



(51) International Patent Classification:

A61K 38/46 (2006.01) C12N 15/00 (2006.01)

A61K 48/00 (2006.01) A61P 21/00 (2006.01)

(21) International Application Number:

PCT/US2022/033797

(22) International Filing Date:

16 June 2022 (16.06.2022)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

63/211,791 17 June 2021 (17.06.2021) US

(71) Applicant: EPICRISPR BIOTECHNOLOGIES, INC.

[US/US]; 7000 Shoreline Court, Suite 100, South San Francisco, California 94080 (US).

(72) Inventors: COLLIN DE L'HORTET, Alexandra Sylvie;

7000 Shoreline Court, Suite 100, South San Francisco, California 94080 (US). TSOA, Rosemarie Wenting; 7000 Shoreline Court, Suite 100, South San Francisco, California 94080 (US). ADHIKARI, Abhinav; 7000 Shoreline Court, Suite 100, South San Francisco, California 94080 (US).

BOREGOWDA, Siddharaju; 7000 Shoreline Court, Suite 100, South San Francisco, California 94080 (US). **HART, Daniel O.**; 7000 Shoreline Court, Suite 100, South San Francisco, California 94080 (US). **GAINOUS, Thomas Blair**; 7000 Shoreline Court, Suite 100, South San Francisco, California 94080 (US). **CAROSSO, Giovanni**; 7000 Shoreline Court, Suite 100, South San Francisco, California 94080 (US). **YANG, Xiao**; 7000 Shoreline Court, Suite 100, South San Francisco, California 94080 (US). **DALEY, Timothy**; 7000 Shoreline Court, Suite 100, South San Francisco, California 94080 (US). **LUONG, Thao Nguyen Vu**; 7000 Shoreline Court, Suite 100, South San Francisco, California 94080 (US). **LIU, Yanxia**; 7000 Shoreline Court, Suite 100, South San Francisco, California 94080 (US). **KO, Tengyu**; 7000 Shoreline Court, Suite 100, South San Francisco, California 94080 (US). **SALZMAN, Amber Ruth**; 7000 Shoreline Court, Suite 100, South San Francisco, California 94080 (US).

(74) **Agent: LEE, Sungsoo**; Wilson Sonsini Goodrich & Rosati, 650 Page Mill Road, Palo Alto, California 94304 (US).

(81) **Designated States** (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ,

(54) Title: SYSTEMS AND METHODS FOR REGULATING ABERRANT GENE EXPRESSIONS

DUX4 Repression by dCas9KAL

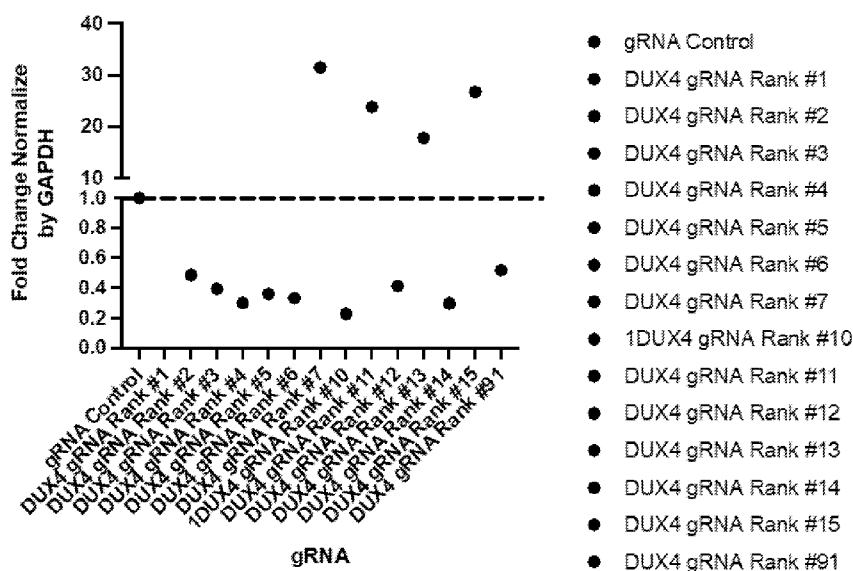


FIG. 2

(57) **Abstract:** The disclosure provides systems, compositions, methods for regulating aberrant expression of a target gene in cell (e.g., a muscle cell), to treat or ameliorate a disease or a condition in a subject (e.g., muscular dystrophy, such as Facioscapulohumeral Muscular Dystrophy (FSHD)).

CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO,
DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,
HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH,
KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA,
MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI,
NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU,
RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM,
TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM,
ZW.

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

Published:

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

SYSTEMS AND METHODS FOR REGULATING ABERRANT GENE EXPRESSIONS

CROSS REFERENCE

[0001] This application claims the benefit of U.S. Provisional Application No. 63/211,791, filed June 17, 2021, which is incorporated herein by reference in its entirety.

BACKGROUND

[0002] Aberrant expression of one or more genes can lead to a disease or a condition. In some cases, aberrant expression of a germinal transcription factor in a muscle cell can in a subject can lead to muscular dystrophy. For example, aberrant expression of a transcription factor in a muscle cell (e.g., aberrant expression of DUX4 in a skeletal muscle cell) can lead to Facioscapulohumeral Muscular Dystrophy (FSHD).

SUMMARY

[0003] Transiently modifying aberrant expression of a target gene in a cell may not be sufficient to treat or cure a disease that is manifested by the aberrant expression of the target gene. Thus, there remains a substantial need for systems and methods to modify the aberrant expression of the target gene and sustain the modified expression level of the target gene for an extended period of time.

[0004] In an aspect, the present disclosure provides a system for regulating aberrant expression of a target gene in a muscle cell, comprising: a heterologous polypeptide comprising a nuclease, wherein the nuclease has a length that is less than or equal to about 900 amino acids; and a guide nucleic acid molecule configured to form a complex with the heterologous polypeptide, wherein the guide nucleic acid molecule exhibits specific binding to a target polynucleotide sequence at or adjacent to, a D4Z4 repeat array in the muscle cell, wherein, upon formation of the complex, the complex is capable of binding the target polynucleotide sequence, to effect modification of an expression level and/or a methylation level of the target gene in the muscle cell, wherein the target gene is within the D4Z4 repeat array.

[0005] In some embodiments of any of the systems disclosed herein, upon formation of the complex, the modified expression level and/or methylation level of the target gene in the muscle cell is sustained for at least about 2 days. In some embodiments of any of the systems disclosed herein, the modified expression level and/or methylation level of the

target gene is sustained for at least about 3 days, 4 days, 5 days, 6 days, 1 week, 2 weeks, 2 weeks, 4 weeks, or 2 months. In some embodiments of any of the systems disclosed herein, the modified expression level and/or methylation level of the target gene is sustained for at least about 17 days. In some embodiments of any of the systems disclosed herein, the modified expression level and/or methylation level of the target gene is sustained for at least about 18 days.

