the patient.

[0003] The basis of gene therapy is to supply a transcription cassette with an active gene product (sometimes referred to as a transgene), *e.g.*, that can result in a positive gain-of-function effect, a negative loss-of-function effect, or another outcome. Such outcomes can be attributed to expression of a therapeutic protein such as an antibody, a functional enzyme, or a fusion protein. Gene therapy can also be used to treat a disease or malignancy caused by other factors. Human monogenic disorders can be treated by the delivery and expression of a normal gene to the target cells. Delivery and expression of a corrective gene in the patient's target cells can be carried out via numerous methods, including the use of engineered viruses and viral gene delivery vectors.

[0004] The liver is directly or indirectly involved in many essential processes and is affected by numerous inherited diseases. Therefore, many inherited diseases could be effectively treated by targeting the liver, using gene transfer approaches. However, there are challenges that remain associated with liver-directed gene therapy, including efficiently targeting hepatocytes, maintaining stability of the vector genome, and achieving persistent high level expression. Among the many virus-derived vectors available (*e.g.*, recombinant retrovirus, recombinant lentivirus, recombinant adenovirus, and the like), recombinant adeno-associated virus (rAAV) has gained popularity as a versatile vector in gene therapy. Liver-directed gene therapy clinical trials with AAV vectors have reported clinical efficacy data (Rodriguez-Marquez *et al.*, Expert Opinion on Biological Therapy Volume 21, 2021 - Issue 6). While clinical advances have been made using rAAV vectors for Factor IX (FIX) expression in the liver, the use of rAAV for FVIII expression in hemophilia A patients has been challenging due to ineffective biosynthesis of human FVIII (hFVIII). rAAV vectors produce capsids that have limited space to

encapsulate nucleic acids. FVIII is a large glycoprotein, and the rAAV sequences necessary to encode and express FVIII generally exceed the packaging capacity of the capsid.

[0005] Recombinant capsid-free AAV vectors can be obtained as an isolated linear nucleic acid molecule comprising an expressible transgene and promoter regions flanked by two wild-type AAV inverted terminal repeat sequences (ITRs) including the Rep binding and terminal resolution sites. These recombinant AAV vectors are devoid of AAV capsid protein encoding sequences, and can be single-stranded, double-stranded or duplex with one or both ends covalently linked through the two wild-type ITR palindrome sequences (*e.g.*, WO2012/123430, U.S. Patent 9,598,703). They avoid many of the problems of AAV-mediated gene therapy in that the transgene capacity is much higher, transgene expression onset is rapid, and the patient immune system recognizes the DNA molecules as a virus to be cleared.

[0006] Non-viral gene therapy is assumed to be less toxic for the host and safer in terms of gene delivery compared to a viral vector. One example of a non-viral gene therapy, closed-ended DNA (“ceDNA”) vectors, has many attractive features for gene-based therapy. For example, ceDNA vectors have no packaging constraints imposed by the limiting space within the viral capsid. ceDNA vectors represent a viable eukaryotically-produced alternative to prokaryote-produced plasmid DNA vectors, as opposed to encapsulated AAV genomes. This permits the insertion of control elements, *e.g.*, large transgenes, multiple transgenes, regulatory switched, and incorporation of the native genetic regulatory elements of the transgene, if desired.

[0007] In most living organisms, and especially in eukaryotes with large genome sizes, however, there does not appear to be a driving force to limit enhancer/promoter size, and therefore most endogenous enhancer/promoters span hundreds, and more often thousands, of base pairs (bp) of DNA. Due to their size, these endogenous natural gene enhancers/promoters are generally not amenable to inclusion in gene therapy products due to size limitations.

[0008] Regardless of viral or non-viral delivery, there remains a need for a technology that permits robust expression of a therapeutic protein, such as a liver-specific therapeutic protein, in a cell, tissue or subject, to improve the efficiency and safety of treatment of a genetic disease or disorder.

SUMMARY

[0009] The present disclosure has applied a range of bioinformatic analyses to identify a novel and inventive set of non-natural modifications to a native liver-specific Serpin enhancer region that surprisingly resulted in acute expression level and improved sequence characteristics known to impact expression durability of gene product.

[0010] The disclosure also provides an evolutionary conservation analysis to selective removal of CpGs in the enhancer without disrupting function. Enhancers are often combined in series to drive higher levels of transcription initiation. However, the principals underlying optimal number and orientation of enhancer regions remain not well understood. Spacing between transcription factor

binding sites is likely a key selection attribute that impacts function, especially considering that DNA is a helix such that number of nucleotides between binding sites also changes their rotational spatial orientation. As described herein, a range of enhancer combinations were tested for improved function, including different numbers of enhancers and nucleotide spacer content. Bioinformatic analysis was used to guide the sequence selection of sequence substitutions tested.

[0011] The technology described herein relates to liver-specific nucleic acid expression cassettes comprising specific regulatory elements (enhancer-promoter combination) that have been improved to enhance liver-specific gene expression, such that the native *cis*-regulatory region has been optimized to minimize CpG content and to enhance spacer optimization, and a vector, either a viral vector (*e.g.*, an AAV-based vector), or a non-viral vector (*e.g.*, a ceDNA vector).

[0012] As disclosed herein, the liver-specific expression cassette surprisingly promotes substantially increased protein expression in the liver and in liver cells than in other tissue types, while retaining tissue specificity. In some embodiments, the liver-specific regulatory elements (*e.g.*, enhancer-promoter combination) can be included in a viral vector (such as an adeno-associated virus vector (AAV)) or a non-viral vector a capsid-free (*e.g.*, non-viral) DNA vector with covalently-closed ends (referred to herein as a “closed-ended DNA vector” or a “ceDNA vector”) in operative combination with a heterologous nucleic acid sequence encoding a protein of interest to promote expression of the protein of interest, for example, in liver tissue and /or cells. An advantage of the promoters of the present disclosure is that the enhancer -promoters can be designed and selected for the amount of expression of gene product by the vector, while also ensuring that the amount of promoter is not immunogenic. In some embodiments, the vector (*e.g.*, the AAV vector or ceDNA vector) provides effective expression of the protein of interest at doses that are not predicted to cause immunogenicity in humans. In some embodiments, the vector (*e.g.*, the AAV vector or ceDNA vector) provides effective expression of the protein of interest at doses that are not predicted to cause toxicity in humans. The improvements described herein can be generalized to the improved expression of any transgene (*e.g.*, AAV, ceDNA).

[0013] In a first aspect, the disclosure relates to a liver-specific nucleic acid regulatory element comprising a nucleic acid sequence having at least 93% identity to any one of SEQ ID NOs: 1-80, 138 or 139. In one embodiment, the nucleic acid sequence has at least 94% identity to any one of SEQ ID NOs: 1-80, 138 or 139. In one embodiment, the nucleic acid sequence has at least 95% identity to any one of SEQ ID NOs: 1-80, 138 or 139. In one embodiment, the nucleic acid sequence has at least 96% identity to any one of SEQ ID NOs: 1-80, 138 or 139. In one embodiment, the nucleic acid sequence has at least 97% identity to any one of SEQ ID NOs: 1-80, 138 or 139. In one embodiment, the nucleic acid sequence has at least 98% identity to any one of SEQ ID NOs: 1-80, 138 or 139. In one embodiment, the nucleic acid sequence has at least 99% identity to any one of SEQ ID NOs: 1-80, 138 or 139. In one embodiment, the nucleic acid consists of any one of SEQ ID NOs: 1-80, 138 or 139. In one embodiment, the nucleic acid sequence comprises any one of SEQ ID NOs: 1-80, 138 or 139.

[0014] In a first aspect, the disclosure relates to a liver-specific nucleic acid regulatory element comprising a nucleic acid sequence having at least 95% identity to SEQ ID NO: 131. In one embodiment, the nucleic acid sequence has at least 96% identity to SEQ ID NO: 131. In one embodiment, the nucleic acid sequence has at least 97% identity to SEQ ID NO: 131. In one embodiment, the nucleic acid sequence has at least 98% identity to SEQ ID NO: 131. In one embodiment, the nucleic acid sequence has at least 99% identity to SEQ ID NO: 131. In one embodiment, the nucleic acid sequence comprises SEQ ID NO: 131. In one embodiment, the nucleic acid sequence consists of SEQ ID NO: 131.

[0015] In a first aspect, the disclosure relates to a liver-specific nucleic acid regulatory element comprising a nucleic acid sequence having at least 95% identity to SEQ ID NO: 122. In one embodiment, the nucleic acid sequence has at least 96% identity to SEQ ID NO: 122. In one embodiment, the nucleic acid sequence has at least 97% identity to SEQ ID NO: 122. In one embodiment, the nucleic acid sequence has at least 98% identity to SEQ ID NO: 122. In one embodiment, the nucleic acid sequence has at least 99% identity to SEQ ID NO: 122. In one embodiment, the nucleic acid sequence comprises SEQ ID NO: 122. In one embodiment, the nucleic acid consists of SEQ ID NO: 122.

[0016] In a first aspect, the disclosure relates to a liver-specific nucleic acid regulatory element comprising a nucleic acid sequence having at least 95% identity to SEQ ID NO: 81. In one embodiment, the nucleic acid sequence has at least 96% identity to SEQ ID NO: 81. In one embodiment, the nucleic acid sequence has at least 97% identity to SEQ ID NO: 81. In one embodiment, the nucleic acid sequence has at least 98% identity to SEQ ID NO: 81. In one embodiment, the nucleic acid sequence has at least 99% identity to SEQ ID NO: 81. In one embodiment, the nucleic acid sequence comprises SEQ ID NO: 81. In one embodiment, the nucleic acid consists of SEQ ID NO: 81.

[0017] In a first aspect, the disclosure relates to a liver-specific nucleic acid regulatory element comprising a nucleic acid sequence having at least 95% identity to SEQ ID NO: 82. In one embodiment, the nucleic acid sequence has at least 96% identity to SEQ ID NO: 82. In one embodiment, the nucleic acid sequence has at least 97% identity to SEQ ID NO: 82. In one embodiment, the nucleic acid sequence has at least 98% identity to SEQ ID NO: 82. In one embodiment, the nucleic acid sequence has at least 99% identity to SEQ ID NO: 82. In one embodiment, the nucleic acid sequence comprises SEQ ID NO: 82. In one embodiment, the nucleic acid consists of SEQ ID NO: 82.

[0018] In a first aspect, the disclosure relates to a liver-specific nucleic acid regulatory element comprising a nucleic acid sequence having at least 95% identity to SEQ ID NO: 83. In one embodiment, the nucleic acid sequence has at least 96% identity to SEQ ID NO: 83. In one embodiment, the nucleic acid sequence has at least 97% identity to SEQ ID NO: 83. In one embodiment, the nucleic acid sequence has at least 98% identity to SEQ ID NO: 83. In one embodiment, the nucleic acid sequence has at least 99% identity to SEQ ID NO: 83. In one embodiment, the nucleic acid sequence comprises SEQ ID NO: 83. In one embodiment, the nucleic acid consists of SEQ ID NO: 83.

[0019] In another aspect, disclosed herein is a liver-specific nucleic acid regulatory element consisting essentially of a nucleic acid sequence set forth in any one of Table 10, Table 11, Table 12, or Table 13.

[0020] In another aspect, disclosed herein is a liver-specific nucleic acid regulatory element comprising a nucleic acid sequence set forth in any one of Table 10, Table 11, Table 12, or Table 13.

[0021] In another aspect, disclosed herein is a liver-specific nucleic acid regulatory element comprising a nucleic acid sequence having at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% identity to a sequence set forth in any one of Table 10, Table 11, Table 12, or Table 13.

[0022] In one embodiment, the element comprises at least two nucleic acid sequences set forth in any one of Table 10, Table 11, Table 12, or Table 13. In one embodiment, the two nucleic acid sequences are identical. In one embodiment, the element comprises three (3) nucleic acid sequences set forth in any one of Table 10, Table 11, Table 12, or Table 13, optionally wherein the three sequences are identical. In one embodiment, the element consists essentially of two (2) to ten (10) nucleic acid sequences set forth in any one of Table 10, Table 11, Table 12, or Table 13.

[0023] In one embodiment, the element comprises a spacer placed between the nucleic acid sequences set forth in any one of Table 10, Table 11, Table 12, or Table 13. In one embodiment, the spacer is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 base pairs long.

[0024] In one embodiment, the element comprises a nucleic acid sequence at least 95%, 96%, 97%, 98% or 99% identical to:

GGGGGAGGCTGCTGGTGAATATTAACCAAGGTCACCCAGTTATCGGAGGAGCAAACAG
GGGCAAAGTCCAC (SEQ ID NO: 223),

GGGGGAAGCTACTGGTGAATATTAACCAAGGTCACCCAGTTATCAGGGAGCAAACAGGA
GCAAAGTCCAT (SEQ ID NO: 1381),

GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCCGTTATCGGAGGAGCAAACAAGGG
CTAAGTCCAC (SEQ ID NO: 1073), or

GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAACAAGGG
CAAAGTCCAC (SEQ ID NO: 1113).

[0025] In one embodiment, the element comprises a nucleic acid consisting of:

GGGGGAGGCTGCTGGTGAATATTAACCAAGGTCACCCAGTTATCGGAGGAGCAAACAG
GGGCAAAGTCCAC (SEQ ID NO: 223),

GGGGGAAGCTACTGGTGAATATTAACCAAGGTCACCCAGTTATCAGGGAGCAAACAGGA
GCAAAGTCCAT (SEQ ID NO: 1381),

GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCCGTTATCGGAGGAGCAAACAAGGG
CTAAGTCCAC (SEQ ID NO: 1073), or

GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAACAAGGG
CAAAGTCCAC (SEQ ID NO: 1113).

[0026] In another aspect, the disclosure relates to a liver-specific nucleic acid regulatory element comprising a nucleic acid sequence having at least 85% identity to any one of SEQ ID NOs: 81, 82, 122, 83 or 85. In one embodiment, the nucleic acid sequence has at least 90% identity to any one of SEQ ID NOs: 81, 82, 122, 83 or 85. In one embodiment, the nucleic acid sequence has at least 91% identity to any one of SEQ ID NOs: 81, 82, 122, 83 or 85. In one embodiment, the nucleic acid sequence has at least 92% identity to any one of SEQ ID NOs: 81, 82, 122, 83 or 85. In one embodiment, the nucleic acid sequence has at least 93% identity to any one of SEQ ID NOs: 81, 82, 122, 83 or 85. In one embodiment, the nucleic acid sequence has at least 94% identity to any one of SEQ ID NOs: 81, 82, 122, 83 or 85. In one embodiment, the nucleic acid sequence has at least 95% identity to any one of SEQ ID NOs: 81, 82, 122, 83 or 85. In one embodiment, the nucleic acid sequence has at least 96% identity to any one of SEQ ID NOs: 81, 82, 122, 83 or 85. In one embodiment, the nucleic acid sequence has at least 97% identity to any one of SEQ ID NOs: 81, 82, 122, 83 or 85. In one embodiment, the nucleic acid sequence has at least 98% identity to any one of SEQ ID NOs: 81, 82, 122, 83 or 85. In one embodiment, the nucleic acid sequence has at least 99% identity to any one of SEQ ID NOs: 81, 82, 122, 83 or 85. In one embodiment, the nucleic acid sequence comprises any one of SEQ ID NOs: 81, 82, 122, 83 or 85. In one embodiment, the nucleic acid sequence consists of any one of SEQ ID NOs: 81, 82, 122, 83 or 85.

[0027] In another aspect, the disclosure provides a liver-specific expression cassette comprising at least one liver-specific regulatory element of any one of the aspects and embodiments herein. In one embodiment, the liver-specific expression cassette further comprises a liver-specific promoter operably linked to a transgene. In one embodiment, two or more nucleotides separate each liver-specific nucleic acid regulatory element. In one embodiment, 5 or more nucleotides separate each liver-specific nucleic acid regulatory element. In one embodiment, 10 or more nucleotides separate each liver-specific nucleic acid regulatory element. In one embodiment, 15 or more nucleotides separate each liver-specific nucleic acid regulatory element. In one embodiment, 20 or more nucleotides separate each liver-specific nucleic acid regulatory element. In one embodiment, 25 or more nucleotides separate each liver-specific nucleic acid regulatory element. In one embodiment, between 2 and 30 nucleotides separate each liver-specific regulatory element, *e.g.*, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 15, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 or 30 nucleotides.

[0028] In another aspect, the disclosure provides a liver-specific expression cassette comprising at least three repeats of a liver-specific nucleic acid regulatory element and a liver-specific promoter operably linked to a transgene, wherein the liver-specific nucleic acid regulatory element comprises a nucleic acid sequence having at least 95% identity to any one of SEQ ID NOs: 81-137, and wherein two or more nucleotides separate each liver-specific nucleic acid regulatory element.

[0029] In one embodiment, between 2 and 30 nucleotides separate each regulatory element. In one embodiment, between 2 and 10, between 5 and 15, between 10 and 15, between 10 and 20, between 15 and 25, between 20 and 30 or between 25 and 30 nucleotides separate each regulatory element. In one

embodiment, 5 nucleotides separate each regulatory element. In one embodiment, 11 nucleotides separate each regulatory element. In one embodiment, 30 nucleotides separate each regulatory element. In one embodiment, the liver-specific expression cassette comprises two, three, four, or five repeats of the liver-specific nucleic acid regulatory element. In one embodiment, the liver-specific expression cassette comprises six, seven, eight, nine or ten repeats of the liver-specific nucleic acid regulatory element. In one embodiment, liver-specific expression cassette comprises one or more FOXA and HNF4 transcription factor consensus sites. In one embodiment, the liver-specific nucleic acid regulatory element comprises one or more sites of CpG minimization. In one embodiment, the liver-specific promoter is selected from the group consisting of: a transthyretin (TTR) promoter, minimal TTR promoter (TTRm), an AAT promoter, an albumin (ALB) promoter or minimal promoter, an apolipoprotein A1 (APOA1) promoter or minimal promoter, a complement factor B (CFB) promoter, a ketohexokinase (KHK) promoter, a hemopexin (HPX) promoter or minimal promoter, a nicotinamide N-methyltransferase (NNMT) promoter or minimal promoter, a carboxylesterase 1 (CES1) promoter or minimal promoter, a protein C (PROC) promoter or minimal promoter, an apolipoprotein C3 (APOC3) promoter or minimal promoter, a mannan-binding lectin serine protease 2 (MASP2) promoter or minimal promoter, a hepcidin antimicrobial peptide (HAMP) promoter or minimal promoter, and a serpin peptidase inhibitor, clade C (antithrombin), member 1 (SERPINC1) promoter or minimal promoter. In one embodiment, the promoter comprises any sequence from **Table 1**. In one embodiment, the liver-specific promoter is a TTR promoter or a TTRm promoter. In one embodiment, the transgene encodes a liver-specific therapeutic protein. In one embodiment, the liver-specific therapeutic protein is coagulation factor VIII (FVIII). In one embodiment, the coagulation FVIII comprises a codon optimized nucleic acid sequence. In one embodiment, the coagulation FVIII comprises a nucleic acid sequence having at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% identity to, comprises, or consists of SEQ ID NO: 143.

[0030] In another aspect, the disclosure provides a vector comprising the liver-specific nucleic acid regulatory element of any one of the aspects or embodiments herein or the liver-specific expression cassette according to any one of the aspects or embodiments herein. In one embodiment, the vector is a viral vector or a non-viral vector. In one embodiment, the vector is a plasmid. In one embodiment, the vector is a closed-ended DNA (ceDNA) vector.

[0031] In another aspect, the disclosure provides a pharmaceutical composition comprising the liver-specific expression cassette according to any one of the aspects or embodiments herein or the vector according to any one of the aspects or embodiments herein, and a pharmaceutically acceptable excipient.

[0032] In another aspect, the disclosure provides a method of treating a liver-specific disease or disorder comprising transduction or transfection of the vector according to any one of the aspects and embodiments herein, or the pharmaceutical composition of the aspects or embodiments herein, into a subject. In one embodiment, the subject is a human subject suffering from a genetic disorder. In one

embodiment, the subject has hemophilia A. In one embodiment, the genetic disorder is selected from the group consisting of sickle-cell anemia, melanoma, hemophilia A (clotting factor VIII (FVIII) deficiency) and hemophilia B (clotting factor IX (FIX) deficiency), cystic fibrosis (CFTR), familial hypercholesterolemia (LDL receptor defect), hepatoblastoma, Wilson disease, phenylketonuria (PKU), congenital hepatic porphyria, inherited disorders of hepatic metabolism, Lesch Nyhan syndrome, sickle cell anemia, thalassaemias, xeroderma pigmentosum, Fanconi's anemia, retinitis pigmentosa, ataxia telangiectasia, Bloom's syndrome, retinoblastoma, mucopolysaccharide storage diseases (*e.g.*, Hurler syndrome (MPS Type I), Scheie syndrome (MPS Type I S), Hurler-Scheie syndrome (MPS Type I H-S), Hunter syndrome (MPS Type II), Sanfilippo Types A, B, C, and D (MPS Types III A, B, C, and D), Morquio Types A and B (MPS IVA and MPS IVB), Maroteaux-Lamy syndrome (MPS Type VI), Sly syndrome (MPS Type VII), hyaluronidase deficiency (MPS Type IX)), Niemann-Pick Disease Types A/B, C1 and C2, Fabry disease, Schindler disease, GM2-gangliosidosis Type II (Sandhoff Disease), Tay-Sachs disease, Metachromatic Leukodystrophy, Krabbe disease, Mucopolipidosis Type I, II/III and IV, Sialidosis Types I and II, Glycogen Storage disease Types I and II (Pompe disease), Gaucher disease Types I, II and III, Fabry disease, cystinosis, Batten disease, Aspartylglucosaminuria, Salla disease, Danon disease (LAMP-2 deficiency), Lysosomal Acid Lipase (LAL) deficiency, neuronal ceroid lipofuscinoses (CLN1-8, INCL, and LINCL), sphingolipidoses, galactosialidosis, amyotrophic lateral sclerosis (ALS), Parkinson's disease, Alzheimer's disease, Huntington's disease, spinocerebellar ataxia, spinal muscular atrophy, Friedreich's ataxia, Duchenne muscular dystrophy (DMD), Becker muscular dystrophies (BMD), dystrophic epidermolysis bullosa (DEB), ectonucleotide pyrophosphatase 1 deficiency, generalized arterial calcification of infancy (GACI), Leber Congenital Amaurosis, Stargardt macular dystrophy (ABCA4), ornithine transcarbamylase (OTC) deficiency, Usher syndrome, alpha-1 antitrypsin deficiency, progressive familial intrahepatic cholestasis (PFIC) type I (ATP8B1 deficiency), type II (ABCB11), type III (ABCB4), or type IV (TJP2) and Cathepsin A deficiency.

[0033] In another aspect, disclosed herein is a method of increasing expression capacity of a liver-specific enhancer element comprising the nucleic acid sequence CTAAG, comprising introducing a single nucleotide substitution (T to A) mutation such that the substitution results in the nucleic acid sequence comprising CAAAG.

[0034] In another aspect, disclosed herein is a liver-specific enhancer element comprising a nucleic acid sequence selected from: CAAAG; CAAAGT; CAAAGTC; GCAAAGT; GCAAAG; or GCAAAGTC.

[0035] These and other aspects of the disclosure are described in further detail below.

DESCRIPTION OF DRAWINGS

[0036] **FIG. 1A** and **FIG. 1B** depict sequences and alignment of conserved enhancer regions of human and 20 other vertebrates. 115 non-human vertebrate genomes were assessed for conserved

SERPINA1 enhancer regions using the UCSC multiz100way and multiz30way multiple alignments. Highlighted nucleotides in the aligned sequences represent differences from the human reference sequence.

[0037] FIG. 2 depicts identification of near-consensus binding sites for various transcription factors (TF) in human SERPINA1 enhancer (hSerpEnh) region, including HNF4 and FOXA, which are key regulators of hepatic gene expression. The arrows in the bottom three rows represent TF binding motifs described by Chuah *et al.* (2014). The arrows at the top 15 rows represent transcription factor (TF) binding motifs identified by independent analyses described herein. Positions where the human SERPINA1 sequence differs from the most highly preferred nucleotide in the sequence logos are boxed.

[0038] FIG. 3 depicts multiple bioinformatic analyses employed to inform potential removal of CpG (*i.e.*, CpG ablation). The human SERPINA1 enhancer contains one internal CpG and the potential to form CpGs at its 5' and 3' ends (highlighted in red and boxed in the "hSerpEnh" track). Low sequence conservation, the presence of human SNPs that are not known to be associated with disease, and the absence of predicted TF binding sites were assessed to inform sequence changes to ablate the central CpG and the remove potential for CpG formation at the ends of the sequence.

[0039] FIG. 4A depicts results of the top 11 constructs (plasmid) in a screen of 30 single (1x) variants using luciferase reporter assay (n=3) *in vitro*. Results are grouped by rationally designed enhancer variants (1x TFBS Consensus Variants) or conserved SERPINA1 enhancer regions identified in other species (1x Conserved Genomic Variants). Error bars represent standard deviation.

[0040] FIG. 4B depicts the sequence design of the top variant in this screen, hSerpEnh_FOXA_HNF4_consensus_v1. hSerpEnh_FOXA_HNF4_consensus_v1 was designed by modifying the FOXA and HNF4 motifs identified in the human SERPINA1 enhancer to match their respective consensus sequences (GTGAATA to GTAAACA for FOXA and CTAAGT to CAAACT for HNF4). The internal CpG was ablated by changing the G, which both has lower sequence conservation than the C and is at the position of a human SNP, to an A to match the SNP.

[0041] FIG. 5 depicts results of a screen of 10 multimerized variants in plasmid using an *in vitro* (HepG2 cells) luciferase reporter assay (n=3). Results are grouped by 3x repeats of rationally designed enhancer variants (3x TFBS Variants), 3x repeats of conserved SERPINA1 enhancer regions identified in other species (3x Conserved Variant), 3x repeats of the human SERPINA1 enhancer separated by spacers of varying lengths and sequences (3x hSerpEnh Spacer Variants), and enhancers with varying numbers of repeats (# Repeat Variants). The wild-type human SERPINA1 enhancer are labelled (wt). The comparison between the 3x human SERPINA1 enhancer variant and the 3x top performing variant is boxed. Two sets of technical triplicates were performed for the 1x and 3x human enhancers and the top performing 3x variant (r1, r2). Error bars represent standard deviation.

[0042] FIG. 6A is a schematic for optimization of spacer sequence to improve performance of hSerpEnh variant repeats. The length and sequence of spacers between SERPINA1 enhancer variant repeats were modified to screen for sequences that improved enhancer function. Spacers of length 2, 3,

5, 11, and 30 were designed to prevent introduction of CpGs or ATGs that may create cryptic translation start sites. 11 nt and 30 nt spacers that contain consensus FOXA and HNF4 binding sites were also designed and tested.

[0043] **FIG. 6B** depicts three main configurations of enhancer elements for screening of improved enhancer variants. The enhancer variants were tested in two main configurations: (1) as a single copy of the enhancer variant upstream of the transthyretin (TTR) promoter, TTR 5' UTR, and the minute virus of mice (MVM) intron or (2) as three copies of the enhancer variant upstream of the TTR enhancer, TTR promoter, TTR 5' UTR, and the MVM intron.

[0044] **FIGs. 7A-7D** depict expression levels of FVIII constructs having multimeric repeats of Serpin enhancer variants compared to multimeric human Serpin enhancer (3x, 5x and 10x) variants. **FIG. 7A** depicts expression levels of 3x HNF_FOXA_v1 variants having CpG minimization, GC rich regions (I-motif secondary structures) minimization, or Aspacers (no spacer) performed equivalent to the level seen in 3x hSerpEnh. However, HNF4 FOXA v1 variants repeated 10 times (10x) did not exhibit a meaningful level of FVIII (see, *e.g.*, **FIG. 7C**), suggesting that the Serpin enhancer exhibits better performance when it is repeated in a certain number, *e.g.*, 3x to 5x, with 3x preference, but not when it is repeated in an excessive number (*e.g.*, 10x). A consistent observation was made with other Serpin Enhancer elements including, for example, that of bushbaby Serpin enhancer, Chinese tree shrew Serpin enhancer, and human Serpin enhancer (hSerpEnh). In particular, 5x and 10x bushbaby Serpin enhancer element did not exhibit detectable expression levels of FVIII when a plasmid containing the element operably linked to FVIII was injected hydrodynamically into a mouse (**FIG. 7D**).

[0045] **FIGs. 8A-FIG. 8E** depict FVIII expression as measured by FVIII activity obtained from the serum of mice hydrodynamically injected with a plasmid containing various spacer variants (two-nucleotide long spacers (2-mer; **FIG. 8A**), three-nucleotide long spacers (3-mer; **FIG. 8B**), five-nucleotide long spacers (5-mer; **FIG. 8C**), eleven-nucleotide long spacers (11-mer; **FIG. 8D**), and thirty nucleotide long spacers (30-mer; **FIG. 8E**).

[0046] **FIG. 9** depicts a chart showing the result of FVIII expression using various spacer variants of hSerpEnh (2mers and 11mers as spacers) and other Serpin enhancer variants (3x bushbaby Serpin enhancer to 3x Chinese tree shrew Serpin Enhancer). One dose of 50ng plasmid was hydrodynamically injected to Rag2 mice on day 0 with a single terminal collection at day 3 (~72hr post dose).

[0047] **FIG. 10** depicts a chart showing the result of FVIII expression using various spacer variants of hSerpEnh (2mers and 11mers) and other Serpin enhancer variants (3x bushbaby Serpin enhancer to 3x Chinese tree shrew Serpin Enhancer). One dose of ceDNA was hydrodynamically injected to Rag2 mice on day 0 with a single terminal collection at day 3 (~72hr post dose).

[0048] **FIG. 11** depicts an exemplary annotated nucleotide sequence of a plasmid containing a FVIII ceDNA construct comprising 3xBushbaby_Aspacers Serpin enhancer element linked to TTR_{re}, TTR liver-specific promoter, MVM intron, codon optimized B-domain deleted FVIII (hFVIII-F309S-BD226seq124-BDD-F309), WPRE 3'UTR, and bGH (SEQ ID NO: 146).

[0049] **FIG. 12** depicts an annotated nucleotide sequence of a plasmid containing a FVIII ceDNA construct comprising 3x human Serpin enhancer element linked to TTR_e-PromoterSet, Consensus_Kozak, codon optimized hFVIII (hFVIII-F309S-BD226seq124-BDD-F309), PacI_{site}, WPRE_3pUTR, and bGH (SEQ ID NO: 147).

[0050] **FIG. 13** depicts FVIII expression levels in mice dosed hydrodynamically via tail vein injection with ceDNA constructs having various FVIII and Serpin Enhancer combinations, at Day 0 at a low dose of 0.5 mg/kg or a high dose of 2.0 mg/kg ($n = 5$). Factor VIII expression was measured at Days 7, 14, 21, and 28. Expression of FVIII driven from 3x human SerpEnh having 2bp spacer and 11bp spacer were compared with 3x human SerpEnh without a spacer

[0051] **FIG. 14** depicts FVIII expression levels in mice dosed hydrodynamically via tail vein injection with ceDNA constructs having various FVIII and Serpin Enhancer combinations at Day 0 at a dose of 50 ng ($n = 5$). Factor VIII expression was measured at Days 1 and 3.

[0052] **FIG. 15** depicts FVIII expression levels in mice dosed hydrodynamically via tail vein injection at Day 0 at a dose of 10 ng ($n = 5$). Factor VIII expression was measured at Day 3.

[0053] **FIG. 16A** and **FIG. 16B** depict FVIII expression levels in mice treated via hydrodynamic tail vein injection with ceDNA constructs having various FVIII and Serpin Enhancer combination (3x Tibetan antelope SERPINA1; 3x Armadillo CpG minimized SERPINA1; 3x Chinese Tree Shrew and 3x Chinese Tree Shrew CpG minimized; and 3x Bushbaby Aspac_{er}) at Day 0 at three different dose levels: 25 ng/an, 50 ng/an, 100 ng/an ($n = 4$). Factor VIII expression was measured at Day 3.

[0054] **FIG. 17** depicts an annotated map of pHTS002L, a plasmid employed in making a library of enhancer-luciferase constructs.

[0055] **FIGs. 18A-18D** depict comparisons for two biological replicates of barcode counts for each RNA sample normalized to the corresponding barcode counts for an input DNA sample which were mapped back to their associated enhancer sequences (custom MATLAB script).

[0056] **FIG. 19** depicts alignment of multiple SERPINA1 enhancer sequences.

DETAILED DESCRIPTION

[0057] Provided herein are liver-specific promoters, wherein the native *cis*-regulatory region has been optimized to minimize CpG content and to enhance spacer optimization. The liver-specific promoters of the present disclosure represent an improvement over those previously known by providing enhanced efficiency and safety for liver-specific gene therapy.

Definitions

[0058] Unless otherwise defined herein, scientific and technical terms used in connection with the present application shall have the meanings that are commonly understood by those of ordinary skill in the art to which this disclosure belongs. It should be understood that this disclosure is not limited to the particular methodology, protocols, and reagents, *etc.*, described herein and as such can vary. The

terminology used herein is for the purpose of describing particular embodiments only and is not intended to limit the scope of the present disclosure, which is defined solely by the claims. Definitions of common terms in immunology and molecular biology can be found in The Merck Manual of Diagnosis and Therapy, 19th Edition, published by Merck Sharp & Dohme Corp., 2011 (ISBN 978-0-911910-19-3); Robert S. Porter *et al.* (eds.), Fields Virology, 6th Edition, published by Lippincott Williams & Wilkins, Philadelphia, PA, USA (2013), Knipe, D.M. and Howley, P.M. (ed.), The Encyclopedia of Molecular Cell Biology and Molecular Medicine, published by Blackwell Science Ltd., 1999-2012 (ISBN 9783527600908); and Robert A. Meyers (ed.), Molecular Biology and Biotechnology: a Comprehensive Desk Reference, published by VCH Publishers, Inc., 1995 (ISBN 1-56081-569-8); Immunology by Werner Luttmann, published by Elsevier, 2006; Janeway's Immunobiology, Kenneth Murphy, Allan Mowat, Casey Weaver (eds.), Taylor & Francis Limited, 2014 (ISBN 0815345305, 9780815345305); Lewin's Genes XI, published by Jones & Bartlett Publishers, 2014 (ISBN-1449659055); Michael Richard Green and Joseph Sambrook, Molecular Cloning: A Laboratory Manual, 4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., USA (2012) (ISBN 1936113414); Davis *et al.*, Basic Methods in Molecular Biology, Elsevier Science Publishing, Inc., New York, USA (2012) (ISBN 044460149X); Laboratory Methods in Enzymology: DNA, Jon Lorsch (ed.) Elsevier, 2013 (ISBN 0124199542); Current Protocols in Molecular Biology (CPMB), Frederick M. Ausubel (ed.), John Wiley and Sons, 2014 (ISBN 047150338X, 9780471503385), Current Protocols in Protein Science (CPPS), John E. Coligan (ed.), John Wiley and Sons, Inc., 2005; and Current Protocols in Immunology (CPI) (John E. Coligan, ADA M Kruisbeek, David H Margulies, Ethan M Shevach, Warren Strobe, (eds.) John Wiley and Sons, Inc., 2003 (ISBN 0471142735, 9780471142737), the contents of which are all incorporated by reference herein in their entireties.

[0059] As used herein, the terms, “administration,” “administering” and variants thereof refers to introducing a composition or agent (*e.g.*, a therapeutic nucleic acid or an immunosuppressant as described herein) into a subject and includes concurrent and sequential introduction of one or more compositions or agents. “Administration” can refer, *e.g.*, to therapeutic, pharmacokinetic, diagnostic, research, placebo, and experimental methods. “Administration” also encompasses *in vitro* and *ex vivo* treatments. The introduction of a composition or agent into a subject is by any suitable route, including orally, pulmonarily, intranasally, parenterally (intravenously, intramuscularly, intraperitoneally, or subcutaneously), rectally, intralymphatically, intratumorally, or topically. The introduction of a composition or agent into a subject is by electroporation. Administration includes self-administration and the administration by another. Administration can be carried out by any suitable route. A suitable route of administration allows the composition or the agent to perform its intended function. For example, if a suitable route is intravenous, the composition is administered by introducing the composition or agent into a vein of the subject.

[0060] As used herein, the phrases “nucleic acid therapeutic”, “therapeutic nucleic acid” and “TNA” are used interchangeably and refer to any modality of therapeutic using nucleic acids as an active component of therapeutic agent to treat a disease or disorder. As used herein, these phrases refer to RNA-based therapeutics and DNA-based therapeutics. Non-limiting examples of RNA-based therapeutics include mRNA, antisense RNA and oligonucleotides, ribozymes, aptamers, interfering RNAs (RNAi), Dicer-substrate dsRNA, small hairpin RNA (shRNA), guide RNA (gRNA), asymmetrical interfering RNA (aiRNA), microRNA (miRNA). Non-limiting examples of DNA-based therapeutics include minicircle DNA, minigene, viral DNA (*e.g.*, Lentiviral or AAV genome) or non-viral synthetic DNA vectors, closed-ended linear duplex DNA (ceDNA / CELiD), plasmids, bacmids, doggybone (dbDNATM) DNA vectors, minimalistic immunological-defined gene expression (MIDGE)-vector, nonviral ministring DNA vector (linear-covalently closed DNA vector), or dumbbell-shaped DNA minimal vector (“dumbbell DNA”).

[0061] As used herein, an “effective amount” or “therapeutically effective amount” of a therapeutic agent, such as a FVIII therapeutic protein or fragment thereof, is an amount sufficient to produce the desired effect, *e.g.*, treatment or prevention of hemophilia A. Suitable assays for measuring expression of a target gene or target sequence include, *e.g.*, examination of protein or RNA levels using techniques known to those of skill in the art such as dot blots, northern blots, *in situ* hybridization, ELISA, immunoprecipitation, enzyme function, as well as phenotypic assays known to those of skill in the art. However, dosage levels are based on a variety of factors, including the type of injury, the age, weight, sex, medical condition of the patient, the severity of the condition, the route of administration, and the particular active agent employed. Thus, the dosage regimen may vary widely, but can be determined routinely by a physician using standard methods. Additionally, the terms “therapeutic amount”, “therapeutically effective amounts” and “pharmaceutically effective amounts” include prophylactic or preventative amounts of the compositions of the described disclosure. In prophylactic or preventative applications of the described disclosure, pharmaceutical compositions or medicaments are administered to a patient susceptible to, or otherwise at risk of, a disease, disorder or condition in an amount sufficient to eliminate or reduce the risk, lessen the severity, or delay the onset of the disease, disorder or condition, including biochemical, histologic and/or behavioral symptoms of the disease, disorder or condition, its complications, and intermediate pathological phenotypes presenting during development of the disease, disorder or condition. It is generally preferred that a maximum dose be used, that is, the highest safe dose according to some medical judgment. In one embodiment, the disease, disorder or condition is hemophilia A. The terms “dose” and “dosage” are used interchangeably herein.

[0062] As used herein the term “therapeutic effect” refers to a consequence of treatment, the results of which are judged to be desirable and beneficial. A therapeutic effect can include, directly or indirectly, the arrest, reduction, or elimination of a disease manifestation. A therapeutic effect can also

include, directly or indirectly, the arrest reduction or elimination of the progression of a disease manifestation.

[0063] For any therapeutic agent described herein therapeutically effective amount may be initially determined from preliminary *in vitro* studies and/or animal models. A therapeutically effective dose may also be determined from human data. The applied dose may be adjusted based on the relative bioavailability and potency of the administered compound. Adjusting the dose to achieve maximal efficacy based on the methods described above and other well-known methods is within the capabilities of the ordinarily skilled artisan. General principles for determining therapeutic effectiveness, which may be found in Chapter 1 of Goodman and Gilman's The Pharmacological Basis of Therapeutics, 10th Edition, McGraw-Hill (New York) (2001), incorporated herein by reference, are summarized below.

[0064] Pharmacokinetic principles provide a basis for modifying a dosage regimen to obtain a desired degree of therapeutic efficacy with a minimum of unacceptable adverse effects. In situations where the drug's plasma concentration can be measured and related to therapeutic window, additional guidance for dosage modification can be obtained.

[0065] As used herein, the terms "heterologous nucleic acid sequence" and "transgene" are used interchangeably and refer to a nucleic acid of interest (other than a nucleic acid encoding a capsid polypeptide) that is incorporated into and may be delivered and expressed by a ceDNA vector as disclosed herein. In one embodiment, a nucleic acid sequence may be a heterologous nucleic acid sequence. In one embodiment, the term "heterologous nucleic acid" is meant to refer to a nucleic acid (or transgene) that is not present in, expressed by, or derived from the cell or subject to which it is contacted.

[0066] As used herein, the terms "expression cassette" and "transcription cassette" are used interchangeably and refer to a linear stretch of nucleic acids that includes a transgene that is operably linked to one or more promoters or other regulatory sequences sufficient to direct transcription of the transgene, but which does not comprise capsid-encoding sequences, other vector sequences or inverted terminal repeat regions. An expression cassette may additionally comprise one or more *cis*-acting sequences (*e.g.*, promoters, enhancers, or repressors), one or more introns, and one or more post-transcriptional regulatory elements.

[0067] The terms "polynucleotide" and "nucleic acid," used interchangeably herein, refer to a polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides. Thus, this term includes single, double, or multi-stranded DNA or RNA, genomic DNA, cDNA, DNA-RNA hybrids, or a polymer including purine and pyrimidine bases or other natural, chemically or biochemically modified, non-natural, or derivatized nucleotide bases. "Oligonucleotide" generally refers to polynucleotides of between about 5 and about 100 nucleotides of single- or double-stranded DNA. However, for the purposes of this disclosure, there is no upper limit to the length of an oligonucleotide. Oligonucleotides are also known as "oligomers" or "oligos" and may be isolated from genes, or chemically synthesized by methods known in the art. The terms "polynucleotide" and "nucleic

acid” should be understood to include, as applicable to the embodiments being described, single-stranded (such as sense or antisense) and double-stranded polynucleotides. DNA may be in the form of, *e.g.*, antisense molecules, plasmid DNA, DNA-DNA duplexes, pre-condensed DNA, PCR products, vectors (P1, PAC, BAC, YAC, artificial chromosomes), expression cassettes, chimeric sequences, chromosomal DNA, or derivatives and combinations of these groups. DNA may be in the form of minicircle, plasmid, bacmid, minigene, ministring DNA (linear covalently closed DNA vector), closed-ended linear duplex DNA (CELiD or ceDNA), doggybone (dbDNATM) DNA, dumbbell shaped DNA, minimalistic immunological-defined gene expression (MIDGE)-vector, viral vector or nonviral vectors. RNA may be in the form of small interfering RNA (siRNA), Dicer-substrate dsRNA, small hairpin RNA (shRNA), guide RNA (gRNA), asymmetrical interfering RNA (aiRNA), microRNA (miRNA), mRNA, rRNA, tRNA, viral RNA (vRNA), and combinations thereof. Nucleic acids include nucleic acids containing known nucleotide analogs or modified backbone residues or linkages, which are synthetic, naturally occurring, and non-naturally occurring, and which have similar binding properties as the reference nucleic acid. Examples of such analogs and/or modified residues include, without limitation, phosphorothioates, phosphorodiamidate morpholino oligomer (morpholino), phosphoramidates, methyl phosphonates, chiral-methyl phosphonates, 2'-O-methyl ribonucleotides, locked nucleic acid (LNATM), and peptide nucleic acids (PNAs). Unless specifically limited, the term encompasses nucleic acids containing known analogues of natural nucleotides that have similar binding properties as the reference nucleic acid. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (*e.g.*, degenerate codon substitutions), alleles, orthologs, SNPs, and complementary sequences as well as the sequence explicitly indicated.

[0068] “Nucleotides” contain a sugar deoxyribose (DNA) or ribose (RNA), a base, and a phosphate group. Nucleotides are linked together through the phosphate groups.

[0069] “Bases” include purines and pyrimidines, which further include natural compounds adenine, thymine, guanine, cytosine, uracil, inosine, and natural analogs, and synthetic derivatives of purines and pyrimidines, which include, but are not limited to, modifications which place new reactive groups such as, but not limited to, amines, alcohols, thiols, carboxylates, and alkylhalides.

[0070] The term “nucleic acid construct” as used herein refers to a nucleic acid molecule, either single- or double-stranded, which is isolated from a naturally occurring gene or which is modified to contain segments of nucleic acids in a manner that would not otherwise exist in nature or which is synthetic. The term nucleic acid construct is synonymous with the term “expression cassette” when the nucleic acid construct contains the control sequences required for expression of a coding sequence of the present disclosure. An “expression cassette” includes a DNA coding sequence operably linked to a promoter.

[0071] By “hybridizable” or “complementary” or “substantially complementary” it is meant that a nucleic acid (*e.g.*, RNA) includes a sequence of nucleotides that enables it to non-covalently bind, *i.e.*

form Watson-Crick base pairs and/or G/U base pairs, “anneal”, or “hybridize,” to another nucleic acid in a sequence-specific, antiparallel, manner (*i.e.*, a nucleic acid specifically binds to a complementary nucleic acid) under the appropriate *in vitro* and/or *in vivo* conditions of temperature and solution ionic strength. As is known in the art, standard Watson-Crick base-pairing includes: adenine (A) pairing with thymidine (T), adenine (A) pairing with uracil (U), and guanine (G) pairing with cytosine (C). In addition, it is also known in the art that for hybridization between two RNA molecules (*e.g.*, dsRNA), guanine (G) base pairs with uracil (U). For example, G/U base-pairing is partially responsible for the degeneracy (*i.e.*, redundancy) of the genetic code in the context of tRNA anti-codon base-pairing with codons in mRNA. In the context of this disclosure, a guanine (G) of a protein-binding segment (dsRNA duplex) of a subject DNA-targeting RNA molecule is considered complementary to an uracil (U), and *vice versa*. As such, when a G/U base-pair can be made at a given nucleotide position a protein-binding segment (dsRNA duplex) of a subject DNA-targeting RNA molecule, the position is not considered to be non-complementary, but is instead considered to be complementary.

[0072] The terms “peptide,” “polypeptide,” and “protein” are used interchangeably herein, and refer to a polymeric form of amino acids of any length, which can include coded and non-coded amino acids, chemically or biochemically modified or derivatized amino acids, and polypeptides having modified peptide backbones.

[0073] A DNA sequence that “encodes” a particular FVIII protein is a DNA nucleic acid sequence that is transcribed into the particular RNA and/or protein. A DNA polynucleotide may encode an RNA (mRNA) that is translated into protein, or a DNA polynucleotide may encode an RNA that is not translated into protein (*e.g.*, tRNA, rRNA, or a DNA-targeting RNA; also called “non-coding” RNA or ncRNA”).

[0074] As used herein, the term “fusion protein” as used herein refers to a polypeptide which comprises protein domains from at least two different proteins. For example, a fusion protein may comprise (i) a therapeutic protein, or a fragment thereof (*e.g.*, FVIII or a fragment thereof) and (ii) at least one non-GOI protein. Fusion proteins encompassed herein include, but are not limited to, an antibody, or Fc or antigen-binding fragment of an antibody fused to a therapeutic protein (*e.g.*, a FVIII protein), *e.g.*, an extracellular domain of a receptor, ligand, enzyme or peptide. The protein or fragment thereof that is part of a fusion protein can be a monospecific antibody or a bispecific or multispecific antibody.

[0075] As used herein, the term “genomic safe harbor gene” or “safe harbor gene” refers to a gene or loci that a nucleic acid sequence can be inserted such that the sequence can integrate and function in a predictable manner (*e.g.*, express a protein of interest) without significant negative consequences to endogenous gene activity, or the promotion of cancer. In some embodiments, a safe harbor gene is also a loci or gene where an inserted nucleic acid sequence can be expressed efficiently and at higher levels than a non-safe harbor site.

[0076] As used herein, the term “gene delivery” means a process by which foreign DNA is transferred to host cells for applications of gene therapy.

[0077] As used herein, the term “terminal repeat” or “TR” includes any viral terminal repeat or synthetic sequence that comprises at least one minimal required origin of replication and a region comprising a palindrome hairpin structure. A Rep-binding sequence (“RBS”) (also referred to as RBE (Rep-binding element)) and a terminal resolution site (“TRS”) together constitute a “minimal required origin of replication” and thus the TR comprises at least one RBS and at least one TRS. TRs that are the inverse complement of one another within a given stretch of polynucleotide sequence are typically each referred to as an “inverted terminal repeat” or “ITR”. In the context of a virus, ITRs mediate replication, virus packaging, integration and provirus rescue. As was unexpectedly found in the disclosure herein, TRs that are not inverse complements across their full length can still perform the traditional functions of ITRs, and thus the term ITR is used herein to refer to a TR in a ceDNA genome or ceDNA vector that is capable of mediating replication of ceDNA vector. It will be understood by one of ordinary skill in the art that in complex ceDNA vector configurations more than two ITRs or asymmetric ITR pairs may be present. The ITR can be an AAV ITR or a non-AAV ITR, or can be derived from an AAV ITR or a non-AAV ITR. For example, the ITR can be derived from the family *Parvoviridae*, which encompasses parvoviruses and dependoviruses (*e.g.*, canine parvovirus, bovine parvovirus, mouse parvovirus, porcine parvovirus, human parvovirus B-19), or the SV40 hairpin that serves as the origin of SV40 replication can be used as an ITR, which can further be modified by truncation, substitution, deletion, insertion and/or addition. Parvoviridae family viruses consist of two subfamilies: *Parvovirinae*, which infect vertebrates, and *Densovirinae*, which infect invertebrates. Dependoparvoviruses include the viral family of the adeno-associated viruses (AAV) which are capable of replication in vertebrate hosts including, but not limited to, human, primate, bovine, canine, equine and ovine species. For convenience herein, an ITR located 5' to (upstream of) an expression cassette in a ceDNA vector is referred to as a “5' ITR” or a “left ITR”, and an ITR located 3' to (downstream of) an expression cassette in a ceDNA vector is referred to as a “3' ITR” or a “right ITR”.

[0078] A “wild-type ITR” or “WT-ITR” refers to the sequence of a naturally occurring ITR sequence in an AAV or other dependovirus that retains, *e.g.*, Rep binding activity and Rep nicking ability. The nucleic acid sequence of a WT-ITR from any AAV serotype may slightly vary from the canonical naturally occurring sequence due to degeneracy of the genetic code or drift, and therefore WT-ITR sequences encompassed for use herein include WT-ITR sequences as result of naturally occurring changes taking place during the production process (*e.g.*, a replication error).

[0079] As used herein, the term “substantially symmetrical WT-ITRs” or a “substantially symmetrical WT-ITR pair” refers to a pair of WT-ITRs within a single ceDNA genome or ceDNA vector that are both wild type ITRs that have an inverse complement sequence across their entire length. For example, an ITR can be considered to be a wild-type sequence, even if it has one or more nucleotides that deviate from the canonical naturally occurring sequence, so long as the changes do not affect the

properties and overall three-dimensional structure of the sequence. In some aspects, the deviating nucleotides represent conservative sequence changes. As one non-limiting example, a sequence that has at least 95%, 96%, 97%, 98%, or 99% sequence identity to the canonical sequence (as measured, *e.g.*, using BLAST at default settings), and also has a symmetrical three-dimensional spatial organization to the other WT-ITR such that their 3D structures are the same shape in geometrical space. The substantially symmetrical WT-ITR has the same A, C-C' and B-B' loops in 3D space. A substantially symmetrical WT-ITR can be functionally confirmed as WT by determining that it has an operable Rep binding site (RBE or RBE') and terminal resolution site (TRS) that pairs with the appropriate Rep protein. One can optionally test other functions, including transgene expression under permissive conditions.

[0080] As used herein, the phrases of “modified ITR” or “mod-ITR” or “mutant ITR” are used interchangeably herein and refer to an ITR that has a mutation in at least one or more nucleotides as compared to the WT-ITR from the same serotype. The mutation can result in a change in one or more of A, C, C', B, B' regions in the ITR, and can result in a change in the three-dimensional spatial organization (*i.e.* its 3D structure in geometric space) as compared to the 3D spatial organization of a WT-ITR of the same serotype.

[0081] As used herein, the term “asymmetric ITRs” also referred to as “asymmetric ITR pairs” refers to a pair of ITRs within a single ceDNA genome or ceDNA vector that are not inverse complements across their full length. As one non-limiting example, an asymmetric ITR pair does not have a symmetrical three-dimensional spatial organization to their cognate ITR such that their 3D structures are different shapes in geometrical space. Stated differently, an asymmetrical ITR pair have the different overall geometric structure, *i.e.*, they have different organization of their A, C-C' and B-B' loops in 3D space (*e.g.*, one ITR may have a short C-C' arm and/or short B-B' arm as compared to the cognate ITR). The difference in sequence between the two ITRs may be due to one or more nucleotide addition, deletion, truncation, or point mutation. In one embodiment, one ITR of the asymmetric ITR pair may be a wild-type AAV ITR sequence and the other ITR a modified ITR as defined herein (*e.g.*, a non-wild-type or synthetic ITR sequence). In another embodiment, neither ITRs of the asymmetric ITR pair is a wild-type AAV sequence and the two ITRs are modified ITRs that have different shapes in geometrical space (*i.e.*, a different overall geometric structure). In some embodiments, one mod-ITRs of an asymmetric ITR pair can have a short C-C' arm and the other ITR can have a different modification (*e.g.*, a single arm, or a short B-B' arm *etc.*) such that they have different three-dimensional spatial organization as compared to the cognate asymmetric mod-ITR.

[0082] As used herein, the term “symmetric ITRs” refers to a pair of ITRs within a single ceDNA genome or ceDNA vector that are wild-type or mutated (*e.g.*, modified relative to wild-type) dependoviral ITR sequences and are inverse complements across their full length. In one non-limiting example, both ITRs are wild type ITRs sequences from AAV2. In another example, neither ITRs are wild type ITR AAV2 sequences (*i.e.*, they are a modified ITR, also referred to as a mutant ITR), and

can have a difference in sequence from the wild type ITR due to nucleotide addition, deletion, substitution, truncation, or point mutation. For convenience herein, an ITR located 5' to (upstream of) an expression cassette in a ceDNA vector is referred to as a "5' ITR" or a "left ITR", and an ITR located 3' to (downstream of) an expression cassette in a ceDNA vector is referred to as a "3' ITR" or a "right ITR".

[0083] As used herein, the terms "substantially symmetrical modified-ITRs" or a "substantially symmetrical mod-ITR pair" refers to a pair of modified-ITRs within a single ceDNA genome or ceDNA vector that are both that have an inverse complement sequence across their entire length. For example, the modified ITR can be considered substantially symmetrical, even if it has some nucleic acid sequences that deviate from the inverse complement sequence so long as the changes do not affect the properties and overall shape. As one non-limiting example, a sequence that has at least 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity to the canonical sequence (as measured using BLAST at default settings), and also has a symmetrical three-dimensional spatial organization to their cognate modified ITR such that their 3D structures are the same shape in geometrical space. Stated differently, a substantially symmetrical modified-ITR pair have the same A, C-C' and B-B' loops organized in 3D space. In some embodiments, the ITRs from a mod-ITR pair may have different reverse complement nucleic acid sequences but still have the same symmetrical three-dimensional spatial organization – that is both ITRs have mutations that result in the same overall 3D shape. For example, one ITR (*e.g.*, 5' ITR) in a mod-ITR pair can be from one serotype, and the other ITR (*e.g.*, 3' ITR) can be from a different serotype, however, both can have the same corresponding mutation (*e.g.*, if the 5' ITR has a deletion in the C region, the cognate modified 3' ITR from a different serotype has a deletion at the corresponding position in the C' region), such that the modified ITR pair has the same symmetrical three-dimensional spatial organization. In such embodiments, each ITR in a modified ITR pair can be from different serotypes (*e.g.*, AAV1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, and 12) such as the combination of AAV2 and AAV6, with the modification in one ITR reflected in the corresponding position in the cognate ITR from a different serotype. In one embodiment, a substantially symmetrical modified ITR pair refers to a pair of modified ITRs (mod-ITRs) so long as the difference in nucleic acid sequences between the ITRs does not affect the properties or overall shape and they have substantially the same shape in 3D space. As a non-limiting example, a mod-ITR that has at least 95%, 96%, 97%, 98% or 99% sequence identity to the canonical mod-ITR as determined by standard means well known in the art such as BLAST (Basic Local Alignment Search Tool), or BLASTN at default settings, and also has a symmetrical three-dimensional spatial organization such that their 3D structure is the same shape in geometric space. A substantially symmetrical mod-ITR pair has the same A, C-C' and B-B' loops in 3D space, *e.g.*, if a modified ITR in a substantially symmetrical mod-ITR pair has a deletion of a C-C' arm, then the cognate mod-ITR has the corresponding deletion of the C-C' loop and also has a similar 3D structure of the remaining A and B-B' loops in the same shape in geometric space of its cognate mod-ITR.

[0084] The term “flanking” refers to a relative position of one nucleic acid sequence with respect to another nucleic acid sequence. Generally, in the sequence ABC, B is flanked by A and C. The same is true for the arrangement AxBxC. Thus, a flanking sequence precedes or follows a flanked sequence but need not be contiguous with, or immediately adjacent to the flanked sequence. In one embodiment, the term flanking refers to terminal repeats at each end of the linear duplex ceDNA vector.

[0085] As used herein, the terms “treat,” “treating,” and/or “treatment” include abrogating, substantially inhibiting, slowing or reversing the progression of a condition, substantially ameliorating clinical symptoms of a condition, or substantially preventing the appearance of clinical symptoms of a condition, obtaining beneficial or desired clinical results. In one embodiment, the condition is hemophilia A. Treating further refers to accomplishing one or more of the following: (a) reducing the severity of the disorder; (b) limiting development of symptoms characteristic of the disorder(s) being treated; (c) limiting worsening of symptoms characteristic of the disorder(s) being treated; (d) limiting recurrence of the disorder(s) in patients that have previously had the disorder(s); and (e) limiting recurrence of symptoms in patients that were previously asymptomatic for the disorder(s). Beneficial or desired clinical results, such as pharmacologic and/or physiologic effects include, but are not limited to, preventing the disease, disorder or condition from occurring in a subject that may be predisposed to the disease, disorder or condition but does not yet experience or exhibit symptoms of the disease (prophylactic treatment), alleviation of symptoms of the disease, disorder or condition, diminishment of extent of the disease, disorder or condition, stabilization (*i.e.*, not worsening) of the disease, disorder or condition, preventing spread of the disease, disorder or condition, delaying or slowing of the disease, disorder or condition progression, amelioration or palliation of the disease, disorder or condition, and combinations thereof, as well as prolonging survival as compared to expected survival if not receiving treatment.

[0086] As used herein, the term “increase,” “enhance,” “raise” (and like terms) generally refers to the act of increasing, either directly or indirectly, a concentration, level, function, activity, or behavior relative to the natural, expected, or average, or relative to a control condition.

[0087] As used herein, the term “minimize,” “reduce,” “decrease,” and/or “inhibit” (and like terms) generally refers to the act of reducing, either directly or indirectly, a concentration, level, function, activity, or behavior relative to the natural, expected, or average, or relative to a control condition.

[0088] As used herein, the term “ceDNA genome” refers to an expression cassette that further incorporates at least one inverted terminal repeat region. A ceDNA genome may further comprise one or more spacer regions. In some embodiments the ceDNA genome is incorporated as an intermolecular duplex polynucleotide of DNA into a plasmid or viral genome.

[0089] As used herein, the term “ceDNA spacer region” refers to an intervening sequence that separates functional elements in the ceDNA vector or ceDNA genome. In some embodiments, ceDNA spacer regions keep two functional elements at a desired distance for optimal functionality. In some embodiments, ceDNA spacer regions provide or add to the genetic stability of the ceDNA genome

within *e.g.*, a plasmid or baculovirus. In some embodiments, ceDNA spacer regions facilitate ready genetic manipulation of the ceDNA genome by providing a convenient location for cloning sites and the like. For example, in certain aspects, an oligonucleotide “polylinker” containing several restriction endonuclease sites, or a non-open reading frame sequence designed to have no known protein (*e.g.*, transcription factor) binding sites can be positioned in the ceDNA genome to separate the *cis* – acting factors, *e.g.*, inserting a 6mer, 12mer, 18mer, 24mer, 48mer, 86mer, 176mer, *etc.* between the terminal resolution site and the upstream transcriptional regulatory element. Similarly, the spacer may be incorporated between the polyadenylation signal sequence and the 3’-terminal resolution site.

[0090] As used herein, the terms “Rep binding site,” “Rep binding element,” “RBE” and “RBS” are used interchangeably and refer to a binding site for Rep protein (*e.g.*, AAV Rep 78 or AAV Rep 68) which upon binding by a Rep protein permits the Rep protein to perform its site-specific endonuclease activity on the sequence incorporating the RBS. An RBS sequence and its inverse complement together form a single RBS. RBS sequences are known in the art, and include, for example, 5’-GCGCGCTCGCTCGCTC-3’ (SEQ ID NO: 140), an RBS sequence identified in AAV2. Any known RBS sequence may be used in the embodiments of the disclosure, including other known AAV RBS sequences and other naturally known or synthetic RBS sequences. Without being bound by theory it is thought that the nuclease domain of a Rep protein binds to the duplex nucleic acid sequence GCTC, and thus the two known AAV Rep proteins bind directly to and stably assemble on the duplex oligonucleotide, 5’-(GCGC)(GCTC)(GCTC)(GCTC)-3’ (SEQ ID NO: 140). In addition, soluble aggregated conformers (*i.e.*, undefined number of inter-associated Rep proteins) dissociate and bind to oligonucleotides that contain Rep binding sites. Each Rep protein interacts with both the nitrogenous bases and phosphodiester backbone on each strand. The interactions with the nitrogenous bases provide sequence specificity whereas the interactions with the phosphodiester backbone are non- or less-sequence specific and stabilize the protein-DNA complex.

[0091] As used herein, the terms “terminal resolution site” and “TRS” are used interchangeably herein and refer to a region at which Rep forms a tyrosine-phosphodiester bond with the 5’ thymidine generating a 3’ OH that serves as a substrate for DNA extension via a cellular DNA polymerase, *e.g.*, DNA pol delta or DNA pol epsilon. Alternatively, the Rep-thymidine complex may participate in a coordinated ligation reaction. In some embodiments, a TRS minimally encompasses a non-base-paired thymidine. In some embodiments, the nicking efficiency of the TRS can be controlled at least in part by its distance within the same molecule from the RBS. When the acceptor substrate is the complementary ITR, then the resulting product is an intramolecular duplex. TRS sequences are known in the art, and include, for example, 5’-GGTTGA-3’, the hexanucleotide sequence identified in AAV2. Any known TRS sequence may be used in the embodiments of the disclosure, including other known AAV TRS sequences and other naturally known or synthetic TRS sequences such as AGTT (SEQ ID NO: 1690), GGTTGG, AGTTGG, AGTTGA, and other motifs such as RRTTRR.

[0092] As used herein, the term “ceDNA-plasmid” refers to a plasmid that comprises a ceDNA genome as an intermolecular duplex.

[0093] As used herein, the term “ceDNA-bacmid” refers to an infectious baculovirus genome comprising a ceDNA genome as an intermolecular duplex that is capable of propagating in *E. coli* as a plasmid, and so can operate as a shuttle vector for baculovirus.

[0094] As used herein, the term “ceDNA-baculovirus” refers to a baculovirus that comprises a ceDNA genome as an intermolecular duplex within the baculovirus genome.

[0095] As used herein, the terms “ceDNA-baculovirus infected insect cell” and “ceDNA-BIIC” are used interchangeably, and refer to an invertebrate host cell (including, but not limited to an insect cell (*e.g.*, an Sf9 cell)) infected with a ceDNA-baculovirus.

[0096] As used herein, the term “ceDNA” refers to capsid-free closed-ended linear double stranded (*ds*) duplex DNA for non-viral gene transfer, synthetic or otherwise. Detailed description of ceDNA is described in International application of PCT/US2017/020828, filed March 3, 2017, the entire contents of which are expressly incorporated herein by reference. Certain methods for the production of ceDNA comprising various inverted terminal repeat (ITR) sequences and configurations using cell-based methods are described in Example 1 of International applications PCT/US18/49996, filed September 7, 2018, and PCT/US2018/064242, filed December 6, 2018 each of which is incorporated herein in its entirety by reference. Certain methods for the production of synthetic ceDNA vectors comprising various ITR sequences and configurations are described, *e.g.*, in International application PCT/US2019/14122, filed January 18, 2019, the entire content of which is incorporated herein by reference.

[0097] As used herein, the term “closed-ended DNA vector” refers to a capsid-free DNA vector with at least one covalently closed end and where at least part of the vector has an intramolecular duplex structure.

[0098] As used herein, the terms “ceDNA vector” and “ceDNA” are used interchangeably and refer to a closed-ended DNA vector comprising at least one terminal palindrome. In some embodiments, the ceDNA comprises two covalently-closed ends.

[0099] As used herein, the term “neDNA” or “nicked ceDNA” refers to a closed-ended DNA having a nick or a gap of 1-100 base pairs in a stem region or spacer region 5' upstream of an open reading frame (*e.g.*, a promoter and transgene to be expressed).

[0100] As used herein, the terms “gap” refers to a discontinued portion of synthetic DNA vector of the present disclosure, creating a stretch of single stranded DNA portion in otherwise double stranded ceDNA. The gap can be 1 base-pair to 100 base-pair long in length in one strand of a duplex DNA. Typical gaps, designed and created by the methods described herein and synthetic vectors generated by the methods can be, for example, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48,

49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59 or 60 bp long in length. Exemplified gaps in the present disclosure can be 1 bp to 10 bp long, 1 to 20 bp long, 1 to 30 bp long in length.

[00101] As defined herein, “reporters” refer to proteins that can be used to provide detectable read-outs. Reporters generally produce a measurable signal such as fluorescence, color, or luminescence. Reporter protein coding sequences encode proteins whose presence in the cell or organism is readily observed. For example, fluorescent proteins cause a cell to fluoresce when excited with light of a particular wavelength, luciferases cause a cell to catalyze a reaction that produces light, and enzymes such as β -galactosidase convert a substrate to a colored product. Exemplary reporter polypeptides useful for experimental or diagnostic purposes include, but are not limited to β -lactamase, β -galactosidase (LacZ), alkaline phosphatase (AP), thymidine kinase (TK), green fluorescent protein (GFP) and other fluorescent proteins, chloramphenicol acetyltransferase (CAT), luciferase, and others well known in the art.

[00102] As used herein, the terms “sense” and “antisense” refer to the orientation of the structural element on the polynucleotide. The sense and antisense versions of an element are the reverse complement of each other.

[00103] As used herein, the term “synthetic AAV vector” and “synthetic production of AAV vector” refers to an AAV vector and synthetic production methods thereof in an entirely cell-free environment.

[00104] As used herein, “reporters” refer to proteins that can be used to provide detectable read-outs. Reporters generally produce a measurable signal such as fluorescence, color, or luminescence. Reporter protein coding sequences encode proteins whose presence in the cell or organism is readily observed. For example, fluorescent proteins cause a cell to fluoresce when excited with light of a particular wavelength, luciferases cause a cell to catalyze a reaction that produces light, and enzymes such as β -galactosidase convert a substrate to a colored product. Exemplary reporter polypeptides useful for experimental or diagnostic purposes include, but are not limited to β -lactamase, β -galactosidase (LacZ), alkaline phosphatase (AP), thymidine kinase (TK), green fluorescent protein (GFP) and other fluorescent proteins, chloramphenicol acetyltransferase (CAT), luciferase, and others well known in the art.

[00105] As used herein, the term “effector protein” refers to a polypeptide that provides a detectable read-out, either as, for example, a reporter polypeptide, or more appropriately, as a polypeptide that kills a cell, *e.g.*, a toxin, or an agent that renders a cell susceptible to killing with a chosen agent or lack thereof. Effector proteins include any protein or peptide that directly targets or damages the host cell’s DNA and/or RNA. For example, effector proteins can include, but are not limited to, a restriction endonuclease that targets a host cell DNA sequence (whether genomic or on an extrachromosomal element), a protease that degrades a polypeptide target necessary for cell survival, a DNA gyrase inhibitor, and a ribonuclease-type toxin. In some embodiments, the expression of an effector protein controlled by a synthetic biological circuit as described herein can participate as a factor in another

synthetic biological circuit to thereby expand the range and complexity of a biological circuit system's responsiveness.

[00106] Transcriptional regulators refer to transcriptional activators and repressors that either activate or repress transcription of a gene of interest, such as FVIII. Promoters are regions of nucleic acid that initiate transcription of a particular gene. Transcriptional activators typically bind nearby to transcriptional promoters and recruit RNA polymerase to directly initiate transcription. Repressors bind to transcriptional promoters and sterically hinder transcriptional initiation by RNA polymerase. Other transcriptional regulators may serve as either an activator or a repressor depending on where they bind and cellular and environmental conditions. Non-limiting examples of transcriptional regulator classes include, but are not limited to homeodomain proteins, zinc-finger proteins, winged-helix (forkhead) proteins, and leucine-zipper proteins.

[00107] As used herein, a "repressor protein" or "inducer protein" is a protein that binds to a regulatory sequence element and represses or activates, respectively, the transcription of sequences operatively linked to the regulatory sequence element. Preferred repressor and inducer proteins as described herein are sensitive to the presence or absence of at least one input agent or environmental input. Preferred proteins as described herein are modular in form, comprising, for example, separable DNA-binding and input agent-binding or responsive elements or domains.

[00108] As used herein, "carrier" includes any and all solvents, dispersion media, vehicles, coatings, diluents, antibacterial and antifungal agents, isotonic and absorption delaying agents, buffers, carrier solutions, suspensions, colloids, and the like. The use of such media and agents for pharmaceutically active substances is well known in the art. Supplementary active ingredients can also be incorporated into the compositions. The phrase "pharmaceutically-acceptable" refers to molecular entities and compositions that do not produce a toxic, an allergic, or similar untoward reaction when administered to a host.

[00109] As used herein, an "input agent responsive domain" is a domain of a transcription factor that binds to or otherwise responds to a condition or input agent in a manner that renders a linked DNA binding fusion domain responsive to the presence of that condition or input. In one embodiment, the presence of the condition or input results in a conformational change in the input agent responsive domain, or in a protein to which it is fused, that modifies the transcription-modulating activity of the transcription factor.

[00110] The term "*in vivo*" refers to assays or processes that occur in or within an organism, such as a multicellular animal. In some of the aspects described herein, a method or use can be said to occur "*in vivo*" when a unicellular organism, such as a bacterium, is used. The term "*ex vivo*" refers to methods and uses that are performed using a living cell with an intact membrane that is outside of the body of a multicellular animal or plant, *e.g.*, explants, cultured cells, including primary cells and cell lines, transformed cell lines, and extracted tissue or cells, including blood cells, among others. The term "*in vitro*" refers to assays and methods that do not require the presence of a cell with an intact membrane,

such as cellular extracts, and can refer to the introducing of a programmable synthetic biological circuit in a non-cellular system, such as a medium not comprising cells or cellular systems, such as cellular extracts.

[00111] The term “promoter,” as used herein, refers to any nucleic acid sequence that regulates the expression of another nucleic acid sequence by driving transcription of the nucleic acid sequence, which can be a target gene, *e.g.*, heterologous target gene, encoding a protein or an RNA. Promoters can be constitutive, inducible, repressible, tissue-specific, or any combination thereof. A promoter is a control region of a nucleic acid sequence at which initiation and rate of transcription of the remainder of a nucleic acid sequence are controlled. A promoter can also contain genetic elements at which regulatory proteins and molecules can bind, such as RNA polymerase and other transcription factors. In some embodiments of the aspects described herein, a promoter can drive the expression of a transcription factor that regulates the expression of the promoter itself. Within the promoter sequence will be found a transcription initiation site, as well as protein binding domains responsible for the binding of RNA polymerase. Eukaryotic promoters will often, but not always, contain “TATA” boxes and “CAT” boxes. Various promoters, including inducible promoters, may be used to drive the expression of transgenes in the ceDNA vectors disclosed herein. A promoter sequence may be bounded at its 3’ terminus by the transcription initiation site and extends upstream (5’ direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background.

[00112] In one embodiment, the promoter contained in the nucleic acid expression cassettes and vectors disclosed herein is a liver-specific promoter.

[00113] The term “liver-specific promoter” encompasses any promoter that confers liver-specific expression to a (trans)gene. Non-limiting examples of liver-specific promoters are provided on the Liver-specific Gene Promoter Database (LSPD, rulai.cshl.edu/LSPD/), and include, for example, the transthyretin (TTR) promoter or TTR-minimal promoter (TTRm), the alpha 1-antitrypsin (AAT) promoter, the albumin (ALB) promoter or minimal promoter, the apolipoprotein A1 (APOA1) promoter or minimal promoter, the complement factor B (CFB) promoter, the ketohexokinase (KHK) promoter, the hemopexin (HPX) promoter or minimal promoter, the nicotinamide N-methyltransferase (NNMT) promoter or minimal promoter, the (liver) carboxylesterase 1 (CES1) promoter or minimal promoter, the protein C (PROC) promoter or minimal promoter, the apolipoprotein C3 (APOC3) promoter or minimal promoter, the mannan-binding lectin serine protease 2 (MASP2) promoter or minimal promoter, the hepcidin antimicrobial peptide (HAMP) promoter or minimal promoter, and the serpin peptidase inhibitor, clade C (antithrombin), member 1 (SERPINC1) promoter or minimal promoter.

[00114] In some embodiments, the promoter is a mammalian liver-specific promoter, in particular a murine or human liver-specific promoter.

[00115] The term “enhancer” as used herein refers to a cis-acting regulatory sequence (*e.g.*, 50-1,500 base pairs) that binds one or more proteins (*e.g.*, activator proteins, or transcription factor) to increase transcriptional activation of a nucleic acid sequence. Enhancers can be positioned up to 1,000,000 base

pars upstream of the gene start site or downstream of the gene start site that they regulate. An enhancer can be positioned within an intronic region, or in the exonic region of an unrelated gene.

[00116] A promoter can be said to drive expression or drive transcription of the nucleic acid sequence that it regulates. The phrases “operably linked,” “operatively positioned,” “operatively linked,” “under control,” and “under transcriptional control” indicate that a promoter is in a correct functional location and/or orientation in relation to a nucleic acid sequence it regulates to control transcriptional initiation and/or expression of that sequence. An “inverted promoter,” as used herein, refers to a promoter in which the nucleic acid sequence is in the reverse orientation, such that what was the coding strand is now the non-coding strand, and vice versa. Inverted promoter sequences can be used in various embodiments to regulate the state of a switch. In addition, in various embodiments, a promoter can be used in conjunction with an enhancer.

[00117] A promoter can be one naturally associated with a gene or sequence, as can be obtained by isolating the 5' non-coding sequences located upstream of the coding segment and/or exon of a given gene or sequence. Such a promoter can be referred to as “endogenous.” Similarly, in some embodiments, an enhancer can be one naturally associated with a nucleic acid sequence, located either downstream or upstream of that sequence.

[00118] In some embodiments, a coding nucleic acid segment is positioned under the control of a “recombinant promoter” or “heterologous promoter,” both of which refer to a promoter that is not normally associated with the encoded nucleic acid sequence it is operably linked to in its natural environment. A recombinant or heterologous enhancer refers to an enhancer not normally associated with a given nucleic acid sequence in its natural environment. Such promoters or enhancers can include promoters or enhancers of other genes; promoters or enhancers isolated from any other prokaryotic, viral, or eukaryotic cell; and synthetic promoters or enhancers that are not “naturally occurring,” *i.e.*, comprise different elements of different transcriptional regulatory regions, and/or mutations that alter expression through methods of genetic engineering that are known in the art. In addition to producing nucleic acid sequences of promoters and enhancers synthetically, promoter sequences can be produced using recombinant cloning and/or nucleic acid amplification technology, including PCR, in connection with the synthetic biological circuits and modules disclosed herein (see, *e.g.*, U.S. Pat. No. 4,683,202, U.S. Pat. No. 5,928,906, each incorporated herein by reference). Furthermore, it is contemplated that control sequences that direct transcription and/or expression of sequences within non-nuclear organelles such as mitochondria, chloroplasts, and the like, can be employed as well.

[00119] As described herein, an “inducible promoter” is one that is characterized by initiating or enhancing transcriptional activity when in the presence of, influenced by, or contacted by an inducer or inducing agent. An “inducer” or “inducing agent,” as defined herein, can be endogenous, or a normally exogenous compound or protein that is administered in such a way as to be active in inducing transcriptional activity from the inducible promoter. In some embodiments, the inducer or inducing agent, *i.e.*, a chemical, a compound or a protein, can itself be the result of transcription or expression of

a nucleic acid sequence (*i.e.*, an inducer can be an inducer protein expressed by another component or module), which itself can be under the control of an inducible promoter. In some embodiments, an inducible promoter is induced in the absence of certain agents, such as a repressor. Examples of inducible promoters include but are not limited to, tetracycline, metallothionine, ecdysone, mammalian viruses (*e.g.*, the adenovirus late promoter; and the mouse mammary tumor virus long terminal repeat (MMTV-LTR)) and other steroid-responsive promoters, rapamycin responsive promoters and the like.

[00120] The terms “DNA regulatory sequences,” “control elements,” and “regulatory elements,” used interchangeably herein, refer to transcriptional and translational control sequences, such as promoters, enhancers, polyadenylation signals, terminators, protein degradation signals, and the like, that provide for and/or regulate transcription of a non-coding sequence (*e.g.*, DNA-targeting RNA) or a coding sequence (*e.g.*, site-directed modifying polypeptide, or Cas9/Csn1 polypeptide) and/or regulate translation of an encoded polypeptide.

[00121] Regulatory elements comprise at least one transcription factor binding site (TFBS), more in particular at least one binding site for a tissue-specific transcription factor, most particularly at least one binding site for a liver-specific transcription factor. Typically, regulatory elements as used herein increase or enhance promoter-driven gene expression when compared to the transcription of the gene from the promoter alone, without the regulatory elements. Thus, regulatory elements particularly comprise enhancer sequences, although it is to be understood that the regulatory elements enhancing transcription are not limited to typical far upstream enhancer sequences, but may occur at any distance of the gene they regulate. Indeed, it is known in the art that sequences regulating transcription may be situated either upstream (*e.g.*, in the promoter region) or downstream (*e.g.*, in the 3'UTR) of the gene they regulate *in vivo*, and may be located in the immediate vicinity of the gene or further away. Although regulatory elements as disclosed herein typically are naturally occurring sequences, combinations of (parts of) such regulatory elements or several copies of a regulatory element, *i.e.*, non-naturally occurring sequences, are themselves also envisaged as regulatory element. Regulatory elements as used herein may be part of a larger sequence involved in transcriptional control, *e.g.*, part of a promoter sequence. However, regulatory elements alone are typically not sufficient to initiate transcription, but require a promoter to this end.

