



US 20240052368A1

(19) **United States**

(12) **Patent Application Publication**
Majumder et al.

(10) **Pub. No.: US 2024/0052368 A1**

(43) **Pub. Date: Feb. 15, 2024**

(54) **GENERATION OF NEXT GENERATION RECOMBINANT AAV GENE THERAPY VECTORS THAT ADOPT 3D CONFORMATION**

(71) Applicant: **WISCONSIN ALUMNI RESEARCH FOUNDATION**, Madison, WI (US)

(72) Inventors: **Kinjal Majumder**, Madison, WI (US); **Clairine Larsen**, Madison, WI (US)

(21) Appl. No.: **18/332,380**

(22) Filed: **Jun. 9, 2023**

Related U.S. Application Data

(60) Provisional application No. 63/350,769, filed on Jun. 9, 2022.

Publication Classification

(51) **Int. Cl.**
C12N 15/86 (2006.01)

(52) **U.S. Cl.**
CPC *C12N 15/86* (2013.01); *C12N 2750/14143* (2013.01); *C12N 2750/14152* (2013.01); *A61K 48/00* (2013.01)

(57) **ABSTRACT**

The present invention provides constructs for producing modified recombinant adeno-associated virus (rAAV) vectors. The constructs comprise one or more CCCTC-binding factor (CTCF) binding sites, which facilitate DNA looping and promote efficient transgene expression. Also provided are modified rAAV virus particles comprising these constructs, methods for producing the modified rAAV virus particles, and methods of using the modified rAAV virus particles to deliver a transgene to a subject.

Specification includes a Sequence Listing.

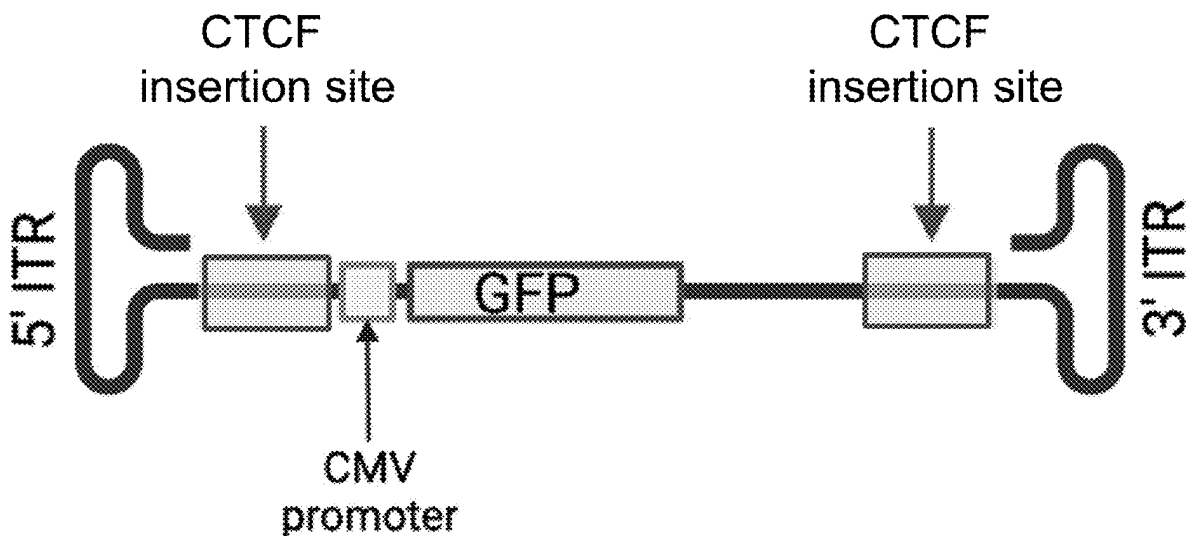
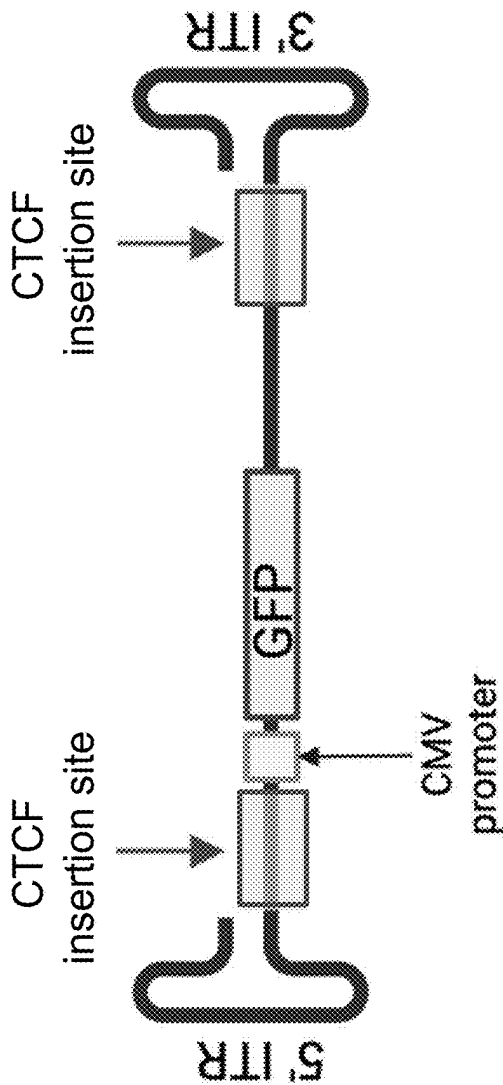


FIG. 1



FIGS. 2A-2B

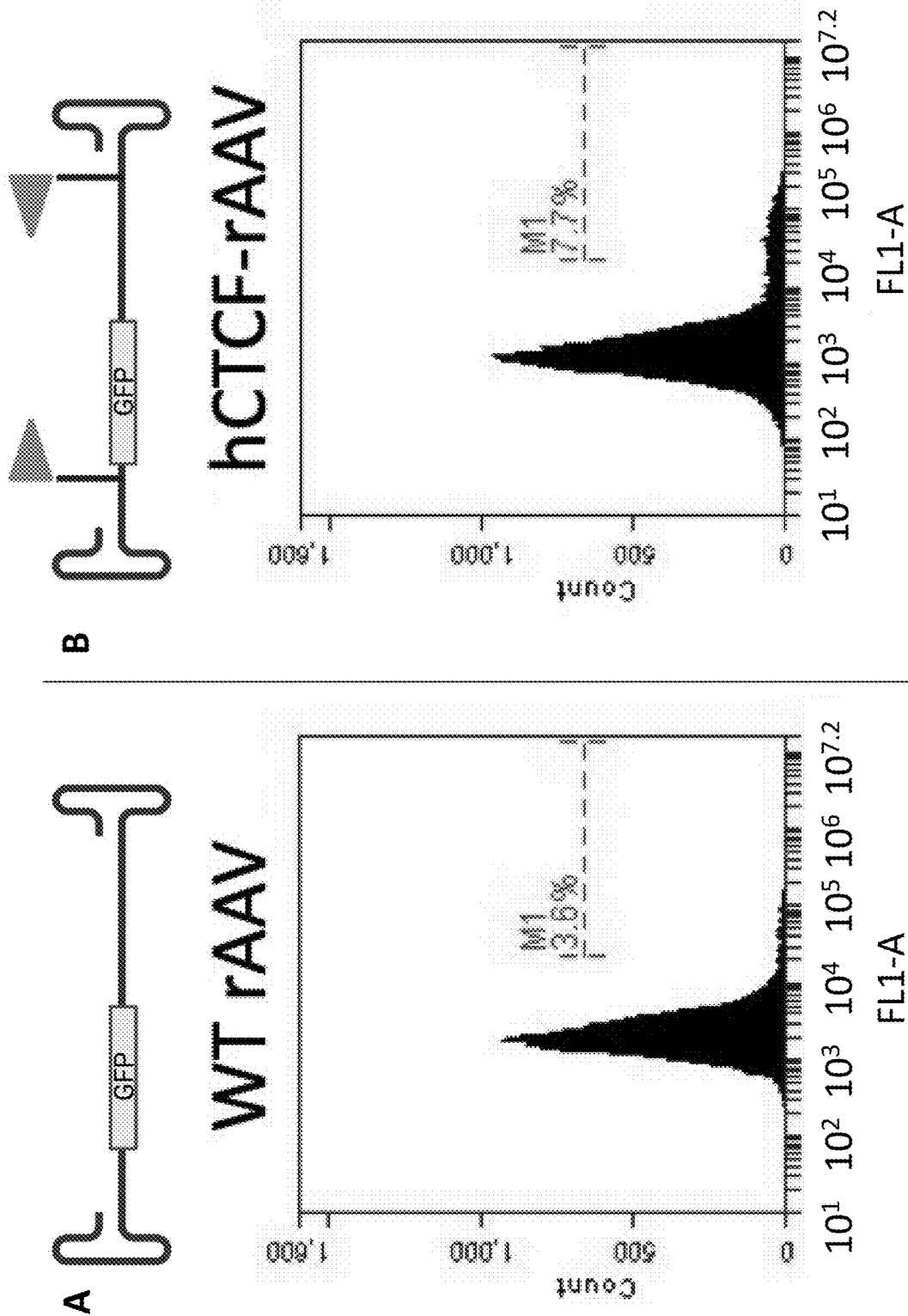


FIG. 3

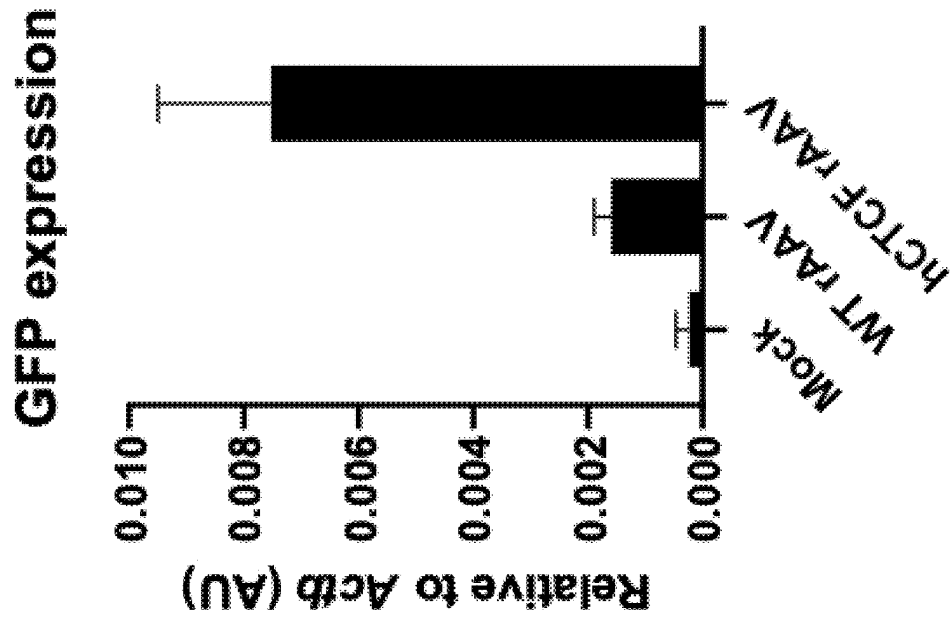
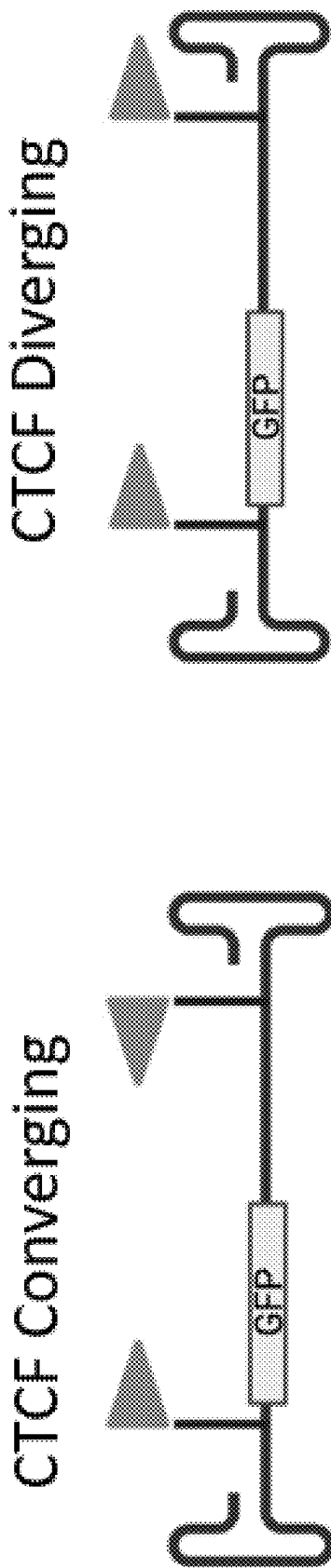


FIG. 4



FIGS. 5A-5C

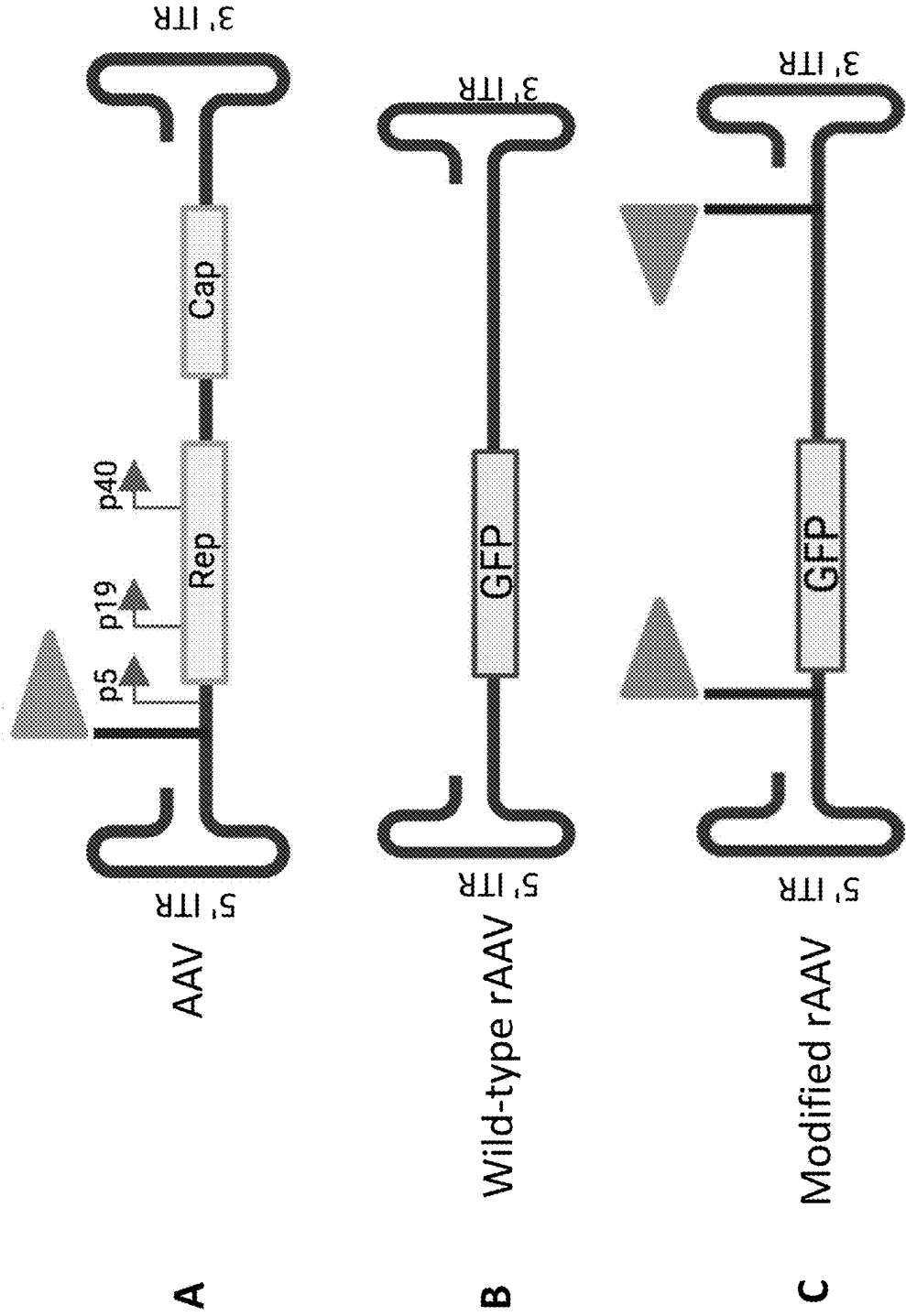
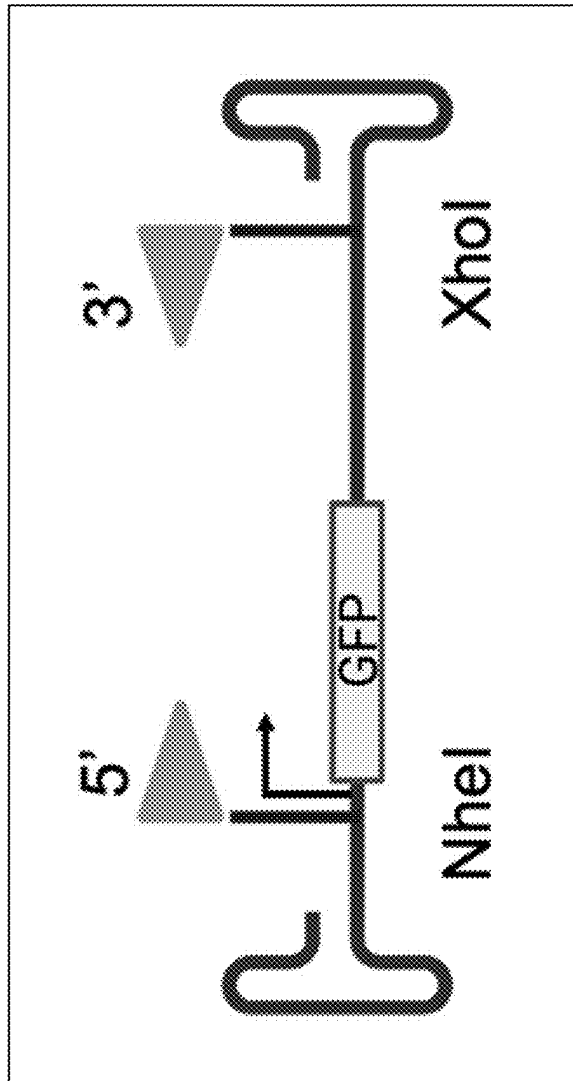


FIG. 6



FIGS. 7A-7E

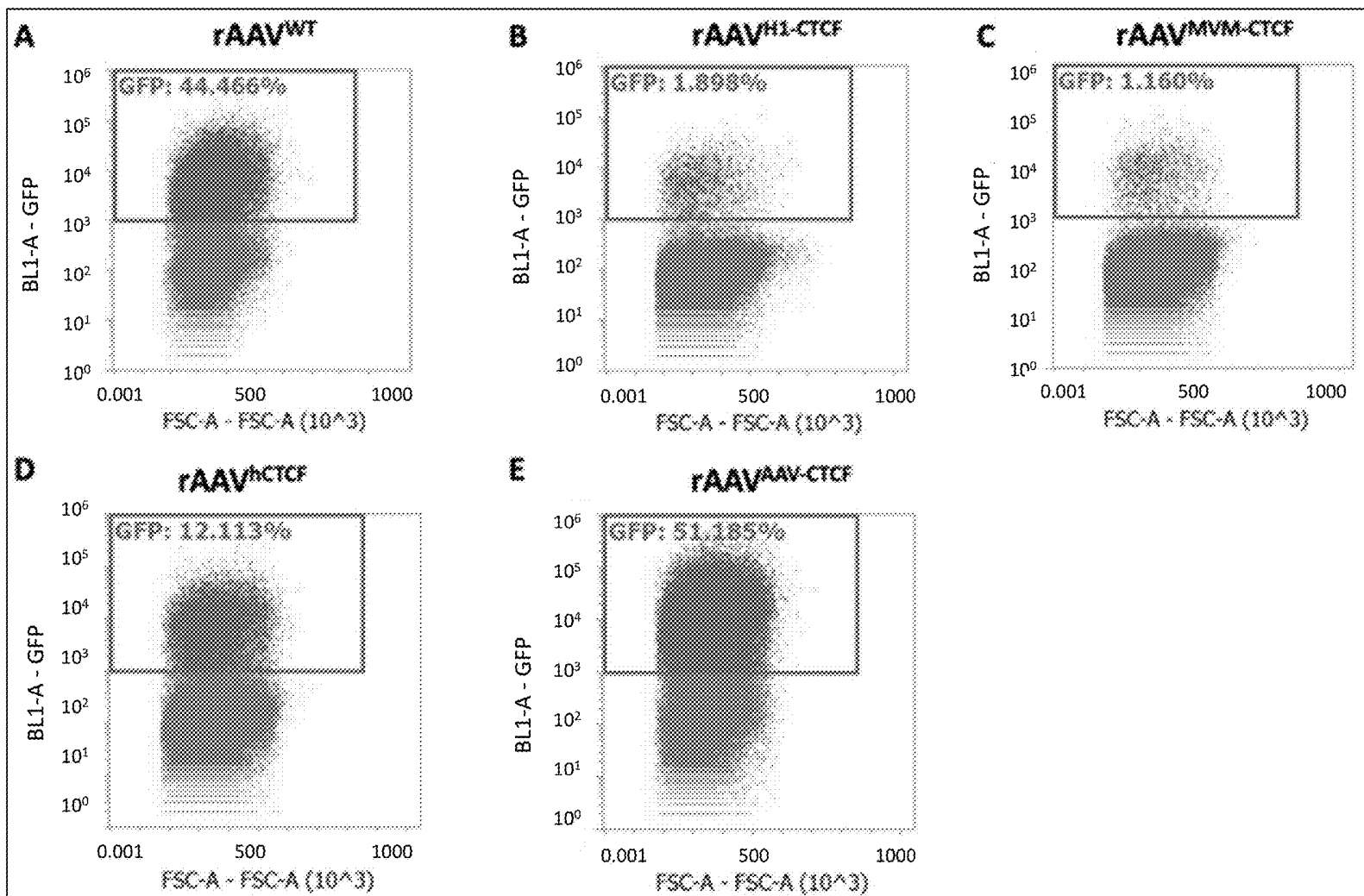
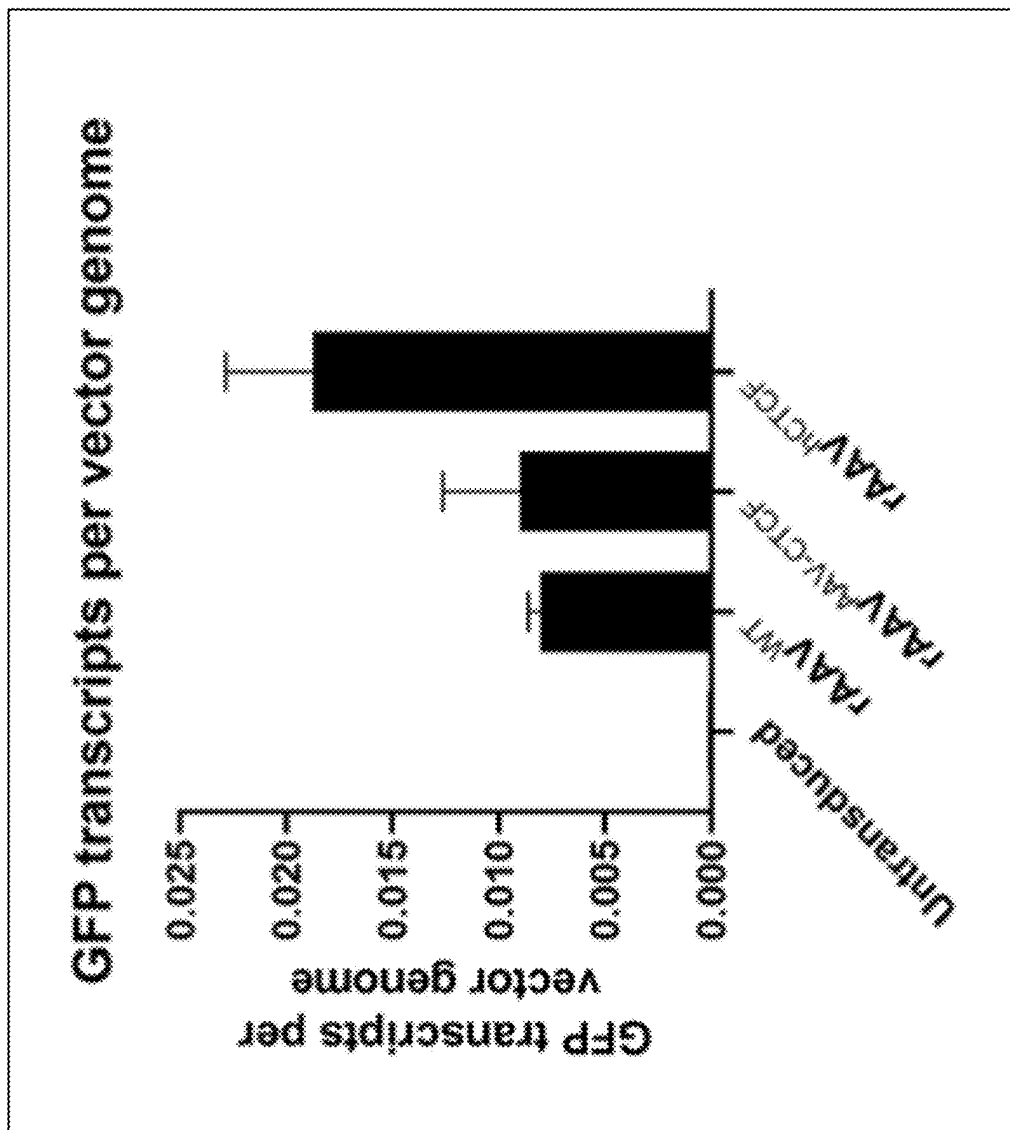


FIG. 8



**GENERATION OF NEXT GENERATION
RECOMBINANT AAV GENE THERAPY
VECTORS THAT ADOPT 3D
CONFORMATION**

**CROSS-REFERENCE TO RELATED
APPLICATIONS**

[0001] This application claims priority to U.S. Provisional Application No. 63/350,769 filed on Jun. 9, 2022, the content of which is incorporated by reference in its entirety.

**STATEMENT REGARDING FEDERALLY
SPONSORED RESEARCH**

[0002] This invention was made with government support under AI148511 awarded by the National Institutes of Health. The government has certain rights in the invention.

**REFERENCE TO AN ELECTRONIC SEQUENCE
LISTING**

[0003] The contents of the electronic sequence listing (960296.04414.xml; Size: 100,339 bytes; and Date of Creation: Jun. 20, 2023) is herein incorporated by reference in its entirety.

BACKGROUND

[0004] Recombinant adeno associated virus vectors (rAAV) offer great potential for use as gene therapy vectors for the treatment of monogenic diseases, such as muscular dystrophy and spinal muscular atrophy. rAAV gene therapy vectors offer several advantages over other types of viral vectors due to (1) their ability to persist long-term as a largely unintegrated expression platform, and (2) their inability to elicit significant innate immune responses in the host. However, the widespread use of rAAV vectors in the clinic is limited by our lack of knowledge about how the rAAV genome is chromatinized, where in the host nucleus it persists long-term, and how the vector genome navigates the nuclear milieu. Current gene therapy applications utilize high doses of rAAV vectors (10^{12} - 10^{13} viral genomes per kg) to ensure proper transgene expression. High doses increase production costs and increase the risk for oncogenic integration and toxicity.

[0005] Accordingly, there remains a need in the art for rAAV vectors that drive robust transgene expression in target cells.

SUMMARY

[0006] In a first aspect, the present disclosure provides a construct for producing a recombinant adeno-associated virus (rAAV) vector. The construct comprises: a 5' inverted terminal repeat (ITR), a first CCCTC-binding factor (CTCF) binding site, a promoter, a transgene, and a 3' ITR. In embodiments, the construct further comprises a second CTCF binding site. In embodiments, the construct comprises from 5' to 3': the 5' inverted terminal repeat (ITR), the first CCCTC-binding factor (CTCF) binding site, the promoter, the transgene, the second CTCF binding site, and the 3' ITR. In embodiments, the second CTCF binding site is in the convergent orientation relative to the first CTCF binding site. In embodiments, the CTCF binding site(s) are from a human or a virus.