[0006] In some embodiments of any of the systems disclosed herein, the muscle cell is in a subject having or is suspected of having facioscapulohumeral muscular dystrophy (FSHD). In some embodiments of any of the systems disclosed herein, the target gene is Dux4.

[0007] In some embodiments of any of the systems disclosed herein, the nuclease has a length that is less than or equal to about 800 amino acids. In some embodiments of any of the systems disclosed herein, the nuclease has a length that is less than or equal to about 750 amino acids.

[0008] In some embodiments of any of the systems disclosed herein, the nuclease is UnlCas12f1 or a modified variant thereof. In some embodiments of any of the systems disclosed herein, the nuclease comprises an amino acid sequence that is at least about 80%, at least about 90%, at least about 95%, or at least about 99% identical to the polypeptide sequence of SEQ ID NO: 43. In some embodiments of any of the systems disclosed herein, the nuclease comprises an amino acid sequence that is at least about 80%, at least about 90%, at least about 95%, or at least about 99% identical to the polypeptide sequence of SEQ ID NO: 44.

[0009] In some embodiments of any of the systems disclosed herein, the heterologous polypeptide further comprises a transcriptional regulator. In some embodiments of any of the systems disclosed herein, the transcriptional regulator comprises at least one methyltransferases. In some embodiments of any of the systems disclosed herein, the transcriptional regulator comprises at least one DNA Methyltransferases (DNMT). In some embodiments of any of the systems disclosed herein, the transcriptional regulator comprises DNMT-A or DNMT-L. In some embodiments of any of the systems disclosed herein, the transcriptional regulator comprises (i) DNMT-A or DNMT-L and (ii) KRAB or a variant of KRAB. In some embodiments of any of the systems disclosed herein, the transcriptional regulator comprises (i) DNMT-L and (ii) KRAB or a variant of KRAB. In some embodiments of any of the systems disclosed herein, the transcriptional regulator comprises DNMT-A, DNMT-L, and KRAB or a variant of KRAB. In some embodiments

of any of the systems disclosed herein, the transcriptional regulator comprises DNMT-L or KRAB or variant of KRAB. In some embodiments of any of the systems disclosed herein, the transcriptional regulator comprises KRAB or a variant of KRAB. In some embodiments of any of the systems disclosed herein, the transcriptional regulator comprises a plurality of different transcriptional regulators.

[0010] In some embodiments of any of the systems disclosed herein, the modification of the expression level and/or the methylation level of the target gene effects downregulation of a downstream gene of the target gene, wherein the downstream gene comprises one or more members selected from the group consisting of ZSCAN4, LEUTX, MBD3L2, TRIM48, and TRIM43.

[0011] In some embodiments of any of the systems disclosed herein, the modification of the expression level and/or the methylation level of the target gene effects downregulation of an apoptosis marker in the muscle cell. In some embodiments of any of the systems disclosed herein, the apoptosis marker comprises Caspase 3.

[0012] In some embodiments of any of the systems disclosed herein, the complex effects the modification of the expression level of the target gene in the muscle gene. In some embodiments of any of the systems disclosed herein, the modification of the expression level results in downregulation of the target gene.

[0013] In some embodiments of any of the systems disclosed herein, the complex effects the modification of the methylation level of the target gene in the muscle gene. In some embodiments of any of the systems disclosed herein, the modification of the methylation level results in downregulation of the target gene.

[0014] In some embodiments of any of the systems disclosed herein, the nuclease is a deactivated nuclease.

[0015] In another aspect, the present disclosure provides a composition comprising any of the systems disclosed herein.

[0016] In another aspect, the present disclosure provides a viral vector comprising any of the systems or any of the compositions disclosed herein.

[0017] In some embodiments of any of the viral vectors disclosed herein, the viral vector comprises an adeno-associated virus (AAVs), a retrovirus, a lentivirus, a poxvirus, or an adenovirus. In some embodiments of any of the viral vectors disclosed herein, the AAV comprises a AAV serotype RH74 AAV.

[0018] In another aspect, the present disclosure provides a method for regulating aberrant expression of a target gene in a muscle cell, the method comprising (a) contacting the

muscle cell with a complex comprising (i) a heterologous polypeptide comprising a nuclease, wherein the nuclease has a length that is less than or equal to about 900 amino acids and (ii) a guide nucleic acid molecule exhibiting specific binding to a target polynucleotide sequence at or adjacent to, a D4Z4 repeat array in the muscle cell; and (b) upon the contacting, binding the target gene with the complex to effect modification of an expression level and/or a methylation level of the target gene in the muscle cell, wherein the target gene is within the D4Z4 repeat array.

[0019] In some embodiments of any of the methods disclosed herein, upon formation of the complex, the modified expression level and/or methylation level of the target gene in the muscle cell is sustained for at least about 2 days. In some embodiments of any of the methods disclosed herein, the modified expression level and/or methylation level of the target gene is sustained for at least about 3 days, 4 days, 5 days, 6 days, 1 week, 2 weeks, 2 weeks, 4 weeks, or 2 months. In some embodiments of any of the methods disclosed herein, the modified expression level and/or methylation level of the target gene is sustained for at least about 17 days. In some embodiments of any of the methods disclosed herein, the modified expression level and/or methylation level of the target gene is sustained for at least about 18 days.

[0020] In some embodiments of any of the methods disclosed herein, the contacting comprises injecting a composition comprising the complex to a subject in need thereof, wherein the subject has or is suspected of having facioscapulohumeral muscular dystrophy (FSHD). In some embodiments of any of the methods disclosed herein, the target gene is Dux4.

[0021] In some embodiments of any of the methods disclosed herein, the nuclease has a length that is less than or equal to about 800 amino acids. In some embodiments of any of the methods disclosed herein, the nuclease has a length that is less than or equal to about 750 amino acids.

[0022] In some embodiments of any of the methods disclosed herein, the nuclease is UnlCas12f1 or a modified variant thereof.

[0023] In some embodiments of any of the methods disclosed herein, the nuclease comprises an amino acid sequence that is at least about 80%, at least about 90%, at least about 95%, or at least about 99% identical to the polypeptide sequence of SEQ ID NO: 43. In some embodiments of any of the methods disclosed herein, the nuclease comprises an amino acid sequence that is at least about 80%, at least about 90%, at least about 95%, or at least about 99% identical to the polypeptide sequence of SEQ ID NO: 44.