[00122] In one embodiment, the one or more regulatory elements contained in the nucleic acid expression cassettes and vectors disclosed herein are preferably liver-specific. Non-limiting examples of liver-specific regulatory elements are disclosed in WO 2009/130208, incorporated by reference in its entirety herein. Another example of a liver-specific regulatory element is a regulatory element derived from the transthyretin (TTR) gene, also referred to herein as “TTRe.” “Liver-specific expression”, as used herein, refers to the preferential or predominant expression of a (trans)gene (as RNA and/or polypeptide) in the liver as compared to other tissues. In one embodiment, at least 50% of the (trans)gene expression occurs within the liver. According to some embodiments, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 99% or

100% of the (trans)gene expression occurs within the liver. In one embodiment, liver-specific expression entails that there is no ‘leakage’ of expressed gene product to other organs, such as spleen, muscle, heart and/or lung. It is to be understood that, where liver-specific is mentioned in the context of expression, hepatocyte-specific expression is also explicitly envisaged. Similarly, where tissue-specific expression is used in the application, cell-type specific expression of the cell type(s) predominantly making up the tissue is also envisaged.

[00123] As used herein, the term “liver cells” encompasses the cells predominantly populating the liver and encompasses mainly hepatocytes, oval cells, liver sinusoidal endothelial cells (LSEC) and cholangiocytes (epithelial cells forming the bile ducts).

[00124] “Operably linked” refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. For instance, a promoter is operably linked to a coding sequence if the promoter affects its transcription or expression. An “expression cassette” includes a DNA sequence, *e.g.*, heterologous DNA sequence, that is operably linked to a promoter or other regulatory sequence sufficient to direct transcription of the transgene in the ceDNA vector. Suitable promoters include, for example, tissue specific promoters or promoters of AAV origin.

[00125] The term “subject” as used herein refers to a human or animal, to whom treatment, including prophylactic treatment, with the ceDNA vector according to the present disclosure, is provided. Usually, the animal is a vertebrate such as, but not limited to a primate, rodent, domestic animal or game animal. Primates include but are not limited to, chimpanzees, cynomolgous monkeys, spider monkeys, and macaques, *e.g.*, Rhesus. Rodents include mice, rats, woodchucks, ferrets, rabbits and hamsters. Domestic and game animals include, but are not limited to, cows, horses, pigs, deer, bison, buffalo, feline species, *e.g.*, domestic cat, canine species, *e.g.*, dog, fox, wolf, avian species, *e.g.*, chicken, emu, ostrich, and fish. In certain embodiments of the aspects described herein, the subject is a mammal, *e.g.*, a primate or a human. A subject can be male or female. Additionally, a subject can be an infant or a child. In some embodiments, the subject can be a neonate or an unborn subject, *e.g.*, the subject is *in utero*. Preferably, the subject is a mammal. The mammal can be a human, non-human primate, mouse, rat, dog, cat, horse, or cow, but is not limited to these examples. Mammals other than humans can be advantageously used as subjects that represent animal models of diseases and disorders. In addition, the methods and compositions described herein can be used for domesticated animals and/or pets. A human subject can be of any age, gender, race or ethnic group, *e.g.*, Caucasian (white), Asian, African, black, African American, African European, Hispanic, Mideastern, *etc.* In some embodiments, the subject can be a patient or other subject in a clinical setting. In some embodiments, the subject is already undergoing treatment. In some embodiments, the subject is an embryo, a fetus, neonate, infant, child, adolescent, or adult. In some embodiments, the subject is a human fetus, human neonate, human infant, human child, human adolescent, or human adult. In some embodiments, the subject is an animal embryo, or non-human embryo or non-human primate embryo. In some embodiments, the subject is a human embryo.

[00126] The term “control” as used herein is meant to refer to a reference standard. In one embodiment, a control may be a negative control sample obtained from a healthy patient. According to other embodiments, the control is a positive control sample obtained from a patient diagnosed with a genetic disease or disorder (*e.g.*, hemophilia). In one embodiment, the control is a historical control or a standard reference value or a range of values (such as a previously tested control sample, such as a group of hemophilia A patients with a known prognosis or outcome, or a group of samples representing baseline or normal values).

[00127] A difference between a test sample and a control can be an increase or, conversely, a decrease. The difference can be a qualitative difference or a quantitative difference, for example, a statistically significant difference. In some examples, a difference is an increase or decrease, relative to a control, of at least about 5%, such as at least about 10%, at least about 20%, at least about 30%, by less than about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 100%, at least about 150%, at least about 200%, at least about 250%, at least about 300%, at least about 350%, at least about 400%, at least about 500% or more than 500 %.

[00128] As used herein, the term “host cell”, includes any cell type that is susceptible to transformation, transfection, transduction, and the like with a nucleic acid construct or cDNA expression vector of the present disclosure. As non-limiting examples, a host cell can be an isolated primary cell, pluripotent stem cells, CD34⁺ cells), induced pluripotent stem cells, or any of a number of immortalized cell lines (*e.g.*, HepG2 cells). Alternatively, a host cell can be an *in situ* or *in vivo* cell in a tissue, organ or organism.

[00129] The term “exogenous” refers to a substance present in a cell other than its native source. The term “exogenous” when used herein can refer to a nucleic acid (*e.g.*, a nucleic acid encoding a polypeptide) or a polypeptide that has been introduced by a process involving the hand of man into a biological system such as a cell or organism in which it is not normally found, and one wishes to introduce the nucleic acid or polypeptide into such a cell or organism. Alternatively, “exogenous” can refer to a nucleic acid or a polypeptide that has been introduced by a process involving the hand of man into a biological system such as a cell or organism in which it is found in relatively low amounts and one wishes to increase the amount of the nucleic acid or polypeptide in the cell or organism, *e.g.*, to create ectopic expression or levels. In contrast, the term “endogenous” refers to a substance that is native to the biological system or cell.

[00130] The term “sequence identity” refers to the relatedness between two nucleic acid sequences. For purposes of the present disclosure, the degree of sequence identity between two deoxyribonucleotide sequences is determined using the Needleman-Wunsch algorithm (Needleman and Wunsch, 1970, *supra*) as implemented in the Needle program of the EMBOSS package (EMBOSS: The European Molecular Biology Open Software Suite, Rice *et al.*, 2000, *supra*), preferably version 3.0.0 or later. The optional parameters used are gap open penalty of 10, gap extension penalty of 0.5, and the EDNAFULL (EMBOSS version of NCBI NUC4.4) substitution matrix. The output of Needle labeled

“longest identity” (obtained using the -nobrief option) is used as the percent identity and is calculated as follows: (Identical Deoxyribonucleotides.times.100)/(Length of Alignment-Total Number of Gaps in Alignment). The length of the alignment is preferably at least 10 nucleotides, preferably at least 25 nucleotides more preferred at least 50 nucleotides and most preferred at least 100 nucleotides.

[00131] The term “homology” or “homologous” as used herein is defined as the percentage of nucleotide residues that are identical to the nucleotide residues in the corresponding sequence on the target chromosome, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity. Alignment for purposes of determining percent nucleotide sequence homology can be achieved in various ways that are within the skill in the art, for instance, using publicly available computer software such as BLAST, BLAST-2, ALIGN, ClustalW2 or Megalign (DNASTAR) software. Those skilled in the art can determine appropriate parameters for aligning sequences, including any algorithms needed to achieve maximal alignment over the full length of the sequences being compared. In some embodiments, a nucleic acid sequence (*e.g.*, DNA sequence), for example of a homology arm, is considered “homologous” when the sequence is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or more, identical to the corresponding native or unedited nucleic acid sequence (*e.g.*, genomic sequence) of the host cell.

[00132] The term “heterologous,” as used herein, means a nucleotide or polypeptide sequence that is not found in the native nucleic acid or protein, respectively. A heterologous nucleic acid sequence may be linked to a naturally-occurring nucleic acid sequence (or a variant thereof) (*e.g.*, by genetic engineering) to generate a chimeric nucleotide sequence encoding a chimeric polypeptide. A heterologous nucleic acid sequence may be linked to a variant polypeptide (*e.g.*, by genetic engineering) to generate a nucleotide sequence encoding a fusion variant polypeptide.

[00133] A “vector” or “expression vector” is a replicon, such as plasmid, bacmid, phage, virus, virion, or cosmid, to which another DNA segment, *i.e.* an “insert”, may be attached so as to bring about the replication of the attached segment in a cell. A vector can be a nucleic acid construct designed for delivery to a host cell or for transfer between different host cells. A vector can include nucleic acid sequences that allow it to replicate in a host cell, such as an origin of replication. A vector can also include one or more selectable marker genes and other genetic elements. The term “vector” encompasses any genetic element that is capable of replication when associated with the proper control elements and that can transfer gene sequences to cells. In some embodiments, a vector can be an expression vector or recombinant vector. In some embodiments, the vector is an expression vector that contains the regulatory sequences necessary to allow transcription and translation of the inserted gene(s). In some embodiments, the vector is a cDNA vector. In some embodiments, the vector is an AAV vector. In some embodiments, the vector is a retroviral gamma vector, a lentiviral vector, or an adenoviral vector.

[00134] As used herein, the term “expression vector” refers to a vector that directs expression of an RNA or polypeptide from sequences linked to transcriptional regulatory sequences on the vector. The sequences expressed will often, but not necessarily, be heterologous to the cell. An expression vector may comprise additional elements, for example, the expression vector may have two replication systems, thus allowing it to be maintained in two organisms, for example in human cells for expression and in a prokaryotic host for cloning and amplification. The term “expression” refers to the cellular processes involved in producing RNA and proteins and as appropriate, secreting proteins, including where applicable, but not limited to, for example, transcription, transcript processing, translation and protein folding, modification and processing. “Expression products” include RNA transcribed from a gene, and polypeptides obtained by translation of mRNA transcribed from a gene. The term “gene” means the nucleic acid sequence which is transcribed (DNA) to RNA *in vitro* or *in vivo* when operably linked to appropriate regulatory sequences. The gene may or may not include regions preceding and following the coding region, *e.g.*, 5’ untranslated (5’ UTR) or “leader” sequences and 3’ UTR or “trailer” sequences, as well as intervening sequences (introns) between individual coding segments (exons).

[00135] By “recombinant vector” is meant a vector that includes a nucleic acid sequence, *e.g.*, heterologous nucleic acid sequence, or “transgene” that is capable of expression *in vivo*. It should be understood that the vectors described herein can, in some embodiments, be combined with other suitable compositions and therapies. In some embodiments, the vector is episomal. The use of a suitable episomal vector provides a means of maintaining the nucleotide of interest in the subject in high copy number extra chromosomal DNA thereby eliminating potential effects of chromosomal integration.

[00136] The phrase “genetic disease” as used herein refers to a disease, partially or completely, directly or indirectly, caused by one or more abnormalities in the genome, especially a condition that is present from birth. The abnormality may be a mutation, an insertion or a deletion. The abnormality may affect the coding sequence of the gene or its regulatory sequence. The genetic disease may be, but not limited to DMD, hemophilia, cystic fibrosis, Huntington's chorea, familial hypercholesterolemia (LDL receptor defect), hepatoblastoma, Wilson's disease, congenital hepatic porphyria, inherited disorders of hepatic metabolism, Lesch Nyhan syndrome, sickle cell anemia, thalassemia, xeroderma pigmentosum, Fanconi's anemia, retinitis pigmentosa, ataxia telangiectasia, Bloom's syndrome, retinoblastoma, and Tay-Sachs disease.

[00137] As used herein the term “comprising” or “comprises” is used in reference to compositions, methods, and respective component(s) thereof, that are essential to the method or composition, yet open to the inclusion of unspecified elements, whether essential or not.

[00138] As used herein the term “consisting essentially of” refers to those elements required for a given embodiment. The term permits the presence of elements that do not materially affect the basic and novel or functional characteristic(s) of that embodiment. The use of “comprising” indicates inclusion rather than limitation.

[00139] The term “consisting of” refers to compositions, methods, and respective components thereof as described herein, which are exclusive of any element not recited in that description of the embodiment.

[00140] As used herein the term “consisting essentially of” refers to those elements required for a given embodiment. The term permits the presence of additional elements that do not materially affect the basic and novel or functional characteristic(s) of that embodiment of the disclosure.

[00141] As used in this specification and the appended claims, the singular forms “a,” “an,” and “the” include plural references unless the context clearly dictates otherwise. Thus, for example, references to “the method” includes one or more methods, and/or steps of the type described herein and/or which will become apparent to those persons skilled in the art upon reading this disclosure and so forth. Similarly, the word “or” is intended to include “and” unless the context clearly indicates otherwise. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of this disclosure, suitable methods and materials are described below. The abbreviation, “*e.g.*”, is derived from the Latin *exempli gratia* and is used herein to indicate a non-limiting example. Thus, the abbreviation “*e.g.*” is synonymous with the term “for example.”

[00142] Groupings of alternative elements or embodiments of the disclosure disclosed herein are not to be construed as limitations. Each group member can be referred to and claimed individually or in any combination with other members of the group or other elements found herein. One or more members of a group can be included in, or deleted from, a group for reasons of convenience and/or patentability. When any such inclusion or deletion occurs, the specification is herein deemed to contain the group as modified thus fulfilling the written description of all Markush groups used in the appended claims.

[00143] In some embodiments of any of the aspects, the disclosure described herein does not concern a process for cloning human beings, processes for modifying the germ line genetic identity of human beings, uses of human embryos for industrial or commercial purposes or processes for modifying the genetic identity of animals which are likely to cause them suffering without any substantial medical benefit to man or animal, and also animals resulting from such processes.

[00144] Other terms are defined herein within the description of the various aspects of the disclosure.

[00145] All patents and other publications; including literature references, issued patents, published patent applications, and co-pending patent applications; cited throughout this application are expressly incorporated herein by reference for the purpose of describing and disclosing, for example, the methodologies described in such publications that might be used in connection with the technology described herein. These publications are provided solely for their disclosure prior to the filing date of the present application. Nothing in this regard should be construed as an admission that the inventors are not entitled to antedate such disclosure by virtue of prior disclosure or for any other reason. All statements as to the date or representation as to the contents of these documents is based on the

information available to the applicants and does not constitute any admission as to the correctness of the dates or contents of these documents.

[00146] The description of embodiments of the disclosure is not intended to be exhaustive or to limit the disclosure to the precise form disclosed. While specific embodiments of, and examples for, the disclosure are described herein for illustrative purposes, various equivalent modifications are possible within the scope of the disclosure, as those skilled in the relevant art will recognize. For example, while method steps or functions are presented in a given order, alternative embodiments may perform functions in a different order, or functions may be performed substantially concurrently. The teachings of the disclosure provided herein can be applied to other procedures or methods as appropriate. The various embodiments described herein can be combined to provide further embodiments. Aspects of the disclosure can be modified, if necessary, to employ the compositions, functions and concepts of the above references and application to provide yet further embodiments of the disclosure. Moreover, due to biological functional equivalency considerations, some changes can be made in protein structure without affecting the biological or chemical action in kind or amount. These and other changes can be made to the disclosure in light of the detailed description. All such modifications are intended to be included within the scope of the appended claims.

[00147] Specific elements of any of the foregoing embodiments can be combined or substituted for elements in other embodiments. Furthermore, while advantages associated with certain embodiments of the disclosure have been described in the context of these embodiments, other embodiments may also exhibit such advantages, and not all embodiments need necessarily exhibit such advantages to fall within the scope of the disclosure.

II. Expression Cassettes Optimized for Liver-Specific Expression

[00148] The present disclosure provides liver-specific expression cassettes to enhance transcription in liver tissue and/or cells. As discussed in the Examples, the present disclosure provides a novel set of non-natural modifications to a native liver-specific enhancer region that unexpectedly increase acute protein expression level and improve sequence characteristics known to impact protein expression durability.

[00149] In one embodiment, the liver-specific expression cassette provided herein comprises an enhancer nucleic acid sequence. In one embodiment, the liver-specific expression cassette provided herein comprises more than one repeated enhancer nucleic acid sequences. In one embodiment, the liver-specific expression cassette provided herein comprises two repeated enhancer nucleic acid sequences. In one embodiment, the liver-specific expression cassette provided herein comprises three repeated enhancer nucleic acid sequences. In one embodiment, the liver-specific expression cassette provided herein comprises five repeated enhancer nucleic acid sequences. In one embodiment, the liver-specific expression cassette provided herein comprises between two and 10 repeated enhancer nucleic acid sequences. In one embodiment, the liver-specific expression cassette provided herein comprises ten repeated enhancer nucleic acid sequences. In one embodiment, the liver-specific

expression cassette provided herein comprises between 3 and 10 repeated enhancer nucleic acid sequences. In one embodiment, the liver-specific expression cassette comprises more than three repeated enhancer nucleic acid sequences.

[00150] In one embodiment, the liver-specific expression cassette comprises two or more repeated enhancer nucleic acid sequences, wherein at least two nucleic acids (2 mer) separate each repeated enhancer nucleic acid sequence. In one embodiment, the liver-specific expression cassette comprises three or more repeated enhancer nucleic acid sequences, wherein at least two nucleic acids separate each repeated enhancer nucleic acid sequence. In one embodiment, the liver-specific expression cassette comprises five or more repeated enhancer nucleic acid sequences, wherein at least two nucleic acids separate each repeated enhancer nucleic acid sequence. In one embodiment, the liver-specific expression cassette comprises ten or more repeated enhancer nucleic acid sequences, wherein at least two nucleic acids separate each repeated enhancer nucleic acid sequence. In one embodiment, the liver-specific expression cassette comprises about 3 to 10 repeated enhancer nucleic acid sequences, wherein at least two nucleic acids separate each repeated enhancer nucleic acid sequence.

[00151] In one embodiment, the liver-specific expression cassette provided herein comprises two or more repeated enhancer nucleic acid sequences, wherein at least three nucleic acids (3 mer) separate each repeated enhancer nucleic acid sequence. In one embodiment, the liver-specific expression cassette comprises three or more repeated enhancer nucleic acid sequences, wherein at least three nucleic acids separate each repeated enhancer nucleic acid sequence. In one embodiment, the liver-specific expression cassette comprises five or more repeated enhancer nucleic acid sequences, wherein at least three nucleic acids separate each repeated enhancer nucleic acid sequence. In one embodiment, the liver-specific expression cassette comprises ten or more repeated enhancer nucleic acid sequences, wherein at least three nucleic acids separate each repeated enhancer nucleic acid sequence. In one embodiment, the liver-specific expression cassette comprises between 3 and 10 repeated enhancer nucleic acid sequences, wherein at least three nucleic acids separate each repeated enhancer nucleic acid sequence.

[00152] In one embodiment, the liver-specific expression cassette provided herein comprises two or more repeated enhancer nucleic acid sequences, wherein at least five nucleic acids (5 mer) separate each repeated enhancer nucleic acid sequence. In one embodiment, the liver-specific expression cassette comprises three or more repeated enhancer nucleic acid sequences, wherein at least five nucleic acids separate each repeated enhancer nucleic acid sequence. In one embodiment, the liver-specific expression cassette comprises five or more repeated enhancer nucleic acid sequences, wherein at least five nucleic acids separate each repeated enhancer nucleic acid sequence. In one embodiment, the liver-specific expression cassette comprises ten or more repeated enhancer nucleic acid sequences, wherein at least five nucleic acids separate each repeated enhancer nucleic acid sequence. In one embodiment, the liver-specific expression cassette comprises about 3 to 10 repeated enhancer nucleic

acid sequences, wherein at least five nucleic acids separate each repeated enhancer nucleic acid sequence.

[00153] In one embodiment, the liver-specific expression cassette provided herein comprises two or more repeated enhancer nucleic acid sequences, wherein at least 11 nucleic acids (11 mer) separate each repeated enhancer nucleic acid sequence. In one embodiment, the liver-specific expression cassette comprises three or more repeated enhancer nucleic acid sequences, wherein at least 11 nucleic acids separate each repeated enhancer nucleic acid sequence. In one embodiment, the liver-specific expression cassette comprises five or more repeated enhancer nucleic acid sequences, wherein at least 11 nucleic acids separate each repeated enhancer nucleic acid sequence. In one embodiment, the liver-specific expression cassette comprises ten or more repeated enhancer nucleic acid sequences, wherein at least 11 nucleic acids separate each repeated enhancer nucleic acid sequence. In one embodiment, the liver-specific expression cassette comprises about 3 to 10 repeated enhancer nucleic acid sequences, wherein at least 11 nucleic acids separate each repeated enhancer nucleic acid sequence.

[00154] In one embodiment, the liver-specific expression cassette provided herein comprises two or more repeated enhancer nucleic acid sequences, wherein at least 30 nucleic acids (30 mer) separate each repeated enhancer nucleic acid sequence. In one embodiment, the liver-specific expression cassette comprises three or more repeated enhancer nucleic acid sequences, wherein at least 30 nucleic acids separate each repeated enhancer nucleic acid sequence. In one embodiment, the liver-specific expression cassette comprises five or more repeated enhancer nucleic acid sequences, wherein at least 30 nucleic acids separate each repeated enhancer nucleic acid sequence. In one embodiment, the liver-specific expression cassette comprises ten or more repeated enhancer nucleic acid sequences, wherein at least 30 nucleic acids separate each repeated enhancer nucleic acid sequence. In one embodiment, the liver-specific expression cassette provided herein comprises about 3 to 10 repeated enhancer nucleic acid sequences, wherein at least 30 nucleic acids separate each repeated enhancer nucleic acid sequence.

[00155] In one embodiment, the liver-specific expression cassette provided herein comprises two or more repeated enhancer nucleic acid sequences, wherein between about 2 and 30 nucleic acids separate each repeated enhancer nucleic acid sequence. In one embodiment, the liver-specific expression cassette provided herein comprises three or more repeated enhancer nucleic acid sequences, wherein between about 2 and 30 nucleic acids separate each repeated enhancer nucleic acid sequence. In one embodiment, the liver-specific expression cassette provided herein comprises five or more repeated enhancer nucleic acid sequences, wherein between about 2 and 30 nucleic acids separate each repeated enhancer nucleic acid sequence. In one embodiment, the liver-specific expression cassette provided herein comprises ten or more repeated enhancer nucleic acid sequences, wherein between about 2 and 30 nucleic acids separate each repeated enhancer nucleic acid sequence. In one embodiment, the liver-specific expression cassette comprises about 3 to 10 repeated enhancer nucleic acid sequences, wherein between about 2 and 30 nucleic acids separate each repeated enhancer nucleic acid sequence.

[00156] In one embodiment, the enhancer nucleic acid sequences are further operably linked to a liver-specific promoter and a transgene. In one embodiment, the liver-specific promoter is a human liver-specific promoter.

[00157] In one embodiment, the liver-specific promoter is selected from the group consisting of a minimal TTR promoter (TTRm), an AAT promoter, an albumin (ALB) promoter or minimal promoter, an apolipoprotein A1 (APOA1) promoter or minimal promoter, a complement factor B (CFB) promoter, a ketohexokinase (KHK) promoter, a hemopexin (HPX) promoter or minimal promoter, a nicotinamide N-methyltransferase (NNMT) promoter or minimal promoter, a carboxylesterase 1 (CES1) promoter or minimal promoter, a protein C (PROC) promoter or minimal promoter, an apolipoprotein C3 (APOC3) promoter or minimal promoter, a mannan-binding lectin serine protease 2 (MASP2) promoter or minimal promoter, a hepcidin antimicrobial peptide (HAMP) promoter or minimal promoter, or a serpin peptidase inhibitor, clade C (antithrombin), member 1 (SERPINC1) promoter or minimal promoter.

[00158] In some embodiments, a promoter may also be a promoter from a human gene. The promoter may also be a tissue specific promoter, such as a liver-specific promoter, such as human alpha 1-antitrypsin (HAAT). In one embodiment, the promoter may be synthetic.

[00159] Non-limiting examples of suitable promoters for use in accordance with the present disclosure include any of the promoters described herein, or any of the following:

[00160] In one embodiment, the promoter is hAAT core, the human alpha 1 antitrypsin (hAAT) promoter (Core promoter sequence from human A1AT gene). In one embodiment, the hAAT promoter comprises the sequence set forth as SEQ ID NO: 210 below:

```
GATCTTGCTACCAGTGGAACAGCCACTAAGGATTCTGCAGTGAGAGCAGAGGGCCAGCT
AAGTGGTACTCTCCAGAGACTGTCTGACTCACGCCACCCCTCCACCTTGGACACAGGA
CGCTGTGGTTTCTGAGCCAGGTACAATGACTCCTTTTCGGTAAGTGCAGTGGAAGCTGTAC
ACTGCCCAGGCAAAGCGTCCGGGCAGCGTAGGCGGGCGACTCAGATCCCAGCCAGTGGA
CTTAGCCCCGTGTTTGCTCCTCCGATAACTGGGGTGACCTTGGTTAATATTACACAGCAGCC
TCCCCCGTTGCCCCTCTGGATCCACTGCTTAAATACGGACGAGGACAGG
(SEQ ID NO: 210)
```

[00161] In one embodiment, the promoter comprises a nucleic acid sequence at least about 85% identical to SEQ ID NO: 210. In one embodiment, the promoter comprises a nucleic acid sequence at least about 90% identical to SEQ ID NO: 210. In one embodiment, the promoter comprises a nucleic acid sequence at least about 95% identical to SEQ ID NO: 210. In one embodiment, the promoter comprises a nucleic acid sequence at least about 96% identical to SEQ ID NO: 210. In one embodiment, the promoter comprises a nucleic acid sequence at least about 97% identical to SEQ ID NO: 210. In one embodiment, the promoter comprises a nucleic acid sequence at least about 98% identical to SEQ ID NO: 210. In one embodiment, the promoter comprises a nucleic acid sequence at least about 99%

identical to SEQ ID NO: 210. In one embodiment, the promoter consists of the nucleic acid sequence of SEQ ID NO: 210.

[00162] In one embodiment, the promoter is the minimal transthyretin promoter (TTRm). In one embodiment, the TTRm promoter comprises the sequence set forth as SEQ ID NO: 211 below:

GTCTGTCTGCACATTTTCGTAGAGCGAGTGTTCGATACTCTAATCTCCCTAGGCAAGGTT
CATATTTGTGTAGGTTACTTATTCTCCTTTTGTGACTAAGTCAATAATCAGAATCAGCAG
GTTTGGAGTCAGCTTGGCAGGGATCAGCAGCCTGGGTTGGAAGGAGGGGGTATAAAAGC
CCCTTCACCAGGAGAAGCCGTC (SEQ ID NO: 211)

[00163] In one embodiment, the promoter comprises a nucleic acid sequence at least about 85% identical to SEQ ID NO: 211. In one embodiment, the promoter comprises a nucleic acid sequence at least about 90% identical to SEQ ID NO: 211. In one embodiment, the promoter comprises a nucleic acid sequence at least about 95% identical to SEQ ID NO: 211. In one embodiment, the promoter comprises a nucleic acid sequence at least about 96% identical to SEQ ID NO: 211. In one embodiment, the promoter comprises a nucleic acid sequence at least about 97% identical to SEQ ID NO: 211. In one embodiment, the promoter comprises a nucleic acid sequence at least about 98% identical to SEQ ID NO: 211. In one embodiment, the promoter comprises a nucleic acid sequence at least about 99% identical to SEQ ID NO: 211. In one embodiment, the promoter consists of the nucleic acid sequence of SEQ ID NO: 211.

[00164] In one embodiment, the promoter is hAAT_core_C06, a CpG minimized version of the hAAT core promoter (A1AT gene promoter). In one embodiment, the hAAT promoter comprises the sequence set forth as SEQ ID NO: 212 below:

GATCTTGCTACCAGTGGAACAGCCACTAAGGATTCTGCAGTGAGAGCAGAGGGCCAGCT
AAGTGGTACTCTCCAGAGACTGTCTGACTCATGCCACCCCCTCCACCTTGGACACAGGA
CACTGTGGTTTCTGAGCCAGGTACAATGACTCCTTTTGGTAAGTGCAGTGGAAGCTGTAC
ACTGCCCAGGCAAAGTGTCCGGGCAGCGTAGGCGGGCGACTCAGATCCCAGCCAGTGGA
CTTAGCCCCGTGTTGCTCCTCCGATAACTGGGGTGACCTTGGTTAATATTCACCAGCAGCC
TCCCCCGTTGCCCCCTCTGGATCCACTGCTTAAATACGGACGAGGACAGG (SEQ ID NO:
212).

[00165] In one embodiment, the promoter comprises a nucleic acid sequence at least about 85% identical to SEQ ID NO: 212. In one embodiment, the promoter comprises a nucleic acid sequence at least about 90% identical to SEQ ID NO: 212. In one embodiment, the promoter comprises a nucleic acid sequence at least about 95% identical to SEQ ID NO: 212. In one embodiment, the promoter comprises a nucleic acid sequence at least about 96% identical to SEQ ID NO: 212. In one embodiment, the promoter comprises a nucleic acid sequence at least about 97% identical to SEQ ID NO: 212. In one embodiment, the promoter comprises a nucleic acid sequence at least about 98% identical to SEQ ID NO: 212. In one embodiment, the promoter comprises a nucleic acid sequence at least about 99%

identical to SEQ ID NO: 212. In one embodiment, the promoter consists of the nucleic acid sequence of SEQ ID NO: 212.

[00166] In one embodiment, the promoter is hAAT_core_C07, a CpG minimized version of the hAAT core promoter (A1AT gene promoter). In one embodiment, the hAAT promoter comprises the sequence set forth as SEQ ID NO: 213 below:

GATCTTGCTACCAAGTGGAAACAGCCACTAAGGATTCTGCAGTGAGAGCAGAGGGCCAGCT
AAGTGGTACTCTCCAGAGACTGTCTGACTCACGCCACCCCCTCCACCTTGGACACAGGA
CGCTGTGGTTTCTGAGCCAGGTACAATGACTCCTTTTCGGTAAGTGCAGTGGAAGCTGTAC
ACTGCCCAGGCAAAGCGTCCGGGCAGCGTAGGCGGGCGACTCAGATCCCAGCCAGTGGA
CTTAGCCCCCTGTTTGCTCCTCTGATAACTGGGGTGACCTTGGTTAATATTCACCAGCAGCC
TCCCCTGTTGCCCCTCTGGATCCACTGCTTAAATACGGACAAGGACAGG (SEQ ID NO: 213)

[00167] In one embodiment, the promoter comprises a nucleic acid sequence at least about 85% identical to SEQ ID NO: 213. In one embodiment, the promoter comprises a nucleic acid sequence at least about 90% identical to SEQ ID NO: 213. In one embodiment, the promoter comprises a nucleic acid sequence at least about 95% identical to SEQ ID NO: 213. In one embodiment, the promoter comprises a nucleic acid sequence at least about 96% identical to SEQ ID NO: 213. In one embodiment, the promoter comprises a nucleic acid sequence at least about 97% identical to SEQ ID NO: 213. In one embodiment, the promoter comprises a nucleic acid sequence at least about 98% identical to SEQ ID NO: 213. In one embodiment, the promoter comprises a nucleic acid sequence at least about 99% identical to SEQ ID NO: 213. In one embodiment, the promoter consists of the nucleic acid sequence of SEQ ID NO: 213.

[00168] In one embodiment, the promoter is hAAT_core_C08, a CpG minimized version of the hAAT core promoter (A1AT gene promoter). In one embodiment, the hAAT promoter comprises the sequence set forth as SEQ ID NO: 214 below:

GATCTTGCTACCAAGTGGAAACAGCCACTAAGGATTCTGCAGTGAGAGCAGAGGGCCAGCT
AAGTGGTACTCTCCAGAGACTGTCTGACTCACGCCACCCCCTCCACCTTGGACACAGGA
CGCTGTGGTTTCTGAGCCAGGTACAATGACTCCTTTTCGGTAAGTGCAGTGGAAGCTGTAC
ACTGCCCAGGCAAAGCGTCTGGGCAGCATAGGCAGGCGACTCAGATCCCAGCCAGTGGA
CTTAGCCCCCTGTTTGCTCCTCCGATAACTGGGGTGACCTTGGTTAATATTCACCAGCAGCC
TCCCCCGTTGCCCCTCTGGATCCACTGCTTAAATACGGACGAGGACAGG (SEQ ID NO: 214)

[00169] In one embodiment, the promoter comprises a nucleic acid sequence at least about 85% identical to SEQ ID NO: 214. In one embodiment, the promoter comprises a nucleic acid sequence at least about 90% identical to SEQ ID NO: 214. In one embodiment, the promoter comprises a nucleic acid sequence at least about 95% identical to SEQ ID NO: 214. In one embodiment, the promoter comprises a nucleic acid sequence at least about 96% identical to SEQ ID NO: 214. In one embodiment, the promoter comprises a nucleic acid sequence at least about 97% identical to SEQ ID NO: 214. In one embodiment, the promoter comprises a nucleic acid sequence at least about 98% identical to SEQ

ID NO: 214. In one embodiment, the promoter comprises a nucleic acid sequence at least about 99% identical to SEQ ID NO: 214. In one embodiment, the promoter consists of the nucleic acid sequence of SEQ ID NO: 214.

[00170] In one embodiment, the promoter is hAAT_core_C09, a CpG minimized version of the hAAT core promoter (A1AT gene promoter). In one embodiment, the hAAT promoter comprises the sequence set forth as SEQ ID NO: 215 below:

GATCTTGCTACCACTGGAACAGCCACTAAGGATTCTGCAGTGAGAGCAGAGGGCCAGCT
AAGTGGTACTCTCCAGAGACTGTCTGACTCACGCCACCCCCTCCACCTTGGACACAGGA
CGCTGTGGTTTCTGAGCCAGGTACAATGACTCCTTTTCGGTAAGTGCAGTGGAAGCTGTAC
ACTGCCCAGGCAAAGCGTCCGGGCAGCGTAGGCGGGCGACTCAGATCCCAGCCAGTGGA
CTTAGCCCCCTGTTTGCTCCTCTGATAACTGGGGTGACCTTGGTTAATATTCACCAGCAGCC
TCCCCCGTTGCCCCTCTGGATCCACTGCTTAAATACAGACGAGGACAGG (SEQ ID NO: 215)

[00171] In one embodiment, the promoter comprises a nucleic acid sequence at least about 85% identical to SEQ ID NO: 215. In one embodiment, the promoter comprises a nucleic acid sequence at least about 90% identical to SEQ ID NO: 215. In one embodiment, the promoter comprises a nucleic acid sequence at least about 95% identical to SEQ ID NO: 215. In one embodiment, the promoter comprises a nucleic acid sequence at least about 96% identical to SEQ ID NO: 215. In one embodiment, the promoter comprises a nucleic acid sequence at least about 97% identical to SEQ ID NO: 215. In one embodiment, the promoter comprises a nucleic acid sequence at least about 98% identical to SEQ ID NO: 215. In one embodiment, the promoter comprises a nucleic acid sequence at least about 99% identical to SEQ ID NO: 215. In one embodiment, the promoter consists of the nucleic acid sequence of SEQ ID NO: 215.

[00172] In one embodiment, the promoter is hAAT_core_C10, a CpG minimized version of the hAAT core promoter (A1AT gene promoter). In one embodiment, the hAAT promoter comprises the sequence set forth as SEQ ID NO: 216 below:

GATCTTGCTACCACTGGAACAGCCACTAAGGATTCTGCAGTGAGAGCAGAGGGCCAGCT
AAGTGGTACTCTCCAGAGACTGTCTGACTCACGCCACCCCCTCCACCTTGGACACAGGA
CGCTGTGGTTTCTGAGCCAGGTACAATGACTCCTTTTCGGTAAGTGCAGTGGAAGCTGTAC
ACTGCCCAGGCAAAGCGTCCGGGCAGCGTAGGCGGGCGACTCAGATCCCAGCCAGTGGA
CTTAGCCCCCTGTTTGCTCCTCTGATAACTGGGGTGACCTTGGTTAATATTCACCAGCAGCC
TCCCCCGTTGCCCCTCTGGATCCACTGCTTAAATACAGACGAGGACAGG (SEQ ID NO: 216)

[00173] In one embodiment, the promoter comprises a nucleic acid sequence at least about 85% identical to SEQ ID NO: 216. In one embodiment, the promoter comprises a nucleic acid sequence at least about 90% identical to SEQ ID NO: 216. In one embodiment, the promoter comprises a nucleic acid sequence at least about 95% identical to SEQ ID NO: 216. In one embodiment, the promoter comprises a nucleic acid sequence at least about 96% identical to SEQ ID NO: 216. In one embodiment, the promoter comprises a nucleic acid sequence at least about 97% identical to SEQ ID NO: 216. In

one embodiment, the promoter comprises a nucleic acid sequence at least about 98% identical to SEQ ID NO: 216. In one embodiment, the promoter comprises a nucleic acid sequence at least about 99% identical to SEQ ID NO: 216. In one embodiment, the promoter consists of the nucleic acid sequence of SEQ ID NO: 216.

[00174] In one embodiment, the promoter is hAAT_core_truncated, 5p truncated hAAT core promoter derived from hAAT_core (SEQ ID NO: 210). In one embodiment, the hAAT promoter comprises the sequence set forth as SEQ ID NO: 217 below:

GATCTTGCTACCAGTGGAACAGCCACTAAGGATTCTGCAGTGAGAGCAGAGGGCCAGCT
AAGTGGTACTCTCCCAGAGACTGTCTGACTCACGCCACCCCCTCCACCTTGGACACAGGA
CGCTGTGGTTTCTGAGCCAGGTACAATGACTCCTTTTCGGTAAAGTGCAAGCTGTAC
ACTGCCCAGGCAAAGCGTCCGGGCAGCGTAGGCGGGCGACTCAGATCCCAGCCAGTGGA
CTTAGCCCCCTGTTTGCTCCTCTGATAACTGGGGTGACCTTGGTTAATATTCACCAGCAGCC
TCCCCCGTTGCCCCTCTGGATCCACTGCTTAAATACAGACGAGGACAGG (SEQ ID NO: 217)

[00175] In one embodiment, the promoter comprises a nucleic acid sequence at least about 85% identical to SEQ ID NO: 217. In one embodiment, the promoter comprises a nucleic acid sequence at least about 90% identical to SEQ ID NO: 217. In one embodiment, the promoter comprises a nucleic acid sequence at least about 95% identical to SEQ ID NO: 217. In one embodiment, the promoter comprises a nucleic acid sequence at least about 96% identical to SEQ ID NO: 217. In one embodiment, the promoter comprises a nucleic acid sequence at least about 97% identical to SEQ ID NO: 217. In one embodiment, the promoter comprises a nucleic acid sequence at least about 98% identical to SEQ ID NO: 217. In one embodiment, the promoter comprises a nucleic acid sequence at least about 99% identical to SEQ ID NO: 217. In one embodiment, the promoter consists of the nucleic acid sequence of SEQ ID NO: 217.

[00176] **Table 1** below lists core promoter sequences, and their corresponding SEQ ID NOs, that can be implemented in ceDNA FVIII therapeutics described herein.

Table 1. Core Promoters

	Name	Description	SEQ ID NO.
GE-015	hAAT_core	Core promoter sequence from human A1AT gene	210
GE-1121	TTRm	Core promoter sequence from mouse Transthyretin gene	211
GE-1133	hAAT_core_C06	CpG minimized version of the hAAT core promoter (A1AT gene promoter)	212
GE-1134	hAAT_core_C07	CpG minimized version of the hAAT core promoter (A1AT gene promoter)	213
GE-1135	hAAT_core_C08	CpG minimized version of the hAAT core promoter (A1AT gene promoter)	214
GE-1136	hAAT_core_C09	CpG minimized version of the hAAT core promoter (A1AT gene promoter)	215

GE-1137	hAAT_core_C10	CpG minimized version of the hAAT core promoter (A1AT gene promoter) (also referred to as hAAT(979))	216
GE-1170	hAAT_core_truncated	5p truncated hAAT core promoter derived from GE-015	217

[00177] According to particular embodiments, the promoter is selected from the group consisting of: human alpha 1-antitrypsin (hAAT) promoter (including the CpG minimized hAAT(979) promoter (CpGmin hAAT_core_C10) and other CpGmin_hAAT promoters like hAAT_core_C06; hAAT_core_C07; hAAT_core_C08; and hAAT_core_C09) and the transthyretin (TTR) liver-specific promoter.

[00178] In one embodiment, the TTRm comprises SEQ ID NO: 211. In one embodiment, the serpin enhancer comprises SEQ ID NO: 19. In one embodiment, the TTRm 5'UTR comprises SEQ ID NO: 141 (ACACAGATCCACAAGCTCCTG).

[00179] In one embodiment, the CpGmin_hAAT promoter comprises a sequence selected from any one of SEQ ID NOs 212, 213, 214, 215 or 216.

[00180] In one embodiment, the enhancer is selected from the group consisting of: a SERPIN enhancer (SerpEnh), human SERPINA1 enhancer, Hepatic Nuclear Factor 4 binding site (HNF4), the transthyretin (TTRe) gene enhancer (TTRe), the Hepatic Nuclear Factor 1 binding site (HNF1), Human apolipoprotein E/C-I liver-specific enhancer (ApoE_Enh), the enhancer region from Pro-albumin gene (ProEnh).

[00181] In one embodiment, the enhancer is a SERPINA1 enhancer. In one embodiment, the enhancer is a SERPINA1 enhancer variant, selected from a nucleic acid sequence as set forth in **Table 4**, herein. In one embodiment, the SERPINA1 enhancer comprises a sequence having at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% identity to, comprises or consists of any one of the nucleic acid sequences set forth in **Table 2**, herein.

[00182] According to further embodiments, the enhancer is a human SERPIN1A enhancer. According to still further embodiments, the human SERPIN1A enhancer comprises SEQ ID NO: 81 shown below.

GGGGGAGGCTGCTGGTGAATATTAACCAAGGTCACCCAGTTATCGGAGGAGCAAACAG
GGGCTAAGTCCAC (SEQ ID NO: 81)

[00183] In one embodiment, the enhancer is a Chinese Tree Shrew SERPINA1 enhancer. According to further embodiments, the Chinese Tree Shrew SERPINA1 enhancer comprises SEQ ID NO: 82 shown below.

GGAGGCTGTTGGTGAATATTAACCAAGGTCACCCAGTTATCGGAGGAGCAAACAAGGG
CTAAGTCCAC (SEQ ID NO: 82)

[00184] In one embodiment, the enhancer is a Chinese Tree Shrew SERPINA1 enhancer. According to further embodiments, the Chinese Tree Shrew SERPINA1 enhancer comprises SEQ ID NO: 122 shown below.

GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAACAAGGG
CTAAGTCCAC (SEQ ID NO: 122)

[00185] In one embodiment, the enhancer is a Bushbaby SERPINA1 enhancer. According to further embodiments, the Bushbaby SERPINA1 enhancer comprises SEQ ID NO: 83 shown below.

GGGGGAAGCTACTGGTGAATATTAACCAAGGTCACCCAGTTATCAGGGAGCAAACAGGA
GCTAAGTCCAT (SEQ ID NO: 83)

[00186] In one embodiment, the enhancer is a HNF4 enhancer. In one embodiment, the enhancer is HNF4. According to further embodiments, the HNF4 enhancer comprises SEQ ID NO: 84 shown below.

GAGGGAGGCTGCTGGTGAATATTAACCAAGGTCACCTCAGTTATCAGAGGAGCAAACAGGGG
CAAAGTCCAT (SEQ ID NO: 84)

[00187] In one embodiment, the enhancer is HNF4_FOXA. According to further embodiments, the HNF4_FOXA enhancer comprises SEQ ID NO: 85 shown below.

GGGGGAGGCTGCTGGTAAACATTAACCAAGGTCACCCCAGTTATCAGAGGAGCAAACAGGG
GCAAAGTCCAC (SEQ ID NO: 85)

[00188] CpG dinucleotides are undesirable for gene therapy applications. CpGs can impact expression durability through stimulation of the innate immune system and through methylation-based silencing. Accordingly, in some embodiments, CpGs are removed from the enhancer nucleic acid sequences. In one embodiment, internal CpGs are removed.

[00189] In one embodiment, the enhancer comprises human SERPINA1 enhancer, wherein CpG dinucleotides have been minimized.

[00190] In one embodiment, the enhancer comprises Chinese Tree Shrew SERPINA1 enhancer, wherein CpG dinucleotides have been minimized.

[00191] In one embodiment, the enhancer comprises Bushbaby SERPINA1 enhancer, wherein CpG dinucleotides have been minimized.

[00192] In one embodiment, the enhancer comprises HNF4, wherein CpG dinucleotides have been minimized.

[00193] In one embodiment, the enhancer comprises HNF4_FOXA, wherein CpG dinucleotides have been minimized.

[00194] In one embodiment, the enhancer comprises human SERPINA1 enhancer, wherein poly-C/poly-G have been minimized.

[00195] In one embodiment, the enhancer comprises Chinese Tree Shrew SERPINA1 enhancer, wherein poly-C/poly-G have been minimized.

[00196] In one embodiment, the enhancer comprises Bushbaby SERPINA1 enhancer, wherein poly-C/poly-G have been minimized.

[00197] In one embodiment, the enhancer comprises HNF4, wherein poly-C/poly-G have been minimized.

[00198] In one embodiment, the enhancer comprises HNF4_FOXA, wherein poly-C/poly-G have been minimized.

[00199] In one embodiment, the enhancer comprises human SERPINA1 enhancer, wherein CpG dinucleotides and poly-C/poly-G have been minimized.

[00200] In one embodiment, the enhancer comprises Chinese Tree Shrew SERPINA1 enhancer, wherein CpG dinucleotides and poly-C/poly-G have been minimized.

[00201] In one embodiment, the enhancer comprises Bushbaby SERPINA1 enhancer, wherein CpG dinucleotides and poly-C/poly-G have been minimized.

[00202] In one embodiment, the enhancer comprises HNF4, wherein CpG dinucleotides and poly-C/poly-G have been minimized.

[00203] In one embodiment, the enhancer comprises HNF4_FOXA, wherein CpG dinucleotides and poly-C/poly-G have been minimized.

[00204] In some embodiments, the enhancer is selected from a sequence shown in **Table 2**, below.

Table 2. Enhancers

Name	Description	Sequence	SEQ ID NO:
3x_HNF4_FOXA_v1	3x repeat of the Human SERPINA1 enhancer with FOXA & HNF4 consensus sites ("C" spacer in bold)	GGGGGAGGCTGCTGGTAAACATTAACCAAGGT CACCCCAGTTATCAGAGGAGCAAACAGGGGCA AAGTCCACCGGGGGAGGCTGCTGGTAAACATT AACCAAGGTCACCCCAGTTATCAGAGGAGCAA ACAGGGGGCAAAGTCCACCGGGGGAGGCTGCTG GTAAACATTAACCAAGGTCACCCCAGTTATCA GAGGAGCAAACAGGGGCAAAGTCCAC	1
3x_HNF4_FOXA_v1_CpGmin	3x repeat of HNF4_FOXA_v1 with CpG minimization ("A" spacer in bold)	AGGGGAGGCTGCTGGTAAACATTAACCAAGGT CACCCCAGTTATCAGAGGAGCAAACAGGGGCA AAGTCCACAGGGGGAGGCTGCTGGTAAACATT AACCAAGGTCACCCCAGTTATCAGAGGAGCAA ACAGGGGGCAAAGTCCACAGGGGGAGGCTGCTG GTAAACATTAACCAAGGTCACCCCAGTTATCA GAGGAGCAAACAGGGGCAAAGTCCAT	2
3x_HNF4_FOXA_v1_SecondaryS truct_min_v 1	3x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v1 ("C" spacer in bold)	GAGGGAGGCTGCTGGTAAACATTAACCAAGGT CACCCCAGTTATCAGAGGAGCAAACAGGGGCAA AGTCCACCGAGGGAGGCTGCTGGTAAACATTA ACCAAGGTCACCCCAGTTATCAGAGGAGCAAAC AGGGGGCAAAGTCCACCGAGGGAGGCTGCTGGT AAACATTAACCAAGGTCACCCCAGTTATCAGAG GAGCAAACAGGGGCAAAGTCCAC	3

3x_HNF4_FOXA_v1_SecondaryS truct_min_v1_CpG_min	3x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization and CpG minimization v1 ("A" spacer in bold)	AGGGAGGCTGCTGGTAAACATTAACCAAGGTC ACCCAGTTATCAGAGGAGCAAACAGGGGCAAA GTCCACAGAGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGAGCAAACA GGGGCAAAGTCCACAGAGGGAGGCTGCTGGTA AACATTAACCAAGGTCACCCAGTTATCAGAGG AGCAAACAGGGGCAAAGTCCAT	4
3x_HNF4_FOXA_v1_SecondaryS truct_min_v2	3x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v2 ("C" spacer)	GGGGGAGGCTGCTGGTAAACATTAACCAAGGT CACCTCAGTTATCAGAGGAGCAAACAGGGACA AAGTCCACCGGGGGAGGCTGCTGGTAAACATT AACCAAGGTCACCTCAGTTATCAGAGGAGCAA ACAGGGACAAAGTCCACCGGGGGAGGCTGCTG GTAAACATTAACCAAGGTCACCTCAGTTATCAG AGGAGCAAACAGGGACAAAGTCCAC	5
3x_HNF4_FOXA_v1_SecondaryS truct_min_v2_CpG_min	3x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization and CpG minimization v2 ("A" spacer)	AGGGGAGGCTGCTGGTAAACATTAACCAAGGT CACCTCAGTTATCAGAGGAGCAAACAGGGACA AAGTCCACAGGGGGAGGCTGCTGGTAAACATT AACCAAGGTCACCTCAGTTATCAGAGGAGCAA ACAGGGACAAAGTCCACAGGGGGAGGCTGCTG GTAAACATTAACCAAGGTCACCTCAGTTATCAG AGGAGCAAACAGGGACAAAGTCCACA	6
3x_HNF4_FOXA_v1_SecondaryS truct_min_v3	3x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v3 ("C" spacer)	GGGAGGCTGCTGGTAAACATTAACCAAGGTCA CCCCAGTTATCAGAGGAGCAAACAAGGGCAAA GTCCACCGGGAGGCTGCTGGTAAACATTAACC AAGGTCACCCCAGTTATCAGAGGAGCAAACAA GGGCAAAGTCCACCGGGAGGCTGCTGGTAAAC ATTAACCAAGGTCACCCCAGTTATCAGAGGAG CAAACAAGGGCAAAGTCCAC	7
3x_HNF4_FOXA_v1_SecondaryS truct_min_v3_CpG_min	3x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization and CpG minimization v3 ("A" spacer)	AGGGAGGCTGCTGGTAAACATTAACCAAGGTC ACCCAGTTATCAGAGGAGCAAACAAGGGCAA AGTCCACAGGGAGGCTGCTGGTAAACATTAAC CAAGGTCACCCCAGTTATCAGAGGAGCAAACA AGGGCAAAGTCCACAGGGAGGCTGCTGGTAA CATTAACCAAGGTCACCCCAGTTATCAGAGGA GCAAACAAGGGCAAAGTCCACA	8
3x_HNF4_FOXA_v1_SecondaryS truct_min_v4_Aspacers (no spacer inbetween the repeats)	3x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v4 (2585)	AGGAGGAGGCTGCTGGTAAACATTAACCAAGG TCACCTCAGTTATCAGAGGAGCAAACAGGGGC AAAGTCCACAGGAGGAGGCTGCTGGTAAACAT TAACCAAGGTCACCTCAGTTATCAGAGGAGCA AACAGGGGCAAAGTCCACAGGAGGAGGCTGCT GGTAAACATTAACCAAGGTCACCTCAGTTATCA GAGGAGCAAACAGGGGCAAAGTCCACA	9
3x_HNF4_FOXA_v1_SecondaryS truct_min_v5_Aspacers ("A" spacer inbetween the repeats)	3x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v5	AGGGGGAGGCTGCTGGTAAACATTAACCAAGG TCACCTCAGTTATCAGAGGAGCAAACAGGTGC AAAGTCCACAGGGGAGGCTGCTGGTAAACAT TAACCAAGGTCACCTCAGTTATCAGAGGAGCA AACAGGTGCAAAGTCCACAGGGGAGGCTGCT GGTAAACATTAACCAAGGTCACCTCAGTTATCA GAGGAGCAAACAGGTGCAAAGTCCACA	10

3x_HNF4_FOXA_v1_SecondaryS truct_min_v6_Aspacers (“A” spacer inbetween the repeats)	3x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v6	AGGAGGAGGCTGCTGGTAAACATTAACCAAGG TCACCCAGTTATCAGAGGAGCAAACAGGTGC AAAGTCCACAGGAGGAGGCTGCTGGTAAACAT TAACCAAGGTCACCCAGTTATCAGAGGAGCA AACAGGTGCAAAGTCCACAGGAGGAGGCTGCT GGTAAACATTAACCAAGGTCACCCAGTTATC AGAGGAGCAAACAGGTGCAAAGTCCACA	11
3x_Chinese TreeShrew	3x repeat of the Chinese Tree Shrew SERPINA1 enhancer (“C” spacer inbetween the repeats)	GGAGGCTGTTGGTGAATATTAACCAAGGTCAC CTCAGTTATCGGAGGAGCAAACAAGGGCTAAG TCCACCGGAGGCTGTTGGTGAATATTAACCA GGTCACCTCAGTTATCGGAGGAGCAAACAAGG GCTAAGTCCACCGGAGGCTGTTGGTGAATATT AACCAAGGTCACCTCAGTTATCGGAGGAGCAA ACAAGGGCTAAGTCCAC	12
3x_Chinese TreeShrew_ CpGmin	3x repeat of the Chinese Tree Shrew SERPINA1 enhancer with CpG minimization (no spacer)	AGGAGGCTGTTGGTGAATATTAACCAAGGTCA CCTCAGTTATCAGAGGAGCAAACAAGGGCTAA GTCCACAGGAGGCTGTTGGTGAATATTAACCA AGGTCACCTCAGTTATCAGAGGAGCAAACAAG GGCTAAGTCCACAGGAGGCTGTTGGTGAATAT TAACCAAGGTCACCTCAGTTATCAGAGGAGCA AACCAAGGGCTAAGTCCACA	13
3x_hSerpEn h_Aspacers	3x repeat of the human SERPINA1 enhancer with 1 adenine between repeats (“A” spacer)	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCAGTTATCGGAGGAGCAAACAGGGGCT AAGTCCACAGGGGAGGCTGCTGGTGAATATT AACCAAGGTCACCCAGTTATCGGAGGAGCAA ACAGGGGCTAAGTCCACAGGGGAGGCTGCTG GTGAATATTAACCAAGGTCACCCAGTTATCGG AGGAGCAAACAGGGGCTAAGTCCAC	14
3x_Bushba by_Aspacer s	3x repeat of the Bushbaby SERPINA1 enhancer with adenine nucleotide spacer (no spacer)	<u>AGGGGAAGCTACTGGTGAATATTAACCAAGGT</u> <u>CACCCAGTTATCAGGGAGCAAACAGGAGCTAA</u> <u>GTCCATAGGGGGAAGCTACTGGTGAATATTAACC</u> <u>AAGGTACCCAGTTATCAGGGAGCAAACAGGAGC</u> <u>TAAGTCCATAGGGGGAAGCTACTGGTGAATAT</u> <u>TAACCAAGGTCACCCAGTTATCAGGGAGCAA</u> <u>ACAGGAGCTAAGTCCAT</u>	15
5x_HNF4_FOXA_v1	5x repeat of HNF4_FOXA_v1 (“C” spacer)	GGGGGAGGCTGCTGGTAAACATTAACCAAGGT CACCCAGTTATCAGAGGAGCAAACAGGGGCA AAGTCCACCGGGGAGGCTGCTGGTAAACATT AACCAAGGTCACCCAGTTATCAGAGGAGCAA ACAGGGGCAAAGTCCACCGGGGAGGCTGCTG GTAAACATTAACCAAGGTCACCCAGTTATCA GAGGAGCAAACAGGGGCAAAGTCCACCGGGG GAGGCTGCTGGTAAACATTAACCAAGGTCACC CCAGTTATCAGAGGAGCAAACAGGGGCAAAGT CCACCGGGGAGGCTGCTGGTAAACATTAACC AAGGTCACCCAGTTATCAGAGGAGCAAACAG GGCAAAGTCCAC	16

5x_HNF4_FOXA_v1_SecondaryS truct_min_v1	5x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v1 ("C" spacer)	GAGGGAGGCTGCTGGTAAACATTAACCAAGGT CACCCAGTTATCAGAGGAGCAAACAGGGGCAA AGTCCACCGAGGGAGGCTGCTGGTAAACATTA ACCAAGGTCACCCAGTTATCAGAGGAGCAAAC AGGGGCAAAGTCCACCGAGGGAGGCTGCTGGT AAACATTAACCAAGGTCACCCAGTTATCAGAG GAGCAAACAGGGGCAAAGTCCACCGAGGGAG GCTGCTGGTAAACATTAACCAAGGTCACCCAG TTATCAGAGGAGCAAACAGGGGCAAAGTCCAC CGAGGGAGGCTGCTGGTAAACATTAACCAAGG TCACCCAGTTATCAGAGGAGCAAACAGGGGCA AAGTCCAC	17
5x_HNF4_FOXA_v1_SecondaryS truct_min_v1_CpG_min	5x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization and CpG minimization v1 ("AG" spacer)	AGGGAGGCTGCTGGTAAACATTAACCAAGGTC ACCCAGTTATCAGAGGAGCAAACAGGGGCAAA GTCCACAGAGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGAGCAAACA GGGGCAAAGTCCACAGAGGGAGGCTGCTGGTA AACATTAACCAAGGTCACCCAGTTATCAGAGG AGCAAACAGGGGCAAAGTCCACAGAGGGAGG CTGCTGGTAAACATTAACCAAGGTCACCCAGTT ATCAGAGGAGCAAACAGGGGCAAAGTCCACA GAGGGAGGCTGCTGGTAAACATTAACCAAGGT CACCCAGTTATCAGAGGAGCAAACAGGGGCAA AGTCCAT	18
5x_HNF4_FOXA_v1_SecondaryS truct_min_v2	5x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v2 ("C" spacer)	GGGGAGGCTGCTGGTAAACATTAACCAAGGT CACCTCAGTTATCAGAGGAGCAAACAGGGACA AAGTCCACCGGGGGAGGCTGCTGGTAAACATT AACCAAGGTCACCTCAGTTATCAGAGGAGCAA ACAGGGACAAAGTCCACCGGGGGAGGCTGCTG GTAAACATTAACCAAGGTCACCTCAGTTATCAG AGGAGCAAACAGGGACAAAGTCCACCGGGGG AGGCTGCTGGTAAACATTAACCAAGGTCACCT CAGTTATCAGAGGAGCAAACAGGGACAAAGTC CACCGGGGGAGGCTGCTGGTAAACATTAACCA AGGTCACCTCAGTTATCAGAGGAGCAAACAGG GACAAAGTCCAC	19
5x_HNF4_FOXA_v1_SecondaryS truct_min_v2_CpG_min	5x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization and CpG minimization v2 ("A" spacer)	AGGGAGGCTGCTGGTAAACATTAACCAAGGT CACCTCAGTTATCAGAGGAGCAAACAGGGACA AAGTCCACAGGGGGAGGCTGCTGGTAAACATT AACCAAGGTCACCTCAGTTATCAGAGGAGCAA ACAGGGACAAAGTCCACAGGGGGAGGCTGCTG GTAAACATTAACCAAGGTCACCTCAGTTATCAG AGGAGCAAACAGGGACAAAGTCCACAGGGGG AGGCTGCTGGTAAACATTAACCAAGGTCACCT CAGTTATCAGAGGAGCAAACAGGGACAAAGTC CACAGGGGGAGGCTGCTGGTAAACATTAACCA AGGTCACCTCAGTTATCAGAGGAGCAAACAGG GACAAAGTCCACA	20

5x_HNF4_FOXA_v1_SecondaryS truct_min_v3	5x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v3 ("C" spacer)	GGGAGGCTGCTGGTAAACATTAACCAAGGTCA CCCCAGTTATCAGAGGAGCAAACAAGGGCAA GTCCACCGGGAGGCTGCTGGTAAACATTAACC AAGGTCACCCCAGTTATCAGAGGAGCAAACAA GGGCAAAGTCCACCGGGAGGCTGCTGGTAAAC ATTAACCAAGGTACCCCCAGTTATCAGAGGAG CAAACAAGGGCAAAGTCCACCGGGAGGCTGCT GGTAAACATTAACCAAGGTACCCCCAGTTATC AGAGGAGCAAACAAGGGCAAAGTCCACCGGG AGGCTGCTGGTAAACATTAACCAAGGTACCCC CAGTTATCAGAGGAGCAAACAAGGGCAAAGTC CAC	21
5x_HNF4_FOXA_v1_SecondaryS truct_min_v3_CpG_min	5x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization and CpG minimization v3	AGGGAGGCTGCTGGTAAACATTAACCAAGGTC ACCCAGTTATCAGAGGAGCAAACAAGGGCAA AGTCCACAGGGAGGCTGCTGGTAAACATTAAC CAAGGTCACCCCAGTTATCAGAGGAGCAAACA AGGGCAAAGTCCACAGGGAGGCTGCTGGTAA CATTAACCAAGGTACCCCCAGTTATCAGAGGA GCAAACAAGGGCAAAGTCCACAGGGAGGCTGC TGGTAAACATTAACCAAGGTACCCCCAGTTATC AGAGGAGCAAACAAGGGCAAAGTCCACAGGG AGGCTGCTGGTAAACATTAACCAAGGTACCCC CAGTTATCAGAGGAGCAAACAAGGGCAAAGTC CACA	22
5x_HNF4_FOXA_v1_SecondaryS truct_min_v4_Aspacers	5x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v4	AGGAGGAGGCTGCTGGTAAACATTAACCAAGG TCACCTCAGTTATCAGAGGAGCAAACAGGGGC AAAGTCCACAGGAGGAGGCTGCTGGTAAACAT TAACCAAGGTACCTCAGTTATCAGAGGAGCA AACAGGGGCAAAGTCCACAGGAGGAGGCTGCT GGTAAACATTAACCAAGGTACCTCAGTTATCA GAGGAGCAAACAGGGGCAAAGTCCACAGGAG GAGGCTGCTGGTAAACATTAACCAAGGTCACC TCAGTTATCAGAGGAGCAAACAGGGGCAAAGT CCACAGGAGGAGGCTGCTGGTAAACATTAACC AAGGTCACCTCAGTTATCAGAGGAGCAAACAG GGCAAAGTCCACA	23
5x_HNF4_FOXA_v1_SecondaryS truct_min_v5_Aspacers	5x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v5	AGGGGAGGCTGCTGGTAAACATTAACCAAGG TCACCTCAGTTATCAGAGGAGCAAACAGGTGC AAAGTCCACAGGGGAGGCTGCTGGTAAACAT TAACCAAGGTACCTCAGTTATCAGAGGAGCA AACAGGTGCAAAGTCCACAGGGGAGGCTGCT GGTAAACATTAACCAAGGTACCTCAGTTATCA GAGGAGCAAACAGGTGCAAAGTCCACAGGGG GAGGCTGCTGGTAAACATTAACCAAGGTCACC TCAGTTATCAGAGGAGCAAACAGGTGCAAAGT CCACAGGGGAGGCTGCTGGTAAACATTAACC AAGGTCACCTCAGTTATCAGAGGAGCAAACAG GTCAAAGTCCACA	24

5x_HNF4_FOXA_v1_SecondaryStructure_min_v6_Aspacers	5x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v6	AGGAGGAGGCTGCTGGTAAACATTAACCAAGGT TCACCCAGTTATCAGAGGAGCAAACAGGTGC AAAGTCCACAGGAGGAGGCTGCTGGTAAACAT TAACCAAGGTCACCCCAGTTATCAGAGGAGCA AACAGGTGCAAAGTCCACAGGAGGAGGCTGCT GGTAAACATTAACCAAGGTCACCCCAGTTATC AGAGGAGCAAACAGGTGCAAAGTCCACAGGA GGAGGCTGCTGGTAAACATTAACCAAGGTCAC CCCAGTTATCAGAGGAGCAAACAGGTGCAAAG TCCACAGGAGGAGGCTGCTGGTAAACATTAAC CAAGGTCACCCCAGTTATCAGAGGAGCAAACA GGTGCAAAGTCCACA	25
5x_ChineseTreeShrew	5x repeat of the Chinese Tree Shrew SERPINA1 enhancer	GGAGGCTGTTGGTGAATATTAACCAAGGTCAC CTCAGTTATCGGAGGAGCAAACAAGGGCTAAG TCCACCGGAGGCTGTTGGTGAATATTAACCA GGTCACCTCAGTTATCGGAGGAGCAAACAAGG GCTAAGTCCACCGGAGGCTGTTGGTGAATATTA ACCAAGGTCACCTCAGTTATCGGAGGAGCAA CAAGGGCTAAGTCCACCGGAGGCTGTTGGTGA ATATTAACCAAGGTCACCTCAGTTATCGGAGG AGCAAACAAGGGCTAAGTCCACCGGAGGCTGT TGGTGAATATTAACCAAGGTCACCTCAGTTATC GGAGGAGCAAACAAGGGCTAAGTCCAC	26
5x_ChineseTreeShrew_CpGmin	5x repeat of the Chinese Tree Shrew SERPINA1 enhancer with CpG minimization	AGGAGGCTGTTGGTGAATATTAACCAAGGTCA CCTCAGTTATCAGAGGAGCAAACAAGGGCTAA GTCCACAGGAGGCTGTTGGTGAATATTAACCA AGGTCACCTCAGTTATCAGAGGAGCAAACAAG GGCTAAGTCCACAGGAGGCTGTTGGTGAATAT TAACCAAGGTCACCTCAGTTATCAGAGGAGCA AACAAGGGCTAAGTCCACAGGAGGCTGTTGGT GAATATTAACCAAGGTCACCTCAGTTATCAGA GGAGCAAACAAGGGCTAAGTCCACAGGAGGCT GTTGGTGAATATTAACCAAGGTCACCTCAGTTA TCAGAGGAGCAAACAAGGGCTAAGTCCACA	27
5x_Bushbaby_Aspacer s	5x repeat of the Bushbaby SERPINA1 enhancer with adenine nucleotide spacer	AGGGGAAGCTACTGGTGAATATTAACCAAGGT CACCCAGTTATCAGGGAGCAAACAGGAGCTAA GTCCATAGGGGGAAGCTACTGGTGAATATTAA CCAAGGTCACCCAGTTATCAGGGAGCAAACAG GAGCTAAGTCCATAGGGGGAAGCTACTGGTGA ATATTAACCAAGGTCACCCAGTTATCAGGGAG CAAACAGGAGCTAAGTCCATAGGGGGAAGCTA CTGGTGAATATTAACCAAGGTCACCCAGTTATC AGGGAGCAAACAGGAGCTAAGTCCATAGGGGG AAGCTACTGGTGAATATTAACCAAGGTCACCC AGTTATCAGGGAGCAAACAGGAGCTAAGTCCA T	28
5x_hSerpEnh	5x repeat of the human SERPINA1 enhancer	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGCT AAGTCCACCGGGGGAGGCTGCTGGTGAATATT AACCAAGGTCACCCCAGTTATCGGAGGAGCAA ACAGGGGCTAAGTCCACCGGGGGAGGCTGCTG GTGAATATTAACCAAGGTCACCCCAGTTATCGG AGGAGCAAACAGGGGCTAAGTCCACCGGGGGA GGCTGCTGGTGAATATTAACCAAGGTCACCCC	29

		AGTTATCGGAGGAGCAAACAGGGGCTAAGTCC ACCGGGGAGGCTGCTGGTGAATATTAACCAA GGTACCCCAGTTATCGGAGGAGCAAACAGGG GCTAAGTCCAC	
10x_HNF4 _FOXA_v1	10x repeat of HNF4_FOXA_v1	GGGGGAGGCTGCTGGTAAACATTAACCAAGGT CACCCCAGTTATCAGAGGAGCAAACAGGGGCA AAGTCCACCGGGGAGGCTGCTGGTAAACATT AACCAAGGTCACCCCAGTTATCAGAGGAGCAA ACAGGGGCAAAGTCCACCGGGGAGGCTGCTG GTAAACATTAACCAAGGTCACCCCAGTTATCA GAGGAGCAAACAGGGGCAAAGTCCACCGGGG GAGGCTGCTGGTAAACATTAACCAAGGTCACC CCAGTTATCAGAGGAGCAAACAGGGGCAAAGT CCACCGGGGAGGCTGCTGGTAAACATTAACC AAGGTCACCCCAGTTATCAGAGGAGCAAACAG GGCAAAGTCCACCGGGGAGGCTGCTGGTAA ACATTAACCAAGGTCACCCCAGTTATCAGAGG AGCAAACAGGGGCAAAGTCCACCGGGGAGG CTGCTGGTAAACATTAACCAAGGTCACCCCAGT TATCAGAGGAGCAAACAGGGGCAAAGTCCACC GGGGGAGGCTGCTGGTAAACATTAACCAAGGT CACCCCAGTTATCAGAGGAGCAAACAGGGGCA AAGTCCACCGGGGAGGCTGCTGGTAAACATT AACCAAGGTCACCCCAGTTATCAGAGGAGCAA ACAGGGGCAAAGTCCACCGGGGAGGCTGCTG GTAAACATTAACCAAGGTCACCCCAGTTATCA GAGGAGCAAACAGGGGCAAAGTCCAC	30
10x_HNF4 _FOXA_v1 _Secondary Struct_min_ v1	10x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v1	GAGGGAGGCTGCTGGTAAACATTAACCAAGGT CACCCAGTTATCAGAGGAGCAAACAGGGGCAA AGTCCACCGAGGGAGGCTGCTGGTAAACATTA ACCAAGGTCACCCAGTTATCAGAGGAGCAAAC AGGGGCAAAGTCCACCGAGGGAGGCTGCTGGT AAACATTAACCAAGGTCACCCAGTTATCAGAG GAGCAAACAGGGGCAAAGTCCACCGAGGGAG GCTGCTGGTAAACATTAACCAAGGTCACCCAG TTATCAGAGGAGCAAACAGGGGCAAAGTCCAC CGAGGGAGGCTGCTGGTAAACATTAACCAAGG TCACCCAGTTATCAGAGGAGCAAACAGGGGCA AAGTCCACCGAGGGAGGCTGCTGGTAAACATT AACCAAGGTCACCCAGTTATCAGAGGAGCAAA CAGGGGCAAAGTCCACCGAGGGAGGCTGCTGG TAAACATTAACCAAGGTCACCCAGTTATCAGA GGAGCAAACAGGGGCAAAGTCCACCGAGGGA GGCTGCTGGTAAACATTAACCAAGGTCACCCA GTTATCAGAGGAGCAAACAGGGGCAAAGTCCA CCGAGGGAGGCTGCTGGTAAACATTAACCAAG GTCACCCAGTTATCAGAGGAGCAAACAGGGGC AAAGTCCACCGAGGGAGGCTGCTGGTAAACAT TAACCAAGGTCACCCAGTTATCAGAGGAGCAA ACAGGGGCAAAGTCCAC	31
10x_HNF4 _FOXA_v1 _Secondary	10x repeat of HNF4_FOXA_v1 with poly-C/poly-G	AGGGAGGCTGCTGGTAAACATTAACCAAGGTC ACCCAGTTATCAGAGGAGCAAACAGGGGCAAA GTCCACAGAGGGAGGCTGCTGGTAAACATTAA	

Struct_min_v1_CpG_min	minimization and CpG minimization v1	CCAAGGTCACCCAGTTATCAGAGGAGCAAACA GGGGCAAAGTCCACAGAGGGAGGCTGCTGGTA AACATTAACCAAGGTCACCCAGTTATCAGAGG AGCAAACAGGGGCAAAGTCCACAGAGGGAGG CTGCTGGTAAACATTAACCAAGGTCACCCAGTT ATCAGAGGAGCAAACAGGGGCAAAGTCCACAG AGGGAGGCTGCTGGTAAACATTAACCAAGGTC ACCCAGTTATCAGAGGAGCAAACAGGGGCAA GTCCACAGAGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGAGCAAACA GGGGCAAAGTCCACAGAGGGAGGCTGCTGGTA AACATTAACCAAGGTCACCCAGTTATCAGAGG AGCAAACAGGGGCAAAGTCCACAGAGGGAGG CTGCTGGTAAACATTAACCAAGGTCACCCAGTT ATCAGAGGAGCAAACAGGGGCAAAGTCCACAG AGGGAGGCTGCTGGTAAACATTAACCAAGGTC ACCCAGTTATCAGAGGAGCAAACAGGGGCAA GTCCACAGAGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGAGCAAACA GGGGCAAAGTCCAT	32
10x_HNF4_FOXA_v1_Secondary Struct_min_v2	10x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v2	GGGGGAGGCTGCTGGTAAACATTAACCAAGGT CACCTCAGTTATCAGAGGAGCAAACAGGGACA AAGTCCACCGGGGAGGCTGCTGGTAAACATT AACCAAGGTCACCTCAGTTATCAGAGGAGCAA ACAGGGACAAAGTCCACCGGGGAGGCTGCTG GTAAACATTAACCAAGGTCACCTCAGTTATCAG AGGAGCAAACAGGGACAAAGTCCACCGGGG AGGCTGCTGGTAAACATTAACCAAGGTCACCT CAGTTATCAGAGGAGCAAACAGGGACAAAGTC CACCGGGGAGGCTGCTGGTAAACATTAACCA AGGTCACCTCAGTTATCAGAGGAGCAAACAGG GACAAAGTCCACCGGGGAGGCTGCTGGTAA CATTAACCAAGGTCACCTCAGTTATCAGAGGA GCAAACAGGGACAAAGTCCACCGGGGAGGCT GCTGGTAAACATTAACCAAGGTCACCTCAGTTA TCAGAGGAGCAAACAGGGACAAAGTCCACCGG GGGAGGCTGCTGGTAAACATTAACCAAGGTCA CCTCAGTTATCAGAGGAGCAAACAGGGACAAA GTCCACCGGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCTCAGTTATCAGAGGAGCAAAC AGGGACAAAGTCCACCGGGGAGGCTGCTGGT AAACATTAACCAAGGTCACCTCAGTTATCAGA GGAGCAAACAGGGACAAAGTCCAC	33
10x_HNF4_FOXA_v1_Secondary Struct_min_v2_CpG_min	10x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization and CpG minimization v2	AGGGGAGGCTGCTGGTAAACATTAACCAAGGT CACCTCAGTTATCAGAGGAGCAAACAGGGACA AAGTCCACAGGGGAGGCTGCTGGTAAACATT AACCAAGGTCACCTCAGTTATCAGAGGAGCAA ACAGGGACAAAGTCCACAGGGGAGGCTGCTG GTAAACATTAACCAAGGTCACCTCAGTTATCAG AGGAGCAAACAGGGACAAAGTCCACAGGGG AGGCTGCTGGTAAACATTAACCAAGGTCACCT CAGTTATCAGAGGAGCAAACAGGGACAAAGTC CACAGGGGAGGCTGCTGGTAAACATTAACCA AGGTCACCTCAGTTATCAGAGGAGCAAACAGG GACAAAGTCCACAGGGGAGGCTGCTGGTAA	34

		CATTAACCAAGGTCACCTCAGTTATCAGAGGA GCAAACAGGGACAAAGTCCACAGGGGGAGGCT GCTGGTAAACATTAACCAAGGTCACCTCAGTTA TCAGAGGAGCAAACAGGGACAAAGTCCACAGG GGGAGGCTGCTGGTAAACATTAACCAAGGTCA CCTCAGTTATCAGAGGAGCAAACAGGGACAAA GTCCACAGGGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCTCAGTTATCAGAGGAGCAAAC AGGGACAAAGTCCACAGGGGGAGGCTGCTGGT AAACATTAACCAAGGTCACCTCAGTTATCAGA GGAGCAAACAGGGACAAAGTCCACA	
10x_HNF4 _FOXA_v1 _Secondary Struct_min_ v3	10x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v3	GGGAGGCTGCTGGTAAACATTAACCAAGGTCA CCCCAGTTATCAGAGGAGCAAACAAGGGCAAA GTCCACCGGGAGGCTGCTGGTAAACATTAACC AAGGTCACCCAGTTATCAGAGGAGCAAACA GGGCAAAGTCCACCGGGAGGCTGCTGGTAAAC ATTAACCAAGGTCACCCAGTTATCAGAGGAG CAAACAAGGGCAAAGTCCACCGGGAGGCTGCT GGTAAACATTAACCAAGGTCACCCAGTTATC AGAGGAGCAAACAAGGGCAAAGTCCACCGGG AGGCTGCTGGTAAACATTAACCAAGGTCACCC CAGTTATCAGAGGAGCAAACAAGGGCAAAGTC CACCGGGAGGCTGCTGGTAAACATTAACCAAG GTCACCCAGTTATCAGAGGAGCAAACAAGGG CAAAGTCCACCGGGAGGCTGCTGGTAAACATT AACCAAGGTCACCCAGTTATCAGAGGAGCAA ACAAGGGCAAAGTCCACCGGGAGGCTGCTGGT AAACATTAACCAAGGTCACCCAGTTATCAGA GGAGCAAACAAGGGCAAAGTCCACCGGGAGG CTGCTGGTAAACATTAACCAAGGTCACCCAGT TATCAGAGGAGCAAACAAGGGCAAAGTCCACC GGGAGGCTGCTGGTAAACATTAACCAAGGTCA CCCCAGTTATCAGAGGAGCAAACAAGGGCAAA GTCCAC	35
10x_HNF4 _FOXA_v1 _Secondary Struct_min_ v3_CpG_min	10x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization and CpG minimization v3	AGGGAGGCTGCTGGTAAACATTAACCAAGGTC ACCCAGTTATCAGAGGAGCAAACAAGGGCAA AGTCCACAGGGAGGCTGCTGGTAAACATTAAC CAAGGTCACCCAGTTATCAGAGGAGCAAACA AGGGCAAAGTCCACAGGGAGGCTGCTGGTAA CATTAACCAAGGTCACCCAGTTATCAGAGGA GCAAACAAGGGCAAAGTCCACAGGGAGGCTGC TGGTAAACATTAACCAAGGTCACCCAGTTATC AGAGGAGCAAACAAGGGCAAAGTCCACAGGG AGGCTGCTGGTAAACATTAACCAAGGTCACCC CAGTTATCAGAGGAGCAAACAAGGGCAAAGTC CACAGGGAGGCTGCTGGTAAACATTAACCAAG GTCACCCAGTTATCAGAGGAGCAAACAAGGG CAAAGTCCACAGGGAGGCTGCTGGTAAACATT AACCAAGGTCACCCAGTTATCAGAGGAGCAA ACAAGGGCAAAGTCCACAGGGAGGCTGCTGGT AAACATTAACCAAGGTCACCCAGTTATCAGA GGAGCAAACAAGGGCAAAGTCCACAGGGAGG CTGCTGGTAAACATTAACCAAGGTCACCCAGT TATCAGAGGAGCAAACAAGGGCAAAGTCCACA GGGAGGCTGCTGGTAAACATTAACCAAGGTCA	36