[0007] In embodiments, the virus is selected from the group consisting of: adeno-associated virus (AAV), minute virus of mice (MVM), H1 parvovirus, MmuPV, B19, canine parvovirus, human cytomegalovirus (HCMV)/human herpesvirus 5 strain Merlin, human alphaherpesvirus 1, human herpesvirus 4 type 2 (Epstein-Barr virus type 2), HPV16, herpes simplex virus (HSV), and herpes B virus (HBV).

[0008] In embodiments, the CTCF binding site(s) comprise a sequence selected from: SEQ ID NOs:1-28. In embodiments, the first CTCF binding site comprises SEQ ID NO:1 and the second CTCF binding site comprises SEQ ID NO:42. In embodiments, the first and/or second CTCF binding site comprises multiple CTCF binding sequences. In embodiments, the first and/or second CTCF binding site comprises five CTCF binding sequences. In embodiments, the first CTCF binding site comprises SEQ ID NO: 3.

[0009] In a second aspect, the present invention provides host cells transduced with a construct described herein.

[0010] In a third aspect, the present invention provides rAAV virus particles comprising a construct described herein.

[0011] In a fourth aspect, the present invention provides packaging cell lines for producing the virus particles described herein.

[0012] In a fifth aspect, the present invention provides a method for producing a modified rAAV virus particle. The method comprises: (a) transducing a host cell with a plasmid comprising a construct described herein, a packaging plasmid, and a helper plasmid; (b) collecting the supernatant and the cells from culture; and (c) isolating virus particles from the supernatant and cells. In embodiments, the method further comprises concentrating the virus particles.

[0013] In a sixth aspect, the present invention provides a method of delivering a transgene to a subject in need thereof. The method comprises: administering a modified rAAV virus particle described herein to the subject. In embodiments, the transgene is expressed in a greater proportion of the subject's cells when it is delivered in the modified rAAV vector as compared to when it is delivered in a wild-type rAAV vector. In embodiments, the transgene is expressed at higher levels when it is delivered in the modified rAAV vector as compared to when it is delivered in a wild-type rAAV vector.

BRIEF DESCRIPTION OF THE DRAWINGS

[0014] FIG. 1 is a schematic showing how the modified recombinant adeno-associated virus (rAAV) vectors tested in the Examples were generated. A wild-type rAAV vector comprising a green fluorescent protein (GFP) transgene operably linked to a cytomegalovirus (CMV) promoter was modified via insertion of a first CTCF binding site between the 5' inverted terminal repeat (ITR) and the CMV promoter and a second CTCF binding site between the GFP transgene and the 3' ITR.

[0015] FIGS. 2A-2B shows the results of a fluorescence-activated cell sorting (FACS) analysis measuring GFP expression in HEK 293 cells transduced with either a (FIG. 2A) wild-type rAAV vector comprising the GFP transgene (WT rAAV) or (FIG. 2B) a modified version of the rAAV vector in which the GFP transgene is flanked by convergent human CTCF binding sites (hCTCF rAAV). These results show that 7.7% of the cells transduced with hCTCF rAAV

expressed GFP above background levels, whereas only 3.6% of the cells transduced with WT rAAV expressed GFP above background levels.

[0016] FIG. 3 shows the results of a quantitative reverse transcription PCR (RT-qPCR) analysis measuring GFP expression in HEK 293 cells transduced with either WT rAAV or hCTCF rAAV. Mock infected cells (mock) were also analyzed to serve as a negative control. Values were normalized to the levels of housekeeping gene Actb.

[0017] FIG. 4 is a schematic depicting the predicted outcomes of inserting CTCF binding sites into rAAV vectors in both the convergent and divergent orientations. When the CTCF binding sites are inserted in the convergent orientation, CTCF binding and dimerization brings together distal DNA elements and results in looping of the intervening sequence. Published data suggests that chromatin loops preferentially form between CTCF binding sites oriented in a convergent manner.

[0018] FIGS. 5A-5C is a schematic depicting the difference between an adeno associated virus (AAV) vector (FIG. 5A), a wild-type recombinant adeno-associated virus (rAAV) vector (FIG. 5B), and a modified rAAV vector (FIG. 5C). The triangles represent CTCF binding sites.

[0019] FIG. 6 is a schematic of rAAV vectors indicating the locations where the CTCF sites have been inserted (designated as 5' and 3'; corresponding the NheI and XhoI restriction enzyme sites). The flags indicate CTCF binding elements and their orientation (convergent or divergent) is shown by their direction.

[0020] FIGS. 7A-7E show the FACS analysis of 293T cells transduced for 24 hours with rAAV without insertions (FIG. 7A), and rAAV with CTCF inserts from H1 (FIG. 7B), MVM (FIG. 7C), human (FIG. 7D), and AAV (FIG. 7E). The cells were monitored for levels of GFP positivity. Live cells were first selected by gating on forward and side scatter, which were then assessed for GFP positivity.

[0021] FIG. 8 shows the number of GFP transcripts generated per input vector genome. This was computed from rAAV-transduced 293T cells for 24 hours using qRT-PCR. PCR primers were used to determine the ratio of GFP mRNA molecules to that of input vector genomes in the target cells.

DETAILED DESCRIPTION

[0022] The present disclosure provides constructs for producing modified recombinant adeno-associated virus (rAAV) vectors that have improved properties, including increased transgene expression. The constructs comprise one or more CCCTC-binding factor (CTCF) binding sites, which facilitate DNA looping and promote efficient transgene expression. Also provided are modified rAAV virus particles comprising these constructs, methods for producing the modified rAAV virus particles, and methods of using the modified rAAV virus particles to deliver a transgene to a subject.

[0023] Recombinant AAV (rAAV) vectors are the platforms of choice for gene therapy to express therapeutic transgenes, and have been designed from Adeno-Associated Viruses (AAVs), that are single-stranded DNA viruses'. Recombinant AAV gene therapy vectors have been designed from AAV parvoviruses by removing all genomic elements, retaining only the Inverted Terminal Repeats (ITRs), which are required to package the transgene in the vector capsid². AAV packaging signals. The resulting rAAV vectors do not

contain any of the transcriptional regulatory elements in AAV viruses that regulate AAV gene expression, and as a result do not regulate rAAV expression. This has led to the use of rAAV vectors at high doses in clinical settings.

[0024] Expression of genes on the human genome, and from DNA virus genomes, are regulated by packaging of DNA around histones, forming chromatin^{3, 4}. The accessibility of chromatin is regulated by barrier elements, bound by the host protein CTCF. Additionally, CTCF-bound cellular elements loop out intervening DNA molecules, regulating gene expression from promoters by facilitating their 3D interaction with enhancer elements³. The inventors have discovered that the AAV genome is folded into a distinct topological conformation akin to the three-dimensional (3D) structure of the eukaryotic genome, that formation of this 3D structure is required for efficient AAV gene expression, and that formation of the 3D structure is facilitated by binding of the transcription factor CCCTC-binding factor (CTCF) to regulatory elements in the AAV genome. In view of this discovery, the inventors have engineered novel modifications into an rAAV vector that facilitate the formation of 3D structures. Namely, they have introduced one or more binding sites for CTCF into the construct. As demonstrated in the Examples, the modified rAAV vectors drive at least two-fold higher levels of transgene expression in twice as many transduced target cells compared to their wild-type rAAV counterpart, providing surprisingly better transduction results. Thus, the use of the modified rAAV vectors of the present invention improve the use of AAV vectors in gene therapies by reducing the amount of vector that must be administered, which (1) decreases production costs and ultimately increases access to gene therapies, and (2) improves the safety of gene therapies by reducing the chances of oncogenic integration and toxicity.

Constructs

[0025] In a first aspect, the present disclosure provides constructs for producing a modified recombinant adeno-associated virus (rAAV) vector. The constructs comprise: a 5' inverted terminal repeat (ITR), a first CCCTC-binding factor (CTCF) binding site, a promoter, a transgene, and a 3' ITR.

[0026] Adeno associated viruses (AAV) are non-pathogenic viruses that belong to the genus Dependoparvovirus. AAV are small, nonenveloped viruses that have a linear single-stranded DNA genome that is approximately 4.7 kilobases (kb) in size. Their genomes encode two distinct sets of proteins: the non-structural replication (Rep) proteins, and the capsid (Cap) proteins that form the structure into which the genome is packaged (FIG. 5A). AAV viruses are replication defective, meaning that the production of AAV virus requires coinfection with helper virus(es). AAV offer several advantages for use as gene therapy vectors: AAV-based gene therapy vectors cause a very mild immune response, can infect both dividing and quiescent cells, and persist in an extrachromosomal state without integrating into the genome of the host cell.

[0027] As used herein, a "recombinant adeno-associated virus (rAAV) vector" is an AAV vector in which the Rep/Cap genes and their regulatory sequences have been replaced with a transgene, as depicted in FIG. 5B. As used herein, the term "modified rAAV vector" is used to describe an rAAV vector into which one or more CTCF binding sites has been introduced (FIG. 5C), whereas a "wild-type rAAV vector" is

an rAAV vector that lacks CTCF binding sites. Wild-type rAAV vectors genomes persist as linear DNA molecules in the host nucleus, and their expression is regulated solely by the transcriptional regulatory elements (e.g., promoters, enhancers) included in the vector. In contrast, CTCF binding to the CTCF binding sites included in the modified rAAV vectors results in recruitment of transcription factors and/or DNA looping, which can both facilitate more efficient transgene expression. The rAAV vectors of the present invention may comprise a sequence selected from: SEQ ID NOs:29-40, or a sequence having at least 90% identity to any one of SEQ ID NOs:29-40 (Table 4).

[0028] As used herein, the term “construct” refers to a recombinant polynucleotide, i.e., a polynucleotide that was formed artificially by combining at least two polynucleotide components from different sources (natural or synthetic). For example, the constructs described herein comprise the coding region of a transgene of interest operably linked to a promoter that (1) is associated with another gene found within the same genome, (2) from the genome of a different species, or (3) is synthetic. Constructs can be generated using conventional recombinant DNA methods. The constructs described herein are single stranded polynucleotides that comprise inverted terminal repeats on their 5' and 3' ends.

[0029] Constructs may be part of a vector. When referring to a nucleic acid molecule alone, the term “vector” is used herein to describe a nucleic acid molecule capable of transporting another nucleic acid to which it is linked. In contrast, the term “viral vector”, “AAV vector”, or “rAAV vector” is used to describe a virus particle that is used to deliver genetic material (e.g., the constructs of the present invention) into cells.

[0030] The constructs of the present invention comprise 5' and 3' inverted terminal repeats. “Inverted terminal repeats (ITRs)” are palindromic G-C-rich inverted repeats found on each end of the single stranded AAV genome, which self-base-pair to form unique AAV genome structures. ITRs contain several cis-acting elements that are involved in the initiation of viral DNA replication, as well as binding motifs for cellular transcription factors. Thus, the inclusion of ITRs in the constructs of the present invention allows the constructs to be incorporated into an AAV particle and replicated for viral production.

[0031] The constructs of the present invention also comprise one or more CCCTC-binding factor (CTCF) binding sites. CTCF is a transcription factor that regulates the 3D structure of chromatin. CTCF brings specific DNA loci together, forming chromatin loops. Because the 3D structure of DNA influences the regulation of genes, CTCF's activity influences the gene expression. For example, CTCF binding can bridge together promoters and transcription factor-bound enhancers to facilitate transcription initiation. In many cases, two CTCF proteins bound to distinct binding sites dimerize to bring together distal DNA elements. However, in some cases, a single CTCF binding site is sufficient for genome looping. For example, the single CTCF binding site found in the AAV2 genome forms a loop with a region found 2 kb downstream. In these single site instances, CTCF interacts with a different set of architectural proteins, i.e., cohesin and mediator.

[0032] As used herein, the term “CTCF binding site” refers to a region of DNA that comprises one or more CTCF binding sequences (i.e., DNA sequences to which CTCF

binds). The inventors have generated constructs in which one or more CTCF binding sites (e.g., a first and second CTCF binding site) each comprise five CTCF binding sequences. Thus, in some embodiments, the first and/or second CTCF binding site comprises multiple CTCF binding sequences. For example, the first and/or second CTCF binding site may comprise 2, 3, 4, 5, 6, 7, 8, 9, 10, or more CTCF binding sequences.

[0033] The constructs of the present invention also comprise a promoter. As used herein, the term “promoter” refers to a DNA sequence that regulates the transcription of a polynucleotide. Typically, a promoter is a regulatory region that is capable of binding RNA polymerase and initiating transcription of a downstream sequence. However, a promoter may be located at the 5' or 3' end, within a coding region, or within an intron of a gene that it regulates. Promoters may be derived in their entirety from a native gene, may be composed of elements derived from multiple regulatory sequences found in nature, or may comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, at different stages of development, or in response to different environmental conditions. A promoter is “operably linked” to a polynucleotide if the promoter is connected to the polynucleotide such that it may affect transcription of the polynucleotide.

[0034] The constructs of the present invention also comprise a transgene of interest. As used herein, the term “transgene” or “transgene of interest” refers to a gene or genetic material that one wishes to transfer into an organism or a cell thereof. A transgene may encode any protein or functional RNA of interest. Suitable transgenes include those that encode a therapeutic product. For example, the transgene may encode a protein that is lacking due to a genetic disorder or may encode a small interfering RNA (siRNA) that downregulates the expression of a protein that is overexpressed or ectopically expressed due to a genetic disorder. Any suitable transgene for use in gene therapy is contemplated for use in the present disclosure.

[0035] In the Examples, the inventors modified a wild-type rAAV vector comprising a green fluorescent protein (GFP) transgene operably linked to a cytomegalovirus (CMV) promoter by inserting a first CTCF binding site between the 5' ITR and the CMV promoter and a second CTCF binding site between the GFP transgene and the 3' ITR as depicted in FIG. 1. Thus, in some embodiments, the constructs further comprise a second CTCF binding site. In specific embodiments, the constructs comprise from 5' to 3': the 5' inverted terminal repeat (ITR), the first CCCTC-binding factor (CTCF) binding site, the promoter, the transgene, the second CTCF binding site, and the 3' ITR.

[0036] Convergence/divergence of the CTCF binding sites refers to a 5' to 3' directionality of CTCF protein binding, and does not refer to the palindromic or non-palindromic nature of the sequences. The inventors have generated constructs in which the two CTCF binding sites are in a convergent orientation as well as constructs in which the two CTCF binding sites are in a divergent orientation. As used herein, the term “convergent orientation” describes two CTCF binding sites that are oriented towards each other, and the term “divergent orientation” describes two CTCF binding sites that are oriented in the same direction or away from each other (see FIG. 4). Published data suggests that chromatin loops preferentially form between CTCF binding sites

oriented in a convergent manner. Thus, in some embodiments, the second CTCF binding site is in the convergent orientation relative to the first CTCF binding site.

[0037] The CTCF binding sites used in the constructs of the present invention can be from any organism. The inventors have identified a series of suitable CTCF binding sites that are natively found in humans and various viruses. The sequences of these binding sites are provided in Tables 2 and 3. Thus, in some embodiments, the CTCF binding site(s) are from a human (e.g., SEQ ID NOs: 1 and 2). In other embodiments, the CTCF binding site(s) are from a virus selected from the group consisting of adeno-associated virus (AAV; e.g., SEQ ID NO: 3), minute virus of mice (MVM; e.g., SEQ ID NOs: 4-6), H1 parvovirus (e.g., SEQ ID NOs: 7-9), mouse papillomavirus (MmuPV) (e.g., SEQ ID NO: 10), B19 (e.g., SEQ ID NO: 11), canine parvovirus (e.g., SEQ ID NO: 12), human cytomegalovirus (HCMV)/human herpesvirus 5 strain Merlin (e.g., SEQ ID NO: 13), human alphaherpesvirus 1 (e.g., SEQ ID NOs: 14-16), human herpesvirus 4 type 2 (Epstein-Barr virus type 2; e.g., SEQ ID NOs: 17-19), human papillomavirus (HPV)16 (e.g., SEQ ID NO: 20), and herpes B virus (HBV) (e.g., SEQ ID NOs: 21 and 22). In some embodiments, the CTCF binding site(s) comprise a sequence selected from: SEQ ID NOs:1-28, or a sequence having at least 90% identity to a sequence selected from: SEQ ID NOs:1-28.

[0038] In Example 1, the inventors inserted the human CTCF binding sequence of SEQ ID NO:1 into the 5' end of the rAAV construct and inserted the CTCF binding sequence of SEQ ID NO:2 into the 3' end in the convergent orientation. Thus, in some embodiments, the first CTCF binding site comprises SEQ ID NO:1 and the second CTCF binding site comprises SEQ ID NO:2. In one of the constructs of Example 2, the inventors inserted the AAV CTCF binding sequence of SEQ ID NO:3 into the 5' end of the rAAV construct. Thus, in some embodiments, the construct includes one CTCF binding site of SEQ ID NO:3. In another of the constructs of Example 2, the inventors inserted the human CTCF binding sequence of SEQ ID NO:1 into the 5' end of the rAAV construct and inserted the human CTCF binding sequence of SEQ ID NO:42 into the 3' end in the convergent orientation. Thus, in some embodiments, the first CTCF binding site comprises SEQ ID NO:1 and the second CTCF binding site comprises SEQ ID NO:42.

[0039] Protein and nucleic acid sequence identities are evaluated using the Basic Local Alignment Search Tool ("BLAST") which is well known in the art (Karlin and Altschul, 1990, Proc. Natl. Acad. Sci. USA 87: 2267-2268; Altschul et al., 1997, Nucl. Acids Res. 25: 3389-3402). The BLAST programs identify homologous sequences by identifying similar segments, which are referred to herein as "high-scoring segment pairs," between a query amino or nucleic acid sequence and a test sequence which is preferably obtained from a protein or nucleic acid sequence database. Preferably, the statistical significance of a high-scoring segment pair is evaluated using the statistical significance formula (Karlin and Altschul, 1990), the disclosure of which is incorporated by reference in its entirety. The BLAST programs can be used with the default parameters or with modified parameters provided by the user.

[0040] "Percentage of sequence identity" is determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions

or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison, and multiplying the result by 100 to yield the percentage of sequence identity.

[0041] The term "substantial identity" of polynucleotide sequences means that a polynucleotide comprises a sequence that has at least 85% sequence identity to the SEQ ID. Alternatively, percent identity can be any integer from 85% to 100%. More preferred embodiments include at least: 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% compared to a reference sequence using the programs described herein; preferably BLAST using standard parameters, as described. These values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide sequences by taking into account codon degeneracy, amino acid similarity, reading frame positioning, and the like.

[0042] "Substantial identity" of amino acid sequences for purposes of this invention normally means polypeptide sequence identity of at least 85%. Preferred percent identity of polypeptides can be any integer from 85% to 100%. More preferred embodiments include at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%.

Host Cells, Virus Particles, and Packaging Cell Lines

[0043] In a second aspect, the present invention provides host cells transduced with a construct described herein. As used herein, the term "host cell" refers to any prokaryotic or eukaryotic cell that contains a construct of the present invention. This term also includes cells that have been genetically engineered such that a construct of the present invention is integrated into its genome. The host cell can be a cell line that is used for producing the AAV vectors for use as a gene therapy. Suitable host cells include mammalian cells, including human cells.

[0044] The terms "transduced," "transfected," and "transformed" all refer to processes by which an exogenous nucleic acid is introduced into a host cell. The term "transduced" specifically refers to the process by which a virus transfers a nucleic acid into a host cell. Plasmids may be used to transfect the construct into a host cell for AAV production along with the helper viruses.

[0045] In a third aspect, the present invention provides rAAV virus particles comprising a construct described herein. As used herein, the term "virus particle" refers to a virion consisting of nucleic acid surrounded by a protective protein coat called a capsid.

[0046] To generate viral particles, the constructs comprising the rAAV vector are cloned into a plasmid for expression in a host cell. Viral particles may then be generated by helper virus-free co-transfection of HEK 293T cells with three plasmids: (1) an AAV vector comprising a construct of the present invention, (2) a packaging plasmid carrying the AAV Rep and Cap genes, and (3) a helper plasmid carrying the AAV helper functions. For a detailed description of viral production methods, see Ayuso et al. (Gene Ther 17(4):503-

10, 2010), which is hereby incorporated by reference in its entirety. Other suitable methods for producing AAV virus particles are well known and understood in the art.

[0047] In a fourth aspect, the present invention provides packaging cell lines for producing the virus particles described herein. The term “packaging cell line” is used to refer to a cell line that provides all the proteins necessary for AAV virus production and maturation. Suitable packaging cell lines for use with the present invention include, without limitation, mammalian cells and human cell lines. For example, suitable cell lines include, but are not limited to, HEK 293T cells and HEK 293 cell variants. The packaging cell line should be selected with the method of viral production in mind. For example, cells that have strong adhesion properties should be selected for growth in culture plates, whereas cells lacking adhesion properties should be selected for growth in suspension culture. In some embodiments, the packaging cell line comprises the complement of any genes that have been functionally deleted in the virus particle used to produce the virus, allowing replication incompetent viral particles to be produced.

Method for Producing a Modified rAAV Virus Particle

[0048] In a fifth aspect, the present invention provides methods for producing a modified rAAV virus particle. The methods comprise: (a) transducing a host cell with a plasmid comprising a construct described herein, a packaging plasmid, and a helper plasmid; (b) collecting the supernatant and the cells from culture; and (c) isolating virus particles from the supernatant and cells.

[0049] A “plasmid” is a small circular DNA molecule that can replicate independently from chromosomal DNA. In nature, plasmids are commonly found in bacteria, and artificial plasmids are widely used as vectors in molecular cloning.

[0050] In the present methods, host cells (e.g., packaging cell lines) are transfected with three plasmids: a plasmid comprising a construct described herein, a packaging plasmid, and a helper plasmid. The term “packaging plasmid” refers to a plasmid that encodes components of the AAV proteins. For rAAV production, the packaging plasmid may encode the AAV genes Rep and Cap. The term “helper plasmid” refers to a plasmid that encodes adenovirus helper functions. Proteins encoded by all three plasmids that are transfected into the host cell in the present methods are required for rAAV production and AAV replication, as is well known in the art.

[0051] Virus can be isolated from the supernatant and/or from lysed cells by methods known and understood in the art. Suitable methods for isolating virus from cell culture include, but are not limited to, cesium chloride density gradient centrifugation and affinity purification (e.g., using a porous matrix modified to retain the virus).

[0052] In some embodiments, the methods further comprise concentrating the virus. Suitable methods for concentrating virus include, but are not limited to, ultracentrifugation and dialysis.

[0053] In some embodiments, the methods further comprise dialyzing the supernatant. For some applications, it may be advantageous to replace the cell culture media present in the supernatant with a solution that is better for long-term storage. Suitable solutions for storage include, but are not limited to, phosphate-buffered saline (PBS), PBS

with plutonic acid, saline adjusted to pH 7-7.4 with or without pluronic acid (0.001-0.01%), and Ringer's lactate solution. However, any biocompatible, osmotically balanced, neutral pH fluid should be suitable for storage.

Method for Delivering a Transgene

[0054] In a sixth aspect, the present invention provides methods of delivering a transgene to a subject in need thereof. The methods comprise: administering a modified rAAV virus particle described herein to the subject. By “delivering a transgene” we mean that the methods result in transgene expression in one or more of the subject's cells.

[0055] As used herein, the term “administering” refers to any method of providing a pharmaceutical preparation to a subject. Such methods are well known to those skilled in the art and include, but are not limited to, oral administration, transdermal administration, administration by inhalation, nasal administration, topical administration, intravaginal administration, ophthalmic administration, intraaural administration, intracerebral administration, rectal administration, sublingual administration, buccal administration, and parenteral administration, including injectable such as intravenous administration, intra-arterial administration, intramuscular administration, intradermal administration, intrathecal administration, and subcutaneous administration. Administration can be continuous or intermittent. In some embodiments, the virus particle is administered by vascular injection.

[0056] In some embodiments, the virus particle is administered with a pharmaceutically acceptable carrier. “Pharmaceutically acceptable carriers” are known in the art and include, but are not limited to, diluents, preservatives, solubilizers, emulsifiers, liposomes, nanoparticles, and adjuvants. Pharmaceutically acceptable carriers may be aqueous or non-aqueous solutions, suspensions, and emulsions. Examples of nonaqueous solvents are propylene glycol, polyethylene glycol, vegetable oils such as olive oil, and injectable organic esters such as ethyl oleate. Aqueous carriers include isotonic solutions, alcoholic/aqueous solutions, emulsions, and suspensions, including saline and buffered media.

[0057] Ideally, the virus particles are administered in a therapeutically effective amount. The term “therapeutically effective amount” refers to an amount sufficient to effect beneficial or desirable biological or clinical results. Methods for determining an effective means of administration and dosage are well known to those of skill in the art and will vary with the formulation used for therapy and the subject (e.g., species, age, health, etc.) being treated. Single or multiple administrations can be carried out with the dose level and pattern being selected by the treating physician. In some embodiments, the virus particle is administered at a dose of 1×10^{12} viral genome/kg (vg/kg) or less.

[0058] In the Examples, the inventors demonstrate that the introduction of two convergent CTCF binding sites into an rAAV vector comprising a GFP transgene causes the vector to drive GFP expression at higher levels and in a greater proportion of transduced cells. Thus, in some embodiments, the transgene is expressed in a greater proportion of the subject's cells when it is delivered in the modified rAAV vector as compared to when it is delivered in a wild-type rAAV vector. For example, the transgene may be expressed in 1.5 times, 2 times, 3 times, 4 times, or 5 times as many cells as compared to with a wild-type rAAV vector. In some

embodiments, the transgene is expressed at higher levels when it is delivered in the modified rAAV vector as compared to when it is delivered in a wild-type rAAV vector. For example, the transgene may be expressed at 1.5 times, 2 times, 3 times, 4 times, or 5 times the level that it is expressed at a wild-type rAAV vector.

[0059] Transgene expression can be detected using any suitable method known in the art. For example, when the transgene encodes a protein, the protein product may be detected using an enzyme-linked immunoassay (ELISA), dot blot, western blot, flow cytometry, mass spectrometry, or chromatographic method. When the transgene encodes a functional RNA, the RNA product may be detected using reverse transcription and polymerase chain reaction (RT-PCR) or Northern blotting.

[0060] It should be apparent to those skilled in the art that many additional modifications besides those already described are possible without departing from the inventive concepts. In interpreting this disclosure, all terms should be interpreted in the broadest possible manner consistent with the context. Variations of the term “comprising” should be interpreted as referring to elements, components, or steps in a non-exclusive manner, so the referenced elements, components, or steps may be combined with other elements, components, or steps that are not expressly referenced. Embodiments referenced as “comprising” certain elements are also contemplated as “consisting essentially of” and “consisting of” those elements. The term “consisting essentially of” and “consisting of” should be interpreted in line with the MPEP and relevant Federal Circuit interpretation. The transitional phrase “consisting essentially of” limits the scope of a claim to the specified materials or steps “and those that do not materially affect the basic and novel characteristic(s)” of the claimed invention. “Consisting of” is a closed term that excludes any element, step or ingredient not specified in the claim. For example, with regard to sequences “consisting of” refers to the sequence listed in the SEQ ID NO. and does refer to larger sequences that may contain the SEQ ID as a portion thereof.

[0061] As used in this specification and the claims, the singular forms “a,” “an,” and “the” include plural forms unless the context clearly dictates otherwise. Thus, the indefinite articles “a” and “an,” as used herein in the specification and in the claims should be understood to mean “at least one”, unless clearly indicated to the contrary.

[0062] As used herein, “about”, “approximately”, “substantially”, and “significantly” will be understood by persons of ordinary skill in the art and will vary to some extent on the context in which they are used. If there are uses of the term which are not clear to persons of ordinary skill in the art given the context in which it is used, “about” and “approximately” will mean up to plus or minus 10% of the particular term and “substantially” and “significantly” will mean more than plus or minus 10% of the particular term. Where ranges are stated, the endpoints are included within the range unless otherwise stated or otherwise evident from the context.