- [0024]** In some embodiments of any of the methods disclosed herein, the heterologous polypeptide further comprises a transcriptional regulator. In some embodiments of any of the methods disclosed herein, the transcriptional regulator comprises at least one methyltransferases. In some embodiments of any of the methods disclosed herein, the transcriptional regulator comprises at least one DNA Methyltransferases (DNMT). In some embodiments of any of the methods disclosed herein, the transcriptional regulator comprises DNMT-A or DNMT-L. In some embodiments of any of the methods disclosed herein, the transcriptional regulator comprises (i) DNMT-A or DNMT-L and (ii) KRAB or a variant of KRAB. In some embodiments of any of the methods disclosed herein, the transcriptional regulator comprises (i) DNMT-L and (ii) KRAB or a variant of KRAB. In some embodiments of any of the methods disclosed herein, the transcriptional regulator comprises DNMT-A, DNMT-L, and KRAB or variant of KRAB. In some embodiments of any of the methods disclosed herein, the transcriptional regulator comprises DNMT-L or KRAB or a variant of KRAB. In some embodiments of any of the methods disclosed herein, the transcriptional regulator comprises KRAB or a variant of KRAB. In some embodiments of any of the methods disclosed herein, the transcriptional regulator comprises a plurality of different transcriptional regulators.
- [0025]** In some embodiments of any of the methods disclosed herein, the modification of the expression level and/or the methylation level of the target gene effects downregulation of a downstream gene of the target gene, wherein the downstream gene comprises one or more members selected from the group consisting of ZSCAN4, LEUTX, MBD3L2, TRIM48, and TRIM43.
- [0026]** In some embodiments of any of the methods disclosed herein, the modification of the expression level and/or the methylation level of the target gene effects downregulation of an apoptosis marker in the muscle cell. In some embodiments of any of the methods disclosed herein, the apoptosis marker comprises Caspase 3.
- [0027]** In some embodiments of any of the methods disclosed herein, the complex effects the modification of the expression level of the target gene in the muscle gene.
- [0028]** In some embodiments of any of the methods disclosed herein, the modification of the expression level results in downregulation of the target gene.
- [0029]** In some embodiments of any of the methods disclosed herein, the complex effects the modification of the methylation level of the target gene in the muscle gene. In some embodiments of any of the methods disclosed herein, the modification of the methylation level results in downregulation of the target gene.

[0030] In some embodiments of any of the methods disclosed herein, the nuclease is a deactivated nuclease.

[0031] In another aspect, the present disclosure provides a system for regulating aberrant expression of a target gene in a muscle cell, the system comprising: a heterologous actuator moiety coupled to a gene regulator, wherein the heterologous actuator moiety is capable of forming a complex with the target gene in the muscle cell, and wherein the gene regulator is capable of modifying an expression level and/or a methylation level of the target gene in the muscle cell, wherein, upon formation of the complex, the modified expression level and/or methylation level of the target gene in the muscle cell is sustained for at least about 2 days.

[0032] In some embodiments of any of the systems disclosed herein, the sustained modified expression level and/or the methylation level of the target gene is characterized by maintaining at least about 70%, 75%, 80%, 85%, 90%, 95%, 99%, or 100% of the modified expression level and/or methylation level of the target gene.

[0033] In some embodiments of any of the systems disclosed herein, the modified expression level and/or methylation level of the target gene is sustained for at least about 3 days, 4 days, 5 days, 6 days, 1 week, 2 weeks, 3 weeks, 4 weeks, or 2 months.

[0034] In some embodiments of any of the systems disclosed herein, the modified expression level of the target gene is a decreased expression level of the target gene.

[0035] In some embodiments of any of the systems disclosed herein, the modified methylation level of the target gene is an increased degree of methylation of the target gene.

[0036] In some embodiments of any of the systems disclosed herein, the gene regulator comprises an epigenetic regulator. In some embodiments of any of the systems disclosed herein, the epigenetic regulator comprises a chromatin modifier. In some embodiments of any of the systems disclosed herein, the epigenetic regulator comprises at least one methyltransferases. In some embodiments of any of the systems disclosed herein, the epigenetic regulator comprises at least one DNA Methyltransferases (DNMT). In some embodiments of any of the systems disclosed herein, the epigenetic regulator comprises DNMT-A or DNMT-L. In some embodiments of any of the systems disclosed herein, the epigenetic regulator comprises (i) DNMT-A or DNMT-L and (ii) KRAB or a variant of KRAB. In some embodiments of any of the systems disclosed herein, the epigenetic regulator comprises DNMT-A, DNMT-L, and KRAB or a variant of KRAB. In some embodiments of any of the systems disclosed herein, the epigenetic regular comprises of

KRAB or a variant of KRAB. In some embodiments of any of the systems disclosed herein, the gene regulator comprises a plurality of different gene regulators.

[0037] In some embodiments of any of the systems disclosed herein, the system further comprises a guide nucleic acid molecule capable of directing the heterologous actuator moiety to the target gene, to form the complex.

[0038] In some embodiments of any of the systems disclosed herein, the heterologous actuator moiety is capable of forming a complex with a first portion of the muscle-regulating gene, and wherein the system further comprises an additional heterologous actuator moiety coupled to an additional gene regulator, wherein the additional heterologous actuator moiety is capable of forming a complex with a second portion of the muscle-regulating gene.

[0039] In some embodiments of any of the systems disclosed herein, the system further comprises an additional guide nucleic acid molecule capable of directing the additional heterologous actuator moiety to the second portion of the muscle-regulating gene.

[0040] In some embodiments of any of the systems disclosed herein, the target gene is a transcription factor.

[0041] In some embodiments of any of the systems disclosed herein, the target gene is within a D4Z4 repeat array. In some embodiments of any of the systems disclosed herein, the target gene encodes DUX4.

[0042] In some embodiments of any of the systems disclosed herein, the target gene is not C9orf72.

[0043] In some embodiments of any of the systems disclosed herein, the muscle cell is a skeletal muscle cell.

[0044] In some embodiments of any of the systems disclosed herein, the heterologous actuator moiety or the additional heterologous actuator moiety comprises an endonuclease. In some embodiments of any of the systems disclosed herein, the heterologous actuator moiety or the additional heterologous actuator moiety comprises a CRISPR-Cas protein. In some embodiments of any of the systems disclosed herein, the heterologous actuator moiety or the additional heterologous actuator moiety comprises a dCas protein. In some embodiments of any of the systems disclosed herein, the guide nucleic acid molecule or the additional guide nucleic acid molecule comprises a guide RNA molecule.