		CCCCAGTTATCAGAGGAGCAAACAAGGGCAAA GTCCACA	
10x_hSerpE nh	10x repeat of the human SERPINA1 enhancer ("C" spacer)	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGGCT AAGTCCACCGGGGGAGGCTGCTGGTGAATATT AACCAAGGTCACCCCAGTTATCGGAGGAGCAA ACAGGGGGCTAAGTCCACCGGGGGAGGCTGCTG GTGAATATTAACCAAGGTCACCCCAGTTATCGG AGGAGCAAACAGGGGGCTAAGTCCACCGGGGG AGGCTGCTGGTGAATATTAACCAAGGTCACCC CAGTTATCGGAGGAGCAAACAGGGGGCTAAGTC CACCGGGGGAGGCTGCTGGTGAATATTAACCA AGGTCACCCCAGTTATCGGAGGAGCAAACAGG GGCTAAGTCCACCGGGGGAGGCTGCTGGTGAA TATTAACCAAGGTCACCCCAGTTATCGGAGGA GCAAACAGGGGGCTAAGTCCACCGGGGGAGGCT GCTGGTGAATATTAACCAAGGTCACCCCAGTTA TCGGAGGAGCAAACAGGGGGCTAAGTCCACCGG GGGAGGCTGCTGGTGAATATTAACCAAGGTCA CCCCAGTTATCGGAGGAGCAAACAGGGGGCTAA GTCCACCGGGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAGTTATCGGAGGAGCAAAC AGGGGGCTAAGTCCACCGGGGGAGGCTGCTGGT GAATATTAACCAAGGTCACCCCAGTTATCGGA GGAGCAAACAGGGGGCTAAGTCCAC	37
10x_Bushb aby_Aspace rs	10x repeat of the Bushbaby SERPINA1 enhancer with <i>adenenine</i> <i>nucleotide spacer</i>	AGGGGAAGCTACTGGTGAATATTAACCAAGGT CACCCAGTTATCAGGGAGCAAACAGGAGCTAA GTCCATAGGGGGAAGCTACTGGTGAATATTAA CCAAGGTCACCCAGTTATCAGGGAGCAAACAG GAGCTAAGTCCATAGGGGGAAGCTACTGGTGA ATATTAACCAAGGTCACCCAGTTATCAGGGAG CAAACAGGAGCTAAGTCCATAGGGGGAAGCTA CTGGTGAATATTAACCAAGGTCACCCAGTTATC AGGGAGCAAACAGGAGCTAAGTCCATAGGGGG AAGCTACTGGTGAATATTAACCAAGGTCACCC AGTTATCAGGGAGCAAACAGGAGCTAAGTCCA TAGGGGGAAGCTACTGGTGAATATTAACCAAG GTCACCCAGTTATCAGGGAGCAAACAGGAGCT AAGTCCATAGGGGGAAGCTACTGGTGAATATT AACCAAGGTCACCCAGTTATCAGGGAGCAAAC AGGAGCTAAGTCCATAGGGGGAAGCTACTGGT GAATATTAACCAAGGTCACCCAGTTATCAGGG AGCAAACAGGAGCTAAGTCCATAGGGGGAAGC TACTGGTGAATATTAACCAAGGTCACCCAGTTA TCAGGGAGCAAACAGGAGCTAAGTCCATAGGG GGAAGCTACTGGTGAATATTAACCAAGGTCAC	38

		CCAGTTATCAGGGAGCAAACAGGAGCTAAGTC CAT	
Bushbaby_ HN4F/FOX v1_HNF4m od	Bushbaby SERPINA1 enhancer, FOXA_HNF4_v1 enhancer, HNF4 consensus binding site enhancer	GGGGGAAGCTACTGGTGAATATTAACCAAGGT CACCCAGTTATCAGGGAGCAAACAGGAGCTAA GTCCATAGGGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGAGCAAAC AGGGGCAAAGTCCACAAGGGGAGGCTGCTGGT GAATATTAACCAAGGTCACCTCAGTTATCAGA GGAGCAAACAGGGGCAAAGTCCAT	39
HNF4mod_ BushbabyM od_HN4F/F OXv1	HNF4 consensus binding site enhancer, Bushbaby SERPINA1 enhancer, FOXA_HNF4_v1 enhancer	AGAGGGAGGCTGCTGGTGAATATTAACCAAGG TCACCTCAGTTATCAGAGGAGCAAACAGGGGC AAAGTCCATAGAGGGAAGCTACTGGTGAATAT TAACCAAGGTCACCCAGTTATCAGGGAGCAAA CAGGAGCTAAGTCCATAGGGGGAGGCTGCTGG TAAACATTAACCAAGGTCACCCAGTTATCAG AGGAGCAAACAGGGGCAAAGTCCAC	40
3x_hSerpEn h_2mer_spa cers_v1	3x repeat of hSerpEnh with 2mer spacers v1 (bold underlined)	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCAGTTATCGGAGGAGCAAACAGGGGCT AAGTCCACAAGGGGGAGGCTGCTGGTGAATAT TAACCAAGGTCACCCAGTTATCGGAGGAGCA AACAGGGGCTAAGTCCACAGGGGGAGGCTGC TGGTGAATATTAACCAAGGTCACCCAGTTATC GGAGGAGCAAACAGGGGCTAAGTCCAC	41
3x_hSerpEn h_2mer_spa cers_v2	3x repeat of hSerpEnh with 2mer spacers v2 (bold underlined)	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCAGTTATCGGAGGAGCAAACAGGGGCT AAGTCCACAAGGGGGAGGCTGCTGGTGAATAT TAACCAAGGTCACCCAGTTATCGGAGGAGCA AACAGGGGCTAAGTCCACCTGGGGGAGGCTGC TGGTGAATATTAACCAAGGTCACCCAGTTATC GGAGGAGCAAACAGGGGCTAAGTCCAC	42
3x_hSerpEn h_2mer_spa cers_v3	3x repeat of hSerpEnh with 2mer spacers v3 (bold underlined)	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCAGTTATCGGAGGAGCAAACAGGGGCT AAGTCCACAAGGGGGAGGCTGCTGGTGAATAT TAACCAAGGTCACCCAGTTATCGGAGGAGCA AACAGGGGCTAAGTCCACTAGGGGGAGGCTGC TGGTGAATATTAACCAAGGTCACCCAGTTATC GGAGGAGCAAACAGGGGCTAAGTCCAC	43
3x_hSerpEn h_2mer_spa cers_v4	3x repeat of hSerpEnh with 2mer spacers v4 (bold underlined)	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCAGTTATCGGAGGAGCAAACAGGGGCT AAGTCCACAAGGGGGAGGCTGCTGGTGAATAT TAACCAAGGTCACCCAGTTATCGGAGGAGCA AACAGGGGCTAAGTCCACTTGGGGGAGGCTGC	44

		TGGTGAATATTAACCAAGGTCACCCCAGTTATC GGAGGAGCAAACAGGGGCTAAGTCCAC	
3x_hSerpEn h_2mer_spa cers_v5	3x repeat of hSerpEnh with 2mer spacers v5 (bold underlined)	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGCT AAGTCCAC <u>CA</u> GGGGGAGGCTGCTGGTGAATAT TAACCAAGGTCACCCCAGTTATCGGAGGAGCA AACAGGGGCTAAGTCCAC <u>AA</u> GGGGGAGGCTGC TGGTGAATATTAACCAAGGTCACCCCAGTTATC GGAGGAGCAAACAGGGGCTAAGTCCAC	45
3x_hSerpEn h_2mer_spa cers_v6	3x repeat of hSerpEnh with 2mer spacers v6 (bold underlined)	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGCT AAGTCCAC <u>CA</u> GGGGGAGGCTGCTGGTGAATAT TAACCAAGGTCACCCCAGTTATCGGAGGAGCA AACAGGGGCTAAGTCCAC <u>CT</u> GGGGGAGGCTGC TGGTGAATATTAACCAAGGTCACCCCAGTTATC GGAGGAGCAAACAGGGGCTAAGTCCAC	46
3x_hSerpEn h_2mer_spa cers_v7	3x repeat of hSerpEnh with 2mer spacers v7 (bold underlined)	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGCT AAGTCCAC <u>CA</u> GGGGGAGGCTGCTGGTGAATAT TAACCAAGGTCACCCCAGTTATCGGAGGAGCA AACAGGGGCTAAGTCCAC <u>TA</u> GGGGGAGGCTGC TGGTGAATATTAACCAAGGTCACCCCAGTTATC GGAGGAGCAAACAGGGGCTAAGTCCAC	47
3x_hSerpEn h_2mer_spa cers_v8	3x repeat of hSerpEnh with 2mer spacers v8 (bold underlined)	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGCT AAGTCCAC <u>CA</u> GGGGGAGGCTGCTGGTGAATAT TAACCAAGGTCACCCCAGTTATCGGAGGAGCA AACAGGGGCTAAGTCCAC <u>TT</u> GGGGGAGGCTGC TGGTGAATATTAACCAAGGTCACCCCAGTTATC GGAGGAGCAAACAGGGGCTAAGTCCAC	48
3x_hSerpEn h_2mer_spa cers_v9	3x repeat of hSerpEnh with 2mer spacers v9 (bold underlined)	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGCT AAGTCCAC <u>CT</u> GGGGGAGGCTGCTGGTGAATAT TAACCAAGGTCACCCCAGTTATCGGAGGAGCA AACAGGGGCTAAGTCCAC <u>AA</u> GGGGGAGGCTGC TGGTGAATATTAACCAAGGTCACCCCAGTTATC GGAGGAGCAAACAGGGGCTAAGTCCAC	49
3x_hSerpEn h_2mer_spa cers_v10	3x repeat of hSerpEnh with 2mer spacers v10 (bold underlined)	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGCT AAGTCCAC <u>CT</u> GGGGGAGGCTGCTGGTGAATAT TAACCAAGGTCACCCCAGTTATCGGAGGAGCA AACAGGGGCTAAGTCCAC <u>CA</u> GGGGGAGGCTGC TGGTGAATATTAACCAAGGTCACCCCAGTTATC GGAGGAGCAAACAGGGGCTAAGTCCAC	50
3x_hSerpEn h_2mer_spa cers_v11	3x repeat of hSerpEnh with 2mer spacers v11 (bold underlined)	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGCT AAGTCCAC <u>CT</u> GGGGGAGGCTGCTGGTGAATAT TAACCAAGGTCACCCCAGTTATCGGAGGAGCA AACAGGGGCTAAGTCCAC <u>TA</u> GGGGGAGGCTGC	51

		TGGTGAATATTAACCAAGGTCACCCCAGTTATC GGAGGAGCAAACAGGGGCTAAGTCCAC	
3x_hSerpEn h_2mer_spa cers_v12	3x repeat of hSerpEnh with 2mer spacers v12 (bold underlined)	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGCT AAGTCCAC <u>CT</u> GGGGGAGGCTGCTGGTGAATAT TAACCAAGGTCACCCCAGTTATCGGAGGAGCA AACAGGGGCTAAGTCCAC <u>TT</u> GGGGGAGGCTGC TGGTGAATATTAACCAAGGTCACCCCAGTTATC GGAGGAGCAAACAGGGGCTAAGTCCAC	52
3x_hSerpEn h_2mer_spa cers_v13	3x repeat of hSerpEnh with 2mer spacers v13 (bold underlined)	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGCT AAGTCCAC <u>TA</u> GGGGGAGGCTGCTGGTGAATAT TAACCAAGGTCACCCCAGTTATCGGAGGAGCA AACAGGGGCTAAGTCCAC <u>AA</u> GGGGGAGGCTGC TGGTGAATATTAACCAAGGTCACCCCAGTTATC GGAGGAGCAAACAGGGGCTAAGTCCAC	53
3x_hSerpEn h_2mer_spa cers_v14	3x repeat of hSerpEnh with 2mer spacers v14 (bold underlined)	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGCT AAGTCCAC <u>TA</u> GGGGGAGGCTGCTGGTGAATAT TAACCAAGGTCACCCCAGTTATCGGAGGAGCA AACAGGGGCTAAGTCCAC <u>CA</u> GGGGGAGGCTGC TGGTGAATATTAACCAAGGTCACCCCAGTTATC GGAGGAGCAAACAGGGGCTAAGTCCAC	54
3x_hSerpEn h_2mer_spa cers_v15	3x repeat of hSerpEnh with 2mer spacers v15 (bold underlined)	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGCT AAGTCCAC <u>TA</u> GGGGGAGGCTGCTGGTGAATAT TAACCAAGGTCACCCCAGTTATCGGAGGAGCA AACAGGGGCTAAGTCCAC <u>CT</u> GGGGGAGGCTGC TGGTGAATATTAACCAAGGTCACCCCAGTTATC GGAGGAGCAAACAGGGGCTAAGTCCAC	55
3x_hSerpEn h_2mer_spa cers_v16	3x repeat of hSerpEnh with 2mer spacers v16 (bold underlined)	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGCT AAGTCCAC <u>TA</u> GGGGGAGGCTGCTGGTGAATAT TAACCAAGGTCACCCCAGTTATCGGAGGAGCA AACAGGGGCTAAGTCCAC <u>TT</u> GGGGGAGGCTGC TGGTGAATATTAACCAAGGTCACCCCAGTTATC GGAGGAGCAAACAGGGGCTAAGTCCAC	56
3x_hSerpEn h_2mer_spa cers_v17	3x repeat of hSerpEnh with 2mer spacers v17 (bold underlined)	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGCT AAGTCCAC <u>TT</u> GGGGGAGGCTGCTGGTGAATAT TAACCAAGGTCACCCCAGTTATCGGAGGAGCA AACAGGGGCTAAGTCCAC <u>AA</u> GGGGGAGGCTGC TGGTGAATATTAACCAAGGTCACCCCAGTTATC GGAGGAGCAAACAGGGGCTAAGTCCAC	57
3x_hSerpEn h_2mer_spa cers_v18	3x repeat of hSerpEnh with 2mer spacers v18 (bold underlined)	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGCT AAGTCCAC <u>TT</u> GGGGGAGGCTGCTGGTGAATAT TAACCAAGGTCACCCCAGTTATCGGAGGAGCA AACAGGGGCTAAGTCCAC <u>CA</u> GGGGGAGGCTGC	58

		TGGTGAATATTAACCAAGGTCACCCCAGTTATC GGAGGAGCAAACAGGGGCTAAGTCCAC	
3x_hSerpEn h_2mer_spa cers_v19	3x repeat of hSerpEnh with 2mer spacers v19 (bold underlined)	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGCT AAGTCCAC <u>TT</u> GGGGGAGGCTGCTGGTGAATAT TAACCAAGGTCACCCCAGTTATCGGAGGAGCA AACAGGGGCTAAGTCCAC <u>CT</u> GGGGGAGGCTGC TGGTGAATATTAACCAAGGTCACCCCAGTTATC GGAGGAGCAAACAGGGGCTAAGTCCAC	59
3x_hSerpEn h_2mer_spa cers_v20	3x repeat of hSerpEnh with 2mer spacers v20 (bold underlined)	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGCT AAGTCCAC <u>TT</u> GGGGGAGGCTGCTGGTGAATAT TAACCAAGGTCACCCCAGTTATCGGAGGAGCA AACAGGGGCTAAGTCCAC <u>TA</u> GGGGGAGGCTGC TGGTGAATATTAACCAAGGTCACCCCAGTTATC GGAGGAGCAAACAGGGGCTAAGTCCAC	60
3x_hSerpEn h_3mer_spa cers_v1	3x repeat of hSerpEnh with 3mer spacers v1 (bold underlined)	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGCT AAGTCCACTTAGGGGGAGGCTGCTGGTGAATA TTAACCAAGGTCACCCCAGTTATCGGAGGAGC AAACAGGGGCTAAGTCCACTGTGGGGAGGCT GCTGGTGAATATTAACCAAGGTCACCCCAGTTA TCGGAGGAGCAAACAGGGGCTAAGTCCAC	61
3x_hSerpEn h_3mer_spa cers_v2	3x repeat of hSerpEnh with 3mer spacers v2 (bold underlined)	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGCT AAGTCCAC <u>AGA</u> GGGGGAGGCTGCTGGTGAATA TTAACCAAGGTCACCCCAGTTATCGGAGGAGC AAACAGGGGCTAAGTCCACT <u>TGA</u> GGGGGAGGCT GCTGGTGAATATTAACCAAGGTCACCCCAGTTA TCGGAGGAGCAAACAGGGGCTAAGTCCAC	62
3x_hSerpEn h_3mer_spa cers_v3	3x repeat of hSerpEnh with 3mer spacers v3 (bold underlined)	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGCT AAGTCCAC <u>ACT</u> GGGGGAGGCTGCTGGTGAATA TTAACCAAGGTCACCCCAGTTATCGGAGGAGC AAACAGGGGCTAAGTCCAC <u>CAA</u> GGGGGAGGCT GCTGGTGAATATTAACCAAGGTCACCCCAGTTA TCGGAGGAGCAAACAGGGGCTAAGTCCAC	63
3x_hSerpEn h_5mer_spa cers_v1	3x repeat of hSerpEnh with 5mer spacers v1 (bold underlined)	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGCT AAGTCCAC <u>ACATA</u> GGGGGAGGCTGCTGGTGAA TATTAACCAAGGTCACCCCAGTTATCGGAGGA GCAAACAGGGGCTAAGTCCAC <u>CTGTAG</u> GGGGGA GGCTGCTGGTGAATATTAACCAAGGTCACCCC AGTTATCGGAGGAGCAAACAGGGGCTAAGTCC AC	64
3x_hSerpEn h_5mer_spa cers_v2	3x repeat of hSerpEnh with 5mer spacers v2 (bold underlined)	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGCT AAGTCCAC <u>AACA</u> GGGGGAGGCTGCTGGTGAA TATTAACCAAGGTCACCCCAGTTATCGGAGGA GCAAACAGGGGCTAAGTCCAC <u>CATCA</u> GGGGGA	

		GGCTGCTGGTGAATATTAACCAAGGTCACCCC AGTTATCGGAGGAGCAAACAGGGGCTAAGTCC AC	65
3x_hSerpEn h_5mer_spa cers_v3	3x repeat of hSerpEnh with 5mer spacers v3 (bold underlined)	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGCT AAGTCCAC <u>CAATT</u> GGGGGAGGCTGCTGGTGAA TATTAACCAAGGTCACCCCAGTTATCGGAGGA GCAAACAGGGGCTAAGTCCAC <u>TTGCT</u> GGGGGA GGCTGCTGGTGAATATTAACCAAGGTCACCCC AGTTATCGGAGGAGCAAACAGGGGCTAAGTCC AC	66
3x_hSerpEn h_11mer_sp acers_v1	3x repeat of hSerpEnh with 11mer spacers v1 (bold underlined)	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGCT AAGTCCAC <u>CCTTGGGACCA</u> GGGGGAGGCTGC TGGTGAATATTAACCAAGGTCACCCCAGTTATC GGAGGAGCAAACAGGGGCTAAGTCCACA <u>AAGC</u> <u>TGTTCCA</u> GGGGGAGGCTGCTGGTGAATATTA CCAAGGTCACCCCAGTTATCGGAGGAGCAAAC AGGGGCTAAGTCCAC	67
3x_hSerpEn h_11mer_sp acers_v2	3x repeat of hSerpEnh with 11mer spacers v2 (bold underlined)	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGCT AAGTCCAC <u>AGGCTGGTTGA</u> GGGGGAGGCTGC TGGTGAATATTAACCAAGGTCACCCCAGTTATC GGAGGAGCAAACAGGGGCTAAGTCCAC <u>TGATA</u> <u>ATAGCT</u> GGGGGAGGCTGCTGGTGAATATTAAC CAAGGTCACCCCAGTTATCGGAGGAGCAAACA GGGGCTAAGTCCAC	68
3x_hSerpEn h_11mer_sp acers_v3	3x repeat of hSerpEnh with 11mer spacers v3 (bold underlined)	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGCT AAGTCCAC <u>CATTCTGCTTT</u> GGGGGAGGCTGCT GGTGAATATTAACCAAGGTCACCCCAGTTATCG GAGGAGCAAACAGGGGCTAAGTCCAC <u>TTGATT</u> <u>AAGAA</u> GGGGGAGGCTGCTGGTGAATATTAACC AAGGTCACCCCAGTTATCGGAGGAGCAAACAG GGGCTAAGTCCAC	69
3x_hSerpEn h_11mer_sp acers_HNF 4former_sp acers_FOX Afor	3x repeat of hSerpEnh with 11mer spacers (bold underlined) with HNF4 binding site in orientation 1 & FOXA binding site in orientation 1	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGCT AAGTCCAC <u>AACAAAGTCCA</u> GGGGGAGGCTGCT GGTGAATATTAACCAAGGTCACCCCAGTTATCG GAGGAGCAAACAGGGGCTAAGTCCAC <u>CTTGTA</u> <u>AACA</u> GGGGGAGGCTGCTGGTGAATATTAACC AAGGTCACCCCAGTTATCGGAGGAGCAAACAG GGGCTAAGTCCAC	70
3x_hSerpEn h_11mer_sp acers_HNF 4former_sp acers_FOX Arev	3x repeat of hSerpEnh with 11mer spacers (bold underlined) with HNF4 binding site in orientation 1 & FOXA binding site in orientation 2	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGCT AAGTCCAC <u>TGCAAAGTCCT</u> GGGGGAGGCTGCT GGTGAATATTAACCAAGGTCACCCCAGTTATCG GAGGAGCAAACAGGGGCTAAGTCCAC <u>AGTGTT</u> <u>TACA</u> GGGGGAGGCTGCTGGTGAATATTAACC AAGGTCACCCCAGTTATCGGAGGAGCAAACAG GGGCTAAGTCCAC	71

3x_hSerpEnh_11mer_spacers_HNF4revmer_spacers_FOX Afor	3x repeat of hSerpEnh with 11mer spacers (bold underlined) with HNF4 binding site in orientation 2 & FOXA binding site in orientation 1	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGGCT AAGTCCAC <u>AGGACTTTGAA</u> GGGGGAGGCTGCT GGTGAATATTAACCAAGGTCACCCCAGTTATCG GAGGAGCAAACAGGGGGCTAAGTCCAC <u>AGTGT</u> <u>AAACA</u> AGGGGAGGCTGCTGGTGAATATTAAC CAAGGTCACCCCAGTTATCGGAGGAGCAAACA GGGGCTAAGTCCAC	72
3x_hSerpEnh_11mer_spacers_HNF4revmer_spacers_FOX Arev	3x repeat of hSerpEnh with 11mer spacers (bold underlined) with HNF4 binding site in orientation 2 & FOXA binding site in orientation 2	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGGCT AAGTCCAC <u>TGGACTTTGGT</u> GGGGGAGGCTGCT GGTGAATATTAACCAAGGTCACCCCAGTTATCG GAGGAGCAAACAGGGGGCTAAGTCCAC <u>TCTGTT</u> <u>TACA</u> AGGGGAGGCTGCTGGTGAATATTAACC AAGGTCACCCCAGTTATCGGAGGAGCAAACAG GGGCTAAGTCCAC	73
3x_hSerpEnh_30mer_spacers_v1	3x repeat of hSerpEnh with 30mer spacers v1 (bold underlined)	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGGCT AAGTCCAC <u>CTGCTTGACATCTGCAGTAATCT</u> <u>TTGATT</u> AGGGGAGGCTGCTGGTGAATATTAAC CCAAGGTCACCCCAGTTATCGGAGGAGCAAAC AGGGGGCTAAGTCCAC <u>CTCTGATACTTTGAT</u> <u>CTAGTCTACTGCT</u> GGGGGAGGCTGCTGGTGA TATTAACCAAGGTCACCCCAGTTATCGGAGGA GCAAACAGGGGGCTAAGTCCAC	74
3x_hSerpEnh_30mer_spacers_v2	3x repeat of hSerpEnh with 30mer spacers v2 (bold underlined)	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGGCT AAGTCCAC <u>CACTTGATTTAATCATACTACT</u> <u>TAGCA</u> AGGGGAGGCTGCTGGTGAATATTAAC CAAGGTCACCCCAGTTATCGGAGGAGCAAACA GGGGCTAAGTCCAC <u>TAACATCTTACAACTAA</u> <u>AGTTAGATAGTA</u> AGGGGAGGCTGCTGGTGAAT ATTAACCAAGGTCACCCCAGTTATCGGAGGAG CAAACAGGGGGCTAAGTCCAC	75
3x_hSerpEnh_30mer_spacers_v3	3x repeat of hSerpEnh with 30mer spacers v3 (bold underlined)	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGGCT AAGTCCAC <u>ATAGAAGAATTTCTTACATTGTGT</u> <u>GAACCT</u> GGGGGAGGCTGCTGGTGAATATTAAC CAAGGTCACCCCAGTTATCGGAGGAGCAAACA GGGGCTAAGTCCAC <u>ATTGAAGTGCAAAGTCA</u> <u>CTAATATTAAGCA</u> AGGGGAGGCTGCTGGTGA TATTAACCAAGGTCACCCCAGTTATCGGAGGA GCAAACAGGGGGCTAAGTCCAC	76
3x_hSerpEnh_30mer_spacers_HNF4former_spacers_FOX Afor	3x repeat of hSerpEnh with 30mer spacers (bold underlined) with HNF4 binding site in orientation 1 & FOXA binding site in orientation 1	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGGCT AAGTCCAC <u>ATAATTAAAGTCAAAGTCCTCAC</u> <u>TGCTAGT</u> GGGGGAGGCTGCTGGTGAATATTAAC CCAAGGTCACCCCAGTTATCGGAGGAGCAAAC AGGGGGCTAAGTCCAC <u>ACAATTAGAGCTGTAA</u> <u>ACATAATTTGTGTA</u> AGGGGAGGCTGCTGGTGA ATATTAACCAAGGTCACCCCAGTTATCGGAGG AGCAAACAGGGGGCTAAGTCCAC	77

3x_hSerpEnh_30mer_spacers_HNF4former_spacers_FOX Arev	3x repeat of hSerpEnh with 30mer spacers (bold underlined) with HNF4 binding site in orientation 1 & FOXA binding site in orientation 2	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGGCT AAGTCCAC <u>TTATTTGCACTCAAAGTCCACTTT</u> <u>ATTACAG</u> GGGGGAGGCTGCTGGTGAATATTAAC CAAGGTCACCCCAGTTATCGGAGGAGCAAACA GGGGCTAAGTCCAC <u>TCAATCATAAGTGTTTAC</u> <u>AAGTTTAAGATT</u> GGGGGAGGCTGCTGGTGAAT ATTAACCAAGGTCACCCCAGTTATCGGAGGAG CAAACAGGGGGCTAAGTCCAC	78
3x_hSerpEnh_30mer_spacers_HNF4revmer_spacers_FOX Afor	3x repeat of hSerpEnh with 30mer spacers (bold underlined) with HNF4 binding site in orientation 2 & FOXA binding site in orientation 1	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGGCT AAGTCCAC <u>AGTTGCTGTGTGGACTTTGTTCAC</u> <u>TGCAAGA</u> GGGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAGTTATCGGAGGAGCAAAC AGGGGCTAAGTCCAC <u>AACAGCATATTTGTAAA</u> <u>CAGTTCTATTAGT</u> GGGGGAGGCTGCTGGTGAA TATTAACCAAGGTCACCCCAGTTATCGGAGGA GCAAACAGGGGGCTAAGTCCAC	79
3x_hSerpEnh_30mer_spacers_HNF4revmer_spacers_FOX Arev	3x repeat of hSerpEnh with 30mer spacers (bold underlined) with HNF4 binding site in orientation 2 & FOXA binding site in orientation 2	GGGGGAGGCTGCTGGTGAATATTAACCAAGGT CACCCCAGTTATCGGAGGAGCAAACAGGGGGCT AAGTCCAC <u>ATTAACATTTGGGACTTTGGTTA</u> <u>ACAACA</u> GGGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAGTTATCGGAGGAGCAAAC AGGGGCTAAGTCCAC <u>CAGAGACTTATTGTTTA</u> <u>CAGCTAACTATCT</u> GGGGGAGGCTGCTGGTGAA TATTAACCAAGGTCACCCCAGTTATCGGAGGA GCAAACAGGGGGCTAAGTCCAC	80
3x_Tibetan_antelope_SERPINA1_enhancer	3 repeats of SERPINA1 enhancer from tibetan antelope, separated by T.	GGGGGAGGCTGCTGGTAAACATTAACCAAGGT CACCCCAGTTATCAGAGGAACAAACAAGGACT AAGTCCAT <u>TGGGGGAGGCTGCTGGTAAACATT</u> AACCAAGGTCACCCCAGTTATCAGAGGAACAA ACAAGGACTAAGTCCAT <u>TGGGGGAGGCTGCTG</u> <u>GTAAACATTAACCAAGGTCACCCCAGTTATCA</u> <u>GAGGAACAAACAAGGACTAAGTCCAT</u>	138
3x_Armadillo_CpGmin_SERPINA1_enhancer	3 repeats of SERPINA1 enhancer from armadillo with CpG removed, separated by T.	GGGGGAGGCTGCTAGTGAACATTAACCAAGGT CACCCAGTTATCAGAGGAGCAAACAGGGGACTA AGTCCACT <u>TGGGGGAGGCTGCTAGTGAACATTA</u> ACCAAGGTCACCCAGTTATCAGAGGAGCAAAC AGGGACTAAGTCCACT <u>TGGGGGAGGCTGCTAGT</u> <u>GAACATTAACCAAGGTCACCCAGTTATCAGAG</u> <u>GAGCAAACAGGGGACTAAGTCCAC</u>	139

[00205] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of the human SERPINA1 enhancer with FOXA & HNF4 consensus sites. In certain embodiment, the regulatory element comprising the 3x repeat of the human SERPINA1 enhancer with FOXA & HNF4 consensus sites comprises SEQ ID NO: 1.

[00206] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of HNF4_FOXA_v1 with CpG minimization. In certain embodiment,

the regulatory element comprising the 3x repeat of HNF4_FOXA_v1 with CpG minimization comprises SEQ ID NO: 2.

[00207] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v1. In certain embodiment, the regulatory element comprising the 3x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v1 comprises SEQ ID NO: 3.

[00208] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization and CpG minimization v1. In certain embodiment, the regulatory element comprising the 3x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization and CpG minimization v1 comprises SEQ ID NO: 4.

[00209] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v2. In certain embodiment, the regulatory element comprising the 3x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v2 comprises SEQ ID NO: 5.

[00210] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization and CpG minimization v2. In certain embodiment, the regulatory element comprising the 3x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization and CpG minimization v2 comprises SEQ ID NO: 6.

[00211] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v3. In certain embodiment, the regulatory element comprising the 3x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v3 comprises SEQ ID NO: 7.

[00212] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization and CpG minimization v3. In certain embodiment, the regulatory element comprising the 3x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization and CpG minimization v3 comprises SEQ ID NO: 8.

[00213] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v4. In certain embodiment, the regulatory element comprising the 3x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v4 comprises SEQ ID NO: 9.

[00214] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v5. In certain embodiment, the regulatory element comprising the 3x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v5 comprises SEQ ID NO: 10.

[00215] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v6. In certain embodiment, the regulatory element comprising the 3x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v6 comprises SEQ ID NO: 11.

[00216] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of the Chinese Tree Shrew SERPINA1 enhancer. In certain embodiment, the regulatory element comprising the 3x repeat of the Chinese Tree Shrew SERPINA1 enhancer comprises SEQ ID NO: 12.

[00217] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of the Chinese Tree Shrew SERPINA1 enhancer with CpG minimization. In certain embodiment, the regulatory element comprising the 3x repeat of the Chinese Tree Shrew SERPINA1 enhancer with CpG minimization comprises SEQ ID NO: 13.

[00218] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of the human SERPINA1 enhancer with 1 adenine between repeats. In certain embodiment, the regulatory element comprising the 3x repeat of the human SERPINA1 enhancer with 1 adenine between repeats comprises SEQ ID NO: 14.

[00219] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of the Bushbaby SERPINA1 enhancer with adenine nucleotide spacer. In certain embodiment, the regulatory element comprising the 3x repeat of the Bushbaby SERPINA1 enhancer with adenine nucleotide spacer comprises SEQ ID NO: 15.

[00220] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 5x repeat of HNF4_FOXA_v1. In certain embodiment, the regulatory element comprising the 5x repeat of HNF4_FOXA_v1 comprises SEQ ID NO: 16.

[00221] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 5x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v1. In certain embodiment, the regulatory element comprising the 5x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v1 comprises SEQ ID NO: 17.

[00222] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 5x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization and CpG minimization v1. In certain embodiment, the regulatory element comprising the 5x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization and CpG minimization v1 comprises SEQ ID NO: 18.

[00223] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 5x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v2. In certain embodiment, the regulatory element comprising the 5x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v2 comprises SEQ ID NO: 19.

[00224] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 5x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization and CpG minimization v2. In certain embodiment, the regulatory element comprising the 5x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization and CpG minimization v2 comprises SEQ ID NO: 20.

[00225] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 5x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v3. In certain embodiment, the regulatory element comprising the 5x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v3 comprises SEQ ID NO: 21.

[00226] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 5x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization and CpG minimization v3. In certain embodiment, the regulatory element comprising the 5x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization and CpG minimization v3 comprises SEQ ID NO: 22.

[00227] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 5x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v4. In certain embodiment, the regulatory element comprising the 5x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v4 comprises SEQ ID NO: 23.

[00228] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 5x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v5. In certain embodiment, the regulatory element comprising the 5x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v5 comprises SEQ ID NO: 24.

[00229] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 5x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v6. In certain embodiment, the regulatory element comprising the 5x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v6 comprises SEQ ID NO: 25.

[00230] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 5x repeat of the Chinese Tree Shrew SERPINA1 enhancer. In certain embodiment, the regulatory element comprising the 5x repeat of the Chinese Tree Shrew SERPINA1 enhancer comprises SEQ ID NO: 26.

[00231] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 5x repeat of the Chinese Tree Shrew SERPINA1 enhancer with CpG minimization. In certain embodiment, the regulatory element comprising the 5x repeat of the Chinese Tree Shrew SERPINA1 enhancer with CpG minimization comprises SEQ ID NO: 27.

[00232] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 5x repeat of the Bushbaby SERPINA1 enhancer with adenine nucleotide

spacer. In certain embodiment, the regulatory element comprising the 5x repeat of the Bushbaby SERPINA1 enhancer with adenine nucleotide spacer comprises SEQ ID NO: 28.

[00233] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 5x repeat of the human SERPINA1 enhancer. In certain embodiment, the regulatory element comprising the 5x repeat of the human SERPINA1 enhancer comprises SEQ ID NO: 29.

[00234] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 10x repeat of HNF4_FOXA_v1. In certain embodiment, the regulatory element comprising the 10x repeat of HNF4_FOXA_v1 comprises SEQ ID NO: 30.

[00235] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 10x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v1. In certain embodiment, the regulatory element comprising the 10x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v1 comprises SEQ ID NO: 31.

[00236] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 10x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization and CpG minimization v1. In certain embodiment, the regulatory element comprising the 10x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization and CpG minimization v1 comprises SEQ ID NO: 32.

[00237] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 10x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v2. In certain embodiment, the regulatory element comprising the 10x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v2 comprises SEQ ID NO: 33.

[00238] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 10x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization and CpG minimization v2. In certain embodiment, the regulatory element comprising the 10x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization and CpG minimization v2 comprises SEQ ID NO: 34.

[00239] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 10x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v3. In certain embodiment, the regulatory element comprising the 10x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization v3 comprises SEQ ID NO: 35.

[00240] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 10x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization and CpG minimization v3. In certain embodiment, the regulatory element comprising the 10x repeat of HNF4_FOXA_v1 with poly-C/poly-G minimization and CpG minimization v3 comprises SEQ ID NO: 36.

[00241] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 10x repeat of the human SERPINA1 enhancer. In certain embodiment, the regulatory element comprising the 10x repeat of the human SERPINA1 enhancer comprises SEQ ID NO: 37.

[00242] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 10x repeat of the Bushbaby SERPINA1 enhancer with adenine nucleotide spacer. In certain embodiment, the regulatory element comprising the 10x repeat of the Bushbaby SERPINA1 enhancer with adenine nucleotide spacer comprises SEQ ID NO: 38.

[00243] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising Bushbaby SERPINA1 enhancer, FOXA_HNF4_v1 enhancer, HNF4 consensus binding site enhancer. In certain embodiment, the regulatory element comprising the Bushbaby SERPINA1 enhancer, FOXA_HNF4_v1 enhancer, HNF4 consensus binding site enhancer comprises SEQ ID NO: 39.

[00244] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising HNF4 consensus binding site enhancer, Bushbaby SERPINA1 enhancer, FOXA_HNF4_v1 enhancer. In certain embodiment, the regulatory element comprising the HNF4 consensus binding site enhancer, Bushbaby SERPINA1 enhancer, FOXA_HNF4_v1 enhancer comprises SEQ ID NO: 40.

[00245] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 2mer spacers v1. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 2mer spacers v1 comprises SEQ ID NO: 41.

[00246] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 2mer spacers v2. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 2mer spacers v2 comprises SEQ ID NO: 42.

[00247] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 2mer spacers v3. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 2mer spacers v3 comprises SEQ ID NO: 43.

[00248] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 2mer spacers v4. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 2mer spacers v4 comprises SEQ ID NO: 44.

[00249] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 2mer spacers v5. In certain embodiment, the

regulatory element comprising the 3x repeat of hSerpEnh with 2mer spacers v5 comprises SEQ ID NO: 45.

[00250] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 2mer spacers v6. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 2mer spacers v6 comprises SEQ ID NO: 46.

[00251] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 2mer spacers v7. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 2mer spacers v7 comprises SEQ ID NO: 47.

[00252] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 2mer spacers v8. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 2mer spacers v8 comprises SEQ ID NO: 48.

[00253] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 2mer spacers v9. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 2mer spacers v9 comprises SEQ ID NO: 49.

[00254] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 2mer spacers v10. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 2mer spacers v10 comprises SEQ ID NO: 50.

[00255] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 2mer spacers v11. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 2mer spacers v11 comprises SEQ ID NO: 51.

[00256] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 2mer spacers v12. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 2mer spacers v12 comprises SEQ ID NO: 52.

[00257] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 2mer spacers v13. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 2mer spacers v13 comprises SEQ ID NO: 53.

[00258] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 2mer spacers v14. In certain embodiment, the

regulatory element comprising the 3x repeat of hSerpEnh with 2mer spacers v14 comprises SEQ ID NO: 54.

[00259] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 2mer spacers v15. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 2mer spacers v15 comprises SEQ ID NO: 55.

[00260] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 2mer spacers v16. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 2mer spacers v16 comprises SEQ ID NO: 56.

[00261] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 2mer spacers v17. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 2mer spacers v17 comprises SEQ ID NO: 57.

[00262] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 2mer spacers v18. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 2mer spacers v18 comprises SEQ ID NO: 58.

[00263] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 2mer spacers v19. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 2mer spacers v19 comprises SEQ ID NO: 59.

[00264] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 2mer spacers v20. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 2mer spacers v20 comprises SEQ ID NO: 60.

[00265] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 3mer spacers v1. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 3mer spacers v1 comprises SEQ ID NO: 61.

[00266] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 3mer spacers v2. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 3mer spacers v2 comprises SEQ ID NO: 62.

[00267] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 3mer spacers v3. In certain embodiment, the

regulatory element comprising the 3x repeat of hSerpEnh with 3mer spacers v3 comprises SEQ ID NO: 63.

[00268] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 5mer spacers v1. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 5mer spacers v1 comprises SEQ ID NO: 64.

[00269] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 5mer spacers v2. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 5mer spacers v2 comprises SEQ ID NO: 65.

[00270] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 5mer spacers v3. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 5mer spacers v3 comprises SEQ ID NO: 66.

[00271] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 11mer spacers v1. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 11mer spacers v1 comprises SEQ ID NO: 67.

[00272] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 11mer spacers v2. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 11mer spacers v2 comprises SEQ ID NO: 68.

[00273] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 11mer spacers v3. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 11mer spacers v3 comprises SEQ ID NO: 69.

[00274] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 11mer spacers with HNF4 binding site in orientation 1 & FOXA binding site in orientation 1. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 11mer spacers with HNF4 binding site in orientation 1 & FOXA binding site in orientation 1 comprises SEQ ID NO: 70.

[00275] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 11mer spacers with HNF4 binding site in orientation 1 & FOXA binding site in orientation 2. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 11mer spacers with HNF4 binding site in orientation 1 & FOXA binding site in orientation 2 comprises SEQ ID NO: 71.

[00276] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 11mer spacers with HNF4 binding site in orientation 2 & FOXA binding site in orientation 1. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 11mer spacers with HNF4 binding site in orientation 2 & FOXA binding site in orientation 1 comprises SEQ ID NO: 72.

[00277] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 11mer spacers with HNF4 binding site in orientation 2 & FOXA binding site in orientation 2. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 11mer spacers with HNF4 binding site in orientation 2 & FOXA binding site in orientation 2 comprises SEQ ID NO: 73.

[00278] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 30mer spacers v1. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 30mer spacers v1 comprises SEQ ID NO: 74.

[00279] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 30mer spacers v2. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 30mer spacers v2 comprises SEQ ID NO: 75.

[00280] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 30mer spacers v3. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 30mer spacers v3 comprises SEQ ID NO: 76.

[00281] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 30mer spacers with HNF4 binding site in orientation 1 & FOXA binding site in orientation 1. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 30mer spacers with HNF4 binding site in orientation 1 & FOXA binding site in orientation 1 comprises SEQ ID NO: 77.

[00282] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 30mer spacers with HNF4 binding site in orientation 1 & FOXA binding site in orientation 2. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 30mer spacers with HNF4 binding site in orientation 1 & FOXA binding site in orientation 2 comprises SEQ ID NO: 78.

[00283] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 30mer spacers with HNF4 binding site in orientation 2 & FOXA binding site in orientation 1. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 30mer spacers with HNF4 binding site in orientation 2 & FOXA binding site in orientation 1 comprises SEQ ID NO: 79.

[00284] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3x repeat of hSerpEnh with 30mer spacers with HNF4 binding site in orientation 2 & FOXA binding site in orientation 2. In certain embodiment, the regulatory element comprising the 3x repeat of hSerpEnh with 30mer spacers with HNF4 binding site in orientation 2 & FOXA binding site in orientation 2 comprises SEQ ID NO: 80.

[00285] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3 repeats of SERPINA1 enhancer derived from tibetan antelope, separated by T. In certain embodiment, the regulatory element comprising the 3x repeat of Tibetan antelope SERPINA1 comprises SEQ ID NO:138.

[00286] In some embodiments, the expression cassette comprises a regulatory element (*e.g.*, an enhancer) comprising 3 repeats of SERPINA1 enhancer derived from armadillo with minimum CpG and separated by T. In certain embodiment, the regulatory element comprising the 3x repeat of Tibetan antelope SERPINA1 comprises SEQ ID NO:139.

[00287] In one embodiment, the disclosure provides an expression cassette comprising any one of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, SEQ ID NO: 40, SEQ ID NO: 41, SEQ ID NO: 42, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 45, SEQ ID NO: 46, SEQ ID NO: 47, SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 52, SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 59, SEQ ID NO: 60, SEQ ID NO: 61, SEQ ID NO: 62, SEQ ID NO: 63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO: 69, SEQ ID NO: 70, SEQ ID NO: 71, SEQ ID NO: 72, SEQ ID NO: 73, SEQ ID NO: 74, SEQ ID NO: 75, SEQ ID NO: 76, SEQ ID NO: 77, SEQ ID NO: 78, SEQ ID NO: 79, SEQ ID NO: 80, SEQ ID NO: 138, or SEQ ID NO: 139.

[00288] In one embodiment, the expression cassette comprises a nucleic acid sequence that is at least 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical to any one of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, SEQ ID NO: 40, SEQ ID NO: 41, SEQ ID NO: 42, SEQ ID NO: 43, SEQ ID NO: 44,

SEQ ID NO: 45, SEQ ID NO: 46, SEQ ID NO: 47, SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 52, SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 59, SEQ ID NO: 60, SEQ ID NO: 61, SEQ ID NO: 62, SEQ ID NO: 63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO: 69, SEQ ID NO: 70, SEQ ID NO: 71, SEQ ID NO: 72, SEQ ID NO: 73, SEQ ID NO: 74, SEQ ID NO: 75, SEQ ID NO: 76, SEQ ID NO: 77, SEQ ID NO: 78, SEQ ID NO: 79, SEQ ID NO: 80, SEQ ID NO: 138, or SEQ ID NO: 139.

[00289] In one embodiment, the disclosure provides an expression cassette consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, SEQ ID NO: 40, SEQ ID NO: 41, SEQ ID NO: 42, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 45, SEQ ID NO: 46, SEQ ID NO: 47, SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 52, SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 59, SEQ ID NO: 60, SEQ ID NO: 61, SEQ ID NO: 62, SEQ ID NO: 63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO: 69, SEQ ID NO: 70, SEQ ID NO: 71, SEQ ID NO: 72, SEQ ID NO: 73, SEQ ID NO: 74, SEQ ID NO: 75, SEQ ID NO: 76, SEQ ID NO: 77, SEQ ID NO: 78, SEQ ID NO: 79, SEQ ID NO: 80, SEQ ID NO: 138, or SEQ ID NO: 139.

[00290] The disclosed expression cassettes can be used in any situation where liver-specific transcription is desired. In various embodiments, any of the expression cassettes, including one or more of the enhancers, the spacers, the promoters, of the disclosure can be included in a viral vector (*e.g.*, an AAV vector) or a non-viral vector (*e.g.*, a cDNA vector) for gene therapy methods in which liver-specific expression of a transgene is desired, such as liver-specific expression of a clotting factor (*e.g.*, as described herein).

III. Viral vectors

[00291] In one embodiment, the disclosure relates to recombinant viral vectors comprising a nucleic acid sequence of a liver-specific promoter as described herein, in operative combination with a heterologous nucleic acid sequence encoding a therapeutic protein.

[00292] In one embodiment, the vector comprises a viral nucleic acid sequence of greater than 10, 20, 30, 40, 50, 100, or 200 nucleotides. In certain embodiments, the sequence of a viral nucleic acid comprises a human adeno-associated virus (hAAV) of serotypes 1, 2, 3B, 4, 5, 6, 7, 8, 9, or combinations or variants thereof, which it generally comprises an inverted terminal repeat of AAV.

[00293] In one embodiment, the disclosure provides a viral particle, *e.g.*, a viral capsid comprising a vector as disclosed herein, *e.g.*, wherein the vector is packaged in a capsid. The capsid can be a recombinant or chimeric capsid or particle, for example a capsid that has amino acid sequences that are a combination of AAV pseudotypes for VP 1, VP2 or VP3. An AAV capsid VP can be derived from a human AAV gene or animal AAV gene, or combinations with genetically modified alterations, *i.e.*, AAV isolated from infected human cells or a non-human primate. Animal AAVs include those derived from birds, cattle, pigs, mice, *etc.* In one embodiment, the capsid may have amino acid sequences that are genetically modified or synthetic capsids identified by methods such as directed evolution or rational design.

[00294] In one embodiment, the vector is incompetent for replication within a human host, for example, the vector does not encode a viral polymerase.

[00295] In one embodiment, the liver-specific expression cassette comprises a sequence having at least 50, 60, 70, 75, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% sequence identity with any one of SEQ ID NOs: 1- 80, 138 or 139 as set forth above.

Expresion of a Protein from an AAV Vector

[00296] In one embodiment, the nucleic acid sequences and promoters of the disclosure are useful in the production of AAV vectors. AAV belongs to the *Parvoviridae* family and the *Dependovirus* genus. AAV is a small, enveloped virus that packages a single-stranded, linear DNA genome. Both the sense and antisense AAV DNA strands are packaged in AAV capsids with the same frequency.

[00297] The AAV genome is characterized by two inverted terminal repeats (ITRs) flanking two open reading frames (ORFs). In the AAV2 genome, for example, the first 125 nucleotides of the ITR are a palindrome, which folds back on itself to maximize base pairing and forms a T-shaped hairpin structure. The other 20 bases of the ITR, called sequence D, they remain unpaired. ITRs are cis-acting sequences important for AAV DNA replication; ITR is the origin of replication and serves as a primer for the synthesis of the second chain by DNA polymerase. The double-stranded DNA formed during this synthesis, which is called the replicating monomer, is used for a second round of self-priming replication and forms a replicating dimer. These double-stranded intermediates are processed using a chain-shifting mechanism, resulting in single-stranded DNA that is used for packaging and double-stranded DNA that is used for transcription. Located within the ITR are the Rep binding elements and a terminal resolution site (TRS). These characteristics are used by the viral regulatory protein Rep during AAV replication to process double-stranded intermediates. In addition to its role in AAV replication, ITR is also essential for AAV genome packaging, transcription, down-regulation under non-permissive conditions, and site-specific integration (Daya and Berns, Clin Microbiol Rev 21 (4): 583-593, 2008).

[00298] The AAV's left ORF contains the Rep gene, which encodes four proteins: Rep78, Rep 68, Rep52, and Rep40. The right ORF contains the Cap gene, which produces three viral capsid proteins (VP1, VP2, and VP3). The AAV capsid contains 60 viral capsid proteins arranged in icosahedral

symmetry. VP1, VP2 and VP3 are present in a 1: 1: 10 molar ratio (Daya and Berns, Clin Microbiol Rev 21 (4): 583-593, 2008).

[00299] AAV vectors generally contain a transgene expression cassette between ITRs that replaces the rep and cap genes. Vector particles are produced by cotransfecting cells with a plasmid containing the vector genome and a packaging / helper construct that expresses the rep and cap proteins in trans. During infection, the genomes of AAV vectors enter the cell nucleus and can persist in multiple molecular states. A common result is the conversion of the AAV genome to a double-stranded circular episome by synthesis of the second strand or pairing with the complementary strand.

[00300] In the context of AAV vectors, the disclosed vectors generally have a recombinant genome that It comprises the following structure:

[00301] (5'ITR of AAV) - (promoter) - (transgene) - (3' ITR of AAV)

[00302] As discussed above, these recombinant AAV vectors contain a transgene expression cassette between the ITRs that replaces the rep and cap genes. Vector particles are produced, for example, by cotransfecting cells with a plasmid containing the recombinant vector genome and a packaging / helper construct that expresses the rep and cap proteins in trans.

[00303] The AAV ITRs, and other selected AAV components described herein, can be readily selected from any AAV serotype, including, without limitation, AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9 and function variants thereof. These ITRs or other AAV components can be easily isolated using techniques available to those skilled in the art from an AAV serotype. Said AAV can be isolated or obtained from academic, commercial or public sources (for example, the American Type Culture Collection, Manassas, Va.). Alternatively, AAV sequences can be obtained through synthetic means or other suitable means by reference to published sequences such as those available in the literature or in databases such as, for example, GenBank, PubMed or the like.

[00304] In one embodiment, the nucleic acids of the disclosure are part of an expression cassette or transgene. See for example, US Patent Application Publication 20150139953. The expression cassette is comprised of a transgene and regulatory sequences, eg , for example a promoter and 5 'and 3' AVV inverted terminal repeats (ITRs). In a desirable embodiment, ITRs of AAV serotype 2 or 8 are used. However, ITRs can be selected from other suitable serotypes. An expression cassette is generally packaged in a capsid protein and delivered to a selected host cell.

[00305] In one embodiment, the disclosure provides a method of generating a recombinant adeno-associated virus (AAV) having an AAV serotype capsid, or a portion thereof. Such a method involves culturing a host cell that contains a nucleic acid sequence encoding an adeno-associated virus (AAV) serotype capsid protein; a functional rep gene; an expression cassette consisting of AAV inverted terminal repeats (ITRs) and a transgene; and enough ancillary functions to allow for packaging of the expression cassette into the AAV capsid protein. See for example, U.S. Patent Application Publication 20150139953.

[00306] Components for culture in the host cell to package an AAV expression cassette into an AAV capsid can be provided to the host cell in trans. Alternatively, one or more of the components (*e.g.*, expression cassette, rep sequences, cap sequences, and / or helper functions) can be provided by a stable host cell that has been engineered to contain one or more of the components.

[00307] In one embodiment, the disclosure relates to recombinant vectors comprising a liver-specific promoter nucleic acid sequence of the disclosure in operative combination with the transgene. The transgene is a nucleic acid sequence, heterologous to the vector sequences that flank the transgene, that encodes a protein, *e.g.*, a therapeutic protein, or other product of interest. The nucleic acid coding sequence is operably linked to regulatory components in a manner that allows transcription, translation and / or transgene expression in a host cell.

[00308] A typical transgene is a sequence that encodes a product that is useful in biology and medicine, such as proteins, peptides, RNA, enzymes, dominant negative mutants, or catalytic RNA. Desirable RNA molecules include mRNA, tRNA, dsRNA, ribosomal RNA, catalytic RNA, siRNA, guide RNA (gRNA), microRNA, small hairpin RNA, trans-splice RNA, and antisense RNA. An example of a useful RNA sequence is a sequence that inhibits or extinguishes the expression of a targeted nucleic acid sequence in the treated animal.

[00309] The transgene can be used to correct or improve genetic deficiencies, which may include deficiencies in which normal genes are expressed at lower than normal levels or deficiencies in which the functional gene product is not expressed. A preferred type of transgenic sequence encodes a therapeutic protein or polypeptide that is expressed in a host cell. The disclosure further contemplates the use of multiple transgenes, for example, to correct or improve a genetic defect caused by a multi-subunit protein. In certain situations, a different transgene can be used to encode each subunit of a protein, or to encode different peptides or proteins. This is desirable when the size of the DNA encoding the protein subunit is large, for example, for an immunoglobulin, platelet-derived growth factor, or a dystrophin protein. In order for the cell to produce the multi-subunit protein, a cell is infected with the recombinant virus that contains each of the different subunits. Alternatively, different subunits of a protein may be encoded by the same transgene.

[00310] The expression cassette can be carried in any suitable viral vector which is supplied to a host cell. The plasmids useful in the present disclosure can be engineered to be suitable for replication and, optionally, integration in prokaryotic cells, mammalian cells, or both. These plasmids (or other vectors carrying the AAV 5' ITR-heterologous molecule-3' ITR) contain sequences that allow replication of the expression cassette in eukaryotes and / or prokaryotes and selection markers for these systems. Preferably, the molecule that carries the expression cassette is transfected into the cell, where it may exist transiently. Alternatively, the expression cassette (carrying the 5' ITR of AAV-heterologous molecule-3' ITR) can be stably integrated into the genome of the host cell, either chromosomally or as an episome. In certain embodiments, the expression cassette may be present in multiple copies,

optionally in head-to-head, head-to-tail, or tail-to-tail concatamers. Suitable transfection techniques are known and can be easily used to deliver the expression cassette to the host cell.

[00311] In general, when the vector comprising the expression cassette is delivered by transfection, the vector and the relative amounts of vector DNA can be adjusted to the host cells, taking into account factors such as the selected vector, the delivery method and selected host cells. In addition to the expression cassette, the host cell contains the sequences that drive the expression of the AAV capsid protein in the host cell and the rep sequences of the same serotype as the AAV ITR serotype found in the expression cassette, or a cross-complement serotype. Although the molecules that provide rep and cap may exist in the host cell transiently (*i.e.*, through transfection), it is preferred that one or both of the rep and cap proteins and the promoter (s) that control their expression they are stably expressed in the host cell, for example, as an episome or by integration into the chromosome of the host cell.

[00312] The packaging host cell generally also contains helper functions for packaging the rAAV of the disclosure. Optionally, these functions can be supplied by a herpesvirus. More desirably, the necessary auxiliary functions are each provided from a source of human or non-human primate adenoviruses, such as those described above and / or available from a variety of sources, including the American Type Culture Collection (ATCC), Manassas, Va. (USA). The desired auxiliary functions can be provided using any means that allows their expression in a cell.

[00313] Introduction into the vector host cell can be accomplished by any means known in the art or as disclosed above, including transfection, infection, electroporation, liposome delivery, membrane fusion techniques, high-speed DNA coated microgranules, infection viral or protoplast fusion, among others. One or more of the adenoviral genes can be stably integrated into the genome of the host cell, stably expressed as episomes, or transiently expressed. All gene products can be expressed transiently, at an episome, or stably integrated, or some of the gene products can be stably expressed while others are transiently expressed. Furthermore, promoters for each of the adenoviral genes can be independently selected from a constitutive promoter, an inducible promoter, or a natural adenoviral promoter. Promoters may be regulated by a specific physiological state of the organism or the cell (*i.e.*, by the state of differentiation or in replicating or quiescent cells) or by exogenously added factors, for example.

[00314] The introduction of the molecules (such as plasmids or viruses) into the host cell can be accomplished using techniques known to the person skilled in the art. In a preferred embodiment, conventional transfection techniques, eg, transfection or electroporation with CaPO₄, and / or infection by adenovirus / AAV hybrid vectors are used in cell lines such as the HEK 293 human embryonic kidney cell line (a cell line human kidney containing functional adenovirus E1 genes that provide trans-acting E1 proteins).

[00315] One of skill in the art will readily understand that AAV techniques can be adapted for use in these and other viral vector systems for gene delivery *in vitro*, *ex vivo*, or *in vivo*. In certain embodiments, the disclosure contemplates the use of nucleic acids and vectors disclosed herein in a

variety of rAAV and non-rAAV vector systems. Such vector systems can include, for example, lentiviruses, retroviruses, poxviruses, vaccinia viruses, and adenoviral systems, among others.

[00316] In certain embodiments, the protein is a FVIII or FIX or a variant thereof as described herein. In certain embodiments, the codon and promoter optimization schemes disclosed herein could be used for any gene therapy with AAV targeting the liver. Other metabolic diseases caused by liver enzyme deficiencies and the expression of these functional proteins are contemplated.

[00317] In certain embodiments the nucleic acid sequence encoding a therapeutic protein comprises codons that are used or differentially represented in highly expressed genes within the liver or other specific tissue compared to the use of codons from the entire coding region of the human genome and avoid codons that are underrepresented in the liver or other specific tissue.

IV. Non-Viral Vectors

[00318] In one embodiment, the expression cassettes described herein are useful in the production of non-vectors.

[00319] In one embodiment, the expression cassettes described herein are useful in the production of ceDNA vectors. In one embodiment, the disclosure provides the expression and/or production of a therapeutic protein (*e.g.*, a liver-specific protein, *e.g.*, a FVIII protein) in a cell, *e.g.*, a liver cell, from a non-viral DNA vector, *e.g.*, a ceDNA vector as described herein. In particular, ceDNA vectors for expression of a therapeutic protein (*e.g.*, a FVIII protein) comprise a pair of ITRs (*e.g.*, symmetric or asymmetric as described herein) and between the ITR pair, a nucleic acid encoding a therapeutic protein (*e.g.*, a FVIII protein) operatively linked to a promoter or regulatory sequence. A distinct advantage of ceDNA vectors for expression of a therapeutic protein (*e.g.*, a FVIII protein) over traditional AAV vectors, and even lentiviral vectors, is that there is no size constraint for the nucleic acid sequences, *e.g.*, heterologous nucleic acid sequences, encoding a desired protein. Even a full length 6.8kb FVIII protein can be expressed from a single ceDNA vector. Thus, ceDNA vectors described herein can be used to express a therapeutic FVIII protein in a subject in need thereof, *e.g.*, a subject with hemophilia A.

[00320] In general, a ceDNA vector for expression of a therapeutic protein as disclosed herein, comprises in the 5' to 3' direction: a first adeno-associated virus (AAV) inverted terminal repeat (ITR), a nucleic acid sequence of interest (for example an expression cassette as described herein) and a second AAV ITR. The ITR sequences selected from any of: (i) at least one WT ITR and at least one modified AAV inverted terminal repeat (mod-ITR) (*e.g.*, asymmetric modified ITRs); (ii) two modified ITRs where the mod-ITR pair have a different three-dimensional spatial organization with respect to each other (*e.g.*, asymmetric modified ITRs), or (iii) symmetrical or substantially symmetrical WT-WT ITR pair, where each WT-ITR has the same three-dimensional spatial organization, or (iv) symmetrical or substantially symmetrical modified ITR pair, where each mod-ITR has the same three-dimensional spatial organization.

[00321] As one will appreciate, the ceDNA vector technologies can be adapted to any level of complexity or can be used in a modular fashion, where expression of different components of a therapeutic protein (*e.g.*, a FVIII protein) can be controlled in an independent manner. For example, it is specifically contemplated that the ceDNA vector technologies described here can be as simple as using a single ceDNA vector to express a single gene sequence a therapeutic protein (*e.g.*, a FVIII protein) or can be as complex as using multiple ceDNA vectors, where each vector expresses multiple FVIII therapeutic proteins or associated co-factors or accessory proteins that are each independently controlled by different promoters. The following embodiments are specifically contemplated and can be adapted by one of skill in the art as desired.

[00322] In one embodiment, a single ceDNA vector can be used to express a single component of a therapeutic protein (*e.g.*, a FVIII protein). Alternatively, a single ceDNA vector can be used to express multiple components (*e.g.*, at least 2) of a therapeutic protein (*e.g.*, a FVIII protein) under the control of a single promoter (*e.g.*, a strong promoter), optionally using an IRES sequence(s) to ensure appropriate expression of each of the components, *e.g.*, co-factors or accessory proteins.

[00323] As one of skill in the art will appreciate, it is often desirable to express components of a therapeutic protein (*e.g.*, a FVIII protein) at different expression levels, thus controlling the stoichiometry of the individual components expressed to ensure efficient protein folding and combination in the cell. Additional variations of ceDNA vector technologies can be envisioned by one of skill in the art or can be adapted from protein production methods using conventional vectors.

[00324] Certain methods for the production of a ceDNA vector for expression of a therapeutic protein (*e.g.*, a FVIII protein) comprising an asymmetrical ITR pair or symmetrical ITR pair as defined herein is described in section IV of International application PCT/US18/49996 filed September 7, 2018, which is incorporated herein in its entirety by reference. In some embodiments, a ceDNA vector for expression of a therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein can be produced using insect cells, as described herein. In alternative embodiments, a ceDNA vector for expression of a therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein can be produced synthetically and in some embodiments, in a cell-free method, as disclosed on International Application PCT/US19/14122, filed January 18, 2019, which is incorporated herein in its entirety by reference.

[00325] As described herein, in one embodiment, a ceDNA vector for expression of a therapeutic protein (*e.g.*, a FVIII protein) can be obtained, for example, by the process comprising the steps of: a) incubating a population of host cells (*e.g.*, insect cells) harboring the polynucleotide expression construct template (*e.g.*, a ceDNA-plasmid, a ceDNA-Bacmid, and/or a ceDNA-baculovirus), which is devoid of viral capsid coding sequences, in the presence of a Rep protein under conditions effective and for a time sufficient to induce production of the ceDNA vector within the host cells, and wherein the host cells do not comprise viral capsid coding sequences; and b) harvesting and isolating the ceDNA vector from the host cells. The presence of Rep protein induces replication of the vector polynucleotide with a modified ITR to produce the ceDNA vector in a host cell. However, no viral particles (*e.g.*, AAV

virions) are expressed. Thus, there is no size limitation such as that naturally imposed in AAV or other viral-based vectors.

[00326] The presence of the ceDNA vector isolated from the host cells can be confirmed by digesting DNA isolated from the host cell with a restriction enzyme having a single recognition site on the ceDNA vector and analyzing the digested DNA material on a non-denaturing gel to confirm the presence of characteristic bands of linear and continuous DNA as compared to linear and non-continuous DNA.

[00327] In yet another aspect, the disclosure provides for use of host cell lines that have stably integrated the DNA vector polynucleotide expression template (ceDNA template) into their own genome in production of the non-viral DNA vector, *e.g.*, as described in Lee, L. et al. (2013) Plos One 8(8): e69879. Preferably, Rep is added to host cells at an MOI of about 3. When the host cell line is a mammalian cell line, *e.g.*, HEK293 cells, the cell lines can have polynucleotide vector template stably integrated, and a second vector such as herpes virus can be used to introduce Rep protein into cells, allowing for the excision and amplification of ceDNA in the presence of Rep and helper virus.

[00328] In one embodiment, the host cells used to make the ceDNA vectors for expression of a therapeutic protein (*e.g.*, a FVIII protein) as described herein are insect cells, and baculovirus is used to deliver both the polynucleotide that encodes Rep protein and the non-viral DNA vector polynucleotide expression construct template for ceDNA. In some embodiments, the host cell is engineered to express Rep protein.

[00329] The ceDNA vector is then harvested and isolated from the host cells. The time for harvesting and collecting ceDNA vectors described herein from the cells can be selected and optimized to achieve a high-yield production of the ceDNA vectors. For example, the harvest time can be selected in view of cell viability, cell morphology, cell growth, *etc.* In one embodiment, cells are grown under sufficient conditions and harvested a sufficient time after baculoviral infection to produce ceDNA vectors but before a majority of cells start to die because of the baculoviral toxicity. The DNA vectors can be isolated using plasmid purification kits such as Qiagen Endo-Free Plasmid kits. Other methods developed for plasmid isolation can be also adapted for DNA vectors. Generally, any nucleic acid purification methods can be adopted.

[00330] The DNA vectors can be purified by any means known to those of skill in the art for purification of DNA. In one embodiment, ceDNA vectors are purified as DNA molecules. In another embodiment, the ceDNA vectors are purified as exosomes or microparticles.

[00331] The presence of the ceDNA vector for expression of a therapeutic protein (*e.g.*, a FVIII protein) can be confirmed by digesting the vector DNA isolated from the cells with a restriction enzyme having a single recognition site on the DNA vector and analyzing both digested and undigested DNA material using gel electrophoresis to confirm the presence of characteristic bands of linear and continuous DNA as compared to linear and non-continuous DNA.

ceDNA Plasmid

[00332] A ceDNA-plasmid is a plasmid used for later production of a ceDNA vector for expression of a therapeutic protein (*e.g.*, a FVIII protein). In some embodiments, a ceDNA-plasmid can be constructed using known techniques to provide at least the following as operatively linked components in the direction of transcription: (1) a modified 5' ITR sequence; (2) an expression cassette as described herein comprising any one of SEQ ID NOs: 1-80, 138 and 139 and comprising a therapeutic transgene; and (3) a modified 3' ITR sequence, where the 3' ITR sequence is symmetric relative to the 5' ITR sequence. In some embodiments, the expression cassette flanked by the ITRs comprises a cloning site for introducing an exogenous sequence. The expression cassette replaces the rep and cap coding regions of the AAV genomes.

[00333] In one aspect, a ceDNA vector for expression of therapeutic protein (*e.g.*, a FVIII protein) is obtained from a plasmid, referred to herein as a "ceDNA-plasmid" encoding in this order: a first adeno-associated virus (AAV) inverted terminal repeat (ITR), an expression cassette as described herein comprising any one of SEQ ID NOs: 1-80, 138 and 139 and comprising a therapeutic transgene, and a mutated or modified AAV ITR, wherein said ceDNA-plasmid is devoid of AAV capsid protein coding sequences. In alternative embodiments, the ceDNA-plasmid encodes in this order: a first (or 5') modified or mutated AAV ITR, an expression cassette comprising a transgene, and a second (or 3') modified AAV ITR, wherein said ceDNA-plasmid is devoid of AAV capsid protein coding sequences, and wherein the 5' and 3' ITRs are symmetric relative to each other. In alternative embodiments, the ceDNA-plasmid encodes in this order: a first (or 5') modified or mutated AAV ITR, an expression cassette as described herein comprising any one of SEQ ID NOs: 1-80, 138 and 139 and comprising a therapeutic transgene, and a second (or 3') mutated or modified AAV ITR, wherein said ceDNA-plasmid is devoid of AAV capsid protein coding sequences, and wherein the 5' and 3' modified ITRs are have the same modifications (*i.e.*, they are inverse complement or symmetric relative to each other).

[00334] In a further embodiment, the ceDNA-plasmid system is devoid of viral capsid protein coding sequences (*i.e.*, it is devoid of AAV capsid genes but also of capsid genes of other viruses). In addition, in a particular embodiment, the ceDNA-plasmid is also devoid of AAV Rep protein coding sequences. Accordingly, in a preferred embodiment, ceDNA-plasmid is devoid of functional AAV cap and AAV rep genes GG-3' for AAV2) plus a variable palindromic sequence allowing for hairpin formation.

[00335] A ceDNA-plasmid of the present disclosure can be generated using natural nucleotide sequences of the genomes of any AAV serotypes well known in the art. In one embodiment, the ceDNA-plasmid backbone is derived from the AAV1, AAV2, AAV3, AAV4, AAV5, AAV 5, AAV7, AAV8, AAV9, AAV10, AAV 11, AAV12, AAVrh8, AAVrh10, AAV-DJ, and AAV-DJ8 genome. *e.g.*, NCBI: NC 002077; NC 001401; NC001729; NC001829; NC006152; NC 006260; NC 006261; Kotin and Smith, The Springer Index of Viruses, available at the URL maintained by Springer. In a particular embodiment, the ceDNA-plasmid backbone is derived from the AAV2 genome. In another

particular embodiment, the ceDNA-plasmid backbone is a synthetic backbone genetically engineered to include at its 5' and 3' ITRs derived from one of these AAV genomes.

[00336] A ceDNA-plasmid can optionally include a selectable or selection marker for use in the establishment of a ceDNA vector-producing cell line. In one embodiment, the selection marker can be inserted downstream (*i.e.*, 3') of the 3' ITR sequence. In another embodiment, the selection marker can be inserted upstream (*i.e.*, 5') of the 5' ITR sequence. Appropriate selection markers include, for example, those that confer drug resistance. Selection markers can be, for example, a blasticidin S-resistance gene, kanamycin, geneticin, and the like. In a preferred embodiment, the drug selection marker is a blasticidin S-resistance gene.

[00337] An exemplary ceDNA (*e.g.*, rAAV0) vector for expression of a therapeutic protein (*e.g.*, a FVIII protein) is produced from an rAAV plasmid. A method for the production of a rAAV vector, can comprise: (a) providing a host cell with a rAAV plasmid as described above, wherein both the host cell and the plasmid are devoid of capsid protein encoding genes, (b) culturing the host cell under conditions allowing production of an ceDNA genome, and (c) harvesting the cells and isolating the AAV genome produced from said cells.

Exemplary method of making the ceDNA vectors from ceDNA plasmids

[00338] Methods for making capsid-less ceDNA vectors for expression of a therapeutic protein (*e.g.*, a FVIII protein) are also provided herein, notably a method with a sufficiently high yield to provide sufficient vector for *in vivo* experiments.

[00339] In some embodiments, a method for the production of a ceDNA vector for expression of a therapeutic protein (*e.g.*, a FVIII protein) comprises the steps of: (1) introducing the nucleic acid construct comprising an expression cassette and two symmetric ITR sequences into a host cell (*e.g.*, Sf9 cells), (2) optionally, establishing a clonal cell line, for example, by using a selection marker present on the plasmid, (3) introducing a Rep coding gene (either by transfection or infection with a baculovirus carrying said gene) into said insect cell, and (4) harvesting the cell and purifying the ceDNA vector. The nucleic acid construct comprising an expression cassette and two ITR sequences described above for the production of ceDNA vector can be in the form of a ceDNA plasmid, or Bacmid or Baculovirus generated with the ceDNA plasmid as described below. The nucleic acid construct can be introduced into a host cell by transfection, viral transduction, stable integration, or other methods known in the art.

Cell lines

[00340] Host cell lines used in the production of a ceDNA vector for expression of a therapeutic protein (*e.g.*, a FVIII protein) can include insect cell lines derived from *Spodoptera frugiperda*, such as Sf9 Sf21, or *Trichoplusia ni* cell, or other invertebrate, vertebrate, or other eukaryotic cell lines including mammalian cells. Other cell lines known to an ordinarily skilled artisan can also be used, such as HEK293, Huh-7, HeLa, HepG2, Hep1A, 911, CHO, COS, MeWo, NIH3T3, A549, HT1 180, monocytes, and mature and immature dendritic cells. Host cell lines can be transfected for stable expression of the ceDNA-plasmid for high yield ceDNA vector production.

[00341] ceDNA-plasmids can be introduced into Sf9 cells by transient transfection using reagents (*e.g.*, liposomal, calcium phosphate) or physical means (*e.g.*, electroporation) known in the art. Alternatively, stable Sf9 cell lines which have stably integrated the ceDNA-plasmid into their genomes can be established. Such stable cell lines can be established by incorporating a selection marker into the ceDNA -plasmid as described above. If the ceDNA -plasmid used to transfect the cell line includes a selection marker, such as an antibiotic, cells that have been transfected with the ceDNA-plasmid and integrated the ceDNA-plasmid DNA into their genome can be selected for by addition of the antibiotic to the cell growth media. Resistant clones of the cells can then be isolated by single-cell dilution or colony transfer techniques and propagated.

Isolating and Purifying ceDNA vectors

[00342] ceDNA-vectors for expression of a therapeutic protein (*e.g.*, a FVIII protein) disclosed herein can be obtained from a producer cell expressing AAV Rep protein(s), further transformed with a ceDNA-plasmid, ceDNA-bacmid, or ceDNA-baculovirus. Plasmids useful for the production of ceDNA vectors include plasmids that encode a therapeutic protein (*e.g.*, a FVIII protein), or plasmids encoding one or more REP proteins.

[00343] In one aspect, a polynucleotide encodes the AAV Rep protein (Rep 78 or 68) delivered to a producer cell in a plasmid (Rep-plasmid), a bacmid (Rep-bacmid), or a baculovirus (Rep-baculovirus). The Rep-plasmid, Rep-bacmid, and Rep-baculovirus can be generated by methods described above.

[00344] Methods to produce a ceDNA vector for expression of a therapeutic protein (*e.g.*, a FVIII protein) are described herein. Expression constructs used for generating a ceDNA vector for expression of a therapeutic protein (*e.g.*, a FVIII protein) as described herein can be a plasmid (*e.g.*, ceDNA-plasmids), a Bacmid (*e.g.*, ceDNA-bacmid), and/or a baculovirus (*e.g.*, ceDNA-baculovirus). By way of an example only, a ceDNA-vector can be generated from the cells co-infected with ceDNA-baculovirus and Rep-baculovirus. Rep proteins produced from the Rep-baculovirus can replicate the ceDNA-baculovirus to generate ceDNA-vectors. Alternatively, ceDNA vectors for expression of a therapeutic protein (*e.g.*, a FVIII protein) can be generated from the cells stably transfected with a construct comprising a sequence encoding the AAV Rep protein (Rep78/52) delivered in Rep-plasmids, Rep-bacmids, or Rep-baculovirus. CeDNA-Baculovirus can be transiently transfected to the cells, be replicated by Rep protein and produce ceDNA vectors.

[00345] The bacmid (*e.g.*, ceDNA-bacmid) can be transfected into permissive insect cells such as Sf9, Sf21, Tni (*Trichoplusia ni*) cell, High Five cell, and generate ceDNA-baculovirus, which is a recombinant baculovirus including the sequences comprising the symmetric ITRs and the expression cassette. ceDNA-baculovirus can be again infected into the insect cells to obtain a next generation of the recombinant baculovirus. Optionally, the step can be repeated once or multiple times to produce the recombinant baculovirus in a larger quantity.

[00346] The time for harvesting and collecting ceDNA vectors for expression of a therapeutic protein (*e.g.*, a FVIII protein) as described herein from the cells can be selected and optimized to achieve

a high-yield production of the ceDNA vectors. For example, the harvest time can be selected in view of cell viability, cell morphology, cell growth, *etc.* Usually, cells can be harvested after sufficient time after baculoviral infection to produce ceDNA vectors (*e.g.*, ceDNA vectors) but before majority of cells start to die because of the viral toxicity. The ceDNA-vectors can be isolated from the Sf9 cells using plasmid purification kits such as Qiagen ENDO-FREE PLASMID® kits. Other methods developed for plasmid isolation can be also adapted for ceDNA vectors. Generally, any art-known nucleic acid purification methods can be adopted, as well as commercially available DNA extraction kits.

[00347] Alternatively, purification can be implemented by subjecting a cell pellet to an alkaline lysis process, centrifuging the resulting lysate and performing chromatographic separation. As one non-limiting example, the process can be performed by loading the supernatant on an ion exchange column (*e.g.*, SARTOBIND Q®) which retains nucleic acids, and then eluting (*e.g.*, with a 1.2 M NaCl solution) and performing a further chromatographic purification on a gel filtration column (*e.g.*, 6 fast flow GE). The capsid-free AAV vector is then recovered by, *e.g.*, precipitation.

[00348] In some embodiments, ceDNA vectors for expression of a therapeutic protein (*e.g.*, a FVIII protein) can also be purified in the form of exosomes, or microparticles. It is known in the art that many cell types release not only soluble proteins, but also complex protein/nucleic acid cargoes via membrane microvesicle shedding (Cocucci et al, 2009; EP 10306226.1) Such vesicles include microvesicles (also referred to as microparticles) and exosomes (also referred to as nanovesicles), both of which comprise proteins and RNA as cargo. Microvesicles are generated from the direct budding of the plasma membrane, and exosomes are released into the extracellular environment upon fusion of multivesicular endosomes with the plasma membrane. Thus, ceDNA vector-containing microvesicles and/or exosomes can be isolated from cells that have been transduced with the ceDNA-plasmid or a bacmid or baculovirus generated with the ceDNA-plasmid.

[00349] Microvesicles can be isolated by subjecting culture medium to filtration or ultracentrifugation at 20,000 x g, and exosomes at 100,000 x g. The optimal duration of ultracentrifugation can be experimentally-determined and will depend on the particular cell type from which the vesicles are isolated. Preferably, the culture medium is first cleared by low-speed centrifugation (*e.g.*, at 2000 x g for 5-20 minutes) and subjected to spin concentration using, *e.g.*, an AMICON® spin column (Millipore, Watford, UK). Microvesicles and exosomes can be further purified via FACS or MACS by using specific antibodies that recognize specific surface antigens present on the microvesicles and exosomes. Other microvesicle and exosome purification methods include, but are not limited to, immunoprecipitation, affinity chromatography, filtration, and magnetic beads coated with specific antibodies or aptamers. Upon purification, vesicles are washed with, *e.g.*, phosphate-buffered saline. One advantage of using microvesicles or exosome to deliver ceDNA-containing vesicles is that these vesicles can be targeted to various cell types by including on their membrane proteins recognized by specific receptors on the respective cell types. (See also EP 10306226)

[00350] Another aspect of the disclosure herein relates to methods of purifying ceDNA vectors from host cell lines that have stably integrated a ceDNA construct into their own genome. In one embodiment, ceDNA vectors are purified as DNA molecules. In another embodiment, the ceDNA vectors are purified as exosomes or microparticles.