[0063] The phrase “such as” should be interpreted as “for example, including.” Moreover the use of any and all exemplary language, including but not limited to “such as”, is intended merely to better illuminate the invention and does not pose a limitation on the scope of the invention unless otherwise claimed.

[0064] In those instances where a convention analogous to “at least one of A, B and C, etc.” is used, in general such a construction is intended in the sense of one having ordinary skill in the art would understand the convention (e.g., “a system having at least one of A, B and C” would include but not be limited to systems that have A alone, B alone, C alone, A and B together, A and C together, B and C together, and/or A, B, and C together.). It will be further understood by those within the art that virtually any disjunctive word and/or phrase presenting two or more alternative terms, whether in the description or figures, should be understood to contemplate the possibilities of including one of the terms, either of the terms, or both terms. For example, the phrase “A or B” will be understood to include the possibilities of “A” or “B” or “A and B.” Multiple elements listed with “and/or” should be construed in the same fashion, i.e., “one or more” of the elements so conjoined. Other elements may optionally be present other than the elements specifically identified by the “and/or” clause, whether related or unrelated to those elements specifically identified.

[0065] As used herein in the specification and in the claims, “or” should be understood to have the same meaning as “and/or” as defined above. For example, when separating items in a list, “or” or “and/or” shall be interpreted as being inclusive, i.e., the inclusion of at least one, but also including more than one, of a number or list of elements, and, optionally, additional unlisted items. Only terms clearly indicated to the contrary, such as “only one of” or “exactly one of,” or, when used in the claims, “consisting of,” will refer to the inclusion of exactly one element of a number or list of elements. In general, the term “or” as used herein shall only be interpreted as indicating exclusive alternatives (i.e. “one or the other but not both”) when preceded by terms of exclusivity, such as “either,” “one of,” “only one of,” or “exactly one of” “Consisting essentially of,” when used in the claims, shall have its ordinary meaning as used in the field of patent law.

[0066] All language such as “up to,” “at least,” “greater than,” “less than,” and the like, include the number recited and refer to ranges which can subsequently be broken down into ranges and subranges. A range includes each individual member. Thus, for example, a group having 1-3 members refers to groups having 1, 2, or 3 members. Similarly, a group having 6 members refers to groups having 1, 2, 3, 4, or 6 members, and so forth.

[0067] The modal verb “may” refers to the preferred use or selection of one or more options or choices among the several described embodiments or features contained within the same. Where no options or choices are disclosed regarding a particular embodiment or feature contained in the same, the modal verb “may” refers to an affirmative act regarding how to make or use and aspect of a described embodiment or feature contained in the same, or a definitive decision to use a specific skill regarding a described embodiment or feature contained in the same. In this latter context, the modal verb “may” has the same meaning and connotation as the auxiliary verb “can.”

[0068] In the foregoing description, it will be readily apparent to one skilled in the art that varying substitutions and modifications may be made to the invention disclosed herein without departing from the scope and spirit of the invention. The invention illustratively described herein suitably may be practiced in the absence of any element or elements, limitation or limitations which is not specifically

disclosed herein. The terms and expressions which have been employed are used as terms of description and not of limitation, and there is no intention that in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the invention. Thus, it should be understood that although the present invention has been illustrated by specific embodiments and optional features, modification and/or variation of the concepts herein disclosed may be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of this invention. No language in the specification should be construed as indicating any non-claimed element as essential to the practice of the invention. It should be understood that descriptions of exemplary embodiments are not intended to limit the invention to the particular forms disclosed, but on the contrary, the intention is to cover all modifications, equivalents and alternatives falling within the spirit and scope of the invention as defined by the appended claims. It should also be understood that, unless clearly indicated to the contrary, in any methods claimed herein that include more than one step or act, the order of the steps or acts of the method is not necessarily limited to the order in which the steps or acts of the method are recited. All methods described herein can be performed in any suitable order unless otherwise indicated herein or otherwise clearly contradicted by context.

[0069] The invention will be more fully understood upon consideration of the following non-limiting examples.

EXAMPLES

Example 1

[0070] In the following Example, the inventors introduce binding sites for the chromatin loop-forming protein CCCTC binding factor (CTCF) into an rAAV vector at sites flanking the rAAV transgene (FIG. 1). They demonstrate that the introduction of these CTCF binding sites enhances transgene expression.

Generation of Modified rAAV Vectors

[0071] Human CTCF: The most prevalent CTCF binding site in the human genome was previously identified via chromatin immunoprecipitation sequencing (ChIP-Seq) (Rao et al., Cell, 2015). This study identified pairs of CTCF binding sites that facilitate genome looping, and generated the forward consensus CTCF sequence 5'-CCACNAGGTGGCAG-3' (SEQ ID NO:24) and the reverse consensus CTCF sequence 5'-CTGC-CACCTNGTGG-3' (SEQ ID NO:25). The inventors cloned CTCF binding sequences into an rAAV plasmid comprising a GFP transgene operably linked to a CMV promoter, which was obtained from Addgene (rAAV-GFP; plasmid #105530). Specifically, a human forward CTCF binding sequence (5'-CCACAAGGTGGCGC-3'; SEQ ID NO:1) was inserted in the 5' end of the rAAV vector between the 5' ITR and the CMV promoter, at base pair 205 of the positive-sense strand. A human reverse CTCF binding sequence (5'-CCACCAGGGGGCGG-3'; SEQ ID NO:2) was inserted just downstream of the 3' ITR, at base pair 2477 of the negative-sense strand, in the convergent orientation. Additionally, a human reverse CTCF binding sequence (5'-GGCGGGGACCACC-3'; SEQ ID NO:26) was inserted in

the divergent orientation at that same location. The sequences of the constructs were confirmed via sequencing (Functional Biosciences).

[0072] Viral CTCF: The wild-type AAV2 genome was screened for the presence of CTCF binding sites using the in-silico prediction tool JASPAR (Stormo et al., Quant. Biol, 2013). The inventors discovered that wild-type AAV has a native CTCF binding site (5'-TTGCGACAC-CATGTGGTCA-3'; SEQ ID NO:3) at the 5' end of the AAV genome positioned between the 5' ITR and the p5 promoter (base pairs 166-185) on the positive-sense strand. The inventors detected CTCF binding at this site using ChIP-qPCR. They then inserted this CTCF sequence into the rAAV genome between the 5' ITR and the CMV promoter, at base pair 205 of the positive-sense strand. They also generated a reverse CTCF binding sequence from this native AAV sequence (i.e., by generating the reverse complement of this sequence) and inserted it in rAAV just upstream of the 3' ITR, at base pair 2477 of the negative-sense strand. In the convergent orientation, this reverse CTCF sequence is 5'-AACGCTGTGGTACACCAGT-3' (SEQ ID NO:27) and in the divergent orientation, this sequence is 5'-TGACCA-CATGGTGTGCGAA-3' (SEQ ID NO:28).

Results

[0073] HEK 293 cells were transduced with rAAV vectors comprising a green fluorescent protein (GFP) transgene. The cells were transduced with either a wild-type rAAV vector (comprising no CTCF sequences) or a modified rAAV vector comprising convergent human CTCF binding sequences (Forward: 5'-CCACAAGGTGGCGC-3' (SEQ ID NO:1); Reverse: 5'-CCACCAGGGGGCGG-3' (SEQ ID NO:2)) at an MOI of 2,500 viral genomes/cell for 24 hours. A negative control of mock infected cells was used. Cells were collected and resuspended in PBS buffer, and a fluorescence-activated cell sorting (FACS) analysis was performed to measure transgene (i.e., GFP) expression on the cell surface of live cells. The results of this analysis suggest that the addition of the CTCF binding sequences enhances the efficiency transgene expression, as a greater number of the cells transduced with the modified rAAV vector were GFP-positive as compared to the cells transduced with the wild-type rAAV vector (FIG. 2).

[0074] The samples were then subjected to RNA extraction, and qRT-PCR was performed using primers that amplify GFP transcripts to quantify transgene expression. The expression levels were normalized to the levels of the housekeeping gene Actb, and the relative GFP expression levels were compared. The inventors found that GFP expression was significantly enhanced in the cells that were transduced with the modified rAAV vector as compared to the wild-type rAAV vector, suggesting a causal relationship between the CTCF binding sites and increased transgene expression (FIG. 3).

Example 2

Additional Viral CTCF Binding Sites

[0075] Other DNA viruses and viruses in the parvovirus family have native CTCF binding sites. These CTCF sequences may be able to facilitate looping in rAAVs. For example, minute virus of mice (MVM) is a parvovirus that contains a validated CTCF binding site that is involved in

RNA processing and gene expression (Viruses 12(12): 1368, 2020). Many CTCF binding sites have been identified in Kaposi's sarcoma-associated herpesvirus (KSHV) and a single CTCF site has been identified in both Epstein-Barr virus (EBV) and human papillomavirus (HPV). Additionally, the herpes simplex virus type 1 (HSV-1) genome contains CTCF-mediated looping structures that are induced by viral infection.

[0076] The inventors have cloned CTCF binding site sequences from other parvovirus genomes (e.g. MVM, AAV subtypes, H1 parvovirus, CPV, and B19) and well-characterized DNA viruses (e.g. HCMV, HSV, EBV, HPV and HBV) into the rAAV vector to determine their impact on transgene expression. The CTCF insertion sequences used in the vectors are listed in Table 1.

Identification of Cis-Elements

[0077] The inventors have scanned the genomes of DNA viruses, including parvoviruses such as AAV, MVM, H1, B19, CPV, as well as herpesviruses such as EBV, HSV, HCMV and tumor viruses such as HPV16 and HBV, to identify CTCF binding sites in-silico using the JASPAR online database of transcription factor binding sites⁵. These online screens identified the viral CTCF binding elements in DNA viruses. The inventors additionally identified published CTCF sites on the human genome that have been previously identified using CTCF ChIP-seq genome-wide³. Generation of rAAV Vectors Containing Chromatin Modifying Cis-Elements

[0078] The inventors cloned the identified CTCF binding elements into the 5' end of the rAAV vector expressing a GFP transgene from a CMV promoter as shown in FIG. 6 (labelled as 5' insert into the NheI restriction enzyme site). They additionally cloned these CTCF sequences into the 3' CTCF insert site, downstream of the poly-A tail (labelled as 3' CTCF insert into the XhoI restriction enzyme site). These sequence orientations were varied according to their forward version (labelled as F in Table 1) and in the reverse orientation (labelled as R in Table 1). A subset of the sequence inserts contained multiple CTCF binding elements (designated by multiple F's and R's in Table 1). Convergent CTCF orientations in Table 1 are labelled as "con" and non-convergent CTCF orientations are designated as "noncon".

TABLE 1

Location and sequences of rAAV gene therapy containing vector constructs TCF insertions that have been successfully generated			
Vector	CTCF site	5' CTCF insert sequence (position 200 of SEQ ID NO: 29, on the origin NheI site)	3' CTCF insert sequence (position 2472 of SEQ ID NO: 29, on the XhoI site)
<u>AAV</u>			
<u>CTCF</u>			
1	5F_C2	TTGCGACACCATGTGGTCA (SEQ ID NO: 3)	
2	5F_3R con	TTGCGACACCATGTGGTCA (SEQ ID NO: 3)	ACTGGTGTACCACAG CGTT (SEQ ID NO: 41)

TABLE 1-continued

Location and sequences of rAAV gene therapy containing vector constructs TCF insertions that have been successfully generated			
Vector	CTCF site	5' CTCF insert sequence (position 200 of SEQ ID NO: 29, on the origin NheI site)	3' CTCF insert sequence (position 2472 of SEQ ID NO: 29, on the XhoI site)
3	5F_3F noncon	TTGCGACACCATGTGGTCA (SEQ ID NO: 3)	TTGCGACACCATGTG GTCA (SEQ ID NO: 3)
<u>hCTCF</u>			
4	5F_C1	CCACAAGGTGGCGC (SEQ ID NO: 1)	
5	5F_3R con	CCACAAGGTGGCGC (SEQ ID NO: 1)	CCGCCCCCTGGTGG (SEQ ID NO: 42) (reverse complement of SEQ ID NO: 2)
6	5F_3F noncon	CCACAAGGTGGCGC (SEQ ID NO: 1)	GGTGGTCCCCGCC (SEQ ID NO: 43)
<u>MVM#4</u>			
7	5F	TTGCTCACTAGATGGCGCT C (SEQ ID NO: 44)	
8	5R	CTCGCGGTAGATCACTCGT T (SEQ ID NO: 4)	
<u>MVM#5</u>			
9	5F	CCACCACTAAATGGCATT C TT (SEQ ID NO: 5)	
10	5R	TTCTTACGGTAAATCACCA CC (SEQ ID NO: 45)	
11	5F_3F	CCACCACTAAATGGCATT C TT (SEQ ID NO: 5)	TTCTTACGGTAAATC ACCACC (SEQ ID NO: 46)
12	5F_3R	CCACCACTAAATGGCATT C TT (SEQ ID NO: 5)	CCACCACTAAATGGC ATTCTT (SEQ ID NO: 5)
<u>H1</u>			
<u>Parvo</u>			
13	5F	AGTCCACCAAGGGACGGAG (SEQ ID NO: 9)	
14	5F_3F	AGTCCACCAAGGGACGGAG (SEQ ID NO: 9)	GAGGCAGGGAACCAC CTGA (SEQ ID NO: 47)
15	5F_3R	AGTCCACCAAGGGACGGAG (SEQ ID NO: 9)	AGTCCACCAAGGGAC GGAG (SEQ ID NO: 9)
16	5F_3RF	AGTCCACCAAGGGACGGAG (SEQ ID NO: 9)	AGTCCACCAAGGGAC GGAGGAGGCAGGGAA CCACTGA (SEQ ID NO: 48)

TABLE 1-continued

Location and sequences of rAAV gene therapy containing vector constructs TCF insertions that have been successfully generated			
Vector site number	CTCF origin	5' CTCF insert sequence (position 200 of SEQ ID NO: 29, on the NheI site)	3' CTCF insert sequence (position 2472 of SEQ ID NO: 29, on the XhoI site)
17	5R	GAGGCAGGGAACCCACCTGA (SEQ ID NO: 49)	
	HBV CTCF 1		
18	5RR	ACCTAGGTTGACCACCAGC ACCTAGGTTGACCACCAGC (SEQ ID NO: 50)	
19	5RRR	ACCTAGGTTGACCACCAGC ACCTAGGTTGACCACCAGC ACCTAGGTTGACCACCAGC (SEQ ID NO: 51)	
	HBV CTCF 2		
20	5F	TCTACAGCATGGGGCAGAA (SEQ ID NO: 22)	
	HBV CTCF 1		
21	5RF	ACCTAGGTTGACCACCAGC CGACCACCAGTTGGATCCA (SEQ ID NO: 52)	
22	5RFPR	ACCTAGGTTGACCACCAGC CGACCACCAGTTGGATCCA CGACCACCAGTTGGATCCA CGACCACCAGTTGGATCCA (SEQ ID NO: 53)	
	EBV CTCF2		
23	5FFFFF	CACCCAACAGGTGGTGAAA CACCCAACAGGTGGTGAAA CACCCAACAGGTGGTGAAA CACCCAACAGGTGGTGAAA CACCCAACAGGTGGTGAAA (SEQ ID NO: 54)	
24	5RRRRR R	AAAGTGGTGGACAACCCAC AAAGTGGTGGACAACCCAC AAAGTGGTGGACAACCCAC AAAGTGGTGGACAACCCAC AAAGTGGTGGACAACCCAC AAAGTGGTGGACAACCCAC (SEQ ID NO: 55)	
	EBV CTCF3		
25	5R	ACCGTGGTGTACCACGGTT (SEQ ID NO: 56)	

TABLE 1-continued

Location and sequences of rAAV gene therapy containing vector constructs TCF insertions that have been successfully generated			
Vector site number	CTCF origin	5' CTCF insert sequence (position 200 of SEQ ID NO: 29, on the NheI site)	3' CTCF insert sequence (position 2472 of SEQ ID NO: 29, on the XhoI site)
26	5FF	TTGGCACCATGTGGTGCCA TTGGCACCATGTGGTGCCA (SEQ ID NO: 57)	
	CPV CTCF		
27	5F	CAACCAGGAGGTGAAAATC (SEQ ID NO: 12)	
28	5R	CTAAAAGTGGAGGACCAAC (SEQ ID NO: 58)	
29	5FF	CAACCAGGAGGTGAAAATC CAACCAGGAGGTGAAAATC (SEQ ID NO: 59)	
	HSV CTCF2		
30	5F	CCACCGCGGGGGGGCGCG (SEQ ID NO: 15)	
31	5FF	CCACCGCGGGGGGGCGCG CCACCGCGGGGGGGCGCG (SEQ ID NO: 60)	
32	5FFF	CCACCGCGGGGGGGCGCG CCACCGCGGGGGGGCGCG CCACCGCGGGGGGGCGCG (SEQ ID NO: 61)	
	B19		
33	5F	ATACTGGGGATAACCACC (SEQ ID NO: 62)	
34	5FF	ATACTGGGGATAACCACC ATACTGGGGATAACCACC (SEQ ID NO: 63)	
	HCMV		
35	5F	CGACCCGCACATGGCGCTG (SEQ ID NO: 13)	
	HPV16		
36	5F	TAACCACCAGGTGGTGCCA (SEQ ID NO: 20)	
37	5RR	ACCGTGGTGGACCACCAAT ACCGTGGTGGACCACCAAT (SEQ ID NO: 64)	

Vector Production

[0079] rAAV vectors were produced in HEK 293T cells by cotransfecting them with Rep/Cap plasmids (expressing AAV Rep and Cap proteins) and pHelper plasmids (expressing essential Adenovirus proteins such as E1, E2, E4ORF6 and VA-RNA) for 6-7 days. Vectors were harvested from the producer cells by rapid freeze/thaw cycles, DNase treated

and transduced into target 293T cells⁶. These cells were assessed for GFP expression by FACS and qRT-PCR as described below.

Assessing the Number of Cells That are GFP Positive

[0080] rAAV-GFP vectors were used to transduce HEK 293T cells for 24 hours. They were subsequently monitored for GFP positivity using FACS analysis after gating on the live cells by forward scatter and side scatter. As shown in FIG. 7A, 44.4% of cells transduced with the wild-type rAAV vector without insertions were GFP positive at 24 hpi. Cells transduced with vectors containing the CTCF inserts from H1 parvovirus and MVM parvovirus were respectively 1.8% and 1.1% GFP positive (FIG. 7B; corresponding to Vector number 13 in Table 1 and FIG. 7C; corresponding to Vector number 7 in Table 1). Cells transduced with human CTCF binding elements in a convergent orientation were 12% GFP positive (FIG. 7D, corresponding to Vector number 5 in Tables 1 and 4 (SEQ ID NO: 37)). Strikingly, cells transduced with an rAAV vector containing the AAV CTCF binding element were 51% GFP positive (FIG. 7E, corresponding to Vector number 1 in Tables 1 and 4 (SEQ ID NO: 29)). Notably, the H1, MVM and AAV CTCF insertions were in the 5' end of the genome only, whereas the hCTCF

insertion was at both ends of the genome in a convergent orientation. These findings suggested that CTCF binding site insertions modulate the transduction efficiency of gene therapy vectors in the host.

Vector Expression Efficiency per Vector Genome

[0081] To monitor the ability of novel rAAV gene therapy vectors to express in transduced cells, the inventors normalized the GFP transcript levels generated in target cells to that of input vector genomes. They compared the mRNA molecules per input vector in the current iteration of rAAV vectors to that of the novel constructs, focusing on the constructs containing the convergent hCTCF sites (Vector number 5 in Table 1) and the AAV CTCF sites (Vector number 1 in Table 1). Compared with the current rAAV vectors, rAAV^{AAV-CTCF} yielded similar levels of GFP mRNA per vector whereas rAAV^{hCTCF} vectors expressed at double these levels (FIG. 8). These findings indicate that the CTCF binding elements other than those derived from AAV in rAAV vectors are able to increase the expression capacity of rAAV genomes in individual cells as well as increase the number of cells capable of expressing the rAAV genome.

[0082] Examples of other viral CTCF binding sequences include those listed in Tables 2 and 3. Sequences of rAAV vectors are provided in Table 4.

TABLE 2

Native CTCF binding site sequences found in human and viral genomes				
SEQ ID NO:	CTCF sequence (5' to 3')	Organism	Location in genome	Strand (Fwd/Rev)
1	CCACAAGGTGGCGC	Human	chr1	Fwd
2	CCACCAGGGGGCGG	Human	chr1	Rev
3	TTGCGACACCATGTGGTCA	AAV	166	Fwd
4	CTCGCGGTAGATCACTCGT T	Minute virus of mice (MVM)	990	Rev
5	CCACCCTAAATGGCATTC TT	Minute virus of mice (MVM)	4460	Fwd
6	CTCGCGGTAGATCACTCGT T	Minute virus of mice (MVM)	990	Rev
7	TGCTCACTAGATGGCGCTC	H1 parvovirus	900	Rev
8	TGCACAGCAGAGGACTCTG	H1 parvovirus	2682	Fwd
9	AGTCCACCAAGGGACGGAG	H1 parvovirus	828	Fwd
10	TGAACAGTAGGAGTCAGTT	MmuPV	2966	Fwd
11	CCACCAATAGGGGGTCATA	B19	1516	Rev
12	CAACCAGGAGGTGAAAATC	Canine parvovirus	1220	Fwd
13	CGACCCGCACATGGCGCTG	HCMV; Human herpesvirus 5 strain Merlin	2205	Fwd
14	AGACCACAGGTGGCGCAC	Human alphaherpesvirus 1	1068	Rev
15	CCACCAGGGGGGGGGCGG	Human alphaherpesvirus 1	655	Rev

TABLE 2-continued

Native CTCF binding site sequences found in human and viral genomes				
SEQ ID NO:	CTCF sequence (5' to 3')	Organism	Location in genome	Strand (Fwd/Rev)
16	CGGGCGGCCGGGGGGGGCG	Human alphaherpesvirus 1	543	Rev
17	TGGCCAAAAGACGGCGGTT	Human herpesvirus 4 type 2 (Epstein-Barr virus type 2)	2226	Fwd
18	CACCCAACAGGTGGTGAAA	Human herpesvirus 4 type 2 (Epstein-Barr virus type 2)	2020	Fwd
19	TTGGACCATGTGGTGCCA	Human herpesvirus 4 type 2 (Epstein-Barr virus type 2)	407	Fwd
20	TAACCACCAGGTGGTGCCA	HPV16	2053	Fwd
21	CGACCACCAGTTGGATCCA	HBV	1079	Fwd
22	TCTACAGCATGGGGCAGAA	HBV	1027	Fwd

TABLE 3

Other CTCF binding site sequences used in the Examples		
SEQ ID NO:	CTCF sequence (5' to 3')	Description
24	CCACNAGGTGGCAG	Human forward CTCF consensus sequence
25	CTGCCACCTNGTGG	Human reverse CTCF consensus sequence
26	GGCGGGGACCACC	Human reverse CTCF sequence in the divergent orientation (i.e., SEQ ID NO: 2 in reverse order)
27	AACGCTGTGGTACA CCAGT	AAV Reverse CTCF sequence, generated for use the native AAV CTCF sequence, convergent orientation
28	TGACCACATGGTGT CGCAA	AAV Reverse CTCF sequence, generated for use the native AAV CTCF sequence, divergent orientation (i.e., SEQ ID NO: 27 in reverse order)

TABLE 4

rAAV sequences. The inserted CTCF binding sites are underlined.	
SEQ ID NO:	Description and Sequence
29	rAAV with forward AAV CTCF sequence knocked in on positive-sense strand (Vector 1, Table 1) CCCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGC GACCTTTGGTCGCCCGGCCTCAGTGAGCGAGCGAGCGCAGAGAGGGA GTGGCCAACTCCATCACTAGGGGTTCTTGTAGTTAATGATTAACCCGC CATGCTACTTATCTACGTAGCCATGCTCTAGGAAGATCGGAATTCGCCC TTAAGCTAGTTGCGACACCATGTGGTCACTAGCTAGTTATTAATAGTAA TCAATTACGGGGTCATTAGTTTCATAGCCCATATATGGAGTTCGCGGTTA CATAACTTACGGTAAATGGCCCGCTGGCTGACCGCCCAACGACCCCG CCCATTGACGTCATAATGACGTATGTTCCCATAGTAACGCCAATAGGG ACTTTCATTGACGTCATGGGTGGAGTATTTACGGTAAACTGCCCACT TGGCAGTACATCAAGTGTATCATATGCAAGTACGCCCCCTATTGACGT

TABLE 4-continued

RAAV sequences. The inserted CTCF binding sites are underlined.

SEQ ID NO:	Description and Sequence
	<p>CAATGACGGTAATGGCCCGCTGGCATTATGCCAGTACATGACCTTA TGGGACTTTCCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTA CCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTT TGACTCACGGGGATTTCCTAAGTCTCCACCCATTGACGTCATGGGAGT TTGTTTTGGCACC AAAATCAACGGGACTTTCCAAAATGTCGTAACAACT CCGCCCATGACGCAAAATGGGCGTAGGCGTGTACGGTGGGAGGCTTA TATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCTGCAGAAGTTGGTC GTGAGGCACTGGGCGGTAAGTATCAAGGTTACAAGACAGGTTTAAGGA GACCAATAGAAAATGGGCTTGTGAGACAGAGAAGACTCTTGCGTTTCT GATAGGCACCTATTGGTCTTACTGACATCCACTTTGCTTTCTCTCCAC AGGTGTCCAGGCGGCCGCGCATGGTGAGCAAGGGCGAGGAGCTGTTACC GGGTGGTGCCCATCTGGTTCGAGCTGGACGGCGACGTAACCGGCCACA AGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCT GACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCC ACCTCGTGACACCCTGACCTACGGCGTGCAGTGCTTACGGCGTACC CCGACCACATGAAGCAGCAGCACTTCTTCAAGTCCGCGATGCCCGAAGG CTACGTCCAGGAGCGCACCATCTTCTCAAGGACGACGGCAACTACAAG ACCCGCGCGAGGTGAAGTTCGAGGGCGACACCTGGTGAACCGCATCG AGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAA GCTGGAGTACAACATAACAGCCACAACGCTCTATATCATGGCCGACAAG CAGAAGAACGGCATCAAGGTGAACCTCAAGTCCGCCACAACATCGAGG ACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCATCGG CGACGGCCCGTGTCTGCTGCCCGACAACTACCTGAGCACCACGATCC GCCCTGAGCAAGACCCCAACGAGAAGCGCGATCACATGGTCTTGTCTGG AGTTCGTGACCCGCGCCGGGATCACTCTCGGCATGACGAGCTGTACAA GATAAAGCTTGGATCCAATCAACCTCTGGATTACAAAATTTGTGAAAG ATTGACTGGTATTCTTAACATAGTGTCTCCTTTTACGCTATGTGGATA GCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCA TTTTCTCCTCCTTGTATAAATCCTGGTGTCTGCTCTTTATGAGGAGT GTGGCCCGTTGTGAGCAACGTTGGCGTGGTGTGCACTGTGTTGTGTC GCAACCCCACTGGTGGGGCATTGCCACCACCTGTAGCTCCTTTCCG GGACTTTGCTTTCCCTCCTTATTGGCACGGCGGAACATCGCCCGC CTGCTTTGCCCGTGTGAGCAGGGGCTCGGCTGTGGGCACTGACAA TCCGTTGGTGTGTCGGGAAATCATGCTCCTTTCTTGGCTGCTCGCCT GTGTTGCCACCTGGATTCTGCGCGGACGCTCCTTCTGCTACGCTCCTTC GGCCCTCAATCCAGCGGACCTTCTTCCGCGGCGTGTGCGGCTGTG CGGCTCTTCCGCGTCTTCGAGATCTGCCTCGACTGTGCTTCTAGTTG CCAGCCATCTGTTGTTTGGCCCTCCCGTGCCTTCTTGCACCTGGAA GGTGCCACTCCCACTGTCTTTCCATAAAAATGAGGAAATGCACTGCG ATTGCTGAGTAGGTGTATTCTATTCTGGGGGTGGGGTGGGGCAGGA CAGCAAGGGGAGGATTGGGAAGACAATAGCAGGCATGTGGGGACTCG ATCGAGTTAAGGGCGAATTCGATAAGGATCTTCTAGAGCATGGCTA CGTAGATAAGTAGCATGGCGGGTTAATCATTAACACAGGAACCCCTA GTGATGGAGTTGGCCACTCCTCTCTGCGCGCTCGCTCGCTCACTGAG CCGGGCGACCAAGGTGCGCCGACGCGGGCTTTGCCCGGGCGGCTC AGTGAGCGAGCGAGCGCGCAG</p>
30	<p>RAAV with reverse AAV CTCF sequence knocked in on positive-sense strand CCGCTCGCTCGCTACTGAGCCGCGCCGGGCAAAGCCCGGGCGTGGGGC GACCTTTGGTCCCGGCCCTCAGTGAGCGAGCGAGCGCAGAGAGGGA GTGGCAACTCCATCACTAGGGGTTCCTTGTAGTTAATGATTAACCCCG CATGCTACTTATCTACGTAGCCATGCTCTAGGAAGATCGGAATTCGCC TTAAGCTAGCTAGCTAGTTATTAATAGTAAATCAATTACGGGGTCATTAG TTCATAGCCATATATGGAGTTCGGCGTTACATAACTTACGGTAAATGG CCGCGCTGGCTGACCGCCCAACGACCCCGCCCATGACGTCATAAATG ACGTATGTTCCCATAGTAACGCCAATAGGGACTTCCATGACGTCAT GGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTA TCATATGCCAAGTACGCCCCCTATTGACGTCATGACGGTAAATGGCCC GCCTGGCATTATGCCAGTACATGACCTTATGGGACTTTCCTACTTGGC AGTACATCTACGTATTAGTATCGCTATTACCATGGTGTGCGGTTTTG CGAGTACATCAATGGGCGTGGATAGCGGTTTGAATCACGGGGATTTC AGTCTCCACCCATTGACGTCATGGGAGTTTGTGTTGGCACC AAAATC AACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCATGACGCCAAAT GGCGGTAGGCGTGTACGGTGGGAGGTCATATAAGCAGAGCTGGTTTA GTGAACCGTCAGATCTGCAGAAGTTGGTGTGAGGCACTGGGCGAGTA AGTATCAAGGTACAGACAGGTTTAAGGAGACCAATAGAAAATGGGCT TGTCGAGACAGAGAAGACTCTTGCCTTCTGATAGGCACCTATTGGTCT TACTGACATCCACTTTGCCTTCTCTCCACAGGTGTCAGGGCGGCCG ATGGTGAAGCAAGGGCGAGGAGCTGTTCCCGGGGTGGTGCCTCCTGG TCGAGCTGGACGGCGACGTAACGGCCACAAGTTCAGCGTGTCCGGCGA GGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCGAAAGTTTCATCTGC</p>

TABLE 4-continued

rAAV sequences. The inserted CTCF binding sites are underlined.