[0045] In another aspect, the present disclosure provides a method for regulating aberrant expression of a target gene in a muscle cell, the method comprising: (a) contacting the

muscle cell with a heterologous actuator moiety coupled to a gene regulator, wherein the heterologous actuator moiety is capable of forming a complex with the target gene in the muscle cell, and wherein the gene regulator is capable of modifying an expression level and/or a methylation level of the target gene in the muscle cell; and (b) upon formation of the complex, sustaining the modified expression level and/or methylation level of the target gene in the muscle cell for at least about 2 days.

[0046] In some embodiments of any of the methods disclosed herein, the sustained modified expression level and/or the methylation level of the target gene is characterized by maintaining at least about 70%, 75%, 80%, 85%, 90%, 95%, 99%, or 100% of the modified expression level and/or methylation level of the target gene.

[0047] In some embodiments of any of the methods disclosed herein, the modified expression level and/or methylation level of the target gene is sustained for at least about 3 days, 4 days, 5 days, 6 days, 1 week, 2 weeks, 3 weeks, 4 weeks, or 2 months.

[0048] In some embodiments of any of the methods disclosed herein, the modified expression level of the target gene is a decreased expression level of the target gene.

[0049] In some embodiments of any of the methods disclosed herein, the modified methylation level of the target gene is an increased degree of methylation of the target gene.

[0050] In some embodiments of any of the methods disclosed herein, the gene regulator comprises an epigenetic regulator. In some embodiments of any of the methods disclosed herein, the epigenetic regulator comprises a chromatin modifier. In some embodiments of any of the methods disclosed herein, the epigenetic regulator comprises at least one methyltransferases. In some embodiments of any of the methods disclosed herein, the epigenetic regulator comprises at least one DNA Methyltransferases (DNMT). In some embodiments of any of the methods disclosed herein, the epigenetic regulator comprises DNMT-A or DNMT-L. In some embodiments of any of the methods disclosed herein, the epigenetic regulator comprises (i) DNMT-A or DNMT-L and (ii) KRAB or a variant of KRAB. In some embodiments of any of the methods disclosed herein, the epigenetic regulator comprises DNMT-A, DNMT-L, and KRAB or a variant of KRAB. In some embodiments of any of the methods disclosed herein, the epigenetic regular comprises of KRAB or a variant of KRAB. In some embodiments of any of the methods disclosed herein, the gene regulator comprises a plurality of different gene regulators.

[0051] In some embodiments of any of the methods disclosed herein, the method further comprises contacting the muscle cell with a guide nucleic acid molecule capable of

directing the heterologous actuator moiety to the target gene, to form the complex.

[0052] In some embodiments of any of the methods disclosed herein, the heterologous actuator moiety is capable of forming a complex with a first portion of the muscle-regulating gene, and wherein the method further comprises contacting the muscle cell with an additional heterologous actuator moiety coupled to an additional gene regulator, wherein the additional heterologous actuator moiety is capable of forming a complex with a second portion of the muscle-regulating gene.

[0053] In some embodiments of any of the methods disclosed herein, the method further comprises contacting the muscle cell with an additional guide nucleic acid molecule capable of directing the additional heterologous actuator moiety to the second portion of the muscle-regulating gene.

[0054] In some embodiments of any of the methods disclosed herein, the target gene is a transcription factor.

[0055] In some embodiments of any of the methods disclosed herein, the target gene is within a D4Z4 repeat array. In some embodiments of any of the methods disclosed herein, the target gene encodes DUX4.

In some embodiments of any of the methods disclosed herein, the target gene is not C9orf72.

[0056] In some embodiments of any of the methods disclosed herein, the muscle cell is a skeletal muscle cell.

[0057] In some embodiments of any of the methods disclosed herein, the heterologous actuator moiety or the additional heterologous actuator moiety comprises an endonuclease. In some embodiments of any of the methods disclosed herein, the heterologous actuator moiety or the additional heterologous actuator moiety comprises a CRISPR-Cas protein. In some embodiments of any of the methods disclosed herein, the heterologous actuator moiety or the additional heterologous actuator moiety comprises a dCas protein. In some embodiments of any of the methods disclosed herein, the guide nucleic acid molecule or the additional guide nucleic acid molecule comprises a guide RNA molecule.

[0058] Additional aspects and advantages of the present disclosure will become readily apparent to those skilled in this art from the following detailed description, wherein only illustrative embodiments of the present disclosure are shown and described. As will be realized, the present disclosure is capable of other and different embodiments, and its several details are capable of modifications in various obvious respects, all without departing from the disclosure. Accordingly, the drawings and description are to be

regarded as illustrative in nature, and not as restrictive.

INCORPORATION BY REFERENCE

[0059] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[0060] The novel features of the disclosure are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present disclosure will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[0061] **FIG. 1** provides different target polynucleotide sequences (e.g., Rank #1 through Rank #91) between two CpG islands within a D4Z4 repeat array that encodes DUX4.

[0062] **FIG. 2** provides regulation of DUX4 expression in a target cell population (e.g., lymphoblasts) by a heterologous actuator moiety coupled to a gene regulator (e.g., dCas-KRAB-DNMT3A-DNMT3L) that is complexed with various guide RNA molecules target polynucleotide sequences (e.g., Rank #1 through Rank #91) within the D4Z4 repeat array that encodes DUX4.

[0063] **FIG. 3A** depicts the gene expression of DUX4 and DUX4-target genes in immortalized patient-derived human FSHD skeletal myoblasts (SkM) cells (12ABIC/12A and 15ABIC/15A). The gene expression of DUX4 and DUX4-target genes is measured in 12ABIC and 15ABIC undifferentiated cells, 12ABIC and 15ABIC cells after 2 days of differentiation, and 12ABIC and 15ABIC cells after 7 days of differentiation. Each shade of gray on the graph depicts the gene expression for a different gene corresponding to the legend on the right. **FIG. 3B** depicts the proportion of apoptotic cells in FSHD myoblasts 12ABIC and 15ABIC (right column) compared to their healthy sibling control myoblasts, 12UBIC and 15VBIC, respectively (left column) after two days of differentiation. The white dots in the images on the left represent apoptotic cells. The graph on the right depicts the percentage of apoptotic cells in the 12ABIC, 15ABIC, 12UBIC, and 15VBIC cell cultures after two days of differentiation shown in the images on the left. DAPI stain is used to stain the for nuclei. **FIG. 3C** depicts the percentage of apoptotic cells in

12ABIC, 15ABIC, 12UBIC, and 15VBIC cells after seven days of differentiation. Percentage of apoptotic cells are measured on day 0, day 1, day 2, and day 7 of differentiation. **FIG. 3D** depicts the expression of MYHC in 12ABIC, 15ABIC, 12UBIC, and 15 VBIC cells after 7 days of differentiation. Myosin Heavy Chain (MYHC) is a marker for muscle cell differentiation. The white dots indicate expression of MYHC. **FIG. 3E** depicts the expression level of MYOG, MYH2, and MYMK in 12ABIC, 15ABIC, 12UBIC, and 15VBIC cells after 7 days of differentiation. MYOG is a myogenic regulatory factor that regulates skeletal muscle differentiation and MyoMaker (MYMK) is a marker for muscle cell differentiation. DAPI stain is used to stain the for nuclei. Expression level for the 12ABIC and 15ABIC cells is measured on day 2 and day 7 of differentiation. 12A UD and 15A UD: undifferentiated, proliferating control myoblasts. Dark gray bars depicts MYOG expression levels, light gray bars depicts MYH2 expression levels, and gray bars depicts MYMK expression levels.