[00351] FIG. 5 of International application PCT/US18/49996 shows a gel confirming the production of ceDNA from multiple ceDNA-plasmid constructs using the method described in the Examples.

V. Exemplary Recombinant Vectors

[00352] The nucleic acid sequences disclosed herein are useful in the production of expression plasmid, viral (AAV and rAAV) and non-viral vectors (ceDNA), and are also useful as antisense delivery vectors, gene therapy vectors, gene editing vectors (gRNA), or vaccine vectors.

[00353] In one embodiment, the disclosure provides a viral gene delivery vector comprising any one of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, SEQ ID NO: 40, SEQ ID NO: 41, SEQ ID NO: 42, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 45, SEQ ID NO: 46, SEQ ID NO: 47, SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 52, SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 59, SEQ ID NO: 60, SEQ ID NO: 61, SEQ ID NO: 62, SEQ ID NO: 63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO: 69, SEQ ID NO: 70, SEQ ID NO: 71, SEQ ID NO: 72, SEQ ID NO: 73, SEQ ID NO: 74, SEQ ID NO: 75, SEQ ID NO: 76, SEQ ID NO: 77, SEQ ID NO: 78, SEQ ID NO: 79, SEQ ID NO: 80, SEQ ID NO: 138 or SEQ ID NO: 139 operably linked to a liver-specific promoter and a therapeutic transgene. In one embodiment, the disclosure provides a viral gene delivery vector comprising a nucleic acid sequence that is at least 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical to any one of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, SEQ ID NO: 40, SEQ ID NO: 41, SEQ ID NO: 42, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 45, SEQ ID NO: 46, SEQ ID NO: 47, SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 52, SEQ ID NO:

53, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 59, SEQ ID NO: 60, SEQ ID NO: 61, SEQ ID NO: 62, SEQ ID NO: 63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO: 69, SEQ ID NO: 70, SEQ ID NO: 71, SEQ ID NO: 72, SEQ ID NO: 73, SEQ ID NO: 74, SEQ ID NO: 75, SEQ ID NO: 76, SEQ ID NO: 77, SEQ ID NO: 78, SEQ ID NO: 79, SEQ ID NO: 80, SEQ ID NO: 138, or SEQ ID NO: 139 operably linked to a liver-specific promoter and a therapeutic transgene. In one embodiment, the disclosure provides a viral gene delivery vector consisting of any one of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, SEQ ID NO: 40, SEQ ID NO: 41, SEQ ID NO: 42, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 45, SEQ ID NO: 46, SEQ ID NO: 47, SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 52, SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 59, SEQ ID NO: 60, SEQ ID NO: 61, SEQ ID NO: 62, SEQ ID NO: 63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO: 69, SEQ ID NO: 70, SEQ ID NO: 71, SEQ ID NO: 72, SEQ ID NO: 73, SEQ ID NO: 74, SEQ ID NO: 75, SEQ ID NO: 76, SEQ ID NO: 77, SEQ ID NO: 78, SEQ ID NO: 79, SEQ ID NO: 80, SEQ ID NO: 138, or SEQ ID NO: 139 operably linked to a liver-specific promoter and a therapeutic transgene. In one embodiment, the disclosure provides a non-viral gene delivery vector comprising any one of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, SEQ ID NO: 40, SEQ ID NO: 41, SEQ ID NO: 42, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 45, SEQ ID NO: 46, SEQ ID NO: 47, SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 52, SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 59, SEQ ID NO: 60, SEQ ID NO: 61, SEQ ID NO: 62, SEQ ID NO: 63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO: 69, SEQ ID NO: 70, SEQ ID NO: 71, SEQ ID NO: 72, SEQ ID NO: 73, SEQ ID NO: 74, SEQ ID NO: 75, SEQ ID NO: 76, SEQ ID NO: 77, SEQ ID NO: 78, SEQ ID NO: 79, SEQ ID NO: 80, SEQ ID NO: 138, or SEQ ID NO: 139 operably linked to a liver-specific promoter and a therapeutic transgene. In one embodiment, the disclosure provides a non-viral gene delivery

vector comprising a nucleic acid sequence that is at least 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical to any one of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, SEQ ID NO: 40, SEQ ID NO: 41, SEQ ID NO: 42, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 45, SEQ ID NO: 46, SEQ ID NO: 47, SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 52, SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 59, SEQ ID NO: 60, SEQ ID NO: 61, SEQ ID NO: 62, SEQ ID NO: 63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO: 69, SEQ ID NO: 70, SEQ ID NO: 71, SEQ ID NO: 72, SEQ ID NO: 73, SEQ ID NO: 74, SEQ ID NO: 75, SEQ ID NO: 76, SEQ ID NO: 77, SEQ ID NO: 78, SEQ ID NO: 79, SEQ ID NO: 80, SEQ ID NO: 138, or SEQ ID NO: 139 operably linked to a liver-specific promoter and a therapeutic transgene. In one embodiment, the disclosure provides a non-viral gene delivery vector consisting of comprising any one of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, SEQ ID NO: 40, SEQ ID NO: 41, SEQ ID NO: 42, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 45, SEQ ID NO: 46, SEQ ID NO: 47, SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 52, SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 59, SEQ ID NO: 60, SEQ ID NO: 61, SEQ ID NO: 62, SEQ ID NO: 63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO: 69, SEQ ID NO: 70, SEQ ID NO: 71, SEQ ID NO: 72, SEQ ID NO: 73, SEQ ID NO: 74, SEQ ID NO: 75, SEQ ID NO: 76, SEQ ID NO: 77, SEQ ID NO: 78, SEQ ID NO: 79, SEQ ID NO: 80, SEQ ID NO: 138, or SEQ ID NO: 139 operably linked to a liver-specific promoter and a therapeutic transgene.

[00354] In one embodiment, the nucleic acids of the disclosure can be part of any genetic element (vector) that can be supplied to a host cell, for example, naked DNA, a plasmid, phage, transposon, cosmid, episome, a protein in a non-viral delivery vehicle (*e.g.*, a lipid-based transporter), viruses, *etc.* that transfer the sequences carried on them.

[00355] In one embodiment, a vector can be a lentivirus-based vector (containing genes or lentiviral sequences), for example, having nucleic acid sequences derived from VSVG or GP64 pseudotypes or both.

[00356] According to some aspects, the disclosure refers to virus particles, *e.g.*, capsids, that contain the nucleic acid sequences encoding the expression cassettes and proteins disclosed herein. Viral particles, capsids, and recombinant vectors are useful in delivering a heterologous gene or other nucleic acid sequences to a target cell. Nucleic acids can be easily used in a variety of vector systems, capsids, and host cells. In one embodiment, the nucleic acids are in vectors contained within a capsid comprising terminal protection proteins, including AAV capsid proteins vp1, vp2, vp3 and hypervariable regions.

Exemplary Therapeutic protein (e.g., a FVIII protein)

In particular, a ceDNA vector for expression of a therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein can encode, for example, but is not limited to a FVIII protein, as well as variants, and/or active fragments thereof, for use in the treatment, prophylaxis, and/or amelioration of one or more symptoms of hemophilia A. In one aspect, the hemophilia A is a human hemophilia A.

FVIII therapeutic proteins and fragments thereof

[00357] Essentially any version of the FVIII therapeutic protein or fragment thereof (*e.g.*, functional fragment) can be encoded by and expressed in and from a viral or non-viral vector as described herein. One of skill in the art will understand that FVIII therapeutic protein includes all splice variants and orthologs of the Therapeutic protein (*e.g.*, a FVIII protein). FVIII therapeutic protein includes intact molecules as well as fragments (*e.g.*, functional) thereof.

[00358] In one embodiment, the nucleic acid sequence encoding the protein comprises a higher percentage of liver cell specific amino acid codons compared to the general use of human codons. According to some aspects, the disclosure provides methods of treating a subject diagnosed with a genetic disease or disorder that results in the expression of a mutated or truncated non-functional protein by administering an effective amount of a vector disclosed herein (*e.g.*, an AAV vector or a ceDNA vector) to express a functional liver protein.

Factor VIII

[00359] Factor VIII is the nonenzymatic cofactor to the activated clotting factor IX (FIXa), which, when proteolytically activated, interacts with FIXa to form a tight noncovalent complex that binds to and activates factor X (FX).

[00360] The Factor VIII gene or protein can also be referred to as F8, Coagulation Factor VIII, Procoagulant Component, Antihemophilic Factor, F8C, AHF, DXS1253E, FVIII, HEMA, or F8B. Expression of the Factor VIII gene is tissue-specific and is mostly observed in liver cells. The highest level of the mRNA and Factor VIII proteins has been detected in liver sinusoidal cells; significant amounts of Factor VIII are also present in hepatocytes and in Kupffer cells (resident macrophages of liver sinusoids). Moderate levels of Factor VIII protein are detectable in the serum and plasma. Low to moderate levels of Factor VIII protein are expressed in fetal brain, retina, kidney and testis.

[00361] Factor VIII mRNA is expressed throughout many tissues of the body, including bone marrow, whole blood, white blood cells, lymph nodes, thymus, brain, cerebral cortex, cerebellum, retina, spinal cord, tibial nerve, heart, artery, smooth muscle, skeletal muscle, small intestine, colon, adipocytes, kidney, liver, lung, spleen, stomach, esophagus, bladder, pancreas, thyroid, salivary gland, adrenal gland, pituitary gland, breast, skin, ovary, uterus, placenta, prostate, and testis. The FVIII gene localized on the long arm of the X chromosome occupies a region approximately 186 kbp long and consists of 26 exons (69–3,106 bp) and introns (from 207 bp to 32.4 kbp). The total length of the coding sequence of this gene is 9 kbp.

[00362] The mature factor VIII polypeptide comprises the A1–A2–B–A3–C1–C2 structural domains. Three acidic subdomains, which are denoted as a1–a3 – A1(a1)–A2(a2)–B–(a3)A3–C1–C2, localize at the boundaries of A domains and play a significant role in the interaction between FVIII and other proteins (in particular, with thrombin). Mutations in these subdomains reduce the level of factor VIII activation by thrombin.

[00363] The factor VIII protein (Coagulation factor VIII isoform) is a preproprotein [Homo sapiens]; Accession number: NP_000123.1 (2351 aa) and has the following sequence:

```
MQIELSTCFFLCLLRFCFSATRRYYLGAVELSWDYMQSDLGELPVDARFPFRVPSFPFNTSVVYKCTL
FVEFTDHLFENIAKPRPPWMLLGPITQAEVYDITVITLKNMASHPVSLHAVGVSYWKASEGAEYDDQTS
QREKEDDKVFPGGSHTYVWQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCREGLAKEK
TQTLHKFILLFAVFDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNRSLPGLIGCHRKSVMYWHV
IGMGTTPVEHSIFLEGHTFLVRNHRQASLEISPITFLTAQTLLMDLGQFLFCHISSHQHDGMEAYVKV
DSCPEEPQLRMKNNEEAEDYDDDLTDSEMDVVRFDNDNSPSFIQIRSVAKKHPKTWVHYIAAEEDWDY
APLVLAPDDRSYKSQYLNNGPQRIGRKYKKVRFMAYTDETFTKREAIQHESGILGPLLYGEVGDITLLII
FKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPRCLTRY
SSFVNMERDLASGLIGPLLCYKESVDQGRNQIMSDKRNILFVSFDENRSWYLTENIQRFNPAGVQ
LEDPEFQASNIMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKHKMVYEDTTLT
PFSGETVFMSENPGLWILGCHNSDFRNRGMTALLKVSSCDKNTGDYEDSYEDISAYLLSKNNAIEPR
SFSQNSRHPSTRQKQFNATTIPENDIEKTDPPFAHRTMPMKIQNVSSDILLMLLRQSPTPHGLSLSDLQ
EAKYETFSDDPSPGAIDSNNLSSEMTHFRPQLHHSMDVMFTPEGLQLRLNEKLGTAAATELKKLDFKV
SSTSNNLISTIPSDNLAAGTDNTSSLGPPSMPVHYDSQLDITLFGKKSSPLTESGGPLSLSEENND SKL
LESGLMNSQESSWGKNVSSSTESGRLEFKGKRAHGPAALLTKDNALFKVSISSLKTNKTSNNSATNRKTHID
GPSLLIENSPSVQNILESDETFKKVTPLIHDRMLMDKNATALRLNHMSNKTSSKNMEMVQKKEGPI
PPDAQNPDMSSFFKMLFLPESARWIQRTGKNSLNSGQGPSKQLVSLGPEKSVEGQNFLLSEKNKVVVGK
GEFTKDVGLKEMVFPSSRNFLTLNLDNLHENNTHNQEKIQQEIEKKETLIQENVVLPQIHTVTGTFKNF
MKNLFLSTRQNVESYDGAYAPVLQDFRSLNDSTNRKHTAHFSKKGEEENLEGLGNQTKQIVEKYA
CTTRISPNTSQQNFVTQRSKRALKQFRLPLEETELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKE
KGAITQSPSLSDCLTRSHSIPQANRSLPIAKVSSFPSIRPIYLTRVLFQDNSSHLPAASYRKKDSGVQE
SSHFLQGAKKNNLSLAILTLEMTGDQREVGSLGTSATNSVTYKKVENTVLPKPDLPKTSGKVELLPKVH
IYQKDLFPFTETSNNGSPGHLDLVEGSLQGTGAIKWNEANRPGKVPFLRVATESSAKTPSKLLDPLAWD
```

NHYGTQIPKEEWKSQEKSPKTAFFKKKDTILSLNACESNHAIAAINEGQNKPEIEVTWAKQGRTERLCS
 QNPPVLKRHQREITRTTLQSDQEEIDYDDTISVEMKKEDFDIYDEENQSPRSFQKKTRHYFIAAVERL
 WDYGMSSSPHVLNRNAQSGSVPQFKKVVVFQEFDTGSFTQPLYRGELNEHLGLLGPYIRAEVEDNIMVTF
 RNQASRPYSFYSSLISYEEDQRQGAEPKRNFKVKNETKTYFWKVQHMAPTKDEFDCKAWAYFSDVDLE
 KDVHSGLIGPLLVCHTNTLNPAHGRQVTVQEFALFFTIFDETKSWYFTENMERNCRAPCNIQMEDPTFK
 ENYRFHAINGYIMDTLPGLVMAQDQIRWYLLSMGSNENIHSIHFSGHVFTVRKKEEYKMALYNLYPGV
 FETVEMLPKAGIWRVECLIGEHLHAGMSTFLVYSNKCQTPLGMASGHIRDFQITASGQYQWAPKLA
 RLHYSGSINAWSTKEPFSWIKVDLLAPMIIHGKTKQGARQKFSSLYISQFIIMYSLDGKKWQTYRGNST
 GTLMVFFGNVDSSGIKHNIFNPPIIARYIRLHPHYSIRSTLRMELMGCDLNSCSMPLGMESKAISDAQ
 ITASSYFTNMFATWSPSKARLHLQGRSNAWRPQVNNPKEWLQVDFQKTMKVTGVTQGVKSLLTSMYVK
 EFLISSQDGHQWTLFFQNGKVKVFQGNQDSFTPVVNSLDPPLLTRYLRHPQSWVHQIALRMEVLGCE
 AQDLY (SEQ ID NO: 142)

[00364] In one embodiment, FVIII therapeutic protein can be a “therapeutic protein variant,” which refers to the FVIII therapeutic protein having an altered amino acid sequence, composition or structure as compared to its corresponding native FVIII therapeutic protein. In one embodiment, FVIII is a functional version (*e.g.*, wild type Therapeutic protein (*e.g.*, a FVIII protein)). It may also be useful to express a mutant version of Therapeutic protein (*e.g.*, a FVIII protein) such as a point mutation (F309 mutation) or deletion mutation (*e.g.*, B domain deleted and/or single chain recombinant FVIII) as described in many examples herein. FVIII therapeutic protein expressed from the ceDNA vectors may further comprise a sequence/moiety that confers an additional functionality, such as fluorescence, enzyme activity, or secretion signal. In one embodiment, an FVIII therapeutic protein variant comprises a non-native tag sequence for identification (*e.g.*, an immunotag) to allow it to be distinguished from endogenous FVIII therapeutic protein in a recipient host cell.

[00365] It is well within the abilities of one of skill in the art to take a known and/or publicly available protein sequence of *e.g.*, FVIII therapeutic protein and reverse engineer a cDNA sequence to encode such a protein. The cDNA can then be codon optimized to match the intended host cell and inserted into a vector as described herein.

[00366] In one embodiment, the FVIII therapeutic protein encoding sequence can be derived from an existing host cell or cell line, for example, by reverse transcribing mRNA obtained from the host and amplifying the sequence using PCR.

Vectors expressing FVIII proteins

[00367] A ceDNA vector having one or more sequences encoding a desired FVIII therapeutic protein can comprise regulatory sequences such as promoters, secretion signals, introns, polyA regions, and enhancers to maximize expression of the FVIII therapeutic protein when delivered to a desired cell or tissue. At a minimum, a ceDNA vector comprises one or more nucleic acid sequences encoding the FVIII therapeutic protein or functional fragment thereof.

[00368] In some embodiments, the ceDNA vector comprises a codon optimized FVIII sequence. In some embodiments, the ceDNA vector comprises a codon optimized FVIII sequence as shown in FIGs. 11 and 12 (hFVIII-F309S-BD226seq124-BDD-F309) In some embodiments, the ceDNA vector comprises an FVIII sequence comprising the nucleic acid sequence as set forth in SEQ ID NO: 143 as shown below: ceDNA 1651 ORF sequence (GE-715; hFVIII-F309S-BD226seq124-BDD-F309)

```
ATGCAGATTGAGCTGAGCACCTGCTTCTTCCTGTGCCTGCTGAGGTTCTGCTTCTCTGCCACCAGGAGA
TACTACCTGGGGGCTGTGGAGCTGAGCTGGGACTACATGCAGTCTGACCTGGGGGAGCTGCCTGTGGAT
GCCAGGTTCCCCCAGAGTGCCCAAGAGCTTCCCTTCAACACCTCTGTGGTGTACAAGAAGACCCTG
TTTGTGGAGTTCACTGACCACCTGTTCAACATTGCCAAGCCCAGGCCCCCTGGATGGGCCTGCTGGGC
CCCACCATCCAGGCTGAGGTGTATGACACTGTGGTGATCACCTGAAGAACATGGCCAGCCACCCTGTG
AGCCTGCATGCTGTGGGGGTGAGCTACTGGAAGGCCTCTGAGGGGGCTGAGTATGATGACCAGACCAGC
CAGAGGGAGAAGGAGGATGACAAGGTGTTCCCTGGGGGCAGCCACACCTATGTGTGGCAGGTGCTGAAG
GAGAATGGCCCCATGGCCTCTGACCCCTGTGCCTGACCTACAGCTACCTGAGCCATGTGGACCTGGTG
AAGGACCTGAACTCTGGCCTGATTGGGGCCCTGCTGGTGTGCAGGGAGGGCAGCCTGGCCAAGGAGAAG
ACCCAGACCCTGCACAAGTTCATCCTGCTGTTTGCTGTGTTTGATGAGGGCAAGAGCTGGCACTCTGAA
ACCAAGAACAGCCTGATGCAGGACAGGGATGCTGCCTCTGCCAGGGCCTGGCCAAGATGCACACTGTG
AATGGCTATGTGAACAGGAGCCTGCCTGGCCTGATTGGCTGCCACAGGAAGTCTGTGTACTGGCATGTG
ATTGGCATGGGCACCAACCCCTGAGGTGCACAGCATCTTCCTGGAGGGCCACACCTTCCTGGTCAGGAAC
CACAGGCAGGCCAGCCTGGAGATCAGCCCCATCACCTTCCTGACTGCCAGACCCTGCTGATGGACCTG
GGCCAGTTCCTGCTGTTCTGCCACATCAGCAGCCACCAGCATGATGGCATGGAGGCCTATGTGAAGGTG
GACAGCTGCCCTGAGGAGCCCCAGCTGAGGATGAAGAACAATGAGGAGGCTGAGGACTATGATGATGAC
CTGACTGACTCTGAGATGGATGTGGTGAGGTTTGATGATGACAACAGCCCCAGCTTCATCCAGATCAGG
TCTGTGGCCAAGAAGCACCCCAAGACCTGGGTGCACTACATTGCTGCTGAGGAGGAGGACTGGGACTAT
GCCCCCTGGTGCTGGCCCCCTGATGACAGGAGCTACAAGAGCCAGTACCTGAACAATGGCCCCCAGAGG
ATTGGCAGGAAGTACAAGAAGGTCAGGTTTCATGGCCTACACTGATGAAACCTTCAAGACCAGGGAGGCC
ATCCAGCATGAGTCTGGCATCCTGGGCCCCCTGCTGTATGGGGAGGTGGGGGACACCCTGCTGATCATC
TTCAAGAACCAGGCCAGCAGGCCCTACAACATCTACCCCATGGCATCACTGATGTGAGGCCCCCTGTAC
AGCAGGAGGCTGCCCAAGGGGTGAAGCACCTGAAGGACTTCCCCATCCTGCCTGGGGAGATCTTCAAG
TACAAGTGGACTGTGACTGTGGAGGATGGCCCCACCAAGTCTGACCCAGGTGCCTGACCAGATACTAC
AGCAGCTTTGTGAACATGGAGAGGGACCTGGCCTCTGGCCTGATTGGCCCCCTGCTGATCTGCTACAAG
GAGTCTGTGGACCAGAGGGGCAACCAGATCATGTCTGACAAGAGGAATGTGATCCTGTTCTCTGTGTTT
GATGAGAACAGGAGCTGGTACCTGACTGAGAACATCCAGAGGTTCTGCCCAACCCTGCTGGGGTGCAG
CTGGAGGACCCTGAGTTCCAGGCCAGCAACATCATGCACAGCATCAATGGCTATGTGTTTGACAGCCTG
CAGCTGTCTGTGTGCCTGCATGAGGTGGCCTACTGGTACATCCTGAGCATTGGGGCCAGACTGACTTC
CTGTCTGTGTTCTTCTCTGGCTACACCTTCAAGCACAAAGATGGTGTATGAGGACACCCTGACCCTGTTT
CCCTTCTCTGGGGAGACTGTGTTTCATGAGCATGGAGAACCCTGGCCTGTGGATTCTGGGCTGCCACAAC
TCTGACTTCAGGAACAGGGGCATGACTGCCCTGCTGAAAGTCTCCAGCTGTGACAAGAACACTGGGGAC
```

TACTATGAGGACAGCTATGAGGACATCTCTGCCTACCTGCTGAGCAAGAACAATGCCATTGAGCCCAGG
AGCTTCAGCCAGAATAGCAGGCACCCAGCACCAGGCAGAAGCAGTTCAATGCCACCACCATCCCAGAG
AATACCACCTGCGAGTCTGACCAGGAGGAGATTGACTATGATGACACCATCTCTGTGGAGATGAAGAAG
GAGGACTTTGACATCTACGACGAGGACGAGAACCAGAGCCCCAGGAGCTTCCAGAAGAAGACCAGGCAC
TACTTCATTGCTGCTGTGGAGAGGCTGTGGGACTATGGCATGAGCAGCAGCCCCCATGTGCTGAGGAAC
AGGGCCCAGTCTGGCTCTGTGCCCCAGTTCAAGAAGGTGGTGTTCAGGAGTTCACTGATGGCAGCTTC
ACCCAGCCCCCTGTACAGAGGGGAGCTGAATGAGCACCTGGGCCTGCTGGGCCCCCTACATCAGGGCTGAG
GTGGAGGACAACATCATGGTGACCTTCAGGAACCAGGCCAGCAGGCCCTACAGCTTCTACAGCAGCCTG
ATCAGCTATGAGGAGGACCAGAGGCAGGGGGCTGAGCCCAGGAAGAACTTTGTGAAGCCCAATGAAACC
AAGACCTACTTCTGGAAGGTGCAGCACCATGGCCCCACCAAGGATGAGTTTGACTGCAAGGCCTGG
GCCTACTTCTCTGATGTGGACCTGGAGAAGGATGTGCACTCTGGCCTGATTGGCCCCCTGCTGGTGTGC
CACACCAACACCCTGAACCCTGCCCATGGCAGGCAGGTGACTGTGCAGGAGTTTGCCCTGTTCTTCACC
ATCTTTGATGAAACCAAGAGCTGGTACTTCACTGAGAACATGGAGAGGAACTGCAGGGCCCCCTGCAAC
ATCCAGATGGAGGACCCACCTTCAAGGAGAACTACAGGTTCCATGCCATCAATGGCTACATCATGGAC
ACCCTGCCTGGCCTGGTGATGGCCAGGACCAGAGGATCAGGTGGTACCTGCTGAGCATGGGCAGCAAT
GAGAACATCCACAGCATCCAATTCTCTGGCCATGTGTTCACTGTGAGGAAGAAGGAGGAGTACAAGATG
GCCCTGTACAACCTGTACCCTGGGGTGTGTGAGACTGTGGAGATGCTGCCCAGCAAGGCTGGCATCTGG
AGGGTGGAGTGCCTGATTGGGGAGCACCTGCATGCTGGCATGAGCACCTGTTCTCTGGTGTACAGCAAC
AAGTGCCAGACCCCCCTGGGCATGGCCTCTGGCCACATCAGGGACTTCCAGATCACTGCCTCTGGCCAG
TATGGCCAGTGGGCCCCCAAGCTGGCCAGGCTGCACTACTCTGGCAGCATCAATGCCTGGAGCACCAAG
GAGCCCTTCAGCTGGATCAAGGTGGACCTGCTGGCCCCATGATCATCCATGGCATCAAGACCCAGGGG
GCCAGGCAGAAGTTCAGCAGCCTGTACATCAGCCAGTTCATCATCATGTACAGCCTGGATGGCAAGAAG
TGGCAGACCTACAGGGGCAACAGCACTGGCACCTGATGGTGTCTTTGGCAATGTGGACAGCTCTGGC
ATCAAGCACAACATCTTCAACCCCCCATCATTGCCAGATACATCAGGCTGCACCCACCCACTACAGC
ATCAGGAGCACCTGAGGATGGAGCTGATGGGCTGTGACCTGAACAGCTGCAGCATGCCCTGGGCATG
GAGAGCAAGGCCATCTCTGATGCCCAGATCACTGCCAGCAGCTACTTCACCAACATGTTTGCCACCTGG
AGCCCCAGCAAGGCCAGGCTGCACCTGCAGGGCAGGAGCAATGCCTGGAGGCCCCAGGTCAACAACCCC
AAGGAGTGGCTGCAGGTGGACTTCCAGAAGACCATGAAGGTGACTGGGGTGACCACCCAGGGGGTGAAG
AGCCTGCTGACCAGCATGTATGTGAAGGAGTTCCTGATCAGCAGCAGCCAGGATGGCCACCAGTGGACC
CTGTTCTTCCAGAATGGCAAGGTGAAGGTGTTCCAGGGCAACCAGGACAGCTTCAACCCCTGTGGTGAAC
AGCCTGGACCCCCCCTGCTGACCAGATACCTGAGGATTACCCCCAGAGCTGGGTGCACCAGATTGCC
CTGAGGATGGAGGTGCTGGGCTGTGAGGCCAGGACCTGTACTGA (SEQ ID NO: 143)

[00369] In some embodiments, the ceDNA vector comprises an FVIII sequence that is at least 85% identical to the nucleic acid sequence as set forth in SEQ ID NO: 143. In some embodiments, the ceDNA vector comprises an FVIII sequence that is at least 90% identical to the nucleic acid sequence as set forth in SEQ ID NO: 143. In some embodiments, the ceDNA vector comprises an FVIII sequence that is at least 95% identical to the nucleic acid sequence as set forth in SEQ ID NO: 143. In some embodiments,

the ceDNA vector comprises an FVIII sequence that is at least 96% identical to the nucleic acid sequence as set forth in SEQ ID NO: 143. In some embodiments, the ceDNA vector comprises an FVIII sequence that is at least 97% identical to the nucleic acid sequence as set forth in SEQ ID NO: 143. In some embodiments, the ceDNA vector comprises an FVIII sequence that is at least 98% identical to the nucleic acid sequence as set forth in SEQ ID NO: 143. In some embodiments, the ceDNA vector comprises an FVIII sequence that is at least 99% identical to the nucleic acid sequence as set forth in SEQ ID NO: 143. In some embodiments, the ceDNA vector comprises an FVIII sequence that consists of SEQ ID NO: 143.

FVIII therapeutic proteins and uses thereof for the treatment of hemophilia A

[00370] The viral and non-viral vectors comprising the expression cassettes described herein can be used to deliver a liver-specific therapeutic protein (*e.g.*, a FVIII protein) for treatment of hemophilia A associated with inappropriate expression of the liver-specific therapeutic protein (*e.g.*, a FVIII protein) and/or mutations within the liver-specific therapeutic protein (*e.g.*, a FVIII protein).

[00371] The vectors as described herein can be used to express any desired FVIII therapeutic protein. Exemplary therapeutic FVIII therapeutic proteins include but are not limited to any therapeutic protein (*e.g.*, a FVIII protein), or portion thereof, expressed by, *e.g.*, a nucleic acid at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 143.

[00372] In one embodiment, the expressed FVIII therapeutic protein is functional for the treatment of a hemophilia A. In some embodiments, FVIII therapeutic protein does not cause an immune system reaction.

[00373] In another embodiment, the vectors encoding FVIII therapeutic protein or fragment thereof (*e.g.*, functional fragment) can be used to generate a chimeric protein. Thus, it is specifically contemplated herein that a vector expressing a chimeric protein can be administered to *e.g.*, to any one or more tissues selected from: liver, kidneys, gallbladder, prostate, adrenal gland. In some embodiments, when a vector that has been engineered to express FVIII is administered to an infant, or administered to a subject in utero, one can administer the vector to any one or more tissues selected from: liver, adrenal gland, heart, intestine, lung, and stomach, or to a liver stem cell precursor thereof for the *in vivo* or *ex vivo* treatment of hemophilia A.

Hemophilia

[00374] Hemophilia A is a genetic deficiency in clotting factor VIII, which causes increased bleeding and usually affects males. In the majority of cases, it is inherited as an X-linked recessive trait, though there are cases which arise from spontaneous mutations. In terms of the symptoms of hemophilia A, there are internal or external bleeding episodes. Individuals with more severe hemophilia suffer more severe and more frequent bleeding, while others with mild hemophilia typically suffer more minor symptoms except after surgery or serious trauma. Moderate hemophiliacs have variable symptoms which manifest along a spectrum between severe and mild forms.

[00375] Current treatments to prevent bleeding in people with hemophilia A involve Factor VIII medication. Most individuals with severe hemophilia require regular supplementation with intravenous

recombinant or plasma concentrate Factor VIII. Recombinant blood clotting factor VIII is one of the most complex proteins for industrial manufacturing due to the low efficiency of its gene transcription, massive intracellular loss of its proprotein during post-translational processing, and the instability of the secreted protein. Mild hemophiliacs can manage their condition with desmopressin, a drug which releases stored factor VIII from blood vessel walls.

[00376] There are many complications related to treatment of hemophilia A. In children, an easily accessible intravenous port can be inserted to minimize frequent traumatic intravenous cannulation. However, these ports are associated with high infection rate and a risk of clots forming at the tip of the catheter, rendering it useless. Viral infections can be common in hemophiliacs due to frequent blood transfusions which put patients at risk of acquiring blood borne infections, such as HIV, hepatitis B and hepatitis C. Prion infections can also be transmitted by blood transfusions. Another therapeutic complication of hemophilia A is the development of inhibitor antibodies against factor VIII due to frequent infusions. These develop as the body recognizes the infused factor VIII as foreign, as the body does not produce its own copy. In these individuals, activated factor VII, a precursor to factor VIII in the coagulation cascade, can be infused as a treatment for hemorrhage in individuals with hemophilia and antibodies against replacement factor VIII.

Coagulation Cascade

[00377] Coagulation, also known as clotting, is the process by which blood changes from a liquid to a gel, forming a blood clot. It potentially results in hemostasis, the cessation of blood loss from a damaged vessel, followed by repair. The mechanism of coagulation involves activation, adhesion and aggregation of platelets along with deposition and maturation of fibrin. Disorders of coagulation are disease states which can result in bleeding (hemorrhage or bruising) or obstructive clotting (thrombosis).

[00378] Coagulation begins almost instantly after an injury to the blood vessel has damaged the endothelium lining the blood vessel. Exposure of blood to the subendothelial space initiates two processes: changes in platelets, and the exposure of subendothelial tissue factor to plasma Factor VII, which ultimately leads to fibrin formation. Platelets immediately form a plug at the site of injury; this is called primary hemostasis. Secondary hemostasis occurs simultaneously: additional coagulation factors or clotting factors beyond Factor VII (including Factor VIII) respond in a complex cascade to form fibrin strands, which strengthen the platelet plug.

[00379] The coagulation cascade of secondary hemostasis has two initial pathways which lead to fibrin formation. These are the contact activation pathway (also known as the intrinsic pathway), and the tissue factor pathway (also known as the extrinsic pathway), which both lead to the same fundamental reactions that produce fibrin. The primary pathway for the initiation of blood coagulation is the tissue factor (extrinsic) pathway. The pathways are a series of reactions, in which a zymogen (inactive enzyme precursor) of a serine protease and its glycoprotein co-factor are activated to become active components that then catalyze the next reaction in the cascade, ultimately resulting in cross-

linked fibrin. Coagulation factors are generally indicated by Roman numerals, with a lowercase a appended to indicate an active form.

[00380] The coagulation factors are generally serine proteases (enzymes), which act by cleaving downstream proteins. The exceptions are tissue factor, FV, FVIII, FXIII. Tissue factor, FV and FVIII are glycoproteins, and Factor XIII is a transglutaminase. The coagulation factors circulate as inactive zymogens. The coagulation cascade is therefore classically divided into three pathways. The tissue factor and contact activation pathways both activate the “final common pathway” of factor X, thrombin and fibrin.

[00381] The main role of the tissue factor (extrinsic) pathway is to generate a “thrombin burst”, a process by which thrombin, the most important constituent of the coagulation cascade in terms of its feedback activation roles, is released very rapidly. FVIIa circulates in a higher amount than any other activated coagulation factor. The process includes the following steps:

[00382] Step 1: Following damage to the blood vessel, FVII leaves the circulation and comes into contact with tissue factor (TF) expressed on tissue-factor-bearing cells (stromal fibroblasts and leukocytes), forming an activated complex (TF-FVIIa).

[00383] Step 2: TF-FVIIa activates FIX and FX.

[00384] Step 3: FVII is itself activated by thrombin, FXIa, FXII and FXa.

[00385] Step 4: The activation of FX (to form FXa) by TF-FVIIa is almost immediately inhibited by tissue factor pathway inhibitor (TFPI).

[00386] Step 5: FXa and its co-factor FVa form the prothrombinase complex, which activates prothrombin to thrombin.

[00387] Step 6: Thrombin then activates other components of the coagulation cascade, including FV and FVIII (which forms a complex with FIX), and activates and releases FVIII from being bound to von Willebrand factor (vWF).

[00388] Step 7: FVIIIa is the co-factor of FIXa, and together they form the “tenase” complex, which activates FX; and so the cycle continues.

[00389] The contact activation (intrinsic) pathway begins with formation of the primary complex on collagen by high-molecular-weight kininogen (HMWK), prekallikrein, and FXII (Hageman factor). Prekallikrein is converted to kallikrein and FXII becomes FXIIa. FXIIa converts FXI into FXIa. Factor XIa activates FIX, which with its co-factor FVIIIa form the tenase complex, which activates FX to FXa. The minor role that the contact activation pathway has in initiating clot formation can be illustrated by the fact that patients with severe deficiencies of FXII, HMWK, and prekallikrein do not have a bleeding disorder. Instead, contact activation system is more involved in inflammation, and innate immunity.

[00390] The final common pathway shared by the intrinsic and extrinsic coagulation pathways involves the conversion of prothrombin into thrombin and fibrinogen into fibrin. Thrombin has a large array of functions, not only the conversion of fibrinogen to fibrin, the building block of a hemostatic plug. In addition, it is the most important platelet activator and on top of that it activates Factors VIII

and V and their inhibitor protein C (in the presence of thrombomodulin), and it activates Factor XIII, which forms covalent bonds that crosslink the fibrin polymers that form from activated monomers.

[00391] Following activation by the contact factor or tissue factor pathways, the coagulation cascade is maintained in a prothrombotic state by the continued activation of FVIII and FIX to form the tenase complex, until it is down-regulated by the anticoagulant pathways.

[00392] In some embodiments, a vector for expression of a therapeutic protein (*e.g.*, a FVIII protein) comprising an expression cassette as disclosed herein can also encode co-factors or other polypeptides, sense or antisense oligonucleotides, or RNAs (coding or non-coding; *e.g.*, siRNAs, shRNAs, micro-RNAs, and their antisense counterparts (*e.g.*, antagoMiR)) that can be used in conjunction with the Therapeutic protein (*e.g.*, a FVIII protein) expressed from the ceDNA. Additionally, expression cassettes comprising sequence encoding an Therapeutic protein (*e.g.*, a FVIII protein) can also include an exogenous sequence that encodes a reporter protein to be used for experimental or diagnostic purposes, such as β -lactamase, β -galactosidase (LacZ), alkaline phosphatase, thymidine kinase, green fluorescent protein (GFP), chloramphenicol acetyltransferase (CAT), luciferase, and others well known in the art.

[00393] In one embodiment, the ceDNA vector comprises a nucleic acid sequence to express the therapeutic protein (*e.g.*, a FVIII protein) that is functional for the treatment of hemophilia A. In a preferred embodiment, the therapeutic protein (*e.g.*, a FVIII protein) does not cause an immune system reaction, unless so desired.

VI. Pharmaceutical Compositions

[00394] In another aspect, pharmaceutical compositions are provided. The pharmaceutical composition comprises a ceDNA vector for expression of a therapeutic protein (*e.g.*, a FVIII protein) as described herein and a pharmaceutically acceptable carrier or diluent.

[00395] The viral and non-viral vectors for expression of a therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein can be incorporated into pharmaceutical compositions suitable for administration to a subject for *in vivo* delivery to cells, tissues, or organs of the subject. Typically, the pharmaceutical composition comprises a viral or non-viral vector (*e.g.*, an AAV vector, a ceDNA vector) as disclosed herein and a pharmaceutically acceptable carrier. For example, the vectors for expression of a therapeutic protein (*e.g.*, a FVIII protein) as described herein can be incorporated into a pharmaceutical composition suitable for a desired route of therapeutic administration (*e.g.*, parenteral administration). Passive tissue transduction via high pressure intravenous or intra-arterial infusion, as well as intracellular injection, such as intranuclear microinjection or intracytoplasmic injection, are also contemplated.

[00396] In one embodiment, pharmaceutical compositions for therapeutic purposes can be formulated as a solution, microemulsion, dispersion, liposomes, or other ordered structure suitable to high vector concentration, in particular, high ceDNA vector concentration. Sterile injectable solutions can be prepared by incorporating the vector compound in the required amount in an appropriate buffer

with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization including a vector can be formulated to deliver a transgene in the nucleic acid to the cells of a recipient, resulting in the therapeutic expression of the transgene or donor sequence therein. The composition can also include a pharmaceutically acceptable carrier.

[00397] Pharmaceutically active compositions comprising a vector (*e.g.*, an AAV vector, a ceDNA vector) for expression of a therapeutic protein (*e.g.*, a FVIII protein) can be formulated to deliver a transgene for various purposes to the cell, *e.g.*, cells of a subject.

[00398] Pharmaceutical compositions for therapeutic purposes typically must be sterile and stable under the conditions of manufacture and storage. The composition can be formulated as a solution, microemulsion, dispersion, liposomes, or other ordered structure suitable to high vector, in particular, high ceDNA vector concentration. Sterile injectable solutions can be prepared by incorporating the ceDNA vector compound in the required amount in an appropriate buffer with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization.

[00399] Ator for expression of therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein can be incorporated into a pharmaceutical composition suitable for topical, systemic, intra-amniotic, intrathecal, intracranial, intra-arterial, intravenous, intralymphatic, intraperitoneal, subcutaneous, tracheal, intra-tissue (*e.g.*, intramuscular, intracardiac, intrahepatic, intrarenal, intracerebral), intrathecal, intravesical, conjunctival (*e.g.*, extra-orbital, intraorbital, retroorbital, intraretinal, subretinal, choroidal, sub-choroidal, intrastromal, intracameral and intravitreal), intracochlear, and mucosal (*e.g.*, oral, rectal, nasal) administration. Passive tissue transduction via high pressure intravenous or intraarterial infusion, as well as intracellular injection, such as intranuclear microinjection or intracytoplasmic injection, are also contemplated.

[00400] In some aspects, the methods provided herein comprise delivering one or more vectors for expression of therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein to a host cell. Also provided herein are cells produced by such methods, and organisms (such as animals, plants, or fungi) comprising or produced from such cells. Methods of delivery of nucleic acids can include lipofection, nucleofection, microinjection, biolistics, liposomes, immunoliposomes, polycation or lipid:nucleic acid conjugates, naked DNA, and agent-enhanced uptake of DNA. Lipofection is described in *e.g.*, U.S. Pat. Nos. 5,049,386, 4,946,787; and 4,897,355, the contents of each of which are incorporated by reference in their entireties herein) and lipofection reagents are sold commercially (*e.g.*, TRANSFECTAM™ and LIPFECTIN™). Delivery can be to cells (*e.g.*, *in vitro* or *ex vivo* administration) or target tissues (*e.g.*, *in vivo* administration).

[00401] Various techniques and methods are known in the art for delivering nucleic acids to cells. For example, nucleic acids, such as ceDNA for expression of therapeutic protein (*e.g.*, a FVIII protein) can be formulated into lipid nanoparticles (LNPs), lipidoids, liposomes, lipid nanoparticles, lipoplexes, or core-shell nanoparticles. Typically, LNPs are composed of nucleic acid (*e.g.*, ceDNA) molecules, one or more ionizable or cationic lipids (or salts thereof), one or more non-ionic or neutral

lipids (*e.g.*, a phospholipid), a molecule that prevents aggregation (*e.g.*, PEG or a PEG-lipid conjugate), and optionally a sterol (*e.g.*, cholesterol).

[00402] Another method for delivering nucleic acids, such as ceDNA for expression of therapeutic protein (*e.g.*, a FVIII protein) to a cell is by conjugating the nucleic acid with a ligand that is internalized by the cell. For example, the ligand can bind a receptor on the cell surface and internalized via endocytosis. The ligand can be covalently linked to a nucleotide in the nucleic acid. Exemplary conjugates for delivering nucleic acids into a cell are described, example, in WO2015/006740, WO2014/025805, WO2012/037254, WO2009/082606, WO2009/073809, WO2009/018332, WO2006/112872, WO2004/090108, WO2004/091515 and WO2017/177326, the contents of each of which are incorporated by reference in their entireties herein.

[00403] Nucleic acids, such as ceDNA vectors for expression of therapeutic protein (*e.g.*, a FVIII protein) can also be delivered to a cell by transfection. Useful transfection methods include, but are not limited to, lipid-mediated transfection, cationic polymer-mediated transfection, or calcium phosphate precipitation. Transfection reagents are well known in the art and include, but are not limited to, TurboFect Transfection Reagent (Thermo Fisher Scientific), Pro-Ject Reagent (Thermo Fisher Scientific), TRANSPASS™ P Protein Transfection Reagent (New England Biolabs), CHARIOT™ Protein Delivery Reagent (Active Motif), PROTEOJUICE™ Protein Transfection Reagent (EMD Millipore), 293fectin, LIPOFECTAMINE™ 2000, LIPOFECTAMINE™ 3000 (Thermo Fisher Scientific), LIPOFECTAMINE™ (Thermo Fisher Scientific), LIPOFECTIN™ (Thermo Fisher Scientific), DMRIE-C, CELLFECTIN™ (Thermo Fisher Scientific), OLIGOFECTAMINE™ (Thermo Fisher Scientific), LIPOFECTACE™, FUGENE™ (Roche, Basel, Switzerland), FUGENE™ HD (Roche), TRANSFECTAM™ (Transfectam, Promega, Madison, Wis.), TFX-10™ (Promega), TFX-20™ (Promega), TFX-50™ (Promega), TRANSFECTIN™ (BioRad, Hercules, Calif.), SILENTFECT™ (Bio-Rad), Effectene™ (Qiagen, Valencia, Calif.), DC-chol (Avanti Polar Lipids), GENEPORTER™ (Gene Therapy Systems, San Diego, Calif.), DHARMAFECT 1™ (Dharmacon, Lafayette, Colo.), DHARMAFECT 2™ (Dharmacon), DHARMAFECT 3™ (Dharmacon), DHARMAFECT 4™ (Dharmacon), ESCORT™ III (Sigma, St. Louis, Mo.), and ESCORT™ IV (Sigma Chemical Co.). Nucleic acids, such as ceDNA, can also be delivered to a cell via microfluidics methods known to those of skill in the art.

[00404] Vectors (*e.g.*, AAV vectors or ceDNA vectors) for expression of therapeutic protein (*e.g.*, a FVIII protein) as described herein can also be administered directly to an organism for transduction of cells *in vivo*. Administration is by any of the routes normally used for introducing a molecule into ultimate contact with blood or tissue cells including, but not limited to, injection, infusion, topical application and electroporation. Suitable methods of administering such nucleic acids are available and well known to those of skill in the art, and, although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective reaction than another route.

[00405] Methods for introduction of a nucleic acid vector ceDNA vector for expression of therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein can be delivered into hematopoietic stem cells, for example, by the methods as described, for example, in U.S. Pat. No. 5,928,638, the contents of which is incorporated by reference in its entirety herein.

VII. Methods of Use

[00406] A non-viral or viral vector for expression of a therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein can also be used in a method for the delivery of a nucleic acid sequence of interest (*e.g.*, encoding a therapeutic protein (*e.g.*, a FVIII protein)) to a target cell (*e.g.*, a host cell). In some embodiments, the method comprises a method for delivering a therapeutic protein (*e.g.*, a FVIII protein) to a cell of a subject in need thereof and treating hemophilia A. The disclosure allows for the *in vivo* expression of the therapeutic protein (*e.g.*, a FVIII protein) encoded in the ceDNA vector in a cell in a subject such that therapeutic effect of the expression of the therapeutic protein (*e.g.*, a FVIII protein) occurs. These results are seen with both *in vivo* and *in vitro* modes of vector delivery.

[00407] In some embodiments, the disclosure provides a method for the delivery of a therapeutic protein (*e.g.*, a FVIII protein) in a cell of a subject in need thereof, comprising multiple administrations of the vector of the disclosure encoding said therapeutic protein (*e.g.*, a FVIII protein). In some embodiments, the ceDNA vectors of the disclosure do not induce an immune response like that typically observed against encapsidated viral vectors, such that a multiple administration strategy will likely have greater success in a ceDNA-based system. The ceDNA vector are administered in sufficient amounts to transfect the cells of a desired tissue and to provide sufficient levels of gene transfer and expression of the therapeutic protein (*e.g.*, a FVIII protein) without undue adverse effects.

[00408] The disclosure also provides for a method of treating hemophilia A in a subject comprising introducing into a target cell in need thereof (in particular a muscle cell or tissue) of the subject a therapeutically effective amount of a vector as described herein, optionally with a pharmaceutically acceptable carrier. While the vector can be introduced in the presence of a carrier, such a carrier is not required. The ceDNA vector selected comprises a nucleic acid sequence encoding an therapeutic protein (*e.g.*, a FVIII protein) useful for treating hemophilia A.

[00409] The compositions and vectors provided herein can be used to deliver a therapeutic protein (*e.g.*, a FVIII protein) for various purposes. In some embodiments, the transgene encodes an Therapeutic protein (*e.g.*, a FVIII protein) that is intended to be used for research purposes, *e.g.*, to create a somatic transgenic animal model harboring the transgene, *e.g.*, to study the function of the therapeutic protein (*e.g.*, a FVIII protein) product. In another example, the transgene encodes a therapeutic protein (*e.g.*, a FVIII protein) that is intended to be used to create an animal model of hemophilia A. In some embodiments, the encoded therapeutic protein (*e.g.*, a FVIII protein) is useful for the treatment or prevention of hemophilia A states in a mammalian subject. The therapeutic protein (*e.g.*, a FVIII protein) can be transferred (*e.g.*, expressed in) to a patient in a sufficient amount to treat hemophilia A associated with reduced expression, lack of expression or dysfunction of the gene.

[00410] In principle, the expression cassette can include a nucleic acid or any transgene that encodes a therapeutic protein (*e.g.*, a FVIII protein) that is either reduced or absent due to a mutation or which conveys a therapeutic benefit when overexpressed is considered to be within the scope of the disclosure. Preferably, noninserted bacterial DNA is not present and preferably no bacterial DNA is present in the ceDNA compositions provided herein.

[00411] In another aspect, multiple vectors expressing different proteins or the same therapeutic protein (*e.g.*, a FVIII protein) but operatively linked to different promoters or cis-regulatory elements can be delivered simultaneously or sequentially to the target cell, tissue, organ, or subject. Therefore, this strategy can allow for the gene therapy or gene delivery of multiple proteins simultaneously. It is also possible to separate different portions of a therapeutic protein (*e.g.*, a FVIII protein) into separate vectors (*e.g.*, different domains and/or co-factors required for functionality of a therapeutic protein (*e.g.*, a FVIII protein)) which can be administered simultaneously or at different times, and can be separately regulatable, thereby adding an additional level of control of expression of a therapeutic protein (*e.g.*, a FVIII protein).

[00412] The disclosure also provides for a method of treating hemophilia A in a subject comprising introducing into a target cell in need thereof (in particular a muscle cell or tissue) of the subject a therapeutically effective amount of a ceDNA vector as disclosed herein, optionally with a pharmaceutically acceptable carrier. While the ceDNA vector can be introduced in the presence of a carrier, such a carrier is not required. The ceDNA vector implemented comprises a nucleic acid sequence of interest useful for treating the hemophilia A. In particular, the ceDNA vector may comprise a desired exogenous DNA sequence operably linked to control elements capable of directing transcription of the desired polypeptide, protein, or oligonucleotide encoded by the exogenous DNA sequence when introduced into the subject. The ceDNA vector can be administered via any suitable route as provided above, and elsewhere herein.

VIII. Methods of Delivery

[00413] In some embodiments, non-viral and viral vector for expression of a therapeutic protein as described herein can be delivered to a target cell *in vitro* or *in vivo* by various suitable methods. Vectors alone can be applied or injected. According to embodiments, the vectors can be delivered to a cell without the help of a transfection reagent or other physical means. Alternatively, according to other embodiments, the vectors for expression of a therapeutic protein (*e.g.*, a FVIII protein) can be delivered using any art-known transfection reagent or other art-known physical means that facilitates entry of DNA into a cell, *e.g.*, liposomes, alcohols, polylysine- rich compounds, arginine-rich compounds, calcium phosphate, microvesicles, microinjection, electroporation and the like.

[00414] One aspect of the technology described herein relates to a method of delivering a therapeutic protein (*e.g.*, a FVIII protein) to a cell. Typically, for *in vivo* and *in vitro* methods, a non-viral or viral vector for expression of a therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein may be introduced into the cell using the methods as disclosed herein, as well as other methods known in the

art. A vector for expression of a therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein are preferably administered to the cell in a biologically-effective amount. If the vector is administered to a cell *in vivo* (*e.g.*, to a subject), a biologically-effective amount of the vector is an amount that is sufficient to result in transduction and expression of the therapeutic protein (*e.g.*, a FVIII protein) in a target cell.

[00415] Exemplary modes of administration of a vector composition for expression of therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein includes oral, rectal, transmucosal, intranasal, inhalation (*e.g.*, via an aerosol), buccal (*e.g.*, sublingual), vaginal, intrathecal, intraocular, transdermal, intraendothelial, in utero (or in ovo), parenteral (*e.g.*, intravenous, subcutaneous, intradermal, intracranial, intramuscular [including administration to skeletal, diaphragm and/or cardiac muscle], intrapleural, intracerebral, and intraarticular). Administration can be systemically or direct delivery to the liver or elsewhere (*e.g.*, any kidneys, gallbladder, prostate, adrenal gland, heart, intestine, lung, and stomach).

[00416] Administration can be topical (*e.g.*, to both skin and mucosal surfaces, including airway surfaces, and transdermal administration), intralymphatic, and the like, as well as direct tissue or organ injection (*e.g.*, but not limited to, liver, but also to eye, muscles, including skeletal muscle, cardiac muscle, diaphragm muscle, or brain).

[00417] Methods for introduction of a nucleic acid vector for expression of therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein can be delivered into hematopoietic stem cells, for example, by the methods as described, for example, in U.S. Pat. No. 5,928,638, the contents of which is incorporated by reference in its entirety herein.

[00418] Administration of the vectors described herein (*e.g.*, AAV, ccdNA) can be to any site in a subject, including, without limitation, a site selected from the group consisting of the liver and/or also eyes, brain, a skeletal muscle, a smooth muscle, the heart, the diaphragm, the airway epithelium, the kidney, the spleen, the pancreas, the skin.

[00419] The most suitable route in any given case will depend on the nature and severity of the condition being treated, ameliorated, and/or prevented and on the nature of the particular vector that is being used.

[00420] In one embodiment, delivery is to the liver. The vectors comprising the nucleic acids disclosed herein can be delivered to the liver through the hepatic artery, portal vein, or intravenously to produce therapeutic levels of therapeutic proteins or clotting factors in the blood. The capsid or vector is preferably suspended in a physiologically compatible transporter, and can be administered to a human or non-human mammalian patient. A person skilled in the art can easily select suitable transporters in view of the indication for which the transfer virus is directed. For example, a suitable carrier includes saline, which can be formulated with a variety of buffer solutions (*eg*, phosphate buffered saline). Other illustrative carriers include sterile saline, lactose, sucrose, calcium phosphate, gelatin, dextran, agar, pectin, sesame oil, and water.

[00421] In some embodiments, cells are removed from a subject, a ceDNA vector for expression of a therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein is introduced therein, and the cells are then replaced back into the subject. Methods of removing cells from subject for treatment *ex vivo*, followed by introduction back into the subject are known in the art (see, *e.g.*, U.S. Pat. No. 5,399,346; the disclosure of which is incorporated herein in its entirety). Alternatively, a ceDNA vector is introduced into cells from another subject, into cultured cells, or into cells from any other suitable source, and the cells are administered to a subject in need thereof.

[00422] Cells transduced with a ceDNA vector for expression of a therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein are preferably administered to the subject in a "therapeutically-effective amount" in combination with a pharmaceutical carrier. Those skilled in the art will appreciate that the therapeutic effects need not be complete or curative, as long as some benefit is provided to the subject.

[00423] In some embodiments, a ceDNA vector for expression of therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein can encode a therapeutic protein (*e.g.*, a FVIII protein) as described herein (sometimes called a transgene or heterologous nucleic acid sequence) that is to be produced in a cell *in vitro*, *ex vivo*, or *in vivo*. For example, in contrast to the use of the ceDNA vectors described herein in a method of treatment as discussed herein, in some embodiments a ceDNA vector for expression of Therapeutic protein (*e.g.*, a FVIII protein) may be introduced into cultured cells and the expressed Therapeutic protein (*e.g.*, a FVIII protein) isolated from the cells, *e.g.*, for the production of antibodies and fusion proteins. In some embodiments, the cultured cells comprising a ceDNA vector for expression of Therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein can be used for commercial production of antibodies or fusion proteins, *e.g.*, serving as a cell source for small or large scale biomanufacturing of antibodies or fusion proteins. In alternative embodiments, a ceDNA vector for expression of Therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein is introduced into cells in a host non-human subject, for *in vivo* production of antibodies or fusion proteins, including small scale production as well as for commercial large scale Therapeutic protein (*e.g.*, a FVIII protein) production.

[00424] The ceDNA vectors for expression of Therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein can be used in both veterinary and medical applications. Suitable subjects for *ex vivo* gene delivery methods as described above include both avians (*e.g.*, chickens, ducks, geese, quail, turkeys and pheasants) and mammals (*e.g.*, humans, bovines, ovines, caprines, equines, felines, canines, and lagomorphs), with mammals being preferred. Human subjects are most preferred. Human subjects include neonates, infants, juveniles, and adults.

Dose ranges

[00425] Provided herein are methods of treatment comprising administering to the subject an effective amount of a composition comprising a vector encoding a therapeutic protein (*e.g.*, a FVIII protein) as described herein. As will be appreciated by a skilled practitioner, the term "effective amount" refers to the amount of the composition administered that results in expression of the therapeutic protein (*e.g.*, a FVIII protein) in a "therapeutically effective amount" for the treatment of hemophilia A.

[00426] *In vivo* and/or *in vitro* assays can optionally be employed to help identify optimal dosage ranges for use. The precise dose to be employed in the formulation will also depend on the route of administration, and the seriousness of the condition, and should be decided according to the judgment of the person of ordinary skill in the art and each subject's circumstances. Effective doses can be extrapolated from dose-response curves derived from *in vitro* or animal model test systems.

[00427] s for expression of therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein is administered in sufficient amounts to transfect the cells of a desired tissue and to provide sufficient levels of gene transfer and expression without undue adverse effects. Conventional and pharmaceutically acceptable routes of administration include, but are not limited to, those described above in the "Administration" section, such as direct delivery to the selected organ (*e.g.*, intraportal delivery to the liver), oral, inhalation (including intranasal and intratracheal delivery), intraocular, intravenous, intramuscular, subcutaneous, intradermal, intratumoral, and other parental routes of administration. Routes of administration can be combined, if desired.

[00428] The dose of the amount of a vector for expression of a therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein required to achieve a particular "therapeutic effect," will vary based on several factors including, but not limited to: the route of nucleic acid administration, the level of gene or RNA expression required to achieve a therapeutic effect, the specific disease or disorder being treated, and the stability of the gene(s), RNA product(s), or resulting expressed protein(s). One of skill in the art can readily determine a vector dose range to treat a patient having a particular disease or disorder based on the aforementioned factors, as well as other factors that are well known in the art.

[00429] Dosage regime can be adjusted to provide the optimum therapeutic response. For example, the oligonucleotide can be repeatedly administered, *e.g.*, several doses can be administered daily, or the dose can be proportionally reduced as indicated by the exigencies of the therapeutic situation. One of ordinary skill in the art will readily be able to determine appropriate doses and schedules of administration of the subject oligonucleotides, whether the oligonucleotides are to be administered to cells or to subjects.

[00430] An FVIII therapeutic protein can be expressed in a subject for at least 1 week, at least 2 weeks, at least 1 month, at least 2 months, at least 6 months, at least 12 months/one year, at least 2 years, at least 5 years, at least 10 years, at least 15 years, at least 20 years, at least 30 years, at least 40 years, at least 50 years or more. Long-term expression can be achieved by repeated administration of the ceDNA vectors described herein at predetermined or desired intervals.

[00431] The duration of treatment depends upon the subject's clinical progress and responsiveness to therapy. Continuous, relatively low maintenance doses are contemplated after an initial higher therapeutic dose.

Unit dosage forms

[00432] In some embodiments, the pharmaceutical compositions comprising a viral or non-viral vector comprising an expression cassette as described herein, for expression of a therapeutic protein

(*e.g.*, a FVIII protein) as disclosed herein can conveniently be presented in unit dosage form. A unit dosage form will typically be adapted to one or more specific routes of administration of the pharmaceutical composition. In some embodiments, the unit dosage form is adapted for droplets to be administered directly to the eye. In some embodiments, the unit dosage form is adapted for administration by inhalation. In some embodiments, the unit dosage form is adapted for administration by a vaporizer. In some embodiments, the unit dosage form is adapted for administration by a nebulizer. In some embodiments, the unit dosage form is adapted for administration by an aerosolizer. In some embodiments, the unit dosage form is adapted for oral administration, for buccal administration, or for sublingual administration. In some embodiments, the unit dosage form is adapted for intravenous, intramuscular, or subcutaneous administration. In some embodiments, the unit dosage form is adapted for subretinal injection, suprachoroidal injection or intravitreal injection.

[00433] In some embodiments, the unit dosage form is adapted for intrathecal or intracerebroventricular administration. In some embodiments, the pharmaceutical composition is formulated for topical administration. The amount of active ingredient which can be combined with a carrier material to produce a single dosage form will generally be that amount of the compound which produces a therapeutic effect.

IX. Methods of Treatment

[00434] The technology described herein also demonstrates methods for making, as well as methods of using the disclosed viral and non-viral vectors for expression of a therapeutic protein in a variety of ways, including, for example, *ex vivo*, *ex situ*, *in vitro* and *in vivo* applications, methodologies, diagnostic procedures, gene editing and/or gene therapy regimens to treat a subject suffering from a genetic disorder.

[00435] According to some embodiments, the subject is a human. According to some embodiments, the genetic disorder is selected from the group consisting of sickle-cell anemia, melanoma, hemophilia A (clotting factor VIII (FVIII) deficiency) and hemophilia B (clotting factor IX (FIX) deficiency), cystic fibrosis (CFTR), familial hypercholesterolemia (LDL receptor defect), hepatoblastoma, Wilson disease, phenylketonuria (PKU), congenital hepatic porphyria, inherited disorders of hepatic metabolism, Lesch Nyhan syndrome, sickle cell anemia, thalassaemias, xeroderma pigmentosum, Fanconi's anemia, retinitis pigmentosa, ataxia telangiectasia, Bloom's syndrome, retinoblastoma, mucopolysaccharide storage diseases (*e.g.*, Hurler syndrome (MPS Type I), Scheie syndrome (MPS Type I S), Hurler-Scheie syndrome (MPS Type I H-S), Hunter syndrome (MPS Type II), Sanfilippo Types A, B, C, and D (MPS Types III A, B, C, and D), Morquio Types A and B (MPS IVA and MPS IVB), Maroteaux-Lamy syndrome (MPS Type VI), Sly syndrome (MPS Type VII), hyaluronidase deficiency (MPS Type IX)), Niemann-Pick Disease Types A/B, C1 and C2, Fabry disease, Schindler disease, GM2-gangliosidosis Type II (Sandhoff Disease), Tay-Sachs disease, Metachromatic Leukodystrophy, Krabbe disease, Mucopolipidosis Type I, II/III and IV, Sialidosis Types I and II, Glycogen Storage disease Types I and II (Pompe disease), Gaucher disease Types I, II and III, Fabry disease, cystinosis, Batten disease,

Aspartylglucosaminuria, Salla disease, Danon disease (LAMP-2 deficiency), Lysosomal Acid Lipase (LAL) deficiency, neuronal ceroid lipofuscinoses (CLN1-8, INCL, and LINCL), sphingolipidoses, galactosialidosis, amyotrophic lateral sclerosis (ALS), Parkinson's disease, Alzheimer's disease, Huntington's disease, spinocerebellar ataxia, spinal muscular atrophy, Friedreich's ataxia, Duchenne muscular dystrophy (DMD), Becker muscular dystrophies (BMD), dystrophic epidermolysis bullosa (DEB), ectonucleotide pyrophosphatase 1 deficiency, generalized arterial calcification of infancy (GACI), Leber Congenital Amaurosis, Stargardt macular dystrophy (ABCA4), ornithine transcarbamylase (OTC) deficiency, Usher syndrome, alpha-1 antitrypsin deficiency, progressive familial intrahepatic cholestasis (PFIC) type I (ATP8B1 deficiency), type II (ABCB11), type III (ABCB4), or type IV (TJP2) and Cathepsin A deficiency. According to some embodiments, the genetic disorder is Leber congenital amaurosis (LCA). According to some embodiments, the LCA is LCA10. According to some embodiments, the genetic disorder is Niemann-Pick disease. According to some embodiments, the genetic disorder is Stargardt macular dystrophy. According to some embodiments, the genetic disorder is glucose-6-phosphatase (G6Pase) deficiency (glycogen storage disease type I) or Pompe disease (glycogen storage disease type II). According to some embodiments, the genetic disorder is hemophilia A (Factor VIII deficiency). According to some embodiments, the genetic disorder is hemophilia B (Factor IX deficiency). According to some embodiments, the genetic disorder is hunter syndrome (Mucopolysaccharidosis II). According to some embodiments, the genetic disorder is cystic fibrosis. According to some embodiments, the genetic disorder is dystrophic epidermolysis bullosa (DEB). According to some embodiments, the genetic disorder is phenylketonuria (PKU). According to some embodiments, the genetic disorder is progressive familial intrahepatic cholestasis (PFIC). According to some embodiments, the genetic disorder is Wilson disease. According to some embodiments, the genetic disorder is Gaucher disease Type I, II or III.

[00436] In one embodiment, the expressed therapeutic protein (*e.g.*, a FVIII protein) expressed from a vector as disclosed herein is functional for the treatment of disease. In a preferred embodiment, the therapeutic protein (*e.g.*, a FVIII protein) does not cause an immune system reaction, unless so desired.

[00437] Provided herein is a method of treating hemophilia A in a subject comprising introducing into a target cell in need thereof (for example, a muscle cell or tissue, or other affected cell type) of the subject a therapeutically effective amount of a ceDNA vector for expression of therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein, optionally with a pharmaceutically acceptable carrier. While the vector can be introduced in the presence of a carrier, such a carrier is not required. The vector implemented comprises a nucleic acid sequence encoding a therapeutic protein (*e.g.*, a FVIII protein) as described herein useful for treating the disease. In particular, a ceDNA vector for expression of therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein may comprise a desired therapeutic protein (*e.g.*, a FVIII protein) DNA sequence operably linked to control elements capable of directing transcription of the desired therapeutic protein (*e.g.*, a FVIII protein) encoded by the exogenous DNA sequence when introduced into the subject. The ceDNA vector for expression of therapeutic protein

(*e.g.*, a FVIII protein) as disclosed herein can be administered via any suitable route as provided above, and elsewhere herein.

[00438] Disclosed herein are ceDNA vector compositions and formulations for expression of Therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein that include one or more of the ceDNA vectors of the present disclosure together with one or more pharmaceutically-acceptable buffers, diluents, or excipients. Such compositions may be included in one or more diagnostic or therapeutic kits, for diagnosing, preventing, treating or ameliorating one or more symptoms of hemophilia A. In one aspect the disease, injury, disorder, trauma or dysfunction is a human disease, injury, disorder, trauma or dysfunction.

[00439] Another aspect of the technology described herein provides a method for providing a subject in need thereof with a diagnostically- or therapeutically-effective amount of a viral or non-viral vector for expression of therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein, the method comprising providing to a cell, tissue or organ of a subject in need thereof, an amount of the vector as disclosed herein; and for a time effective to enable expression of the therapeutic protein (*e.g.*, a FVIII protein) from the vector thereby providing the subject with a diagnostically- or a therapeutically-effective amount of the therapeutic protein (*e.g.*, a FVIII protein) expressed by the vector. In a further aspect, the subject is human.

[00440] Another aspect of the technology described herein provides a method for diagnosing, preventing, treating, or ameliorating at least one or more symptoms of hemophilia A, a disorder, a dysfunction, an injury, an abnormal condition, or trauma in a subject. In an overall and general sense, the method includes at least the step of administering to a subject in need thereof one or more of the disclosed ceDNA vector for a therapeutic protein (*e.g.*, a FVIII protein) production, in an amount and for a time sufficient to diagnose, prevent, treat or ameliorate the one or more symptoms of the disease, disorder, dysfunction, injury, abnormal condition, or trauma in the subject. In such an embodiment, the subject can be evaluated for efficacy of the therapeutic protein (*e.g.*, a FVIII protein), or alternatively, detection of the therapeutic protein (*e.g.*, a FVIII protein) or tissue location (including cellular and subcellular location) of the therapeutic protein (*e.g.*, a FVIII protein) in the subject. As such, the ceDNA vector for expression of therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein can be used as an *in vivo* diagnostic tool, *e.g.*, for the detection of cancer or other indications. In a further aspect, the subject is human.

[00441] Another aspect is use of a viral or non-viral vector for expression of a therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein as a tool for treating or reducing one or more symptoms of hemophilia A or disease states. There are a number of inherited diseases in which defective genes are known, and typically fall into two classes: deficiency states, usually of enzymes, which are generally inherited in a recessive manner, and unbalanced states, which may involve regulatory or structural proteins, and which are typically but not always inherited in a dominant manner. For unbalanced disease states, a vector for expression of a therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein can be

used to create hemophilia A state in a model system, which could then be used in efforts to counteract the disease state. Thus, the vector for expression of a therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein permit the treatment of genetic diseases. As used herein, hemophilia A state is treated by partially or wholly remedying the deficiency or imbalance that causes the disease or makes it more severe.

[00442] As used herein, the term “therapeutically effective amount” is an amount of an expressed FVIII therapeutic protein, or functional fragment thereof that is sufficient to produce a statistically significant, measurable change in expression of a disease biomarker or reduction in a given disease symptom (see “Efficacy Measurement” below). Such effective amounts can be gauged in clinical trials as well as animal studies for a given ceDNA composition.

[00443] The efficacy of a given treatment for hemophilia A, can be determined by the skilled clinician. However, a treatment is considered “effective treatment,” as the term is used herein, if any one or all of the signs or symptoms of the disease or disorder is/are altered in a beneficial manner, or other clinically accepted symptoms or markers of disease are improved, or ameliorated, *e.g.*, by at least 10% following treatment with a viral or non-viral vector encoding FVIII, or a functional fragment thereof. Efficacy can also be measured by failure of an individual to worsen as assessed by stabilization of the disease, or the need for medical interventions (*i.e.*, progression of the disease is halted or at least slowed). Methods of measuring these indicators are known to those of skill in the art and/or described herein. Treatment includes any treatment of a disease in an individual or an animal (some non-limiting examples include a human, or a mammal) and includes: (1) inhibiting the disease, *e.g.*, arresting, or slowing progression of the disease or disorder; or (2) relieving the disease, *e.g.*, causing regression of symptoms; and (3) preventing or reducing the likelihood of the development of the disease, or preventing secondary diseases/disorders associated with the disease, such as liver or kidney failure. An effective amount for the treatment of a disease means that amount which, when administered to a mammal in need thereof, is sufficient to result in effective treatment as that term is defined herein, for that disease.

[00444] Efficacy of an agent can be determined by assessing physical indicators that are particular to hemophilia A. Standard methods of analysis of hemophilia A indicators are known in the art.

Host cells

[00445] In some embodiments, a non-viral or viral vector for expression of a therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein delivers the therapeutic protein (*e.g.*, a FVIII protein) transgene into a subject host cell.

[00446] In some embodiments, the cells are hepatic (*i.e.*, liver) cells.

[00447] In some embodiments, the cells are photoreceptor cells. In some embodiments, the cells are RPE cells. In some embodiments, the subject host cell is a human host cell, including, for example blood cells, stem cells, hematopoietic cells, CD34⁺ cells, cancer cells, vascular cells, muscle cells, pancreatic cells, neural cells, ocular or retinal cells, epithelial or endothelial cells, dendritic cells,

fibroblasts, or any other cell of mammalian origin, including, without limitation, lung cells, cardiac cells, pancreatic cells, intestinal cells, diaphragmatic cells, renal (*i.e.*, kidney) cells, neural cells, blood cells, bone marrow cells, or any one or more selected tissues of a subject for which gene therapy is contemplated. In one aspect, the subject host cell is a human host cell.

[00448] The present disclosure also relates to recombinant host cells as mentioned above, including a non-viral or viral vector as disclosed herein, for expression of a therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein. Thus, one can use multiple host cells depending on the purpose as is obvious to the skilled artisan. A construct or a vector for expression of a therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein including donor sequence is introduced into a host cell so that the donor sequence is maintained as a chromosomal integrant as described earlier. The term host cell encompasses any progeny of a parent cell that is not identical to the parent cell due to mutations that occur during replication. The choice of a host cell will to a large extent depend upon the donor sequence and its source.

[00449] The host cell may also be a eukaryote, such as a mammalian, insect, plant, or fungal cell. In one embodiment, the host cell is a human cell (*e.g.*, a primary cell, a stem cell, or an immortalized cell line). In some embodiments, the host cell can be administered a vector for expression of a therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein *ex vivo* and then delivered to the subject after the gene therapy event. A host cell can be any cell type, *e.g.*, a somatic cell or a stem cell, an induced pluripotent stem cell, or a blood cell, *e.g.*, T-cell or B-cell, or bone marrow cell. In certain embodiments, the host cell is an allogenic cell. For example, T-cell genome engineering is useful for cancer immunotherapies, disease modulation such as HIV therapy (*e.g.*, receptor knock out, such as CXCR4 and CCR5) and immunodeficiency therapies. MHC receptors on B-cells can be targeted for immunotherapy. In some embodiments, gene modified host cells, *e.g.*, bone marrow stem cells, *e.g.*, CD34⁺ cells, or induced pluripotent stem cells can be transplanted back into a patient for expression of a therapeutic protein.

Additional diseases for gene therapy

[00450] In general, a viral or non-viral vector as described herein for expression of a therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein can be used to deliver any therapeutic protein in accordance with the description above to treat, prevent, or ameliorate the symptoms associated with aberrant protein expression or gene expression in a subject.

[00451] In some embodiments, a viral or non-viral vector for expression of a therapeutic protein as disclosed herein can be used to deliver a therapeutic protein to skeletal, cardiac or diaphragm muscle, for production of a therapeutic protein for secretion and circulation in the blood or for systemic delivery to other tissues to treat, ameliorate, and/or prevent a disease or disorder characterized by aberrant gene expression.

Testing for successful gene expression using a ceDNA vector

[00452] Assays well known in the art can be used to test the efficiency of gene delivery of a therapeutic protein (*e.g.*, a FVIII protein) by a vector can be performed in both *in vitro* and *in vivo*

models. Levels of the expression of the therapeutic protein (*e.g.*, a FVIII protein) can be assessed by one skilled in the art by measuring mRNA and protein levels of the therapeutic protein (*e.g.*, a FVIII protein) (*e.g.*, reverse transcription PCR, western blot analysis, and enzyme-linked immunosorbent assay (ELISA)). In one embodiment, expression cassette comprises a reporter protein that can be used to assess the expression of the therapeutic protein (*e.g.*, a FVIII protein), for example by examining the expression of the reporter protein by fluorescence microscopy or a luminescence plate reader. For *in vivo* applications, protein function assays can be used to test the functionality of a given therapeutic protein (*e.g.*, a FVIII protein) to determine if gene expression has successfully occurred. One skilled will be able to determine the best test for measuring functionality of a therapeutic protein (*e.g.*, a FVIII protein) expressed by the ceDNA vector *in vitro* or *in vivo*.

[00453] It is contemplated herein that the effects of gene expression of a therapeutic protein (*e.g.*, a FVIII protein) from the vector in a cell or subject can last for at least 1 month, at least 2 months, at least 3 months, at least four months, at least 5 months, at least six months, at least 10 months, at least 12 months, at least 18 months, at least 2 years, at least 5 years, at least 10 years, at least 20 years, or can be permanent.

[00454] In some embodiments, a therapeutic protein (*e.g.*, a FVIII protein) in the expression cassette, expression construct, or non-viral or viral vector described herein can be codon optimized for the host cell. As used herein, the term “codon optimized” or “codon optimization” refers to the process of modifying a nucleic acid sequence for enhanced expression in the cells of the vertebrate of interest, *e.g.*, mouse or human (*e.g.*, humanized), by replacing at least one, more than one, or a significant number of codons of the native sequence (*e.g.*, a prokaryotic sequence) with codons that are more frequently or most frequently used in the genes of that vertebrate. Various species exhibit particular bias for certain codons of a particular amino acid. Typically, codon optimization does not alter the amino acid sequence of the original translated protein. Optimized codons can be determined using *e.g.*, Aptagen’s GENE FORGE® codon optimization and custom gene synthesis platform (Aptagen, Inc.) or another publicly available database.

Determining Efficacy by Assessing therapeutic protein Expression from the vector

[00455] Essentially any method known in the art for determining protein expression can be used to analyze expression of a therapeutic protein (*e.g.*, a FVIII protein) from a viral or non-viral vector. Non-limiting examples of such methods/assays include enzyme-linked immunoassay (ELISA), affinity ELISA, ELISPOT, serial dilution, flow cytometry, surface plasmon resonance analysis, kinetic exclusion assay, mass spectrometry, Western blot, immunoprecipitation, and PCR.

[00456] For assessing a therapeutic protein (*e.g.*, a FVIII protein) expression *in vivo*, a biological sample can be obtained from a subject for analysis. Exemplary biological samples include a biofluid sample, a body fluid sample, blood (including whole blood), serum, plasma, urine, saliva, a biopsy and/or tissue sample *etc.* A biological sample or tissue sample can also refer to a sample of tissue or fluid isolated from an individual including, but not limited to, tumor biopsy, stool, spinal fluid, pleural

fluid, nipple aspirates, lymph fluid, the external sections of the skin, respiratory, intestinal, and genitourinary tracts, tears, saliva, breast milk, cells (including, but not limited to, blood cells), tumors, organs, and also samples of *in vitro* cell culture constituent. The term also includes a mixture of the above-mentioned samples. The term "sample" also includes untreated or pretreated (or pre-processed) biological samples. In some embodiments, the sample used for the assays and methods described herein comprises a serum sample collected from a subject to be tested.

X. Various applications of viral and non-viral vectors

[00457] As disclosed herein, the viral and non-viral vectors for expression of a therapeutic protein as described herein can be used to express a therapeutic protein for a range of purposes. In one embodiment, the vector expressing a therapeutic protein (*e.g.*, a FVIII protein) can be used to create a somatic transgenic animal model harboring the transgene, *e.g.*, to study the function or disease progression of hemophilia A. In some embodiments, a cDNA vector expressing a therapeutic protein (*e.g.*, a FVIII protein) is useful for the treatment, prevention, or amelioration of hemophilia A states or disorders in a mammalian subject.

[00458] In some embodiments the therapeutic protein (*e.g.*, a FVIII protein) can be expressed from the vector in a subject in a sufficient amount to treat a disease associated with increased expression, increased activity of the gene product, or inappropriate upregulation of a gene.

[00459] In some embodiments the therapeutic protein (*e.g.*, a FVIII protein) can be expressed from the vector in a subject in a sufficient amount to treat hemophilia A with a reduced expression, lack of expression or dysfunction of a protein.