SEQ ID NO:	Description and Sequence
	<p>ACCACCGCAAGCTGCCCGTCCCTGGCCACCCCTCGTGACCACCCCTGA CCTACGGCGTGCAAGTCTTCAGCCGCTACCCCGACCATGAAGCAGCA CGACTTCTCAAGTCCGCCATGCCGGAAGGCTACGTCAGGAGCGCAC ATCTTCTCAAGGACGCGCAACTACAAGACCCGCGCGAGGTGAAGT TCGAGGGCGACACCCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTT CAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACTACA AGCCACAACGCTTATATCATGGCCGACAAGCAGAGAAGACGGCATCAAG TGAACCTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGC CGACCACTACCAGCAGAACACCCCATCGGCAGCGGCCCGTGTCTGTG CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCA ACGAGAAGCGCGATCACATGGTCTGTGGAGTTCGTGACCGCCCGCGG GATCACTCTCGGCATGGACGAGCTGTACAAGTAATAAGCTTGGATCCAA TCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAAC TATGTTGCTCCTTTTACCGTATGTGGATACGCTGCTTTAATGCCTTTGT ATCATGCTATTGCTTCCGATATGGCTTTTATTTCTCCTCTGTATAAA ATCCTGGTTGCTGTCTCTTTATGAGGAGTTGGCCCGTTGTCAGGC CGTGGCGTGGTGTGCACTGTGTTGCTGACGCAACCCCACTGGTTGGG GCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTCCCCCT CCCTATTGCCACGGCGGAACCTCATCGCCGCTGCCTTGCCCGCTGTGG ACAGGGGCTCGGCTGTGGGCACTGACAATCCGTGGTGTGTGCGGGGA AATCATCGTCTCTTCTGGTGTCTGCTGTGGTGGCCACCTGGATTCT GCGCGGACGTCCTTCTGCTACGTCCTTCCGGCCCAATCCAGCGGAC CTTCTTCCCGCGGCTGTGCGCGCTCTGCGGCTCTTCCGCGCTTTC GAGATCTGCCTCGACTGTGCTTCTAGTTGCCAGCCATCTGTTGTTGTC CCCTCCCCCGTGCCTTCTTGACCTGGAAGGTGCCACTCCCACTGTCC TTTCTAATAAATGAGGAAATGTCATCGCATGTCTGAGTAGGTTGTC TTCTATTCTGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGG GAAGACAATAGCAGGCATGCTGGGACTCGAAGTGGTGTACACAGCGT TTGAGTTAAGGGCGAATCCCGATAAGGATCTTCTAGAGCATGGCTA CGTAGATAAGTAGCATGGCGGGTTAATCATTAACTACAAGGAACCCCTA GTGATGGAGTTGGCCACTCCTCTCTGCGCGCTCGCTCACTGAGG CGGGCGACCAAGGTCGCCCAGCGCCCGGGCTTTGCCCGGGCGGCTC AGTGAGCGAGCGAGCGCGCAG</p>
31	<p>rAAV with AAV CTCF convergent sequence knocked in on positive-sense strand CCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAGCCCGGGCGTCCGGC GACCTTTGGTCCGCCCTCAGTGAGCGAGCGAGCGCGAGAGAGGGA GTGGCCAACTCATCACTAGGGGTTCTTGTAGTTAATGATTAACCCGC CATGCTACTTATCTACGTAGCCATGCTCTAGGAAGATCGAATTCGCC TTAAGCTAGTTGCGACACCATGTGGTCACTAGCTAGTTATTAATAGTAA TCAATTACGGGGTCAATTAGTTATAGCCCATATATGGAGTTCCGCGTTA CATAACTTACGGTAAATGGCCCGCTGGCTGACCGCCCAACGACCCCG CCCATGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGG ACTTTCCATTGACGTCAATGGGTGGAGTATTACGGTAAACTGCCCACT TGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGT CAATGACGTTAAATGGCCCGCTGGCATTATGCCAGTACATGACCTTA TGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTA CCATGGTGTATGCGGTTTGGCAGTACATCAATGGGCGTGGATAGCGGT TGACTCACGGGATTTCCAAGTCTCCACCCATTGACGTCAATGGGAGT TTGTTTTGGCACAAAATCAACGGGACTTCCAAAATGTCGTAACAAC CCGCCCATTTGACGCAAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTA TATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCTGCAAGAAGTTGGT GTGAGGCACTGGGAGGTAAGTATCAAGTTACAAGACAGGTTTAAGGA GACCAATAGAACTGGGCTTGTGAGACAGAGAAGACTCTTGCCTTCT GATAGGCACCTATTGGTCTTACTGACATCCACTTTGCCTTCTCTCCAC AGGTGTCAGGCGGCCCGCATGGTGAAGGCGAGGAGCTGTTTACC GGGTTGGTGCCTATCTGGTGCAGCTGGACGGCGAGTAAACGGCCACA AGTTCAAGCTGTCCGGCAGGGCGAGGGCGATGCCACTACGGCAAGT GACCTGAAGTTCATCTGACACCGGGCAAGCTGCCCGTGCCTGGCCC ACCCCTGTCACCCCTGACCTACGGCGTGCAGTCTTACGGCGTACC CCGACACATGAAAGCAGCACGACTTCTTCAAGTCCGCGATGCCGAAG CTACGTCAGGAGCGCACCATCTTCTCAAGGACGACGGCAACTACAAG ACCCCGCGGAGGTGAAGTTCGAGGGCGACACCTGGTGAACCCGATCG AGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCTGGGGCACA GCTGGAGTACAACATAACAGCCACAACGCTTATATCATGGCCGACAAG CAGAAGAACGGCATCAAGGTGAACCTCAAGATCCGCCACAACATCGAGG ACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCATCGG CGACGGCCCGTGTCTGCTCCCGACAACCACTACCTGAGCACCCAGTCC GCCCTGAGCAAAGACCCCAACGAGAAGCGCATCACATGGTCTGTCTGG AGTTCGTGACCGCGCGGGATCACTCTCGGCATGGACGAGCTGTACAA GTAATAAGCTTGGATCCAATCAACCTCTGGATTACAAAATTTGTGAAAG</p>

TABLE 4-continued

rAAV sequences. The inserted CTCF binding sites are underlined.	
SEQ ID NO:	Description and Sequence
	<p>ATTGACTGGTATTCTTAACATGTTGCTCCTTTTACGCTATGTGGATAC GCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCA TTTTCTCCTCCTTGTATAAATCCTGGTTGCTGCTCTTTATGAGGAGTT GTGGCCGTTGTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTTTGGCTGAC GCAACCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCG GGACTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACATCGCCGC CTGCCCTTGCCCGTCTGGACAGGGGCTCGGCTGTGGGCACTGACAA TCCGTGGTGTGTCGGGGAAATCATCGTCTTTCCCTGGCTGCTCGCCT GTGTTGCCACCTGGATTCTGCGCGGGACGCTCCTTCTGCTACGTCCTTC GGCCCTCAATFCAGCGGACCTTCCCTCCGCGGCCCTGCTGCCGGCTCTG CGGCCTCTTCCGCGTCTTCGAGATCTGCCTCGACTGTGCCTTCTAGTTG CCAGCCATCTGTTGTTTGGCCCTCCCCCGTGCCTTCCCTGACCCCTGGAA GGTGCCACTCCCACTGCTTCTTAATAAAAATGAGGAAATGCAATCGC ATTGCTGAGTAGGTGTCATTCTATTCTGGGGGTGGGGGGGCGAGGAC AGCAAGGGGGAGGATTGGGAAGACAAATAGCAGGCATGCTGGGACTCGA ACTGGTGTACCCACAGCGTTTCGAGTTAAGGGCGAATTCGATAAGGAT CTTCTTAGAGCATGGCTACGTAGATAAGTAGCATGGCGGGTTAATCATT AACTACAAGGAACCCCTAGTGTAGGAGTTGGCCACTCCTCTCTGCGCG CTCGCTCGCTCACTGAGGCGGGCGACCAAAGGTCGCCGACGCGCCGGG CTTGGCCGGGCGGCTCAGTGAGCGAGCGAGCGCGCAG</p>
32	<p>rAAV with AAV CTCF divergent sequence knocked in on positive sense strand</p> <p>CCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGC GACCTTTGGTTCGCCCGGCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGA GTGGCAACTCCATCACTAGGGGTCTCCTGTAGTTAATGATTAACCCCG CATGCTACTTATCTACGTAGCCATGCTCTAGGAAGATCGGAATTCGCC TTAAGCTAGTTGCGACACCATGTGGTCACTAGCTAGTTATTAATAGTAA TCAATTACGGGGTCATTAGTTTCATAGCCCATATATGGAGTTCCGCGTTA CATAACTTACGGTAAATGGCCGCTGGCTGACCGCCCAACGACCCCG CCCATGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGG ACTTCCATTGACGTCAATGGGTGGAGTATTACGGTAAACTGCCCACT TGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCTTATTGACGT CAATGACGGTAAATGGCCCGCTGGCATATGCCAGTACATGACCTTA TGGGACTTTCCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTA CCATGGTGTATGCGGTTTGGCAGTACATCAATGGCGGTGGATAGCGGTT TGACTCACGGGGATTTCCAAGTCTCCACCCATTGACGTCAATGGGAGT TTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAAC CTGCCCCATTGACGCAAAATGGCGGTAGGCGTGTACGGTGGGAGGTCTA TATAAGCAGAGCTGGTTAGTGAACCGTCAGATCCTGCAGAAGTTGGTC GTGAGGCACTGGGCGGTAAGTATCAAGGTTACAAGACAGGTTAAGGA GACCAATAGAAAATGGGCTTGTGAGACAGAGAAGACTCTTGCCTTTCT GATAGGCACCTATTGGTCTTACTGACATCCACTTTGCTTTCTCTCCAC AGGTGTCAGGCGGCCCATGGTGGTGGCAGGCGAGGAGCTGTTTCC CGGGTGGTGCCCATCTGGTTCGAGCTGGACGGCAGCTAAACGGCCACA AGTTCAGCGTGTCCGGCAGGGCGAGGGCGATGCCACTACGGCAAGT GACCCGTAAGTTTCTGACACCACCGGCAAGCTGCCCGTGCCTGGCC ACCCCTGTCGACCCCTGACCTACGGCGTGCAGTCTTACGCGCTACC CCGACCATGAAAGCAGCAGCTTCTCAAGTCCGCCATGCCCGAAGG CTACGTCAGGAGCGCACCATCTTCTCAAGGACGACGGCAACTACAAG ACCCGCGCGAGGTGAAGTTCGAGGGCGACCCCTGGTGAACCGCATCG AGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCTGGGGCACAA GCTGGAGTACAACATAACAAGCCACAACGCTTATATCATGGCCGACAAG CAGAAGAACGGCATCAAGGTGAACCTCAAGATCCGCCACAACATCGAGG ACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACCCCCATCCGG CGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCAGTCC GCCCTGAGCAAGACCCCAACGAGAAGCGCATCAATGGTCTGCTGG AGTTTCGTGACCCGCGCGGGATCACTCTCGGCATGGACGAGCTGTACAA GTAATAAGCTTGGATCCAATCAACCTCTGGATTACAAAATTTGTGAAAG ATTGACTGGTATTCTTAACATGTTGCTCCTTTTACGCTATGTGGATAC GCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCA TTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTT GTGGCCCGTGTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTTGCTGAC GCAACCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCG GGACTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACATCGCCCG CTGCCTTGCCCGTGTGGACAGGGGCTCGGCTGTGGGCACTGACAA TCCGTGGTGTGTCGGGGAAATCATCGTCTTTCCCTGGCTGCTCGCCT GTGTTGCCACCTGGATTCTGCGCGGGACGCTCCTTCTGCTACGTCCTTC GGCCCTCAATFCAGCGGACCTTCCCTCCGCGGCCCTGCTGCCGGCTCTG CGGCCTCTTCCGCGTCTTCGAGATCTGCCTCGACTGTGCCTTCTAGTTG CCAGCCATCTGTTGTTTGGCCCTCCCCCGTGCCTTCTTGCACCTGGAA GGTGCCACTCCCACTGCTTCTTAATAAAAATGAGGAAATGCAATCGC</p>

TABLE 4-continued

rAAV sequences. The inserted CTCF binding sites are underlined.

SEQ ID NO:	Description and Sequence
	<p>ATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGTGGGGCAGGA CAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGACTCG ATTGCGACACCATGTGGTCATCGAGTTAAGGGCGAATTCGCCGATAAGGA TCTTCTAGAGCATGGCTACGTAGATAAGTAGCATGGCGGGTAAATCAT TAACTACAAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGC GCTCGCTCGCTCACTGAGGCCGGGGACCAAAGGTCGCCCGACGCCCGC GCTTTGCCGGGGCGCCTCAGTGAGCGAGCGAGCGCGCAG</p>
33	<p>rAAV with AAV CTCF repeat convergent sequence knocked in on positive-sense strand</p> <p>CCGCTCGCTCGCTCACTGAGGCCGCCGGGCAAGCCCGGGCGTGGGGC GACCTTTGGTCGCCCCGCCCTCAGTGAGCGAGCGAGCGCGAGAGAGGGA GTGGCCAACCTCCTACTAGGGGTTCTTGTAGTTAATGATTAACCCGC CATGCTACTTATCTACGTAGCCATGCTCTAGGAAGATCGAATTCGCC TTAAGTAGTGGCGACACCATGTTGGTCAATGGCAGACCATGTTGGTCATT <u>GCGACACCATGTTGGTCAATGGCAGACCATGTTGGTCAATGGCAGACCATG</u> <u>TGGTCACTAGCTAGTTAATAAGTAATCAATACGGGGTCAATAGTTC</u> ATAGCCCATATATGGAGTTCGCCGTACATAACTTACGGTAAATGGCCC GCCTGGCTGACCGCCCAACGACCCCGCCCATGACGTCAATAATGACG TATGTTCCCATAGTAACGCCAATAGGGACTTCCATTGACGTCAATGGG TGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCA TATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGC TGGCATTATGCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGT ACATCTACGTATTAGTCATCGCTATTACCATGGTGTATGCGGTTTGGCA GTACATCAATGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGT CTCCACCCCATGACGTCAATGGGAGTTTGTTTTGGCACAAAATCAAC GGGACTTCCAAAATGTCGTAACAACCTCCGCCCATGACGCAATGGG CGGTAGGCGTGTACGGTGGGAGGTTATATAAGCAGAGCTGGTTTATGTG AACCGTCAGATCCTGCAGAAGTTGGTCGTGAGGCACTGGGCAGGTAAGT ATCAAGGTTACAGACAGGTTTAAAGGAGACCAATAGAACTGGGCTTGT CGAGACAGAGAAGACTCTTGCCTTCTGATAGGCACCTATTGGTCTTAC TGACATCCACTTTGCCTTTCTCTCACAGGTGTCCAGGCGGCCCGCATG GTGAGCAAGGGCGAGGAGCTGTTCACCGGGTGGTGCCTCATCTGGTCCG AGCTGGACGGCGACGTAACCGGCCACAAGTTCAGCGTGTCCGGCGAGGG CGAGGGCGATGCCACCTACGGCAAGCTGACCCCTGAAGTTCATCTGCACC ACCGCAAGCTGCCCGTGCCTTGGCCACCCCTCGTGACCACCCCTGACCT ACGGCGTGCAGTCTTACGCCGCTACCCCGACCATGAAGCAGCACGA CTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCAGGAGCGCACCATC TTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCCG AGGGGACACCCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTCAA GAGGACCGGCAACATCCTGGGGCACAAGCTGGAGTACAACACAGC CACAACTGCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGA ACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGA CCACTACCAGCAGAACACCCCATCGGCGACGGCCCGTGTCTGTGCC GACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAGAACCCCAACG AGAAGCGCGATCACATGGTCCGTGCTGGAGTTCGTGACCGCCCGGGAT CACTCTCGCATGGACGAGCTGTACAAGTAATAAGCTTGGATCCAATCA ACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTAT GTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATC ATGCTATTGCTTCCCGTATGGCTTTTCAATTTTCTCCTCTTGTATAAATC CTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTGAGGCAACGT GGCGTGGTGTGCACTGTGTTTGTGACGCAACCCCACTGGTTGGGGCA TTGCCACCACTGTGACGCTCTTCCGGGACTTTCGCTTCCCCCTCCC TATTGCCACGGCGGAACCTACGCGCCTGCGCTTGGCCGCTGCTGGACA GGGGCTCGGCTGTGGGCACTGACAATTCGGTGGTGTGTCGGGGAAAT CATGCTCCTTCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGGC CGGGACGCTCTTCTGCTACGCTCCCTTCGGCCCTCAATCCAGCGGACCTT CCTTCCCGGGCCTGCTGCGGCTCTGCGCCCTTTCGCGCTTTCGAG ATCTGCCTCGACTGTGCTTCTAGTTGCCAGCCATCTGTGTTTGGCCC TCCCCCGTGCCTTCTTACCCCTGGAAGGTGCCACTCCCACTGCTCCTTT CCTAATAAAATGAGGAAATTCATCGCAATGCTGAGTAGGTGTCATTCT TATTCTGGGGGTGGGTGGGGCAGGACAGCAAGGGGGAGGATGGGAA GACAATAGCAGGCATGCTGGGACTCGAAGTGGTGTACACAGCGTTAC TGGTGTACACAGCGTTACTGGTGTACACAGCGTTACTGGTGTACAC AGCGTTACTGGTGTACACAGCGTTTTCGAGTTAAGGGCAATTCGCGAT AAGGATCTTCTTAGAGCATGGCTACGTAGATAAGTAGCATGGCGGGTTA ATCATTAACCTACAAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTC TGCGCGCTCGCTCGCTCACTGAGGCCGGGGACCAAAGTTCGCCCGACG CCCGGGCTTTGCCCGGGCGCCTCAGTGAGCGAGCGAGCGCGCAG</p>

TABLE 4-continued

rAAV sequences. The inserted CTCF binding sites are underlined.

SEQ ID NO: Description and Sequence

34 rAAV with AAV CTCF forward on positive-sense and reverse on negative-sense strand, convergent
 CCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGC
 GACCTTTGGTCGCCCGGCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGA
 GTGGCCAACCTCCTACTAGGGGTTCTCTGTAGTTAATGATTAACCCGC
 CATGCTACTTATCTACGTAGCCATGCTCTAGGAAGATCGGAATTCGCC
 TTAAGCTAGTTGCGACACCATGTGGTCACTAGCTAGTTATTAATAGTAA
 TCAATTACGGGGTCATTAGTTTCATAGCCCATATATGGAGTTCGCGTTA
 CATAACTTACGGTAAATGGCCGCTGGCTGACCGCCCAACGACCCCG
 CCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGG
 ACTTCCATTGACGTCAATGGGTGGAGTATTACGGTAAACTGCCCACT
 TGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCTATTGACGT
 CAATGACGGTAAATGGCCCGCTGGCATATGCCAGTACATGACCTTA
 TGGGACTTTCCACTTGGCAGTACATCTACGTATTAGTCATCGCTATTA
 CCATGGTGTATGCGGTTTGGCAGTACATCAATGGCGGTGGATAGCGGTT
 TGACTCAGCGGGATTTCAGTCTCCACCCATTGACGTCAATGGGAGT
 TTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAAC
 TCGCCCCATTGACGCAAAATGGCGGTAGGCGTGTACGGTGGGAGGTCTA
 TATAAGCAGAGCTGGTTAGTGAACCGTCAGATCCGCAAGAGTTGGTC
 GTGAGGCACTGGGCAGGTAAGTATCAAGGTTACAAGACAGGTTAAGGA
 GACCAATAGAACTGGGCTTGTGAGACAGAGAAGACTTTCGCTTTCT
 GATAGGCACCTATTGGTCTTACTGACATCCACTTTGCCCTTCTCTCCAC
 AGTGTCCAGGCGGCCCATGGTGGCAGGCGAGGAGCTGTTTCCAC
 GGGGTGGTGCCTCCTGGTCGAGCTGGACGGCAGCTAAACGGCCACA
 AGTTGACGCTGTCCGGCAGGGCGAGGGCGATGCCACTACGGCAAGCT
 GACCCCTGAAGTTTACTGTGACACCGGCCAAGCTGCCCGTCCCTGGCC
 ACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGTTCAGCCGCTACC
 CCGACCATGAAAGCAGCAGCACTTCTCAAGTCCCGCATGCCGGAAGG
 CTACGTCAGGAGCGCACCATCTTCTCAAGGACGACGGCAACTACAAG
 ACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCG
 AGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACCTCGGGGCACAA
 GCTGGAGTACAACCTACAACAGCCACAACGTCTATATCATGGCCGACAAG
 CAGAAGAACGGCATCAAGGTGAACCTCAAGATCCGCCACAACATCGAGG
 ACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAAACCCCATCGG
 CGACGGCCCCGTGCTGTGCCCCGACAACCCTACCTGAGCACCAGTCC
 GCCCTGAGCAAGACCCCAACGAGAAGCGCATCAATGGTCTGCTGCG
 AGTTGCGTACCGCCCGGGGATCACTCTCGGCATGGACGAGCTGTACAA
 GTAATAAGCTTGGATCCAATCAACCTCTGGATTACAAAATTTGTGAAAG
 ATTGACTGGTATTCTTAACATATGTGTCTCTTTTACGCTATGTGGATAC
 GCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCA
 TTTTCTCCCTTGTATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTT
 GTGGCCCGTGTGAGCAACGTGGCGTGGTGTGCACTGTGTTGTGTCGAC
 GCAACCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCG
 GACTTTTCGCTTTCCCTCCCTATTGCAACGGCGGAACCTCATCGCCG
 CTGCCTTGCCTGCTGACAGGGGCTCGGCTGTGGGCACTGACAAT
 TCCGTGGTGTGTCGGGAAATCATGCTCTTCTTCTGGTGTGCTGCTG
 GTGTTGCCACCTGGATTCTGCGCGGACGCTCCTTCTGCTACGTCCTTC
 GGCCCTCAATCCAGCGGACCTTCCCTCCCGCGGCTGCTGCCGCTCTG
 CGCCCTCTCCCGCTCTTCGAGATCTGCCCTCGACTGTGCTTCTAGTTG
 CCAGCCATCTGTTGTTGCCCCCTCCCGTGCCTTCTTACCCCTGGAA
 GGTGCCACTCCCACTGCTTCCCTAATAAAATGAGGAAATGCAATCGC
 ATTGCTGAGTAGGTGTCAATCTATTCTGGGGGTGGGGTGGGGCAGGA
 CAGCAAGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGACTCG
 ATGACCACATGGTGTGCAATCGAGTTAAGGGCGAATCCCGATAAGGA
 TCTTCTAGAGCATGGCTACGTAGATAAGTAGCATGGCGGGTAAATCAT
 TAACACAGGAACCCCTAGTGTAGGAGTTGGCCACTCCCTCTCTGCGC
 GCTCGCTCGCTCACTGAGGCCGGCGACCAAAGGTCCGCCGACGCCCGG
 GCTTTGCCCGGGCGGCTCAGTGAGCGAGCGAGCGCGCAG

35 rAAV with forward human CTCF sequence knocked in on positive-sense strand
 CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTC
 GGGCGACCTTTGGTCGCCCGGCTCAGTGAGCGAGCGAGCGCGCAGAGA
 GGGAGTGGCCAACCTCCTACTAGGGGTTCTCTGTAGTTAATGATTAAC
 CGCCCATGCTACTTATCTACGTAGCCATGCTCTAGGAAGATCGGAATTC
 GCCCTTAAGCTAGCCACAAGGTGGCGCTAGCTAGTTATTAATAGTAA
 CAATTACGGGGTCAATAGTTTCATAGCCCATATATGGAGTTCGCGGTTAC
 ATAACCTACGGTAAATGGCCCGCTGGCTGACCGCCCAACGACCCCGC
 CCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGA
 CTTTCCATTGACGTCAATGGGTGGAGTATTACGGTAAACTGCCCACTT
 GGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCTTATTGACGTC

TABLE 4-continued

rAAV sequences. The inserted CTCF binding sites are underlined.