[0064] **FIG. 4** depicts the design of multiple gRNAs in relation to the D4Z4 repeat region. The multiple DUX4-targeting gRNAs are designed to span across the DZ4Z repeat region. The DZ4Z repeat region and the DUX4 gene locations in relation to each other is shown at the bottom of **FIG. 4**. The newly designed gRNAs are shown at the top of **FIG. 4**.

[0065] **FIG. 5** depicts the Cas12f effector-modulator vector design. The expression of the Cas12f variant, KRAB domain, and DNMT3L domain are under the control of a muscle-specific promoter, CK8e. The expression of sgRNA spacer sequence with scaffold driven by RNA polymerase III is under the control of a human U6g promoter. The vector additionally includes a modified WPRE and polyadenylation regulatory sequences.

[0066] **FIG. 6A** depicts the relative expression level of DUX4 in 12ABIC FSHD myoblasts that stably express the Cas12f-KRAB effector-modulator after 78 gRNAs were nucleofected into the 12ABIC myoblasts. Following nucleofection, the cells are cultured in differentiation conditions for 7 days before the gene expression of DUX4 is measured. The 78 gRNAs tested are listed on the x-axis and the y-axis represents the relative fold expression of DUX4. The expression level of DUX4 was normalized with the expression of control gene HPRT1. **FIG. 6B** depicts the relative expression level of DUX4 in 12ABIC FSHD myoblasts that stably express the Cas12f-KRAB effector-modulator after 78 gRNAs were nucleofected into the 12ABIC myoblasts. Following nucleofection, the cells are cultured in differentiation conditions for 7 days before the gene expression of DUX4 and the DUX4-target gene, MBD3L2, is measured.

[0067] **FIG. 7A** depicts the repression of DUX4 and DUX4-target genes, DBET/DUX4,

MBD3L2, and TRIM48, in immortalized patient-derived FSHD myoblasts transfected with six gRNAs and a Cas12f effector-modulator. The Cas12f effector-modulator expresses a Cas12f variant, a KRAB domain, and a DNMT-KLa domain. One of the six sgRNA is a control sgRNA (Empty/trcr) which did not target the DZ4Z repeat region. Expression level of MYOG is measured in the cells to assay if the differentiation ability of DUX4 sgRNA transfected cells is similar to control sgRNA transfected myoblasts. Expression level of DUX4, DUX4-target genes, and MYOG is measured 17 days post transfection. **FIG. 7B** depicts the repression of DUX4 and DUX4-target genes, DBET/DUX4, MBD3L2, and TRIM48, in immortalized patient-derived FSHD myoblasts transfected with six gRNAs and a Cas12f effector-modulator. The Cas12f effector-modulator expresses a Cas12f variant, a KRAB domain, and a DNMT-KLb domain. One of the six sgRNA is a control sgRNA (Empty) which did not target the DZ4Z repeat region. Expression level of MYOG is measured in the cells to assay if the differentiation ability of DUX4 sgRNA transfected cells is similar to control sgRNA transfected myoblasts. Expression level of DUX4, DUX4-target genes, and MYOG is measured 18 days post transfection.

[0068] FIGs. 8A and 8B depict the apoptosis level of FSHD-patient derived myoblasts transfected with Cas12f effector-modulator and DUX4-targeting gRNA. The percentage of apoptotic-positive cells is measured after two days of differentiation following transfection. The images in **FIG. 8A** depicts the proportion of apoptotic cells in control 12UBIC cells and 12ABIC cells transfected with the Cas12f effector-modulator and DUX4-targeting gRNA. The white dots depict apoptotic cells. The graphs in **FIG. 8B** depict the percentage of apoptotic cells measured in the images on the left, as well as the percentage of apoptotic cells in 12ABIC cells transfected with either a DUX4-targeting gRNA or a control gRNA, which does not target DUX4. DAPI stain is used to stain the for nuclei.

[0069] FIG. 9 depicts the workflow for an *ex vivo* FSHD model. The *ex vivo* model cultures immortalized healthy sibling control cells and FSHD skeletal myoblasts and then engineers the cells into 3D tissues. The 3D tissues are treated with either a control AAV or a AAV with the Cas12f effector-modulator vector. The 3D tissues are then tested for phenotypic differences in mechanical force, tetanic force, and fatigue, in addition to measuring 3D tissue morphology and gene expression profile.

[0070] FIG. 10 depicts the workflow for an *in vivo* xenograft model. The *in vivo* model begins with treating mice legs with irradiation and TA muscle cardiotoxin to prepare for

the transplantation of human myoblast cells into the mice's leg. Following transplantation, the mice are treated with either a control AAV or a AAV with the Cas12f effector-modulator vector. At designated time points, the mice are euthanized, and the xenograft and tissue samples are collected for analysis. The collected xenograft is fixed, sectioned, and stained with Hematoxylin and eosin. The remaining tissues are used for gene expression assays, as well as determining AAV tropism within the mice.

DETAILED DESCRIPTION

[0071] Aberrant expression of one or more genes can lead to a disease or a condition. The aberrant expression can be characterized by aberrantly low expression level of the gene(s). Alternatively, the aberrant expression can be characterized by aberrantly high expression level of the gene(s). In some cases, the gene(s) can be genetically modified (e.g., via action of endonucleases, such as CRISPR-Cas enzymes) to reverse the aberrant expression (e.g., for treatment of Duchenne muscular dystrophy (DMD)). Alternatively, the aberrant expression can be transiently modified without genetically modifying such gene(s) of interest, e.g., by targeting the gene(s) with gene effectors (e.g., deactivated CRISPR-Cas enzyme that is coupled to a gene effector). Transiently modifying aberrant expression of a target gene in a cell may not be sufficient to treat or cure a disease that is manifested by the aberrant expression of the target gene. Thus, in some embodiments, the present disclosure provides systems and methods for modifying the aberrant expression of the target gene, such that the modified expression level of the target gene may be sustained for an extended period of time.