[00460] It will be appreciated by one of ordinary skill in the art that the transgene may not be an open reading frame of a gene to be transcribed itself; instead it may be a promoter region or repressor region of a target gene, and the cDNA vector may modify such region with the outcome of so modulating the expression of the FVIII gene.

[00461] The compositions and viral and non-viral vectors for expression of a therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein can be used to deliver a therapeutic protein (*e.g.*, a FVIII protein) for various purposes as described above.

[00462] In some embodiments, the transgene encodes one or more therapeutic proteins which are useful for the treatment, amelioration, or prevention of hemophilia A states in a mammalian subject. The therapeutic protein (*e.g.*, a FVIII protein) expressed by the vector is administered to a patient in a sufficient amount to treat hemophilia A associated with an abnormal gene sequence, which can result in any one or more of the following: increased protein expression, over activity of the protein, reduced expression, lack of expression or dysfunction of the target gene or protein.

[00463] In some embodiments, the vectors for expression of a therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein are envisioned for use in diagnostic and screening methods, whereby a therapeutic protein (*e.g.*, a FVIII protein) is transiently or stably expressed in a cell culture system, or alternatively, a transgenic animal model.

[00464] Another aspect of the technology described herein provides a method of transducing a population of mammalian cells with a ceDNA vector for expression of a therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein. In an overall and general sense, the method includes at least the step of introducing into one or more cells of the population, a composition that comprises an effective amount of one or more of the ceDNA vectors for expression of a therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein.

[00465] Additionally, the present disclosure provides compositions, as well as therapeutic and/or diagnostic kits that include one or more of the disclosed ceDNA vectors for expression of Therapeutic protein (*e.g.*, a FVIII protein) as disclosed herein or ceDNA compositions, formulated with one or more additional ingredients, or prepared with one or more instructions for their use.

[00466] A cell to be administered a ceDNA vector for expression of a therapeutic protein as disclosed herein may be of any type, including but not limited to neural cells (including cells of the peripheral and central nervous systems, in particular, brain cells), lung cells, retinal cells, epithelial cells (*e.g.*, gut and respiratory epithelial cells), muscle cells, dendritic cells, pancreatic cells (including islet cells), hepatic cells, myocardial cells, bone cells (*e.g.*, bone marrow stem cells), hematopoietic stem cells, spleen cells, keratinocytes, fibroblasts, endothelial cells, prostate cells, germ cells, and the like. Alternatively, the cell may be any progenitor cell. As a further alternative, the cell can be a stem cell (*e.g.*, neural stem cell, liver stem cell). As still a further alternative, the cell may be a cancer or tumor cell. Moreover, the cells can be from any species of origin, as indicated above.

Production and Purification of ceDNA vectors expressing a therapeutic protein

[00467] The viral and non-viral vectors disclosed herein are to be used to produce a therapeutic protein (*e.g.*, a FVIII protein) either *in vitro* or *in vivo*. The therapeutic protein (*e.g.*, a FVIII protein) that is produced in this manner can be isolated, tested for a desired function, and purified for further use in research or as a therapeutic treatment.

[00468] Each system of protein production has its own advantages/disadvantages. While proteins produced *in vitro* can be easily purified and can proteins in a short time, proteins produced *in vivo* can have post-translational modifications, such as glycosylation.

[00469] A therapeutic protein produced using viral and non-viral vectors described herein can be purified using any method known to those of skill in the art, for example, ion exchange chromatography, affinity chromatography, precipitation, or electrophoresis.

[00470] A therapeutic protein produced by the methods and compositions described herein can be tested for binding to the desired target protein.

[00471] The technology described herein is further illustrated by the following examples which in no way should be construed as being further limiting. It should be understood that this disclosure is not limited to the particular methodology, protocols, and reagents, *etc.*, described herein and as such can vary. The terminology used herein is for the purpose of describing particular embodiments only and is not intended to limit the scope of the present disclosure, which is defined solely by the claims.

EXAMPLES

[00472] The following examples are provided by way of illustration not limitation.

Example 1: *In silico* analyses for identification of potentially improved human liver-specific promoter

[00473] The human SERPINA1 enhancer (hSerpEnh) is often used to drive liver-specific gene expression (Chuah *et al.* (2014) Mol Ther 22(9): 1605–1613). Multiple bioinformatic analyses were used to inform modification of the hSerpEnh for improved function and are described below.

Analysis of evolutionary conservation

[00474] *Cis*-regulatory regions with similar sequence and similar sequence contexts often have conserved function but distinct performance attributes. A curated collection of more than 100 vertebrate genomes were analyzed to identify a set of predicted functionally conserved enhancers with divergent sequence. The function of a range of these enhancer elements were assessed to identify higher-expressing modules. Selections were also prioritized based on amount of CpG content and poly C and poly G sequence motifs.

[00475] 20 homologous sequences of human SERPINA1 enhancer region were identified and selected (see **FIGs. 1A-1B**) to screen for SerpEnh variants with improved expression characteristics. These sequences are listed in **Table 4**. **FIG. 1A** and **FIG. 1B** depict SERPINA1 sequences and alignment of conserved enhancer regions of human and 20 other vertebrates. 115 non-human vertebrate genomes were initially assessed for conserved SERPINA1 enhancer regions using the UCSC multiz100way and multiz30way multiple alignments. Depicted in FIGs 1A-1B are the conserved SERPINA1 enhancer regions from 20 vertebrates with > 90% identity to the human SERPINA1 enhancer, which are mapped to the human SERPINA1 enhancer sequence with Geneious. Highlighted nucleotides in the aligned sequences represent differences from the human reference sequence.

Identification and modification of non-Consensus Transcription Factor Binding Sites (TFBS)

[00476] Transcription factor binding sites can be identified by *in silico* analysis and represent one sequence across a family of possible functional sites that is often divergent from the known consensus sequence. Several important liver-specific transcription factor binding sites were identified that diverged from consensus.

The hSerpEnh contains near-consensus binding sites for many transcription factors, including HNF4 and FOXA, which are key regulators of hepatic gene expression (*see FIG. 2*). Orange arrows represent TF binding motifs described by Chuah *et al.* (2014). Red arrows represent TF binding motifs identified by our independent analysis (**FIG. 2**). FIMO (Bailey *et al.* (2009) Nucleic Acids Res 37: W202-W208) was used to scan the human SERPINA1 enhancer sequence with position weight matrices for TFs generated by the ENCODE Project. Representative motif matches with $p < 1e-4$ are displayed. Sequence logos representing position weight matrices for FOXA (JASPAR MA0148.3), ERR2 (HOCOMOCO ERR2_HUMAN.H11MO.0.A), and HNFA (HNF4A_HUMAN.H11MO.0.A) are shown above the

corresponding motifs identified in our analysis (**FIG. 2**). Positions where the human SERPINA1 sequence differs from the most highly preferred nucleotide in the sequence logos are boxed and highlighted in red (**FIG. 2**).

[00477] Off- consensus nucleotides were modified to reinstate the consensus sequences based on the hypothesis that they would result in higher affinity for the transcription factor and drive higher levels of transcription initiation.

CpG content minimization

[00478] Promoters often contain CpG dinucleotides that are undesirable for gene therapy applications. CpGs can impact expression durability through stimulation of the innate immune system and through methylation-based silencing. Nevertheless, removal of CpGs from *cis*-regulatory regions is non-trivial as they often play important functional roles in driving expression.

Multiple bioinformatic analyses were employed to inform removal of CpG from hSerpEnh (*i.e.*, CpG ablation) (see **FIG. 3**). The evolutionary conservation analysis provided a rational path for selective removal of CpGs in the enhancer without disrupting function. Enhancer regions from diverse species that did not contain some or all CpGs but were likely to maintain function were identified. The human SERPINA1 enhancer contains one internal CpG and the potential to form CpGs at its 5' and 3' ends (highlighted in red and boxed in the "hSerpEnh" track). Low sequence conservation, the presence of human SNPs that are not known to be associated with disease, and the absence of predicted transcription factor (TF) binding sites were assessed to inform sequence changes to ablate the central CpG and the remove potential for CpG formation at the ends of the sequence.

[00479] Further, it was hypothesized that positions within the human SERPINA1 enhancer that are poorly conserved between species are less consequential to the enhancer's function, and thus better targets for sequence modifications that lead to CpG ablations. To assess sequence conservation at each position in the human SERPINA1 enhancer, 115 non-human vertebrate genomes were evaluated for conserved SERPINA1 enhancer elements in the UCSC multiz100way and multiz30way multiple alignments. Of these 115 genomes, 43 contained conserved SERPINA1 enhancer regions, which were aligned using the MUSCLE alignment algorithm. The "Conservation" track displays the mean pairwise identity between all pairs of nucleotides in the aligned sequences at each position in the enhancer. Green bars represent 100% identity and dark yellow bars represent 30 to < 100% identity. Nucleotides that differed from the human sequence, but were utilized in the aligned position in other genomes were preferentially used for CpG ablations. It was further hypothesized that positions in CpGs that contain non-disease associated SNPs in the human population would be preferable targets for CpG ablation, and that changing the sequence to match non-disease associated SNPs would minimize changes to enhancer function. The "SNPs" track depicts SNPs within the human SERPINA1 enhancer that are cataloged in the 1000 Genomes Project and dbSNP that are not known to be associated with disease. Changes were made to ensure that the SERPINA1 sequence to ablate CpGs did not interfere with predicted transcription binding (TF) sites. The top track depicts selected TF motifs from our motif

analysis (described in **FIG. 2**), in which it was found that TF motif clusters and the motifs for TFs that play key roles in hepatic expression had minimal overlap with the internal CpG.

[00480] CpG-free elements were either tested directly for function or used as a reference for making functionally permissive substitutions in the native human enhancer region.

[00481] The variants of SerpEnh generated from the bioinformatic analyses above are listed in the tables below, *e.g.*, **Table 4**.

[00482] The results are described in the following Examples.

Table 3: Selected conserved SERPINA1 enhancer variants from human and 20 other vertebrates

Name	Species	SERPINA1 enhancer region sequence	SEQ ID NO:
hSerpEnh	Human	GGGGAGGCTGCTGGTGAATATTAAACCAAGGTCAACCCAGTTATC GGAGGAGCAAAACAGGGCTAAGTCCAC	137
SerpEnh_Rhesus	Rhesus	GGGGAGGCTGCTGGTGAATATTAAACCAAGGTCAACCCAGTTACC GGAGGAGCAAAACAGGGACTAAGTCCAC	117
SerpEnh_Squirrel_monkey	Squirrel monkey	GGGGATGCTGCTGGTGAATATTAAACCAAGGTCAACCCAGTTACC GGAGGAGCAAAACAGGGCTAAGTCCAC	118
SerpEnh_Bactrian_camel	Bactrian camel	GGGGAGGCTGCTGGTGAATATTAAACCAAGGTCAACCCAGTTATC GGAGGAGCAAAACAGGGACTAAGTCCAT	119
SerpEnh_Ferret	Ferret	GGGGAGGCTGCTGGTGAATATTAACTAAGGTCAACCCAGTTATC GGAGGAGCAAAACAGGGACTAAGTCCAG	120
SerpEnh_Mouse_lemur	Mouse lemur	GAGGGAGGCGCTGGTGAATATTAAACCAAGGTCAACCCAGTTATCG GGAGGCAAAACAGGGCTAAGTCCAT	121
SerpEnh_Chinese_tree_shrew	Chinese tree shrew	GGAGGCTGTTGGTGAATATTAAACCAAGGTCAACCCAGTTATCGGA GGAGCAAAACAGGGCTAAGTCCAC	82
SerpEnh_Prairie_vole	Prairie vole	GGGGAGTCTGCTAGTGAATATTAAACCAAGGTCAACCCAGTTATC GGAGGAGCAAAACAGAGAGGGACTAAGTCCAT	123
SerpEnh_Cat	Cat	GGGGAGGCTGCTGGTGAATATTAACTAAGGTCAACCCAGTTATC AGAGGAGCAAAATAGGGACTAAGTCCAT	124
SerpEnh_Panda	Panda	GGGGAGGCTGCTGGTGAATATTAACTAAGGTCAACCCAGTTATC AGAGGAGCAAAACAGGGACTAAGTCCAG	125
SerpEnh_David's_myotis	David's myotis	GGGGAGGCTGCTGGTGAATATTAAACCAAGGTCAACCTCAGTTATC AGAGGAGCAAAACAGGGACTAAGTCCAT	126
SerpEnh_Coquerel's_sifaka	Coquerel's sifaka	GAGGGAGGCGCTGCTGGTGAATATTAAACCAAGGTCAACCCAGTTATCG GGGAGCAAAACAGGGCTAAGTCCAT	127
SerpEnh_Dog	Dog	GGGGTGGTGGTGGTGAATATTAAACCAAGGTCAACCCCGTTATC GGAGGAGCAAAACAGGGACTAAGTCCAT	128
SerpEnh_Armadillo	Armadillo	GGGGAGGCTGCGAGTGAACATTAAACCAAGGTCAACCCAGTTATCA GAGGAGCAAAACAGGGACTAAGTCCAC	129
SerpEnh_Dolphin	Dolphin	GTGGGAGGCTGCTGGTGAATATTAAACCAAGGTCAACCCAGTTATC AGAGGAGTAAACAGGGACTAAGTCCAC	130

SerpEnh_Bushbaby	Bushbaby	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCAACCCAGTTATCA GGGAGCAACACAGGAGCTAAGTCCAT	131
SerpEnh_Lesser_Egyptian_jerboa	Lesser Egyptian jerboa	GGGGAATCTGCTAGTGAATATTAAACCAAGGTCCCGCAGTTATTG GAGGAGCAACACAGCAGGAGCTAAGTCCAA	132
SerpEnh_Rabbit	Rabbit	GGGCGAGCTGCAGGTGAATATTAAACCAAGGTCAACGCCAGTTATCG GAGGAGCAACACAGGAGTTAAGTCCAC	133
SerpEnh_Tibetan_antelope	Tibetan antelope	GGGGGAGGCTGCTGGTAAACATTAAACCAAGGTCAACCCAGTTATC AGAGGAACAAACAAGGACTAAGTCCAT	134
SerpEnh_Big_brown_bat	Big brown bat	GGGGGAGGCTGCTGGTGAATATTAAACCAAGGTCAACTCAGTTATC AGAGGAGCAACACAGGACTAAGTCCAT	135
SerpEnh_Starnosed_mole	Starnosed mole	TGGGGAGGCTGCTGGTGAATATTAACTAAGGTCACTCCAGTTATC TGGGGAGCAACACAGGGACTAAGTCCAT	136

Table 4: Variants of human SERPINA1 enhancer (hSerpEnh) based on bioinformatic analyses

Name	Description	SERPINA1 enhancer region sequence	SEQ ID NO:
hSerpEnh_100_vertebrate_consensus_v1	hSerpEnh with modifications based on the consensus sequence from the UCSC 100-vertebrate and 27-primate multiple alignments version 1	GGGGGAGGCTGCTGGTGAATATTAAACCAAGGTCAACCCCA GTTATCAGAGGAGCAACACAGGGACTAAGTCCAT	111
hSerpEnh_100_vertebrate_consensus_v2	hSerpEnh with modifications based on the consensus sequence from the UCSC 100-vertebrate and 27-primate multiple alignments version 2	GGGGGAGGCTGCTGGTGAATATTAAACCAAGGTCAACCCCA GTTACCAGAGGAGCAACACAGGGACTAAGTCCAT	112
hSerpEnh_100_vertebrate_consensus_v3	hSerpEnh with modifications based on the consensus sequence from the UCSC 100-vertebrate and 27-primate multiple alignments version 3	GGGGGAGGCTGCTGGTGAATATTAAACCAAGGTCAACCCCA GTTATCAGAGGAGCAACACAGGGACTAAGTTCAT	113
hSerpEnh_100_vertebrate_consensus_v4	hSerpEnh with modifications based on the consensus sequence from the UCSC 100-vertebrate and 27-primate multiple alignments version 4	AGGGGAGGCTGCTGGTGAATATTAAACCAAGGTCAACCCCA GTTATCAGAGGAGCAACACAGGGACTAAGTCCAT	114

hSerpEnh_FOXA_ consensus_v1	HSerpEnh with FOXA consensus site version 1	AGGGAGGCTGCTGGTAAACATTAACCAAGGTCAACCCCA GTTATCAGAGGAGCAAAACAGGGGCTAAGTCCAC	86
hSerpEnh_FOXA_ consensus_v2	HSerpEnh with FOXA consensus site version 2	AGGGAGGCTGCTGGTAAATATTAACCAAGGTCAACCCCA GTTATCAGAGGAGCAAAACAGGGGCTAAGTCCAT	87
hSerpEnh_FOXA_H NF4_consensus_v1	HSerpEnh with FOXA & HNF4 consensus sites version 1	GGGGAGGCTGCTGGTAAACATTAACCAAGGTCAACCCCA GTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC	88
hSerpEnh_FOXA_H NF4_consensus_v2	HSerpEnh with FOXA & HNF4 consensus sites version 2	AGGGAGGCTGCTGGTAAATATTAACCAAGGTCAACCCCA GTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC	89
hSerpEnh_HNF1_ consensus_v1	HSerpEnh with HNF1 consensus site version 1	AGGGAGGCTGCTGGTAAATGATTAACTAAGGTCAACCCCA AGTTATCAGAGGAGCAAAACAGGGGCTAAGTCCAC	90
hSerpEnh_HNF1_ consensus_v2	HSerpEnh with HNF1 consensus site version 2	AGGGAGGCTGCTGGTAAATGATTAACTAAGGTCAACCCCA AGTTATCAGAGGAGCAAAACAGGGGCTAAGTCCAC	91
hSerpEnh_HNF1_H NF4_consensus_v1	HSerpEnh with HNF1 & HNF4 consensus sites version 1	GGGGAGGCTGCTGGTAAATGATTAACTAAGGTCAACCCCA AGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC	92
hSerpEnh_HNF1_H NF4_consensus_v2	HSerpEnh with HNF1 & HNF4 consensus sites version 2	GGGGAGGCTGCTGGTAAATGATTAACTAAGGTCAACCCCA AGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC	93
hSerpEnh_HNF4_ consensus_v1	HSerpEnh with HNF4 consensus site version 1	AGGGAGGCTGCTGGTGAATATTAACCAAGGTCAACCCCA GTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAT	94
hSerpEnh_HNF4_ consensus_v2	HSerpEnh with HNF4 consensus site version 2	AGGGAGGCTGCTGGTGAATATTAACCAAGGTCAACCTCA GTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC	95
hSerpEnh_human_S NPs_v1	HSerpEnh with modifications based on non- disease associated human SNPs version 1	AGAGGAGGCTGCTGGTGAATATTAACCAAGGTCAACCCCA GTTATCAGAGGAGCAAAACAGGGGCTAAGTCCAC	96
hSerpEnh_human_S NPS_v2	HSerpEnh with modifications based on non- disease associated human SNPs version 2	AGAGAAGGCTGCTGGTGAATATTAACCAAGGTCAACCCCA GTTATCGGAGGAGCAAAACAGGGGCTAAGTCCAC	97
hSerpEnh_low_TFBS _and_end_regions_v1	HSerpEnh with modifications to end regions and regions with fewer predicted transcription factor binding sites version 1	GGGGAGGCTGCTGGTGAATATTAACCAAGGTCAAGCACA GTTATCAGAGGAGCAAAACAGGGGCTAAGTCCAT	98
hSerpEnh_low_TFBS _and_end_regions_v2	HSerpEnh with modifications to end regions and regions with fewer predicted transcription factor binding sites version 2	GGGGAGGCTGCTGGTGAATATTAACCAAGGTCAACTCA GTTATCAGAGGAGCAAAACAGGGGCTAAGTCCAT	99
hSerpEnh_low_TFBS _and_end_regions_v3	HSerpEnh with modifications to end regions and regions with fewer predicted transcription factor binding sites version 3	AGGGAGGCTGCTGGTGAATATTAACCAAGGTCAACCTCA GTTATCAGAGGAGCAAAACAGGGGCTAAGTCCAT	100

hSerpEnh_low_TFBS _and_end_regions_v4	H_SerpEnh with modifications to end regions and regions with fewer predicted transcription factor binding sites version 4	AGGGAGGGCTGCTGGTGAATATTAAACCAAGGTCAGCCCCA GTTATTGGAGGAGCAAAACAGGGGCTAAGTCCAT	101
hSerpEnh_low_TFBS _and_end_regions_v5	H_SerpEnh with modifications to end regions and regions with fewer predicted transcription factor binding sites v5	AGGGAGGGCTGCTGGTGAATATTAAACCAAGGTCAGCCCCA GTTATTAGAGGAGCAAAACAGGGGCTAAGTCCAT	102
hSerpEnh_low_TFBS _and_end_regions_v6	H_SerpEnh with modifications to end regions and regions with fewer predicted transcription factor binding sites v6	AGGGAGGGCTGCTGGTGAATATTAAACCAAGGTCAGCCCCA GTTACTGGAGGAGCAAAACAGGGGCTAAGTCCAT	103
hSerpEnh_low_TFBS _region_v1	H_SerpEnh with modifications to regions with fewer predicted transcription factor binding sites version 1	GGGGAGGGCTGCTGGTGAATATTAAACCAAGGTCAGCACA GTTACCAGAGGAGCAAAACAGGGGCTAAGTCCAC	104
hSerpEnh_low_TFBS _region_v2	H_SerpEnh with modifications to regions with fewer predicted transcription factor binding sites version 2	GGGGAGGGCTGCTGGTGAATATTAAACCAAGGTCAGTCA GTTACCAGAGGAGCAAAACAGGGGCTAAGTCCAC	105
hSerpEnh_low_TFBS _region_v3	H_SerpEnh with modifications to regions with fewer predicted transcription factor binding sites version 3	GGGGAGGGCTGCTGGTGAATATTAAACCAAGGTCAGCCCCA GTTACTAGGGGAGCAAAACAGGGGCTAAGTCCAC	106
hSerpEnh_low_TFBS _region_v4	H_SerpEnh with modifications to regions with fewer predicted transcription factor binding sites version 4	GGGGAGGGCTGCTGGTGAATATTAAACCAAGGTCAGCCCCA GTTACTAGAGGAACAAACAGGGGCTAAGTCCAC	107
hSerpEnh_low_TFBS _region_v5	H_SerpEnh with modifications to regions with fewer predicted transcription factor binding sites v5	GGGGAGGGCTGCTGGTGAATATTAAACCAAGGTCAGCCCCA GTTATTAGGGGAACAAACAGGGGCTAAGTCCAC	108
hSerpEnh_11_NHP_ consensus_v1	H_SerpEnh with modifications based on the Rhesus SERPINA1 enhancer sequence, which is shared with at least 10 other non-human primate version 1	GGGGAGGGCTGCTGGTGAATATTAAACCAAGGTCAGCCCCA GTTACCAGAGGAGCAAAACAGGGACTAAGTTTAC	109
hSerpEnh_11_NHP_ consensus_v2	H_SerpEnh with modifications based on the Rhesus SERPINA1 enhancer sequence, which is shared with at least 10 other non-human primate version 2	GGGGAGGGCTGCTGGTGAATATTAAACCAAGGTCAGCCCCA GTTACCCGAGGAGCAAAACAGGGACTAAGTTTAT	110
hSerpEnh_end_region ns_v1	H_SerpEnh with modifications to end regions version 1	AAGGGAGGGCTGCTGGTGAATATTAAACCAAGGTCAGCCCCA GTTATCGGAGGAGCAAAACAGGGGCTAAGTTTAT	115

hSerpEnh_end_regio ns_ v2	H SerpEnh with modifications to end regions version 2	AAGGAGGCTGCTGGTGAATATTAAACCAAGGTCACCCCA GTTATCGGAGGAGCAAAACAGGGACTAAGTCCAT	116
---------------------------------	---	---	-----

Example 2: *In vitro* activity of single variants of human SERPINA1 enhancer*Dual luciferase transient transfection assay*

[00483] Promega ViaFect Transfection: Efficacy of enhancer variants were evaluated *in vitro* using luciferase reporter assays. Expression plasmids containing enhancer variants were transfected into HepG2 cells using Promega ViaFect Transfection. Briefly, 24 hr before beginning transfections, 25,000-30,000 HepG2 cells/well were seeded in 96-well collagen-I coated plates in 100 uL DMEM + 10% FBS and incubated at 37C with 5% humidity. DNA master mixes for each experimental plasmid to be transfected were prepared. Transfections were performed in triplicate (3 wells/plasmid), unless otherwise noted. For each well to be transfected, 1 ng NanoLuc pNL1.1.TK[Nluc/TK] plasmid, 67 ng experimental firefly luciferase plasmid and 133 ng pGEM@-3Zf(-) carrier plasmid were mixed and brought to a volume of 9.2 uL with Opti-MEM. Next, 0.8 uL room temperature ViaFect per well was added to each DNA master mix and incubated 5 - 20 min at room temperature. Each well was transfected with 10 uL of ViaFect/Opti-MEM/NanoLuc mastermix and incubated at 37°C with 5% humidity for 24 hr prior to performing luminescence assay. Benchmarking plasmids, 1x hSerpEnh-Firefly luciferase or 3x hSerpEnh-Firefly luciferase, were included on every plate.

[00484] Promega NanoGlo Dual Luciferase Assay: 24 hr post-transfection, media was replaced with 80 uL room temperature PBS to prevent phenol red from interfering with the assay. Plates were allowed to equilibrate to room temperature. ONE-Glo™ EX Reagent was prepared as follows: the contents of one bottle of ONE-Glo™ EX Luciferase Assay Buffer was transferred to one bottle of ONE-Glo™ EX Luciferase Assay Substrate and mixed by inversion until the substrate was thoroughly dissolved. 80 uL of ONE-Glo™ EX Reagent was added to each well. Samples were mixed by shaking on an orbital shaker for 3 min at 500 rpm. 140 uL of lysed cells was transferred to a white 96-well plate to minimize cross-talk between wells and absorption of the emitted light. Firefly luciferase luminescence was measured on a SpectraMax M5 plate reader. The NanoDLR™ Stop & GloR Substrate was diluted 1:100 into an appropriate volume of room-temperature NanoDLR™ Stop & GloR Buffer and mixed by inversion. 70 uL of NanoDLR™ Stop & GloR Reagent was added to each well, shaken for 3 min at 700 rpms and incubated for an additional 7 min. NanoLuc luminescence was measured on SpectraMax M5 plate reader.

Screening for single enhancer variants that outperforms the human SERPINA1 enhancer

[00485] In a first round of screening, 30 variants including 20 conserved sequences from other organisms and 10 TFBS consensus variants which were placed in a plasmid were tested *in vitro* using the luciferase reporter assay as described herein. Data from the top 11 constructs are shown in **FIG. 4A**. **FIG. 4A** depicts results of the top 11 constructs in a screen of 30 single (1x) variants using luciferase reporter assay (n=3). Results are grouped by rationally designed enhancer variants (1x TFBS Consensus Variants) or conserved SERPINA1 enhancer regions identified in other species (1x Conserved Genomic Variants). The human SERPINA1 enhancer is shown far left. Error bars represent standard deviation.

As shown in **FIG. 4A**, in a plasmid, hSerpEnh_FOXA_HNF4_consensus_v1, performed almost 2 times better than the benchmark hSerpEnh.

[00486] FIG. 4B depicts the sequence design of the top variant in this screen, hSerpEnh_FOXA_HNF4_consensus_v1. The top variant was designed by modifying the FOXA and HNF4 motifs identified in the human SERPINA1 enhancer to match their respective consensus sequences (GTGAATA to GTAAACA for FOXA and CTAAGT to CAAACT for HNF4). The internal CpG was ablated by changing the G, which both has lower sequence conservation than the C and is at the position of a human SNP, to an A to match the SNP. The FOXA and HNF4 binding sites were modified to match the consensus.

Example 3: *In vitro* activity of multimerized human SERPINA1 enhancer variants

[00487] Selected multimerized enhancer variants from **Table 4** were screened using the luciferase reporter assay (described in Example 2) to identify variants with enhanced performance compared to the human SERPINA1 enhancer.

[00488] In a screen of 10 variants, the 1x version of hSerpEnh_FOXA_HNF4_consensus_v1 performed similarly to 3x hSerpEnh. The 3x version of hSerpEnh_FOXA_HNF4_consensus_v1 performed 1.6 times better than the 3x version of hSerpEnh and 3 times better than a single hSerpEnh (see **FIG. 6**). **FIG. 5** depicts results of the screen of 10 multimerized variants using a luciferase reporter assay (n=3). Results are grouped by 3x repeats of rationally designed enhancer variants (3x TFBS Variants), 3x repeats of conserved SERPINA1 enhancer regions identified in other species (3x Conserved Variant), 3x repeats of the human SERPINA1 enhancer separated by spacers of varying lengths and sequences (3x hSerpEnh Spacer Variants), and enhancers with varying numbers of repeats (# Repeat Variants). Results for the wild-type human SERPINA1 enhancer are shown in red. The comparison between the 3x human SERPINA1 enhancer variant and the 3x top performing variant is boxed. Two sets of technical triplicates were performed for the 1x and 3x human enhancers and the top performing 3x variant (r1, r2). Error bars represent standard deviation.

Number of Enhancer Repeats and Length of Spacers

[00489] Like above, enhancers are often combined in series (multimerized / repeated enhancer sequences) to drive higher levels of transcription initiation. However, the principals underlying optimal number of repeats and orientation of enhancer regions are not well understood. Spacing between each iteration of repeated enhancers was hypothesized to be an attribute that impacts function, especially considering that DNA is a helix such that number of nucleotides between binding sites also changes their rotational spatial orientation. Spacers of different length between enhancers were tested. The length and sequence of spacers between SERPINA1 enhancer variant repeats were modified to screen for sequences that improved enhancer function. Spacers of length 2, 3, 5, 11, and 30 were designed to prevent introduction of CpGs or ATGs that may create cryptic translation start sites. 11 nt and 30 nt

spacers that contain consensus FOXA and HNF4 binding sites were also designed and tested. (see, *e.g.*, **FIG. 6A**).

[00490] A range of enhancer combinations for improved function, including various multimer enhancers and nucleotide spacer content, were tested in a dual luciferase transient transfection assay. Three main configurations as shown in **FIG. 6B** were tested: a single human Serpin enhancer (1x hSerpEnh), a 3x human Serpin enhancer (3x hSerpEnh) with spacers between the enhancer repeats and the transthyretin gene enhancer (TTRe) (3x hSerpEnh-TTRe), and multiple enhancers with spacers between enhancer repeats (*e.g.*, 3x, 5x or 10x hSerpEnh). Variants with multiple enhancers for screening are shown in **Table 4**, above.

To determine the optimal number of repeats, HNF4_FOXA_v1, the top variant from the enhancer screen (FOXA_HNF4_consensus_v1), was placed in an array of 3, 5 or 10 repeats (*e.g.*, 3x HNF4 FOXA v1; 5x HNF4 FOXA v1; and 10x HNF4 FOXA v1) to drive expression of FVIII from a plasmid. One dose of 50ng plasmid containing FVIII cDNA sequence was transfected into HepG2 cells. As shown in **FIG. 7A, 7B**, 3x and 5x variants performed better than HNF4 FOXA v1 variants repeated 10 times (10x) which did not exhibit a meaningful level of FVIII (see, *e.g.*, **FIG. 7C**), suggesting that the Serpin enhancer exhibits superior performances when it is repeated in certain number, *e.g.*, 3x to 5x, but not when it is repeated in an excessive number (*e.g.*, 10x). A consistent observation was made with other Serpin Enhancer elements including, for example, that of bushbaby Serpin enhancer, Chinese tree shrew Serpin enhancer, and human Serpin enhancer (hSerpEnh)(**FIG. 7D**). The FVIII open reading frame (ORF) sequence used here was b-domain deleted codon optimized sequence as set forth in **SEQ ID NO: 143** (hFVIII-F309S-BD226seq124-BDD-F309). Exemplary DNA constructs containing SEQ ID NO: 143 and enhancers / promoters of the present disclosure are shown in **FIGs. 11 and 12**.

[00491] Further, spacers having difference number of nucleotides and sequence were tested to determine whether these repeated enhancer elements are sensitive to their spatial orientation created by a spacer as well as characteristics of the spacer created by spacer sequences. The length and sequence of spacers between hSERPINA1 enhancer repeats were modified to screen for sequences that improved enhancer function. As shown in **FIGs. 8A-E**, the hSerpEnh elements were sensitive to length as well as sequences of the spacers placed. In plasmid-mediated FVIII expression in HepG2 cells,) two nucleotide spacers generally exhibited improved activity as compared to 3x human Serpin enhancer with a single nucleotide spacer (3x hSerpEnh-TTRe with "C" spacer). A certain level of spacer-sequence driven dependency was also observed as the activity seen with constructs having the similar length of spacer exhibited widely expression different profile depending on the DNA sequence (see 11-mer in **FIG. 8D**). Surprisingly, one spacer having 11 nucleotides performed exceedingly better than other 11-mers or other length spacers like three nucleotide spacers (see **FIG. 8B**) or 5 nucleotide spacers (see **FIG. 8C**). This spacer (version 3) was one of variants of 11-mer spacers having the sequence of

“TGCAAAGTCCT” (SEQ ID NO: 144) and/or “AGTGTTTACAA” (SEQ ID NO: 145) as shown in SEQ ID NO:71.

[00492] To determine whether Serpin enhancer variants (SerpEnh derived from Bushbaby or Chinese Tree Shrew), plasmid FVIII constructs containing 3xBushbaby SerpEnh having adenine (A) spacers or 3x Chinese Tree Shrew were injected hydrodynamically into Rag 2 mice to drive expression of FVIII (HDI tail vein injection of 50ng plasmid containing FVIII cdDNA sequence on day 0 with a single blood collection at day 7 for the measurement of FVIII activity). Surprisingly, the 3x Serpin A enhancer sequence derived from Bushbaby having a single nucleotide (adenine) as spacers exhibited increased FVIII expression (**FIG. 9**). Consistent with the observations above, 3x human Serpin enhancer sequence with 11mer spacers also exhibited increased expression of FVIII as compared to 3x hSerpEhn (**FIG. 9**). It was also noted that 3x Chinese Tree Shrew enhancers that has 3 missing nucleotides in its 5' end as compared to the human SerpEnh sequence (see **FIG. 1**) also exhibited an equivalent level of expression as compared to those of various 3x human serpin enhancer constructs (**FIG. 9**).

[00493] To determine whether the capacity of these enhancers could vary in a platform-dependent manner (*e.g.*, plasmid *v.* closed-ended DNA), corresponding cdDNA vectors representative of the experimental results shown in **FIGs. 7-9** were prepared and hydrodynamically injected into Rag2 mice as described above. **FIG. 10** shows the result of FVIII expression from spacer variants of hSerpEnh (2mers and 11mers) and Serpin enhancer variants (3x bushbaby Serpin enhancer and 3x Chinese tree shrew Serpin enhancer). Surprisingly, in the cdDNA platform, all of the SerpEnh variants tested (*i.e.*, 3x Bushbaby SerpEnh variant, 3x Chinese Tree shrew SerpEnh variant, 11mer spacer variants and 2mer spacer variants) exhibited equivalent or superior FVIII expression profiles as compared to that of 3x hSerpEnh, suggesting that these enhancers can be successfully implemented to drive expression of a therapeutic protein like FVIII *in vivo*.

[00494] The following nucleotide sequence is a cdDNA-plasmid sequence comprising a left ITR : spacer : bushbaby serpin enhancer (3xBushbaby_Aspacers) : TTRe (TTR enhancer) : TTR liver-specific promoter : MVM intron : B-domain deleted FVIII : WPRE 3'UTR : bGH : spacer right ITR: right ITR. Detailed annotations for this construct are shown in **FIG. 11**.

```
AAAGTAGCCGAAGATGACGGTTTGTACATGGAGTTGGCAGGATGTTTGATTAAAAACATAACAGGAA
GAAAAATGCCCCGCTGTGGGCGGACAAAATAGTTGGGAAGTGGGAGGGGTGGAAATGGAGTTTTTAAG
GATTATTTAGGGAAGAGTGACAAAATAGATGGGAAGTGGGTGTAGCGTCGTAAGCTAATACGAAAATT
AAAAATGACAAAATAGTTTGGAACTAGATTTCACTTATCTGGTTCGGATCTCCTAGGCCTGCAGGCAG
CTGCGCGCTCGCTCGCTCACTGAGGCCGCGCGGGCAAAGCCCGGGCGTGGGCGACCTTTGGTTCGCCC
GGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCCTTGTAGT
TAATGATTAAACCCGCCATGCTACTTATCGCGGCCGCAGGGGAAGCTACTGGTGAATATTAACCAAGGT
CACCCAGTTATCAGGGAGCAAACAGGAGCTAAGTCCATAGGGGAAGCTACTGGTGAATATTAACCAA
```

GGTCACCCAGTTATCAGGGAGCAAACAGGAGCTAAGTCCATAGGGGGAAGCTACTGGTGAATATTAAC
CAAGGTCACCCAGTTATCAGGGAGCAAACAGGAGCTAAGTCCATGGTACCCACTGGGAGGATGTTGAG
TAAGATGGAAAATACTGATGACCCTTGAGAGACAGAGTATTAGGACATGTTGAACAGGGGCCGGG
CGATCAGCAGGTAGCTCTAGAGGATCCCCGTCTGTCTGCACATTTCTAGAGCGAGTGTTCCGATACT
CTAATCTCCCTAGGCAAGGTTTCATATTTGTGTAGGTTACTTATTCTCCTTTTGTTGACTAAGTCAATA
ATCAGAATCAGCAGGTTTGGAGTCAGCTTGGCAGGGATCAGCAGCCTGGGTTGGAAGGAGGGGGTATA
AAAGCCCCCTTACCAGGAGAAGCCGTCACACAGATCCACAAGCTCCTGAAGAGGTAAGGGTTTAAGGG
ATGGTTGGTTGGTGGGGTATTAATGTTTAATTACCTGGAGCACCTGCCTGAAATCACTTTTTTTTTCAGG
TTGGTTTAAACGCCGCCACCATGCAGATTGAGCTGAGCACCTGCTTCTTCTGTGCCTGCTGAGGTTT
TGCTTCTCTGCCACCAGGAGATACTACCTGGGGGCTGTGGAGCTGAGCTGGGACTACATGCAGTCTGA
CCTGGGGGAGCTGCCTGTGGATGCCAGGTTCCCCCAGAGTGCCCAAGAGCTTCCCCTTCAACACCT
CTGTGGTGTACAAGAAGACCCTGTTTGTGGAGTTCACTGACCACCTGTTCAACATTGCCAAGCCCAGG
CCCCCTGGATGGGCCTGCTGGGCCCCACCATCCAGGCTGAGGTGTATGACACTGTGGTGATCACCTT
GAAGAATATGGCCAGCCACCCTGTGAGCCTGCATGCTGTGGGGGTGAGCTACTGGAAGGCCTCTGAGG
GGGCTGAGTATGATGACCAGACCAGCCAGAGGGAGAAGGAGGATGACAAGGTGTTCCCTGGGGGCAGC
CACACCTATGTGTGGCAGGTGCTGAAGGAGAATGGCCCCATGGCCTCTGACCCCTGTGCCTGACCTA
CAGCTACCTGAGCCATGTGGACCTGGTGAAGGACCTGAACTCTGGCCTGATTGGGGCCCTGCTGGTGT
GCAGGGAGGGCAGCCTGGCCAAGGAGAAGACCCAGACCCTGCACAAGTTCATCCTGCTGTTTGCTGTG
TTTGATGAGGGCAAGAGCTGGCACTCTGAAACCAAGAACAGCCTGATGCAGGACAGGGATGCTGCCTC
TGCCAGGGCCTGGCCCCAAGATGCACACTGTGAATGGCTATGTGAACAGGAGCCTGCCTGGCCTGATTG
GCTGCCACAGGAAGTCTGTGTACTGGCATGTGATTGGCATGGGCACCACCCCTGAGGTGCACAGCATC
TTCCTGGAGGGCCACACCTTCTGGTCAGGAACCACAGGCAGGCCAGCCTGGAGATCAGCCCCATCAC
CTTCTGACTGCCCAGACCCTGCTGATGGACCTGGGCCAGTTCTGCTGTTCTGCCACATCAGCAGCC
ACCAGCATGATGGCATGGAGGCCTATGTGAAGGTGGACAGCTGCCCTGAGGAGCCCCAGCTGAGGATG
AAGAACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATGGATGTGGTGAGGTT
TGATGATGACAACAGCCCCAGCTTCATCCAGATCAGGTCTGTGGCCAAGAAGCACCCCAAGACCTGGG
TGCACTACATTGCTGCTGAGGAGGAGGACTGGGACTATGCCCCCTGGTGTGGCCCCCTGATGACAGG
AGCTACAAGAGCCAGTACCTGAACAATGGCCCCCAGAGGATTGGCAGGAAGTACAAGAAGGTGAGGTT
CATGGCCTACACTGATGAAACCTTCAAGACCAGGGAGGCCATCCAGCATGAGTCTGGCATCCTGGGCC
CCCTGCTGTATGGGGAGGTGGGGGACACCCTGCTGATCATCTTCAAGAACCAGGCCAGGCCCCCTAC
AACATCTACCCCCATGGCATCACTGATGTGAGGCCCTGTACAGCAGGAGGCTGCCCCAAGGGGGTGAA
GCACCTGAAGGACTTCCCCATCCTGCCTGGGGAGATCTTCAAGTACAAGTGGACTGTGACTGTGGAGG
ATGGCCCCACCAAGTCTGACCCAGGTGCCTGACCAGATACTACAGCAGCTTGTGAACATGGAGAGG
GACCTGGCCTCTGGCCTGATTGGCCCCCTGCTGATCTGCTACAAGGAGTCTGTGGACCAGAGGGGCAA
CCAGATCATGTCTGACAAGAGGAATGTGATCCTGTTCTCTGTGTTTGATGAGAACAGGAGCTGGTACC
TGACTGAGAACATCCAGAGGTTCTGCCAACCTGCTGGGGTGCAGCTGGAGGACCCTGAGTTCCAG
GCCAGCAACATCATGCACAGCATCAATGGCTATGTGTTTGACAGCCTGCAGCTGTCTGTGTGCCTGCA

TGAGGTGGCCTACTGGTACATCCTGAGCATTGGGGCCCAGACTGACTTCCTGTCTGTGTTCTTCTCTG
GCTACACCTTCAAGCACAAAGATGGTGTATGAGGACACCCTGACCCTGTTCCCCCTTCTCTGGGGAGACT
GTGTTTCATGAGCATGGAGAACCCTGGCCTGTGGATTCTGGGCTGCCACAACCTCTGACTTCAGGAACAG
GGGCATGACTGCCCTGCTGAAAGTCTCCAGCTGTGACAAGAACAACCTGGGGACTACTATGAGGACAGCT
ATGAGGACATCTCTGCCTACCTGCTGAGCAAGAACAATGCCATTGAGCCCAGGAGCTTCAGCCAGAAT
AGCAGGCACCCCAGCACCCAGGCAGAAGCAGTTCAATGCCACCACCATCCCAGAGAATACCACCCTGCA
GTCTGACCAGGAGGAGATTGACTATGATGACACCATCTCTGTGGAGATGAAGAAGGAGGACTTTGACA
TCTACGACGAGGACGAGAACCAGAGCCCCAGGAGCTTCCAGAAGAAGACCAGGCACTACTTCATTGCT
GCTGTGGAGAGGCTGTGGGACTATGGCATGAGCAGCAGCCCCATGTGCTGAGGAACAGGGCCCAGTC
TGGCTCTGTGCCCCAGTTCAAGAAGGTGGTGTTCAGGAGTTCACTGATGGCAGCTTCACCCAGCCCC
TGTACAGAGGGGAGCTGAATGAGCACCTGGGCCTGCTGGGCCCCCTACATCAGGGCTGAGGTGGAGGAC
AACATCATGGTGACCTTCAGGAACCAGGCCAGCAGGCCCTACAGCTTCTACAGCAGCCTGATCAGCTA
TGAGGAGGACCAGAGGCAGGGGGCTGAGCCCAGGAAGAACTTTGTGAAGCCCAATGAAACCAAGACCT
ACTTCTGGAAGGTGCAGCACCATGGCCCCCACCAAGGATGAGTTTGACTGCAAGGCCTGGGCCTAC
TTCTCTGATGTGGACCTGGAGAAGGATGTGCACTCTGGCCTGATTGGCCCCCTGCTGGTGTGCCACAC
CAACACCTGAAACCTGCCCCATGGCAGGCAGGTGACTGTGCAGGAGTTTGCCCTGTTCTTCACCATCT
TTGATGAAACCAAGAGCTGGTACTTCACTGAGAACATGGAGAGGAACTGCAGGGCCCCCTGCAACATC
CAGATGGAGGACCCACCTTCAAGGAGAACTACAGGTTCCATGCCATCAATGGCTACATCATGGACAC
CCTGCCTGGCCTGGTGATGGCCCAGGACCAGAGGATCAGGTGGTACCTGCTGAGCATGGGCAGCAATG
AGAACATCCACAGCATCCACTTCTCTGGCCATGTGTTCACTGTGAGGAAGAAGGAGGAGTACAAGATG
GCCCTGTACAACCTGTACCCTGGGGTGTGTTGAGACTGTGGAGATGCTGCCCAGCAAGGCTGGCATCTG
GAGGGTGGAGTGCTGATTGGGGAGCACCTGCATGCTGGCATGAGCACCTGTTCTTGGTGTACAGCA
ACAAGTGCCAGACCCCCCTGGGCATGGCCTCTGGCCACATCAGGGACTTCCAGATCACTGCCTCTGGC
CAGTATGGCCAGTGGGGCCCCAAGCTGGCCAGGCTGCACTACTCTGGCAGCATCAATGCCTGGAGCAC
CAAGGAGCCCTTCAGCTGGATCAAGGTGGACCTGCTGGCCCCCATGATCATCCATGGCATCAAGACCC
AGGGGGCCAGGCAGAAGTTTCAGCAGCCTGTACATCAGCCAGTTTCATCATCATGTACAGCCTGGATGGC
AAGAAGTGGCAGACCTACAGGGGCAACAGCACTGGCACCCCTGATGGTGTCTTTGGCAATGTGGACAG
CTCTGGCATCAAGCACAAACATCTTCAACCCCCCATCATTGCCAGATACATCAGGCTGCACCCACCC
ACTACAGCATCAGGAGCACCTGAGGATGGAGCTGATGGGCTGTGACCTGAACAGCTGCAGCATGCCC
CTGGGCATGGAGAGCAAGGCCATCTCTGATGCCAGATCACTGCCAGCAGCTACTTCACCAACATGTT
TGCCACCTGGAGCCCCAGCAAGGCCAGGCTGCACCTGCAGGGCAGGAGCAATGCCTGGAGGCCCCAGG
TCAACAACCCCAAGGAGTGGCTGCAGGTGGACTTCCAGAAGACCATGAAGGTGACTGGGGTGACCACC
CAGGGGGTGAAGAGCCTGCTGACCAGCATGTATGTGAAGGAGTTCTGATCAGCAGCAGCCAGGATGG
CCACCAGTGGACCTGTTCTTCCAGAATGGCAAGGTGAAGGTGTTCCAGGGCAACCAGGACAGCTTCA
CCCCGTGTTGTGAACAGCCTGGACCCCCCTGCTGACCAGATACCTGAGGATTACCCCCAGAGCTGG
GTGCACCAGATTGCCCTGAGGATGGAGGTGCTGGGCTGTGAGGCCAGGACCTGTACTGATTAATTA
GAGCATCTTACCGCCATTTATTCCCATATTTGTTCTGTTTTCTTGATTGGGTATACATTTAAATGT

TAATAAAACAAAATGGTGGGGCAATCATTTACATTTTTAGGGATATGTAATTACTAGTTCAGGTGTAT
TGCCACAAGACAAACATGTTAAGAACTTTCCCGTTATTTACGCTCTGTTCCCTGTTAATCAACCTCTG
GATTACAAAATTTGTGAAAGATTGACTGATATTCTTAACTATGTTGCTCCTTTTACGCTGTGTGGATA
TGCTGCTTTATAGCCTCTGTATCTAGCTATTGCTTCCCGTACGGCTTTCGTTTTCTCCTCCTTGTATA
AATCCTGGTTGCTGTCTCTTTTAGAGGAGTTGTGGCCCGTTGTCCGTCAACGTGGCGTGGTGTGCTCT
GTGTTTGTGACGCAACCCCCACTGGCTGGGGCATTGCCACCACCTGTCAACTCCTTTCTGGGACTTT
CGCTTTCCCCCTCCCGATCGCCACGGCAGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGG
CTAGGTTGCTGGGCACTGATAAATCCGTGGTGTGTCTGTGCCTTCTAGTTGCCAGCCATCTGTTGTT
TGCCCCCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCTTTCTTAATAAAATGA
GGAAATTGCATCGCATTGTCTGAGTAGGTGTCTATTCTATTCTGGGGGTGGGGTGGGGCAGGACAGCA
AGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTCTAGAGCAT
GGCTACGTAGATAAGTAGCATGGCGGGTTAATCATTAACTACACCTGCAGGAGGAACCCCTAGTGATG
GAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTGCGCCGACG
CCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGGGGCGCGCCTCGAGCCATGGTGTCT
AGCAGCTGATGCATAGCATGCGGTACCGGGAGATGGGGGAGGCTAACTGAAACACGGAAGGAGACAAT
ACCGGAAGGAACCCGCGCTATGACGGCAATAAAAAGACAGAATAAAACGCACGGGTGTTGGGTGCTTT
GTTCATAAACGCGGGGTTGCGTCCCAGGGCTGGCACTCTGTGATACCCACCGAGACCCCAATTGGGA
CCAATACGCCCAGTTCCTTTTCCCCACCCCAACCCCAAGTTCGGGTGAAGGCCAGGGCTCG
CAGCCAACGTGCGGGCGGCAAGCCCTGCCATAGCCACTACGGGTACGTAGGCCAACCACTAGAACTAT
AGCTAGAGTCCTGGGCGAACAACAGATGCTCGCCTTCCAGAAAACCGAGGATGCGAACCACTTCATCC
GGGGTCAGCACCAACCGGCAAGCGCCGCGACGGCCGAGGTCTACCGATCTCCTGAAGCCAGGGCAGATC
CGTGACAGCACCTTGCCGTAGAAGAACAGCAAGGCCGCAATGCCTGACGATGCGTGAGACCGAAA
CCTTGCGCTCGTTGCCAGCCAGGACAGAAATGCCTCGACTTCGCTGCTGCCAAGGTTGCCGGGTGA
CGCACACCGTGGAACCGGATGAAGGCACGAACCCAGTTGACATAAGCCTGTTGCGTTGCTAAACTGTA
ATGCAAGTAGCGTATGCGCTCACGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCG
CAGTGGCGGTTTTTCATGGCTTGTTATGACTGTTTTTTGTACAGTCTATGCCTCGGGCATCCAAGCAG
CAAGCGGTTACGCCGTGGGTGCGATGTTTGATGTTATGGAGCAGCAACGATGTTACGCAGCAGCAACG
ATGTTACGCAGCAGGGCAGTCGCCCTAAACAAAGTTAGGTGGCTCAAGTATGGGCATCATTGACACA
TGTAGGCTCGGCCCTGACCAAGTCAAATCCATGCGGGCTGCTCTTGATCTTTTCGGTTCGTGAGTTCGG
AGACGTAGCCACCTACTCCCAACATCAGCCGGAAGTCCGATTACCTCGGGAAGTTCCTCGTAGTAAGA
CATTTCATCGCGCTTGCTGCCTTCGACCAAGAAGCGGTTGTTGGCGCTCTCGCGGCTTACGTTCTGCC
AGGTTTGAGCAGCCGCGTAGTGAGATCTATATCTATGATCTCGCAGTCTCCGGCGAGCACCGGAGGCA
GGGCATTGCCACCGCGCTCATCAATCTCCTCAAGCATGAGGCCAACGCGCTTGGTGCTTATGTGATCT
ACGTGCAAGCAGATTACGGTGACGATCCCGCAGTGGCTCTCTATACAAAGTTGGGCATACGGGAAGAA
GTGATGCACTTTGATATCGACCCAAGTACCGCCACCTAACAATTCGTTCAAGCCGAGATCGGCTTCCC
GGCCGCGGAGTTGTTGCGTAAATTGTCAACGCCGCGAATATAGTCTTTACCATGCCCTTGCCACG
CCCTCTTTAATACGACGGGCAATTTGCACTTCAGAAAATGAAGAGTTTGCTTTAGCCATAACAAAAG

TCCAGTATGCTTTTTTACAGCATAACTGGACTGATTTTCAGTTTACAACCTATTCTGTCTAGTTTAAAGAC
TTTATTGTCATAGTTTAGATCTATTTTGTTCAGTTTAAAGACTTTATTGTCCGCCACACCCGCTTACG
CAGGGCATCCATTTATTACTCAACCGTAACCGATTTTGCCAGGTTACGCGGCTGGTCTGCGGTGTGAA
ATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCTCTTCCGCTTCCTCGCTCACTGACTC
GCTGCGCTCGGTCGTTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAAATACGGTTATCCA
CAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAA
AAGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTC
AAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATAACCAGGCGTTTCCCCCTGGAAGCTCCCTCG
TGCGCTCTCCTGTTCCGACCCTGCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTG
GCGCTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTTCGCTCCAAGCTGGGCTG
TGTGCACGAACCCCCCGTTTCAGCCCGACCGCTGCGCCTTATCCGTAACCTATCGTCTTGAGTCCAACC
CGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTA
GGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTAT
CTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAGAGAGTTGGTAGCTCTTGATCCGGCAACAAACCA
CCGCTGGTAGCGGTGGTTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAA
GATCCTTTGATCTTTTTCTACGGGGTCTGACGCTCAGTGGAAACGAAAACCTCACGTTAAGGGATTTTGGT
CATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCT
AAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCG
ATCTGTCTATTTTCGTTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGG
CTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAG
CAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAG
TCTATTAATTGTTGCCGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGC
CATTGCTACAGGCATCGTGGTGTACGCTCGTCGTTTGGTATGGCTTCATTACGCTCCGGTTCCCAAC
GATCAAGGCGAGTTACATGATCCCCATGTTGTGCAAAAAGCGGTTAGCTCCTTCGGTCTCCGATC
GTTGTGCAAGTAAGTTGGCCGAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTAC
TGTATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGT
GTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACT
TTAAAAGTGCTCATCATTGGAACGTTCTTCGGGGCGAAAACCTCTCAAGGATCTTACCGCTGTTGAG
ATCCAGTTCGATGTAACCCACTCGTGACCCAACTGATCTTCAGCATCTTTTACTTTTACCAGCGTTT
CTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAGGGAATAAGGGCGACACGGAAATGTTGA
ATACTCATACTCTTCTTTTTCAATATTATTGAAGCATTATCAGGGTTATTGTCTCATGAGCGGATA
CATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCAC
CTGAAATTGTAAACGTTAATATTTTGTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTT
AACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAGAATAGACCGAGATAGGGTTGAGTGT
TGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCG
TCTATCAGGGCGATGGCCCACTACGTGAACCATACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGT
AAAGCACTAAATCGGAACCTAAAGGGAGCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGT

GGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAAGTGTAGCGGTACAGC
TGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCCCATTCGCCATTCAGG
CTGCAAATAAGCGTTGATATTCAGTCAATTACAAACATTAATAACGAAGAGATGACAGAAAAATTTTC
ATTCTGTGACAGAGAA (SEQ ID NO: 146)

[00495] The following nucleotide sequence is a ceDNA-plasmid sequence comprising a left ITR :
spacer : 3x hSerpEnh : TTRe (TTR enhancer) : TTR liver-specific promoter : MVM intron : B-domain
deleted FVIII : WPRE 3'UTR : bGH : spacer right ITR: right ITR. Detailed annotations for this ceDNA
vector are shown in **FIG. 12**.

CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGACC
TTTGGTCGCCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGG
TTCCTTGTAGTTAATGATTAACCCGCCATGCTACTTATCGCGGCCGCGGGGAGGCTGCTGGTGAATA
TTAACCAAGGTCACCCAGTTATCGGAGGAGCAAACAGGGGCTAAGTCCACCGGGGGAGGCTGCTGGT
GAATATTAACCAAGGTCACCCAGTTATCGGAGGAGCAAACAGGGGCTAAGTCCACCGGGGGAGGCTG
CTGGTGAATATTAACCAAGGTCACCCAGTTATCGGAGGAGCAAACAGGGGCTAAGTCCACGGTACCC
ACTGGGAGGATGTTGAGTAAGATGGAAACTACTGATGACCCTTGCGAGAGACAGAGTATTAGGACATG
TTTGAACAGGGGCCGGGCGATCAGCAGGTAGCTCTAGAGGATCCCCGTCTGTCTGCACATTTTCGTAGA
GCGAGTGTTCCGATACTCTAATCTCCCTAGGCAAGGTTTATATTTGTGTAGGTTACTTATTCTCCTTT
TGTTGACTAAGTCAATAATCAGAATCAGCAGGTTTGGAGTCAGCTTGGCAGGGATCAGCAGCCTGGGT
TGGAAGGAGGGGGTATAAAAGCCCCCTTACCAGGAGAAGCCGTCACACAGATCCACAAGCTCCTGAAG
AGGTAAGGGTTTAAGGGATGTTTGGTTGGTGGGGTATTAATGTTTAATTACCTGGAGCACCTGCCTGA
AATCACTTTTTTTTTCAGGTTGGTTTTAAACGCCGCCACCATGCAGATTGAGCTGAGCACCTGCTTCTTCC
TGTGCTGCTGAGGTTCTGCTTCTCTGCCACCAGGAGATACTACCTGGGGGCTGTGGAGCTGAGCTGG
GACTACATGCAGTCTGACCTGGGGGAGCTGCCTGTGGATGCCAGGTTCCCCCAGAGTGCCCAAGAG
CTTCCCCTTCAACACCTCTGTGGTGTACAAGAAGACCCTGTTTGTGGAGTTCACTGACCACCTGTTCA
ACATTGCCAAGCCCAGGCCCCCTGGATGGGCTGCTGGGCCCCACCATCCAGGCTGAGGTGTATGAC
ACTGTGGTGATCACCTGAAGAACATGGCCAGCCACCCTGTGAGCCTGCATGCTGTGGGGGTGAGCTA
CTGGAAGGCTCTGAGGGGGCTGAGTATGATGACCAGACCAGCCAGAGGAGAAGGAGGATGACAAGG
TGTTCCCTGGGGGCAGCCACACCTATGTGTGGCAGGTGCTGAAGGAGAATGGCCCCATGGCCTCTGAC
CCCCGTGTCCTGACCTACAGCTACCTGAGCCATGTGGACCTGGTGAAGGACCTGAACTCTGGCCTGAT
TGGGGCCCTGCTGGTGTGCAGGGAGGGCAGCCTGGCCAAGGAGAAGACCCAGACCCTGCACAAGTTCA
TCCTGCTGTTTGCTGTGTTTGATGAGGGCAAGAGCTGGCACTCTGAAACCAAGAACAGCCTGATGCAG
GACAGGGATGCTGCCTCTGCCAGGGCCTGGCCCAAGATGCACACTGTGAATGGCTATGTGAACAGGAG
CCTGCCTGGCCTGATTGGCTGCCACAGGAAGTCTGTGTACTGGCATGTGATTGGCATGGGCACCAACC
CTGAGGTGCACAGCATCTTCTGGAGGGCCACACCTTCTGGTTCAGGAACCACAGGCAGGCCAGCCTG
GAGATCAGCCCCATCACCTTCTGACTGCCAGACCCTGCTGATGGACCTGGGCCAGTTCCTGCTGTT
CTGCCACATCAGCAGCCACCAGCATGATGGCATGGAGGCCTATGTGAAGGTGGACAGCTGCCCTGAGG
AGCCCCAGCTGAGGATGAAGAAACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAG

ATGGATGTGGTGAGGTTTGTATGATGACAACAGCCCCAGCTTCATCCAGATCAGGTCTGTGGCCAAGAA
GCACCCCAAGACCTGGGTGCACTACATTGCTGCTGAGGAGGAGGACTGGGACTATGCCCCCTGGTGC
TGGCCCCTGATGACAGGAGCTACAAGAGCCAGTACCTGAACAATGGCCCCCAGAGGATTGGCAGGAAG
TACAAGAAGGTCAGGTTTCATGGCCTACACTGATGAAACCTTCAAGACCAGGGAGGCCATCCAGCATGA
GTCTGGCATCCTGGGCCCCCTGCTGTATGGGGAGGTGGGGGACACCCCTGCTGATCATCTTCAAGAACC
AGGCCAGCAGGCCCTACAACATCTACCCCCATGGCATCACTGATGTGAGGCCCTGTACAGCAGGAGG
CTGCCCAAGGGGGTGAAGCACCTGAAGGACTTCCCCATCCTGCCTGGGGAGATCTTCAAGTACAAGTG
GACTGTGACTGTGGAGGATGGCCCCACCAAGTCTGACCCCAGGTGCCTGACCAGATACTACAGCAGCT
TTGTGAACATGGAGAGGGACCTGGCCTCTGGCCTGATTGGCCCCCTGCTGATCTGCTACAAGGAGTCT
GTGGACCAGAGGGGCAACCAGATCATGTCTGACAAGAGGAATGTGATCCTGTTCTCTGTGTTTGTATGA
GAACAGGAGCTGGTACCTGACTGAGAACATCCAGAGGTTCTGCCCAACCCCTGCTGGGGTGCAGCTGG
AGGACCCCTGAGTTCCAGGCCAGCAACATCATGCACAGCATCAATGGCTATGTGTTTGACAGCCTGCAG
CTGTCTGTGTGCCTGCATGAGGTGGCCTACTGGTACATCCTGAGCATTGGGGCCCAGACTGACTTCCT
GTCTGTGTTCTTCTCTGGCTACACCTTCAAGCACAAAGATGGTGTATGAGGACACCCTGACCCTGTTCC
CCTTCTCTGGGGAGACTGTGTTTCATGAGCATGGAGAACCCTGGCCTGTGGATTCTGGGCTGCCACAAC
TCTGACTTCAGGAACAGGGGCATGACTGCCCTGCTGAAAGTCTCCAGCTGTGACAAGAACAAGTGGGA
CTACTATGAGGACAGCTATGAGGACATCTCTGCCTACCTGCTGAGCAAGAACAATGCCATTGAGCCCA
GGAGCTTCAGCCAGAATAGCAGGCACCCAGCACCAGGCAGAAGCAGTTCAATGCCACCACCATCCCA
GAGAATACCACCCCTGCAGTCTGACCAGGAGGAGATTGACTATGATGACACCATCTCTGTGGAGATGAA
GAAGGAGGACTTTGACATCTACGACGAGGACGAGAACCAGAGCCCCAGGAGCTTCCAGAAGAAGACCA
GGCACTACTTCATTGCTGCTGTGGAGAGGCTGTGGGACTATGGCATGAGCAGCAGCCCCCATGTGCTG
AGGAACAGGGCCAGTCTGGCTCTGTGCCCCAGTTCAAGAAGGTGGTGTTCAGGAGTTCACTGATGG
CAGCTTCACCCAGCCCCCTGTACAGAGGGGAGCTGAATGAGCACCTGGGCCTGCTGGGCCCCCTACATCA
GGGCTGAGGTGGAGGACAACATCATGGTGACCTTCAGGAACCAGGCCAGCAGGCCCTACAGCTTCTAC
AGCAGCCTGATCAGCTATGAGGAGGACCAGAGGCAGGGGGCTGAGCCCAGGAAGAACTTTGTGAAGCC
CAATGAAACCAAGACCTACTTCTGGAAGGTGCAGCACCACATGGCCCCCACCAGGATGAGTTTGACT
GCAAGGCCTGGGCCTACTTCTCTGATGTGGACCTGGAGAAGGATGTGCACTCTGGCCTGATTGGCCCC
CTGCTGGTGTGCCACACCAACACCCTGAACCCTGCCATGGCAGGCAGGTGACTGTGCAGGAGTTTGC
CCTGTTCTTACCATCTTTGATGAAACCAAGAGCTGGTACTTCACTGAGAACATGGAGAGGAACTGCA
GGGCCCCCTGCAACATCCAGATGGAGGACCCACCTTCAAGGAGAACTACAGGTTCCATGCCATCAAT
GGCTACATCATGGACACCCTGCCTGGCCTGGTGTATGGCCAGGACCAGAGGATCAGGTGGTACCTGCT
GAGCATGGGCAGCAATGAGAACATCCACAGCATCCACTTCTCTGGCCATGTGTTCACTGTGAGGAAGA
AGGAGGAGTACAAGATGGCCCTGTACAACCTGTACCCTGGGGTGTGTTGAGACTGTGGAGATGCTGCCC
AGCAAGGCTGGCATCTGGAGGGTGGAGTGCCTGATTGGGGAGCACCTGCATGCTGGCATGAGCACCCCT
GTTCTGTGTACAGCAACAAGTGCCAGACCCCCCTGGGCATGGCCTCTGGCCACATCAGGGACTTCC
AGATCACTGCCTCTGGCCAGTATGGCCAGTGGGCCCCCAAGCTGGCCAGGCTGCACTACTCTGGCAGC
ATCAATGCCTGGAGCACCAAGGAGCCCTTCACTGGATCAAGGTGGACCTGCTGGCCCCCATGATCAT

CCATGGCATCAAGACCCAGGGGGCCAGGCAGAAGTTCAGCAGCCTGTACATCAGCCAGTTCATCATCA
 TGTACAGCCTGGATGGCAAGAAGTGGCAGACCTACAGGGGCAACAGCACTGGCACCCCTGATGGTGTTC
 TTTGGCAATGTGGACAGCTCTGGCATCAAGCACAACATCTTCAACCCCCCATCATTGCCAGATACAT
 CAGGCTGCACCCACCCACTACAGCATCAGGAGCACCTGAGGATGGAGCTGATGGGCTGTGACCTGA
 ACAGCTGCAGCATGCCCCTGGGCATGGAGAGCAAGGCCATCTCTGATGCCCAGATCACTGCCAGCAGC
 TACTTCACCAACATGTTTGCCACCTGGAGCCCCAGCAAGGCCAGGCTGCACCTGCAGGGCAGGAGCAA
 TGCCTGGAGGCCCCAGGTCAACAACCCCAAGGAGTGGCTGCAGGTGGACTTCCAGAAGACCATGAAGG
 TGACTGGGGTGACCACCCAGGGGGTGAAGAGCCTGCTGACCAGCATGTATGTGAAGGAGTTCCTGATC
 AGCAGCAGCCAGGATGGCCACCAGTGGACCCTGTTCTTCCAGAATGGCAAGGTGAAGGTGTTCCAGGG
 CAACCAGGACAGCTTCACCCCTGTGGTGAACAGCCTGGACCCCCCCTGCTGACCAGATACCTGAGGA
 TTCACCCCAAGAGCTGGGTGCACCAGATTGCCCTGAGGATGGAGGTGCTGGGCTGTGAGGCCCAGGAC
 CTGTACTGATTAATTAAGAGCATCTTACCGCCATTTATTCCCATATTTGTTCTGTTTTTCTTGATTTG
 GGTATACATTTAAATGTTAATAAAACAAAATGGTGGGGCAATCATTTACATTTTTAGGGATATGTAAT
 TACTAGTTCAGGTGATTGCCACAAGACAAACATGTTAAGAACTTTCCCGTTATTTACGCTCTGTTC
 CTGTTAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGATATTCTTAACATGTTGCTCCT
 TTTACGCTGTGTGGATATGCTGCTTTATAGCCTCTGTATCTAGCTATTGCTTCCCGTACGGCTTTTCGT
 TTTCTCCTCCTTGATAAATCCTGGTTGCTGTCTCTTTTAGAGGAGTTGTGGCCCGTTGTCCGTCAAC
 GTGGCGTGGTGTGCTCTGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCATTGCCACCACCTGTCAA
 CTCCTTTCTGGGACTTTTCGCTTTCCCCCTCCCGATCGCCACGGCAGAACTCATCGCCGCTGCCTTGC
 CCGCTGCTGGACAGGGGCTAGGTTGCTGGGCACTGATAATTCCGTGGTGTGTCTGTGCCTTCTAGTT
 GCCAGCCATCTGTTGTTTGCCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTC
 CTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGG
 GGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCT
 CTATGGCTCTAGAGCATGGCTACGTAGATAAGTAGCATGGCGGGTTAATCATTAACTACACCTGCAGG
 AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGA
 CCAAAGGTCGCCCAGCCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG (SEQ
 ID NO: 147)

Example 4: ceDNA with human SERPINA1 enhancer spacer variants exhibited at least equivalent or superior Factor VIII *in vivo* expression in Rag2 mice when formulated as LNP compositions

[00496] The objective of this study was to determine and compare the effect of LNP-formulated ceDNA on *in vivo* expression in male Rag2 mice. where The ceDNA comprising a Factor VIII transgene, as regulated by: (i) 3x version of hSerpEnh enhancer (3x_SerpEnh –control; 1 basepair spacer “1mer”); (ii) 3x version of hSerpEnh enhancer with 2-mer spacers (3x_hSerpEnh_”2mer” 2 bp spacers_v9) placed between hSerpEnh enhancer element repeats; or (iii) 3x version of hSerpEnh

enhancer with 11-mer spacers (3x_hSerpEnh_11mer spacers_FOXA) placed between hSerpEnh element repeats (*see Table 5*).

Table 5. Test material administration

Group No.	No. of Animals	Test Material	Dose Levels (mg/kg)	Dose Volume (mL/kg)	Dosing Regimen ROA	Terminal Time Point
1	5	PBS	0	5	Once on Day 0 by IV	Day 42
2	5	3x_hSerpEnh (Control)	0.5			
3	5	3x_hSerpEnh (Control)	2.0			
4	5	3x_hSerpEnh_2mer_spacers_v9	0.5			
5	5	3x_hSerpEnh_2mer_spacers_v9	2.0			
6	5	3x_hSerpEnh_11mer_spacers_FOXA	0.5			
7	5	3x_hSerpEnh_11mer_spacers_FOXA	2.0			

[00497] The mice were dosed intravenously once at Day 0 at a low dose of 0.5 mg/kg or a high dose of 2.0 mg/kg ($n = 5$) and the Factor VIII expression was measured at Days 7, 14, 21, and 28. As shown in **FIG. 13**, the treated mice exhibited dose-dependent response to the administered LNP formulations comprising various ceDNA 3x hSerpEnh spacer variants listed above in that a high dose of 2.0 mg/kg consistently resulted in higher Factor VIII expression in the mice, as compared to a low dose of 0.5 mg/kg. Additionally, it was observed that the 3x version of hSerpEnh enhancers that have either 2bp or 11 bp spacer exhibited at least equivalent or higher Factor VIII expression in the mice, as compared to the control 3x_hSerpEnh enhancer having a 1bp spacer.

Example 5: ceDNA with Chinese Tree Shrew SERPINA1 enhancer variants exhibited at least equivalent or superior Factor VIII *in vivo* expression in C57BL/6J mice when administered via hydrodynamic tail vein injection

[00498] The objective of this study was to determine and compare the effect of ceDNA on *in vivo* expression in male C57BL/6J mice. The ceDNA comprised a Factor VIII transgene, as regulated by: (i) 3x version of hSerpEnh enhancer (3x_SerpEnh with 1 bp spacer “1mer”) – control); (ii) 3x version of Chinese Tree Shrew SerpEnh enhancer (3x_ChineseTreeShrew with 1bp spacer); (iii) 3x version of hSerpEnh enhancer with HNF4 and FOXA transcription factor consensus sites and secondary structure formation minimization (3x_HNF4_FOXA_v1_SecondaryStruct_min_v2 with 1bp spacer); or (iv) 3x version of Chinese Tree Shrew SerpEnh enhancer with CpG minimization (3x_ChineseTreeShrew_CpG_min with 1bp spacer) (*see Table 6*).

Table 6. Test material administration

Group	No. Animals	Test Material	Dose Levels (µg/an)	Dose Volume (ml/kg)	Route of Administration (ROA)	Dosage Regimen/Frequency
A	5	PBS	N/A	90-100 ml/kg	HDIV	Once on Day 0
B	5	3x_hSerpEnh (Control)	50 ng			
C	5	3x_ChineseTreeShrew				
D	5	3x_HNF4_FOXA_v1_SecondaryStruct_min_v2				
E	5	3x_ChineseTreeShrew_CpG_min				

[00499] The mice were dosed intravenously via hydrodynamic tail vein injection once at Day 0 at a dose of 50 ng ($n = 5$) and the Factor VIII expression was measured at Days 1 and 3. As shown in **FIG. 14**, the 3x version of Chinese Tree Shrew SerpEnh enhancers and 3x version of hSerpEnh enhancer with HNF4 and FOXA transcription factor consensus sites and secondary structure formation minimization exhibited at least equivalent or higher Factor VIII expression in the mice, as compared to the control 3x_hSerpEnh enhancer.

Example 6: ceDNA with Bushbaby SERPINA1 enhancer variants exhibited superior Factor VIII *in vivo* expression in C57BL/6J mice when administered via hydrodynamic tail vein injection

[00500] The objective of this study was to determine and compare the effect of ceDNA on *in vivo* expression in male C57BL/6J mice, whereby the ceDNA comprised a Factor VIII transgene, as regulated by: (i) 3x version of hSerpEnh enhancer (3x_SerpEnh – positive control); (ii) 3x version of Bushbaby SerpEnh enhancer with adenosine spacer between every two copies of the enhancer (3x_Bushbaby_Aspacers – Sample 1); or (iii) 3x version of Bushbaby SerpEnh enhancer with adenosine spacer between every two copies of the enhancer (3x_Bushbaby_Aspacers – Sample 2); (*see Table 7*).

Table 7. Test material administration

Group	No. Animals	Test Material	Dose Levels (µg/an)	Endpoint
A	5	PBS	10 ng on Day 0	Citrate plasma on Day 3
B	5	3x_hSerpEnh (Control)		
C	5	3x_Bushbaby_Aspacers (Sample 1)		
D	5	3x_Bushbaby_Aspacers (Sample 2)		

[00501] The mice were dosed hydrodynamically via tail vein injection once at Day 0 at a dose of 10 ng ($n = 5$) and the Factor VIII expression was measured at Day 3. As shown in **FIG. 15**, the 3x version

of Bushbaby SerpEnh enhancer with adenosine nucleotide spacers exhibited higher Factor VIII expression in the mice, as compared to the control 3x_hSerpEnh enhancer.

Example 7: ceDNA with Armadillo or Tibetan Antelope SERPINA1 enhancer variants exhibited at least equivalent Factor VIII *in vivo* expression in C57BL/6J mice when administered via hydrodynamic tail vein injection

[00502] The objective of this study was to determine and compare the effect of ceDNA on *in vivo* expression in male C57BL/6J mice, whereby the ceDNA comprised a Factor VIII transgene, as regulated by: (i) 3x version of hSerpEnh enhancer (3x_SerpEnh –control); (ii) 3x version of Tibetan Antelope SerpEnh enhancer (3x_Tibetan_antelope_SERPINA1_enhancer); or (iii) 3x version of Armadillo SerpEnh enhancer with CpG minimization (3x_Armadillo_CpGmin_SERPINA1 enhancer) (*see Table 8*).

Table 8. Test material administration

Group No.	No. of Animals	Test Material	Dose Levels (ng/an)	Dose Volume (mL/kg)	Dosing Regimen ROA	Terminal Time Point
1	4	PBS	0	90 – 100 (set volume)	Once on Day 0 by HDIV	Day 3
2	4	3x_hSerpEnh (Control)	25			
3	4		50			
4	4		100			
5	4	3x_Tibetan_antelope_SERPINA1_enhancer	25			
6	4		50			
7	4		100			
8	4	3x_Armadillo_CpGmin_SERPINA1_enhancer	25			
9	4		50			
10	4		100			
11	4	3x_ChineseTreeShrew	25			
12	4		50			
13	4		100			
14	4	3x_ChineseTreeShrew_CpGmin	25			
15	4		50			
16	4		100			
17	4	3x_Bushbaby_Aspacers	25			

18	4		50			
19	4		100			

[00503] The mice were dosed intravenously via hydrodynamic tail vein injection once at Day 0 at three different dose levels: 25 ng/an, 50 ng/an, 100 ng/an ($n = 4$) and the Factor VIII expression was measured at Day 3. As shown in **FIG. 16A** and **FIG. 16B**, the administered ceDNA constructs were dose-responsive in that higher doses of 50 ng/an and 100 ng/an resulted in higher Factor VIII expression in the mice, as compared to a low dose of 25 ng/an. Additionally, it was observed that all tested enhancers exhibited at least equivalent Factor VIII expression in the mice, as compared to the control 3x_hSerpEnh enhancer.

Example 8: High-throughput screening of human SERPINA1 enhancer variants, BushBaby SERPINA1 enhancer variants, Chinese Tree Shrew SERPINA1 enhancer variants, and human SERPINA1 enhancer variants with HNF4 and FOXA transcription factor consensus sites

[00504] The following enhancer sequence variants were generated to evaluate their ability to drive expression by high-throughput expression screening:

(1) all single nucleotide substitution and adjacent di-nucleotide substitution variants for human SERPINA1 enhancer (*e.g.*, single nucleotide variants and adjacent di-nucleotide variants of SEQ ID NO: 81), Chinese Tree Shrew modified SERPINA1 enhancer (*e.g.*, single nucleotide substitution and adjacent di-nucleotide substitution variants for SEQ ID NO: 122), BushBaby SERPINA1 enhancer (*e.g.*, single nucleotide variants and adjacent di-nucleotide variants of SEQ ID NO: 83), and human SERPINA1 enhancer variants with HNF4 and FOXA transcription factor consensus sites HNF4_FOXA enhancer *i.e.*, single nucleotide variants and adjacent di-nucleotide variants of SEQ ID NO: 85);

(2) selected variants of the human SERPINA1 enhancer with four modified nucleotides (*i.e.*, four nucleotide substitution variants of SEQ ID NO: 81); and

(3) 50 random permutations of the human SERPINA1 enhancer used as a negative control (*i.e.*, random permutations of SEQ ID NO: 81). In total, 5,151 unique sequences were screened. To map sequencing reads back to enhancer sequences, all screened sequenced were associated with one or more 10 nucleotide barcodes with a minimum Levenshtein distance of two nucleotides between all barcodes. Original enhancer sequences were each associated with 200 barcodes and all variants were associated with one barcode. Enhancer sequences and barcodes were generated with custom MATLAB scripts.