SEQ ID NO:	Description and Sequence
	<p>AATGACGGTAAATGGCCCGCCTGGCATTATGCCAGTACATGACCTTAT GGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTAC CATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTT GACTCACGGGGATTTCCAAGTCTCCACCCATTGACGTCAATGGGAGTT TGTTTTGGCACAAAATCAACGGGACTTTCCAAAATGTCGTAACAACCTC CGCCCATTTGACGCAAAATGGGCGGTAGGCGTGTACGGTGGGAGGCTAT ATAAGCAGAGCTGGTTTGTGAACCGTCAGATCCTGCAGAAGTTGGTCG TGAGGCATGGGCAGGTAAGTATCAAGGTTACAAGACAGGTTTAAAGGAG ACCAATAGAACTGGGCTTGTGAGACAGAGAAGACTCTTGCCTTCTG ATAGGCACCTATTGGTCTTACTGACATCCACTTTGCCTTCTCTCCACA GGTGTCCAGGCGGCCCATGGTGAGCAAGGGCGAGGAGCTGTTCCACCG GGTGGTGCCTTCTGGTGCAGCTGGACGGCGAGTAAACGGCCACAA GTTACGCGTGTCCGGCAGGGCGAGGGCGATGCCACCTACGGCAAGCTG ACCTGAAGTTCATCTGCACCACGGCAAGCTGCCCGTGCCTGGCCCA CCCTCGTGACCACCTGACCTACGGCGTGCAGTGTTCAGCCGCTACCC CGACCACATGAAGCAGCAGCACTTCTCAAGTCCGCCATGCCCGAAGGC TACGTCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGA CCCGCCGAGGTTGAAGTTCGAGGGCGACACCTGGTGAACCGCATCGA GCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCTGGGGCACAAG CTGGAGTACAACCTACAACAGCCACACGCTATATCATGGCCGACAAGC AGAAGAACGGCATCAAGGTGAAGTTCAGAGTCCGCCACAACATCGAGGA CGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACAACCCCATCGGC GACGGCCCGTGTGTGCTGCCGACAACTACTACCTGAGCACCCAGTCCG CCCTGAGCAAGACCCCAACGAGAAGCGCATCACATGGTCTGCTGGA TTCTGTGACCGCCCGGGATCACTCTCGGCATGGACGAGCTGTACAAG TAATAAGCTTGGATCCAATCAACCTCTGGATTACAAAATTTGTGAAA TTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACG CTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCAT TTTCTCCTCTGTATAAATCTGGTGTGCTCTCTTTATGAGGAGTTG TGGCCCGTGTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTTTGTGACG CAACCCCACTGGTTGGGCAATGCCACCACCTGTGCTCCTTTCCGG GACTTTCGCTTTCCTCCTCCTATTTGCCAGCGGAACTCATCGCCGCT TGCTTTGCCCGTGTGTCGACAGGGCTCGGCTGTTGGGCATGACAATT CCGTGGTGTGTGTCGGGAAATCATCGCTCTTTCTTGGCTGCTCGCCTG TGTGCCACCTGGATTCTGCGCGGGACGCTCTTCTGCTACGCTCCTTTCG GCCCTCAATCCAGCGGACCTTCTTCCGCGCCTGCTGCCGCTCTG GGCTCTTCCGCTCTTCCGAGATCTGCCTCGACTGTGCTTCTAGTTGC CAGCCATCTGTGTTTGGCCCTCCCGGCTCCTTCTGACCTGGAAG GTGCCACTCCCCTGCTCCTTCTAATAAAATGAGGAAATTCATCGCA TTGTCTGAGTAGGTGTCATCTATTCTGGGGGTGGGTGGGGCAGGAC AGCAAGGGGAGGATGGGAAGACAATAGCAGGCATGCTGGGACTCGA TCGAGTTAAGGGCAATTCCTGATAAGGATCTTCTAGAGCATGGCTAC GTAGATAAGTAGCATGGCGGGTAAATCATTAACTACAAGGAACCCCTAG TGATGGAGTTGGCCACTCCTCTCTGCGCGCTCGCTCGCTACTGAGG CGGGCAGCAAAGGTCGCCGACGCCCGGGCTTTGCGGGCGGCCTCA GTGAGCGAGCGAGCGCGAG</p>
36	<p>rAAV with reverse human CTCF sequence knocked in on positive-sense strand CTGCGCGCTCGCTCGCTCACTGAGGCGCCCGGCAAGCCCGGGCGTC GGCGCACCTTTGGTCGCCCCGCTCAGTGAGCGAGCGAGCGCAGAGA GGGAGTGGCCAATCCATCACTAGGGGTTCCCTGTAGTTAATGATTAAC CGCCATGCTACTTATCTACGTAGCCATGCTCTAGGAAGATCGAATTC GCCCTTAAGCTAGCTAGCTAGTTATTAATAGTAATCAATACGGGTCA TTAGTTCAAGCCATATATGGAGTTCCGCGTTACATAACTTACGGTAA ATGGCCCGCTGGCTGACCGCCCAACGACCCCGCCATTGACGTCAT AATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTCCATTGACGT CAATGGGTGGAGTATTTACGGTAAACTGCCACTTGGCAGTACATCAAG TGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATG GCCCGCTGGCATATGCCCAGTACATGACCTTATGGGACTTCTCTACT TGGCAGTACATCTACGTATTAGTCAICGCTATTACATGGTGTGTCGGT TTTGGCAGTACATCAATGGGCGTGGATAGCGGTTGACTCACGGGGATT TCCAAGTCTCCACCCATTGACGTCAATGGGAGTTGTTTGGCACCAA AATCAACGGGACTTTCCAAAATGTCGTAACAACCTCCGCCCATTGACG AAATGGGCGGTAGGCGTGTACGGTGGGAGGCTATATAAGCAGAGCTGG TTTAGTGAACCGTCAGATCTCGACAAGTTGGTGTGAGGCCTGGGGAG GGTAAGTATCAAGGTTACAAGACAGGTTTAAAGGAGACCAATAGAACTG GGCTGTGAGACAGAGAAGACTTTCGCTTCTGTATAGGCACCTATTG GTCTTACTGACATCCACTTTGCCTTCTCTCCACAGGTTGTCAGGCGGC CGCCATGGTGGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGGCCATC CTGGTCGAGCTGGACGGCGAGCTAAACGGCCACAAGTTACGCGTGTCCG CGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCTGAAGTTCAT</p>

TABLE 4-continued

rAAV sequences. The inserted CTCF binding sites are underlined.	
SEQ ID NO:	Description and Sequence
	<p>CTGCACCACCGGCAAGCTGCCCGTGGCCCTGGCCACCCCTCGTGACCACC CTGACCTACGGCGTGCAAGTGCCTTCAGCCGCTACCCCGACCATGAAGC AGCAGGACTTCTTCAAGTCCGCCATGCCGAAGGCTACGTCCAGGAGCG CACCATCTTCTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTG AAGTTCGAGGGCGACACCCCTGGTGAACCGCATCGAGCTGAAGGGCATCG ACTTCAAGGAGGACGGCAACATCCTGGGCGACAAGCTGGAGTACAACATA CAACAGCCACAACGCTATATCATGGCCGACAGCAGAAGAACGGCATC AAGGTGAACCTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGC TCGCCGACCACTACCAGCAGAACCACCCCATCGGCAGCGCCCGTGTCT GCTGCCCGACAACCACTACTGAGCACCAGTCCGCCCTGAGCAAGAC CCCAACGAGAAGCGGATCACATGGTCCCTGGAGTTCGTGACCGCCG CCGGATCACTCTCGGCATGGACGAGCTGTACAAGTAATAAGCTTGGAT CCAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCT TAATATGTTGCTCCTTTTACGCTATGTGGATACGTGCTTAAATGCCT TTGATATGCTATTGCTTCCCGTATGGTTTCATTTTCTCCTCCTTGT ATAAATCCTGGTTGCTGTCTCTTATGAGGAGTTGTGGCCCGTGTGAG GCACGTTGGCGTGGTGTGCACTGTGTTGCTGACGCAACCCCACTGGT TGGGGCATGGCCACCCTGTGAGTCTTTCGGGACTTTCGCTTTCC CCCTCCCTATTGCCACGGCGGAACATCGCCGCTGCCTTGCCTGCTG CTGGACAGGGGCTCGGCTGTGGGCACTGACAAATCCGTGGTGTGTCG GGAAATCATCGTCCCTTCTGGCTGCTCGCCTGTGTGCCACCTGGA TTCTGCGCGGGACGCTCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGC GGACCTTCCCTCCCGCGGCTGCTGCGGCTCTGCGGCTCTTCCGCT CTTCGAGATCTGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGT TTGCCCTCCCCCGTGCCTTCTTGACCTGGAAAGTGCCACTCCCACT GCTCTTTCATAAAATGAGGAAATTCATCGCATTGTCTGAGTAGGT GTCAATCTAATCTGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGA TTGGGAAGACAATAGCAGGCATGCTGGGACTCGACCGCCCTGGTGG TCGAGTTAAGGGCGAATTCGGGATAGGATCTTCTTAGAGCATGGCTAC TAGATAAGTAGCATGGCGGTTAATCATTAACTACAAGGAACCCCTAG TGATGGAGTTGGCCACTCCCTCTCTGCGGCTCGCTCGCTCACTGAGGC CGGGCAGCAAAAGGTGCCCCGACGCCCCGGCTTTGCCCGGGCGGCTCA GTGAGCGAGCGAGCGCGAG</p>
37	<p>rAAV with human CTCF convergent sequence knocked in on positive-sense strand CTGCGCGCTCGCTCGCTCACTGAGGCCGCGCCGGGCAAGCCCGGGCGTC GGGCGACCTTTGGTGCGCCCGCCTCAGTGAAGCGAGCGCGCAGAGA GGGAGTGGCCAACTCCATCACTAGGGGTTCCCTGTAGTTAATGATTAAC CCGCCATGCTACTTATCTACGTAGCCATGCTCTAGGAAGATCGGAATTC GCCCTTAAGCTAGCCACAAGGTGGCGCTAGCTAGTTATTAATAGTAAT CAATTACGGGGTCAATAGTTTATAGCCATATATGGAGTTCCGCGTTAC ATAACTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCGC CCATTGACGCTAATAATGACGTATGTTCCATAGTAACGCCAATAGGGA TTTTCCATTGACGCTAATGGGTGGAGTATTACGGTAAACTGCCCACTT GGAGTACATCAAGTGTATCATATGCCAAGTACGCCCTTATGACGCTC AATGACGGTAAATGGCCCGCCTGGCATTATGCCAGTACATGACCTTAT GGGACTTCTACTTGGCAGTACATCTACGTATTAGTCACTCGCTATTAC CATGGGTATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTT GACTCACGGGGATTTCCAAGTCTCACCCCATTTGACGCTCAATGGGAGTT TGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACCTC CGCCCCATTGACGCAAAATGGGCGGTAGGCGTGTACGGTGGGAGGCTAT ATAAGCAGAGCTGGTTTGTGTAACCGTCAAGTCTTGCAGAAAGTGGTGG TGAGGCACTGGGCGGTAAGTATCAAGGTTACAAGACAGGTTAAGGAG ACCAATAGAAACTGGGCTTGTGAGACAGAGAAGACTCTTGCCTTCTG ATAGGCACCTATGGTCTTACTGACATCCACTTTGCCCTTCTCTCCACA GGTGTCCAGGCGGCCCATGGTGGAGCAAGGGCGAGGAGCTGTACCCG GGTGGTGGCCCATCTGGTTCGAGCTGGACGGCGAGTAAACGGCCACAA GTTACGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG ACCTGAAGTTCATCTGACACCACGGCAAGCTGCCCGTGCCTGGCCCA CCCTCGTGACCACCTGACCTACGGCGTGCAGTGCCTCAGCCGCTACCC CGACCACATGAAGCAGCAGCACTTCTCAAGTCCCGCATGCCCGAAGGC TACGTCACGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGA CCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGA CGTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCTGGGGCAACG CTGGAGTACAACATACAACAGCCACACGCTATATCATGGCCGACAAGC AGAAGAACGGCATCAAGGTGAACCTCAAGATCCGCCACAACATCGAGGA CGGCAGCGTGCAGCTCGCCGACCACTACAGCAGAACACCCCATCGGC GAGGCCCGCTGTGCTGCCGACAACTACCTGAGCACCCAGTCCG CCCTGAGCAAGACCCCAACGAGAAGCGGATCACATGGTCTGCTGGA GTTCTGTGACCGCCCGGGATCACTCTCGCATGGACGAGCTGTACAAG TAATAAGCTTGGATCCAATCAACCTCTGGATTACAAAATTTGTGAAAGA</p>

TABLE 4-continued

rAAV sequences. The inserted CTCF binding sites are underlined.

SEQ ID NO:	Description and Sequence
	<p>TTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACG CTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCAT TTTCTCCTCCTGTATAAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTG TGGCCCGTTGTGAGCAACGTGGCGTGGTGTGCACTGTGTTTGTCTGACG CAACCCCACTGGTTGGGCATTGCCACCACCTGTCAGCTCCTTTCCGG GACTTTCGCTTTCCCTCCTTATGCCACGGCGGAACATCGCCCGCC TGCTTGGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATT CCGTGGTGTGTCGGGAAATCATCGTCCTTTCCTTGGCTGCTCGCCTG TGTGCCACCTGGATTCTGCGCGGGACGCTTCTGCTACGTCCTTTCG GCCTCAATCCAGCGGACCTTCTTCCCGCGCCTGCTGCCGGCTTGC GGCCTTTCGCGCTTTCGAGATCTGCCCTGACTGTGCTTCTAGTTGC CAGCCATCTGTTGTTTGGCCCTCCCGCTGCTTCTTACCTTGAACCTGGAAG GTGCCACTCCCCTGTCCTTCTAATAAAAATGAGGAAATGCATCGCA TTGTCTGAGTAGGTGTCATTCTATTCTGGGGGTGGGTGGGCGAGGAC AGCAAGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGACATCGA CCGCCCTTGGTGGTTCGAGTTAAGGGCGAATCCCGATAAGGATCTTCC TAGAGTAGGCTACGTAGATAAGTAGCATGGCGGGTTAATCATTAACTA CAAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTGCGCGCTCGC TCGCTCACTGAGGCGGGCGGCAACAAGGTCGCCCGACCCCGGGCTTTG CCGGCGGCTCAGTGAGCGAGCGAGCGCGCAG</p>
38	<p>rAAV with human CTCF divergent sequence knocked in on positive sense strand CTGCGCGCTCGCTCACTGAGGCGCCCGGGCAAAGCCCGGGCGTC GGCGACCTTTGGTCCCGGCTCAGTGAGCGAGCGAGCGCAGAGA GGGATGGCCAACTCCATCACTAGGGGTTCTTGTAGTTAATGATTAAC CCGCCATGCTACTTATCTACGTAGCCATGCTTAGGAAGATCGGAATT GCCCTTAAGCTAGCCACAAGTGGCGCTAGCTAGTTATTAATAGTAAT CAATTACGGGTCATTAGTTTACAGCCATATATGGAGTTCCCGGTTAC ATAACTTACGGTAAATGGCCCGCTGGCTGACCGCCCAACGACCCCGC CCATTGACGTCATAATGACGTATGTTCCCATAGTAACGCCAATAGGGA CTTTCATTGACGTCATGGGTGGAGTATTTACGGTAAACTGCCCACT GGCAGTACATCAAGTGTATCATATGCAAGTACGCCCTTATTGACGTC AATGACGTTAAATGGCCCGCTGGCATTATGCCAGTACATGACCTTAT GGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTAC CATGGTGTATGCGGTTTTGGCAGTACATCAATGGCGTGGATAGCGGTTT GACTCAGCGGGATTTCCAAGTCTCCACCCCATGACGTCATGGGAGTT TGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCTGTAACAATC CGCCCATTGACGCAAAATGGCGGTAGGCGTGTACGGTGGGAGGCTAT ATAAGCAGAGCTGGTTAGTGAACCGTCAGATCCTGCAGAAGTTGGTCG TGAGGCCTGGGCAGGTAAGTATCAAGGTTACAAGACAGGTTAAGGAG ACCAATAGAAAATGGGCTTGTGAGACAGAGAAGACTTTGCGTTTCTG ATAGGCACCTATTGGTCTTACTGACATCCACTTTGCTTCTCTCCACA GGTGTCCAGGCGCCGCCATGTTGAGCAAGGGCGAGGAGCTGTTCCACC GGTGGTGGCCATCCTGGTTCGAGCTGGACGGCGACGTAACCGCCACAA GTTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACTACGGCAAGCTG ACCTGAAGTTCATCTGCAACCAGGCAAGCTGCGCTGCTGCGCCCA CCCTCGTGACCACCTGACCTACGGCGTGCAGTGTTCAGCCGCTACCC CGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGC TAGCTCCAGGAGCGCACCATCTTCTTCAAGGACGACGCAACTACAAGA CCCGCGCCGAGGTGAAGTTGAGGGCGACACCTGGTGAACCGCATCGA GCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCTGGGGCAAG CTGGAGTACAACCTACAACAGCCACACGCTCTATATCATGGCCGACAAGC AGAAGAACGGCATCAAGGTGAACCTCAAGATCCGCCACACATCGAGGA CGGCAGCGTGCAGCTCGCCGACCACTACAGCAGAACCCCCATCGGC GACGGCCCCGTGCTGCTGCCGACACCACTACCTGAGCACCCAGTCCG CCCTGAGCAAGACCCCAACGAGAAGCGGATCAGATGGTCTGCTGGA GTTCTGTGACCGCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAG TAATAAGCTTGGATCCAATCAACCTCTGGATTACAAAATTTGTGAAA TTGACTGGTATTCTTAACTATGTTGCTCCTTTACGCTATGTGGATACG CTGCTTTAATGCCTTTGTATCATGCTATGCTTCCCGTATGGCTTTCAT TTTCTCCTCCTGTATAAAATCCTGGTTGCTGTCTTTTATGAGGAGTTG TGGCCCGTTGTGAGCAACGTGGCGTGGTGTGCACTGTGTTTGTCTGACG CAACCCCACTGGTTGGGCATTGCCACCACCTGTCAGCTCCTTTCCGG GACTTTCGCTTTCCCTCCTTATGCCCACGGCGAACTCATCGCCGCT TGCTTGGCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATT CCGTGGTGTGTCGGGAAATCATCGTCCTTTCCTTGGCTGCTCGCCTG TGTGCCACCTGGATTCTGCGCGGGACGCTTCTGCTACGTCCTTTCG GCCTCAATCCAGCGGACCTTCTTCCCGCGCCTGCTGCCGGCTCTGC GGCTCTTCCGCTTTCGAGATCTGCTCGACTGTGCTTCTAGTTGC CAGCCATCTGTTGTTTGGCCCTCCCGCTGCTTCTTACCTGGAAG GTGCCACTCCCCTGTCCTTCTAATAAAAATGAGGAAATGCATCGCA</p>

TABLE 4-continued

rAAV sequences. The inserted CTCF binding sites are underlined.	
SEQ ID NO:	Description and Sequence
	<p>TTGCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGAC AGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGACTCGA GGTGGTCCCCCGCCTCGAGTTAAGGGCGAATCCCATAAGGATCTTCC TAGAGCATGGCTACGTAGATAAGTAGCATGGCGGGTAAATCATTAACTA CAAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCG TCGCTCACTGAGGCCGGGCGACCAAGGTGCGCCGACGCCGGGCTTTG CCCGGGCGCCTCAGTAGCGAGCGAGCGCGCAG</p>
39	<p>rAAV with human CTCF multiple convergent sequence knocked in on positive-sense strand CTGCGCGCTCGCTCGCTCACTGAGGCCGCCGGGCAAGCCCGGGCGTC GGCGACCTTTGGTCGCCCCGCCCTCAGTGAGCGAGCGAGCCGCGCAGAGA GGGAGTGGCCAACCTCCATCACTAGGGGTTCCCTGTAGTTAATGATTAAC CCGCCATGCTACTTATCTACGTAGCCATGCTCTAGGAAGATCGGAATTC GCCTTAAGCTAGCCACAAGGTGGCGCCACAAGGTGGCGCCACAAGG TGGCGCCCAAGGTGGCGCCACAAGGTGGCGCCTAGCTAGTTATTA TAGTAATCAATTACGGGGTCATTAGTTCATAGCCATATATGGAGTTC CGGTTACATAACTTACGGTAAATGGCCCGCTGGCTGACCGCCCAACGA CCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCA ATAGGACTTTCCATTGACGTCAATGGGTGGAGTATTACGGTAAACTG CCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTAT TGACGTCAATGACGGTAAATGGCCCGCTGGCATTATGCCAGTACATG ACCTTATGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCG CTATTACCATGGTGATGCGGTTTGGCAGTACATCAATGGCGTGGATA CGGTTTGTACTCACGGGATTTCCAAGTCTCCACCCCATTTGACGTCAAT GGGAGTTGTTTGGCACCAAAATCAACGGGACTTTCAAAATGTCGTA ACAACCTCCGCCCATTTGACGCAAAATGGCGGTAGGCGTGTACGGTGGGA GGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCTGCAGAAG TTGGTCGTGAGGCCTGGCGAGTAAATCAAGTTACAAGACAGGTT TAAGGAGACCAATAGAAACTGGGCTTGTGAGACAGAGAAGACTCTTGC GTTTCTGATAGGCACCTATTGGTCTTACTGACATCCACTTTGCCCTTCT CTCCACAGGTGTCAGGGCGGCCCATGTTGAGCAAGGGCGAGGAGCTG TTCACCGGGGTGGTGCCATCTGGTTCGAGCTGGACGGCGACGTAAACG GCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGG CAAGCTGACCCGAAAGTTCATCTGCACCACCGGCAAGTGCCTGGTGC TGGCCACCCCTCGTGACCACCCCTGACCTACGGCGTGCAGTGTCTGAC GCTACCCCGACCAATGAAGCAGCAGACTTCTTCAAGTCCGCCATGCC CGAAGGCTACGTCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAAC TACAAGACCCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCCTGGTGAAC GCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCTGGG GCACAAGCTGGAGTACAACACAACAGCCACAACGCTTATATCATGGCC GACAAGCAGAAGAACGGCATCAAGGTGAACCTCAAGATCCGCCACAACA TCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACAGCAGAACACCC CATCGGCGACGGCCCGTGTCTGCTGCCGACAAACACTACCTGAGCACC CAGTCCGCCCTGAGCAAGACCCCAACGAGAAGCGGATCAATGGTCC TGCTGGAGTTCGTGACCCCGCCGGGATCACTCTCGGCATGGACGAGCT GTACAAGTAATAAGCTTGGATCCAATCAACCTTGGATTACAAAATTTG TGAAAGATTGACTGGTATTCTTAATATGTTGCTCCTTTTACGCTATGT GGATACGCTGCTTAAATGCCTTTGTATCATGCTATGCTTCCCGTATGG CTTTCATTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCTTTATGA GGAGTTGTGGCCCGTGTGTCAGGCAACGTGGCGTGGTGTGACTGTGTTT GCTGACGCAACCCCACTGGTTGGGGCATGGCCACCACTGTGAGCTCC TTTCCGGGACTTTCGCTTTCCCTTCCCTATTGACCAGGCGGAACATCAT CGCCGCTGCTTGCCTGCTGGACAGGGGCTCGGCTGTGGGCACT GACAATTCCTGGTGTGTGCGGGAATCATGCTCCTTCTTGGCTGC TCGCCTGTGTGCCACCTGGATTCTGCGGGGACGTCCTTCTGCTACGT CCCTTCGGCCCTCAATCCAGCGGACCTTCTTCCCGGGCTGCTGCGG GCTCTGCGCCCTTCCCGCTCTCGAGATCTGCCCTGACTGTGCCTTC TAGTTGCCAGCCATCTGTTGTGTCCTTCCCGTGCCTTCTTGGACC CTGGAAGGTGCACTCCCACTGTCCTTCTAATAAAATGAGGAATTTG CATCGCATTTGCTGAGTAGGTGTCATTCTATTCTGGGGGTGGGGGGG CAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGG ACTCGACCGCCCTGGTGGCGCCCTGGTGGCGCCCTGGTGGC CGCCCTGGTGGCGCCCTGGTGGTGGTGGTAAAGGGCGAATTTCCCG ATAAGGATCTTCTAGAGCATGGCTACGTAGATAAGTAGCATGGCGGGT TAATCATTAACACAAGGAACCCCTAGTGATGGAGTTGGCCACTCCCT CTGCGCGCTCGCTCGCTCACTGAGGCCGCCGGGCAAGCCCGGGCGTC CGCCCGGCTTTGCCCGGGCGCTCAGTAGCGAGCGAGCGCGCAG</p>
40	<p>rAAV with human CTCF forward on positive-sense and reverse on negative-sense strand, convergent CTGCGCGCTCGCTCGCTCACTGAGGCCGCCGGGCAAGCCCGGGCGTC</p>

TABLE 4-continued

rAAV sequences. The inserted CTCF binding sites are underlined.

SEQ ID NO: Description and Sequence

```

GGGCGACCTTTGGTCGCCCGGCCTCAGTGAGCGAGCGCGCAGAGA
GGGAGTGGCCCACTCCATCACTAGGGGTTCCTTGAGTTAATGATTAAC
CCGCCATGCTACTTATCTACGTAGCCATGCTCTAGGAAGATCGGAATTC
CCCTTAAGCTAGCCACAAGGTGGCGCTAGCTAGTTATTAATAGTAAT
CAATTACGGGGTCATTAGTTCATAGCCATATATGGAGTTCGCGTTAC
ATAACTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCGGC
CCATTGACGTCAATAATGACGTATGTTCCTCATAGTAACGCCAATAGGGA
CTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTT
GGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCTTATTGACGTC
AATGACGGTAAATGGCCCGCCTGGCATTATGCCAGTACATGACCTTAT
GGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTAC
CATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTT
GACTCACGGGGATTTCCAAGTCTCCACCCCATGACGTCAATGGGAGTT
TGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACTC
CGCCCATGACGCAAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTAT
ATAAGCAGAGCTGGTTTGTAGTGAACCGTCAGATCCTGCAGAAGTGGTCG
TGAGGCACTGGGCAGGTAAGTATCAAGGTTACAAGACAGGTTAAGGAG
ACCAATAGAAACTGGGCTTGTGAGACAGAGAAGACTCTTGCGTTTCTG
ATAGGCACCTATTGGTCTTACTGACATCCACTTTGCCTTCTCTCCACA
GGTGCCAGGCGGCCCATGGTGAGCAAGGGCGAGGAGCTGTTACCG
GGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAACGGCCACAA
GTTACGCGTGTCCGGCGAGGGCGAGGGCGATGCCCTACGGCAAGCTG
ACCTGAAGTTCATCTGCACCCCGCAAGCTGCCCGTGCCTGGCCCA
CCCTCGTGACCACCTGACCTACGGCGTGCAGTGCCTCAGCCGCTACCC
CGACCACATGAAGCAGCAGCACTTCTCAAGTCCGCCATGCCCGAAGGC
TAGCTCCAGGAGCGCACCATCTTCTCAAGGACGACGGCACTACAAGA
CCCGCCCGAGGTGAAGTTCGAGGGCGACACCTGGTGAACCGCATCGA
GCTGAAGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACCTACAACAGCCACACCTCTATATCATGGCCGACAAGC
AGAAGAACCGCATCAAGGTGAAGTTCAGATCCGCCACAACATCGAGGA
CGGCAGCGTGCAGCTCGCCGACCCTACCAGCAGAACCCCCATCGGC
GAGCGCCCGTGTGCTGCCGACAACTACCTGAGCACCCAGTCCG
CCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGGA
GTTGCTGACCGCCCGGGATCACTCTCGGCATGGACGAGCTGTACAAG
TAATAAGCTTGGATCCAATCAACCTCTGGATTACAAAATTTGTGAAAAG
TTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACG
CTGCTTAAATGCCCTTGTATCATGCTATTGCTTCCCGTATGGCTTTCAT
TTTCTCCTCCTGTATAAACTCTGGTTGCTGTCTCTTTATGAGGAGTTG
TGCCCGTGTGTCAGGCAACGTGGCGTGGTGTGCATGTTGTTGCTGACG
CAACCCCACTGGTTGGGCAATGGCCACCCTGTGAGTCTCTTCCGG
GACTTTCGCTTTCCTCCCTTATTGCCACGGCGGAACCTCATCGCCGCT
TGCTTTCCTGCTGCTGGACAGGGCTCGGCTGTGGGCACTGACAAT
CCGTGGTGTGTGTCGGGAAATCATGCTCCTTCTTGGCTGCTCGCCTG
TGTTGCCACCTGGATCTGCGCGGGACGTCCTTCTGCTACGTCCTTCCG
GCCTCAATCCAGCGGACCTTCTTCCCGCGCCTGCTGCCGCTCTGC
GGCTCTTCCGCTCTTCGAGATCTGCCCTGACTGTGCTTCTAGTTGC
CAGCCATCTGTTGTTTCCCTCCCTCCCTTCTTCCCTTCCCTTCCCTTCCCT
GTGCCACTCCACTGTCTTCTTAAATAAAATGAGGAAATTCATCGCA
TTGCTGAGTAGGTGTCATCTATTCTGGGGGTGGGGTGGGGCAGGAC
AGCAAGGGGGAGGATGGGAAGACAATAGCAGGCATGCTGGGGACTCGA
GGCGGGGGACCCTCGAGTTAAGGGCAATCCCGATAAGGATCTTC
CTAGAGCATGGCTACGTAGATAAGTAGCATGGCGGTTAATCATTAACT
ACAAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCG
CTCGCTCACTGAGGCCGGGCGACCAAGGTCGCCCAGCCCGGGCTTT
GCCCGGGCGCCTCAGTGAGCGAGCGAGCGCGCAG
    
```

REFERENCES

[0083] 1. Wang D, Tai PWL, Gao G. Adeno-associated virus vector as a platform for gene therapy delivery. *Nat Rev Drug Discov.* 2019;18(5):358-78. doi: 10.1038/s41573-019-0012-9. PubMed PMID: 30710128; PMCID: PMC6927556.