[0072] MODIFICATION OF ABERRANT EXPRESSION OF A TARGET GENE

[0073] The present disclosure provides compositions, systems, and methods thereof for regulating aberrant expression of a target gene in a cell (e.g., a muscle cell). For example, the target gene can be within a D4Z4 repeat array. The target gene can encode at least a portion of DUX4. The compositions, systems, and methods disclosed herein can utilize at least a heterologous polypeptide (e.g., a heterologous actuator moiety, optionally with a heterologous polynucleotide such as a guide nucleic acid molecule) to modify an expression level and/or an epigenetic modification level (e.g., methylation level) of the target gene. For example, the compositions, systems, and methods disclosed herein can utilize a heterologous actuator moiety that is operatively coupled (e.g., covalently or non-covalently coupled) to a heterologous gene effector or regulator (e.g., gene actuator, gene repressor, etc.) to modify an expression level and/or an epigenetic modification level of the

target gene.

[0074] In some cases, the cell can be a muscle cell. A muscle cell as disclosed herein can be any classification of muscle cells at any state of development. The muscle cell can comprise undifferentiated muscle cells (e.g., mononucleated cells, such as muscle stem cells, muscle satellite cells, myoblasts, etc.). Alternatively or in addition to, the muscle cell can comprise differentiated muscle cells (e.g. multinucleated muscle cells, such as myotubes). The muscle cell can be a skeletal muscle cell, a cardiac muscle cell, or a smooth muscle cell. For example, the skeletal muscle cell can be a primary myoblast (e.g., an immortalized primary myoblast cell line). In some cases, the cell can be a non-muscle cell, such as a lymphoblast.

[0075] In some cases, the target gene can be in chromosome number 4 of the cell as disclosed herein. In some cases, the target gene can be in chromosome number 10 of the cell, such as a distal portion of the q (long) arm of the chromosome number 10 of the cell.

[0076] Prior to the modification of the target gene as disclosed herein, the aberrant expression of the target gene can be characterized by an expression level and/or epigenetic modification level (e.g., methylation level) of the target gene that is higher than that in a control cell (e.g., a healthy cell in a healthy subject) by at least about 1%, 5%, 10%, 20%, 50%, 100%, 150%, 200%, 300%, 400%, 500%, or more. The aberrant expression can be characterized by an expression level and/or epigenetic modification level (e.g., methylation level) of the target gene that is higher than that in a control cell (e.g., a healthy cell in a healthy subject) by at most about 500%, 400%, 300%, 200%, 150%, 100%, 50%, 20%, 10%, 5%, 1%, or less.

[0077] Prior to the modification of the target gene as disclosed herein, the aberrant expression of the target gene can be characterized by an expression level and/or epigenetic modification level (e.g., methylation level) of the target gene that is lower than that in a control cell (e.g., a healthy cell in a healthy subject) by at least about 1%, 5%, 10%, 20%, 30%, 40%, 50%, 70%, 99%, or more. The aberrant expression can be characterized by an expression level and/or epigenetic modification level (e.g., methylation level) of the target gene that is lower than that in a control cell (e.g., a healthy cell in a healthy subject) by at most about 100%, 70%, 50%, 40%, 30%, 20%, 10%, 5%, 1%, or less.

[0078] Prior to the modification of the target gene as disclosed herein, the aberrant expression of the target gene can be characterized by a duration of an expression level and/or epigenetic modification level (e.g., methylation level) of the target gene that is longer than that in a control cell (e.g., a healthy cell in a healthy subject) by at least about

1%, 5%, 10%, 20%, 50%, 100%, 150%, 200%, 300%, 400%, 500%, or more. The aberrant expression can be characterized by a duration of an expression level and/or epigenetic modification level (e.g., methylation level) of the target gene that is longer than that in a control cell (e.g., a healthy cell in a healthy subject) by at most about 500%, 400%, 300%, 200%, 150%, 100%, 50%, 20%, 10%, 5%, 1%, or less.

[0079] Prior to the modification of the target gene as disclosed herein, the aberrant expression of the target gene can be characterized by a duration of an expression level and/or epigenetic modification level (e.g., methylation level) of the target gene that is shorter than that in a control cell (e.g., a healthy cell in a healthy subject) by at least about 1%, 5%, 10%, 20%, 30%, 40%, 50%, 70%, 99%, or more. The aberrant expression can be characterized by a duration of an expression level and/or epigenetic modification level (e.g., methylation level) of the target gene that is shorter than that in a control cell (e.g., a healthy cell in a healthy subject) by at most about 100%, 70%, 50%, 40%, 30%, 20%, 10%, 5%, 1%, or less.

[0080] Subsequent to the modification of the target gene as disclosed herein, modification of the aberrant expression of the target gene can be characterized by an increased expression level and/or epigenetic modification level (e.g., methylation level) of the target gene by at least about 1%, 5%, 10%, 20%, 50%, 100%, 150%, 200%, 300%, 400%, 500%, or more, as compared to a control (e.g., without the modification). The modification of the aberrant expression of the target gene can be characterized by an increased expression level and/or epigenetic modification level (e.g., methylation level) of the target gene by at most about 500%, 400%, 300%, 200%, 150%, 100%, 50%, 20%, 10%, 5%, 1%, or less, as compared to a control.

[0081] Subsequent to the modification of the target gene as disclosed herein, modification of the aberrant expression of the target gene can be characterized by a decreased expression level and/or epigenetic modification level (e.g., methylation level) of the target gene by at least about 1%, 5%, 10%, 20%, 30%, 40%, 50%, 70%, 99%, or more, as compared to a control (e.g., without the modification). The modification of the aberrant expression of the target gene can be characterized by a decreased expression level and/or epigenetic modification level (e.g., methylation level) of the target gene by at most about 100%, 70%, 50%, 40%, 30%, 20%, 10%, 5%, 1%, or less.

[0082] Subsequent to the modification of the target gene as disclosed herein, modification of the aberrant expression of the target gene can be characterized by an increased duration of an expression level and/or epigenetic modification level (e.g., methylation level) of the

target gene by at least about 1%, 5%, 10%, 20%, 50%, 100%, 150%, 200%, 300%, 400%, 500%, or more, as compared to a control (e.g., without the modification). The modification of the aberrant expression of the target gene can be characterized by an increased duration of an expression level and/or epigenetic modification level (e.g., methylation level) of the target gene by at most about 500%, 400%, 300%, 200%, 150%, 100%, 50%, 20%, 10%, 5%, 1%, or less, as compared to a control.