[00505] An oligo pool of the enhancers was ordered from Twist Biosciences and the plasmid library, pHTS002L (**FIG. 17**), with luciferase as the reporter gene, was constructed. HepG2 cells (ATCC, VA) were cultured in DMEM (Dulbecco's Modified Eagle Medium) medium (ThermoFisher, MA) with 10% fetal bovine serum (ThermoFisher, MA) at 37 °C. The day before transfection, cells were harvested with 0.25% trypsin (ThermoFisher, MA) and seeded at the density of 600,000 cells per well on 6-well

collagen-coated plates (VWR, PA). 2 ug plasmid were transfected per well with TransfeX (ATCC, VA) according to the manufacture's manual. The cells were harvested 24 hours after transfection and total RNA were extracted with the RNAeasy plus kit (Qiagen, Germany). 1 ug RNA or 10 ng plasmid DNA were used for amplicon production using the primers in **Table 9** with SuperScript™ IV One-Step RT-PCR System (ThermoFisher, MA) according to the manufacture's manual. Amplicons contained the barcode associated with each enhancer. The concentrations of the 6 indexed amplicons were measured with Qubit (ThermoFisher, MA). The amplicons were sequenced with Illumina Miseq (75bp x 2) at MIT BioMicro Center.

Table 9. Primers used for amplicon production

Forward	SEQ ID NO: 148	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCT TCCGATCTGGAGGGAAGATTGCTGTGTGATAG
Reverse 1	SEQ ID NO: 149	CAAGCAGAAGACGGCATACGAGATGTGACTGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT CAG GAA CAG AGC GTA AAT AAC GGG
Reverse 2	SEQ ID NO: 150	CAAGCAGAAGACGGCATACGAGATCTGCAAGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT CAG GAA CAG AGC GTA AAT AAC GGG
Reverse 3	SEQ ID NO: 151	CAAGCAGAAGACGGCATACGAGATACCATGGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT CAG GAA CAG AGC GTA AAT AAC GGG
Reverse 4	SEQ ID NO: 152	CAAGCAGAAGACGGCATACGAGATGAACGTGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT CAG GAA CAG AGC GTA AAT AAC GGG
Reverse 5	SEQ ID NO: 153	CAAGCAGAAGACGGCATACGAGATACTAGTGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT CAG GAA CAG AGC GTA AAT AAC GGG
Reverse 6	SEQ ID NO: 154	CAAGCAGAAGACGGCATACGAGATCGTTACGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT CAG GAA CAG AGC GTA AAT AAC GGG

[00506] Reads were filtered out that (1) did not contain the expected sequence for the 10 nucleotides up- and downstream of the barcode and (2) contained quality scores less than 20 in the barcode (FASTX-Toolkit). Barcode counts for each RNA sample were normalized to the corresponding barcode counts for an input DNA sample and mapped back to their associated enhancer sequences (custom MATLAB script). Comparisons for two biological replicates are shown in **FIGS. 18A-18D**.

[00507] The enhancer variants that exhibited higher expression levels than the expression levels of their respective original sequences (*i.e.*, SEQ ID NOs: 81, 122, 83, or 85) are listed in **Table 10** (human SERPINA1 enhancer variants); **Table 11** (Chinese Tree Shrew SERPINA1 enhancer variants), **Table 12** (BushBaby SERPINA 1 enhancer variants, and **Table 13** (human SERPINA1 enhancer variants with HNF4 and FOXA transcription factor consensus sites). Of note, single nucleotide substitution variants of the human SERPINA1 enhancer, single nucleotide substitution variants of the Chinese Tree Shrew SERPINA1 enhancer, and single nucleotide substitution variants of the Bushbaby SERPINA1 enhancer that each carry a CTAAG -> CAAAG mutation were consistently among the highest expression variants among their respective variant populations (*see FIGS. 18A-18C*). As indicated in **FIG. 19** that shows the alignment of multiple SERPINA1 enhancer sequences, the CAAAG sequence is located in the

HN4 transcription factor consensus site in SEQ ID NO: 85, the sequence of the human SERPINA1 enhancer variants with HN4 and FOXA transcription factor consensus sites HN4_FOXA enhancer. When the CAAAG sequence is modified to CTAAG, the expression levels of the reciprocal variant were compromised (*see* **FIG. 18D**).

Table 10. Single, adjacent di-, or four nucleotide substitution variants of human SERPINA1 enhancer with higher luciferase expression than original sequence
SEQ ID NO: 81

SEQ ID	Human SERPINA1 enhancer variant	Sequence
155	202_Human_monoMut_G1C_n1	CGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
156	204_Human_monoMut_G2A_n1	GAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
157	206_Human_monoMut_G2T_n1	GTGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
158	210_Human_monoMut_G4A_n1	GGGAGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
159	214_Human_monoMut_G5C_n1	GGGCAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
160	216_Human_monoMut_A6C_n1	GGGGCGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
161	220_Human_monoMut_G7C_n1	GGGGACGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
162	225_Human_monoMut_C9A_n1	GGGGAGGATGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
163	226_Human_monoMut_C9G_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
164	229_Human_monoMut_T10C_n1	GGGGAGGCCGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
165	230_Human_monoMut_T10G_n1	GGGGAGGCGGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
166	235_Human_monoMut_C12G_n1	GGGGAGGCTGGTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
167	236_Human_monoMut_C12T_n1	GGGGAGGCTGTTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
168	245_Human_monoMut_G15T_n1	GGGGAGGCTGCTGTTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
169	247_Human_monoMut_T16C_n1	GGGGAGGCTGCTGGCGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C

170	250_Human_monoMut_G17C_n1	GGGGAGGCTGCTGGTCAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
171	251_Human_monoMut_G17T_n1	GGGGAGGCTGCTGGTTAATAATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
172	260_Human_monoMut_T20G_n1	GGGGAGGCTGCTGGTGAAGATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
173	261_Human_monoMut_A21C_n1	GGGGAGGCTGCTGGTGAATCTTTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
174	263_Human_monoMut_A21T_n1	GGGGAGGCTGCTGGTGAATTTTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
175	264_Human_monoMut_T22A_n1	GGGGAGGCTGCTGGTGAATAATAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
176	269_Human_monoMut_T23G_n1	GGGGAGGCTGCTGGTGAATATGAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
177	270_Human_monoMut_A24C_n1	GGGGAGGCTGCTGGTGAATATTCA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
178	274_Human_monoMut_A25G_n1	GGGGAGGCTGCTGGTGAATATTAG CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
179	275_Human_monoMut_A25T_n1	GGGGAGGCTGCTGGTGAATATTAT CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
180	278_Human_monoMut_C26T_n1	GGGGAGGCTGCTGGTGAATAATTAA TCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
181	282_Human_monoMut_A28C_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
182	283_Human_monoMut_A28G_n1	GGGGAGGCTGCTGGTGAATATTAA CCGAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
183	284_Human_monoMut_A28T_n1	GGGGAGGCTGCTGGTGAATATTAA CCTAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
184	295_Human_monoMut_T32C_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGCCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
185	301_Human_monoMut_A34G_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCGCCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
186	303_Human_monoMut_C35A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
187	304_Human_monoMut_C35G_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAGCCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C

188	305_Human_monoMut_C35T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCCAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
189	308_Human_monoMut_C36T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCTCCAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
190	314_Human_monoMut_C38T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCTAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
191	315_Human_monoMut_A39C_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
192	316_Human_monoMut_A39G_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
193	317_Human_monoMut_A39T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCTGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
194	325_Human_monoMut_T42C_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCAGTCTCGGAGGAGCAAAACAGGGGCTAAGTCCA C
195	328_Human_monoMut_A43G_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCAGTTGTGCGGAGGAGCAAAACAGGGGCTAAGTCCA C
196	331_Human_monoMut_T44C_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCAGTTACCGGAGGAGCAAAACAGGGGCTAAGTCCA C
197	332_Human_monoMut_T44G_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCAGTTAGCGGAGGAGCAAAACAGGGGCTAAGTCCA C
198	333_Human_monoMut_C45A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCAGTTATAGGAGGAGCAAAACAGGGGCTAAGTCCA C
199	334_Human_monoMut_C45G_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCAGTTATGGGAGGAGCAAAACAGGGGCTAAGTCCA C
200	337_Human_monoMut_G46C_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
201	344_Human_monoMut_A48T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCAGTTATCGGTGGAGCAAAACAGGGGCTAAGTCCA C
202	345_Human_monoMut_G49A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCAGTTATCGGAAGAGCAAAACAGGGGCTAAGTCCA C
203	346_Human_monoMut_G49C_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCAGTTATCGGACGAGCAAAACAGGGGCTAAGTCCA C
204	347_Human_monoMut_G49T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCAGTTATCGGATGAGCAAAACAGGGGCTAAGTCCA C
205	351_Human_monoMut_A51C_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCAGTTATCGGAGGCGCAAAACAGGGGCTAAGTCCA C

206	353_Human_monoMut_A51T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGAGGTGCAAAACAGGGGCTAAGTCCA C
207	354_Human_monoMut_G52A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGAGGAA CAAAACAGGGGCTAAGTCCA C
208	356_Human_monoMut_G52T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGAGGAT CAAAACAGGGGCTAAGTCCA C
209	371_Human_monoMut_C57T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGAGGAGCAAAATAGGGGCTAAGTCCA C
218	375_Human_monoMut_G59A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGAGGAGCAAAACAGGGGCTAAGTCCA C
219	380_Human_monoMut_G60T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGAGGAGCAAAACAGTGGCTAAGTCCA C
220	381_Human_monoMut_G61A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGAGGAGCAAAACAGGAGCTAAGTCCA C
221	383_Human_monoMut_G61T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGAGGAGCAAAACAGGTGCTAAGTCCA C
222	386_Human_monoMut_G62T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGAGGAGCAAAACAGGGTCTAAGTCCA C
223	390_Human_monoMut_T64A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGAGGAGCAAAACAGGGGCAAAAGTCCA C
224	392_Human_monoMut_T64G_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGAGGAGCAAAACAGGGGCGAAAGTCCA C
225	394_Human_monoMut_A65G_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGAGGAGCAAAACAGGGGCTGAGTCCA C
226	405_Human_monoMut_C69A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGAGGAGCAAAACAGGGGCTAAGTACA C
227	414_Human_monoMut_C72A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGAGGAGCAAAACAGGGGCTAAGTCCA A
228	415_Human_monoMut_C72G_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGAGGAGCAAAACAGGGGCTAAGTCCA G
229	421_Human_diMut_GG1CC_n1	CCGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGAGGAGCAAAACAGGGGCTAAGTCCA C
230	427_Human_diMut_GG2AC_n1	GACGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGAGGAGCAAAACAGGGGCTAAGTCCA C
231	429_Human_diMut_GG2CA_n1	GCAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGAGGAGCAAAACAGGGGCTAAGTCCA C

232	430_Human_diMut_GG2CC_n1	GCCGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
233	439_Human_diMut_GG3CC_n1	GGCCGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
234	440_Human_diMut_GG3CT_n1	GGCTGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
235	443_Human_diMut_GG3TT_n1	GGTTGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
236	444_Human_diMut_GG4AA_n1	GGGAAAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
237	447_Human_diMut_GG4CA_n1	GGGCAAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
238	448_Human_diMut_GG4CC_n1	GGGCCAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
239	452_Human_diMut_GG4TT_n1	GGGTTAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
240	453_Human_diMut_GA5AC_n1	GGGACGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
241	456_Human_diMut_GA5CC_n1	GGGGCCGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
242	461_Human_diMut_GA5TT_n1	GGGGTTGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
243	462_Human_diMut_AG6CA_n1	GGGGCAGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
244	464_Human_diMut_AG6CT_n1	GGGGGCTGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
245	465_Human_diMut_AG6GA_n1	GGGGGAGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
246	468_Human_diMut_AG6TA_n1	GGGGTAGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
247	469_Human_diMut_AG6TC_n1	GGGGTCGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
248	470_Human_diMut_AG6TT_n1	GGGGTTGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
249	471_Human_diMut_GG7AA_n1	GGGGGAAACTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C

250	472_Human_diMut_GG7AC_n1	GGGGAAACCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
251	473_Human_diMut_GG7AT_n1	GGGGAAATCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
252	475_Human_diMut_GG7CC_n1	GGGGACCCCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
253	478_Human_diMut_GG7TC_n1	GGGGATCCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
254	482_Human_diMut_GC8AT_n1	GGGGAGATTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
255	483_Human_diMut_GC8CA_n1	GGGGAGCATGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
256	484_Human_diMut_GC8CG_n1	GGGGAGCGTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
257	488_Human_diMut_GC8TT_n1	GGGGAGTTTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
258	490_Human_diMut_CT9AC_n1	GGGGAGGACGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
259	491_Human_diMut_CT9AG_n1	GGGGAGGAGGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
260	492_Human_diMut_CT9GA_n1	GGGGAGGGAGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
261	493_Human_diMut_CT9GC_n1	GGGGAGGGCGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
262	494_Human_diMut_CT9GG_n1	GGGGAGGGGGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
263	495_Human_diMut_CT9TA_n1	GGGGAGGTAGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
264	496_Human_diMut_CT9TC_n1	GGGGAGGTCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
265	502_Human_diMut_TG10CC_n1	GGGGAGGCCCTCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
266	503_Human_diMut_IG10CT_n1	GGGGAGGCCCTCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
267	504_Human_diMut_TG10GA_n1	GGGGAGGCGACTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C

268	506_Human_diMut_TG10GT_n1	GGGGAGGCGTCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
269	507_Human_diMut_GC11AA_n1	GGGGAGGCTAATGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
270	510_Human_diMut_GC11CA_n1	GGGGAGGCTCATGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
271	511_Human_diMut_GC11CG_n1	GGGGAGGCTCGTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
272	512_Human_diMut_GC11CT_n1	GGGGAGGCTCTTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
273	516_Human_diMut_CT12AA_n1	GGGGAGGCTGAAGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
274	517_Human_diMut_CT12AC_n1	GGGGAGGCTGACGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
275	521_Human_diMut_CT12GG_n1	GGGGAGGCTGGGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
276	522_Human_diMut_CT12TA_n1	GGGGAGGCTGTAGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
277	523_Human_diMut_CT12TC_n1	GGGGAGGCTGTCGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
278	524_Human_diMut_CT12TG_n1	GGGGAGGCTGTGGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
279	528_Human_diMut_IG13CA_n1	GGGGAGGCTGCCAGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
280	544_Human_diMut_GT15AC_n1	GGGGAGGCTGCTGACGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
281	549_Human_diMut_GT15TA_n1	GGGGAGGCTGCTGTAGAAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
282	551_Human_diMut_GT15TG_n1	GGGGAGGCTGCTGTGGAAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
283	552_Human_diMut_TG16AA_n1	GGGGAGGCTGCTGGAAAAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
284	553_Human_diMut_IG16AC_n1	GGGGAGGCTGCTGGACAAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
285	554_Human_diMut_TG16AT_n1	GGGGAGGCTGCTGGATAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C

286	555_Human_diMut_TG16CA_n1	GGGGAGGCTGCTGGCAATATTAACCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
287	557_Human_diMut_IG16CT_n1	GGGGAGGCTGCTGGCTAATATTAACCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
288	558_Human_diMut_IG16GA_n1	GGGGAGGCTGCTGGGAATATTAACCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
289	561_Human_diMut_GA17AC_n1	GGGGAGGCTGCTGGTACATATTAACCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
290	563_Human_diMut_GA17AT_n1	GGGGAGGCTGCTGGTATATTAACCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
291	565_Human_diMut_GA17CG_n1	GGGGAGGCTGCTGGTGCATATTAACCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
292	568_Human_diMut_GA17TG_n1	GGGGAGGCTGCTGGTTGATATTAACCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
293	572_Human_diMut_AA18CT_n1	GGGGAGGCTGCTGGTGCTTATTAACCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
294	578_Human_diMut_AA18TT_n1	GGGGAGGCTGCTGGTGTTTATTAACCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
295	583_Human_diMut_AT19GC_n1	GGGGAGGCTGCTGGTGACATTAACCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
296	584_Human_diMut_AT19GG_n1	GGGGAGGCTGCTGGTGAGGATTAACCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
297	585_Human_diMut_AT19TA_n1	GGGGAGGCTGCTGGTGATAATTAACCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
298	586_Human_diMut_AT19TC_n1	GGGGAGGCTGCTGGTGATCATTAACCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
299	587_Human_diMut_AT19TG_n1	GGGGAGGCTGCTGGTGATGATTAACCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
300	589_Human_diMut_TA20AG_n1	GGGGAGGCTGCTGGTGAAGTTAAACCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
301	591_Human_diMut_TA20CC_n1	GGGGAGGCTGCTGGTGAACCTTAACCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
302	592_Human_diMut_TA20CG_n1	GGGGAGGCTGCTGGTGAACGTTAAACCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
303	593_Human_diMut_TA20CT_n1	GGGGAGGCTGCTGGTGAACCTTAACCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C

304	596_Human_diMut_TA20GT_n1	GGGGAGGCTGCTGGTGAATTTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
305	597_Human_diMut_AT21CA_n1	GGGGAGGCTGCTGGTGAATCATAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
306	601_Human_diMut_AT21GC_n1	GGGGAGGCTGCTGGTGAATGCTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
308	602_Human_diMut_AT21GG_n1	GGGGAGGCTGCTGGTGAATGGTAAC CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
309	604_Human_diMut_AT21TC_n1	GGGGAGGCTGCTGGTGAATTCTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
310	608_Human_diMut_TT22AG_n1	GGGGAGGCTGCTGGTGAATAAGAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
311	614_Human_diMut_TT22GG_n1	GGGGAGGCTGCTGGTGAATAGAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
312	616_Human_diMut_TA23AG_n1	GGGGAGGCTGCTGGTGAATATCGA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
313	619_Human_diMut_TA23CG_n1	GGGGAGGCTGCTGGTGAATATCGA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
314	621_Human_diMut_TA23GC_n1	GGGGAGGCTGCTGGTGAATATGCA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
315	623_Human_diMut_TA23GT_n1	GGGGAGGCTGCTGGTGAATATGTA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
316	624_Human_diMut_AA24CC_n1	GGGGAGGCTGCTGGTGAATATTCCC CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
317	629_Human_diMut_AA24GT_n1	GGGGAGGCTGCTGGTGAATATTGT CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
318	650_Human_diMut_CC26TT_n1	GGGGAGGCTGCTGGTGAATATTAA TTAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
319	667_Human_diMut_AA28TG_n1	GGGGAGGCTGCTGGTGAATATTAA CCTGGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
320	678_Human_diMut_GG30AA_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAAATCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
321	679_Human_diMut_GG30AC_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAACTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
322	682_Human_diMut_GG30CC_n1	GGGGAGGCTGCTGGTGAATATTAA CCAACCTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C

323	688_Human_diMut_GT31AC_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGACACCCAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
324	689_Human_diMut_GT31AG_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGACACCCAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
325	694_Human_diMut_GT31TC_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGTCCACCCAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
326	698_Human_diMut_TC32AT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGATACCCAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
327	704_Human_diMut_TC32GT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTACCCAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
328	705_Human_diMut_CA33AC_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTACCCAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
329	706_Human_diMut_CA33AG_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTAGCCCCAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
330	709_Human_diMut_CA33GG_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTGCCCCAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
331	720_Human_diMut_AC34TA_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCTACCCAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
332	723_Human_diMut_CC35AA_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAACCCAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
323	728_Human_diMut_CC35GT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAAGTCCAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
324	733_Human_diMut_CC36AG_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAAGTCCAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
325	734_Human_diMut_CC36AT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAATCAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
326	736_Human_diMut_CC36GG_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGTCAAGTCCAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
327	737_Human_diMut_CC36GT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGTCAAGTCCAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
328	738_Human_diMut_CC36TA_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGTCAAGTCCAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
329	739_Human_diMut_CC36TG_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGTCAAGTCCAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
340	740_Human_diMut_CC36TT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGTCAAGTCCAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C

341	741_Human_diMut_CC37AA_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCAAAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
342	743_Human_diMut_CC37AT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCATAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
343	746_Human_diMut_CC37GT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCGTAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
344	747_Human_diMut_CC37TA_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCTAAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
345	748_Human_diMut_CC37TG_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCTGAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
346	750_Human_diMut_CA38AC_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCACGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
347	752_Human_diMut_CA38AT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCATGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
348	753_Human_diMut_CA38GC_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCGCGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
349	754_Human_diMut_CA38GG_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCGGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
350	755_Human_diMut_CA38GT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCGTGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
351	756_Human_diMut_CA38TC_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCTCGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
352	757_Human_diMut_CA38TG_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCTGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
353	758_Human_diMut_CA38TT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCTTGTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
354	759_Human_diMut_AG39CA_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAATTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
355	765_Human_diMut_AG39TA_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCTATTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
356	766_Human_diMut_AG39TC_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCTTTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
357	767_Human_diMut_AG39TT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCTTTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
358	772_Human_diMut_GT40CC_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAACCTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C

359	776_Human_diMut_GT40TG_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCATGTATCGAGGAGCAAAACAGGGGCTAAGTCCA C
360	777_Human_diMut_IT41AA_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCACAGAAATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
361	781_Human_diMut_IT41CC_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCACGCCATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
362	787_Human_diMut_TA42AG_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCACCCAGTAGTCGGAGGAGCAAAACAGGGGCTAAGTCCA C
363	790_Human_diMut_TA42CG_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCACCCAGTCGTCGGAGGAGCAAAACAGGGGCTAAGTCCA C
364	794_Human_diMut_TA42GT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCACCCAGTGTTCGGAGGAGCAAAACAGGGGCTAAGTCCA C
365	795_Human_diMut_AT43CA_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCACCCAGTTTCGGAGGAGCAAAACAGGGGCTAAGTCCA C
366	799_Human_diMut_AT43GC_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCACCCAGTTTCGGAGGAGCAAAACAGGGGCTAAGTCCA C
367	802_Human_diMut_AT43TC_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCACCCAGTTTCGGAGGAGCAAAACAGGGGCTAAGTCCA C
368	806_Human_diMut_TC44AT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCACCCAGTTTAA TGGAGGAGCAAAACAGGGGCTAAGTCCA C
369	807_Human_diMut_TC44CA_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCACCCAGTTACAGGAGGAGCAAAACAGGGGCTAAGTCCA C
370	811_Human_diMut_TC44GG_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCACCCAGTTAGGGAGGAGCAAAACAGGGGCTAAGTCCA C
371	815_Human_diMut_CG45AT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCACCCAGTTATATAGGAGCAAAACAGGGGCTAAGTCCA C
372	818_Human_diMut_CG45GT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCACCCAGTTATGTGAGGAGCAAAACAGGGGCTAAGTCCA C
373	821_Human_diMut_CG45TT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCACCCAGTTATTGAGGAGCAAAACAGGGGCTAAGTCCA C
374	822_Human_diMut_GG46AA_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCACCCAGTTATCAAAAGGAGCAAAACAGGGGCTAAGTCCA C
375	824_Human_diMut_GG46AT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCACCCAGTTATCATAGGAGCAAAACAGGGGCTAAGTCCA C
376	826_Human_diMut_GG46CC_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCACCCAGTTATCCAGGAGCAAAACAGGGGCTAAGTCCA C

377	827_Human_diMut_GG46CT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCCTAGGAGCAAAACAGGGGCTAAGTCCA C
378	828_Human_diMut_GG46TA_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCTAAGGAGCAAAACAGGGGCTAAGTCCA C
379	830_Human_diMut_GG46TT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCTTAGGAGCAAAACAGGGGCTAAGTCCA C
380	832_Human_diMut_GA47AG_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGAGGGAGCAAAACAGGGGCTAAGTCCA C
381	833_Human_diMut_GA47AT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGATGGAGCAAAACAGGGGCTAAGTCCA C
382	834_Human_diMut_GA47CC_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGCCGGAGCAAAACAGGGGCTAAGTCCA C
383	836_Human_diMut_GA47CT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGCTGGAGCAAAACAGGGGCTAAGTCCA C
384	839_Human_diMut_GA47TT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGTTGGAGCAAAACAGGGGCTAAGTCCA C
385	840_Human_diMut_AG48CA_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGCAGAGCAAAACAGGGGCTAAGTCCA C
386	842_Human_diMut_AG48CT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGCTGAGCAAAACAGGGGCTAAGTCCA C
387	843_Human_diMut_AG48GA_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGGAGAGCAAAACAGGGGCTAAGTCCA C
388	845_Human_diMut_AG48GT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGGTGAGCAAAACAGGGGCTAAGTCCA C
389	846_Human_diMut_AG48TA_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGTAGAGCAAAACAGGGGCTAAGTCCA C
390	848_Human_diMut_AG48TT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGTTGAGCAAAACAGGGGCTAAGTCCA C
391	849_Human_diMut_GG49AA_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAAAAGCAAAACAGGGGCTAAGTCCA C
392	850_Human_diMut_GG49AC_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAACAGCAAAACAGGGGCTAAGTCCA C
393	851_Human_diMut_GG49AT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAATAGCAAAACAGGGGCTAAGTCCA C
394	852_Human_diMut_GG49CA_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGACAAGCAAAACAGGGGCTAAGTCCA C

395	853_Human_diMut_GG49CC_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGACCAAGCAAAACAGGGGCTAAGTCCA C
396	854_Human_diMut_GG49CT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGACTAGCAAAACAGGGGCTAAGTCCA C
397	855_Human_diMut_GG49TA_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGATAAGCAAAACAGGGGCTAAGTCCA C
398	856_Human_diMut_GG49TC_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGATCAGCAAAACAGGGGCTAAGTCCA C
399	857_Human_diMut_GG49TT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGATTAGCAAAACAGGGGCTAAGTCCA C
400	859_Human_diMut_GA50AG_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGAGGCAAAACAGGGGCTAAGTCCA C
401	866_Human_diMut_GA50TT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGTTGCAAAACAGGGGCTAAGTCCA C
402	867_Human_diMut_AG51CA_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGCCAAGCAAGGGGCTAAGTCCA C
403	868_Human_diMut_AG51CC_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGCCCAAGCAAGGGGCTAAGTCCA C
404	876_Human_diMut_GC52AA_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAAAAACAGGGGCTAAGTCCA C
405	881_Human_diMut_GC52CT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGACTAAACAGGGGCTAAGTCCA C
406	885_Human_diMut_CA53AC_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGACAAACAGGGGCTAAGTCCA C
407	886_Human_diMut_CA53AG_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGAGAAACAGGGGCTAAGTCCA C
408	887_Human_diMut_CA53AT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGATAACAGGGGCTAAGTCCA C
409	891_Human_diMut_CA53TC_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGTCAACAGGGGCTAAGTCCA C
410	892_Human_diMut_CA53TG_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGTGAACAGGGGCTAAGTCCA C
411	898_Human_diMut_AA54GG_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCGGACAGGGGCTAAGTCCA C
412	900_Human_diMut_AA54TC_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCTCACAGGGGCTAAGTCCA C

413	906_Human_diMut_AA55GC_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAGCCAGGGGCTAAGTCCA C
414	933_Human_diMut_AG58GA_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAA CAGGGCTAAGTCCA C
415	941_Human_diMut_GG59AT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAA CCAATGGCTAAGTCCA C
416	943_Human_diMut_GG59CC_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAA CACCCGGCTAAGTCCA C
417	947_Human_diMut_GG59TT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAA CATTTGGCTAAGTCCA C
418	950_Human_diMut_GG60AT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAA CAGATGCTAAGTCCA C
419	955_Human_diMut_GG60TC_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAA CAGTCGCTAAGTCCA C
420	957_Human_diMut_GG61AA_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAA CAGGAACCTAAGTCCA C
421	958_Human_diMut_GG61AC_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAA CAGGACCTAAGTCCA C
422	960_Human_diMut_GG61CA_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAA CAGGCACCTAAGTCCA C
423	964_Human_diMut_GG61TC_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAA CAGGTCCTAAGTCCA C
424	965_Human_diMut_GG61TT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAA CAGGTTCTAAGTCCA C
425	968_Human_diMut_GC62AT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAA CAGGGATTAAAGTCCA C
426	969_Human_diMut_GC62CA_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAA CAGGGCATAAAGTCCA C
427	980_Human_diMut_CT63GG_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAA CAGGGGGGAAGTCCA C
428	981_Human_diMut_CT63TA_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAA CAGGGGTAAAGTCCA C
429	983_Human_diMut_CT63IG_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAA CAGGGGTAAAGTCCA C
430	985_Human_diMut_TA64AG_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAA CAGGGGCAGAGTCCA C

431	986_Human_diMut_TA64AT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCATAGTCCA C
432	994_Human_diMut_AA65CG_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTCGGTCCA C
433	1022_Human_diMut_TC68AT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGATCA C
434	1023_Human_diMut_TC68CA_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGCACA C
435	1028_Human_diMut_TC68GT_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGGTCA C
436	1029_Human_diMut_CC69AA_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTAAA C
437	4267_Human_quadMut_1T_3A_8T_48T_n1	TGAGGAGTCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGTGGAGCAAAACAGGGGCTAAGTCCA C
438	4268_Human_quadMut_1T_3A_8C_52T_n1	TGAGGAGCCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGATCAAAACAGGGGCTAAGTCCA C
439	4269_Human_quadMut_1A_3A_8A_62C_n1	AGAGGAGACTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGCCTAAGTCCA C
440	4272_Human_quadMut_1A_3T_14A_37T_n1	AGTGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
441	4282_Human_quadMut_1T_3A_35A_48G_n1	TGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGGGAGCAAAACAGGGGCTAAGTCCA C
442	4284_Human_quadMut_1A_3T_35T_62A_n1	AGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCCCAGTTATCGGAGGAGCAAAACAGGGACTAAGTCCA C
443	4287_Human_quadMut_1A_3T_37A_46T_n1	AGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCA CAGTTATCTGAGGAGCAAAACAGGGGCTAAGTCCA C
444	4288_Human_quadMut_1A_3T_37T_48T_n1	AGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCTCAGTTATCGGTGGAGCAAAACAGGGGCTAAGTCCA C
445	4289_Human_quadMut_1A_3A_37T_52T_n1	AGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCTCAGTTATCGGAGGATCAAAACAGGGGCTAAGTCCA C
446	4293_Human_quadMut_1A_3A_45T_48T_n1	AGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGTGGAGCAAAACAGGGGCTAAGTCCA C
447	4298_Human_quadMut_1T_3T_46T_52T_n1	TGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCTGAGGATCAAAACAGGGGCTAAGTCCA C
448	4299_Human_quadMut_1T_3T_46T_62A_n1	TGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCTGAGGAGCAAAACAGGGACTAAGTCCA C

449	4310_Human_quadMut_1T_8T_14T_46T_n1	TGGGAGTCTGCTTGTGAATATTAA CCAAGGTCACCC CAGTTATCTGAGGAGCAAA CAGGGGCTAAGTCCA	C
450	4312_Human_quadMut_1T_8C_14T_52T_n1	TGGGAGCCTGCTTGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGATCAAA CAGGGGCTAAGTCCA	C
451	4313_Human_quadMut_1A_8T_14T_62T_n1	AGGGAGTCTGCTTGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGAGCAAA CAGGGTCTAAGTCCA	C
452	4314_Human_quadMut_1A_8C_14A_72A_n1	AGGGAGCCTGCTAGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGAGCAAA CAGGGGCTAAGTCCA	A
453	4315_Human_quadMut_1A_8C_35T_37A_n1	AGGGAGCCTGCTGGTGAATATTAA CCAAGGTCATC CAGTTATCGGAGGAGCAAA CAGGGGCTAAGTCCA	C
454	4317_Human_quadMut_1A_8A_35T_46A_n1	AGGGAGACTGCTGGTGAATATTAA CCAAGGTCATCC CAGTTATCAGAGGAGCAAA CAGGGGCTAAGTCCA	C
455	4318_Human_quadMut_1T_8C_35G_48G_n1	TGGGAGCCTGCTGGTGAATATTAA CCAAGGTCAGCCC CAGTTATCGGGGAGCAAA CAGGGGCTAAGTCCA	C
456	4320_Human_quadMut_1T_8A_35A_62C_n1	TGGGAGACTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGAGCAAA CAGGGCCTAAGTCCA	C
457	4323_Human_quadMut_1A_8A_37A_46A_n1	AGGGAGACTGCTGGTGAATATTAA CCAAGGTCACC CAGTTATCAGAGGAGCAAA CAGGGGCTAAGTCCA	C
458	4325_Human_quadMut_1T_8C_37T_52A_n1	TGGGAGCCTGCTGGTGAATATTAA CCAAGGTCACCT CAGTTATCGGAGGAA CAA CAGGGGCTAAGTCCA	C
459	4340_Human_quadMut_1A_8T_52T_62T_n1	AGGGAGTCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGATCAAA CAGGGTCTAAGTCCA	C
460	4345_Human_quadMut_1T_14A_35G_46T_n1	TGGGAGGCTGCTAGTGAATATTAA CCAAGGTCAGCCC CAGTTATCTGAGGAGCAAA CAGGGGCTAAGTCCA	C
461	4353_Human_quadMut_1T_14T_37A_52T_n1	TGGGAGGCTGCTTGTGAATATTAA CCAAGGTCACC CAGTTATCGGAGGATCAAA CAGGGGCTAAGTCCA	C
462	4371_Human_quadMut_1A_35T_37A_45T_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCATC CAGTTATTGAGGAGCAAA CAGGGGCTAAGTCCA	C
463	4374_Human_quadMut_1A_35A_37A_52T_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAAC CAGTTATCGGAGGATCAAA CAGGGGCTAAGTCCA	C
464	4376_Human_quadMut_1A_35A_37A_72T_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAGCCC CAGTTATTGAGGAGCAAA CAGGGGCTAAGTCCA	T
465	4377_Human_quadMut_1T_35G_45T_46T_n1	TGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAGCCC CAGTTATTGGTGGAGCAAA CAGGGGCTAAGTCCA	C
466	4378_Human_quadMut_1T_35G_45T_48T_n1	TGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAGCCC CAGTTATTGGTGGAGCAAA CAGGGGCTAAGTCCA	C

467	4379_Human_quadMut_1A_35G_45T_52C_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAGCCCAAGTTATTGGAGGACCAAAACAGGGGCTAAGTCCA C
468	4381_Human_quadMut_1A_35T_45T_72A_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCCCAAGTTATTGGAGGACAAAACAGGGGCTAAGTCCA A
469	4382_Human_quadMut_1A_35G_46A_48G_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAGCCCAAGTTATCAGGGGAGCAAAACAGGGGCTAAGTCCA C
470	4383_Human_quadMut_1A_35A_46T_52A_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAAACCCAGTTATCTGAGGAACAAAACAGGGGCTAAGTCCA C
471	4384_Human_quadMut_1A_35T_46T_62C_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCCCAAGTTATCTGAGGAGCAAAACAGGGCCTAAGTCCA C
472	4387_Human_quadMut_1T_35T_48G_62T_n1	TGGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCCCAAGTTATCGGGGAGCAAAACAGGGTCTAAGTCCA C
473	4388_Human_quadMut_1A_35T_48G_72A_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCCCAAGTTATCGGGGAGCAAAACAGGGGCTAAGTCCA A
474	4394_Human_quadMut_1A_37A_45G_52T_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCAACAGTTATGGGAGGATCAAAACAGGGGCTAAGTCCA C
475	4395_Human_quadMut_1T_37T_45G_62A_n1	TGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCTCAGTTATGGGAGGAGCAAAACAGGGACTAAGTCCA C
476	4396_Human_quadMut_1A_37T_45G_72A_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCTCAGTTATGGGAGGAGCAAAACAGGGGCTAAGTCCA A
477	4398_Human_quadMut_1T_37A_46A_52C_n1	TGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCAACAGTTATCAGAGGACCAAAACAGGGGCTAAGTCCA C
478	4401_Human_quadMut_1A_37A_48G_52C_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCAACAGTTATCGGGGGACCAAAACAGGGGCTAAGTCCA C
479	4402_Human_quadMut_1T_37T_48G_62T_n1	TGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCTCAGTTATCGGGGGAGCAAAACAGGGTCTAAGTCCA C
480	4405_Human_quadMut_1A_37A_52A_72T_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCAACAGTTATCGGAGGAACAAAACAGGGGCTAAGTCCA T
481	4407_Human_quadMut_1T_45G_46T_48G_n1	TGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATGTGGGAGCAAAACAGGGGCTAAGTCCA C
482	4410_Human_quadMut_1T_45G_46A_72A_n1	TGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATGAGAGGAGCAAAACAGGGGCTAAGTCCA A
483	4416_Human_quadMut_1T_45T_62C_72A_n1	TGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATGTGGAGGAGCAAAACAGGGCCTAAGTCCA A
484	4417_Human_quadMut_1T_46A_48T_52C_n1	TGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCAGTGGACCAAAACAGGGGCTAAGTCCA C

485	4418_Human_quadMut_1T_46A_48T_62A_n1	TGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCAGTGGAGCAAA CAGGGACTAAGTCCA	C
486	4419_Human_quadMut_1T_46A_48T_72A_n1	TGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCAGTGGAGCAAA CAGGGGCTAAGTCCA	A
487	4423_Human_quadMut_1T_48G_52A_62T_n1	TGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGGGGGAACAAA CAGGGTCTAAGTCCA	C
488	4426_Human_quadMut_1T_52T_62C_72A_n1	TGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGATCAAA CAGGGCCTAAGTCCA	A
489	4435_Human_quadMut_3A_8A_35G_37T_n1	GGAGGACTGCTGGTGAATATTAA CCAAGGTCAGCTCAGTTATCGGAGGAGCAAA CAGGGGCTAAGTCCA	C
490	4439_Human_quadMut_3A_8T_35T_52T_n1	GGAGGACTGCTGGTGAATATTAA CCAAGGTCATCCC CAGTTATCGGAGGATCAAA CAGGGGCTAAGTCCA	C
491	4445_Human_quadMut_3A_8A_37A_52A_n1	GGAGGACTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGAA CAAA CAGGGGCTAAGTCCA	C
492	4448_Human_quadMut_3A_8A_45T_46A_n1	GGAGGACTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATTAGAGGAGCAAA CAGGGGCTAAGTCCA	C
493	4453_Human_quadMut_3T_8C_46T_48G_n1	GGTGGAGCCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCTGGGAGCAAA CAGGGGCTAAGTCCA	C
494	4455_Human_quadMut_3A_8C_46A_62T_n1	GGAGGACTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCAGAGGAGCAAA CAGGGTCTAAGTCCA	C
495	4456_Human_quadMut_3T_8T_46A_72A_n1	GGTGGAGTCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCAGAGGAGCAAA CAGGGGCTAAGTCCA	A
496	4457_Human_quadMut_3T_8A_48G_52A_n1	GGTGGAGACTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGGGGGAACAAA CAGGGGCTAAGTCCA	C
497	4460_Human_quadMut_3T_8A_52C_62T_n1	GGTGGAGACTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGAGCAAA CAGGGTCTAAGTCCA	C
498	4462_Human_quadMut_3T_8T_62T_72T_n1	GGTGGAGTCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGAGCAAA CAGGGTCTAAGTCCA	T
499	4467_Human_quadMut_3T_14T_35G_52T_n1	GGTGGAGGCTGCTTGTGAATATTAA CCAAGGTCAGCCC CAGTTATCGGAGGATCAAA CAGGGGCTAAGTCCA	C
500	4471_Human_quadMut_3T_14A_37A_46A_n1	GGTGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCA CAGTTATCAGAGGAGCAAA CAGGGGCTAAGTCCA	C
501	4481_Human_quadMut_3A_14T_46T_48T_n1	GGAGGAGGCTGCTTGTGAATATTAA CCAAGGTCACCC CAGTTATCTGTGGAGCAAA CAGGGGCTAAGTCCA	C
502	4482_Human_quadMut_3A_14T_46T_52C_n1	GGAGGAGGCTGCTTGTGAATATTAA CCAAGGTCACCC CAGTTATCTGAGGAGCAAA CAGGGGCTAAGTCCA	C

503	4487_Human_quadMut_3T_14T_48G_72A_n1	GGTGAGGCTGCTTGTGAATATTAA CCAAGGTCACCC CAGTTATCGGGGAGCAAAACAGGGGCTAAGTCCA A
504	4488_Human_quadMut_3A_14A_52C_62T_n1	GGAGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGACCAAAACAGGGTCTAAGTCCA C
505	4492_Human_quadMut_3A_35T_37T_46T_n1	GGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCTCAGTTATCTGAGGAGCAAAACAGGGGCTAAGTCCA C
506	4493_Human_quadMut_3T_35T_37T_48G_n1	GGTGAGGCTGCTGGTGAATATTAA CCAAGGTCATCTCAGTTATCGGGGAGCAAAACAGGGGCTAAGTCCA C
507	4494_Human_quadMut_3T_35A_37T_52C_n1	GGTGAGGCTGCTGGTGAATATTAA CCAAGGTCAACTCAGTTATCGGAGGACCAAAACAGGGGCTAAGTCCA C
508	4499_Human_quadMut_3T_35T_45T_52A_n1	GGTGAGGCTGCTGGTGAATATTAA CCAAGGTCATCCCAGTTATTGGAGGAA CAAACAGGGGCTAAGTCCA C
509	4502_Human_quadMut_3A_35G_46A_48G_n1	GGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCAGCCCAGTTATCAGGGGAGCAAAACAGGGGCTAAGTCCA C
510	4503_Human_quadMut_3T_35T_46T_52A_n1	GGTGAGGCTGCTGGTGAATATTAA CCAAGGTCATCCCAGTTATCTGAGGAA CAAACAGGGGCTAAGTCCA C
511	4506_Human_quadMut_3T_35G_48T_52T_n1	GGTGAGGCTGCTGGTGAATATTAA CCAAGGTCAGCCCAGTTATCGGTGGATCAAAACAGGGGCTAAGTCCA C
512	4507_Human_quadMut_3A_35A_48T_62T_n1	GGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCAACCCAGTTATCGGTGGAGCAAAACAGGGTCTAAGTCCA C
513	4509_Human_quadMut_3T_35T_52A_62T_n1	GGTGAGGCTGCTGGTGAATATTAA CCAAGGTCATCCCAGTTATCGGAGGAA CAAACAGGGTCTAAGTCCA C
514	4513_Human_quadMut_3A_37A_45T_48T_n1	GGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCA CAGTTATTTGGTGGAGCAAAACAGGGGCTAAGTCCA C
515	4519_Human_quadMut_3T_37T_46A_62A_n1	GGTGAGGCTGCTGGTGAATATTAA CCAAGGTCACCTCAGTTATCAGAGGAGCAAAACAGGGACTAAGTCCA C
516	4522_Human_quadMut_3A_37T_48T_62T_n1	GGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCTCAGTTATCGGTGGAGCAAAACAGGGTCTAAGTCCA C
517	4529_Human_quadMut_3T_45T_46T_62C_n1	GGTGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAGTTATTTGAGGAGCAAAACAGGGCCCTAAGTCCA C
518	4533_Human_quadMut_3A_45G_48G_72A_n1	GGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATGGGGGAGCAAAACAGGGGCTAAGTCCA A
519	4537_Human_quadMut_3A_46T_48T_52T_n1	GGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCTGTGGATCAAAACAGGGGCTAAGTCCA C
520	4539_Human_quadMut_3A_46T_48G_72G_n1	GGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCTGGGGAGCAAAACAGGGGCTAAGTCCA G

521	4541_Human_quadMut_3A_46T_52T_72A_n1	GGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCTGAGGATCAAA CAGGGGCTAAGTCCA A
522	4543_Human_quadMut_3T_48T_52A_62T_n1	GGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGGTGGAACAAA CAGGGTCTAAGTCCA C
523	4545_Human_quadMut_3A_48G_62T_72A_n1	GGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGGGGAGCAAA CAGGGTCTAAGTCCA A
524	4546_Human_quadMut_3T_52T_62T_72G_n1	GGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGATCAAA CAGGGTCTAAGTCCA G
525	4549_Human_quadMut_8C_14T_35T_46T_n1	GGGGAGCCTGCTTGTGAATATTAA CCAAGGTCACCC CAGTTATCTGAGGAGCAAA CAGGGGCTAAGTCCA C
526	4551_Human_quadMut_8T_14A_35T_52T_n1	GGGGAGTCTGCTAGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGATCAAA CAGGGGCTAAGTCCA C
527	4557_Human_quadMut_8A_14A_37T_52C_n1	GGGGAGACTGCTAGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGACCAAA CAGGGGCTAAGTCCA C
528	4562_Human_quadMut_8A_14T_45G_52T_n1	GGGGAGACTGCTTGTGAATATTAA CCAAGGTCACCC CAGTTATGGGAGGATCAAA CAGGGGCTAAGTCCA C
529	4563_Human_quadMut_8C_14A_45T_62T_n1	GGGGAGCCTGCTAGTGAATATTAA CCAAGGTCACCC CAGTTATTGGAGGACCAAA CAGGGTCTAAGTCCA C
530	4566_Human_quadMut_8T_14A_46A_52A_n1	GGGGAGTCTGCTAGTGAATATTAA CCAAGGTCACCC CAGTTATCAGAGGAACAAA CAGGGGCTAAGTCCA C
531	4569_Human_quadMut_8A_14A_48T_52A_n1	GGGGAGACTGCTAGTGAATATTAA CCAAGGTCACCC CAGTTATCGGTGGAACAAA CAGGGGCTAAGTCCA C
532	4576_Human_quadMut_8C_35A_37T_46T_n1	GGGGAGCCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCTGAGGAGCAAA CAGGGGCTAAGTCCA C
533	4579_Human_quadMut_8C_35G_37A_62A_n1	GGGGAGCCTGCTGGTGAATATTAA CCAAGGTCAGCA CAGTTATCGGAGGAGCAAA CAGGGACTAAGTCCA C
534	4580_Human_quadMut_8C_35A_37A_72A_n1	GGGGAGCCTGCTGGTGAATATTAA CCAAGGTCAA CACAGTTATCGGAGGAGCAAA CAGGGGCTAAGTCCA A
535	4587_Human_quadMut_8A_35G_46A_52T_n1	GGGGAGACTGCTGGTGAATATTAA CCAAGGTCAGCCC CAGTTATCAGAGGATCAAA CAGGGGCTAAGTCCA C
536	4588_Human_quadMut_8A_35T_46A_62C_n1	GGGGAGACTGCTGGTGAATATTAA CCAAGGTCATCCC CAGTTATCAGAGGAGCAAA CAGGGCCTAAGTCCA C
537	4590_Human_quadMut_8C_35G_48T_52A_n1	GGGGAGCCTGCTGGTGAATATTAA CCAAGGTCAGCCC CAGTTATCGGTGGAACAAA CAGGGGCTAAGTCCA C
538	4592_Human_quadMut_8C_35G_48T_72A_n1	GGGGAGCCTGCTGGTGAATATTAA CCAAGGTCAGCCC CAGTTATCGGTGAGCAAA CAGGGGCTAAGTCCA A

539	4599_Human_quadMut_8A_37T_45G_62T_n1	GGGGGAGACTGCTGGTGAATATTAA CCAAGGTCACCTCAGTTATGGAGGAGCAAAACAGGGTCTAAGTCCA	C
540	4600_Human_quadMut_8T_37T_45G_72T_n1	GGGGGAGTCTGCTGGTGAATATTAA CCAAGGTCACCTCAGTTATGGAGGAGCAAAACAGGGGCTAAGTCCA	T
541	4601_Human_quadMut_8T_37T_46T_48T_n1	GGGGGAGTCTGCTGGTGAATATTAA CCAAGGTCACCTCAGTTATCTGTGGAGCAAAACAGGGGCTAAGTCCA	C
542	4607_Human_quadMut_8T_37A_48G_72T_n1	GGGGGAGTCTGCTGGTGAATATTAA CCAAGGTCACACACAGTTATCGGGGAGCAAAACAGGGGCTAAGTCCA	T
543	4608_Human_quadMut_8C_37T_52C_62T_n1	GGGGGAGCCTGCTGGTGAATATTAA CCAAGGTCACCTCAGTTATCGGAGGACCAAAACAGGGTCTAAGTCCA	C
544	4609_Human_quadMut_8T_37T_52T_72G_n1	GGGGGAGTCTGCTGGTGAATATTAA CCAAGGTCACCTCAGTTATCGGAGGATCAAAACAGGGGCTAAGTCCA	G
545	4616_Human_quadMut_8A_45G_48T_62A_n1	GGGGGAGACTGCTGGTGAATATTAA CCAAGGTCACCCACAGTTATGGGTGGAGCAAAACAGGGACTAAGTCCA	C
546	4619_Human_quadMut_8T_45A_52T_72G_n1	GGGGGAGTCTGCTGGTGAATATTAA CCAAGGTCACCCACAGTTATATAGGAGGATCAAAACAGGGGCTAAGTCCA	G
547	4620_Human_quadMut_8A_45A_62T_72A_n1	GGGGGAGACTGCTGGTGAATATTAA CCAAGGTCACCCACAGTTATAGGAGGAGCAAAACAGGGTCTAAGTCCA	A
548	4622_Human_quadMut_8C_46A_48T_62A_n1	GGGGGAGCCTGCTGGTGAATATTAA CCAAGGTCACCCACAGTTATCAGTGGAGCAAAACAGGGACTAAGTCCA	C
549	4623_Human_quadMut_8A_46T_48T_72A_n1	GGGGGAGACTGCTGGTGAATATTAA CCAAGGTCACCCACAGTTATCTGTGGAGCAAAACAGGGGCTAAGTCCA	A
550	4624_Human_quadMut_8A_46A_52T_62C_n1	GGGGGAGACTGCTGGTGAATATTAA CCAAGGTCACCCACAGTTATCAGAGGATCAAAACAGGGCCTAAGTCCA	C
551	4626_Human_quadMut_8A_46A_62C_72T_n1	GGGGGAGACTGCTGGTGAATATTAA CCAAGGTCACCCACAGTTATCAGAGGAGCAAAACAGGGCCTAAGTCCA	T
552	4628_Human_quadMut_8C_48G_52A_72T_n1	GGGGGAGCCTGCTGGTGAATATTAA CCAAGGTCACCCACAGTTATCGGGGGAACAAACAGGGGCTAAGTCCA	T
553	4630_Human_quadMut_8C_52T_62T_72A_n1	GGGGGAGCCTGCTGGTGAATATTAA CCAAGGTCACCCACAGTTATCGGAGGATCAAAACAGGGTCTAAGTCCA	A
554	4632_Human_quadMut_14A_35G_37A_46A_n1	GGGGGAGGCTGCTAGTGAATATTAA CCAAGGTCAGCACAGTTATCAGAGGAGCAAAACAGGGGCTAAGTCCA	C
555	4633_Human_quadMut_14T_35A_37T_48G_n1	GGGGGAGGCTGCTTGTGAATATTAA CCAAGGTCAACTCAGTTATCGGGGAGCAAAACAGGGGCTAAGTCCA	C
556	4634_Human_quadMut_14A_35A_37A_52T_n1	GGGGGAGGCTGCTAGTGAATATTAA CCAAGGTCACACAGTTATCGGAGGATCAAAACAGGGGCTAAGTCCA	C

557	4635_Human_quadMut_14A_35G_37A_62A_n 1	GGGGAGGCTGCTAGTGAATATTAA CCAAGGTCAGCACAGTTATCGGAGGAGCAAAACAGGGACTAAGTCCA C
558	4638_Human_quadMut_14A_35G_45T_48T_n1	GGGGAGGCTGCTAGTGAATATTAA CCAAGGTCAGCCACAGTTATTGGTGGAGCAAAACAGGGGCTAAGTCCA C
559	4642_Human_quadMut_14T_35A_46A_48G_n 1	GGGGAGGCTGCTTGTGAATATTAA CCAAGGTCAACCCAGTTATCAGGGGAGCAAAACAGGGGCTAAGTCCA C
560	4643_Human_quadMut_14T_35T_46T_52A_n1	GGGGAGGCTGCTTGTGAATATTAA CCAAGGTCATCCCAGTTATCTGAGGAA CAAACAGGGGCTAAGTCCA C
561	4644_Human_quadMut_14A_35T_46A_62A_n1	GGGGAGGCTGCTAGTGAATATTAA CCAAGGTCATCCCAGTTATCAGAGGAGCAAAACAGGGACTAAGTCCA C
562	4647_Human_quadMut_14T_35T_48G_62C_n1	GGGGAGGCTGCTTGTGAATATTAA CCAAGGTCATCCCAGTTATCGGGGGAGCAAAACAGGGCCTAAGTCCA C
563	4649_Human_quadMut_14A_35A_52T_62T_n1	GGGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCCAGTTATCGGAGGATCAAAACAGGGTCTAAGTCCA C
564	4654_Human_quadMut_14T_37A_45T_52A_n1	GGGGAGGCTGCTTGTGAATATTAA CCAAGGTCACCACAGTTATTGGAGGA CAAACAGGGGCTAAGTCCA C
565	4659_Human_quadMut_14A_37A_46A_62T_n1	GGGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCACAGTTATCAGAGGAGCAAAACAGGGTCTAAGTCCA C
566	4660_Human_quadMut_14A_37A_46T_72A_n1	GGGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCACAGTTATCTGAGGAGCAAAACAGGGGCTAAGTCCA A
567	4661_Human_quadMut_14T_37T_48G_52T_n1	GGGGAGGCTGCTTGTGAATATTAA CCAAGGTCACCTCAGTTATCGGGGGATCAAAACAGGGGCTAAGTCCA C
568	4663_Human_quadMut_14T_37A_48G_72G_n 1	GGGGAGGCTGCTTGTGAATATTAA CCAAGGTCACCACAGTTATCGGGGGAGCAAAACAGGGGCTAAGTCCA G
569	4668_Human_quadMut_14T_45T_46T_52T_n1	GGGGAGGCTGCTTGTGAATATTAA CCAAGGTCACCCACAGTTATTGAGGATCAAAACAGGGGCTAAGTCCA C
570	4669_Human_quadMut_14A_45G_46T_62A_n 1	GGGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCCACAGTTATGTGAGGAGCAAAACAGGGACTAAGTCCA C
571	4676_Human_quadMut_14A_45T_62A_72A_n1	GGGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCCACAGTTATTGAGGAGCAAAACAGGGACTAAGTCCA A
572	4678_Human_quadMut_14A_46T_48G_62C_n1	GGGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCCACAGTTATCTGGGAGCAAAACAGGGCCTAAGTCCA C
573	4682_Human_quadMut_14A_46A_62C_72A_n 1	GGGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCCACAGTTATCAGAGGAGCAAAACAGGGCCTAAGTCCA A
574	4683_Human_quadMut_14A_48G_52T_62A_n 1	GGGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCCACAGTTATCGGGGGATCAAAACAGGGACTAAGTCCA C

575	4692_Human_quadMut_35G_37T_46A_48G_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAGCTCAGTTATCAGGGGAGCAAAACAGGGGCTAAGTCCA C
576	4693_Human_quadMut_35A_37A_46T_52C_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAACACACAGTTATCTGAGGACCAAAACAGGGGCTAAGTCCA C
577	4694_Human_quadMut_35T_37A_46T_62A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCACAGTTATCTGAGGACCAAAACAGGGACTAAGTCCA C
578	4695_Human_quadMut_35A_37A_46A_72A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACACAGTTATCAGAGGAGCAAAACAGGGGCTAAGTCCA A
579	4696_Human_quadMut_35G_37T_48G_52T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAGCTCAGTTATCGGGGGATCAAAACAGGGGCTAAGTCCA C
580	4697_Human_quadMut_35T_37A_48T_62T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCACAGTTATCGGTGGAGCAAAACAGGGTCTAAGTCCA C
581	4698_Human_quadMut_35A_37A_48G_72G_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACACAGTTATCGGGGGAGCAAAACAGGGGCTAAGTCCA G
582	4699_Human_quadMut_35A_37A_52T_62C_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACACAGTTATCGGAGGATCAAAACAGGGCCTAAGTCCA C
583	4700_Human_quadMut_35G_37T_52T_72A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAGCTCAGTTATCGGAGGATCAAAACAGGGGCTAAGTCCA A
584	4702_Human_quadMut_35T_45T_46A_48G_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCCAGTTATTAGGGGAGCAAAACAGGGGCTAAGTCCA C
585	4704_Human_quadMut_35G_45T_46T_62C_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAGCCCAGTTATTGAGGAGCAAAACAGGGCCTAAGTCCA C
586	4705_Human_quadMut_35T_45T_46T_72G_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCCCAGTTATTGAGGAGCAAAACAGGGGCTAAGTCCA G
587	4706_Human_quadMut_35A_45G_48T_52T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCAGTTATTTGAGGAGCAAAACAGGGGCTAAGTCCA C
588	4709_Human_quadMut_35A_45A_52T_62T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCAGTTATAGGAGGATCAAAACAGGGTCTAAGTCCA C
589	4712_Human_quadMut_35T_46A_48G_52A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCCCAGTTATCAGGGGAAACAAACAGGGGCTAAGTCCA C
590	4714_Human_quadMut_35A_46T_48T_72T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCAGTTATCTGTGGAGCAAAACAGGGGCTAAGTCCA T
591	4717_Human_quadMut_35T_46A_62C_72A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCCCAGTTATCAGAGGAGCAAAACAGGGCCTAAGTCCA A
592	4718_Human_quadMut_35T_48G_52C_62C_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCCCAGTTATCGGGGGACCAAAACAGGGCCTAAGTCCA C

593	4720_Human_quadMut_35G_48G_62C_72T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAGCCCAGTTATCGGGGAGCAAAACAGGGCCTAAGTCCA	T
594	4723_Human_quadMut_37T_45G_46A_52A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCTCAGTTATCAGAGGAACAAAACAGGGGCTAAGTCCA	C
595	4728_Human_quadMut_37T_45G_48T_72G_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCTCAGTTATCGGTGGAGCAAAACAGGGGCTAAGTCCA	G
596	4730_Human_quadMut_37T_45T_52A_72T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCTCAGTTATTGGAGGAACAAAACAGGGGCTAAGTCCA	T
597	4733_Human_quadMut_37T_46T_48T_62T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCTCAGTTATCTGTGGAGCAAAACAGGGTCTAAGTCCA	C
598	4734_Human_quadMut_37A_46T_48G_72G_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCACAGTTATCTGGGGAGCAAAACAGGGGCTAAGTCCA	G
599	4735_Human_quadMut_37A_46A_52T_62A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCACAGTTATCAGAGGATCAAAACAGGGACTAAGTCCA	C
600	4736_Human_quadMut_37T_46T_52A_72A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCTCAGTTATCTGAGGAACAAAACAGGGGCTAAGTCCA	A
601	4738_Human_quadMut_37T_48G_52A_62A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCTCAGTTATCGGGGGAACAAAACAGGGACTAAGTCCA	C
602	4739_Human_quadMut_37A_48G_52A_72A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCACAGTTATCGGGGGAACAAAACAGGGGCTAAGTCCA	A
603	4740_Human_quadMut_37T_48G_62C_72A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCTCAGTTATCGGGGAGCAAAACAGGGCCTAAGTCCA	A
604	4742_Human_quadMut_45T_46T_48G_52C_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCACAGTTATTGGGGACCAAAACAGGGGCTAAGTCCA	C
605	4751_Human_quadMut_45T_52T_62T_72G_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCACAGTTATTGGAGGATCAAAACAGGGTCTAAGTCCA	G
606	4755_Human_quadMut_46T_52T_62T_72G_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCACAGTTATCTGAGGATCAAAACAGGGTCTAAGTCCA	G
607	4761_Human_quadMut_1T_3T_8T_46A_n1	TGTGAGTCTGCTGGTGAATATTAA CCAAGGTCACCCACAGTTATCAGAGGAGCAAAACAGGGGCTAAGTCCA	C
608	4763_Human_quadMut_1A_3T_8T_52T_n1	AGTGGAGTCTGCTGGTGAATATTAA CCAAGGTCACCCACAGTTATCGGAGGATCAAAACAGGGGCTAAGTCCA	C
609	4766_Human_quadMut_1A_3A_14A_35G_n1	AGAGGAGGCTGCTAGTGAATATTAA CCAAGGTCAGCCACAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA	C
610	4770_Human_quadMut_1A_3A_14A_48T_n1	AGAGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCCACAGTTATCGGTGGAGCAAAACAGGGGCTAAGTCCA	C

611	4771_Human_quadMut_1A_3T_14A_52T_n1	AGTGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGATCAAAACAGGGGCTAAGTCCA C
612	4772_Human_quadMut_1T_3T_14A_62T_n1	TGTGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGAGCAAAACAGGGTCTAAGTCCA C
613	4776_Human_quadMut_1A_3A_35A_48G_n1	AGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGGGGGAGCAAAACAGGGGCTAAGTCCA C
614	4777_Human_quadMut_1A_3A_35T_52C_n1	AGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCC CAGTTATCGGAGGACCAAAACAGGGGCTAAGTCCA C
615	4778_Human_quadMut_1A_3T_35A_62T_n1	AGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGAGCAAAACAGGGTCTAAGTCCA C
616	4779_Human_quadMut_1A_3A_35G_72A_n1	AGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCAGCCC CAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA A
617	4780_Human_quadMut_1T_3A_37T_45T_n1	TGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCTC AGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
618	4783_Human_quadMut_1T_3A_37T_52A_n1	TGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCTC AGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
619	4784_Human_quadMut_1T_3A_37A_62T_n1	TGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCA CAGTTATCGGAGGAGCAAAACAGGGTCTAAGTCCA C
620	4788_Human_quadMut_1T_3A_45A_52C_n1	TGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCC CAGTTATAGGAGGACCAAAACAGGGGCTAAGTCCA C
621	4790_Human_quadMut_1A_3T_45G_72A_n1	AGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCC CAGTTATGGGAGGAGCAAAACAGGGGCTAAGTCCA A
622	4791_Human_quadMut_1T_3T_46A_48T_n1	TGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCC CAGTTATCAGTGGAGCAAAACAGGGGCTAAGTCCA C
623	4792_Human_quadMut_1A_3A_46T_52A_n1	AGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCC CAGTTATCTGAGGAGCAAAACAGGGGCTAAGTCCA C
624	4796_Human_quadMut_1T_3A_48T_62T_n1	TGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCC CAGTTATCGGTGGAGCAAAACAGGGTCTAAGTCCA C
625	4798_Human_quadMut_1T_3A_52A_62C_n1	TGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCC CAGTTATCGGAGGAGCAAAACAGGGCCCTAAGTCCA C
626	4805_Human_quadMut_1T_8A_14T_48T_n1	TGGGGAGACTGCTTGTGAATATTAA CCAAGGTCACCCC CAGTTATCGGTGGAGCAAAACAGGGGCTAAGTCCA C
627	4806_Human_quadMut_1A_8A_14A_52T_n1	AGGGGAGACTGCTAGTGAATATTAA CCAAGGTCACCCC CAGTTATCGGAGGATCAAAACAGGGGCTAAGTCCA C
728	4809_Human_quadMut_1A_8C_35A_37A_n1	AGGGGAGCCTGCTGGTGAATATTAA CCAAGGTCACCA CAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C

629	4810_Human_quadMut_1A_8C_35A_45A_n1	AGGGGAGCCTGCTGGTGAATATTAA CCAAGGTCAACCCAGTTATAGGAGGAGCAAAACAGGGGCTAAGTCCA C
630	4811_Human_quadMut_1A_8T_35G_46A_n1	AGGGGAGTCTGCTGGTGAATATTAA CCAAGGTCAAGCCAGTTATCAGAGGAGCAAAACAGGGGCTAAGTCCA C
631	4812_Human_quadMut_1T_8A_35T_48T_n1	TGGGAGACTGCTGGTGAATATTAA CCAAGGTCAATCCCAGTTATCGGTGGAGCAAAACAGGGGCTAAGTCCA C
632	4819_Human_quadMut_1A_8T_37T_52T_n1	AGGGGAGTCTGCTGGTGAATATTAA CCAAGGTCACTCAGTTATCGGAGGATCAAAACAGGGGCTAAGTCCA C
633	4820_Human_quadMut_1A_8A_37T_72G_n1	AGGGGAGACTGCTGGTGAATATTAA CCAAGGTCACTCAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA G
634	4822_Human_quadMut_1T_8A_45A_48G_n1	TGGGAGACTGCTGGTGAATATTAA CCAAGGTCAACCCAGTTATAGGGGAGCAAAACAGGGGCTAAGTCCA C
635	4826_Human_quadMut_1T_8T_46T_48T_n1	TGGGAGTCTGCTGGTGAATATTAA CCAAGGTCAACCCAGTTATCTGTGGAGCAAAACAGGGGCTAAGTCCA C
636	4833_Human_quadMut_1T_8C_52T_72G_n1	TGGGAGCCTGCTGGTGAATATTAA CCAAGGTCAACCCAGTTATCGGAGGATCAAAACAGGGGCTAAGTCCA G
637	4839_Human_quadMut_1A_14T_35T_52T_n1	AGGGAGGCTGCTTGTGAATATTAA CCAAGGTCAATCCAGTTATCGGAGGATCAAAACAGGGGCTAAGTCCA C
638	4853_Human_quadMut_1A_14A_46A_48G_n1	AGGGAGGCTGCTAGTGAATATTAA CCAAGGTCAACCCAGTTATCAGGGGAGCAAAACAGGGGCTAAGTCCA C
639	4864_Human_quadMut_1A_35A_37A_46T_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAACACACAGTTATCTGAGGAGCAAAACAGGGGCTAAGTCCA C
640	4867_Human_quadMut_1A_35A_37T_62C_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAACTCAGTTATCGGAGGAGCAAAACAGGGCCTAAGTCCA C
641	4872_Human_quadMut_1T_35A_45T_62T_n1	TGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAACCCAGTTATTTGGAGGAGCAAAACAGGGTCTAAGTCCA C
642	4876_Human_quadMut_1A_35A_46A_62C_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAACCCAGTTATCAGAGGAGCAAAACAGGGCCTAAGTCCA C
643	4879_Human_quadMut_1A_35A_48T_62T_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAACCCAGTTATCGGTGGAGCAAAACAGGGTCTAAGTCCA C
644	4880_Human_quadMut_1A_35G_48G_72T_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAAGCCAGTTATCGGGGAGCAAAACAGGGGCTAAGTCCA T
645	4884_Human_quadMut_1A_37T_45T_46T_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACTCAGTTATTTGAGGAGCAAAACAGGGGCTAAGTCCA C
646	4887_Human_quadMut_1A_37A_45T_62C_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAACACAGTTATTTGAGGAGCAAAACAGGGCCTAAGTCCA C

647	4890_Human_quadMut_1T_37A_46T_52T_n1	TGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCA CAGTTATCTGAGGATCAAAACAGGGGCTAAGTCCA	C
648	4891_Human_quadMut_1A_37A_46T_62A_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCA CAGTTATCTGAGGAGCAAAACAGGGACTAAGTCCA	C
649	4892_Human_quadMut_1T_37T_46A_72A_n1	TGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCT CAGTTATCAGAGGAGCAAAACAGGGGCTAAGTCCA	A
650	4894_Human_quadMut_1A_37A_48G_62A_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCA CAGTTATCGGGGAGCAAAACAGGGACTAAGTCCA	C
651	4896_Human_quadMut_1T_37A_52T_62A_n1	TGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCA CAGTTATCGGAGGATCAAAACAGGGACTAAGTCCA	C
652	4898_Human_quadMut_1A_37T_62T_72T_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCT CAGTTATCGGAGGAGCAAAACAGGGTCTAAGTCCA	T
653	4907_Human_quadMut_1A_45G_52A_72A_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGAA CAAACAGGGGCTAAGTCCA	A
654	4911_Human_quadMut_1A_46A_52A_62T_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCAGAGGAA CAAACAGGGTCTAAGTCCA	C
655	4914_Human_quadMut_1A_48T_52T_62C_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGGTGGATCAAAACAGGGCCTAAGTCCA	A
656	4923_Human_quadMut_3A_8A_14A_52A_n1	GGAGGACTGCTAGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGAA CAAACAGGGGCTAAGTCCA	C
657	4926_Human_quadMut_3A_8C_35A_37A_n1	GGAGGACTGCTGGTGAATATTAA CCAAGGTCACCA CAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA	C
658	4932_Human_quadMut_3A_8C_35A_72A_n1	GGAGGACTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA	A
659	4934_Human_quadMut_3A_8C_37T_46T_n1	GGAGGACTGCTGGTGAATATTAA CCAAGGTCACCT CAGTTATCTGAGGAGCAAAACAGGGGCTAAGTCCA	C
660	4935_Human_quadMut_3T_8C_37A_48T_n1	GGAGGACTGCTGGTGAATATTAA CCAAGGTCACCA CAGTTATCGGTGGAGCAAAACAGGGGCTAAGTCCA	C
661	4943_Human_quadMut_3A_8A_46T_48T_n1	GGAGGACTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCTGTGGAGCAAAACAGGGGCTAAGTCCA	C
662	4944_Human_quadMut_3A_8A_46A_52T_n1	GGAGGACTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCAGAGGATCAAAACAGGGGCTAAGTCCA	C
663	4947_Human_quadMut_3T_8T_48T_52T_n1	GGTGGAGTCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGGTGGATCAAAACAGGGGCTAAGTCCA	C
664	4951_Human_quadMut_3T_8A_52A_72G_n1	GGTGGAGTCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGAA CAAACAGGGGCTAAGTCCA	G

665	4952_Human_quadMut_3T_8C_62T_72A_n1	GGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGAGCAAAACAGGGTCTAAGTCCA A
666	4955_Human_quadMut_3T_14T_35T_46A_n1	GGTGGAGGCTGCTTGTGAATATTAA CCAAGGTCATCC CAGTTATCAGAGGAGCAAAACAGGGGCTAAGTCCA C
667	4956_Human_quadMut_3A_14A_35G_48G_n1	GGAGGAGGCTGCTAGTGAATATTAA CCAAGGTCAGCCC CAGTTATCGGGGGAGCAAAACAGGGGCTAAGTCCA C
668	4957_Human_quadMut_3T_14T_35T_52T_n1	GGTGGAGGCTGCTTGTGAATATTAA CCAAGGTCATCC CAGTTATCGGAGGATCAAAACAGGGGCTAAGTCCA C
669	4975_Human_quadMut_3A_14A_48G_62T_n1	GGAGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCCC CAGTTATCGGGGGAGCAAAACAGGGTCTAAGTCCA C
670	4976_Human_quadMut_3A_14A_48T_72A_n1	GGAGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCCC CAGTTATCGGTGGAGCAAAACAGGGGCTAAGTCCA A
671	4977_Human_quadMut_3T_14A_52A_62T_n1	GGTGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCCC CAGTTATCGGAGGAA CAAACAGGGTCTAAGTCCA C
672	4980_Human_quadMut_3T_35A_37T_46T_n1	GGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCAACT CAGTTATCTGAGGAGCAAAACAGGGGCTAAGTCCA C
673	4986_Human_quadMut_3T_35A_45A_48T_n1	GGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCAA CCCCAGTTATAGGTGGAGCAAAACAGGGGCTAAGTCCA C
674	4988_Human_quadMut_3T_35G_45T_62T_n1	GGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCAGCCC CAGTTATTTGGAGGAGCAAAACAGGGTCTAAGTCCA C
675	4991_Human_quadMut_3A_35A_46A_52C_n1	GGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCAA CCCCAGTTATCAGAGGACCAAAACAGGGGCTAAGTCCA C
676	4992_Human_quadMut_3A_35T_46A_62A_n1	GGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCC CAGTTATCAGAGGAGCAAAACAGGGACTAAGTCCA C
677	4993_Human_quadMut_3T_35T_46T_72T_n1	GGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCC CAGTTATCTGAGGAGCAAAACAGGGGCTAAGTCCA T
678	4995_Human_quadMut_3T_35A_48T_62A_n1	GGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCAA CCCCAGTTATCGGTGGAGCAAAACAGGGACTAAGTCCA C
679	4997_Human_quadMut_3A_35A_52T_62T_n1	GGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCAA CCCCAGTTATCGGAGGATCAAAACAGGGTCTAAGTCCA C
680	5000_Human_quadMut_3T_37T_45A_46A_n1	GGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCT CAGTTATAAGAGGAGCAAAACAGGGGCTAAGTCCA C
681	5003_Human_quadMut_3T_37A_45T_62T_n1	GGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCAC CACACAGTTATTTGGAGGAGCAAAACAGGGTCTAAGTCCA C
682	5004_Human_quadMut_3A_37T_46A_48G_n1	GGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCT CAGTTATCAGGGGAGCAAAACAGGGGCTAAGTCCA C

683	5005_Human_quadMut_3A_37A_46T_52T_n1	GGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCA CAGTTATCTGAGGATCAAAACAGGGGCTAAGTCCA	C
684	5006_Human_quadMut_3A_37A_46A_62T_n1	GGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCA CAGTTATCAGAGGAGCAAAACAGGGTCTAAGTCCA	C
685	5008_Human_quadMut_3A_37A_48G_52T_n1	GGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCA CAGTTATCGGGGGATCAAAACAGGGGCTAAGTCCA	C
686	5010_Human_quadMut_3T_37A_48G_72A_n1	GGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCA CAGTTATCGGGGGAGCAAAACAGGGGCTAAGTCCA	A
687	5019_Human_quadMut_3T_45G_48G_62T_n1	GGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGGGGGAGCAAAACAGGGTCTAAGTCCA	C
688	5025_Human_quadMut_3A_46A_48T_62C_n1	GGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCAGTGGAGCAAAACAGGGCCTAAGTCCA	C
689	5027_Human_quadMut_3T_46A_52A_62T_n1	GGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCAGAGGAA CAAACAGGGTCTAAGTCCA	C
690	5030_Human_quadMut_3T_48G_52T_62T_n1	GGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGGGGGATCAAAACAGGGTCTAAGTCCA	C
691	5032_Human_quadMut_3A_48T_62C_72A_n1	GGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGTGGAGCAAAACAGGGCCTAAGTCCA	A
692	5033_Human_quadMut_3T_52T_62C_72G_n1	GGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGATCAAAACAGGGCCTAAGTCCA	G
693	5036_Human_quadMut_8T_14A_35T_46T_n1	GGGGAGTCTGCTAGTGAATATTAA CCAAGGTCATCC CAGTTATCTGAGGAGCAAAACAGGGGCTAAGTCCA	C
694	5037_Human_quadMut_8C_14T_35T_48G_n1	GGGGAGCCTGCTTGTGAATATTAA CCAAGGTCATCC CAGTTATCGGGGGAGCAAAACAGGGGCTAAGTCCA	C
695	5040_Human_quadMut_8T_14A_35G_72G_n1	GGGGAGTCTGCTAGTGAATATTAA CCAAGGTCAGCC CAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA	G
696	5041_Human_quadMut_8T_14A_37T_45T_n1	GGGGAGTCTGCTAGTGAATATTAA CCAAGGTCACCT CAGTTATTTGAGGAGCAAAACAGGGGCTAAGTCCA	C
697	5042_Human_quadMut_8C_14A_37T_46A_n1	GGGGAGCCTGCTAGTGAATATTAA CCAAGGTCACCT CAGTTATCAGAGGAGCAAAACAGGGGCTAAGTCCA	C
698	5043_Human_quadMut_8A_14A_37A_48G_n1	GGGGAGACTGCTAGTGAATATTAA CCAAGGTCACCA CAGTTATCGGGGGAGCAAAACAGGGGCTAAGTCCA	C
699	5051_Human_quadMut_8C_14T_46A_52T_n1	GGGGAGCCTGCTTGTGAATATTAA CCAAGGTCACCC CAGTTATCAGAGGATCAAAACAGGGGCTAAGTCCA	C
700	5052_Human_quadMut_8A_14T_46A_62C_n1	GGGGAGACTGCTTGTGAATATTAA CCAAGGTCACCC CAGTTATCAGAGGAGCAAAACAGGGCCTAAGTCCA	C

701	5056_Human_quadMut_8C_14A_48G_72A_n1	GGGGAGCCTGCTAGTGAATATTAA CCAAGGTCACCC CAGTTATCGGGGAGCAAA CAGGGGCTAAGTCCA	A
702	5057_Human_quadMut_8T_14A_52A_62C_n1	GGGGAGTCTGCTAGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGAA CAAA CAGGGCCTAAGTCCA	C
703	5063_Human_quadMut_8A_35A_37A_52T_n1	GGGGAGACTGCTGGTGAATATTAA CCAAGGTCACAC CAGTTATCGGAGGAT CAAA CAGGGGCTAAGTCCA	C
704	5071_Human_quadMut_8T_35A_46T_48T_n1	GGGGAGTCTGCTGGTGAATATTAA CCAAGGTCAA CCCCAGTTATCTGTGGAGCAAA CAGGGGCTAAGTCCA	C
705	5072_Human_quadMut_8A_35G_46T_52T_n1	GGGGAGACTGCTGGTGAATATTAA CCAAGGTCAGCCC CAGTTATCTGAGGAT CAAA CAGGGGCTAAGTCCA	C
706	5073_Human_quadMut_8T_35G_46T_62T_n1	GGGGAGTCTGCTGGTGAATATTAA CCAAGGTCAGCCC CAGTTATCTGAGGAGCAAA CAGGGTCTAAGTCCA	C
707	5074_Human_quadMut_8T_35G_46T_72A_n1	GGGGAGTCTGCTGGTGAATATTAA CCAAGGTCAGCCC CAGTTATCTGAGGAGCAAA CAGGGGCTAAGTCCA	A
708	5075_Human_quadMut_8T_35T_48G_52T_n1	GGGGAGTCTGCTGGTGAATATTAA CCAAGGTCATCCC CAGTTATCGGGGAT CAAA CAGGGGCTAAGTCCA	C
709	5080_Human_quadMut_8A_35T_62T_72G_n1	GGGGAGACTGCTGGTGAATATTAA CCAAGGTCATCCC CAGTTATCGGAGGAGCAAA CAGGGTCTAAGTCCA	G
710	5081_Human_quadMut_8A_37A_45A_46T_n1	GGGGAGACTGCTGGTGAATATTAA CCAAGGTCACC CAGTTATATGAGGAGCAAA CAGGGGCTAAGTCCA	C
711	5083_Human_quadMut_8T_37T_45T_52T_n1	GGGGAGTCTGCTGGTGAATATTAA CCAAGGTCACCT CAGTTATTGGAGGAT CAAA CAGGGGCTAAGTCCA	C
712	5085_Human_quadMut_8T_37T_45T_72T_n1	GGGGAGTCTGCTGGTGAATATTAA CCAAGGTCACCT CAGTTATTGGAGGAGCAAA CAGGGGCTAAGTCCA	T
713	5086_Human_quadMut_8A_37T_46T_48G_n1	GGGGAGACTGCTGGTGAATATTAA CCAAGGTCACCT CAGTTATCTGGGAGCAAA CAGGGGCTAAGTCCA	C
714	5087_Human_quadMut_8A_37A_46A_52A_n1	GGGGAGACTGCTGGTGAATATTAA CCAAGGTCACC CAGTTATCAGAGGAA CAAA CAGGGGCTAAGTCCA	C
715	5088_Human_quadMut_8A_37A_46A_62C_n1	GGGGAGACTGCTGGTGAATATTAA CCAAGGTCACC CAGTTATCAGAGGAGCAAA CAGGGCCTAAGTCCA	C
716	5089_Human_quadMut_8T_37A_46T_72A_n1	GGGGAGTCTGCTGGTGAATATTAA CCAAGGTCACC CAGTTATCTGAGGAGCAAA CAGGGGCTAAGTCCA	A
717	5090_Human_quadMut_8A_37A_48G_52T_n1	GGGGAGACTGCTGGTGAATATTAA CCAAGGTCACC CAGTTATCGGGGAT CAAA CAGGGGCTAAGTCCA	C
718	5092_Human_quadMut_8T_37T_48T_72G_n1	GGGGAGTCTGCTGGTGAATATTAA CCAAGGTCACCT CAGTTATCGGTGGAGCAAA CAGGGGCTAAGTCCA	G