[0084] 2. Colella P, Ronzitti G, Mingozzi F. Emerging Issues in AAV-Mediated. *Mol Ther Methods Clin Dev.* 2018;8:87-104. Epub 20171201. doi: 10.1016/j.omtm.2017.11.007. PubMed PMID: 29326962; PMCID: PMC5758940.

[0085] 3. Rao S S, Huntley M H, Durand N C, Stamenova E K, Bochkov I D, Robinson J T, Sanborn A L, Machol I, Omer A D, Lander E S, Aiden E L. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell.* 2014;159(7):1665-80. Epub 2014/12/11. doi: 10.1016/j.cell.2014.11.021. PubMed PMID: 25497547; PMCID: PMC5635824.

[0086] 4. Phillips-Cremins J E, Sarnia M E, Sanyal A, Gerasimova T I, Lajoie B R, Bell J S, Ong C T, Hookway T A, Guo C, Sun Y, Bland M J, Wagstaff W, Dalton S, McDevitt T C, Sen R, Dekker J, Taylor J,

Corces V G. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell*. 2013;153(6):1281-95. Epub 2013/05/23. doi: 10.1016/j.cell.2013.04.053. PubMed PMID: 23706625; PMCID: PMC3712340.

Mathelier A. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res*. 2022;50(D1):D165-D73. doi: 10.1093/nar/gkab1113. PubMed PMID: 34850907; PMCID: PMC8728201.

[0087] 5. Castro-Mondragon J A, Riudavets-Puig R, Rauluseviciute I, Lemma R B, Turchi L, Blanc-Mathieu R, Lucas J, Boddie P, Khan A, Manosalva Pérez N, Fomes O, Leung T Y, Aguirre A, Hammal F, Schmelter D, Baranasic D, Ballester B, Sandelin A, Lenhard B, Vandepoele K, Wasserman W W, Parcy F,

[0088] 6. Boftsi M, Whittle F B, Wang J, Shepherd P, Burger L R, Kaifer K A, Lorson C L, Joshi T, Pintel D J, Majumder K. The adeno-associated virus 2 (AAV2) genome and rep 68/78 proteins interact with cellular sites of DNA damage. *Hum Mol Genet*. 2021. Epub 2021/10/16. doi: 10.1093/hmg/ddab300. PubMed PMID: 34652429.

SEQUENCE LISTING

```

Sequence total quantity: 64
SEQ ID NO: 1          moltype = DNA length = 14
FEATURE              Location/Qualifiers
source                1..14
                     mol_type = genomic DNA
                     organism = Homo sapiens

SEQUENCE: 1
ccacaagggtg gcgc                                         14

SEQ ID NO: 2          moltype = DNA length = 14
FEATURE              Location/Qualifiers
source                1..14
                     mol_type = genomic DNA
                     organism = Homo sapiens

SEQUENCE: 2
ccaccagggg gcgg                                         14

SEQ ID NO: 3          moltype = DNA length = 19
FEATURE              Location/Qualifiers
source                1..19
                     mol_type = genomic DNA
                     organism = Adeno-associated virus

SEQUENCE: 3
ttgcgacacc atgtggtca                                     19

SEQ ID NO: 4          moltype = DNA length = 20
FEATURE              Location/Qualifiers
source                1..20
                     mol_type = genomic DNA
                     organism = Rodent protoparvovirus 1

SEQUENCE: 4
ctcgcggtag atcactcgtt                                    20

SEQ ID NO: 5          moltype = DNA length = 21
FEATURE              Location/Qualifiers
source                1..21
                     mol_type = genomic DNA
                     organism = Rodent protoparvovirus 1

SEQUENCE: 5
ccaccactaa atggcattct t                                  21

SEQ ID NO: 6          moltype = DNA length = 20
FEATURE              Location/Qualifiers
source                1..20
                     mol_type = genomic DNA
                     organism = Rodent protoparvovirus 1

SEQUENCE: 6
ctcgcggtag atcactcgtt                                    20

SEQ ID NO: 7          moltype = DNA length = 19
FEATURE              Location/Qualifiers
source                1..19
                     mol_type = genomic DNA
                     organism = H1 parvovirus

SEQUENCE: 7
tgctcactag atggcgctc                                     19

SEQ ID NO: 8          moltype = DNA length = 19
FEATURE              Location/Qualifiers

```

-continued

source	1..19 mol_type = genomic DNA organism = H1 parvovirus	
SEQUENCE: 8		
tgcacagcag aggactctg		19
SEQ ID NO: 9	moltype = DNA length = 19 Location/Qualifiers	
FEATURE		
source	1..19 mol_type = genomic DNA organism = H1 parvovirus	
SEQUENCE: 9		
agtcaccaa gggacggag		19
SEQ ID NO: 10	moltype = DNA length = 19 Location/Qualifiers	
FEATURE		
source	1..19 mol_type = genomic DNA organism = Mouse papillomavirus	
SEQUENCE: 10		
tgaacagtag gagtcagtt		19
SEQ ID NO: 11	moltype = DNA length = 19 Location/Qualifiers	
FEATURE		
source	1..19 mol_type = genomic DNA organism = Primate erythroparvovirus 1	
SEQUENCE: 11		
ccaccaatag ggggtcata		19
SEQ ID NO: 12	moltype = DNA length = 19 Location/Qualifiers	
FEATURE		
source	1..19 mol_type = genomic DNA organism = Carnivore protoparvovirus 1	
SEQUENCE: 12		
caaccaggag gtgaaaatc		19
SEQ ID NO: 13	moltype = DNA length = 19 Location/Qualifiers	
FEATURE		
source	1..19 mol_type = genomic DNA organism = Human beta herpesvirus 5	
SEQUENCE: 13		
cgaccgcgac atggcgctg		19
SEQ ID NO: 14	moltype = DNA length = 19 Location/Qualifiers	
FEATURE		
source	1..19 mol_type = genomic DNA organism = Human alpha herpesvirus 1	
SEQUENCE: 14		
agaccaccag gtggcgcac		19
SEQ ID NO: 15	moltype = DNA length = 19 Location/Qualifiers	
FEATURE		
source	1..19 mol_type = genomic DNA organism = Human alpha herpesvirus 1	
SEQUENCE: 15		
ccaccggcgg gggcggcg		19
SEQ ID NO: 16	moltype = DNA length = 19 Location/Qualifiers	
FEATURE		
source	1..19 mol_type = genomic DNA organism = Human alpha herpesvirus 1	
SEQUENCE: 16		
cggcggcgcg gggcggcg		19
SEQ ID NO: 17	moltype = DNA length = 19 Location/Qualifiers	
FEATURE		
source	1..19 mol_type = genomic DNA organism = Human gamma herpesvirus 4	
SEQUENCE: 17		

-continued

tggccaaaag acggcggtt		19
SEQ ID NO: 18	moltype = DNA length = 19	
FEATURE	Location/Qualifiers	
source	1..19	
	mol_type = genomic DNA	
	organism = Human gammaherpesvirus 4	
SEQUENCE: 18		
caccaaacag gtggtgaaa		19
SEQ ID NO: 19	moltype = DNA length = 19	
FEATURE	Location/Qualifiers	
source	1..19	
	mol_type = genomic DNA	
	organism = Human gammaherpesvirus 4	
SEQUENCE: 19		
ttggcaccat gtggtgcca		19
SEQ ID NO: 20	moltype = DNA length = 19	
FEATURE	Location/Qualifiers	
source	1..19	
	mol_type = genomic DNA	
	organism = Alphapapillomavirus 9	
SEQUENCE: 20		
taaccaccag gtggtgcca		19
SEQ ID NO: 21	moltype = DNA length = 19	
FEATURE	Location/Qualifiers	
source	1..19	
	mol_type = genomic DNA	
	organism = Hepatitis B virus	
SEQUENCE: 21		
cgaccaccag ttgatcca		19
SEQ ID NO: 22	moltype = DNA length = 19	
FEATURE	Location/Qualifiers	
source	1..19	
	mol_type = genomic DNA	
	organism = Hepatitis B virus	
SEQUENCE: 22		
tctacagcat gggcgagaa		19
SEQ ID NO: 23	moltype = length =	
SEQUENCE: 23		
000		
SEQ ID NO: 24	moltype = DNA length = 14	
FEATURE	Location/Qualifiers	
source	1..14	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 24		
ccacnagggtg gcag		14
SEQ ID NO: 25	moltype = DNA length = 14	
FEATURE	Location/Qualifiers	
source	1..14	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 25		
ctgccacctn gtgg		14
SEQ ID NO: 26	moltype = DNA length = 14	
FEATURE	Location/Qualifiers	
source	1..14	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 26		
ggcgggggac cacc		14
SEQ ID NO: 27	moltype = DNA length = 19	
FEATURE	Location/Qualifiers	
source	1..19	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 27		

-continued

```

aacgctgtgg tacaccagt 19

SEQ ID NO: 28          moltype = DNA length = 19
FEATURE              Location/Qualifiers
source               1..19
                    mol_type = other DNA
                    organism = synthetic construct

SEQUENCE: 28
tgaccacatg gtgtcgcaa 19

SEQ ID NO: 29          moltype = DNA length = 2716
FEATURE              Location/Qualifiers
source               1..2716
                    mol_type = other DNA
                    organism = synthetic construct

misc_binding         206..224
                    bound_moiety = CTCF binding sites

SEQUENCE: 29
ccgctcgctc gctcactgag gccgcccggg caaagcccgg gcgtcggggc acctttggtc 60
gcccggcctc agtgagcgag cgagcgcgca gagagggagt ggccaactcc atcactaggg 120
gttccttgta gttaatgatt aaccgcccat gctacttacc tacgtagcca tgctctagga 180
agatcggaat tcgcccctaa gctagttgcy acaccatgtg gtcactagct agttattaat 240
agtaatacaat tacggggctca ttagtctata gcccatatat ggagttccgc gttacataac 300
ttacggtaaa tggcccgcct ggtgaccgcg ccaacgaccg ccgcccattg acgtcaataa 360
tgacgtatgt tcccatagta acgccaatag ggactttcca ttgacgtcaa tgggtggagt 420
atctacggta aactgcccac ttggcagtag atcaagtgta tcatatgcca agtacgcccc 480
ctattgacgt caatgacggg aatggcccgc cctggcatta tgcccagtag atgaccttat 540
gggactttcc tacttgccag tacatctacg tattagtcac cgctattacc atgggtgatgc 600
ggttttggca gtacatcaat gggcgtggat agcggtttga ctcacgggga ttccaagtcc 660
tcccaccatc tgacgtcaat gggagtttgt tttggcacca aaatcaacgg gactttccaa 720
aatgtcgtaa caactccgcc ccattgacgc aaatggggcg taggcgtgta cgggtgggag 780
tctatataag cagagctggt ttagtgaacc gtcagatcct gcagaagtgt gtcgtgaggc 840
actgggcagg taagtatcaa ggttacaaga caggtttaag gagaccaata gaaactgggc 900
ttgtcgagac agagaagact cttgcgtttc tgataggcac ctattggtct tactgacatc 960
cactttgect ttctctccac aggtgtccag gggcccgcga ttggtgagcaa gggcgaggag 1020
ctgttcaccg ggggtggtgcc catcctggtc gagctggacg gcgacgtaaa cggccacaag 1080
ttcagcgtgt ccggcagagg cgagggcgat gccacctacg gcaagctgac cctgaagttc 1140
atctgcacca cgggcaagct gcccggtccc tggcccaccg tcgtgaccac cctgacctac 1200
ggcgtgcagt gcttcagccg ctaccccagc cacatgaagc agcacgactt ctccaagtcc 1260
gccatgcccg aaggctacgt ccaggagcgc accatcttct tcaaggacga cggcaactac 1320
aagaccgcgc ccgaggtgaa gttcgagggc gacaccctgg tgaaccgcat cgagctgaag 1380
ggcatcgact tcaaggagga cggcaacatc ctggggcaca agctggagta caactacaac 1440
agccacaacg tctatatcat gggcgacaag cagaagaacg gcatcaaggt gaactcaag 1500
atcccacaca acatcgagga cggcagcgtg cagctcgcgc accactacca gcagaacacc 1560
cccatcggcg acggcccctg gctgctgccc gacaaccact acctgagcac ccagtcggcc 1620
ctgagcaaaag accccaacga gaagcgcgat cacatggctc tgctggagtt cgtgaccgcc 1680
gccgggatac ctctcggcat ggacgagctg tacaagtaat aagcttggat ccaatcaacc 1740
tctggattac aaaatttgtg aaagattgac tggattctct aactatggtg ctctctttac 1800
gctatgtgga tacgctgctt taatgccttt gtatcatgct attgctccc gtatggcttt 1860
cattttctcc tecttgata aatcctgggt gctgtctctt tatgaggagt tgtggcccgt 1920
tgtagggcaa cgtggcgtgg tgtgactgt gtttgctgac gcaaccccca ctggttgggg 1980
cattgccacc acctgtcagc tcttttccgg gactttcgct tccccctcc ctattgccac 2040
ggcggaaact atcgcgccct gccttgcggc ctgctggaca ggggctcggc tgttgggca 2100
tgacaattcc gtgggtgtgt cggggaatc atcgtcctt ccttggctgc tcgctgtgt 2160
tgccacctgg attctgcgcy ggaagctcct ctgctacgtc ccttcggccc tcaatccagc 2220
ggaccttctc tcccggggcc tgetgcccgc tctgcccct ctcccgctc ttcgagatct 2280
gcctcgactg tgccctctag ttgccagcca tctgtgtgt gccctccccc cgtgccctcc 2340
ttgacctggg aaggtgccac tcccactgtc ctttctaat aaaaatgagga aattgcatcg 2400
cattgtctga gtaggtgtca ttctattctg gggggtgggg tggggcagga cagcaagggg 2460
gaggtattgg aagacaatag caggcatgct ggggactcga tcgagttaa ggcgaattcc 2520
cgataaggat cttcctagag atggctacg tagataagta gcatggcggg ttaatcatta 2580
actacaagga accctagctg atggagttgg ccaatccctc tctgcgcgct cgtcgtctca 2640
ctgaggcccg gcgaccaaa gtcgcccgcg gccccgggctt tgcccggggc gcctcagtga 2700
gcgagcgagc gcgacg 2716

SEQ ID NO: 30          moltype = DNA length = 2716
FEATURE              Location/Qualifiers
source               1..2716
                    mol_type = other DNA
                    organism = synthetic construct

misc_binding         2482..2500
                    bound_moiety = CTCF binding site

SEQUENCE: 30
ccgctcgctc gctcactgag gccgcccggg caaagcccgg gcgtcggggc acctttggtc 60
gcccggcctc agtgagcgag cgagcgcgca gagagggagt ggccaactcc atcactaggg 120
gttccttgta gttaatgatt aaccgcccat gctacttacc tacgtagcca tgctctagga 180

```

-continued

```

agatcggaat tgcgctttaa gctagctagc tagttattaa tagtaatcaa ttacggggct 240
attagttcat agcccatata tggagttccg cgttacataa cttacggtaa atggcccggc 300
tggctgaccg cccaacgacc cccgcccatt gacgtcaata atgacgtatg ttcccatagt 360
aacgccaata gggactttcc attgacgtca atgggtggag tatttacggg aaactgocca 420
cttggcagta catcaagtgt atcatatgcc aagtacgccc cctattgacg tcaatgacgg 480
taaattggcc gcctggcatt atgcccagta catgacctta tgggactttc ctacttgga 540
gtacatctac gtatttagtca tgcgtattac catgggtgat cggttttggc agtacaatcaa 600
tggcgctgga tagcgttttg actcacgggg atttccaagt ctcccaccca ttgacgtcaa 660
tgggagtttg ttttggcacc aaaatcaacg ggactttcca aaatgtcgta acaactccgc 720
cccattgacg caaatggggcg gtaggcgtgt acgggtgggag gtctataata gcagagctgg 780
tttagtgaac cgtcagatcc tgcagaagt ggtcgtgagg cactgggcag gtaagtatca 840
aggttacaag acaggtttaa ggagaccaat agaaactggg cttgtcgaga cagagaagac 900
tcttgcgttt ctgataggca cctattggtc ttactgacat ccaacttggc tttctctcca 960
caggtgtcca ggcggcggcc atgggtgagca agggcgaggg gctgttacc ggggtgggtg 1020
ccatcctggt cgagctggac ggcgacgtaa acggccacaa gttcagcgtg tccggcgagg 1080
gcgagggcga gtcacacttc ggaacgtga cctgaagtt catctgacc accggcaagc 1140
tgcgctgccc ctggcccacc ctgctgacca cctgacctca cggcgtgacg tcttccagcc 1200
gtaccccga ccacatgaag cagcacgact tcttcaagtc gcgcatgccc gaaggtctag 1260
tcaggagcg caccatcttc tccaaggacg acggcaacta caagaccgc gccgaggtga 1320
agttcgaggg cgacaccctg gtgaaccgca tgcagctgaa gggcatcgac tccaaggagg 1380
acggcaacat cctggggcac aagctggagt acaactaca cagccacaac gtctatatca 1440
tggcgcgaaa gcagaagaac ggcatcaagg tgaacttcaa gatccgccc aacatcgagg 1500
acggcagcgt gcagctcgcc gaccactacc agcagaacac ccccatcgcc gacggcccgc 1560
tgctgctgcc tcacaaccac tacctgagca cccagtccgc cctgagcaaa gaccccaacg 1620
agaagcgcga tcacatggtc ctgctggagt tctgaccgc cgcggggatc actctggca 1680
tggcagcagc gtacaagtta taagcttggg tccaatcaac ctctggatta caaaatttgt 1740
gaaagattga ctggtattct taactatggt gctccttcta cgtatgtgt atacgctgct 1800
ttaatgcctt tgtatcatgc tattgcttcc cgtatgctt tcattttctc ctcttgtat 1860
aaatcctggt tctgtctctc ttatgaggag ttgtggcccg ttgtcaggca acgtggcgtg 1920
tgtgcaactg tgtttgctga cgcaaccccc actggttggg gcattgcccac cacctgtcag 1980
ctcttctccg ggactttcgc tttcccctc cctattgcca cggcggaact catcgccgc 2040
tgcttgcgcc gctgctggac aggggctcgg ctggtgggca ctgacaattc cgtggtgttg 2100
tcggggaaat catcgtcctt tccttggctg ctgcctctgt ttgccacctg gattctgcgc 2160
gggacgtcct tctgctactg ccttccggcc ctcaatccag cggaccttcc tcccggcgc 2220
ctgctgcccg ctctgcccgc tcttccgcgt cttcgagatc tgctcagact gtgccttcta 2280
gttgccagcc atctgttgtt tgcccctccc cctgctcttc cttgaccttc gaaggtgcca 2340
ctcccactgt cctttcctaa taaaatgagg aaattgcatc gcattgtctg agtaggtgtc 2400
atcttattct ggggggtggg gtggggcagg acagcaaggg ggaggattgg gaagacaata 2460
gcaggcatgc caactggcgt aactgggtga ccacagcgtt tcgagttaag ggcgaatcc 2520
cgataaggat cttcctagag catggctacg tagataagta gcatggcggg ttaatcatta 2580
actacaagga acccctagtg atggagttgg ccaactccctc tctgcccgtc cgctcgtca 2640
ctgagggcgg gcgaccaaaag gtgcgccgac gcccgggctt tgcccggggc gcctcagtga 2700
gcgagcgagc gcgcag 2716
    
```

```

SEQ ID NO: 31 moltype = DNA length = 2735
FEATURE Location/Qualifiers
source 1..2735
mol_type = other DNA
organism = synthetic construct
misc_binding 207..225
bound_moiety = CTCF binding site
    
```

```

SEQUENCE: 31
ccgctcgtc gctcactgag gccgcccggg caaagcccgg cgcgtcggcg acctttggct 60
gcccggcctc agtgagcgag cgagcgcgca gagaggaggt ggccaactcc atcactaggg 120
gttccttgta gttaatgatt aaccccgcct gctacttacc tacgtagcca tgccttagga 180
agatcggaat tgcgcccata gctagttgag acaccatgtg gtcactagct agttattaat 240
agtaatcaat tacggggtca ttagttcata gcccatatag ggagttccgc gttacataac 300
ttacggtaaa tggcccgcct ggtgaccgc ccaacgaccc ccgcccattg acgtcaataa 360
tgactgatgt tcccatagta acgccaatag ggactttcca ttgacgtcaa tgggtggagt 420
atttacggta aactgcccac ttggcagtac atcaagtgta tcatatgcca agtacgcccc 480
ctattgacgt caatgacggt aaatggcccc cctggcatta tgcccagtac atgacctat 540
gggactttcc cacttggcag tacatctacg tattagtcac cgtattacc atgggtgatg 600
ggttttggca gtacatcaat gggcgtggat agcggtttga ctacggggga tttccaagtc 660
tccaccccat tgacgtcaat gggagtttgt tttggcaca aaatcaacgg gactttccaa 720
aatgtcgtaa caactgccc ccattgacgc aaatggggcg taggcgtgta cgggtgggag 780
tctataaag cagagctggt ttagtgaacc gtcagatcct gcagaagttg gtcgtgaggc 840
actgggcagg taagtatcaa ggttacaaga caggtttaag gagaccaata gaaactgggc 900
tgtcagagac agagaagact cttgcccctt tgataggcac ctattggtct tactgacatc 960
cactttgctt ttctctccac aggtgtccag gcggccgcca tgggtgagcaa gggcgaggag 1020
ctgttaccgc ggggtgggtcc catcctggtc gagctggacg gcgacgtaaa cggccacaag 1080
ttcagcgtgt cccgcgaggg cgagggcgat gccacctacg gcaagctgac cctgaagttc 1140
atctgcacca ccggcaagct gcccggtccc tggcccaccc tctgtgaccac cctgacctac 1200
ggcgtgacgt gcttcagccg ctaccccgcac cacatgaagc agcacgactt cttcaagttc 1260
gcaatgcccg aaggctacgt ccaggagcgc accatctct tcaaggacga gcgcaactac 1320
aagaccgcgc ccgaggtgaa gttcagagggc gacaccctgg tgaaccgcat cgagctgaag 1380
ggcatcgact tcaaggagga cggcaacatc ctggggcaca agctggagta caactacaac 1440
    
```

-continued

```

agccacaacg tctatatacat ggccgacaag cagaagaacg gcatcaaggt gaacttcaag 1500
atccgccaca acatcgagga cggcagcgtg cagctcgccg accactacca gcagaacacc 1560
cccatcggcg acgcccccggt gctgctgccc gacaaccact acctgagcac ccagtcggcc 1620
ctgagcaaaag accccaacga gaagcgcgat cacatggtcc tgetggagtt cgtgaccgcc 1680
gccgggataca ctctcggcat ggaacgagctg tacaagtaat aagcttggat ccaatcaacc 1740
ctcggattac aaaatttggg aaagattgac tggatttctt aactatgttg ctccctttac 1800
gctatgtgga tacgctgctt taatgccttt gtatcatgct attgcttccc gtatggcttt 1860
cattttctcc tccttgata aatcctgggt gctgtctctt tatgaggagt tgtggcccgt 1920
tgtcaggcaa cgtggcgtgg tgtgcaactgt gtttgcctgac gcaacccccca ctggttgggg 1980
cattgccacc acctgtcagc tcctttccgg gactttcgct ttccccctcc ctattgccac 2040
ggcggaaactc atcgccgect gccttgcccc ctgctggaca ggggctcggc tgttgggca 2100
tgacaattcc gtggtgttgt cggggaatc atcgtccttt ccttggctgc tcgctgtgt 2160
tgccacctgg attctgcgct ggaacgtcctt ctgctacgtc ccttcggccc tcaatccagc 2220
ggaccttctc tcccgcggcc tgetgcccgc tctgcggcct cttcccgctc ttcgagatct 2280
gcctcgactg tgccttctag ttgccagcca tctgttgttt gcccctcccc cgtgcttcc 2340
ttgacctggc aaggtgccac tcccactgtc ctttctaat aaaatgagga aattgcatcg 2400
cattgtctga gtagggtgca ttctattctg gggggtgggg tggggcagga cagcaagggg 2460
gaggattggg aagacaatag caggcatgct ggggactcga actgggtgac cacagcgttt 2520
cgagtttaagg gcgaattccc gataaggatc ttcctagagc atggctacgt agataagtag 2580
catggcgggt taatcattaa ctacaaggaa cccctagtga tggagttggc cactccctct 2640
ctgcccgcctc gctcgtcac tgaggccggg cgaccaaagg tcgcccagac cccgggcttt 2700
gccggggcgg cctcagtgag cgagcgagcg cgcag 2735

```

```

SEQ ID NO: 32 moltype = DNA length = 2735
FEATURE Location/Qualifiers
source 1..2735
mol_type = other DNA
organism = synthetic construct
misc_binding 207..225
bound_moiety = CTCF binding site
misc_binding 2505..2520
bound_moiety = CTCF binding site

```

```

SEQUENCE: 32
ccgctcgctc gctcactgag gccgcccggg caaagcccgg gcgtcggggc acctttggtc 60
gcccccctcc agtgagcgag cgagcgcgca gagagggagt ggccaactcc atcactaggg 120
gttccttgta gttaatgatt aaccgcctat gctacttacc tacgtagcca tgccttagga 180
agatcggaat tgcgccctaa gctagttgag acaccatgtg gtcactagct agttatata 240
agtaatacat tacgggggtca ttagttcata gcccatatat ggagtccgcg gttacataac 300
ttacggtaaa tggcccgcct ggctgaccgc ccaacgaccc ccgcccattg acgtcaataa 360
tgacgtatgt tcccatagta acgccaatag ggaacttcca ttgacgtcaa tgggtggagt 420
atttacggta aactgcccac ttggcagtac atcaagtgtc tcatatgcca agtacgccc 480
ctattgacgt caatgacggg aaatggcccg cctggcatta tgcacagtac atgaccttat 540
gggactttcc tacttggcag tacatctacg tattagtcac cgctattacc atgggtgatgc 600
ggttttggca gtacatcaat gggcgtggat agcggtttga ctacggggga ttccaagtcc 660
tccaccctcc tgacgtcaat gggagtttgt tttggcacca aaatcaacgg gactttccaa 720
aatgtcgtaa caactccgccc ccattgacgc aatggggcgg taggcgtgta cggttggagg 780
tctataaag cagagctggg ttagtgaacc gtcagatcct gcagaagttg gtcgtgagcc 840
actgggcagg taagtataca ggttacaaga caggtttaag gagaccaata gaaactgggc 900
ttgtcgagac agagaagact ctgctgtttc tgataggcac ctattgtgct tactgacatc 960
cactttgcct ttctctccac aggtgtccag gcggccgcca tgggtgagcaa gggcgaggag 1020
ctgttcaccg ggggtgtgccc catcctggtc gagctggacg cgcagctaaa cggccacaag 1080
ttcagcgtgt ccggcgaggg cgagggcgat gccacctacg gcaagctgac cctgaagttc 1140
atctgacca ccggcaagct cggcctgccc tggcccaccc tctgtagacc cctgacctac 1200
ggcgtgacgt gcttcagccg ctaccccagc cacatgaagc agcagcactc cttcaagtcc 1260
gccatgcccg aaggetacgt ccaggagcgc accatcttct tcaaggacga cggcaactac 1320
aagaccgcgc ccgaggtgaa gttcgagggc gacaccctgg tgaaccgatc cgagctgaag 1380
ggcatcgact tcaagggagga cggcaacatc ctggggcaca agctggagta caactacaac 1440
agccacaacg tctatatacat ggccgacaag cagaagaacg gcatcaaggt gaacttcaag 1500
atccgccaca acatcgagga cggcagcgtg cagctcgccg accactacca gcagaacacc 1560
cccatcggcg acgcccccggt gctgctgccc gacaaccact acctgagcac ccagtcggcc 1620
ctgagcaaaag accccaacga gaagcgcgat cacatggtcc tgetggagtt cgtgaccgcc 1680
gccgggataca ctctcggcat ggaacgagctg tacaagtaat aagcttggat ccaatcaacc 1740
ctcggattac aaaatttggg aaagattgac tggatttctt aactatgttg ctccctttac 1800
gctatgtgga tacgctgctt taatgccttt gtatcatgct attgcttccc gtatggcttt 1860
cattttctcc tccttgata aatcctgggt gctgtctctt tatgaggagt tgtggcccgt 1920
tgtcaggcaa cgtggcgtgg tgtgcaactgt gtttgcctgac gcaacccccca ctggttgggg 1980
cattgccacc acctgtcagc tcctttccgg gactttcgct ttccccctcc ctattgccac 2040
ggcggaaactc atcgccgect gccttgcccc ctgctggaca ggggctcggc tgttgggca 2100
tgacaattcc gtggtgttgt cggggaatc atcgtccttt ccttggctgc tcgctgtgt 2160
tgccacctgg attctgcgct ggaacgtcctt ctgctacgtc ccttcggccc tcaatccagc 2220
ggaccttctc tcccgcggcc tgetgcccgc tctgcggcct cttcccgctc ttcgagatct 2280
gcctcgactg tgccttctag ttgccagcca tctgttgttt gcccctcccc cgtgcttcc 2340
ttgacctggc aaggtgccac tcccactgtc ctttctaat aaaatgagga aattgcatcg 2400
cattgtctga gtagggtgca ttctattctg gggggtgggg tggggcagga cagcaagggg 2460
gaggattggg aagacaatag caggcatgct ggggactcga ttgcgacacc atgtggctat 2520
cgagtttaagg gcgaattccc gataaggatc ttcctagagc atggctacgt agataagtag 2580