[0083] Subsequent to the modification of the target gene as disclosed herein, modification of the aberrant expression of the target gene can be characterized by a decreased duration of an expression level and/or epigenetic modification level (e.g., methylation level) of the target gene by at least about 1%, 5%, 10%, 20%, 30%, 40%, 50%, 70%, 99%, or more, as compared to a control (e.g., without the modification). The modification of the aberrant expression of the target gene can be characterized by a decreased duration of an expression level and/or epigenetic modification level (e.g., methylation level) of the target gene by at most about 100%, 70%, 50%, 40%, 30%, 20%, 10%, 5%, 1%, or less.

[0084] Subsequent to the modification of the target gene as disclosed herein, the modified expression level and/or epigenetic modification level (e.g., methylation level) of the target gene (e.g., the aberrantly expressed target gene) can be sustained for at least about 1 day, 2 days, 3 days, 4 days, 5 days, 6 days, 7 days, 2 weeks, 3 weeks, 4 weeks, 2 months, 3 months, 4 months, 5 months, 6 months, 7 months, 8 months, 9 months, 10 months, 11 months, 12 months, 2 years, 3 years, 4 years, 5 years, or more. Subsequent to the modification of the target gene as disclosed herein, the modified expression level and/or epigenetic modification level (e.g., methylation level) of the target gene (e.g., the aberrantly expressed target gene) can be sustained for at most about 5 years, 4 years, 3 years, 2 years, 12 months, 11 months, 10 months, 9 months, 8 months, 7 months, 6 months, 5 months, 4 months, 3 months, 2 months, 4 weeks, 3 weeks, 2 weeks, 7 days, 6 days, 5 days, 4 days, 3 days, 2 days, 1 day, or less.

[0085] Subsequent to the modification of the target gene as disclosed herein, the modified expression level and/or epigenetic modification level (e.g., methylation level) of the target gene (e.g., the aberrantly expressed target gene) can be sustained for at least about 1 cell division, at least about 2 cell divisions, at least about 3 cell divisions, at least about 4 cell divisions, at least about 5 cell divisions, at least about 6 cell divisions, at least about 7 cell divisions, at least about 8 cell divisions, at least about 9 cell divisions, at least about 10 cell divisions, at least about 15 cell divisions, at least about 20 cell divisions, at least about 25 cell divisions, at least about 30 cell divisions, at least about 40 cell divisions, at least

about 50 cell divisions, or at least about 100 cell divisions. Subsequent to the modification of the target gene as disclosed herein, the modified expression level and/or epigenetic modification level (e.g., methylation level) of the target gene (e.g., the aberrantly expressed target gene) can be sustained for at most about 100 cell divisions, at most about 50 cell divisions, at most about 40 cell divisions, at most about 30 cell divisions, at most about 25 cell divisions, at most about 20 cell divisions, at most about 15 cell divisions, at most about 10 cell divisions, at most about 9 cell divisions, at most about 8 cell divisions, at most about 7 cell divisions, at most about 6 cell divisions, at most about 5 cell divisions, at most about 4 cell divisions, at most about 3 cell divisions, at most about 2 cell division, or at most about 1 cell division.

[0086] As disclosed herein, non-limiting examples of the epigenetic modification can include methylation, acetylation, phosphorylation, ADP-ribosylation, glycosylation, SUMOylation, ubiquitination, modification of histone structure (e.g., via an ATP hydrolysis-dependent process). For example, the epigenetic modification can result in a modified methylation level of one or more target genes.

[0087] As disclosed herein, the sustained modified expression level and/or epigenetic modification level (e.g., methylation level) of the target gene can be characterized by maintaining at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% of the modified expression level and/or methylation level of the target gene. The sustained modified expression level and/or epigenetic modification level (e.g., methylation level) of the target gene can be characterized by maintaining at most about 100%, 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80%, 75%, or 70% of the modified expression level and/or methylation level of the target gene.

[0088] The systems, compositions, and methods as disclosed herein can be used to treat or ameliorate a disease (e.g., muscular dystrophy, such as Facioscapulohumeral Muscular Dystrophy (FSHD)) of a subject.

[0089] HETEROLOGOUS POLYPEPTIDES

[0090] The heterologous polypeptide as disclosed herein, either alone or in conjunction with one or more co-agents such as a heterologous polynucleotide (e.g., a guide nucleic acid) can be configured to specifically bind a target polynucleotide sequence, to modulate an expression level and/or an epigenetic level of the target gene (e.g., the D4Z4 repeat array) in the target cell, as disclosed herein. The target polynucleotide sequence can be at (e.g., within) the target gene. Alternatively, the target polynucleotide sequence can be adjacent to the target gene. For example, the target polynucleotide sequence can be adjacent to an

end (e.g., a 5' end or a 3' end) of the target gene. The target polynucleotide sequence can be at least about 5 nucleobases, at least about 10 nucleobases, at least about 20 nucleobases, at least about 30 nucleobases, at least about 40 nucleobases, at least about 50 nucleobases, at least about 100 nucleobases, at least about 150 nucleobases, at least about 200 nucleobases, at least about 250 nucleobases, at least about 300 nucleobases, at least about 400 nucleobases, at least about 500 nucleobases, at least about 1,000 nucleobases, at least about 1,500 nucleobases, at least about 2,000 nucleobases, at least about 3,000 nucleobases, at least about 4,000 nucleobases, or at least about 5,000 nucleobases away from the end of the target gene. The target polynucleotide sequence can be at most about 5,000 nucleobases, at most about 4,000 nucleobases, at most about 3,000 nucleobases, at most about 2,000 nucleobases, at most about 1,500 nucleobases, at most about 1,000 nucleobases, at most about 500 nucleobases, at most about 400 nucleobases, at most about 300 nucleobases, at most about 200 nucleobases, at most about 150 nucleobases, at most about 100 nucleobases, at most about 50 nucleobases, at most about 40 nucleobases, at most about 30 nucleobases, at most about 20 nucleobases, at most about 10 nucleobases, or at most about 5 nucleobases away from the end of the target gene.

[0091] Without wishing to be bound by theory, when the target polynucleotide sequence is not within the target gene, the target polynucleotide sequence can interact (e.g., via direct or indirect binding) with at least a portion of the target gene (e.g., a promoter sequence of the target gene), such that binding or targeting of the target polynucleotide sequence by at least the heterologous polypeptide (e.g., by a complex comprising the heterologous polypeptide and the heterologous polynucleotide as disclosed herein) can target the at least the portion of the target gene (e.g., the promoter sequence), to effect the modulation of the expression level and/or the epigenetic level of the target gene in the cell.