719	5093_Human_quadMut_8C_37A_52T_62T_n1	GGGGAGCCTGCTGGTGAATATTAA CCAAGGTCACCA CAGTTATCGGAGGATCAAAACAGGGTCTAAGTCCA C
720	5100_Human_quadMut_8T_45G_48G_52A_n1	GGGGAGTCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATGGGGGAA CAAAACAGGGGCTAAGTCCA C
721	5104_Human_quadMut_8T_45G_52T_72G_n1	GGGGAGTCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATGGGAGGATCAAAACAGGGGCTAAGTCCA G
722	5107_Human_quadMut_8A_46A_48T_62T_n1	GGGGAGACTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCAGTGGAGCAAAACAGGGTCTAAGTCCA C
723	5109_Human_quadMut_8A_46T_52T_62T_n1	GGGGAGACTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCTGAGGATCAAAACAGGGTCTAAGTCCA C
724	5110_Human_quadMut_8C_46T_52C_72A_n1	GGGGAGCCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCTGAGGACCAAAACAGGGGCTAAGTCCA A
725	5115_Human_quadMut_8T_52A_62A_72T_n1	GGGGAGTCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGAA CAAAACAGGGACTAAGTCCA T
726	5119_Human_quadMut_14A_35G_37A_52A_n1	GGGGAGGCTGCTAGTGAATATTAA CCAAGGTCAGCA CAGTTATCGGAGGAA CAAAACAGGGGCTAAGTCCA C
727	5128_Human_quadMut_14T_35T_46A_52A_n1	GGGGAGGCTGCTTGTGAATATTAA CCAAGGTCATCC CAGTTATCAGAGGAA CAAAACAGGGGCTAAGTCCA C
728	5132_Human_quadMut_14A_35G_48T_62T_n1	GGGGAGGCTGCTAGTGAATATTAA CCAAGGTCAGCC CAGTTATCGGTGGAGCAAAACAGGGTCTAAGTCCA C
729	5133_Human_quadMut_14A_35T_48T_72G_n1	GGGGAGGCTGCTAGTGAATATTAA CCAAGGTCATCC CAGTTATCGGTGGAGCAAAACAGGGGCTAAGTCCA G
730	5134_Human_quadMut_14T_35G_52T_62C_n1	GGGGAGGCTGCTTGTGAATATTAA CCAAGGTCAGCC CAGTTATCGGAGGATCAAAACAGGGCCTAAGTCCA C
731	5135_Human_quadMut_14T_35A_52T_72G_n1	GGGGAGGCTGCTTGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGATCAAAACAGGGGCTAAGTCCA G
732	5140_Human_quadMut_14A_37T_45G_62T_n1	GGGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCT CAGTTATGGGAGGAGCAAAACAGGGTCTAAGTCCA C
733	5142_Human_quadMut_14A_37A_46A_48T_n1	GGGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCA CAGTTATCAGTGGAGCAAAACAGGGGCTAAGTCCA C
734	5145_Human_quadMut_14A_37T_46T_72A_n1	GGGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCT CAGTTATCTGAGGAGCAAAACAGGGGCTAAGTCCA A
735	5146_Human_quadMut_14A_37T_48G_52A_n1	GGGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCT CAGTTATCGGGGAA CAAAACAGGGGCTAAGTCCA C
736	5147_Human_quadMut_14A_37T_48T_62C_n1	GGGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCT CAGTTATCGGTGGAGCAAAACAGGGCCTAAGTCCA C

737	5148_Human_quadMut_14A_37T_48G_72A_n 1	GGGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCTCAGTTATCGGGGAGCAAAACAGGGGCTAAGTCCA A
738	5155_Human_quadMut_14T_45T_46T_72G_n1	GGGGAGGCTGCTTGTGAATATTAA CCAAGGTCACCCAGTTATTGAGGACAAAACAGGGGCTAAGTCCA G
739	5157_Human_quadMut_14T_45G_48G_62T_n 1	GGGGAGGCTGCTTGTGAATATTAA CCAAGGTCACCCAGTTATCGGGGAGCAAAACAGGGTCTAAGTCCA C
740	5164_Human_quadMut_14T_46A_52T_62T_n1	GGGGAGGCTGCTTGTGAATATTAA CCAAGGTCACCCAGTTATCAGAGGATCAAAACAGGGTCTAAGTCCA C
741	5165_Human_quadMut_14T_46A_52T_72G_n1	GGGGAGGCTGCTTGTGAATATTAA CCAAGGTCACCCAGTTATCAGAGGATCAAAACAGGGGCTAAGTCCA G
742	5167_Human_quadMut_14A_48G_52A_62C_n 1	GGGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCCAGTTATCGGGGAA CAAAACAGGGCCTAAGTCCA C
743	5171_Human_quadMut_35G_37A_45T_46T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAGCACAGTTATTGAGGACAAAACAGGGGCTAAGTCCA C
744	5175_Human_quadMut_35T_37A_46A_48G_n 1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCACAGTTATCAGGGGAGCAAAACAGGGGCTAAGTCCA C
745	5177_Human_quadMut_35G_37A_46A_72A_n 1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAGCACAGTTATCAGAGGACAAAACAGGGGCTAAGTCCA A
746	5178_Human_quadMut_35G_37A_48G_52A_n 1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAGCACAGTTATCGGGGAA CAAAACAGGGGCTAAGTCCA C
747	5179_Human_quadMut_35T_37T_48T_62A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCTCAGTTATCGGTGGAGCAAAACAGGGACTAAGTCCA C
748	5181_Human_quadMut_35T_37A_52A_62A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCACAGTTATCGGAGGAA CAAAACAGGGACTAAGTCCA C
749	5185_Human_quadMut_35T_45A_46T_52A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCCAGTTATATGAGGAA CAAAACAGGGGCTAAGTCCA C
750	5186_Human_quadMut_35T_45T_46T_62T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCCCAGTTATTGAGGACAAAACAGGGTCTAAGTCCA C
751	5188_Human_quadMut_35A_45A_48T_52A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCAGTTATAGGTGGAACAAAACAGGGGCTAAGTCCA C
752	5194_Human_quadMut_35G_46T_48G_52C_n 1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAGCCAGTTATCTGGGGACAAAACAGGGGCTAAGTCCA C
753	5195_Human_quadMut_35T_46A_48G_62A_n 1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCCCAGTTATCAGGGGAGCAAAACAGGGACTAAGTCCA C
754	5196_Human_quadMut_35A_46T_48G_72A_n 1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCAGTTATCTGGGAGCAAAACAGGGGCTAAGTCCA A

755	5197_Human_quadMut_35A_46T_52C_62T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAACCCAGTTATCTGAGGACCAAAACAGGGTCTAAGTCCA C
756	5200_Human_quadMut_35A_48G_52A_62A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAACCCAGTTATCGGGGGAACAAAACAGGGACTAAGTCCA C
757	5201_Human_quadMut_35A_48T_52T_72A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAACCCAGTTATCGGTGGATCAAAACAGGGGCTAAGTCCA A
758	5202_Human_quadMut_35T_48G_62C_72T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAATCCAGTTATCGGGGGAGCAAAACAGGGCCTAAGTCCA T
759	5203_Human_quadMut_35T_52T_62T_72T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAATCCAGTTATCGGAGGATCAAAACAGGGTCTAAGTCCA T
760	5204_Human_quadMut_37A_45G_46A_48G_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAACACAGTTATGAGGGGAGCAAAACAGGGGCTAAGTCCA C
761	5205_Human_quadMut_37A_45A_46A_62T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAACACAGTTATAAGAGGAGCAAAACAGGGTCTAAGTCCA C
762	5208_Human_quadMut_37T_45T_48G_62T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACTCAGTTATCGGGGAGCAAAACAGGGTCTAAGTCCA C
763	5210_Human_quadMut_37A_45G_52A_62C_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAACACAGTTATGGGAGGAACAAAACAGGGCCTAAGTCCA C
764	5213_Human_quadMut_37A_46A_48G_52T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAACACAGTTATCAGGGGATCAAAACAGGGGCTAAGTCCA C
765	5214_Human_quadMut_37A_46A_48G_62A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAACACAGTTATCAGGGGAGCAAAACAGGGACTAAGTCCA C
766	5216_Human_quadMut_37T_46A_52A_62T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACTCAGTTATCAGAGGGAACAAAACAGGGTCTAAGTCCA C
767	5217_Human_quadMut_37T_46A_52T_72A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACTCAGTTATCAGAGGATCAAAACAGGGGCTAAGTCCA A
768	5219_Human_quadMut_37A_48G_52T_62C_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAACACAGTTATCGGGGGATCAAAACAGGGCCTAAGTCCA C
769	5222_Human_quadMut_37T_52A_62T_72A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACTCAGTTATCGGAGGAACAAAACAGGGTCTAAGTCCA A
770	5223_Human_quadMut_45G_46A_48G_52A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACTCAGTTATGAGGGGAACAAAACAGGGGCTAAGTCCA C
771	5224_Human_quadMut_45T_46T_48T_62C_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACTCAGTTATTTGTGAGCAAAACAGGGCCTAAGTCCA C
772	5228_Human_quadMut_45G_46A_62T_72A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACTCAGTTATGAGAGGAGCAAAACAGGGTCTAAGTCCA A

773	5230_Human_quadMut_45A_48G_52T_72A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATATAGGGGATCAAAACAGGGGCTAAGTCCA A
774	5232_Human_quadMut_45G_52T_62C_72T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATGGAGGATCAAAACAGGGCCTAAGTCCA T
775	5233_Human_quadMut_46A_48G_52C_62T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCAGGGACCAAAACAGGGTCTAAGTCCA C
776	5235_Human_quadMut_46T_48G_62A_72T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCTGGGGAGCAAAACAGGGACTAAGTCCA T
777	5240_Human_quadMut_1T_3A_8A_46T_n1	TGAGGAGACTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCTGAGGAGCAAAACAGGGGCTAAGTCCA C
778	5241_Human_quadMut_1A_3A_8T_48T_n1	AGAGGAGTCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGTGGAGCAAAACAGGGGCTAAGTCCA C
779	5242_Human_quadMut_1A_3A_8A_52T_n1	AGAGGAGACTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGATCAAAACAGGGGCTAAGTCCA C
780	5243_Human_quadMut_1A_3A_8C_62C_n1	AGAGGAGCCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGCCTAAGTCCA C
781	5245_Human_quadMut_1A_3T_14A_35G_n1	AGTGGAGGCTGCTAGTGAATATTAA CCAAGGTCAGCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
782	5248_Human_quadMut_1T_3A_14A_46T_n1	TGAGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCCCAAGTTATCTGAGGAGCAAAACAGGGGCTAAGTCCA C
783	5252_Human_quadMut_1A_3T_14A_72A_n1	AGTGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA A
784	5260_Human_quadMut_1A_3A_37A_46A_n1	AGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCAAGTTATCAGAGGAGCAAAACAGGGGCTAAGTCCA C
785	5262_Human_quadMut_1A_3T_37A_62C_n1	AGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCAAGTTATCTGGAGGAGCAAAACAGGGCCTAAGTCCA C
786	5263_Human_quadMut_1A_3T_37T_72A_n1	AGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCTCAGTTATCTGGAGGAGCAAAACAGGGGCTAAGTCCA A
787	5269_Human_quadMut_1T_3A_46A_52A_n1	TGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCAGAGGAACAAACAGGGGCTAAGTCCA C
788	5271_Human_quadMut_1A_3T_46T_72A_n1	AGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCTGAGGAGCAAAACAGGGGCTAAGTCCA A
789	5274_Human_quadMut_1A_3A_48G_72A_n1	AGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGGGGAGCAAAACAGGGGCTAAGTCCA A
790	5275_Human_quadMut_1T_3A_52T_62C_n1	TGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGATCAAAACAGGGCCTAAGTCCA C

791	5283_Human_quadMut_1A_8C_14A_52A_n1	AGGGGAGCCTGCTAGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGAACAAA CAGGGGCTAAGTCCA C
792	5286_Human_quadMut_1A_8A_35G_37A_n1	AGGGGAGACTGCTGGTGAATATTAA CCAAGGTCAGCACAGTTATCGGAGGAGCAAAA CAGGGGCTAAGTCCA C
793	5287_Human_quadMut_1T_8T_35A_45G_n1	TGGGAGTCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATGGGAGGAGCAAAA CAGGGGCTAAGTCCA C
794	5289_Human_quadMut_1A_8C_35A_48T_n1	AGGGGAGCCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGGTGGAGCAAAA CAGGGGCTAAGTCCA C
795	5296_Human_quadMut_1A_8C_37T_62C_n1	AGGGGAGCCTGCTGGTGAATATTAA CCAAGGTCACCT CAGTTATCGGAGGAGCAAAA CAGGGCCTAAGTCCA C
796	5304_Human_quadMut_1A_8A_46A_52T_n1	AGGGGAGACTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCAGAGGATCAAAA CAGGGGCTAAGTCCA C
797	5313_Human_quadMut_1T_14T_35T_37T_n1	TGGGAGGGCTGCTTGTGAATATTAA CCAAGGTCATCT CAGTTATCGGAGGAGCAAAA CAGGGGCTAAGTCCA C
798	5322_Human_quadMut_1A_14A_37A_52T_n1	AGGGGAGGGCTGCTAGTGAATATTAA CCAAGGTCACCA CAGTTATCGGAGGATCAAAA CAGGGGCTAAGTCCA C
799	5324_Human_quadMut_1A_14A_37T_72T_n1	AGGGGAGGGCTGCTAGTGAATATTAA CCAAGGTCACCT CAGTTATCGGAGGAGCAAAA CAGGGGCTAAGTCCA T
800	5331_Human_quadMut_1T_14T_46T_52T_n1	TGGGAGGGCTGCTTGTGAATATTAA CCAAGGTCACCC CAGTTATCTGAGGATCAAAA CAGGGGCTAAGTCCA C
801	5332_Human_quadMut_1A_14A_46A_62C_n1	AGGGGAGGGCTGCTAGTGAATATTAA CCAAGGTCACCC CAGTTATCAGAGGAGCAAAA CAGGGCCTAAGTCCA C
802	5335_Human_quadMut_1T_14T_52A_62A_n1	TGGGAGGGCTGCTTGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGAACAAA CAGGGACTAAGTCCA C
803	5338_Human_quadMut_1T_35G_37A_45T_n1	TGGGAGGGCTGCTGGTGAATATTAA CCAAGGTCAGCA CAGTTATTTGGAGGAGCAAAA CAGGGGCTAAGTCCA C
804	5339_Human_quadMut_1A_35T_37T_46T_n1	AGGGGAGGGCTGCTGGTGAATATTAA CCAAGGTCATCT CAGTTATCTGAGGAGCAAAA CAGGGGCTAAGTCCA C
805	5342_Human_quadMut_1A_35G_37T_62C_n1	AGGGGAGGGCTGCTGGTGAATATTAA CCAAGGTCAGCT CAGTTATCGGAGGAGCAAAA CAGGGCCTAAGTCCA C
806	5348_Human_quadMut_1T_35G_46T_48T_n1	TGGGAGGGCTGCTGGTGAATATTAA CCAAGGTCAGCC CAGTTATCTGTGGAGCAAAA CAGGGGCTAAGTCCA C
807	5349_Human_quadMut_1A_35G_46A_52T_n1	AGGGGAGGGCTGCTGGTGAATATTAA CCAAGGTCAGCC CAGTTATCAGAGGATCAAAA CAGGGGCTAAGTCCA C
808	5350_Human_quadMut_1T_35G_46A_72T_n1	TGGGAGGGCTGCTGGTGAATATTAA CCAAGGTCAGCC CAGTTATCAGAGGAGCAAAA CAGGGGCTAAGTCCA T

809	5353_Human_quadMut_1T_35G_48G_72T_n1	TGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAGCCCAAGTTATCGGGGAGCAAAACAGGGGCTAAGTCCA T
810	5354_Human_quadMut_1A_35A_52C_62C_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCAGTTATCGGAGGACCAAAACAGGGCCTAAGTCCA C
811	5356_Human_quadMut_1A_35T_62C_72A_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCCCAGTTATCGGAGGAGCAAAACAGGGCCTAAGTCCA A
812	5358_Human_quadMut_1A_37A_45A_52T_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACACAGTTATAGGAGGATCAAAACAGGGGCTAAGTCCA C
813	5361_Human_quadMut_1A_37A_46T_48T_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCACAGTTATCTGTGGAGCAAAACAGGGGCTAAGTCCA C
814	5362_Human_quadMut_1T_37T_46T_52A_n1	TGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCTCAGTTATCTGAGGAA CAAACAGGGGCTAAGTCCA C
815	5372_Human_quadMut_1A_45G_46A_52A_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCAGTTATGAGAGGAA CAAACAGGGGCTAAGTCCA C
816	5377_Human_quadMut_1A_45G_48G_72T_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCAGTTATGAGAGGAA CAAACAGGGGCTAAGTCCA T
817	5378_Human_quadMut_1T_45A_52A_62T_n1	TGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCAGTTATAGGAGGAA CAAACAGGGTCTAAGTCCA C
818	5381_Human_quadMut_1A_46A_48G_62T_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCAGTTATCAGGGGAGCAAAACAGGGTCTAAGTCCA C
819	5382_Human_quadMut_1A_46A_48G_72A_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCAGTTATCAGGGGAGCAAAACAGGGGCTAAGTCCA A
820	5384_Human_quadMut_1T_46T_52T_72A_n1	TGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCAGTTATCTGAGGATCAAAACAGGGGCTAAGTCCA A
821	5386_Human_quadMut_1A_48G_52T_62C_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCAGTTATCGGGGGATCAAAACAGGGCCTAAGTCCA C
822	5388_Human_quadMut_1A_48G_62C_72T_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCAGTTATCGGGGAGCAAAACAGGGCCTAAGTCCA T
823	5392_Human_quadMut_3A_8A_14T_45G_n1	GGAGGAGCTGCTTGTGAATATTAA CCAAGGTCACCCAGTTATGGGAGGAGCAAAACAGGGGCTAAGTCCA C
824	5398_Human_quadMut_3A_8A_35T_37T_n1	GGAGGAGCTGCTGGTGAATATTAA CCAAGGTCACCTCAGTTATCAGAGGAGCAAAACAGGGGCTAAGTCCA C
825	5406_Human_quadMut_3A_8C_37T_46A_n1	GGAGGAGCCTGCTGGTGAATATTAA CCAAGGTCACCTCAGTTATCGGGGAGCAAAACAGGGGCTAAGTCCA C
826	5407_Human_quadMut_3T_8C_37T_48G_n1	GGTGGAGCCTGCTGGTGAATATTAA CCAAGGTCACCTCAGTTATCGGGGAGCAAAACAGGGGCTAAGTCCA C

827	5408_Human_quadMut_3T_8A_37T_52T_n1	GGTGGAGACTGCTGGTGAATATTAA CCAAGGTCACCTCAGTTATCGGAGGATCAAAACAGGGGCTAAGTCCA C
828	5409_Human_quadMut_3T_8T_37A_62A_n1	GGTGGAGTCTGCTGGTGAATATTAA CCAAGGTACCA CACAGTTATCGGAGGAGCAAAACAGGGACTAAGTCCA C
829	5423_Human_quadMut_3A_8A_52T_62A_n1	GGAGGAGACTGCTGGTGAATATTAA CCAAGGTCACCCCAGTTATCGGAGGATCAAAACAGGGACTAAGTCCA C
830	5424_Human_quadMut_3A_8A_52T_72G_n1	GGAGGAGACTGCTGGTGAATATTAA CCAAGGTACCCCAGTTATCGGAGGATCAAAACAGGGGCTAAGTCCA G
831	5425_Human_quadMut_3A_8T_62T_72A_n1	GGAGGAGTCTGCTGGTGAATATTAA CCAAGGTACCCCAGTTATCGGAGGAGCAAAACAGGGTCTAAGTCCA A
832	5430_Human_quadMut_3A_14T_35G_52T_n1	GGAGGAGGCTGCTTGTGAATATTAA CCAAGGTCAGCCCAGTTATCGGAGGATCAAAACAGGGGCTAAGTCCA C
833	5435_Human_quadMut_3T_14T_37A_48G_n1	GGTGGAGGCTGCTTGTGAATATTAA CCAAGGTACCA CACAGTTATCGGGGGAGCAAAACAGGGGCTAAGTCCA C
834	5436_Human_quadMut_3A_14A_37T_52T_n1	GGAGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCTCAGTTATCGGAGGATCAAAACAGGGGCTAAGTCCA C
835	5443_Human_quadMut_3T_14A_45T_72A_n1	GGTGGAGGCTGCTAGTGAATATTAA CCAAGGTACCCCAGTTATTGGAGGAGCAAAACAGGGGCTAAGTCCA A
836	5444_Human_quadMut_3A_14A_46A_48T_n1	GGAGGAGGCTGCTAGTGAATATTAA CCAAGGTACCCCAGTTATCAGTGGAGCAAAACAGGGGCTAAGTCCA C
837	5447_Human_quadMut_3A_14A_46A_72A_n1	GGAGGAGGCTGCTAGTGAATATTAA CCAAGGTACCCCAGTTATCAGAGGAGCAAAACAGGGGCTAAGTCCA A
838	5450_Human_quadMut_3A_14T_48T_72G_n1	GGAGGAGGCTGCTTGTGAATATTAA CCAAGGTACCCCAGTTATCGGTGGAGCAAAACAGGGGCTAAGTCCA G
839	5451_Human_quadMut_3A_14A_52A_62C_n1	GGAGGAGGCTGCTAGTGAATATTAA CCAAGGTACCCCAGTTATCGGAGGAA CAAACAGGGCCTAAGTCCA C
840	5455_Human_quadMut_3A_35A_37A_46A_n1	GGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCAACACAGTTATCAGAGGAGCAAAACAGGGGCTAAGTCCA C
841	5456_Human_quadMut_3A_35A_37A_48G_n1	GGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCAACACAGTTATCGGGGGAGCAAAACAGGGGCTAAGTCCA C
842	5457_Human_quadMut_3T_35T_37T_52C_n1	GGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCACTCAGTTATCGGAGGACCAAAACAGGGGCTAAGTCCA C
843	5465_Human_quadMut_3T_35T_46A_48G_n1	GGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCACTCCAGTTATCAGGGGAGCAAAACAGGGGCTAAGTCCA C
844	5467_Human_quadMut_3T_35A_46A_62C_n1	GGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCAACCCAGTTATCAGAGGAGCAAAACAGGGCCTAAGTCCA C

845	5468_Human_quadMut_3A_35A_46T_72T_n1	GGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCAGTTTATCTGAGGAGCAAAACAGGGGCTAAGTCCA T
846	5481_Human_quadMut_3T_37T_46T_52T_n1	GGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCTCAGTTTATCTGAGGATCAAAACAGGGGCTAAGTCCA C
847	5482_Human_quadMut_3T_37T_46A_62T_n1	GGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCTCAGTTTATCAGAGGAGCAAAACAGGGTCTAAGTCCA C
848	5485_Human_quadMut_3A_37A_48G_62C_n1	GGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACACACAGTTTATCGGGGAGCAAAACAGGGCCTAAGTCCA C
849	5487_Human_quadMut_3T_37A_52T_62C_n1	GGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCACAGTTTATCGGAGGATCAAAACAGGGCCTAAGTCCA C
850	5488_Human_quadMut_3A_37A_52C_72A_n1	GGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCACAGTTTATCGGAGGACCAAAACAGGGGCTAAGTCCA A
851	5491_Human_quadMut_3A_45G_46T_52C_n1	GGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCAGTTTATCTGAGGACCAAAACAGGGGCTAAGTCCA C
852	5492_Human_quadMut_3A_45A_46A_62T_n1	GGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCAGTTTATAAGAGGAGCAAAACAGGGTCTAAGTCCA C
853	5500_Human_quadMut_3A_46T_48G_52T_n1	GGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCAGTTTATCTGGGGATCAAAACAGGGGCTAAGTCCA C
854	5503_Human_quadMut_3A_46A_52T_62C_n1	GGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCAGTTTATCAGAGGATCAAAACAGGGCCTAAGTCCA C
855	5505_Human_quadMut_3A_46A_62C_72A_n1	GGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCAGTTTATCAGAGGAGCAAAACAGGGCCTAAGTCCA A
856	5509_Human_quadMut_8A_14A_35A_37T_n1	GGGGAGACTGCTAGTGAATATTAA CCAAGGTCAACTCAGTTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA C
857	5513_Human_quadMut_8C_14T_35A_62T_n1	GGGGAGCCTGCTTGTGAATATTAA CCAAGGTCACCCAGTTTATCGGAGGAGCAAAACAGGGTCTAAGTCCA C
858	5517_Human_quadMut_8T_14T_37T_48G_n1	GGGGAGTCTGCTTGTGAATATTAA CCAAGGTCACCTCAGTTTATCGGGGAGCAAAACAGGGGCTAAGTCCA C
859	5519_Human_quadMut_8A_14A_37A_62A_n1	GGGGAGACTGCTAGTGAATATTAA CCAAGGTCACCACAGTTTATCGGAGGAGCAAAACAGGGACTAAGTCCA C
860	5526_Human_quadMut_8C_14T_46A_48T_n1	GGGGAGCCTGCTTGTGAATATTAA CCAAGGTCACCCAGTTTATCAGTGGAGCAAAACAGGGGCTAAGTCCA C
861	5527_Human_quadMut_8T_14T_46A_52T_n1	GGGGAGTCTGCTTGTGAATATTAA CCAAGGTCACCCAGTTTATCAGAGGATCAAAACAGGGGCTAAGTCCA C
862	5531_Human_quadMut_8C_14T_48T_62T_n1	GGGGAGCCTGCTTGTGAATATTAA CCAAGGTCACCCAGTTTATCGTGGAGCAAAACAGGGTCTAAGTCCA C

863	5532_Human_quadMut_8C_14T_48G_72A_n1	GGGGAGCCTGCTTGTGAATATTAA CCAAGGTCACCC CAGTTATCGGGGAGCAAA CAGGGGCTAAGTCCA	A
864	5533_Human_quadMut_8T_14T_52A_62C_n1	GGGGAGTCTGCTTGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGAA CAAA CAGGGCCTAAGTCCA	C
865	5537_Human_quadMut_8A_35A_37A_46A_n1	GGGGAGACTGCTGGTGAATATTAA CCAAGGTCACAC CAGTTATCAGAGGAGCAAA CAGGGGCTAAGTCCA	C
866	5540_Human_quadMut_8C_35T_37T_62A_n1	GGGGAGCCTGCTGGTGAATATTAA CCAAGGTCATCT CAGTTATCGGAGGAGCAAA CAGGGACTAAGTCCA	C
867	5546_Human_quadMut_8C_35T_46A_48T_n1	GGGGAGCCTGCTGGTGAATATTAA CCAAGGTCATCCC CAGTTATCAGTGGAGCAAA CAGGGGCTAAGTCCA	C
868	5547_Human_quadMut_8C_35G_46A_52A_n1	GGGGAGCCTGCTGGTGAATATTAA CCAAGGTCAGCCC CAGTTATCAGAGGAA CAAA CAGGGGCTAAGTCCA	C
869	5548_Human_quadMut_8T_35A_46A_62T_n1	GGGGAGTCTGCTGGTGAATATTAA CCAAGGTCACCCC CAGTTATCAGAGGAGCAAA CAGGGTCTAAGTCCA	C
870	5551_Human_quadMut_8C_35T_48G_62T_n1	GGGGAGCCTGCTGGTGAATATTAA CCAAGGTCATCCC CAGTTATCGGGGAGCAAA CAGGGTCTAAGTCCA	C
872	5553_Human_quadMut_8C_35A_52T_62A_n1	GGGGAGCCTGCTGGTGAATATTAA CCAAGGTCAA CCCCAGTTATCGGAGGATCAAA CAGGGACTAAGTCCA	C
873	5554_Human_quadMut_8A_35T_52A_72A_n1	GGGGAGACTGCTGGTGAATATTAA CCAAGGTCATCCC CAGTTATCGGAGGAA CAAA CAGGGGCTAAGTCCA	A
874	5556_Human_quadMut_8A_37T_45T_46A_n1	GGGGAGACTGCTGGTGAATATTAA CCAAGGTCACCT CAGTTATTAGAGGAGCAAA CAGGGGCTAAGTCCA	C
875	5557_Human_quadMut_8C_37T_45G_48T_n1	GGGGAGCCTGCTGGTGAATATTAA CCAAGGTCACCT CAGTTATGGGTGGAGCAAA CAGGGGCTAAGTCCA	C
876	5558_Human_quadMut_8A_37A_45G_52A_n1	GGGGAGACTGCTGGTGAATATTAA CCAAGGTCACCA CAGTTATGGAGGAA CAAA CAGGGGCTAAGTCCA	C
877	5561_Human_quadMut_8C_37A_46T_48G_n1	GGGGAGCCTGCTGGTGAATATTAA CCAAGGTCACCA CAGTTATCTGGGAGCAAA CAGGGGCTAAGTCCA	C
878	5562_Human_quadMut_8C_37A_46T_52T_n1	GGGGAGCCTGCTGGTGAATATTAA CCAAGGTCACCA CAGTTATCTGAGGATCAAA CAGGGGCTAAGTCCA	C
879	5563_Human_quadMut_8C_37T_46A_62T_n1	GGGGAGCCTGCTGGTGAATATTAA CCAAGGTCACCT CAGTTATCAGAGGAGCAAA CAGGGTCTAAGTCCA	C
880	5566_Human_quadMut_8T_37T_48G_62C_n1	GGGGAGTCTGCTGGTGAATATTAA CCAAGGTCACCT CAGTTATCGGGGAGCAAA CAGGGCCTAAGTCCA	C
881	5567_Human_quadMut_8A_37A_52A_62T_n1	GGGGAGACTGCTGGTGAATATTAA CCAAGGTCACCA CAGTTATCGGAGGAA CAAA CAGGGTCTAAGTCCA	C

882	5568_Human_quadMut_8C_37T_52A_72A_n1	GGGGAGCCTGCTGGTGAATATTAA CCAAGGTCACCTCAGTTATCGGAGGAACAAAACAGGGGCTAAGTCCA A
883	5569_Human_quadMut_8A_37T_62T_72A_n1	GGGGAGACTGCTGGTGAATATTAA CCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAGGGTCTAAGTCCA A
884	5572_Human_quadMut_8T_45T_46T_72A_n1	GGGGAGTCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATTGAGGAGCAAAACAGGGGCTAAGTCCA A
885	5573_Human_quadMut_8T_45G_48T_52A_n1	GGGGAGTCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTAIGGGTGGAAACAAAACAGGGGCTAAGTCCA C
886	5576_Human_quadMut_8C_45T_52T_62A_n1	GGGGAGCCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATTGAGGATCAAAACAGGGACTAAGTCCA C
887	5580_Human_quadMut_8C_46T_48G_62T_n1	GGGGAGCCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCTGGGGAGCAAAACAGGGTCTAAGTCCA C
888	5581_Human_quadMut_8A_46T_52C_62C_n1	GGGGAGACTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCTGAGGACCAAAACAGGGCCTAAGTCCA C
889	5584_Human_quadMut_8C_48G_52A_62T_n1	GGGGAGCCTGCTGGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGGGGAACAAAACAGGGTCTAAGTCCA C
890	5590_Human_quadMut_14A_35T_37T_48G_n1	GGGGAGGCTGCTAGTGAATATTAA CCAAGGTCATCTCAGTTATCGGGGAGCAAAACAGGGGCTAAGTCCA C
891	5592_Human_quadMut_14A_35G_37T_62T_n1	GGGGAGGCTGCTAGTGAATATTAA CCAAGGTCAGCTCAGTTATCGGAGGAGCAAAACAGGGTCTAAGTCCA C
892	5594_Human_quadMut_14A_35G_45G_46T_n1	GGGGAGGCTGCTAGTGAATATTAA CCAAGGTCAGCCCAAGTTATGTGAGGAGCAAAACAGGGGCTAAGTCCA C
893	5599_Human_quadMut_14A_35A_46A_48G_n1	GGGGAGGCTGCTAGTGAATATTAA CCAAGGTCAACCCAGTTATCAGGGGAGCAAAACAGGGGCTAAGTCCA C
894	5601_Human_quadMut_14A_35T_46T_62C_n1	GGGGAGGCTGCTAGTGAATATTAA CCAAGGTCATCCCAGTTATCTGAGGAGCAAAACAGGGCCTAAGTCCA C
895	5604_Human_quadMut_14A_35A_48G_62T_n1	GGGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGGGAGCAAAACAGGGTCTAAGTCCA C
896	5606_Human_quadMut_14T_35A_52A_62C_n1	GGGGAGGCTGCTTGTGAATATTAA CCAAGGTCAACCCAGTTATCGGAGGAACAAAACAGGGCCTAAGTCCA C
897	5608_Human_quadMut_14A_35A_62T_72T_n1	GGGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCCCAAGTTATCGGAGGAGCAAAACAGGGTCTAAGTCCA T
898	5618_Human_quadMut_14A_37T_52A_62T_n1	GGGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCTCAGTTATCGGAGGAACAAAACAGGGTCTAAGTCCA C
899	5633_Human_quadMut_14A_46T_48G_72T_n1	GGGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCCCAAGTTATCTGGGGAGCAAAACAGGGGCTAAGTCCA T

900	5634_Human_quadMut_14A_46A_52C_62C_n1	GGGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCC CAGTTATCAGAGGACCAAA CAGGGCCTAAGTCCA	C
901	5635_Human_quadMut_14A_46T_52T_72A_n1	GGGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCC CAGTTATCTGAGGATCAAA CAGGGGCTAAGTCCA	A
902	5640_Human_quadMut_35T_37T_45T_46A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCTCAGTTATTAGAGGACAAA CAGGGGCTAAGTCCA	C
903	5641_Human_quadMut_35T_37A_45A_48T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCA CAGTTATAGGTGGAGCAAA CAGGGGCTAAGTCCA	C
904	5642_Human_quadMut_35T_37T_45T_52T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCTCAGTTATTGGAGGATCAAA CAGGGGCTAAGTCCA	C
905	5645_Human_quadMut_35A_37A_46T_48G_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAA CACAGTTATCTGGGAGCAAA CAGGGGCTAAGTCCA	C
906	5646_Human_quadMut_35G_37A_46A_52C_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAG CACAGTTATCAGAGGACCAAA CAGGGGCTAAGTCCA	C
907	5647_Human_quadMut_35G_37T_46A_62T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAGCTCAGTTATCAGAGGACCAAA CAGGGTCTAAGTCCA	C
908	5648_Human_quadMut_35A_37A_46T_72G_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAA CACAGTTATCTGAGGAGCAAA CAGGGGCTAAGTCCA	G
909	5649_Human_quadMut_35A_37A_48T_52T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAA CACAGTTATCGGTGGATCAAA CAGGGGCTAAGTCCA	C
910	5652_Human_quadMut_35A_37A_52T_62A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAA CACAGTTATCGGAGGATCAAA CAGGGACTAAGTCCA	C
911	5653_Human_quadMut_35A_37T_52T_72G_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAA CACTCAGTTATCGGAGGATCAAA CAGGGGCTAAGTCCA	G
912	5654_Human_quadMut_35T_37T_62T_72T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCTCAGTTATCGGAGGACCAAA CAGGGTCTAAGTCCA	T
913	5655_Human_quadMut_35A_45T_46T_48T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAA CCCCAGTTATTGTGGAGCAAA CAGGGGCTAAGTCCA	C
914	5656_Human_quadMut_35A_45T_46T_52T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAA CCCCAGTTATTGAGGATCAAA CAGGGGCTAAGTCCA	C
915	5658_Human_quadMut_35G_45G_46A_72A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAG CCCCAGTTATGAGAGGACCAAA CAGGGGCTAAGTCCA	A
916	5660_Human_quadMut_35A_45A_48G_62T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAA CCCCAGTTATAGGGGGAGCAAA CAGGGTCTAAGTCCA	C
917	5661_Human_quadMut_35T_45T_48T_72A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAT CCCCAGTTATTGGTGGAGCAAA CAGGGGCTAAGTCCA	A

918	5662_Human_quadMut_35G_45T_52A_62T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAGCC CAGTTATTGGAGGA CAAACACAGGGTCTAAGTCCA C
919	5665_Human_quadMut_35A_46A_48T_52A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCAGTGGAA CAAACAGGGGCTAAGTCCA C
920	5666_Human_quadMut_35A_46A_48G_62C_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCAGGGGAGCAAA CAGGGCCTAAGTCCA C
921	5668_Human_quadMut_35A_46A_52T_62T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCAGAGGATCAAA CAGGGTCTAAGTCCA C
922	5669_Human_quadMut_35T_48G_52A_62T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCCC CAGTTATCGGGGGA CAAACAGGGTCTAAGTCCA C
923	5670_Human_quadMut_35G_48G_52C_72G_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAGCCC CAGTTATCGGGGGA CCAAA CAGGGGCTAAGTCCA G
924	5672_Human_quadMut_35A_52C_62T_72G_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGA CCAAA CAGGGTCTAAGTCCA G
925	5674_Human_quadMut_37T_45G_46T_52A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCT CAGTTATGTAGGAA CAAACAGGGGCTAAGTCCA C
926	5675_Human_quadMut_37A_45G_46A_62T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCA CAGTTATGAGAGGA CCAAA CAGGGTCTAAGTCCA C
927	5683_Human_quadMut_37T_46A_48G_52A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCT CAGTTATCAGGGGA CCAAA CAGGGGCTAAGTCCA C
928	5685_Human_quadMut_37A_46A_52C_62T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCA CAGTTATCAGAGGA CCAAA CAGGGTCTAAGTCCA C
929	5687_Human_quadMut_37T_46T_62T_72T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCT CAGTTATCTGAGGA CCAAA CAGGGTCTAAGTCCA T
930	5689_Human_quadMut_37T_48G_52T_72A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCT CAGTTATCTGAGGA CCAAA CAGGGGCTAAGTCCA A
931	5690_Human_quadMut_37A_48G_62T_72A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCA CAGTTATCGGGGGA CCAAA CAGGGTCTAAGTCCA A
932	5691_Human_quadMut_37T_52T_62A_72T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCT CAGTTATCGGAGGA CCAAA CAGGGACTAAGTCCA T
933	5695_Human_quadMut_45G_46A_52A_62C_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATGAGAGGA CCAAA CAGGGCCTAAGTCCA C
934	5696_Human_quadMut_45G_46A_52A_72G_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATGAGAGGA CCAAA CAGGGGCTAAGTCCA G
935	5702_Human_quadMut_46A_48G_52A_62A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCAGGGGA CCAAA CAGGGACTAAGTCCA C

936	5704_Human_quadMut_46T_48G_62T_72T_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCTGGGAGCAAA CAGGGTCTAAGTCCA	T
937	5705_Human_quadMut_46T_52C_62T_72A_n1	GGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCTGAGGACCAAA CAGGGTCTAAGTCCA	A
938	5708_Human_quadMut_1T_3T_8T_37T_n1	TGTGAGTCTGCTGGTGAATATTAA CCAAGGTCACCT CAGTTATCGGAGGAGCAAA CAGGGGCTAAGTCCA	C
939	5712_Human_quadMut_1A_3T_8C_52A_n1	AGTGGAGCCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGAA CAAA CAGGGGCTAAGTCCA	C
940	5724_Human_quadMut_1A_3A_35T_72A_n1	AGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCCC CAGTTATCGGAGGAGCAAA CAGGGGCTAAGTCCA	A
941	5726_Human_quadMut_1T_3A_37A_48T_n1	TGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCA CAGTTATCGGTGGAGCAAA CAGGGGCTAAGTCCA	C
942	5728_Human_quadMut_1T_3A_37T_62C_n1	TGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCT CAGTTATCGGAGGAGCAAA CAGGGCCTAAGTCCA	C
943	5730_Human_quadMut_1A_3A_45T_46T_n1	AGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCC CAGTTATTGAGGAGCAAA CAGGGGCTAAGTCCA	C
944	5733_Human_quadMut_1T_3A_45A_62T_n1	TGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATAGGAGGAGCAAA CAGGGTCTAAGTCCA	C
945	5736_Human_quadMut_1T_3T_46A_52A_n1	TGTGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCC CAGTTATCAGAGGAA CAAA CAGGGGCTAAGTCCA	C
946	5739_Human_quadMut_1T_3A_48T_52A_n1	TGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGGTGGAA CAAA CAGGGGCTAAGTCCA	C
947	5740_Human_quadMut_1A_3T_48T_62T_n1	AGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCC CAGTTATCGGTGGAGCAAA CAGGGTCTAAGTCCA	C
948	5743_Human_quadMut_1T_3T_52T_72G_n1	TGTGAGGCTGCTGGTGAATATTAA CCAAGGTCACCCC CAGTTATCGGAGGATCAAA CAGGGGCTAAGTCCA	G
949	5745_Human_quadMut_1A_8A_14T_37T_n1	AGGGAGACTGCTTGTGAATATTAA CCAAGGTCACCT CAGTTATCGGAGGAGCAAA CAGGGGCTAAGTCCA	C
950	5754_Human_quadMut_1T_8A_35T_52T_n1	TGGGAGACTGCTGGTGAATATTAA CCAAGGTCATCCC CAGTTATCGGAGGATCAAA CAGGGGCTAAGTCCA	C
951	5759_Human_quadMut_1A_8A_37A_48T_n1	AGGGAGACTGCTGGTGAATATTAA CCAAGGTCACCA CAGTTATCGGTGGAGCAAA CAGGGGCTAAGTCCA	C
952	5760_Human_quadMut_1A_8C_37T_52A_n1	AGGGAGCCTGCTGGTGAATATTAA CCAAGGTCACCT CAGTTATCGGAGGAA CAAA CAGGGGCTAAGTCCA	C
953	5767_Human_quadMut_1A_8C_46T_48G_n1	AGGGAGCCTGCTGGTGAATATTAA CCAAGGTCACCCC CAGTTATCTGGGAGCAAA CAGGGGCTAAGTCCA	C

954	5772_Human_quadMut_1T_8T_48G_62A_n1	TGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGGGGAGCAAA CAGGGACTAAGTCCA	C
955	5776_Human_quadMut_1A_8A_62T_72A_n1	AGGGGAGACTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGAGCAAA CAGGGTCTAAGTCCA	A
956	5779_Human_quadMut_1A_14T_35T_46A_n1	AGGGAGGCTGCTTGTGAATATTAA CCAAGGTCATCC CAGTTATCAGAGGAGCAAA CAGGGGCTAAGTCCA	C
957	5781_Human_quadMut_1T_14A_35A_52A_n1	TGGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGAACAAA CAGGGGCTAAGTCCA	C
958	5787_Human_quadMut_1T_14A_37T_52T_n1	TGGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCT CAGTTATCGGAGGATCAAA CAGGGGCTAAGTCCA	C
959	5791_Human_quadMut_1T_14A_45A_62T_n1	TGGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCC CAGTTATAGGAGGAGCAAA CAGGGTCTAAGTCCA	C
960	5793_Human_quadMut_1T_14A_46T_48T_n1	TGGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCC CAGTTATCTGTGGAGCAAA CAGGGGCTAAGTCCA	C
961	5794_Human_quadMut_1T_14A_46A_52T_n1	TGGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCC CAGTTATCAGAGGATCAAA CAGGGGCTAAGTCCA	C
962	5800_Human_quadMut_1T_14A_52T_62T_n1	TGGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGATCAAA CAGGGTCTAAGTCCA	C
963	5805_Human_quadMut_1T_35G_37T_48G_n1	TGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAGCT CAGTTATCGGGGAGCAAA CAGGGGCTAAGTCCA	C
964	5809_Human_quadMut_1A_35T_45T_46T_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCC CAGTTATTTGAGGAGCAAA CAGGGGCTAAGTCCA	C
965	5810_Human_quadMut_1A_35T_45G_48G_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCC CAGTTATCGGGGAGCAAA CAGGGGCTAAGTCCA	C
966	5814_Human_quadMut_1T_35A_46T_48T_n1	TGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCTGTGGAGCAAA CAGGGGCTAAGTCCA	C
967	5815_Human_quadMut_1A_35T_46A_52C_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCC CAGTTATCAGAGGACCAAA CAGGGGCTAAGTCCA	C
968	5816_Human_quadMut_1T_35A_46T_62T_n1	TGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCTGAGGAGCAAA CAGGGTCTAAGTCCA	C
969	5817_Human_quadMut_1A_35A_46A_72T_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCAGAGGAGCAAA CAGGGGCTAAGTCCA	T
970	5821_Human_quadMut_1A_35G_52A_62T_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCAGCCC CAGTTATCGGAGGAACAAA CAGGGTCTAAGTCCA	C
971	5823_Human_quadMut_1T_35T_62T_72G_n1	TGGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCC CAGTTATCGGAGGAGCAAA CAGGGTCTAAGTCCA	G

972	5824_Human_quadMut_1A_37A_45G_46A_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCA CAGTTATGAGAGGAGCAAAACAGGGGCTAAGTCCA C
973	5827_Human_quadMut_1A_37T_45G_62C_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCTCAGTTATGGAGGAGCAAAACAGGGCCTAAGTCCA C
974	5828_Human_quadMut_1T_37T_45G_72A_n1	TGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCTCAGTTATGGAGGAGCAAAACAGGGGCTAAGTCCA A
975	5829_Human_quadMut_1A_37T_46A_52T_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCTCAGTTATCAGAGGATCAAAACAGGGGCTAAGTCCA C
976	5832_Human_quadMut_1A_37A_48T_52T_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCA CAGTTATCGGTGGATCAAAACAGGGGCTAAGTCCA C
977	5839_Human_quadMut_1A_45G_46T_52T_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATGTGAGGATCAAAACAGGGGCTAAGTCCA C
978	5840_Human_quadMut_1T_45A_46T_62C_n1	TGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATATGAGGAGCAAAACAGGGCCTAAGTCCA C
979	5848_Human_quadMut_1T_46T_48G_52C_n1	TGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCTGGGGACCAAAACAGGGGCTAAGTCCA C
980	5852_Human_quadMut_1T_46T_52T_72T_n1	TGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCTGAGGATCAAAACAGGGGCTAAGTCCA T
981	5855_Human_quadMut_1A_48G_52T_72T_n1	AGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGGGGGATCAAAACAGGGGCTAAGTCCA T
982	5856_Human_quadMut_1T_48T_62C_72A_n1	TGGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGGTGGAGCAAAACAGGGCCTAAGTCCA A
983	5866_Human_quadMut_3T_8A_35T_46A_n1	GGTGAGACTGCTGGTGAATATTAA CCAAGGTCATCC CAGTTATCAGAGGAGCAAAACAGGGGCTAAGTCCA C
984	5867_Human_quadMut_3A_8A_35T_48T_n1	GGAGGAGACTGCTGGTGAATATTAA CCAAGGTCATCC CAGTTATCGGTGGAGCAAAACAGGGGCTAAGTCCA C
985	5868_Human_quadMut_3A_8A_35T_52T_n1	GGAGGAGACTGCTGGTGAATATTAA CCAAGGTCATCC CAGTTATCGGAGGATCAAAACAGGGGCTAAGTCCA C
986	5869_Human_quadMut_3A_8A_35A_62T_n1	GGAGGAGACTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGAGCAAAACAGGGTCTAAGTCCA C
987	5880_Human_quadMut_3A_8T_45G_72A_n1	GGAGGAGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGAGGAAACAAACAGGGGCTAAGTCCA A
988	5882_Human_quadMut_3A_8A_46A_52A_n1	GGTGAGACTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCAGAGGAGCAAAACAGGGGCTAAGTCCA C
990	5883_Human_quadMut_3T_8T_46A_62T_n1	GGTGAGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCAGAGGAGCAAAACAGGGTCTAAGTCCA C

991	5884_Human_quadMut_3T_8C_46A_72A_n1	GGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCAGAGGAGCAAA CAGGGGCTAAGTCCA	A
992	5886_Human_quadMut_3T_8T_48G_62T_n1	GGTGGAGTCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGGGGAGCAAA CAGGGTCTAAGTCCA	C
993	5889_Human_quadMut_3T_8A_52T_72G_n1	GGTGGAGACTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGATCAAA CAGGGGCTAAGTCCA	G
994	5897_Human_quadMut_3A_14A_37T_48T_n1	GGAGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCT CAGTTATCGGTGGAGCAAA CAGGGGCTAAGTCCA	C
995	5899_Human_quadMut_3T_14A_37T_62T_n1	GGTGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCT CAGTTATCGGAGGAGCAAA CAGGGTCTAAGTCCA	C
996	5905_Human_quadMut_3A_14T_45T_72A_n1	GGAGGAGGCTGCTTGTGAATATTAA CCAAGGTCACCC CAGTTATTGGAGGAGCAAA CAGGGGCTAAGTCCA	A
997	5907_Human_quadMut_3T_14A_46T_52T_n1	GGTGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCC CAGTTATCTGAGGATCAAA CAGGGGCTAAGTCCA	C
998	5908_Human_quadMut_3T_14A_46A_62T_n1	GGTGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCC CAGTTATCAGAGGAGCAAA CAGGGTCTAAGTCCA	C
999	5911_Human_quadMut_3T_14T_48T_62C_n1	GGTGGAGGCTGCTTGTGAATATTAA CCAAGGTCACCC CAGTTATCGGTGGAGCAAA CAGGGCCTAAGTCCA	C
1000	5914_Human_quadMut_3T_14A_52A_72A_n1	GGTGGAGGCTGCTAGTGAATATTAA CCAAGGTCACCC CAGTTATCGGAGGAGCAAA CAGGGGCTAAGTCCA	A
1001	5916_Human_quadMut_3A_35A_37T_45G_n1	GGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCAACT CAGTTATGGAGGAGCAAA CAGGGGCTAAGTCCA	C
1002	5917_Human_quadMut_3T_35A_37A_46A_n1	GGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCAA CACACAGTTATCAGAGGAGCAAA CAGGGGCTAAGTCCA	C
1003	5919_Human_quadMut_3A_35T_37A_52T_n1	GGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCATC ACAGTTATCGGAGGATCAAA CAGGGGCTAAGTCCA	C
1004	5921_Human_quadMut_3T_35G_45T_46T_n1	GGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCAGCCC CAGTTATTGAGGAGCAAA CAGGGGCTAAGTCCA	C
1005	5926_Human_quadMut_3T_35T_46A_52C_n1	GGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCC CAGTTATCAGAGGAGCAAA CAGGGGCTAAGTCCA	C
1006	5927_Human_quadMut_3T_35A_46T_62T_n1	GGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCACCC CAGTTATCTGAGGAGCAAA CAGGGTCTAAGTCCA	C
1007	5929_Human_quadMut_3A_35T_48G_52T_n1	GGAGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCC CAGTTATCGGGGGATCAAA CAGGGGCTAAGTCCA	C
1008	5930_Human_quadMut_3T_35T_48T_62C_n1	GGTGGAGGCTGCTGGTGAATATTAA CCAAGGTCATCC CAGTTATCGGTGGAGCAAA CAGGGCCTAAGTCCA	C

100	9	5932_Human_quadMut_3T_35G_52T_72G_n1	GGTGGAGGCTGCTGGTGAATATTAA	CCAAAGGTCAGCCCCAGTTATCGGAGGATCAAAACAGGGGCTAAGTCCA	G
101	0	5933_Human_quadMut_3T_35T_62A_72T_n1	GGTGGAGGCTGCTGGTGAATATTAA	CCAAAGGTCATCCAGTTATCGGAGGAGCAAAACAGGGACTAAGTCCA	T
101	1	5934_Human_quadMut_3T_37A_45G_46T_n1	GGTGGAGGCTGCTGGTGAATATTAA	CCAAAGGTCACCACAGTTATGTGAGGAGCAAAACAGGGGCTAAGTCCA	C
101	2	5936_Human_quadMut_3T_37A_45G_52A_n1	GGTGGAGGCTGCTGGTGAATATTAA	CCAAAGGTCACCACAGTTATGGGAGGAACAAACAGGGGCTAAGTCCA	C
101	3	5941_Human_quadMut_3A_37T_48T_72T_n1	GGAGGAGGCTGCTGGTGAATATTAA	CCAAAGGTCACCTCAGTTATCGGTGGAGCAAAACAGGGGCTAAGTCCA	T
101	4	5944_Human_quadMut_3A_37A_62T_72A_n1	GGAGGAGGCTGCTGGTGAATATTAA	CCAAAGGTCACCACAGTTATCGGAGGAGCAAAACAGGGTCTAAGTCCA	A
101	5	5948_Human_quadMut_3A_45A_46A_72A_n1	GGAGGAGGCTGCTGGTGAATATTAA	CCAAAGGTCACCCCAGTTATAAGAGGAGCAAAACAGGGGCTAAGTCCA	A
101	6	5950_Human_quadMut_3T_45A_48T_62T_n1	GGTGGAGGCTGCTGGTGAATATTAA	CCAAAGGTCACCCCAGTTATAGGTGGAGCAAAACAGGGTCTAAGTCCA	C
101	7	5956_Human_quadMut_3T_46A_48T_72A_n1	GGTGGAGGCTGCTGGTGAATATTAA	CCAAAGGTCACCCCAGTTATCAGTGGAGCAAAACAGGGGCTAAGTCCA	A
101	8	5962_Human_quadMut_3T_52T_62T_72T_n1	GGTGGAGGCTGCTGGTGAATATTAA	CCAAAGGTCACCCCAGTTATCGGAGGATCAAAACAGGGTCTAAGTCCA	T
101	9	5963_Human_quadMut_8T_14A_35G_37T_n1	GGGGAGTCTGCTAGTGAATATTAA	CCAAAGGTCAGCTCAGTTATCGGAGGAGCAAAACAGGGGCTAAGTCCA	C
102	0	5968_Human_quadMut_8T_14T_35A_62T_n1	GGGGAGTCTGCTTGTGAATATTAA	CCAAAGGTCAACCCAGTTATCGGAGGAGCAAAACAGGGTCTAAGTCCA	C
102	1	5971_Human_quadMut_8T_14A_37A_46A_n1	GGGGAGTCTGCTAGTGAATATTAA	CCAAAGGTCACCACAGTTATCAGAGGAGCAAAACAGGGGCTAAGTCCA	C
102	2	5984_Human_quadMut_8A_14A_48G_52T_n1	GGGGGAGACTGCTAGTGAATATTAA	CCAAAGGTCACCCCAGTTATCGGGGGATCAAAACAGGGGCTAAGTCCA	C
102	3	5985_Human_quadMut_8T_14T_48G_62A_n1	GGGGAGTCTGCTTGTGAATATTAA	CCAAAGGTCACCCCAGTTATCGGGGGAGCAAAACAGGGACTAAGTCCA	C
102	4	5990_Human_quadMut_8A_35T_37T_45G_n1	GGGGGAGACTGCTGGTGAATATTAA	CCAAAGGTCATCTCAGTTATGGGAGGAGCAAAACAGGGGCTAAGTCCA	C
102	5	5991_Human_quadMut_8A_35G_37T_46T_n1	GGGGGAGACTGCTGGTGAATATTAA	CCAAAGGTCAGCTCAGTTATCTGAGGAGCAAAACAGGGGCTAAGTCCA	C

102 6	5993_Human_quadMut_8C_35A_37T_52T_n1	GGGGAGCCTGCTGGTGAATATTAAACCAAGGTCAACTCAGTTATCGGAGGATCAAAACAGGGGCTAAGTCCA C
102 7	5995_Human_quadMut_8A_35G_45A_46A_n1	GGGGAGACTGCTGGTGAATATTAAACCAAGGTCAAGCCAGTTCAGTTATAAGAGGAGCAAAACAGGGGCTAAGTCCA C
102 8	5996_Human_quadMut_8C_35A_45T_48T_n1	GGGGAGCCTGCTGGTGAATATTAAACCAAGGTCAACCCAGTTCAGTTATGGTGGAGCAAAACAGGGGCTAAGTCCA C

Table 11. Single or adjacent di- nucleotide substitution variants of Chinese Tree Shrew SERPINA1 enhancer with higher luciferase expression than original sequence SEQ ID NO: 122

SEQ ID NO:	Chinese Tree Shrew SerpEnh variant	Sequence
1029	2290_ChineseTreeShrewMod_monoMut_G2A_n1	GAAGGCTGTTGGTGAATATTAAACCAAGGTCACTCAGTTATCGGAGGAGCAAAACAAAGGGCTAAGTCCAC
1030	2293_ChineseTreeShrewMod_monoMut_A3C_n1	GGCGGCTGTTGGTGAATATTAAACCAAGGTCACTCAGTTATCGGAGGAGCAAAACAAAGGGCTAAGTCCAC
1031	2295_ChineseTreeShrewMod_monoMut_A3T_n1	GGTGGCTGTTGGTGAATATTAAACCAAGGTCACTCAGTTATCGGAGGAGCAAAACAAAGGGCTAAGTCCAC
1032	2296_ChineseTreeShrewMod_monoMut_G4A_n1	GGAAGCTGTTGGTGAATATTAAACCAAGGTCACTCAGTTATCGGAGGAGCAAAACAAAGGGCTAAGTCCAC
1033	2298_ChineseTreeShrewMod_monoMut_G4T_n1	GGATGCTGTTGGTGAATATTAAACCAAGGTCACTCAGTTATCGGAGGAGCAAAACAAAGGGCTAAGTCCAC
1034	2302_ChineseTreeShrewMod_monoMut_C6A_n1	GGAGGATGTTGGTGAATATTAAACCAAGGTCACTCAGTTATCGGAGGAGCAAAACAAAGGGCTAAGTCCAC
1035	2305_ChineseTreeShrewMod_monoMut_T7A_n1	GGAGGCAGTTGGTGAATATTAAACCAAGGTCACTCAGTTATCGGAGGAGCAAAACAAAGGGCTAAGTCCAC
1036	2309_ChineseTreeShrewMod_monoMut_G8C_n1	GGAGGCTCTGGTGAATATTAAACCAAGGTCACTCAGTTATCGGAGGAGCAAAACAAAGGGCTAAGTCCAC
1037	2310_ChineseTreeShrewMod_monoMut_G8T_n1	GGAGGCTTTGGTGAATATTAAACCAAGGTCACTCAGTTATCGGAGGAGCAAAACAAAGGGCTAAGTCCAC
1038	2314_ChineseTreeShrewMod_monoMut_T10A_n1	GGAGGCTGTAGGTGAATATTAAACCAAGGTCACTCAGTTATCGGAGGAGCAAAACAAAGGGCTAAGTCCAC
1039	2316_ChineseTreeShrewMod_monoMut_T10G_n1	GGAGGCTGTGGGTGAATATTAAACCAAGGTCACTCAGTTATCGGAGGAGCAAAACAAAGGGCTAAGTCCAC
1040	2317_ChineseTreeShrewMod_monoMut_G11A_n1	GGAGGCTGTAGTGAATATTAAACCAAGGTCACTCAGTTATCGGAGGAGCAAAACAAAGGGCTAAGTCCAC
1041	2319_ChineseTreeShrewMod_monoMut_G11T_n1	GGAGGCTGTTTGTGAATATTAAACCAAGGTCACTCAGTTATCGGAGGAGCAAAACAAAGGGCTAAGTCCAC
1042	2321_ChineseTreeShrewMod_monoMut_G12C_n1	GGAGGCTGTTGTGAATATTAAACCAAGGTCACTCAGTTATCGGAGGAGCAAAACAAAGGGCTAAGTCCAC
1043	2325_ChineseTreeShrewMod_monoMut_T13G_n1	GGAGGCTGTGGGGAATATTAAACCAAGGTCACTCAGTTATCGGAGGAGCAAAACAAAGGGCTAAGTCCAC
1044	2326_ChineseTreeShrewMod_monoMut_G14A_n1	GGAGGCTGTGGTGAATATTAAACCAAGGTCACTCAGTTATCGGAGGAGCAAAACAAAGGGCTAAGTCCAC
1045	2327_ChineseTreeShrewMod_monoMut_G14C_n1	GGAGGCTGTGGTGAATATTAAACCAAGGTCACTCAGTTATCGGAGGAGCAAAACAAAGGGCTAAGTCCAC
1046	2328_ChineseTreeShrewMod_monoMut_G14T_n1	GGAGGCTGTGGTGAATATTAAACCAAGGTCACTCAGTTATCGGAGGAGCAAAACAAAGGGCTAAGTCCAC

1047	2331_ChineseTreeShrewMod_monoMut_A15T_n1	GGAGGCTGTTGGTGTATATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1048	2332_ChineseTreeShrewMod_monoMut_A16C_n1	GGAGGCTGTTGGTGACTATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1049	2334_ChineseTreeShrewMod_monoMut_A16T_n1	GGAGGCTGTTGGTGATTTATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1050	2336_ChineseTreeShrewMod_monoMut_T17C_n1	GGAGGCTGTTGGTGAACATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1051	2337_ChineseTreeShrewMod_monoMut_T17G_n1	GGAGGCTGTTGGTGAAGATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1052	2344_ChineseTreeShrewMod_monoMut_T20A_n1	GGAGGCTGTTGGTGAATATAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1053	2345_ChineseTreeShrewMod_monoMut_T20C_n1	GGAGGCTGTTGGTGAATATCAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1054	2346_ChineseTreeShrewMod_monoMut_T20G_n1	GGAGGCTGTTGGTGAATATGAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1055	2351_ChineseTreeShrewMod_monoMut_A22G_n1	GGAGGCTGTTGGTGAATATTAGCCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1056	2352_ChineseTreeShrewMod_monoMut_A22T_n1	GGAGGCTGTTGGTGAATATTATCCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1057	2360_ChineseTreeShrewMod_monoMut_A25G_n1	GGAGGCTGTTGGTGAATATTAAACCGAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1058	2363_ChineseTreeShrewMod_monoMut_A26G_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1059	2365_ChineseTreeShrewMod_monoMut_G27A_n1	GGAGGCTGTTGGTGAATATTAAACCAAGTCACTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1060	2369_ChineseTreeShrewMod_monoMut_G28C_n1	GGAGGCTGTTGGTGAATATTAAACCAAGCTCACTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1061	2372_ChineseTreeShrewMod_monoMut_T29C_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGCCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1062	2373_ChineseTreeShrewMod_monoMut_T29G_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGCCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1063	2375_ChineseTreeShrewMod_monoMut_C30G_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTGACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1064	2378_ChineseTreeShrewMod_monoMut_A31G_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTGCGCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1065	2379_ChineseTreeShrewMod_monoMut_A31T_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCTCCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1066	2380_ChineseTreeShrewMod_monoMut_C32A_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCAACTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1067	2381_ChineseTreeShrewMod_monoMut_C32G_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1068	2382_ChineseTreeShrewMod_monoMut_C32T_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1069	2383_ChineseTreeShrewMod_monoMut_C33A_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1070	2385_ChineseTreeShrewMod_monoMut_C33T_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1071	2389_ChineseTreeShrewMod_monoMut_C35A_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACTTAAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1072	2390_ChineseTreeShrewMod_monoMut_C35G_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACTTGAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1073	2392_ChineseTreeShrewMod_monoMut_A36C_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACTCCGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1072	2393_ChineseTreeShrewMod_monoMut_A36G_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACTTCGGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1075	2394_ChineseTreeShrewMod_monoMut_A36T_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACTCTGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1076	2396_ChineseTreeShrewMod_monoMut_G37C_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACTCACTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC

1077	2398_ChineseTreeShrewMod_monoMut_T38A_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCAGATATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1078	2399_ChineseTreeShrewMod_monoMut_T38C_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCAGCTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1079	2400_ChineseTreeShrewMod_monoMut_T38G_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCAGGTAATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1080	2402_ChineseTreeShrewMod_monoMut_T39C_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCAGTCAATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1081	2403_ChineseTreeShrewMod_monoMut_T39G_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCAGTGAATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1082	2405_ChineseTreeShrewMod_monoMut_A40G_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCAGTTGTTCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1083	2407_ChineseTreeShrewMod_monoMut_T41A_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCAGTTAACGGAGGAGCAAAACAAGGGCTAAGTCCAC
1084	2409_ChineseTreeShrewMod_monoMut_T41G_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCAGTTAGCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1085	2411_ChineseTreeShrewMod_monoMut_C42G_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCAGTTATTCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1086	2413_ChineseTreeShrewMod_monoMut_G43A_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCAGTTATTCAGAGGAGCAAAACAAGGGCTAAGTCCAC
1087	2414_ChineseTreeShrewMod_monoMut_G43C_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCAGTTATTCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1088	2416_ChineseTreeShrewMod_monoMut_G44A_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCAGTTATTCGAAGGAGCAAAACAAGGGCTAAGTCCAC
1089	2417_ChineseTreeShrewMod_monoMut_G44C_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCAGTTATTCGCAGGAGCAAAACAAGGGCTAAGTCCAC
1090	2419_ChineseTreeShrewMod_monoMut_A45C_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCAGTTATTCGGCGGAGCAAAACAAGGGCTAAGTCCAC
1091	2422_ChineseTreeShrewMod_monoMut_G46A_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCAGTTATTCGGAAAGAGCAAAACAAGGGCTAAGTCCAC
1092	2423_ChineseTreeShrewMod_monoMut_G46C_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCAGTTATTCGGACAGAGCAAAACAAGGGCTAAGTCCAC
1093	2424_ChineseTreeShrewMod_monoMut_G46T_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCAGTTATTCGGATGAGCAAAACAAGGGCTAAGTCCAC
1094	2425_ChineseTreeShrewMod_monoMut_G47A_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCAGTTATTCGGAGAGCAAAACAAGGGCTAAGTCCAC
1095	2426_ChineseTreeShrewMod_monoMut_G47C_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCAGTTATTCGGAGAGCAAAACAAGGGCTAAGTCCAC
1096	2428_ChineseTreeShrewMod_monoMut_A48C_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCAGTTATTCGGAGGCGCAAAACAAGGGCTAAGTCCAC
1097	2429_ChineseTreeShrewMod_monoMut_A48G_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCAGTTATTCGGAGGCGCAAAACAAGGGCTAAGTCCAC
1098	2430_ChineseTreeShrewMod_monoMut_A48T_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCAGTTATTCGGAGGTGCAAAACAAGGGCTAAGTCCAC
1099	2431_ChineseTreeShrewMod_monoMut_G49A_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCAGTTATTCGGAGGAACAACAAGGGCTAAGTCCAC
1100	2433_ChineseTreeShrewMod_monoMut_G49T_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCAGTTATTCGGAGGATCAAAACAAGGGCTAAGTCCAC
1101	2436_ChineseTreeShrewMod_monoMut_C50T_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCAGTTATTCGGAGGAGTAAACAAGGGCTAAGTCCAC
1102	2437_ChineseTreeShrewMod_monoMut_A51C_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCAGTTATTCGGAGGAGCCAAACAAGGGCTAAGTCCAC
1103	2448_ChineseTreeShrewMod_monoMut_C54T_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCAGTTATTCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1104	2450_ChineseTreeShrewMod_monoMut_A55G_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCAGTTATTCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1105	2453_ChineseTreeShrewMod_monoMut_A56G_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCAGTTATTCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1106	2454_ChineseTreeShrewMod_monoMut_A56T_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCAGTTATTCGGAGGAGCAAAACAATGGGCTAAGTCCAC

1107	2457_ChineseTreeShrewMod_monoMut_G57T_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAATGGCTAAGTCCAC
1108	2459_ChineseTreeShrewMod_monoMut_G58C_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGCGTAAAGTCCAC
1109	2460_ChineseTreeShrewMod_monoMut_G58T_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGTGTAAAGTCCAC
1110	2461_ChineseTreeShrewMod_monoMut_G59A_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGCTAAAGTCCAC
1111	2462_ChineseTreeShrewMod_monoMut_G59C_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGCTAAAGTCCAC
1112	2463_ChineseTreeShrewMod_monoMut_G59T_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGCTAAAGTCCAC
1113	2467_ChineseTreeShrewMod_monoMut_T61A_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCAAGTCCAC
1114	2469_ChineseTreeShrewMod_monoMut_T61G_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCAAGTCCAC
1115	2482_ChineseTreeShrewMod_monoMut_C66A_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTACAC
1116	2484_ChineseTreeShrewMod_monoMut_C66T_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTTCAC
1117	2491_ChineseTreeShrewMod_monoMut_C69A_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAA
1118	2498_ChineseTreeShrewMod_diMut_GG1CC_n1	CCAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1119	2499_ChineseTreeShrewMod_diMut_GG1CT_n1	CTAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1110	2503_ChineseTreeShrewMod_diMut_GA2AC_n1	GACGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1121	2504_ChineseTreeShrewMod_diMut_GA2AG_n1	GAGGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1122	2505_ChineseTreeShrewMod_diMut_GA2AT_n1	GATGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1123	2507_ChineseTreeShrewMod_diMut_GA2CG_n1	GCGGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1124	2508_ChineseTreeShrewMod_diMut_GA2CT_n1	GCTGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1125	2510_ChineseTreeShrewMod_diMut_GA2TG_n1	GTGGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1126	2511_ChineseTreeShrewMod_diMut_GA2TT_n1	GTTGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1127	2512_ChineseTreeShrewMod_diMut_AG3CA_n1	GGCAGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1128	2514_ChineseTreeShrewMod_diMut_AG3CT_n1	GGCTGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1129	2516_ChineseTreeShrewMod_diMut_AG3GC_n1	GGGCGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1130	2517_ChineseTreeShrewMod_diMut_AG3GT_n1	GGGTGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1131	2519_ChineseTreeShrewMod_diMut_AG3TC_n1	GGTCGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1132	2520_ChineseTreeShrewMod_diMut_AG3TT_n1	GGTTGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1133	2521_ChineseTreeShrewMod_diMut_GG4AA_n1	GGAAACTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1134	2523_ChineseTreeShrewMod_diMut_GG4AT_n1	GGAACTCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1135	2524_ChineseTreeShrewMod_diMut_GG4CA_n1	GGACACTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1136	2525_ChineseTreeShrewMod_diMut_GG4CC_n1	GGACCCCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC

1137	2526	ChineseTreeShrewMod_diMut_GG4CT_n1	GGACTCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1138	2527	ChineseTreeShrewMod_diMut_GG4TA_n1	GGATACTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1139	2529	ChineseTreeShrewMod_diMut_GG4TT_n1	GGATTCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1140	2530	ChineseTreeShrewMod_diMut_GC5AA_n1	GGAGAATGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1141	2531	ChineseTreeShrewMod_diMut_GC5AG_n1	GGAGAGTGTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1142	2533	ChineseTreeShrewMod_diMut_GC5CA_n1	GGAGCATGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1143	2534	ChineseTreeShrewMod_diMut_GC5CG_n1	GGAGCGTGTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1144	2536	ChineseTreeShrewMod_diMut_GC5TA_n1	GGAGTATGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1145	2538	ChineseTreeShrewMod_diMut_GC5TT_n1	GGAGTTTGTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1146	2540	ChineseTreeShrewMod_diMut_CT6AC_n1	GGAGGACGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1147	2541	ChineseTreeShrewMod_diMut_CT6AG_n1	GGAGGAGGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1148	2544	ChineseTreeShrewMod_diMut_CT6GG_n1	GGAGGGGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1149	2547	ChineseTreeShrewMod_diMut_CT6TG_n1	GGAGGTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1150	2548	ChineseTreeShrewMod_diMut_TG7AA_n1	GGAGGCAATTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1151	2550	ChineseTreeShrewMod_diMut_TG7AT_n1	GGAGGCAATTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1152	2552	ChineseTreeShrewMod_diMut_TG7CC_n1	GGAGGCCCTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1153	2553	ChineseTreeShrewMod_diMut_TG7CT_n1	GGAGGCCCTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1154	2554	ChineseTreeShrewMod_diMut_TG7GA_n1	GGAGGCGATTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1155	2555	ChineseTreeShrewMod_diMut_TG7GC_n1	GGAGGCGCTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1156	2556	ChineseTreeShrewMod_diMut_TG7GT_n1	GGAGGCGCTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1157	2557	ChineseTreeShrewMod_diMut_GT8AA_n1	GGAGGCTAATGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1158	2558	ChineseTreeShrewMod_diMut_GT8AC_n1	GGAGGCTACTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1159	2559	ChineseTreeShrewMod_diMut_GT8AG_n1	GGAGGCTAGTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1160	2560	ChineseTreeShrewMod_diMut_GT8CA_n1	GGAGGCTCATGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1161	2561	ChineseTreeShrewMod_diMut_GT8CC_n1	GGAGGCTCCTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1162	2563	ChineseTreeShrewMod_diMut_GT8TA_n1	GGAGGCTTATGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1163	2564	ChineseTreeShrewMod_diMut_GT8TC_n1	GGAGGCTTCTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1164	2566	ChineseTreeShrewMod_diMut_TT9AA_n1	GGAGGCTGAAGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1165	2567	ChineseTreeShrewMod_diMut_TT9AC_n1	GGAGGCTGACGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1166	2568	ChineseTreeShrewMod_diMut_TT9AG_n1	GGAGGCTGAGGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC

1167	2574_ChineseTreeShrewMod_diMut_TT9GG_n1	GGAGGCTGGGGTGAATATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1168	2575_ChineseTreeShrewMod_diMut_TG10AA_n1	GGAGGCTGAAGTGAATATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1169	2580_ChineseTreeShrewMod_diMut_TG10CT_n1	GGAGGCTGTCTGTGAATATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1170	2582_ChineseTreeShrewMod_diMut_TG10GC_n1	GGAGGCTGTGCGTGAATATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1171	2586_ChineseTreeShrewMod_diMut_GG11AT_n1	GGAGGCTGTATTGAATATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1172	2591_ChineseTreeShrewMod_diMut_GG11TC_n1	GGAGGCTGTTTCTGAATATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1173	2596_ChineseTreeShrewMod_diMut_GT12CA_n1	GGAGGCTGTGCAGAAATATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1174	2597_ChineseTreeShrewMod_diMut_GT12CC_n1	GGAGGCTGTGCCGAATATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1175	2599_ChineseTreeShrewMod_diMut_GT12TA_n1	GGAGGCTGTGTAGAATATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1176	2600_ChineseTreeShrewMod_diMut_GT12TC_n1	GGAGGCTGTGTGCGAATATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1177	2601_ChineseTreeShrewMod_diMut_GT12TG_n1	GGAGGCTGTGTGGAATATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1178	2602_ChineseTreeShrewMod_diMut_TG13AA_n1	GGAGGCTGTGGAAATATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1179	2603_ChineseTreeShrewMod_diMut_TG13AC_n1	GGAGGCTGTGGACAAATATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1180	2605_ChineseTreeShrewMod_diMut_TG13CA_n1	GGAGGCTGTGGCAATATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1181	2606_ChineseTreeShrewMod_diMut_TG13CC_n1	GGAGGCTGTGGCCAAATATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1183	2608_ChineseTreeShrewMod_diMut_TG13GA_n1	GGAGGCTGTGGGAAATATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1183	2609_ChineseTreeShrewMod_diMut_TG13GC_n1	GGAGGCTGTGGGCAATATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1184	2610_ChineseTreeShrewMod_diMut_TG13GT_n1	GGAGGCTGTGGGTAATATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1185	2611_ChineseTreeShrewMod_diMut_GA14AC_n1	GGAGGCTGTGGTACATATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1186	2614_ChineseTreeShrewMod_diMut_GA14CC_n1	GGAGGCTGTGGTCCATATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1187	2615_ChineseTreeShrewMod_diMut_GA14CG_n1	GGAGGCTGTGGTCGATATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1188	2623_ChineseTreeShrewMod_diMut_AA15GC_n1	GGAGGCTGTGGTGGCTATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1189	2626_ChineseTreeShrewMod_diMut_AA15TC_n1	GGAGGCTGTGGTGCTATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1190	2628_ChineseTreeShrewMod_diMut_AA15TT_n1	GGAGGCTGTGGTGTTTATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1191	2629_ChineseTreeShrewMod_diMut_AT16CA_n1	GGAGGCTGTGGTGACAAATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1192	2633_ChineseTreeShrewMod_diMut_AT16GC_n1	GGAGGCTGTGGTGAGCAATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1193	2635_ChineseTreeShrewMod_diMut_AT16TA_n1	GGAGGCTGTGGTGATAATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1194	2636_ChineseTreeShrewMod_diMut_AT16TC_n1	GGAGGCTGTGGTGATCATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1195	2640_ChineseTreeShrewMod_diMut_TA17AT_n1	GGAGGCTGTGGTGAAATTTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1196	2643_ChineseTreeShrewMod_diMut_TA17CT_n1	GGAGGCTGTGGTGAACCTTTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC

1197	2644_ChineseTreeShrewMod_diMut_TA17GC_n1	GGAGGCTGTTGGTGAAGCTTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1198	2646_ChineseTreeShrewMod_diMut_TA17GT_n1	GGAGGCTGTTGGTGAAGTTTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1199	2647_ChineseTreeShrewMod_diMut_AT18CA_n1	GGAGGCTGTTGGTGAATCATAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1200	2651_ChineseTreeShrewMod_diMut_AT18GC_n1	GGAGGCTGTTGGTGAATGCTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1201	2653_ChineseTreeShrewMod_diMut_AT18TA_n1	GGAGGCTGTTGGTGAATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1202	2654_ChineseTreeShrewMod_diMut_AT18TC_n1	GGAGGCTGTTGGTGAATTTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1203	2664_ChineseTreeShrewMod_diMut_TT19GG_n1	GGAGGCTGTTGGTGAATAGGAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1204	2672_ChineseTreeShrewMod_diMut_TA20GG_n1	GGAGGCTGTTGGTGAATATGACCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1205	2678_ChineseTreeShrewMod_diMut_AA21GG_n1	GGAGGCTGTTGGTGAATATTGGCCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1206	2679_ChineseTreeShrewMod_diMut_AA21GT_n1	GGAGGCTGTTGGTGAATATTGCCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1207	2685_ChineseTreeShrewMod_diMut_AC22CT_n1	GGAGGCTGTTGGTGAATATTACTCAAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1208	2707_ChineseTreeShrewMod_diMut_CA24TC_n1	GGAGGCTGTTGGTGAATATTAACTCAGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1209	2714_ChineseTreeShrewMod_diMut_AA25GG_n1	GGAGGCTGTTGGTGAATATTAAACGGGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1210	2717_ChineseTreeShrewMod_diMut_AA25TG_n1	GGAGGCTGTTGGTGAATATTAAACCTGGGTCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1211	2728_ChineseTreeShrewMod_diMut_GG27AA_n1	GGAGGCTGTTGGTGAATATTAAACCAAAATCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1212	2734_ChineseTreeShrewMod_diMut_GG27TA_n1	GGAGGCTGTTGGTGAATATTAAACCAATATCACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1213	2743_ChineseTreeShrewMod_diMut_GT28TA_n1	GGAGGCTGTTGGTGAATATTAAACCAAGTACACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1214	2748_ChineseTreeShrewMod_diMut_TC29AT_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGATACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1215	2750_ChineseTreeShrewMod_diMut_TC29CG_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1216	2753_ChineseTreeShrewMod_diMut_TC29GG_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGGACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1217	2754_ChineseTreeShrewMod_diMut_TC29GT_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTACCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1218	2756_ChineseTreeShrewMod_diMut_CA30AG_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTAGCCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1219	2758_ChineseTreeShrewMod_diMut_CA30GC_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTGCCCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1220	2759_ChineseTreeShrewMod_diMut_CA30GG_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTGGCCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1221	2763_ChineseTreeShrewMod_diMut_CA30TT_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTTTCCCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1222	2768_ChineseTreeShrewMod_diMut_AC31GG_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTTCGGCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1223	2769_ChineseTreeShrewMod_diMut_AC31GT_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTTCGTCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1224	2770_ChineseTreeShrewMod_diMut_AC31TA_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCTACTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1225	2771_ChineseTreeShrewMod_diMut_AC31TG_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCTGCTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1226	2773_ChineseTreeShrewMod_diMut_CC32AA_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCAATCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC

1227	2774_ChineseTreeShrewMod_diMut_CC32AG_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCAAGTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1228	2778_ChineseTreeShrewMod_diMut_CC32GT_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCAAGTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1229	2779_ChineseTreeShrewMod_diMut_CC32TA_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCAATCAAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1230	2780_ChineseTreeShrewMod_diMut_CC32TG_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCAATGTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1231	2781_ChineseTreeShrewMod_diMut_CC32TT_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCAATTCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1232	2782_ChineseTreeShrewMod_diMut_CT33AA_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCAACAACAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1233	2783_ChineseTreeShrewMod_diMut_CT33AC_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCAACACAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1234	2784_ChineseTreeShrewMod_diMut_CT33AG_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCAACAGAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1235	2785_ChineseTreeShrewMod_diMut_CT33GA_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCAACGACAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1236	2786_ChineseTreeShrewMod_diMut_CT33GC_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCAACGCCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1237	2789_ChineseTreeShrewMod_diMut_CT33TC_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCACTCCAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1238	2791_ChineseTreeShrewMod_diMut_TC34AA_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCAACCAAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1239	2792_ChineseTreeShrewMod_diMut_TC34AG_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCAACAGAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1240	2793_ChineseTreeShrewMod_diMut_TC34AT_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCAACATAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1241	2798_ChineseTreeShrewMod_diMut_TC34GG_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCAACGGAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1242	2799_ChineseTreeShrewMod_diMut_TC34GT_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCAACCTAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1243	2800_ChineseTreeShrewMod_diMut_CA35AC_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCAACCTACGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1244	2801_ChineseTreeShrewMod_diMut_CA35AG_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCAACCTAGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1245	2802_ChineseTreeShrewMod_diMut_CA35AT_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCAACCTATGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1246	2803_ChineseTreeShrewMod_diMut_CA35GC_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCAACCTGCGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1247	2804_ChineseTreeShrewMod_diMut_CA35GG_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCAACCTGGGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1248	2808_ChineseTreeShrewMod_diMut_CA35TT_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCAACCTTTGTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1249	2810_ChineseTreeShrewMod_diMut_AG36CC_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCAACCTCCCTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1250	2811_ChineseTreeShrewMod_diMut_AG36CT_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCAACCTCCTTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1251	2813_ChineseTreeShrewMod_diMut_AG36GC_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCAACCTCGCTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1252	2814_ChineseTreeShrewMod_diMut_AG36GT_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCAACCTCGTTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1253	2815_ChineseTreeShrewMod_diMut_AG36TA_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCAACCTCTATTAACCAAGGAGCAAAACAAGGGCTAAGTCCAC
1254	2816_ChineseTreeShrewMod_diMut_AG36TC_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCAACCTCTCTTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1255	2820_ChineseTreeShrewMod_diMut_GT37AG_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCAACCTCAAGTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1256	2821_ChineseTreeShrewMod_diMut_GT37CA_n1	GGAGGCTGTTGGTGAATATTAACCAAGGTCAACCTCACATATCGGAGGAGCAAAACAAGGGCTAAGTCCAC

1257	2825_ChineseTreeShrewMod_diMut_GI377C_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCATCTATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1258	2831_ChineseTreeShrewMod_diMut_TT38CC_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGCCATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1259	2832_ChineseTreeShrewMod_diMut_TT38CG_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGCGATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1260	2833_ChineseTreeShrewMod_diMut_TT38GA_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGGAAATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1261	2834_ChineseTreeShrewMod_diMut_TT38GC_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGGCAATCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1262	2836_ChineseTreeShrewMod_diMut_TA39AC_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTACTCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1263	2837_ChineseTreeShrewMod_diMut_TA39AG_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTAGTCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1264	2838_ChineseTreeShrewMod_diMut_TA39AT_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTATTTCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1265	2843_ChineseTreeShrewMod_diMut_TA39GG_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTGGTTCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1266	2846_ChineseTreeShrewMod_diMut_AT40CC_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTCCCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1267	2847_ChineseTreeShrewMod_diMut_AT40CG_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTTCGGGAGGAGCAAAACAAGGGCTAAGTCCAC
1268	2848_ChineseTreeShrewMod_diMut_AT40GA_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTGACGGAGGAGCAAAACAAGGGCTAAGTCCAC
1269	2849_ChineseTreeShrewMod_diMut_AT40GC_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTGCCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1270	2851_ChineseTreeShrewMod_diMut_AT40TA_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTTACGGAGGAGCAAAACAAGGGCTAAGTCCAC
1271	2852_ChineseTreeShrewMod_diMut_AT40TC_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTTCCGGAGGAGCAAAACAAGGGCTAAGTCCAC
1272	2855_ChineseTreeShrewMod_diMut_TC41AG_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTAAGGGAGGAGCAAAACAAGGGCTAAGTCCAC
1273	2858_ChineseTreeShrewMod_diMut_TC41CG_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTACGGGAGGAGCAAAACAAGGGCTAAGTCCAC
1274	2862_ChineseTreeShrewMod_diMut_TC41GT_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTAGTGGAGGAGCAAAACAAGGGCTAAGTCCAC
1275	2864_ChineseTreeShrewMod_diMut_CG42AC_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATACGAGGAGCAAAACAAGGGCTAAGTCCAC
1276	2866_ChineseTreeShrewMod_diMut_CG42GA_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATGAGAGGAGCAAAACAAGGGCTAAGTCCAC
1277	2867_ChineseTreeShrewMod_diMut_CG42GC_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATGCGAGGAGCAAAACAAGGGCTAAGTCCAC
1278	2868_ChineseTreeShrewMod_diMut_CG42GT_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATGAGGAGCAAAACAAGGGCTAAGTCCAC
1279	2869_ChineseTreeShrewMod_diMut_CG42TA_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATTAGAGGAGCAAAACAAGGGCTAAGTCCAC
1280	2870_ChineseTreeShrewMod_diMut_CG42TC_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATTCGAGGAGCAAAACAAGGGCTAAGTCCAC
1281	2872_ChineseTreeShrewMod_diMut_GG43AA_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCAAAAGGAGCAAAACAAGGGCTAAGTCCAC
1282	2873_ChineseTreeShrewMod_diMut_GG43AC_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCACAGGAGCAAAACAAGGGCTAAGTCCAC
1283	2874_ChineseTreeShrewMod_diMut_GG43AT_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCATAGGAGCAAAACAAGGGCTAAGTCCAC
1284	2879_ChineseTreeShrewMod_diMut_GG43TC_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCTCAGGAGCAAAACAAGGGCTAAGTCCAC
1285	2880_ChineseTreeShrewMod_diMut_GG43TT_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCTTTAGGAGCAAAACAAGGGCTAAGTCCAC
1286	2882_ChineseTreeShrewMod_diMut_GA44AG_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATTCGAGGAGCAAAACAAGGGCTAAGTCCAC

1287	2884_ChineseTreeShrewMod_diMut_GA44CC_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGCCGGAGCAAAACAAGGGCTAAGTCCAC
1288	2886_ChineseTreeShrewMod_diMut_GA44CT_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGCTGAGCAAAACAAGGGCTAAGTCCAC
1289	2890_ChineseTreeShrewMod_diMut_AG45CA_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGCGAGCAAAACAAGGGCTAAGTCCAC
1290	2891_ChineseTreeShrewMod_diMut_AG45CC_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGCCGAGCAAAACAAGGGCTAAGTCCAC
1291	2892_ChineseTreeShrewMod_diMut_AG45CT_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGCTGAGCAAAACAAGGGCTAAGTCCAC
1292	2893_ChineseTreeShrewMod_diMut_AG45GA_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGGGAGCAAAACAAGGGCTAAGTCCAC
1293	2894_ChineseTreeShrewMod_diMut_AG45GC_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGGGAGCAAAACAAGGGCTAAGTCCAC
1294	2895_ChineseTreeShrewMod_diMut_AG45GT_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGGTGAACAAACAAGGGCTAAGTCCAC
1295	2896_ChineseTreeShrewMod_diMut_AG45TA_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGTAGCAAAACAAGGGCTAAGTCCAC
1296	2897_ChineseTreeShrewMod_diMut_AG45TC_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGTGAGCAAAACAAGGGCTAAGTCCAC
1297	2898_ChineseTreeShrewMod_diMut_AG45TT_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGTTGAGCAAAACAAGGGCTAAGTCCAC
1298	2899_ChineseTreeShrewMod_diMut_GG46AA_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAAAAGCAAAACAAGGGCTAAGTCCAC
1299	2900_ChineseTreeShrewMod_diMut_GG46AC_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAAAGCAAAACAAGGGCTAAGTCCAC
1300	2901_ChineseTreeShrewMod_diMut_GG46AT_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAAAGCAAAACAAGGGCTAAGTCCAC
1301	2902_ChineseTreeShrewMod_diMut_GG46CA_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGACAAAGCAAAACAAGGGCTAAGTCCAC
1302	2903_ChineseTreeShrewMod_diMut_GG46CC_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGACAGCAAAACAAGGGCTAAGTCCAC
1303	2905_ChineseTreeShrewMod_diMut_GG46TA_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGATAAGCAAAACAAGGGCTAAGTCCAC
1304	2906_ChineseTreeShrewMod_diMut_GG46TC_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGATAGCAAAACAAGGGCTAAGTCCAC
1305	2908_ChineseTreeShrewMod_diMut_GA47AC_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGAGCAAAACAAGGGCTAAGTCCAC
1306	2909_ChineseTreeShrewMod_diMut_GA47AG_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGAGGCAAAACAAGGGCTAAGTCCAC
1307	2911_ChineseTreeShrewMod_diMut_GA47CC_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGCGGCAAAACAAGGGCTAAGTCCAC
1308	2912_ChineseTreeShrewMod_diMut_GA47CG_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGGCAAAACAAGGGCTAAGTCCAC
1309	2915_ChineseTreeShrewMod_diMut_GA47TG_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGTGGCAAAACAAGGGCTAAGTCCAC
1310	2917_ChineseTreeShrewMod_diMut_AG48CA_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGCAAAACAAGGGCTAAGTCCAC
1311	2919_ChineseTreeShrewMod_diMut_AG48CT_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGCTCAAAACAAGGGCTAAGTCCAC
1312	2921_ChineseTreeShrewMod_diMut_AG48GC_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGGCCAAACAAGGGCTAAGTCCAC
1313	2922_ChineseTreeShrewMod_diMut_AG48GT_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGGTCAAAACAAGGGCTAAGTCCAC
1314	2923_ChineseTreeShrewMod_diMut_AG48TA_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGTACAAACAAGGGCTAAGTCCAC
1315	2924_ChineseTreeShrewMod_diMut_AG48TC_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGTCCAAACAAGGGCTAAGTCCAC
1316	2928_ChineseTreeShrewMod_diMut_GC49AT_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAAATAACAAGGGCTAAGTCCAC

1317	2931_ChineseTreeShrewMod_diMut_GC49CT_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCTAAACAAGGGCTAAGTCCAC
1318	2934_ChineseTreeShrewMod_diMut_GC49TT_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGATTAACAAGGGCTAAGTCCAC
1319	2940_ChineseTreeShrewMod_diMut_CA50GT_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGGTTAAACAAGGGCTAAGTCCAC
1320	2959_ChineseTreeShrewMod_diMut_AA52TC_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGGATCAACAAGGGCTAAGTCCAC
1321	2973_ChineseTreeShrewMod_diMut_CA54AT_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAAATAGGGCTAAGTCCAC
1322	2977_ChineseTreeShrewMod_diMut_CA54TC_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAAATCAGGGCTAAGTCCAC
1323	2978_ChineseTreeShrewMod_diMut_CA54TG_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAAATGAGGGCTAAGTCCAC
1324	2995_ChineseTreeShrewMod_diMut_AG56TA_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAAATAGGCTAAGTCCAC
1325	3013_ChineseTreeShrewMod_diMut_GG58TA_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAAATAGTACTAAGTCCAC
1326	3014_ChineseTreeShrewMod_diMut_GG58TC_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAAATAGTCTAAGTCCAC
1327	3015_ChineseTreeShrewMod_diMut_GG58TT_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAAATAGTTCTAAGTCCAC
1328	3025_ChineseTreeShrewMod_diMut_CT60AA_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAAATAGGGGAAAAGTCCAC
1329	3028_ChineseTreeShrewMod_diMut_CT60GA_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAAATAGGGGAAAAGTCCAC
1330	3031_ChineseTreeShrewMod_diMut_CT60TA_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAAATAGGGTAAAAGTCCAC
1331	3036_ChineseTreeShrewMod_diMut_TA61AT_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAAATAGGGGATAGTCCAC
1332	3073_ChineseTreeShrewMod_diMut_IC65CA_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAAATAGGGGTAAGCAGCAC
1333	3078_ChineseTreeShrewMod_diMut_IC65GT_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAAATAGGGGTAAGGTCCAC
1334	3079_ChineseTreeShrewMod_diMut_CC66AA_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAAATAGGGGTAAGTAAAC
1335	3090_ChineseTreeShrewMod_diMut_CA67AT_n1	GGAGGCTGTTGGTGAATATTAAACCAAGGTCACCTCAGTTATCGGAGGAGCAAAAATAGGGGTAAGTCAATC

Table 12. Single and adjacent di-nucleotide variants of BushBaby SERPINA1 enhancer with higher luciferase expression than original sequence SEQ ID NO:

83

SEQ ID NO:	Bushbaby SERPINA1 enhancer variant	Sequence
1336	1256_Bushbaby_monoMut_G1A_n1	AGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGGCAAAAACAGGAGCTAAGTCCAT
1337	1262_Bushbaby_monoMut_G3A_n1	GGAGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGGCAAAAACAGGAGCTAAGTCCAT
1338	1263_Bushbaby_monoMut_G3C_n1	GGCGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGGCAAAAACAGGAGCTAAGTCCAT
1339	1270_Bushbaby_monoMut_G5T_n1	GGGGTAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGGCAAAAACAGGAGCTAAGTCCAT
1340	1273_Bushbaby_monoMut_A6T_n1	GGGGTAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGGCAAAAACAGGAGCTAAGTCCAT
1341	1277_Bushbaby_monoMut_G8A_n1	GGGGGAACTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGGCAAAAACAGGAGCTAAGTCCAT

1342	1282_Bushbaby_monoMut_C9T_n1	GGGGAAAGTTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAAGTCCAT
1343	1284_Bushbaby_monoMut_T10C_n1	GGGGAAAGCCACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAAGTCCAT
1344	1286_Bushbaby_monoMut_A11C_n1	GGGGAAAGCTCCTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAAGTCCAT
1345	1287_Bushbaby_monoMut_A11G_n1	GGGGAAAGCTGCTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAAGTCCAT
1346	1288_Bushbaby_monoMut_A11T_n1	GGGGAAAGCTTCTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAAGTCCAT
1347	1294_Bushbaby_monoMut_T13G_n1	GGGGAAAGCTACGGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAAGTCCAT
1348	1300_Bushbaby_monoMut_G15T_n1	GGGGAAAGCTACTGTTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAAGTCCAT
1349	1306_Bushbaby_monoMut_G17T_n1	GGGGAAAGCTACTGGTTAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAAGTCCAT
1350	1310_Bushbaby_monoMut_A19C_n1	GGGGAAAGCTACTGGTGACTATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAAGTCCAT
1351	1311_Bushbaby_monoMut_A19G_n1	GGGGAAAGCTACTGGTGAGTATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAAGTCCAT
1352	1324_Bushbaby_monoMut_T23G_n1	GGGGAAAGCTACTGGTGAATATGAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAAGTCCAT
1353	1330_Bushbaby_monoMut_A25T_n1	GGGGAAAGCTACTGGTGAATATTATCCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAAGTCCAT
1354	1352_Bushbaby_monoMut_C33A_n1	GGGGAAAGCTACTGGTGAATATTAAACCAAGGTAACCCAGTTATCAGGGAGCAAAACAGGAGCTAAAGTCCAT
1355	1359_Bushbaby_monoMut_C35G_n1	GGGGAAAGCTACTGGTGAATATTAAACCAAGGTCAGCCAGTTATCAGGGAGCAAAACAGGAGCTAAAGTCCAT
1356	1360_Bushbaby_monoMut_C35T_n1	GGGGAAAGCTACTGGTGAATATTAAACCAAGGTCATCCAGTTATCAGGGAGCAAAACAGGAGCTAAAGTCCAT
1357	1361_Bushbaby_monoMut_C36A_n1	GGGGAAAGCTACTGGTGAATATTAAACCAAGGTCACACAGTTATCAGGGAGCAAAACAGGAGCTAAAGTCCAT
1358	1362_Bushbaby_monoMut_C36G_n1	GGGGAAAGCTACTGGTGAATATTAAACCAAGGTCACGAGTTATCAGGGAGCAAAACAGGAGCTAAAGTCCAT
1359	1363_Bushbaby_monoMut_C36T_n1	GGGGAAAGCTACTGGTGAATATTAAACCAAGGTCACTCAGTTATCAGGGAGCAAAACAGGAGCTAAAGTCCAT
1360	1365_Bushbaby_monoMut_C37G_n1	GGGGAAAGCTACTGGTGAATATTAAACCAAGGTCACGAGTTATCAGGGAGCAAAACAGGAGCTAAAGTCCAT
1361	1367_Bushbaby_monoMut_A38C_n1	GGGGAAAGCTACTGGTGAATATTAAACCAAGGTAACCCCGTTATCAGGGAGCAAAACAGGAGCTAAAGTCCAT
1362	1368_Bushbaby_monoMut_A38G_n1	GGGGAAAGCTACTGGTGAATATTAAACCAAGGTCACCGGTTATCAGGGAGCAAAACAGGAGCTAAAGTCCAT
1363	1372_Bushbaby_monoMut_G39T_n1	GGGGAAAGCTACTGGTGAATATTAAACCAAGGTCACCCATTATCAGGGAGCAAAACAGGAGCTAAAGTCCAT
1364	1375_Bushbaby_monoMut_T40G_n1	GGGGAAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGGTATCAGGGAGCAAAACAGGAGCTAAAGTCCAT
1365	1380_Bushbaby_monoMut_A42G_n1	GGGGAAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTGTACCCAGTTGTACGGAGCAAAACAGGAGCTAAAGTCCAT
1366	1383_Bushbaby_monoMut_T43C_n1	GGGGAAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTACCCAGTTACCCAGTTACCGAGGAGCAAAACAGGAGCTAAAGTCCAT
1367	1384_Bushbaby_monoMut_T43G_n1	GGGGAAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTACCCAGTTAGCAGGGAGCAAAACAGGAGCTAAAGTCCAT
1368	1390_Bushbaby_monoMut_A45T_n1	GGGGAAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCTGGAGCAAAACAGGAGCTAAAGTCCAT
1369	1393_Bushbaby_monoMut_G46T_n1	GGGGAAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCATCGGAGCAAAACAGGAGCTAAAGTCCAT
1370	1394_Bushbaby_monoMut_G47A_n1	GGGGAAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGAGAGCAAAACAGGAGCTAAAGTCCAT
1371	1396_Bushbaby_monoMut_G47T_n1	GGGGAAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCTAGTCAGCAAAACAGGAGCTAAAGTCCAT

1372	1397_Bushbaby_monoMut_G48A_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGAAGCAAAACAGGAGCTAAGTCCAT
1373	1402_Bushbaby_monoMut_A49T_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGTGCAAAACAGGAGCTAAGTCCAT
1374	1405_Bushbaby_monoMut_G50T_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGATCAAAACAGGAGCTAAGTCCAT
1375	1411_Bushbaby_monoMut_A52T_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCTAAACAGGAGCTAAGTCCAT
1376	1413_Bushbaby_monoMut_A53G_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAGACAGGAGCTAAGTCCAT
1377	1424_Bushbaby_monoMut_G57A_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAAAGAGCTAAGTCCAT
1378	1431_Bushbaby_monoMut_A59G_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGGGCTAAGTCCAT
1379	1432_Bushbaby_monoMut_A59T_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGTGCTAAGTCCAT
1380	1433_Bushbaby_monoMut_G60A_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAACCTAAGTCCAT
1381	1439_Bushbaby_monoMut_T62A_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCAAAAGTCCAT
1382	1441_Bushbaby_monoMut_T62G_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCAAAAGTCCAT
1383	1443_Bushbaby_monoMut_A63G_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1384	1457_Bushbaby_monoMut_C68A_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCAAT
1385	1463_Bushbaby_monoMut_T70A_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAA
1386	1464_Bushbaby_monoMut_T70C_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAC
1387	1467_Bushbaby_diMut_GG1AC_n1	ACGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1388	1470_Bushbaby_diMut_GG1CC_n1	CCGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1389	1476_Bushbaby_diMut_GG2AC_n1	GACGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1390	1478_Bushbaby_diMut_GG2CA_n1	GCAGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1391	1479_Bushbaby_diMut_GG2CC_n1	GCCGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1392	1482_Bushbaby_diMut_GG2TC_n1	GTCCGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1393	1484_Bushbaby_diMut_GG3AA_n1	GGAAGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1394	1495_Bushbaby_diMut_GG4AT_n1	GGGATAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1395	1497_Bushbaby_diMut_GG4CC_n1	GGGCCAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1396	1506_Bushbaby_diMut_GA5CG_n1	GGGGCGAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1397	1507_Bushbaby_diMut_GA5CT_n1	GGGGCTAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1398	1515_Bushbaby_diMut_AA6GG_n1	GGGGGGGGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1399	1516_Bushbaby_diMut_AA6GT_n1	GGGGGGTGTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1400	1519_Bushbaby_diMut_AA6TT_n1	GGGGGTTGTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1401	1521_Bushbaby_diMut_AG7CC_n1	GGGGGACCCCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT

1402	1523_Bushbaby_diMut_AG7GA_n1	GGGGGAGACTACTGGTGAATATTAAACCAAGGTACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1403	1525_Bushbaby_diMut_AG7GT_n1	GGGGGAGTCTACTGGTGAATATTAAACCAAGGTACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1404	1529_Bushbaby_diMut_GC8AA_n1	GGGGGAAAATACTGGTGAATATTAAACCAAGGTACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1405	1531_Bushbaby_diMut_GC8AT_n1	GGGGGAAAATACTGGTGAATATTAAACCAAGGTACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1406	1533_Bushbaby_diMut_GC8CG_n1	GGGGGAACGTAAGTGGTGAATATTAAACCAAGGTACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1407	1535_Bushbaby_diMut_GC8TA_n1	GGGGGAATATACTGGTGAATATTAAACCAAGGTACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1408	1536_Bushbaby_diMut_GC8TG_n1	GGGGGAATGTAAGTGGTGAATATTAAACCAAGGTACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1409	1537_Bushbaby_diMut_GC8TT_n1	GGGGGAATTTACTGGTGAATATTAAACCAAGGTACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1410	1539_Bushbaby_diMut_CT9AC_n1	GGGGGAAGACACTGGTGAATATTAAACCAAGGTACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1411	1540_Bushbaby_diMut_CT9AG_n1	GGGGGAAGAGACTGGTGAATATTAAACCAAGGTACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1412	1541_Bushbaby_diMut_CT9GA_n1	GGGGGAAGGAAGTGGTGAATATTAAACCAAGGTACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1413	1542_Bushbaby_diMut_CT9GC_n1	GGGGGAAGGCACTGGTGAATATTAAACCAAGGTACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1414	1543_Bushbaby_diMut_CT9GG_n1	GGGGGAAGGACTGGTGAATATTAAACCAAGGTACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1415	1544_Bushbaby_diMut_CT9TA_n1	GGGGGAAGTAAGTGGTGAATATTAAACCAAGGTACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1416	1545_Bushbaby_diMut_CT9TC_n1	GGGGGAAGTCACTGGTGAATATTAAACCAAGGTACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1417	1546_Bushbaby_diMut_CT9TG_n1	GGGGGAAGTGAAGTGGTGAATATTAAACCAAGGTACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1418	1550_Bushbaby_diMut_TA10CC_n1	GGGGGAAGCCCTGGTGAATATTAAACCAAGGTACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1419	1551_Bushbaby_diMut_TA10CG_n1	GGGGGAAGCCGCTGGTGAATATTAAACCAAGGTACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1420	1553_Bushbaby_diMut_TA10GC_n1	GGGGGAAGCGCTGGTGAATATTAAACCAAGGTACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1421	1555_Bushbaby_diMut_TA10GT_n1	GGGGGAAGCGTCTGGTGAATATTAAACCAAGGTACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1422	1563_Bushbaby_diMut_AC11TG_n1	GGGGGAAGCTTGTGGTGAATATTAAACCAAGGTACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1423	1564_Bushbaby_diMut_AC11TT_n1	GGGGGAAGCTTTTGGTGAATATTAAACCAAGGTACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1424	1567_Bushbaby_diMut_CT12AG_n1	GGGGGAAGCTAAGGTGAATATTAAACCAAGGTACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1425	1573_Bushbaby_diMut_CT12TG_n1	GGGGGAAGCTATGGTGAATATTAAACCAAGGTACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1426	1574_Bushbaby_diMut_TG13AA_n1	GGGGGAAGCTACAAGTGAATATTAAACCAAGGTACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1427	1585_Bushbaby_diMut_GG14AT_n1	GGGGGAAGCTACTATTGAATATTAAACCAAGGTACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1428	1593_Bushbaby_diMut_GT15AC_n1	GGGGGAAGCTACTGACGAATATTAAACCAAGGTACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1429	1598_Bushbaby_diMut_GT15TA_n1	GGGGGAAGCTACTGTAGAATATTAAACCAAGGTACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1430	1600_Bushbaby_diMut_GT15TG_n1	GGGGGAAGCTACTGTGGAATATTAAACCAAGGTACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1431	1601_Bushbaby_diMut_TG16AA_n1	GGGGGAAGCTACTGGAAAAATATTAAACCAAGGTACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT

1432	1617_Bushbaby_diMut_GA17TG_n1	GGGGGAAGCTACTGTTGATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1433	1622_Bushbaby_diMut_AA18GC_n1	GGGGGAAGCTACTGGTGGCTATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1434	1629_Bushbaby_diMut_AT19CC_n1	GGGGGAAGCTACTGGTGGTGAACCTTAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1435	1642_Bushbaby_diMut_TA20CT_n1	GGGGGAAGCTACTGGTGGTGAACCTTAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1436	1645_Bushbaby_diMut_TA20GT_n1	GGGGGAAGCTACTGGTGGTGAAGTTTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1437	1647_Bushbaby_diMut_AT21CC_n1	GGGGGAAGCTACTGGTGAATCCTTAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1438	1648_Bushbaby_diMut_AT21CG_n1	GGGGGAAGCTACTGGTGAATCGTAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1439	1652_Bushbaby_diMut_AT21TA_n1	GGGGGAAGCTACTGGTGAATTTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1440	1653_Bushbaby_diMut_AT21TC_n1	GGGGGAAGCTACTGGTGAATTTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1441	1669_Bushbaby_diMut_TA23CT_n1	GGGGGAAGCTACTGGTGAATATCTACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1442	1673_Bushbaby_diMut_AA24CC_n1	GGGGGAAGCTACTGGTGAATATTTCCCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1443	1681_Bushbaby_diMut_AA24TT_n1	GGGGGAAGCTACTGGTGAATATTTCCCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1444	1684_Bushbaby_diMut_AC25CT_n1	GGGGGAAGCTACTGGTGAATATTACTCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1445	1686_Bushbaby_diMut_AC25GG_n1	GGGGGAAGCTACTGGTGAATATTAGGCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1446	1690_Bushbaby_diMut_AC25TT_n1	GGGGGAAGCTACTGGTGAATATTCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1447	1716_Bushbaby_diMut_AA28TG_n1	GGGGGAAGCTACTGGTGAATATTAACTGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1448	1727_Bushbaby_diMut_GG30AA_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAAATCACCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1449	1747_Bushbaby_diMut_TC32AT_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGATACCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1450	1758_Bushbaby_diMut_CA33GG_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTGGCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1451	1761_Bushbaby_diMut_CA33TG_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTTGCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1452	1768_Bushbaby_diMut_AC34GT_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCGTCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1453	1772_Bushbaby_diMut_CC35AA_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACACAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1454	1780_Bushbaby_diMut_CC35TT_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCATTCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1455	1785_Bushbaby_diMut_CC36GG_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACGGAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1456	1786_Bushbaby_diMut_CC36GT_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACGTAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1457	1787_Bushbaby_diMut_CC36TA_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCCTAAGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1458	1792_Bushbaby_diMut_CA37AT_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCATGTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1459	1798_Bushbaby_diMut_CA37TT_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCTTGTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1460	1800_Bushbaby_diMut_AG38CC_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCCTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1461	1802_Bushbaby_diMut_AG38GA_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCGATTATCAGGGAGCAAAACAGGAGCTAAGTCCAT

1462	1807_Bushbaby_diMut_AG38TT_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCCTTTTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1463	1809_Bushbaby_diMut_GT39AC_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCAACTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1464	1810_Bushbaby_diMut_GT39AG_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCAAGTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1465	1813_Bushbaby_diMut_GT39CG_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCAACGTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1466	1816_Bushbaby_diMut_GT39TG_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCAATGTATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1467	1818_Bushbaby_diMut TT40AC_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCAAGACATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1468	1822_Bushbaby_diMut TT40CG_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCAAGCGATCAGGGAGCAAAACAGGAGCTAAGTCCAT
1469	1833_Bushbaby_diMut TA41GG_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCAAGTGGTCAGGGAGCAAAACAGGAGCTAAGTCCAT
1470	1835_Bushbaby_diMut_AT42CA_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCAAGTTCACAGGGAGCAAAACAGGAGCTAAGTCCAT
1471	1837_Bushbaby_diMut_AT42CG_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCAAGTTCGCAGGGAGCAAAACAGGAGCTAAGTCCAT
1472	1839_Bushbaby_diMut_AT42GC_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCAAGTTCGCAGGGAGCAAAACAGGAGCTAAGTCCAT
1473	1849_Bushbaby_diMut_TC43CT_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCAAGTACTAGGGAGCAAAACAGGAGCTAAGTCCAT
1474	1851_Bushbaby_diMut_TC43GG_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCAAGTACTAGGGAGCAAAACAGGAGCTAAGTCCAT
1475	1856_Bushbaby_diMut_CA44GC_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCAAGTATATGCGGGAGCAAAACAGGAGCTAAGTCCAT
1476	1871_Bushbaby_diMut_GG46AA_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCAAGTATATCAAAAGAGCAAAACAGGAGCTAAGTCCAT
1477	1873_Bushbaby_diMut_GG46AT_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCAAGTATATCAATGAGCAAAACAGGAGCTAAGTCCAT
1478	1874_Bushbaby_diMut_GG46CA_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCAAGTATATCACAGAGCAAAACAGGAGCTAAGTCCAT
1479	1875_Bushbaby_diMut_GG46CC_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCAAGTATATCACCGAGCAAAACAGGAGCTAAGTCCAT
1480	1877_Bushbaby_diMut_GG46TA_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCAAGTATATCATAGAGCAAAACAGGAGCTAAGTCCAT
1481	1879_Bushbaby_diMut_GG46TT_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCAAGTATATCATTTGAGCAAAACAGGAGCTAAGTCCAT
1482	1880_Bushbaby_diMut_GG47AA_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCAAGTATATCAGAAAAGCAAAACAGGAGCTAAGTCCAT
1483	1881_Bushbaby_diMut_GG47AC_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCAAGTATATCAGACAGCAAAACAGGAGCTAAGTCCAT
1484	1882_Bushbaby_diMut_GG47AT_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCAAGTATATCAGATAGCAAAACAGGAGCTAAGTCCAT
1485	1883_Bushbaby_diMut_GG47CA_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCAAGTATATCAGCAAGCAAAACAGGAGCTAAGTCCAT
1486	1884_Bushbaby_diMut_GG47CC_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCAAGTATATCAGCCAGCAAAACAGGAGCTAAGTCCAT
1487	1888_Bushbaby_diMut_GG47TT_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCAAGTATATCAGTTAGCAAAACAGGAGCTAAGTCCAT
1488	1890_Bushbaby_diMut_GA48AG_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCAAGTATATCAGGAGCAAAACAGGAGCTAAGTCCAT
1489	1909_Bushbaby_diMut_GC50AT_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCAAGTATATCAGGGAGTAATAACAGGAGCTAAGTCCAT
1490	1918_Bushbaby_diMut_CA51AT_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCAAGTATATCAGGGAGATAACAGGAGCTAAGTCCAT
1491	1953_Bushbaby_diMut_CA55AG_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCAAGTATATCAGGGAGCAAAAGGAGCTAAGTCCAT

1492	1959_Bushbaby_diMut_CA55TG_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAATGGGAGCTAAGTCCAT
1493	1988_Bushbaby_diMut_AG59CA_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGCCTAAGTCCAT
1494	1992_Bushbaby_diMut_AG59GC_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGGCCTAAGTCCAT
1495	1993_Bushbaby_diMut_AG59GT_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGGTCTAAGTCCAT
1496	1994_Bushbaby_diMut_AG59TA_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGTACTAAGTCCAT
1497	1995_Bushbaby_diMut_AG59TC_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGTCTTAAGTCCAT
1498	2004_Bushbaby_diMut_GC60TG_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGATGTAAGTCCAT
1499	2012_Bushbaby_diMut_CT61TA_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGTAAAGTCCAT
1500	2016_Bushbaby_diMut_TA62AG_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCAGAGTCCAT
1501	2017_Bushbaby_diMut_TA62AT_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCATAGTCCAT
1502	2022_Bushbaby_diMut_TA62GG_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGGGAGTCCAT
1503	2053_Bushbaby_diMut_TC66AT_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGATCAT
1504	2056_Bushbaby_diMut_TC66CT_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTCTCAT
1505	2059_Bushbaby_diMut_TC66GT_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGGTTCAT
1506	2060_Bushbaby_diMut_CC67AA_n1	GGGGGAAGCTACTGGTGAATATTAAACCAAGGTCACCCAGTTATCAGGGAGCAAAACAGGAGCTAAGTAAAT

CTAAG -> CAAAG single nucleotide substitution variant had the highest expression.

Table 13. Single and adjacent di-nucleotide substitution variants of human SERPINA1 enhancer with HNF4 and FOXA transcription factor consensus sites with higher luciferase expression than original sequence SEQ ID NO: 85

SEQ ID NO:	HNF4_FOXA variant	SERPINA1 enhancer	Sequence
1507	3310_HNF4_FOXA_monoMut_G2C_n1		GCGGGAGGCTGCTGGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1508	3312_HNF4_FOXA_monoMut_G3A_n1		GGAGGAGGCTGCTGGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1509	3315_HNF4_FOXA_monoMut_G4A_n1		GGGAGAGGCTGCTGGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1510	3316_HNF4_FOXA_monoMut_G4C_n1		GGGCGAGGCTGCTGGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1511	3333_HNF4_FOXA_monoMut_T10A_n1		GGGGGAGGCGAGCTGGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC

1512	3334_HNF4_FOXA_monoMut_T10C_n1	GGGGAGGCGCTGGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAACACAGGGGCAAAAGTCCAC
1513	3341_HNF4_FOXA_monoMut_C12T_n1	GGGGAGGCTGTTGGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAACACAGGGGCAAAAGTCCAC
1514	3342_HNF4_FOXA_monoMut_T13A_n1	GGGGAGGCTGCAGGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAACACAGGGGCAAAAGTCCAC
1515	3343_HNF4_FOXA_monoMut_T13C_n1	GGGGAGGCTGCCGGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAACACAGGGGCAAAAGTCCAC
1516	3345_HNF4_FOXA_monoMut_G14A_n1	GGGGAGGCTGCTAGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAACACAGGGGCAAAAGTCCAC
1517	3362_HNF4_FOXA_monoMut_A19T_n1	GGGGAGGCTGCTGGTAAATCATTAACCAAGGTCACCCAGTTATCAGAGGAGCAACACAGGGGCAAAAGTCCAC
1518	3369_HNF4_FOXA_monoMut_T22A_n1	GGGGAGGCTGCTGGTAAACAATAACCAAGGTCACCCAGTTATCAGAGGAGCAACACAGGGGCAAAAGTCCAC
1519	3373_HNF4_FOXA_monoMut_T23C_n1	GGGGAGGCTGCTGGTAAACATCAACCAAGGTCACCCAGTTATCAGAGGAGCAACACAGGGGCAAAAGTCCAC
1520	3375_HNF4_FOXA_monoMut_A24C_n1	GGGGAGGCTGCTGGTAAACATTACCAAGGTCACCCAGTTATCAGAGGAGCAACACAGGGGCAAAAGTCCAC
1521	3376_HNF4_FOXA_monoMut_A24G_n1	GGGGAGGCTGCTGGTAAACATTGACCAAGGTCACCCAGTTATCAGAGGAGCAACACAGGGGCAAAAGTCCAC
1522	3379_HNF4_FOXA_monoMut_A25G_n1	GGGGAGGCTGCTGGTAAACATTAGCCAAGGTCACCCAGTTATCAGAGGAGCAACACAGGGGCAAAAGTCCAC
1523	3383_HNF4_FOXA_monoMut_C26T_n1	GGGGAGGCTGCTGGTAAACATTAACTCAAGGTCACCCAGTTATCAGAGGAGCAACACAGGGGCAAAAGTCCAC
1524	3386_HNF4_FOXA_monoMut_C27T_n1	GGGGAGGCTGCTGGTAAACATTAACTAAGGTCACCCAGTTATCAGAGGAGCAACACAGGGGCAAAAGTCCAC
1525	3408_HNF4_FOXA_monoMut_C35A_n1	GGGGAGGCTGCTGGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAACACAGGGGCAAAAGTCCAC
1526	3409_HNF4_FOXA_monoMut_C35G_n1	GGGGAGGCTGCTGGTAAACATTAAACCAAGGTCAGCCAGTTATCAGAGGAGCAACACAGGGGCAAAAGTCCAC
1527	3410_HNF4_FOXA_monoMut_C35T_n1	GGGGAGGCTGCTGGTAAACATTAAACCAAGGTCATCCAGTTATCAGAGGAGCAACACAGGGGCAAAAGTCCAC
1528	3413_HNF4_FOXA_monoMut_C36T_n1	GGGGAGGCTGCTGGTAAACATTAAACCAAGGTCCTCCAGTTATCAGAGGAGCAACACAGGGGCAAAAGTCCAC
1529	3415_HNF4_FOXA_monoMut_C37G_n1	GGGGAGGCTGCTGGTAAACATTAAACCAAGGTCACCGCAGTTATCAGAGGAGCAACACAGGGGCAAAAGTCCAC
1530	3416_HNF4_FOXA_monoMut_C37T_n1	GGGGAGGCTGCTGGTAAACATTAAACCAAGGTCACCTCAGTTATCAGAGGAGCAACACAGGGGCAAAAGTCCAC
1531	3418_HNF4_FOXA_monoMut_C38G_n1	GGGGAGGCTGCTGGTAAACATTAAACCAAGGTCACCCGAGTTATCAGAGGAGCAACACAGGGGCAAAAGTCCAC
1532	3420_HNF4_FOXA_monoMut_A39C_n1	GGGGAGGCTGCTGGTAAACATTAAACCAAGGTCACCCCGTTATCAGAGGAGCAACACAGGGGCAAAAGTCCAC
1533	3421_HNF4_FOXA_monoMut_A39G_n1	GGGGAGGCTGCTGGTAAACATTAAACCAAGGTCACCCCGTTATCAGAGGAGCAACACAGGGGCAAAAGTCCAC
1534	3422_HNF4_FOXA_monoMut_A39T_n1	GGGGAGGCTGCTGGTAAACATTAAACCAAGGTCACCCCTGTTATCAGAGGAGCAACACAGGGGCAAAAGTCCAC
1535	3423_HNF4_FOXA_monoMut_G40A_n1	GGGGAGGCTGCTGGTAAACATTAAACCAAGGTCACCCCAATTATCAGAGGAGCAACACAGGGGCAAAAGTCCAC
1536	3427_HNF4_FOXA_monoMut_T41C_n1	GGGGAGGCTGCTGGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAACACAGGGGCAAAAGTCCAC
1537	3440_HNF4_FOXA_monoMut_C45T_n1	GGGGAGGCTGCTGGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAACACAGGGGCAAAAGTCCAC
1538	3445_HNF4_FOXA_monoMut_G47C_n1	GGGGAGGCTGCTGGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAACACAGGGGCAAAAGTCCAC
1539	3446_HNF4_FOXA_monoMut_G47T_n1	GGGGAGGCTGCTGGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAACACAGGGGCAAAAGTCCAC
1540	3447_HNF4_FOXA_monoMut_A48C_n1	GGGGAGGCTGCTGGTAAACATTAAACCAAGGTCACCCAGTTATCAGCGGAGCAACACAGGGGCAAAAGTCCAC
1541	3453_HNF4_FOXA_monoMut_G50A_n1	GGGGAGGCTGCTGGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGAAACACAGGGGCAAAAGTCCAC

1542	3456_HNF4_FOXA_monoMut_A51C_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGGCGCAACACAGGGGCAAAAGTCCAC
1543	3459_HNF4_FOXA_monoMut_G52A_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGAACAAACAGGGGCAAAAGTCCAC
1544	3460_HNF4_FOXA_monoMut_G52C_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGACCAAAACAGGGGCAAAAGTCCAC
1545	3461_HNF4_FOXA_monoMut_G52I_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGATCAAAACAGGGGCAAAAGTCCAC
1546	3464_HNF4_FOXA_monoMut_C53T_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGAGTAAACAGGGGCAAAAGTCCAC
1547	3467_HNF4_FOXA_monoMut_A54T_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGAGCTAACAGGGGCAAAAGTCCAC
1548	3478_HNF4_FOXA_monoMut_A58G_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1549	3488_HNF4_FOXA_monoMut_G61T_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGTGCAAAAGTCCAC
1550	3512_HNF4_FOXA_monoMut_C69T_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTTCAC
1551	3519_HNF4_FOXA_monoMut_C72A_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAA
1552	3521_HNF4_FOXA_monoMut_C72T_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAT
1553	3523_HNF4_FOXA_diMut_GG1AC_n1	ACGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1554	3526_HNF4_FOXA_diMut_GG1CC_n1	CCGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1555	3532_HNF4_FOXA_diMut_GG2AC_n1	GACGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1556	3535_HNF4_FOXA_diMut_GG2CC_n1	GCCGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1557	3539_HNF4_FOXA_diMut_GG2TT_n1	GTGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1558	3540_HNF4_FOXA_diMut_GG3AA_n1	GGAAGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1559	3542_HNF4_FOXA_diMut_GG3AT_n1	GGATGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1560	3545_HNF4_FOXA_diMut_GG3CT_n1	GGCTGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1561	3546_HNF4_FOXA_diMut_GG3TA_n1	GGTAGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1562	3550_HNF4_FOXA_diMut_GG4AC_n1	GGGACAGGCTGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1563	3552_HNF4_FOXA_diMut_GG4CA_n1	GGGCAAGGCTGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1564	3560_HNF4_FOXA_diMut_GA5AT_n1	GGGGATGGCTGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1565	3561_HNF4_FOXA_diMut_GA5CC_n1	GGGGCCGGCTGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1566	3562_HNF4_FOXA_diMut_GA5CG_n1	GGGGCGGGCTGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1567	3568_HNF4_FOXA_diMut_AG6CC_n1	GGGGCCCGCTGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1568	3575_HNF4_FOXA_diMut_AG6TT_n1	GGGGGTGCTGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1569	3577_HNF4_FOXA_diMut_GG7AC_n1	GGGGGAACCTGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1570	3584_HNF4_FOXA_diMut_GG7TT_n1	GGGGGATTCTGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1571	3585_HNF4_FOXA_diMut_GC8AA_n1	GGGGGAGAAATGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC

1572	3588_HNF4_FOXA_diMut_GC8CA_n1	GGGGAGCATGCTGGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1573	3589_HNF4_FOXA_diMut_GC8CG_n1	GGGGAGCGTGTGGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1574	3591_HNF4_FOXA_diMut_GC8TA_n1	GGGGAGTATGCTGGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1575	3595_HNF4_FOXA_diMut_CT9AC_n1	GGGGAGGACGCTGGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1576	3598_HNF4_FOXA_diMut_CT9GC_n1	GGGGAGGGCGCTGGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1577	3600_HNF4_FOXA_diMut_CT9TA_n1	GGGGAGGTAGCTGGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1578	3601_HNF4_FOXA_diMut_CT9TC_n1	GGGGAGGTCGCTGGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1579	3602_HNF4_FOXA_diMut_CT9TG_n1	GGGGAGGTGGCTGGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1580	3605_HNF4_FOXA_diMut_TG10AT_n1	GGGGAGGCATCTGGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1581	3608_HNF4_FOXA_diMut_TG10CT_n1	GGGGAGGCCCTCTGGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1582	3609_HNF4_FOXA_diMut_TG10GA_n1	GGGGAGGGACTGGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1583	3613_HNF4_FOXA_diMut_GC11AG_n1	GGGGAGGCTAGTGGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1584	3615_HNF4_FOXA_diMut_GC11CA_n1	GGGGAGGCTCATGGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1585	3618_HNF4_FOXA_diMut_GC11TA_n1	GGGGAGGCTTATGGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1586	3621_HNF4_FOXA_diMut_CT12AA_n1	GGGGAGGCTGAAGGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1587	3622_HNF4_FOXA_diMut_CT12AC_n1	GGGGAGGCTGACGGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1588	3624_HNF4_FOXA_diMut_CT12GA_n1	GGGGAGGCTGGAGGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1589	3625_HNF4_FOXA_diMut_CT12GC_n1	GGGGAGGCTGGCGGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1590	3626_HNF4_FOXA_diMut_CT12GG_n1	GGGGAGGCTGGGGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1591	3627_HNF4_FOXA_diMut_CT12TA_n1	GGGGAGGCTGTAGGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1592	3630_HNF4_FOXA_diMut_TG13AA_n1	GGGGAGGCTGCAAGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1593	3631_HNF4_FOXA_diMut_TG13AC_n1	GGGGAGGCTGCACGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1594	3632_HNF4_FOXA_diMut_TG13AT_n1	GGGGAGGCTGCATGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1595	3635_HNF4_FOXA_diMut_TG13CT_n1	GGGGAGGCTGCCTGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1596	3637_HNF4_FOXA_diMut_TG13GC_n1	GGGGAGGCTGCGCGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1597	3638_HNF4_FOXA_diMut_TG13GT_n1	GGGGAGGCTGCGTGTAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1598	3649_HNF4_FOXA_diMut_GT15AC_n1	GGGGAGGCTGCTGACAAACATTAAACCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1599	3711_HNF4_FOXA_diMut_IT22AA_n1	GGGGAGGCTGCTGGTAAACAAAACCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1600	3713_HNF4_FOXA_diMut_IT22AG_n1	GGGGAGGCTGCTGGTAAACAAGAACCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1601	3726_HNF4_FOXA_diMut_TA23GC_n1	GGGGAGGCTGCTGGTAAACATGCAACCAAGGTCACCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC

1602	3727_HNF4_FOXA_diMut_TA23GG_n1	GGGGAGGCTGCTGGTAAACATGGACCAAGGTCACCCACGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1603	3729_HNF4_FOXA_diMut_AA24CC_n1	GGGGAGGCTGCTGGTAAACATCCCCAAGGTCACCCACGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1604	3730_HNF4_FOXA_diMut_AA24CG_n1	GGGGAGGCTGCTGGTAAACATTCGCCAAGGTCACCCACGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1605	3732_HNF4_FOXA_diMut_AA24GC_n1	GGGGAGGCTGCTGGTAAACATTCGCCAAGGTCACCCACGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1606	3733_HNF4_FOXA_diMut_AA24GG_n1	GGGGAGGCTGCTGGTAAACATTCGCCAAGGTCACCCACGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1607	3734_HNF4_FOXA_diMut_AA24GT_n1	GGGGAGGCTGCTGGTAAACATTCGCCAAGGTCACCCACGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1608	3735_HNF4_FOXA_diMut_AA24TC_n1	GGGGAGGCTGCTGGTAAACATTCGCCAAGGTCACCCACGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1609	3736_HNF4_FOXA_diMut_AA24TG_n1	GGGGAGGCTGCTGGTAAACATTCGCCAAGGTCACCCACGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1610	3738_HNF4_FOXA_diMut_AC25CA_n1	GGGGAGGCTGCTGGTAAACATTAACAAGGTCACCCACGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1611	3740_HNF4_FOXA_diMut_AC25CT_n1	GGGGAGGCTGCTGGTAAACATTAACAAGGTCACCCACGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1612	3743_HNF4_FOXA_diMut_AC25GT_n1	GGGGAGGCTGCTGGTAAACATTAACAAGGTCACCCACGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1613	3763_HNF4_FOXA_diMut_CA27TG_n1	GGGGAGGCTGCTGGTAAACATTAAGTCAAGGTCACCCACGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1614	3814_HNF4_FOXA_diMut_CA33GG_n1	GGGGAGGCTGCTGGTAAACATTAACAAGTGGCCCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1615	3823_HNF4_FOXA_diMut_AC34GG_n1	GGGGAGGCTGCTGGTAAACATTAACAAGGTCGGCCCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1616	3828_HNF4_FOXA_diMut_CC35AA_n1	GGGGAGGCTGCTGGTAAACATTAACAAGGTCAAAACAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1617	3830_HNF4_FOXA_diMut_CC35AT_n1	GGGGAGGCTGCTGGTAAACATTAACAAGGTCAAATCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1618	3831_HNF4_FOXA_diMut_CC35GA_n1	GGGGAGGCTGCTGGTAAACATTAACAAGGTCAGACCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1619	3833_HNF4_FOXA_diMut_CC35GT_n1	GGGGAGGCTGCTGGTAAACATTAACAAGGTCAGTCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1620	3834_HNF4_FOXA_diMut_CC35TA_n1	GGGGAGGCTGCTGGTAAACATTAACAAGGTCATACCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1621	3836_HNF4_FOXA_diMut_CC35TT_n1	GGGGAGGCTGCTGGTAAACATTAACAAGGTCATCCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1622	3838_HNF4_FOXA_diMut_CC36AG_n1	GGGGAGGCTGCTGGTAAACATTAACAAGGTCACAGCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1623	3842_HNF4_FOXA_diMut_CC36GT_n1	GGGGAGGCTGCTGGTAAACATTAACAAGGTCACGTCAAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1624	3845_HNF4_FOXA_diMut_CC36TT_n1	GGGGAGGCTGCTGGTAAACATTAACAAGGTCACCTCAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1625	3846_HNF4_FOXA_diMut_CC37AA_n1	GGGGAGGCTGCTGGTAAACATTAACAAGGTCACCAAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1626	3847_HNF4_FOXA_diMut_CC37AG_n1	GGGGAGGCTGCTGGTAAACATTAACAAGGTCACAGAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1627	3849_HNF4_FOXA_diMut_CC37GA_n1	GGGGAGGCTGCTGGTAAACATTAACAAGGTCACCGAAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1628	3850_HNF4_FOXA_diMut_CC37GG_n1	GGGGAGGCTGCTGGTAAACATTAACAAGGTCACCGGAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1629	3852_HNF4_FOXA_diMut_CC37TA_n1	GGGGAGGCTGCTGGTAAACATTAACAAGGTCACCTAAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1630	3853_HNF4_FOXA_diMut_CC37TG_n1	GGGGAGGCTGCTGGTAAACATTAACAAGGTCACCTGAGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1631	3855_HNF4_FOXA_diMut_CA38AC_n1	GGGGAGGCTGCTGGTAAACATTAACAAGGTCACCCACGTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC

1632	3856_HNF4_FOXA_diMut_CA38AG_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCAGTTATCAGAGGAGCAAA CAGGGGCAAAAGTCCAC
1633	3857_HNF4_FOXA_diMut_CA38AT_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCATGTTATCAGAGGAGCAAA CAGGGGCAAAAGTCCAC
1634	3858_HNF4_FOXA_diMut_CA38GC_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCGCGTTATCAGAGGAGCAAA CAGGGGCAAAAGTCCAC
1635	3860_HNF4_FOXA_diMut_CA38GT_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCGCTGTTATCAGAGGAGCAAA CAGGGGCAAAAGTCCAC
1636	3864_HNF4_FOXA_diMut_AG39CA_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCCACTTATCAGAGGAGCAAA CAGGGGCAAAAGTCCAC
1637	3868_HNF4_FOXA_diMut_AG39GC_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCCGCTTATCAGAGGAGCAAA CAGGGGCAAAAGTCCAC
1638	3870_HNF4_FOXA_diMut_AG39TA_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCCTATTATCAGAGGAGCAAA CAGGGGCAAAAGTCCAC
1639	3871_HNF4_FOXA_diMut_AG39TC_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCCTCTTATCAGAGGAGCAAA CAGGGGCAAAAGTCCAC
1640	3872_HNF4_FOXA_diMut_AG39TT_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCCTTTTATCAGAGGAGCAAA CAGGGGCAAAAGTCCAC
1641	3880_HNF4_FOXA_diMut_GT40TC_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCCACTCTATCAGAGGAGCAAA CAGGGGCAAAAGTCCAC
1642	3881_HNF4_FOXA_diMut_GT40TG_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCCACTGTATCAGAGGAGCAAA CAGGGGCAAAAGTCCAC
1643	3883_HNF4_FOXA_diMut_TT41AC_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCCAAGACATCAGAGGAGCAAA CAGGGGCAAAAGTCCAC
1644	3884_HNF4_FOXA_diMut_TT41AG_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCCAAGAGATCAGAGGAGCAAA CAGGGGCAAAAGTCCAC
1645	3887_HNF4_FOXA_diMut_TT41CG_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCCAAGCGATCAGAGGAGCAAA CAGGGGCAAAAGTCCAC
1646	3889_HNF4_FOXA_diMut_TT41GC_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCCAAGGCATCAGAGGAGCAAA CAGGGGCAAAAGTCCAC
1647	3890_HNF4_FOXA_diMut_TT41GG_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCCAAGGATCAGAGGAGCAAA CAGGGGCAAAAGTCCAC
1648	3897_HNF4_FOXA_diMut_TA42GC_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCCAAGTGTCTCAGAGGAGCAAA CAGGGGCAAAAGTCCAC
1649	3904_HNF4_FOXA_diMut_AT43GC_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCCAAGTGTGTCAGAGGAGCAAA CAGGGGCAAAAGTCCAC
1650	3910_HNF4_FOXA_diMut_TC44AG_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCCAAGTTAAAGAGGAGCAAA CAGGGGCAAAAGTCCAC
1651	3913_HNF4_FOXA_diMut_TC44CG_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCCAAGTTACGAGAGGAGCAAA CAGGGGCAAAAGTCCAC
1652	3917_HNF4_FOXA_diMut_TC44GT_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCCAAGTTAGTAGAGGAGCAAA CAGGGGCAAAAGTCCAC
1653	3934_HNF4_FOXA_diMut_AG46TC_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCCAAGTTATCTCAGAGGAGCAAA CAGGGGCAAAAGTCCAC
1654	3936_HNF4_FOXA_diMut_GA47AC_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCCAAGTTATCAACGGAGCAAA CAGGGGCAAAAGTCCAC
1655	3938_HNF4_FOXA_diMut_GA47AT_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCCAAGTTATCAATGGAGCAAA CAGGGGCAAAAGTCCAC
1656	3940_HNF4_FOXA_diMut_GA47CG_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCCAAGTTATCAGGGGAGCAAA CAGGGGCAAAAGTCCAC
1657	3941_HNF4_FOXA_diMut_GA47CT_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCCAAGTTATCACTGGAGCAAA CAGGGGCAAAAGTCCAC
1658	3945_HNF4_FOXA_diMut_AG48CA_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCCAAGTTATCAGCAGAGCAAA CAGGGGCAAAAGTCCAC
1659	3947_HNF4_FOXA_diMut_AG48CT_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCCAAGTTATCAGCTGAGCAAA CAGGGGCAAAAGTCCAC
1660	3948_HNF4_FOXA_diMut_AG48GA_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCCAAGTTATCAGGAGAGCAAA CAGGGGCAAAAGTCCAC
1661	3949_HNF4_FOXA_diMut_AG48GC_n1	GGGGAGGCTGCTGGTAAACATTAA CCAAGGTCACCCCAAGTTATCAGGCGAGCAAA CAGGGGCAAAAGTCCAC

1662	3950_HNF4_FOXA	diMut_AG48GT_n1	GGGGAGGCTGCTGGTAAACATTAACCAAGGTCACCCACGTTTATCAGGTGAGCAAAACAGGGGCAAAAGTCCAC
1663	3951_HNF4_FOXA	diMut_AG48TA_n1	GGGGAGGCTGCTGGTAAACATTAACCAAGGTCACCCACGTTTATCAGTAGAGCAAAACAGGGGCAAAAGTCCAC
1664	3952_HNF4_FOXA	diMut_AG48TC_n1	GGGGAGGCTGCTGGTAAACATTAACCAAGGTCACCCACGTTTATCAGTCGAGCAAAACAGGGGCAAAAGTCCAC
1665	3953_HNF4_FOXA	diMut_AG48TT_n1	GGGGAGGCTGCTGGTAAACATTAACCAAGGTCACCCACGTTTATCAGTTGAGCAAAACAGGGGCAAAAGTCCAC
1666	3954_HNF4_FOXA	diMut_GG49AA_n1	GGGGAGGCTGCTGGTAAACATTAACCAAGGTCACCCACGTTTATCAGAAAAACAGGGGCAAAAGTCCAC
1667	3955_HNF4_FOXA	diMut_GG49AC_n1	GGGGAGGCTGCTGGTAAACATTAACCAAGGTCACCCACGTTTATCAGAAACAGGGGCAAAAGTCCAC
1668	3956_HNF4_FOXA	diMut_GG49AT_n1	GGGGAGGCTGCTGGTAAACATTAACCAAGGTCACCCACGTTTATCAGAAATAGCAAAACAGGGGCAAAAGTCCAC
1669	3957_HNF4_FOXA	diMut_GG49CA_n1	GGGGAGGCTGCTGGTAAACATTAACCAAGGTCACCCACGTTTATCAGACAAACAGGGGCAAAAGTCCAC
1670	3959_HNF4_FOXA	diMut_GG49CT_n1	GGGGAGGCTGCTGGTAAACATTAACCAAGGTCACCCACGTTTATCAGACTAGCAAAACAGGGGCAAAAGTCCAC
1671	3960_HNF4_FOXA	diMut_GG49TA_n1	GGGGAGGCTGCTGGTAAACATTAACCAAGGTCACCCACGTTTATCAGATAAGCAAAACAGGGGCAAAAGTCCAC
1672	3967_HNF4_FOXA	diMut_GA50CG_n1	GGGGAGGCTGCTGGTAAACATTAACCAAGGTCACCCACGTTTATCAGAGCGGCAAAACAGGGGCAAAAGTCCAC
1673	3968_HNF4_FOXA	diMut_GA50CT_n1	GGGGAGGCTGCTGGTAAACATTAACCAAGGTCACCCACGTTTATCAGAGCTGCAAAACAGGGGCAAAAGTCCAC
1674	3969_HNF4_FOXA	diMut_GA50TC_n1	GGGGAGGCTGCTGGTAAACATTAACCAAGGTCACCCACGTTTATCAGAGTCGCAAAACAGGGGCAAAAGTCCAC
1675	3972_HNF4_FOXA	diMut_AG51CA_n1	GGGGAGGCTGCTGGTAAACATTAACCAAGGTCACCCACGTTTATCAGAGGCAAAACAGGGGCAAAAGTCCAC
1676	3973_HNF4_FOXA	diMut_AG51CC_n1	GGGGAGGCTGCTGGTAAACATTAACCAAGGTCACCCACGTTTATCAGAGGCCCCAAACAGGGGCAAAAGTCCAC
1677	3975_HNF4_FOXA	diMut_AG51GA_n1	GGGGAGGCTGCTGGTAAACATTAACCAAGGTCACCCACGTTTATCAGAGGGACAAACAGGGGCAAAAGTCCAC
1678	3979_HNF4_FOXA	diMut_AG51IC_n1	GGGGAGGCTGCTGGTAAACATTAACCAAGGTCACCCACGTTTATCAGAGGTCCTCAACAGGGGCAAAAGTCCAC
1679	3982_HNF4_FOXA	diMut_GC52AG_n1	GGGGAGGCTGCTGGTAAACATTAACCAAGGTCACCCACGTTTATCAGAGGAAGAAACAGGGGCAAAAGTCCAC
1680	3989_HNF4_FOXA	diMut_GC52TT_n1	GGGGAGGCTGCTGGTAAACATTAACCAAGGTCACCCACGTTTATCAGAGGATTAACAGGGGCAAAAGTCCAC
1682	4000_HNF4_FOXA	diMut_AA54CG_n1	GGGGAGGCTGCTGGTAAACATTAACCAAGGTCACCCACGTTTATCAGAGGAGCCGACAGGGGCAAAAGTCCAC
1682	4007_HNF4_FOXA	diMut_AA54TT_n1	GGGGAGGCTGCTGGTAAACATTAACCAAGGTCACCCACGTTTATCAGAGGAGCTTACAGGGGCAAAAGTCCAC
1683	4012_HNF4_FOXA	diMut_AA55GG_n1	GGGGAGGCTGCTGGTAAACATTAACCAAGGTCACCCACGTTTATCAGAGGAGCAGGAGGGGCAAAAGTCCAC
1684	4033_HNF4_FOXA	diMut_CA57TG_n1	GGGGAGGCTGCTGGTAAACATTAACCAAGGTCACCCACGTTTATCAGAGGAGCAAAATGGGGGCAAAAGTCCAC
1685	4038_HNF4_FOXA	diMut_AG58GA_n1	GGGGAGGCTGCTGGTAAACATTAACCAAGGTCACCCACGTTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1686	4068_HNF4_FOXA	diMut_GG61TA_n1	GGGGAGGCTGCTGGTAAACATTAACCAAGGTCACCCACGTTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCAC
1687	4069_HNF4_FOXA	diMut_GG61TC_n1	GGGGAGGCTGCTGGTAAACATTAACCAAGGTCACCCACGTTTATCAGAGGAGCAAAACAGGTCCAAAGTCCAC
1688	4070_HNF4_FOXA	diMut_GG61TT_n1	GGGGAGGCTGCTGGTAAACATTAACCAAGGTCACCCACGTTTATCAGAGGAGCAAAACAGGTTCAAAGTCCAC
1689	4155_HNF4_FOXA	diMut_AC71GA_n1	GGGGAGGCTGCTGGTAAACATTAACCAAGGTCACCCACGTTTATCAGAGGAGCAAAACAGGGGCAAAAGTCCGA

REFERENCES

[00508] All publications and references, including but not limited to patents and patent applications, cited in this specification and Examples herein are incorporated by reference in their entirety as if each individual publication or reference were specifically and individually indicated to be incorporated by reference herein as being fully set forth. Any patent application to which this application claims priority is also incorporated by reference herein in the manner described above for publications and references.

CLAIMS

1. A liver-specific nucleic acid regulatory element comprising a nucleic acid sequence having at least 93% identity to any one of SEQ ID NOs: 1-80, 138, or 139.
2. The liver-specific nucleic acid regulatory element of claim 1, wherein the nucleic acid sequence has at least 94% identity to any one of SEQ ID NOs: 1-80, 138, or 139.
3. The liver-specific nucleic acid regulatory element of claim 2, wherein the nucleic acid sequence has at least 95% identity to any one of SEQ ID NOs: 1-80, 138, 139.
4. The liver-specific nucleic acid regulatory element of claim 3, wherein the nucleic acid sequence has at least 96% identity to any one of SEQ ID NOs: 1-80, 138, 139 or 155-1689.
5. The liver-specific nucleic acid regulatory element of claim 4, wherein the nucleic acid sequence has at least 97% identity to any one of SEQ ID NOs: 1-80, 138, 139.
6. The liver-specific nucleic acid regulatory element of claim 5, wherein the nucleic acid sequence has at least 98% identity to any one of SEQ ID NOs: 1-80, 138, or 139.
7. The liver-specific nucleic acid regulatory element of claim 6, wherein the nucleic acid sequence has at least 99% identity to any one of SEQ ID NOs: 1-80, 138, or 139.
8. The liver-specific nucleic acid regulatory element of claim 3, wherein the nucleic acid sequence comprises any one of SEQ ID NOs: 1-80, 138, or 139.
9. The liver-specific nucleic acid regulatory element of claim 3, wherein the nucleic acid sequence has at least 95% identity to SEQ ID NO: 131.
10. The liver-specific nucleic acid regulatory element of claim 3, wherein the nucleic acid sequence has at least 96% identity to SEQ ID NO: 131.
11. The liver-specific nucleic acid regulatory element of claim 3, wherein the nucleic acid sequence has at least 97% identity to SEQ ID NO: 131.
12. The liver-specific nucleic acid regulatory element of claim 3, wherein the nucleic acid sequence has at least 98% identity to SEQ ID NO: 131.
13. The liver-specific nucleic acid regulatory element of claim 3, wherein the nucleic acid sequence has at least 99% identity to SEQ ID NO: 131.
14. The liver-specific nucleic acid regulatory element of claim 3, wherein the nucleic acid sequence comprises SEQ ID NO: 131.

15. The liver-specific nucleic acid regulatory element of claim 3, wherein the nucleic acid sequence has at least 95% identity to SEQ ID NO: 122.
16. The liver-specific nucleic acid regulatory element of claim 3, wherein the nucleic acid sequence has at least 96% identity to SEQ ID NO: 122.
17. The liver-specific nucleic acid regulatory element of claim 3, wherein the nucleic acid sequence has at least 97% identity to SEQ ID NO: 122.
18. The liver-specific nucleic acid regulatory element of claim 3, wherein the nucleic acid sequence has at least 98% identity to SEQ ID NO: 122.
19. The liver-specific nucleic acid regulatory element of claim 3, wherein the nucleic acid sequence has at least 99% identity to SEQ ID NO: 122.
20. The liver-specific nucleic acid regulatory element of claim 3, wherein the nucleic acid sequence comprises SEQ ID NO: 122.
21. A liver-specific nucleic acid regulatory element consisting essentially of a nucleic acid sequence set forth in any one of Table 10, Table 11, Table 12, or Table 13.
22. A liver-specific nucleic acid regulatory element comprising a nucleic acid sequence set forth in any one of Table 10, Table 11, Table 12, or Table 13.
23. The liver-specific nucleic acid regulatory element of claim 22, wherein the element comprises at least two nucleic acid sequences set forth in any one of Table 10, Table 11, Table 12, or Table 13.
24. The liver-specific nucleic acid regulatory element of claim 23, wherein the two nucleic acid sequences are identical.
25. The liver-specific nucleic acid regulatory element of claim 22, wherein the element comprises three (3) nucleic acid sequences set forth in any one of Table 10, Table 11, Table 12, or Table 13, optionally wherein the three sequences are identical.
26. The liver-specific nucleic acid regulatory element of claim 22, wherein the element consists essentially of two (2) to ten (10) nucleic acid sequences set forth in any one of Table 10, Table 11, Table 12, or Table 13.

27. The liver-specific nucleic acid regulatory element of any one of claims 23-26, wherein the element comprises a spacer placed between the nucleic acid sequences set forth in any one of Table 10, Table 11, Table 12, or Table 13.

28. The liver-specific nucleic acid regulatory element of claim 27, wherein the spacer is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 base pairs long.

29. The liver-specific nucleic acid regulatory element of claim 22, wherein the element comprises a nucleic acid sequence at least 95%, 96%, 97%, 98% or 99% identical to:

GGGGGAGGCTGCTGGTGAATATTAACCAAGGTCACCCAGTTATCGGAGGAGCAAA
CAGGGGCAAAGTCCAC (SEQ ID NO: 223),
GGGGGAAGCTACTGGTGAATATTAACCAAGGTCACCCAGTTATCAGGGAGCAAACA
GGAGCAAAGTCCAT (SEQ ID NO: 1381),
GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCCGTTATCGGAGGAGCAAACAA
GGGCTAAGTCCAC (SEQ ID NO: 1073), or
GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAACAA
GGGCAAAGTCCAC (SEQ ID NO: 1113).

30. The liver-specific nucleic acid regulatory element of claim 22, wherein the element comprises a nucleic acid consisting of:

GGGGGAGGCTGCTGGTGAATATTAACCAAGGTCACCCAGTTATCGGAGGAGCAAA
CAGGGGCAAAGTCCAC (SEQ ID NO: 223),
GGGGGAAGCTACTGGTGAATATTAACCAAGGTCACCCAGTTATCAGGGAGCAAACA
GGAGCAAAGTCCAT (SEQ ID NO: 1381),
GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCCGTTATCGGAGGAGCAAACAAG
GGCTAAGTCCAC (SEQ ID NO: 1073), or
GGAGGCTGTTGGTGAATATTAACCAAGGTCACCTCAGTTATCGGAGGAGCAAACAA
GGGCAAAGTCCAC (SEQ ID NO: 1113).

31. A liver-specific expression cassette comprising the liver-specific nucleic acid regulatory element of any one of claims 1-30, and a liver-specific promoter operably linked to a transgene.

32. The liver-specific expression cassette of claim 31, wherein two or more nucleotides separate each liver-specific nucleic acid regulatory element.

33. A liver-specific expression cassette comprising at least three repeats of a liver-specific nucleic acid regulatory element and a liver-specific promoter operably linked to a transgene,

wherein the liver-specific nucleic acid regulatory element comprises a nucleic acid sequence having at least 95% identity to any one of SEQ ID NOs: 81-137, and

wherein two or more nucleotides separate each liver-specific nucleic acid regulatory element.

34. The liver-specific expression cassette of claim 33, wherein between 2 and 30 nucleotides separate each regulatory element.

35. The liver-specific expression cassette of claim 34, wherein 5 nucleotides separate each regulatory element.

36. The liver-specific expression cassette of claim 34, wherein 11 nucleotides separate each regulatory element.

37. The liver-specific expression cassette of claim 34, wherein 30 nucleotides separate each regulatory element.

38. The liver-specific expression cassette of any one of claims 33-37, wherein the liver-specific expression cassette comprises two, three, four, or five repeats of the liver-specific nucleic acid regulatory element.

39. The liver-specific expression cassette of any one of claims 33-37, wherein the liver-specific expression cassette comprises six, seven, eight, nine or ten repeats of the liver-specific nucleic acid regulatory element.

40. The liver-specific expression cassette of any one of claims 31-39, comprising one or more FOXA and HNF4 transcription factor consensus sites.

41. The liver-specific expression cassette of any one of claims 31-40, wherein liver-specific nucleic acid regulatory element comprises one or more sites of CpG minimization.

42. The liver-specific expression cassette of any one of claims 31-41, wherein the liver-specific promoter is selected from the group consisting of: a transthyretin (TTR) promoter, minimal TTR promoter (TTRm), an AAT promoter, an albumin (ALB) promoter or minimal promoter, an apolipoprotein A1 (APOA1) promoter or minimal promoter, a complement factor B (CFB) promoter, a ketohexokinase (KHK) promoter, a hemopexin (HPX) promoter or minimal promoter, a nicotinamide N-methyltransferase (NNMT) promoter or minimal promoter, a carboxylesterase 1 (CES1) promoter or minimal promoter, a protein C (PROC) promoter or minimal promoter, an apolipoprotein C3 (APOC3) promoter or minimal promoter, a mannan-

binding lectin serine protease 2 (MASP2) promoter or minimal promoter, a hepcidin antimicrobial peptide (HAMP) promoter or minimal promoter, and a serpin peptidase inhibitor, clade C (antithrombin), member 1 (SERPINC1) promoter or minimal promoter.

43. The liver-specific expression cassette of claim 42, wherein the promoter comprises any sequence from **Table 1**.

44. The liver-specific expression cassette according to claim 42, wherein said liver-specific promoter is a TTR promoter or a TTRm promoter.

45. The liver-specific expression cassette according any one of claims 31-44, wherein the transgene encodes a liver-specific therapeutic protein.

46. The liver-specific expression cassette of claim 45, wherein the liver-specific therapeutic protein is coagulation factor VIII (FVIII).

47. The liver-specific expression cassette of claim 46, wherein the coagulation FVIII comprises a codon optimized nucleic acid sequence.

48. The liver-specific expression cassette of claim 46, wherein the coagulation FVIII comprises a nucleic acid sequence having at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% identity to, comprises, or consists of SEQ ID NO: 143.

49. A vector comprising the liver-specific nucleic acid regulatory element of any one of claims 1-30 or the liver-specific expression cassette according to any one of claims 31-48.

50. The vector of claim 49, wherein the vector is a viral vector or a non-viral vector.

51. The vector of claim 49, wherein the vector is a plasmid.

52. The vector of claim 49 or claim 50, wherein the vector is a closed-ended DNA (ccDNA) vector.

53. A pharmaceutical composition comprising the liver-specific expression cassette according to any one of claims 31-48 or the vector according to any one of claims 49-52, and a pharmaceutically acceptable excipient.

54. A method of treating a liver-specific disease or disorder comprising transduction or transfection of the vector according to any one of claims 49-52, or the pharmaceutical composition of claim 53, into a subject.

55. The method of claim 54, wherein the subject is a human subject suffering from a genetic disorder.

56. The method of claim 54 or claim 55, wherein the subject has hemophilia A.

57. The method of claim 55, wherein the genetic disorder is selected from the group consisting of sickle-cell anemia, melanoma, hemophilia A (clotting factor VIII (FVIII) deficiency) and hemophilia B (clotting factor IX (FIX) deficiency), cystic fibrosis (CFTR), familial hypercholesterolemia (LDL receptor defect), hepatoblastoma, Wilson disease, phenylketonuria (PKU), congenital hepatic porphyria, inherited disorders of hepatic metabolism, Lesch Nyhan syndrome, sickle cell anemia, thalassaemias, xeroderma pigmentosum, Fanconi's anemia, retinitis pigmentosa, ataxia telangiectasia, Bloom's syndrome, retinoblastoma, mucopolysaccharide storage diseases (*e.g.*, Hurler syndrome (MPS Type I), Scheie syndrome (MPS Type I S), Hurler-Scheie syndrome (MPS Type I H-S), Hunter syndrome (MPS Type II), Sanfilippo Types A, B, C, and D (MPS Types III A, B, C, and D), Morquio Types A and B (MPS IVA and MPS IVB), Maroteaux-Lamy syndrome (MPS Type VI), Sly syndrome (MPS Type VII), hyaluronidase deficiency (MPS Type IX)), Niemann-Pick Disease Types A/B, C1 and C2, Fabry disease, Schindler disease, GM2-gangliosidosis Type II (Sandhoff Disease), Tay-Sachs disease, Metachromatic Leukodystrophy, Krabbe disease, Mucopolipidosis Type I, II/III and IV, Sialidosis Types I and II, Glycogen Storage disease Types I and II (Pompe disease), Gaucher disease Types I, II and III, Fabry disease, cystinosis, Batten disease, Aspartylglucosaminuria, Salla disease, Danon disease (LAMP-2 deficiency), Lysosomal Acid Lipase (LAL) deficiency, neuronal ceroid lipofuscinoses (CLN1-8, INCL, and LINCL), sphingolipidoses, galactosialidosis, amyotrophic lateral sclerosis (ALS), Parkinson's disease, Alzheimer's disease, Huntington's disease, spinocerebellar ataxia, spinal muscular atrophy, Friedreich's ataxia, Duchenne muscular dystrophy (DMD), Becker muscular dystrophies (BMD), dystrophic epidermolysis bullosa (DEB), ectonucleotide pyrophosphatase 1 deficiency, generalized arterial calcification of infancy (GACI), Leber Congenital Amaurosis, Stargardt macular dystrophy (ABCA4), ornithine transcarbamylase (OTC) deficiency, Usher syndrome, alpha-1 antitrypsin deficiency, progressive familial intrahepatic cholestasis (PFIC) type I (ATP8B1 deficiency), type II (ABCB11), type III (ABCB4), or type IV (TJP2) and Cathepsin A deficiency.

58. A method of increasing expression capacity of a liver-specific enhancer element

comprising the nucleic acid sequence CTAAG, comprising introducing a single nucleotide substitution (T to A) mutation such that the substitution results in the nucleic acid sequence comprising CAAAG.

59. A liver-specific enhancer element comprising a nucleic acid sequence selected from: CAAAG; CAAAGT; CAAAGTC; GCAAAGT; GCAAAG; or GCAAAGTC.