```

-continued

```

catggcgggt taatcattaa ctacaaggaa ccctagtga tggagttggc cactccctct 2640
ctgcgcgctc gctcgctcac tgaggccggg cgaccaaagg tcgcccgacg cccgggcttt 2700
gcccgggcgg cctcagtgag cgagcgagcg cgacg 2735
    
```

```

SEQ ID NO: 33      moltype = DNA length = 2887
FEATURE           Location/Qualifiers
source            1..2887
                  mol_type = other DNA
                  organism = synthetic construct
misc_binding      207..225
                  bound_moiety = CTCF binding site
misc_binding      283..301
                  bound_moiety = CTCF binding site
misc_binding      2578..2596
                  bound_moiety = CTCF binding site
misc_binding      2616..2634
                  bound_moiety = CTCF binding site
misc_binding      2654..2672
                  bound_moiety = CTCF binding site
misc_binding      245..263
                  bound_moiety = CTCF binding site
    
```

```

SEQUENCE: 33
ccgctcgctc gctcactgag gccgcccggg caaagcccgg gcgtcggggc acctttggtc 60
gcccggcctc agtgagcgag cgagcgcgca gagaggaggt ggccaactcc atcactaggg 120
gttccttgta gttaatgatt aaaccgccat gctacttacc tacgtagcca tgctctagga 180
agatcggaat tcgcccttaa gctagttgag acaccatgtg gtcattgcga caccatgtgg 240
tcattgcgac accatgtggg cattgcgaca ccatgtggtc attgcgacac catgtggtea 300
ctagctagtt attaatagta atcaattacg gggtcattag ttcatagccc atatatggag 360
ttccgcgta cataacttac ggtaaatggc ccgctggct gaccgcccga cgacccccgc 420
ccattgacgt caataatgac gtagttccc atagtaacgc caatagggac ttccattga 480
cgtaaatggg tggagtattt acggtaaac gcccacttgg cagtacatca agtgtatcat 540
atgccaaagta cgcgccctat tgacgtcaat gacggtaaat ggcccgcctg gcattatgcc 600
cagtagcatga ccttatggga ctttcctact tggcagtaga tctacgtatt agtcatcgct 660
attaccatgg tgatcggttt ttggcagtag atcaatgggc gtggatagcg gtttgactca 720
cggggatttc caagtctcca ccccatggac gtcaatggga gtttggtttg gcacaaaaat 780
caacgggact ttccaaaatg tcgtaacaac tccgcccctat tgacgcaaat gggcggtagg 840
cgtgtacggg gggaggtcta tataagcaga gctggtttag tgaaccgtca gatcctgcag 900
aagttggctg tgaggcactg ggcaggttaag tatcaaggtt acaagacagg ttaaggaga 960
ccaatagaaa ctgggcttgt cgagacagag aagactcttg cgtttctgat aggcacctat 1020
tggctctact gacatccact ttgcctttct ctccacaggt gtccaggcgg ccgccatggt 1080
gagcaagggc gaggagctgt tcaccggggg ggtgccctac ctggtcgagc tggacggcga 1140
cgtaaacggc cacaagttca cggtgtccgg cgaggggcag ggcgatgcca cctacggcaa 1200
gctgaccctg aagttcatct gcaccaccgg caagctgccc gtgcccctgg ccaccctcgt 1260
gaccaccctg acctacggcg tgcagtgtt cagccgttac cccgaccaca tgaagcagca 1320
cgactcttc aagtcgcca tgcccgaagg ctacgtccag gagcgacca tcttcttcaa 1380
ggacgacggc actacaaga ccgcgcgcca ggtgaagttc gagggcgaca cctgtgtgaa 1440
ccgcatcgag ctgaagggca tcgaactcaa ggaggacggc aacatcctgg ggcacaagct 1500
ggagtacaac tacaacagcc acaacgtcta tatcatggcc gacaagcaga agaacggcat 1560
caaggtgaac ttcaagatcc gccacaacat cgaggacggc agcgtgcagc tcgcccacca 1620
ctaccagcag aacaccoccca tggcgacggg ccccgctgct ctgcccgaca accactacct 1680
gagcaccagg tccgcccctg gcaagacccc caacgagaag cgcgatcaca tggctctgct 1740
ggagttcgtg accgcccggg ggatcactct cggcatggac gagctgtaca agtaataagc 1800
ttggatccaa tcaacctctg gattacaaaa tttgtgaaag attgactggt attcttaact 1860
atgttgctcc ttttacgcta tgtggatagc ctgctttaat gcctttgcat catgctattg 1920
cttcccgtat ggctttcatt ttctctcctc tgtataaatc ctggttgctg tctctttatg 1980
aggagtgtg gcccgttggc aggcacagtg cgtgggtgtg cactgtgttt gctgacgcaa 2040
ccccactgg ttggggcatt gccaccacct gtcagctcct ttcccggact ttcgctttcc 2100
ccctccctat tgcccaggcg gaactcatcg ccgctgctc tgcccgtgctc tggacagggg 2160
ctcgctgtt gggcactgac aattccgtgg tgttgtcggg gaaatcatcg tctttcctt 2220
ggctgctcgc ctgtgttgcc acctggatc tgccggggac gtcctctgc tacgtccctt 2280
cggccctcaa tccagcggac ctctctccc cggcctgct gcccgtctg cggcctctt 2340
cgcgtcttc agactctgct cgaactgtgc ttctagtgc cagccatctg ttgtttgcc 2400
ctccccctg ccttcttga cctggaagg tgcccactccc actgtccttt cctaataaaa 2460
tgaggaaatt gcctcgatt gctgtagtag gtgtcattct attctggggg gtgggggtgg 2520
gcaggacagc aagggggagg attgggaaga caatagcagg catgctgggg actcgaactg 2580
gtgtaccaca gcgttactgg tgtaccacag cgttactggt gtaccacagc gttactgggt 2640
taccacagcg ttactgtgtg accacagcgt ttcaggttaa gggcgaattc ccgataagga 2700
tcttcctaga gcattggctac gtagataagt agcattggcg gttaatcatt aactacaagg 2760
aaccctagt gatggagtg gccactccct ctctgcgcgc tcgctcgtc actgaggcgg 2820
ggcgacaaa ggtcgcccga cgcgggggct ttgcccgggc ggcctcagtg agcagcgag 2880
cgcgag 2887
    
```

```

SEQ ID NO: 34      moltype = DNA length = 2735
FEATURE           Location/Qualifiers
source            1..2735
                  mol_type = other DNA
    
```

-continued

```

misc_binding      organism = synthetic construct
                  207..225
                  bound_moiety = CTCF binding site
misc_binding      2502..2520
                  bound_moiety = CTCF binding site

SEQUENCE: 34
ccgctcgctc      gctcactgag      gccgcccggg      caaagcccgg      gcgtcggggc      acctttggtc      60
gcccgcctc      agtgagcgag      cgagcgcgca      gagaggaggt      ggccaactcc      atcactaggg      120
gttccttgta      gttaatgatt      aacccgccat      gctacttatac      tacgtagcca      tgctctagga      180
agatcggaat      tcgccccttaa      gctagttgcg      acaccatgtg      gtcactagct      agttattaat      240
agtaatcaat      tacggggctca      ttagttcata      gccatataat      ggagttccgc      gttacataac      300
ttacggtaaa      tggcccgcct      ggtgacccgc      ccaacgaccc      ccgcccattg      acgtcaataa      360
tgacgtatgt      tcccatagta      acgccaatag      ggactttcca      ttgacgtcaa      tgggtggagt      420
atttacggta      aactgcccac      ttggcagtac      atcaagtgtat      tcatatgccaa      agtacgcccc      480
ctattgacgt      caatgacggg      aaatggcccg      cctggcatta      tgcccagtac      atgaccttat      540
gggactttcc      tacttggcag      tacatctaac      tattagtcac      cgctattacc      atgggtatgc      600
ggttttgcca      gtacatcaat      gggcgtggat      agcggtttga      ctcacgggga      ttccaagtcc      660
tccaccccat      tgacgtcaat      gggagtttgt      ttggcaccac      aaatcaacgg      gactttccaa      720
aatgtcgtaa      caactccgcc      ccattgacgc      aaatggggcg      taggcgtgta      cggtggggagg      780
tctatataag      cagagctggg      ttagtgaacc      gtcagatcct      gcagaagtgt      gtcgtgaggc      840
actgggcagg      taagtatcaa      ggttacaaga      caggtttaag      gagaccaata      gaaactgggc      900
ttgtcgagac      agagaagact      cttgcgttcc      tgataggcac      ctattggtct      tactgacatc      960
cactttgcct      ttctctccac      aggtgtccag      cgggcccgca      tgggtgagcaa      gggcggaggag      1020
ctggtcacccg      ggtgtgtgct      catcctggtc      gagctggacg      gcgacgtaaa      cggccacaag      1080
ttcagcgtgt      ccggcgaggg      cgagggcgat      gccacctacg      gcaagctgac      cctgaagttc      1140
atctgcacca      ccggcaagct      gccctgccc      tggcccaccc      tcgtgaccac      cctgacctac      1200
ggcgtgcagt      gcttcagccg      ctaccccagc      cacatgaagc      agcacgactt      cttcaagtcc      1260
gccatgcccc      aaggctacgt      ccaggagcgc      accatcttct      tcaaggacga      cggcaactac      1320
aagaccgcg      ccgaggtgaa      gttcgagggc      gacaccctgg      tgaaccgcat      cgagctgaag      1380
ggcatcgact      tcaagggagga      cggcaacatc      ctggggcaca      agctggagta      caactacaac      1440
agccacaacg      tctatataat      gggcgacaag      cagaagaacg      gcatcaaggt      gaactcaag      1500
atccgccaca      acatcgagga      cggcagcgtg      cagctcgcgc      accactacca      gcagaacacc      1560
cccatcggcg      acgcccctgt      gctgctgccc      gacaaccact      acctgagcac      ccagtcgcgc      1620
ctgagcaaa      accccaacga      gaagcgcgat      cacatggtcc      tgcctggagt      cgtgaccgcc      1680
gccgggata      ctctcggcat      ggacgagctg      tacaagtaat      aagcttggat      ccaatcaacc      1740
tctggattac      aaaatthtgg      aaagattgac      tggattctt      aactatgttg      ctcttttac      1800
gctatgtgga      tacgctgctt      taatgccttt      gtatcatgct      attgctccc      gtatggcttt      1860
cattttctcc      tccttgata      aatcctggtt      gctgtctctt      tatgaggagt      tgtggcccgt      1920
tgtcaggcaa      cgtggcgtgg      tgtgcaactgt      gtttgcctgac      gcaaccccca      ctggttgggg      1980
cattgcccac      acctgtcagc      tcctttccgg      gactttcgct      tccccctcc      ctattgccac      2040
ggcggaaact      atcgccgcct      gccctgccc      ctgctggaca      ggggctcgc      tgttgggca      2100
tgacaattcc      gtgtgtgtgt      cggggaaate      atcgtccttt      ccttggctgc      tcgctctgt      2160
tgcccactgg      attctgcgcg      gacagctcct      ctgctacgtc      ccttcggccc      tcaatccagc      2220
ggacttctct      tcccgcggcc      tgcctgccc      tctgcggcct      cttccgcctc      ttcgagatct      2280
gccctcgactg      tgccctctag      ttgccagcca      tctgttgttt      gccctccc      cgtgctctcc      2340
ttgacctgg      aaggtgccac      tcccactgtc      ctttctcaat      aaaaagagga      aatgtcatcg      2400
cattgtctga      ttaggtgtca      ttctattctg      ggggggtggg      tggggcagga      cagcaagggg      2460
gaggtatggg      aagacaatag      caggcatgct      ggggactcga      tgaccacatg      gtgtcgcaat      2520
cgagttaaag      gcaaatcccc      gataaggatc      ttcctagagc      atggctacgt      agataagtag      2580
catggcgggt      taatcattaa      ctacaaggaa      cccctagtga      tggagttggc      cactccctct      2640
ctgcccgcctc      gctcgtctac      tgaggcccgg      cgaccaaaag      tcgcccagc      cccgggcttt      2700
gcccggcg      cctcagtgag      cgagcgagcg      cgag      2735

```

```

SEQ ID NO: 35      moltype = DNA      length = 2715
FEATURE            Location/Qualifiers
source             1..2715
                  mol_type = other DNA
                  organism = synthetic construct
misc_binding      211..224
                  bound_moiety = CTCFbinding site

SEQUENCE: 35
ctgcccgcctc      gctcgtctac      tgaggcccgc      cgggcaaaag      ccgggcgtgc      ggcgaccttt      60
ggtcgcggcg      cctcagtgag      cgagcagagcg      cgcagagagg      gagtggccaa      ctccatcact      120
aggggttctct      tgtagttaa      gattaaccgc      ccatgctact      tatctacgta      gccatgctct      180
aggaagatcg      gaattcgccc      ttaagctagc      cacaaggtgg      cgcctagcta      gttattaata      240
gtaatcaat      acggggctcat      tagttcatag      cccatataatg      gagttccgcg      ttacataact      300
tacggtaaat      ggcccgcctg      gctgaccgcc      caacgacccc      cgcccattga      cgtcaataat      360
gacgtatggt      cccatagtaa      cgccaatag      gactttccat      tgacgtcaat      ggggtggagta      420
tttacggtaa      actgcccact      tggcagta      tcaagtgtat      catatgccaa      gtacgcccc      480
tattgacgt      aatgacggta      aatggcccgc      ctggcattat      gccacgtaca      tgaccttatg      540
ggacttctct      acttggcagt      acatctact      attagtcatc      gctattacca      tgggtatgct      600
gttttggcag      tacatcaatg      ggcgtggata      gcggtttgac      tcacggggat      ttccaagtct      660
caacccatt      gacgtcaatg      ggagtttgtt      ttggcaccac      aatcaacggg      actttccaaa      720
atgtcgtaac      aactccgcc      cattgacgca      aatggggcgg      aggcgtgtac      ggtgggagggt      780
ctatataagc      agagctggtt      tagtgaaccg      tcagatcctg      cagaagtgtg      tcgtgagga      840
ctgggcaggt      aagtatcaag      gttacaagac      aggtttaagg      agaccaatag      aaactgggct      900

```

-continued

```

tgtcgagaca gagaagactc ttgcgtttct gataggcacc tattggctct actgacatcc 960
actttgcctt tctctccaca ggtgtccagg cggccgccat ggtgagcaag ggcgaggagc 1020
tgttcaccgg ggtggtgccc atcctggtcg agctggacgg cgacgtaaac ggcacacaagt 1080
tcagcgtgtc cggcgagggc gaggcgatg ccacctacgg caagctgacc ctgaagtcca 1140
tctgaccac cggcaagctg cccgtgccct ggcccacct cgtgaccacc ctgacctacg 1200
cgctgcagtg cttcagccgc taccgccacc acatgaagca gcacgacttc ttcaagtccg 1260
ccatgccgga aggtctcgtt caggagcgca ccatcttctt caaggacgac ggcaactaca 1320
agaccccgcg cgaggtgaag ttcaggggcg acaccctggt gaaccgcatc gagctgaagg 1380
gcatcgactt caaggaggac ggcaacatcc tggggcacaa gctggagtac aactacaaca 1440
gccacaacgt ctatatcatg gccgacaagc agaagaacgg catcaaggct aacttcaaga 1500
tccgccacaa catcgaggac ggcagcgtgc agctcgccga ccactaccag cagaacaccc 1560
ccatcggcga cggccccgtg ctgctgcccg acaaccacta cctgagcacc cagtcgcccc 1620
tgagcaaaag ccccacgcgag aagcgcgatc acatggtcct gctggagttc gtgaccgccc 1680
ccgggatcac tctcggcatg gacgagctgt acaagtaata agcttggatc caatcaacct 1740
ctggattaca aaatttctga aagattgact ggtattctta actatgtgc tccttttacg 1800
ctatgtggat acgtgctttt aatgcctttg tatcatgcta ttgcttcccg tatggctttc 1860
atcttctcct ccttgtataa atcctggttg ctgtctcttt atgaggagt gtggcccgtt 1920
gtcaggcaac gtggcgtggt gtgcactgtg tttgctgacg caacccccac tggttggggc 1980
attgccacca cctgtcagct cctttccggg actttcgttt tccccctccc tattgccacg 2040
gcggaactca tcgccgcctg ccttgcccgc tgcctggacag gggctcggct gttgggcact 2100
gacaattccg ttggtgtgtc ggggaaatca tcgtccttc cttggctgct cgcctgtgtt 2160
gccacctgga tctcgcgctg gacgtccttc tgcactgctc cttcggcctc caatccagcg 2220
gaccttcctt ccccgccgct gctgccggct ctgcccctc tcccgctct tcgagatctg 2280
cctcgactgt gccctctagt tgcagcccat ctgtgttttg cccctcccc gtgccctctc 2340
tgacctgga aggtgccact cccactgttc tttcctaata aaatgaggaa attgcatcgc 2400
attgtctgag taggtgtcat tctattctgg ggggtggggg ggggcaggac agcaaggggg 2460
aggattggga agacaatagc aggcctgctg gggactcgat cgagttaagg gcgaattccc 2520
gataaggatc ttcctagagc atggctacgt agataagtag catggcgggt taatcattaa 2580
ctacaaggaa cccctagtga tggagtgggc cactccctct ctgcgcgctc gctcgcctcac 2640
tgaggccggg cgaccaaagg tcgccgacg cccgggcttt gccggggcgg cctcagtgag 2700
cgagcgcgag cgag
    
```

```

SEQ ID NO: 36      moltype = DNA length = 2715
FEATURE           Location/Qualifiers
source            1..2715
                 mol_type = other DNA
                 organism = synthetic construct
                 2487..2500
misc_binding      bound_moiety = CTCF binding site
    
```

```

SEQUENCE: 36
ctgcgcctc tgagccgccc cgggcaaac cggggcgtc ggcgacctt 60
ggtcgcctc cctcagtgag cgagcgagc cgcagagagg gagtggccaa ctccatcact 120
aggggttcct tgtagttaac gataaacccg ccatgctact tatctacgta gccatgctct 180
aggaagatcg gaattcggcc ttaagctagc tagctagtta ttaatagtaa tcaattacgg 240
ggtcattagt tcatagccca tatatggagt tccgcgttac ataacttacg gtaaatggcc 300
cgctcggctg accgcccacg gaccccggcc cattgacgtc aataatgacg tatgttccca 360
tagtaaacgc aatagggact ttccattgac gtcaatgggt ggagtattta cggtaaacgt 420
cccacttggc agtacatcaa gtgtatcata tgccaagtac gcccctatt gacgtcaatg 480
acggtaaatg gcccgccctg cattatgccc agtacatgac cttatgggac tttcctactt 540
ggcagtacat ctacgtatta gtcatcgcta ttaccatggt gatgcggttt tggcagtaaca 600
tcaatgggcg tggatagcgg tttgactcac ggggatttcc aagtctccc cccattgacg 660
tcaatgggag tttgttttgg caccaaaate aacgggactt tccaaaatgt cgtaaacaact 720
cgcgccatt gacgcaaatg ggcggtaggc gtgtacgggt ggaggcttat ataagcagag 780
ctggttagt gaaccgtcag atcctgcaga agttggtcgt gaggcactgg gcaggtaaat 840
atcaaggtta caagacaggt ttaaggagac caatagaaac tgggctgtc gagacagaga 900
agactcttgc gttctgata ggcacctatt ggtcttactg acatccactt tgctttctc 960
tccacaggtg tccagcgccc cgccatggtg agcaagggcg aggagctgt caccggggtg 1020
gtgccatcc ttgctgagct ggacggcgac gtaaacggcc acaagttcag cgtgtccggc 1080
gagggcgagg gcgatgccac ctacggcaag ctgacctga agttcatctg caccaccggc 1140
aagctgcccg tgcctggccc caccctcgtg accaccctga cctacggcgt gcagtgttc 1200
agccgtacc cggaccacat gaagcagcac gacttctca agtccgccat gccgaagggc 1260
tacgtccagg agcgcaacct cttctcaag gacgacggca actacaagac cgcgcgcgag 1320
gtgaagttcg agggcgacac cctggtgaac cgcacgagc tgaagggcat cgactcaag 1380
gaggacggca acatcctggg gcacaagctg gagtacaact acaacagcca caactctat 1440
atcatggcgc acaagcagaa aaggtgaact tcaagatccg ccacaacatc 1500
gaggacggca gcgtgcagct cgcgcaccac taccagcaga acacccccat cggcgacggc 1560
cccgctgtgc tgcgcgacaa ccactacctg agcaccactg ccgcccctgag caaagacccc 1620
aacgagaagc gcgatcacat ggtcctgctg gagttcgtga ccgcccggcg gatcaacttc 1680
ggcatggagc agctgtacaa gtaataagct tggatccaat caacctctg attacaaaat 1740
ttgtgaaga ttgactggta ttcttaacta tgttgcctct tttacgctat gtggatacgc 1800
tgcttaatg cctttgtatc atgctattgc ttcccgtatg gctttcaatt tctcctctc 1860
gtataaatcc tggttgctgt ctctttatga ggagtgtggt cccgttgtca ggcacgtgg 1920
cgtggtgtgc actgtgtttg ctgacgcaac ccccactggt tggggcctg ccaccacctg 1980
tcaagctcct tcgggagctc tcgctttccc cctccctatt gccacggcgg aactcatcgc 2040
cgctgcctt gcccgctgct ggacaggggc tcggctgttg ggcaactgaca attcctggt 2100
gttgtcgggg aaatcatcgt cctttccttg gctgctcgc tgtgttgcca cctgattctc 2160
    
```

-continued

```

gcgcgaggag tccctctgct acgtcccttc ggcctcaat ccagcggacc ttccctccc 2220
cgccctgctg cggcctctgc ggcctcttcc gcgctttcga gatctgcctc gactgtgcct 2280
tctagttgcc agccatctgt tgtttgcccc tccccctgtc ctctcctgac cctggaaggt 2340
gccactccca ctgtcctttc ctaataaaaat gaggaattg catcgattg tctgagtagg 2400
tgtcattcta ttctgggggg tgggggtggg caggacagca agggggagga ttgggaagac 2460
aatagcaggc atgctgggga ctgcaccgcc cctgggtggt cgagtttaagg gcgaattccc 2520
gataaggatc ttccctagagc atggctacgt agataagtag catggcgggt taatcattaa 2580
ctacaaggaa cccctagtga tggagttggc cactcctctc ctgcgcgctc gctcgctcac 2640
tgaggccggg cgaccaaagg tcgcccgaag cccgggcttt gccccggcgg cctcagtgag 2700
cgagcgagcg cgcag
    
```

```

SEQ ID NO: 37          moltype = DNA length = 2729
FEATURE
source                1..2729
                     mol_type = other DNA
                     organism = synthetic construct
misc_binding         211..224
                     bound_moiety = CTCF binding site
    
```

```

SEQUENCE: 37
ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc ccggcgctcg ggcgacctt 60
ggtcgccccg cctcagtgag cgagcgagcg cgcagagagg gactggccaa ctccatcact 120
aggggttctc tgtagttaat gattaaaccg ccatgctact tatctacgta gccatgctct 180
aggaagatcg gaattcgcctc ttaagctagc cacaaggtgg cgctagacta gttattaata 240
gtaatcaatt acggggctcat tagttcatag cccatcatg gagttccggc ttacataact 300
tacggtaaat ggcccgcctg gctgaccgcc caacgacccc cgcccatgga cgtcaataat 360
gacgtatggt cccatagtaa gcccaatagg gactttccat tgacgtcaat ggggtgagta 420
tttacggtaa actgcccact tggcagtaaa tcaagtgat catatgccaa gtacgcccc 480
tattgacgtc aatgacggta aatggccccg ctggcattat gcccagtaaa tgaccttatg 540
ggactttcct acttggcagt acatctacgt attagtcatc gctattacca tgggtgagcg 600
gttttggcag tacatcaatg ggcgtggata ggggtttgac tcacggggat ttccaagtct 660
ccaccccatg gacgtcaatg ggagttggtt ttggcaccaa aatcaacggg actttccaaa 720
atgtcgtaac aactccgccc cattgacgca aatggggcgt aggcgtgtac ggtgggagg 780
ctataataag agagctgggt tagtgaaccg tcagatcctg cagaagtgg tcgtgaggca 840
ctgggcaggt aagtatcaag gttacaagac aggtttaagg agaccaatag aaactgggct 900
tgtcgagaca gagaagactc ttgctgttct gataggcacc tattggtctt actgacatcc 960
actttgcctt tctctccaca ggtgtccagg cggccgccat ggtgagcaag ggcgaggagc 1020
tgttcaccgg ggtggtgccc atcctggtcg agctggacgg cgacgtaaac gccacaaagt 1080
tcagcgtgtc cggcgagggc gaggggcagc ccacctacgg caagctgacc ctgaagtcca 1140
ctgcaaccac cggcaagctg cccgtgccct ggcccaccct cgtgaccacc ctgacctacg 1200
gctgctcagtg ctccagccgc taccgccgac acatgaagca gcacgacttc ttcaagtccg 1260
ccatgcccga aggctacgtc caggagcgca ccatcttctt caaggacgac ggcaactaca 1320
agaccgcgcg cgaggtgaag ttcgaggggc acaccctggt gaaccgcac gagctgaagg 1380
gcatcgactt caaggaggac ggaacatccc tggggcacaa gctggagtag aactacaaca 1440
gcccaaacgt ctatcatg gccgacaagc agaagaacgg catcaagggt aactcaaga 1500
tccgccacaa catcgaggac ggcagcgtgc agctcgccga ccactaccga cagaacaccc 1560
ccatcgggca cggccccctg ctgctgcccc acaaccacta cctgagcacc cagtcggccc 1620
tgagcaaaag ccccaacgag aagcgcgatc acatggtcct gctggagttc gtgaccgccc 1680
ccgggatcac tctcgcatg gacgagctgt acaagtaata agcttggatc caatcaacct 1740
ctggattaca aaatttggta aagattgact ggtattctta actatgttc tccctttacg 1800
ctatggtgat acgctgcttt aatgcccctt tatcatgcta ttgcttccc tatggcttct 1860
atcttctcct ccttgataaa atcctggttg ctgtctcttt atgaggagt gtggcccggt 1920
gtcaggcaac gtggcgtggt gtcactgtgt tttgctgacg caacccccac tgggtggggc 1980
attgccacca cctgtcagct cctttccggg actttcgctt tccccctccc tattgccacg 2040
gcggaactca tcgcccctg ccttgcccgc tgtgggacag gggctcggtt gttgggcact 2100
gacaattccc tgggtgtgtc ggggaaatca tcgtccttcc cttggctgct cgctgtgtt 2160
gccacctgga ttctgcccgg gacgtccttc tgctacgtcc cttcgccct caatccagcg 2220
gacctcctt ccccgccgct gctgcccggc ctgcccgcct ttcccgctct tcgagactct 2280
cctcgactgt gccttctagt tgcagcccat ctggtgtttg cccctcccc gtgcttctc 2340
tgacctgga aggtgccact cccactgtcc tttcctaata aaatgaggaa attgcatcgc 2400
attgtctgag taggtgctat tctattctgg ggggtggggg ggggcaggac agcaaggggg 2460
aggattggga agacaatagc aggcattgct gggactcgac cgccccctg tggtegagtt 2520
aagggcgaat tcccgatagc gatcttcta gagcatggct acgtagataa gtacatggc 2580
gggttaatca ttaactacaa ggaaccctca gtgatggagt tggccactcc ctctctgccc 2640
gctcgctcgc tcactgaggc cggggaccca aaggtcggcc gacgcccggg ctttgcccgg 2700
gcggcctcag tgagcgagcg agcgcgacg
    
```

```

SEQ ID NO: 38          moltype = DNA length = 2729
FEATURE
source                1..2729
                     mol_type = other DNA
                     organism = synthetic construct
misc_binding         211..224
                     bound_moiety = CTCF binding site
misc_binding         2501..2514
                     bound_moiety = CTCF binding site
SEQUENCE: 38
    
```


-continued

```

ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc cggggcgctg ggcgacctt 60
ggtcgccccg cctcagtgag cgagcgagcg cgcagagagg gaggggccaa ctccatcact 120
aggggttcct tgtagttaat gat taaccgg ccatgctact tatctacgta gccatgctct 180
aggaagatcg gaattcgccc ttaagctagc cacaaggtgg cgcccaag gtggcgcca 240
gtaatcaatt acgggggctat tagttcatag cccatataat gagttccgag ttacataact 300
tacggtaaat gggccgctcg gctgaccgcc caacgacccc cgcccatgga cgtcaataat 360
gacgtatggt cccatagtaa cgccaatagg gactttccat tgacgtcaat gggtgagta 420
tttacggtaa actgccact tggcagtaca tcaagtgtat catatgccaa gtacgcccc 480
tattgacgto aatgacggta aatggccccc ctggcattat gccacgtaca tgaccttatg 540
ggactttcct acttggcagt acatctacgt attagtcato gctattacca tggatgagc 600
gttttgccag tacatcaatg ggcgtggata ggggtttgac tcacggggat ttccaagtct 660
ccacccatt gacgtcaatg ggagtttggt ttggaccaca aatcaacggg actttccaaa 720
atgtcgtaac aactccgccc catgacgca aatgggaggg aggcgtgtac ggtgggaggt 780
ctatataaag agagctggt tagtgaaccg tcagatcctg cagaagtgg tcgtgaggca 840
ctggggcaggt aagtatcaag gttacaagac aggtttaagg agaccaatag aaactgggct 900
tgtcgagaca gagaagactc ttgcgtttct gataggcacc tatgggtctt actgacatcc 960
actttgcctt tctctccaca ggtgtccagg cggccgccc atggtgagcaag ggcgaggagc 1020
tgttcaccgg ggtggtgccc atcctggctg agctggacgg cgacgtaaac ggcaccaagt 1080
tcagcgtgto cgggcgatg cgagggcgatg ccacctaccg caagctgacc ctgaagtta 1140
tctgaccac cggcaagctg cccgtgccc gggccaccct cgtgaccacc ctgacctacg 1200
gcgtgacgtg cttcagccgc taccggacc acatgaagca gcacgacttc tcaagtccg 1260
ccatgcccga aggtcactg caggagcgca ccatcttctt caaggacgca ggcactaca 1320
agaccggcgc cgaggtgaag ttccggggcg acacctggt gaaccgcatc gagctgaagg 1380
gcatcgactt caaggaggac ggcacatcc tggggcaca gctggagtac aactacaaca 1440
gccacaacgt ctatatcatg gccgacaagc agaagaacgg catcaagggt aactcaaga 1500
tcggccacaa catcgaggac ggcagcgtgc agctcgccga ccactaccag cagaacaccc 1560
ccatcgcgca cggccccgct ctgctgccc acaaccacta cctgagcacc cagtccgccc 1620
tgagcaaaag ccccaacgag aagcgcgac acatggtcct gctggagttc gtgaccgccc 1680
ccgggatcac tctcgcatg gacgagctgt acaagtaata agcttggatc caatcaacct 1740
ctggattaca aaatttgaag aagattgact ggtattctta actatgttgc tctttttag 1800
ctatgtgat acgctgctt aatgccttg tatcatgcta ttgcttccc tatggcttct 1860
atcttctct cctgtataa atcctggttg ctgtctctt atgaggagt gtggccgctt 1920
gtcaggcaac gtgcccgtgt gtgcaactgt tttgctgac caacccccac tggttggggc 1980
atggccacca cctgtcagct cctttccggg actttcgct tccccctcc tattgcccag 2040
gcggaactca tcgcccctg ccttgcggc gtctggacag gggetcggt gttgggcact 2100
gacaattccg tgggtgtgto ggggaaatca tctgcttct cttggctgct cgcctgtgtt 2160
gccacctgga ttctgcccgg gacgtctctc tgctacgtcc cttcggccct caatccagcg 2220
gaccttctt cccgcccgt gctgcccgt ctgcccctc ttcgctgct tcgagatctg 2280
cctcgaactg gcccttctag tgcagccat ctgtgtttg cccctcccc gtgcttctc 2340
tgacctgga aggtgccact cccactgtcc tttcctaata aaatgaggaa attgcatcgc 2400
attgtctgag taggtgtcat tctattctgg ggggtgggg ggggcaggac agcaaggggg 2460
aggattggga agacaatagc aggcactgct gggactcgag gtggtcccc gctcaggtt 2520
aagggcgaat tcccgataag gatcttcta gagcatggct acgtagataa gtacgatgac 2580
gggttaatca ttaactcaaa ggaacccta gtgatggag tggccactcc cctctgccc 2640
gctcgcctc tcaactgagc cggggcagca aaggtcgccc gacgcccgg ctttgcccgg 2700
gcgccctcag tgagcgagcg agcgcgacg 2720

```

```

SEQ ID NO: 39      moltype = DNA length = 2841
FEATURE
source            Location/Qualifiers
                  1..2841
                  mol_type = other DNA
                  organism = synthetic construct
misc_binding     211..224
                  bound_moiety = CTCF binding site
misc_binding     239..252
                  bound_moiety = CTCF binding site
misc_binding     267..280
                  bound_moiety = CTCF binding site
misc_binding     2557..2570
                  bound_moiety = CTCF binding site
misc_binding     2585..2598
                  bound_moiety = CTCF binding site
misc_binding     2613..2626
                  bound_moiety = CTCF binding site

```

```

SEQUENCE: 39
ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc cggggcgctg ggcgacctt 60
ggtcgccccg cctcagtgag cgagcgagcg cgcagagagg gaggggccaa ctccatcact 120
aggggttcct tgtagttaat gat taaccgg ccatgctact tatctacgta gccatgctct 180
aggaagatcg gaattcgccc ttaagctagc cacaaggtgg cgcccaag gtggcgcca 240
caaggtggcg cccacaaggt ggcgcccaca aggtggcgcc tagctagtta ttaatagtaa 300
caattacgg ggctattagt tcatagccca tatatggagt tccgcttac ataactaacg 360
gtaaatggcc cgccgtgctg accgcccac gacccccgcc cattgacgct aataatgacg 420
tatgttccca tagtaacgcc aatagggact ttccattgac gtcaatgggt ggagtattta 480
cggtaaaactg cccacttggc agtacctcaa gtgtatcata tgccaagtac gccccctatt 540
gacgtcaatg acggtaaatg gcccgctcgt cattatgccc agtacctgac cttatgggac 600
tttctactt ggcagtacat ctacgtatta gtcactgcta ttaecatggt gatgcccgtt 660

```

-continued

tggcagta	tcaatggg	tggatagcg	tttgactcac	ggggatttcc	aagtctccc	720
cccattgac	tcaatggg	tttgttttg	cacccaaatc	aacgggactt	tccaaaatgt	780
cgtaacaact	ccgccccatt	gacgcaaatg	ggcggtagcg	gtgtacggtg	ggaggtctat	840
ataagcagag	ctggtttagt	gaaccgtcag	atcctgcaga	agttggtcgt	gaggcactgg	900
gcaggttaagt	atcaagggtta	caagacaggt	ttaaggagac	caatagaaac	tgggcttgtc	960
gagacagaga	agactcttgc	gtttctgata	ggcacctatt	ggctctactg	acatccactt	1020
tgcccttctc	tccacaggtg	tccaggcggc	cgccatggtg	agcaaggggc	aggagctggt	1080
caccggggtg	gtgccccatc	tggtcgagct	ggacggcgac	gtaaacggcc	acaagttcag	1140
cgtgtccggc	gagggcgagg	gcatgcccac	ctacggcaag	ctgaccctga	agttcatctg	1200
caccaccggc	aagctgcccc	tgccctggcc	caccctcgtg	accaccctga	cctacggcgt	1260
gcagtgcctc	agccgctacc	ccgaccacat	gaagcagcac	gacttcttca	agtccgccat	1320
gcccgaaggc	tacgtccagg	agcgcacccat	cttcttcaag	gacgacggca	actacaagac	1380
ccgcccggag	gtgaagttcg	agggcgacac	cctgggtaac	cgcatcgagc	tgaagggcat	1440
cgacttcaag	gaggacggca	acatcctggg	gcacaagctg	gagtaacaact	acaacagcca	1500
caacgtctat	atcatggccg	acaagcagaa	gaacggcatc	aaggtgaact	tcaagatccg	1560
ccacaacatc	gaggacggca	gcgtgcagct	cgccgaccac	taccagcaga	acacccccat	1620
cgggcagcgc	cccgtgctgc	tgcccgcaca	ccactacctg	agcaccagct	ccgccctgag	1680
caaagacccc	aacgagaagc	gcatcacat	ggtcctgctg	gagttcgtga	ccgcccgccg	1740
gatcactctc	ggcatggagc	agctgtacaa	gtaataagct	tggatccaat	caacctctgg	1800
attacaaaat	ttgtgaaaga	ttgactggta	ttcttaacta	tgttgctcct	tttacgctat	1860
gtggatacgc	tgctttaatg	cctttgtatc	atgctattgc	ttcccgatg	gctttcattt	1920
ttctctcctt	gtataaaatc	tggttgctgt	ctctttatga	ggagttgtgg	cccgttgtea	1980
ggcaacgtgg	ctgagtgctg	actgtgtttg	ctgacgcaac	ccccactggt	tggggcattg	2040
ccaccacctg	tacgtcctt	tcggggactt	tcgctttccc	cctccctatt	gccacggcgg	2100
aactcatcgc	cgctgcctt	gcccgcctgt	ggacaggggc	tcggctgttg	ggcactgaca	2160
attccgtggg	ggtgtcgggg	aaatcatcgt	cctttccttg	gctgctcgcc	tggtgtgcca	2220
cctggattct	cgccgggacg	tccttctgct	acgtcccttc	ggccctcaat	ccagcgggac	2280
ttccttcccc	cgccctgctg	ccggctctgc	ggcctcttcc	gcgtcttcga	gatctgctc	2340
gactgtgctc	tctagttgcc	agccatctgt	tgtttgcccc	tccccctgtc	cttccctgac	2400
cctggaaggt	gccactccca	ctgtcctttc	ctaataaaat	gaggaaaattg	catcgctattg	2460
tctgagtagg	tgtcattcta	ttctgggggg	tggggtgggg	caggacagca	agggggagga	2520
ttgggaagac	aatagcaggc	atgctgggga	ctgacccgcc	ccctgggtggc	cgccccctgg	2580
tgcccgcccc	ctggtggccg	ccccctgggt	gcccggccct	gggtggcgag	ttaagggcga	2640
attcccagata	aggatcttcc	tagagcatgg	ctacgttagat	aagttagcatg	gcgggttaat	2700
cattaactac	aaggaaacccc	tagtgatgga	gttggccact	ccctctctgc	gcgctcgctc	2760
gctcactgag	gcccggggcgc	caaaggtcgc	ccgacgcocg	ggctttgccc	ggcgccctc	2820
agtgagcag	cgagcgcgca	g				2841

SEQ ID NO: 40 moltype = DNA length = 2730
 FEATURE Location/Qualifiers
 source 1..2730
 mol_type = other DNA
 organism = synthetic construct
 misc_binding 211..224
 bound_moiety = CTCF binding site
 misc_binding 2501..2515
 bound_moiety = CTCF binding site

SEQUENCE: 40

ctgcccgcgc	gctcgctcac	tgaggccgcc	cgggcaaaagc	ccgggcgctg	ggcgacctt	60
ggtcgccccg	cctcagtgag	cgagcagagc	cgcagagagg	gagtgcccaa	ctccatcac	120
aggggttctc	tgtagttaat	gattaaccgc	ccatgctact	tatctacgta	gccatgctct	180
aggaagatcg	gaattcgcgc	ttaaagctagc	cacaaggtgg	cgctagctga	gttattaata	240
gtaatcaatt	acggggtcat	tagttcatag	cccatatag	gagttccgcg	ttacataact	300
tacggtaaat	ggcccgctcg	gctgaccgcc	caacgacccc	cgcccatgta	cgtaacaat	360
gacgtatgtt	cccataagta	cgccaatagg	gactttccat	tgacgtcaat	gggtggagta	420
tttacggtaa	actgcccact	tggcagtaca	tcaagtgtat	catatgccaa	gtacgcccc	480
tattgacgtc	aatgacggta	aatggccccg	ctggcattat	gcccagta	tgaccttatg	540
ggactttcct	acttggcagt	acatctacgt	attagtcatc	gctattacca	tggtgatgcg	600
gttttggcag	tacatcaatg	ggcgtggata	gcggtttgac	tcacggggat	ttccaagtct	660
ccacccattt	gacgtcaatg	ggagtttgtt	ttggaccacaa	aatcaacggg	actttccaaa	720
atgtcgtaac	aactccgcgc	cattgacgca	aatggggcgt	agggcgtgtac	gggtggagggt	780
ctataaagc	agagctgggtt	tagtgaaccg	tcagatcctg	cagaagttag	tcgtgaggca	840
ctgggcaggt	aagtatcaag	gttacaagac	agggttaag	agaccaaata	aaactgggct	900
tgctgagaca	gagaagactc	ttgctttct	gataggcacc	tattggtctt	actgacatcc	960
actttgcctt	tacttccaca	gggtgccagg	cgcccgccat	ggtagcaag	ggcgaggagc	1020
tgttcaccgg	gggtgtgccc	atcctggctg	agctggacgg	cgacgtaaac	ggccacaagt	1080
tcagcgtgtc	cgccgagggc	gagggcgatg	ccacctacgg	caagctgacc	ctgaagttca	1140
tctgcaccac	cgccaagctg	cccgctccct	ggccccacct	cgtgaccacc	ctgacctacg	1200
gcgtgcagtg	cttcagccgc	taccccagac	acatgaagca	gcacgacttc	ttcaagtcgg	1260
ccatgcccca	aggctacgtc	caggagcgca	ccatcttctt	caaggacgac	ggcaactaca	1320
agaccggcgc	cgaggtgaa	ttcgaggggc	acacctgggt	gaaccgcac	gagctgaagg	1380
gcatcgactt	caaggaggac	ggcaacatcc	tggggcacaa	gctggagtac	aactacaaca	1440
gccacaacgt	ctatatcatg	gcccagaagc	agaagaacgg	catcaagggt	aactcaaga	1500
tccgccacaa	catcgaggac	ggcagcgtgc	agctcgccga	ccactaccag	cagaacaccc	1560
ccatcgggca	cgcccccgtg	ctgctgcccg	acaaccacta	cctgagcacc	cagtcggccc	1620
tgagcaaa	ccccaacgag	aagcgcgatc	acatggtcct	gctggagtcc	gtgaccggcc	1680

-continued

```

ccgggatcac tctcggcatg gacgagctgt acaagtaata agcttggatc caatcaacct 1740
ctggattaca aaatttgtga aagattgact ggtattctta actatgttgc tccttttaag 1800
ctatgtggat acgctgcttt aatgcctttg tatcatgcta ttgcttcccg tatggctttc 1860
attttctcct ccttgataaa atcctgggtg ctgtctcttt atgaggagt gtggcccggt 1920
gtcaggcaac gtggcgtggt gtgcaactgt tttgtgacg caacccccac tggttggggc 1980
attgccacca cctgtcagct cctttccggg actttcgctt tccccctccc tattgccacg 2040
gcggaactca tcgcccctg ccttgcccgc tgctggacag gggctcggt gttgggcaet 2100
gacaattccg tgggtttgtc ggggaaatca tegtctttc cttggctgct cgctgtgtt 2160
gccacctgga ttctgcgegg gacgtccttc tgctacgtcc ctteggccct caatccagcg 2220
gaccttcctt ccccgcgct getgcccgt ctgcccctc ttecgctct tcgagatctg 2280
cctcgactgt gcctctagt tgcagccat ctggtgtttg cccctcccc gtgccttct 2340
tgacctgga aggtgccact cccactgtcc tttcctaata aaatgaggaa attgcatcg 2400
attgtctgag taggtgtcat tctattctgg ggggtggggg ggggcaggac agcaaggggg 2460
aggattggga agacaatagc aggcactgtg gggactcgag cgggggggac cacctcgagt 2520
taagggcgaa ttcccgataa ggtatcttct agagcatggc tacgtagata agtagcatgg 2580
cgggttaatc attaactaca aggaaccctc agtgatggag ttggccactc cctctctgcg 2640
cgctcgctcg ctactgagg ccgggcgacc aaaggtcgcc cgacgcccgg gctttgcccc 2700
ggcggcctca gtgagcgagc gagcgcgcag 2730

```

```

SEQ ID NO: 41      moltype = DNA length = 19
FEATURE
source            Location/Qualifiers
                  1..19
                  mol_type = genomic DNA
                  organism = Adeno-associated virus
SEQUENCE: 41
actggtgtac cacagcgtt 19

```

```

SEQ ID NO: 42      moltype = DNA length = 14
FEATURE
source            Location/Qualifiers
                  1..14
                  mol_type = genomic DNA
                  organism = Homo sapiens
SEQUENCE: 42
ccgccccctg gtgg 14

```

```

SEQ ID NO: 43      moltype = DNA length = 14
FEATURE
source            Location/Qualifiers
                  1..14
                  mol_type = genomic DNA
                  organism = Homo sapiens
SEQUENCE: 43
ggtggtcccc cgcc 14

```

```

SEQ ID NO: 44      moltype = DNA length = 20
FEATURE
source            Location/Qualifiers
                  1..20
                  mol_type = genomic DNA
                  organism = Rodent protoparvovirus 1
SEQUENCE: 44
ttgctcacta gatggcgctc 20

```

```

SEQ ID NO: 45      moltype = DNA length = 21
FEATURE
source            Location/Qualifiers
                  1..21
                  mol_type = genomic DNA
                  organism = Rodent protoparvovirus 1
SEQUENCE: 45
ttcttacggt aaatcaccac c 21

```

```

SEQ ID NO: 46      moltype = DNA length = 21
FEATURE
source            Location/Qualifiers
                  1..21
                  mol_type = genomic DNA
                  organism = Rodent protoparvovirus 1
SEQUENCE: 46
ttcttacggt aaatcaccac c 21

```

```

SEQ ID NO: 47      moltype = DNA length = 19
FEATURE
source            Location/Qualifiers
                  1..19
                  mol_type = genomic DNA
                  organism = H1 parvovirus
SEQUENCE: 47
gaggcaggga accacctga 19

```

```

SEQ ID NO: 48      moltype = DNA length = 38

```

-continued

FEATURE	Location/Qualifiers	
source	1..38	
	mol_type = genomic DNA	
	organism = H1 parvovirus	
SEQUENCE: 48		
agtccaccaa gggacggagg	aggcagggaa ccacctga	38
SEQ ID NO: 49	moltype = DNA length = 19	
FEATURE	Location/Qualifiers	
source	1..19	
	mol_type = genomic DNA	
	organism = H1 parvovirus	
SEQUENCE: 49		
gaggcaggga accacctga		19
SEQ ID NO: 50	moltype = DNA length = 19	
FEATURE	Location/Qualifiers	
source	1..19	
	mol_type = genomic DNA	
	organism = Hepatitis B virus	
SEQUENCE: 50		
gaggcaggga accacctga		19
SEQ ID NO: 51	moltype = DNA length = 57	
FEATURE	Location/Qualifiers	
source	1..57	
	mol_type = genomic DNA	
	organism = Hepatitis B virus	
SEQUENCE: 51		
acctaggttg accaccagca	cctaggttga ccaccagcac ctaggttgac caccagc	57
SEQ ID NO: 52	moltype = DNA length = 38	
FEATURE	Location/Qualifiers	
source	1..38	
	mol_type = genomic DNA	
	organism = Hepatitis B virus	
SEQUENCE: 52		
acctaggttg accaccagcc	gaccaccagt tggatcca	38
SEQ ID NO: 53	moltype = DNA length = 76	
FEATURE	Location/Qualifiers	
source	1..76	
	mol_type = genomic DNA	
	organism = Hepatitis B virus	
SEQUENCE: 53		
acctaggttg accaccagcc	gaccaccagt tggatccacg accaccagtt ggatccacga	60
ccaccagttg gatcca		76
SEQ ID NO: 54	moltype = DNA length = 95	
FEATURE	Location/Qualifiers	
source	1..95	
	mol_type = genomic DNA	
	organism = Human gammaherpesvirus 4	
SEQUENCE: 54		
cacccaacag gtggtgaaac	acccaacagg tggtgaaaca cccaacaggt ggtgaaacac	60
ccaacaggtg gtgaaacacc	caacaggtg tgaaa	95
SEQ ID NO: 55	moltype = DNA length = 114	
FEATURE	Location/Qualifiers	
source	1..114	
	mol_type = genomic DNA	
	organism = Human gammaherpesvirus 4	
SEQUENCE: 55		
aaagtgggtg acaaccaca	aagtgggtga caaccacaa agtgggtggac aaccacaaa	60
gtgggtggaca acccacaag	tggtggacaa cccacaaagt ggtggacaac ccac	114
SEQ ID NO: 56	moltype = DNA length = 19	
FEATURE	Location/Qualifiers	
source	1..19	
	mol_type = genomic DNA	
	organism = Human gammaherpesvirus 4	
SEQUENCE: 56		
accgtggtgt accacggtt		19
SEQ ID NO: 57	moltype = DNA length = 38	
FEATURE	Location/Qualifiers	

-continued

```

source                1..38
                      mol_type = genomic DNA
                      organism = Human gammaherpesvirus 4
SEQUENCE: 57
ttggcaccat gtggtgceat tggcaccatg tggtgcca                38

SEQ ID NO: 58         moltype = DNA length = 19
FEATURE              Location/Qualifiers
source               1..19
                      mol_type = genomic DNA
                      organism = Carnivore protoparvovirus 1
SEQUENCE: 58
ctaaaagtgg aggaccaac                19

SEQ ID NO: 59         moltype = DNA length = 38
FEATURE              Location/Qualifiers
source               1..38
                      mol_type = genomic DNA
                      organism = Carnivore protoparvovirus 1
SEQUENCE: 59
caaccaggag gtgaaaatcc aaccaggagg tgaaaatc                38

SEQ ID NO: 60         moltype = DNA length = 38
FEATURE              Location/Qualifiers
source               1..38
                      mol_type = genomic DNA
                      organism = Human alphaherpesvirus 1
SEQUENCE: 60
ccaccggcgg ggggcggcgc caccggcggg gggcggcg                38

SEQ ID NO: 61         moltype = DNA length = 57
FEATURE              Location/Qualifiers
source               1..57
                      mol_type = genomic DNA
                      organism = Human alphaherpesvirus 1
SEQUENCE: 61
ccaccggcgg ggggcggcgc caccggcggg gggcggcgcc accggcgggg ggcggcg                57

SEQ ID NO: 62         moltype = DNA length = 19
FEATURE              Location/Qualifiers
source               1..19
                      mol_type = genomic DNA
                      organism = Primate erythroparvovirus 1
SEQUENCE: 62
atactggggg ataaccacc                19

SEQ ID NO: 63         moltype = DNA length = 38
FEATURE              Location/Qualifiers
source               1..38
                      mol_type = genomic DNA
                      organism = Primate erythroparvovirus 1
SEQUENCE: 63
atactggggg ataaccacca tactggggga taaccacc                38

SEQ ID NO: 64         moltype = DNA length = 38
FEATURE              Location/Qualifiers
source               1..38
                      mol_type = genomic DNA
                      organism = Alphapapillomavirus 9
SEQUENCE: 64
accgtggtgg accaccaata ccgtggtgga ccaccaat                38

```

What is claimed:

1. A construct for producing a recombinant adeno-associated virus (rAAV) vector, the construct comprising: a 5' inverted terminal repeat (ITR), a first CCCTC-binding factor (CTCF) binding site, a promoter, a transgene, and a 3' ITR.

2. The construct of claim 1 further comprising a second CTCF binding site.

3. The construct of claim 2, wherein the construct comprises from 5' to 3': the 5' inverted terminal repeat (ITR), the first CCCTC-binding factor (CTCF) binding site, the promoter, the transgene, the second CTCF binding site, and the 3' ITR.

4. The construct of claim 2, wherein the second CTCF binding site is in the convergent orientation relative to the first CTCF binding site.

5. The construct of claim 1, wherein the CTCF binding site(s) are from a human or a virus.

6. The construct of claim 5, wherein the virus is selected from the group consisting of:

adeno-associated virus (AAV), minute virus of mice (MVM), H1 parvovirus, MmuPV, B19, canine parvovirus, human cytomegalovirus (HCMV)/human herpesvirus 5 strain Merlin, human alphaherpesvirus 1, human herpesvirus 4 type 2 (Epstein-Barr virus type 2), HPV16, herpes simplex virus (HSV), and herpes B virus (HBV).

7. The construct of claim 5, wherein the CTCF binding site(s) comprise a sequence selected from: SEQ ID NOs:1-28.

8. The construct of claim 7, wherein the first CTCF binding site comprises SEQ ID NO:1 and the second CTCF binding site comprises SEQ ID NO:42.

9. The construct of claim 1, wherein the first and/or second CTCF binding site comprises multiple CTCF binding sequences.

10. The construct of claim 9, wherein the first and/or second CTCF binding site comprises five CTCF binding sequences.

11. The construct of claim 1, wherein the first CTCF binding site comprises SEQ ID NO: 3.

12. A host cell transduced with the construct of claim 1.

13. A modified rAAV virus particle comprising the construct of claim 1.

14. A packaging cell line for producing the virus particle of claim 13.

15. The packaging cell line of claim 14, wherein the cell line comprises the complement of any genes functionally deleted in the virus particle.

16. A method for producing a modified rAAV virus particle, the method comprising:

a) transducing a host cell with:

i. a plasmid comprising the construct of claim 1,

ii. a packaging plasmid, and

iii. a helper plasmid;

b) collecting the supernatant and the cells from culture; and

c) isolating virus particles from the supernatant and cells.

17. The method of claim 16 further comprising concentrating the virus particles.

18. A method of delivering a transgene to a subject, the method comprising: administering the rAAV virus particle of claim 13 to the subject.

19. The method of claim 18, wherein the transgene is expressed in a greater proportion of the subject's cells when it is delivered in the modified rAAV vector as compared to when it is delivered in a wild-type rAAV vector.

20. The method of claim 18, wherein the transgene is expressed at higher levels when it is delivered in the modified rAAV vector as compared to when it is delivered in a wild-type rAAV vector.

* * * * *