[0092] In some cases, the target polynucleotide sequence can comprise a plurality of target polynucleotide sequences. The plurality of target polynucleotide sequences can be within the target gene. Alternatively, the plurality of target polynucleotide sequences can be outside but adjacent to the target gene, as disclosed herein. Yet in another alternative, the plurality of target polynucleotide sequences can comprise at least one target polynucleotide sequence within the target gene (e.g., within the D4Z4 repeat domain) and at least one additional target polynucleotide sequence that is outside of but adjacent to the target gene. In such case, targeting both the at least one target polynucleotide sequence and the at least one additional target polynucleotide sequence may yield a greater effect (e.g., greater degree of modulation of the expression and/or epigenetic level of the target

gene) (e.g., by at least 0.1-fold, at least 0.5-fold, at least 1-fold, at least 2-fold, at least 3-fold, at least 4-fold, at least 5-fold, at least 10-fold, at least 15-fold, at least 20-fold, or more) as compared to that by targeting only one of the at least one target polynucleotide sequence and the at least one additional target polynucleotide sequence.

[0093] The heterologous polypeptide as disclosed herein can comprise one or more heterologous gene effectors (e.g., gene effectors that are heterologous to a cell comprising the gene effectors and/or another component in a complex of the disclosure). Heterologous gene effectors can comprise domains that are capable of, or are candidates for, modulating expression of a target gene (e.g., a target endogenous gene), for example, activating, repressing, upregulating, downregulating, or stabilizing an expression level or activity level of the gene. Heterologous gene effectors can be heterologous with respect to another component that is present in a complex, for example, a guide moiety (e.g., nuclease and/or guide nucleic acid, as disclosed herein). In some cases, heterologous gene effectors can be heterologous with respect to a host cell they are introduced to.

[0094] A heterologous gene effector can be or can comprise a sequence from any suitable source, for example, an amino acid sequence from a human protein, viral protein, or other protein as disclosed herein. A heterologous gene effector can be or can comprise a sequence from a protein that primarily localized to the nucleus, for example, a member of the human nuclear proteome. A heterologous gene effector can be or can comprise one or more natural amino acid residues. A heterologous gene effector can be or can comprise one or more synthetic amino acid residues.

[0095] A heterologous gene effector can be or can comprise a sequence from a mammalian protein. A heterologous gene effector can be or can comprise a sequence from a human protein. A heterologous gene effector can be or can comprise a sequence from a viral protein. A heterologous gene effector can be or can comprise a sequence from a non-human primate protein. A heterologous gene effector can be or can comprise a sequence from a non-human mammal protein. A heterologous gene effector can be or can comprise a sequence from a non-rodent mammal protein. A heterologous gene effector can be or can comprise a sequence from a plant protein. A heterologous gene effector can be or can comprise a sequence from a pig protein. A heterologous gene effector can be or can comprise a sequence from a lagomorph protein. A heterologous gene effector can be or can comprise a sequence from a canine protein. A heterologous gene effector can be or can comprise a sequence from an avian protein. A heterologous gene effector can be or can comprise a sequence from a reptilian protein. A heterologous gene effector can be or

can comprise a sequence from a bacterial protein. A heterologous gene effector can be or can comprise a sequence from an archaeal protein.

[0096] For example, the amino acid sequence of the heterologous gene effector as disclosed herein may not and need not be derived from a bacterial protein (e.g., may be derived from an archaeal protein). Without wishing to be bound by theory, a subject in need thereof may be treated with a composition comprising the non-bacterial protein-derived heterologous gene effector, such that the composition may not (i) induce a bacterial stimulus in the subject and/or (ii) elicit a bacterial immune response in the subject.

[0097] The heterologous actuator moiety can comprise a nuclease (e.g., an endonuclease). For example, the nuclease can be a CRISPR/Cas protein. The nuclease can have a length that is less than a threshold length. The threshold length can be at most about 1,000 amino acids, at most about 950 amino acids, at most about 900 amino acids, at most about 850 amino acids, at most about 800 amino acids, at most about 750 amino acids, at most about 700 amino acids, at most about 650 amino acids, at most about 600 amino acids, at most about 550 amino acids, at most about 500 amino acids, at most about 450 amino acids, at most about 400 amino acids, at most about 350 amino acids, or at most about 300 amino acids. The threshold length can be at least about 300 amino acids, at least about 350 amino acids, at least about 400 amino acids, at least about 450 amino acids, at least about 500 amino acids, at least about 550 amino acids, at least about 600 amino acids, at least about 650 amino acids, at least about 700 amino acids, at least about 750 amino acids, at least about 800 amino acids, at least about 850 amino acids, at least about 900 amino acids, at least about 950 amino acids, or at least about 1,000 amino acids.

[0098] Without wishing to be bound by theory, using a size of the nuclease to be less than the threshold length can have one or advantages over using a control nuclease having a size greater than the threshold length. When using a delivery vehicle having a limited size (e.g., a limited physical size to entrap the nuclease or a limited expression cassette size, such as a viral genome) can leave sufficient room (or sufficient space within the expression cassette) for one or more co-agents, such as one or more gene regulators (e.g., transcriptional regulator) and/or one or more heterologous polynucleotides (e.g., one or more guide nucleic acid molecules). Alternatively or in addition to, using the nuclease having a size less than or equal to the threshold size can elicit a greater effect on the modulation of the expression level and/or the epigenetic level of the target gene, as compared to the effect on the modulation of the expression level and/or the epigenetic level of the target gene by a control nuclease having a size greater than the threshold size.

[0099] In some examples, the degree of modulation (e.g., increase or decrease) of the expression level and/or the epigenetic level of the target gene by the nuclease as disclosed herein (e.g., having a size less than or equal to the threshold size) can be greater than that by the control nuclease (e.g., having a size greater than the threshold size) by at least or up to about 0.1-fold, at least or up to about 0.5-fold, at least or up to about 1-fold, at least or up to about 2-fold, at least or up to about 3-fold, at least or up to about 4-fold, at least or up to about 5-fold, at least or up to about 6-fold, at least or up to about 7-fold, at least or up to about 8-fold, at least or up to about 9-fold, at least or up to about 10-fold, at least or up to about 15-fold, at least or up to about 20-fold, at least or up to about 25-fold, at least or up to about 30-fold, at least or up to about 40-fold, at least or up to about 50-fold, at least or up to about 60-fold, at least or up to about 70-fold, at least or up to about 80-fold, at least or up to about 90-fold, or at least or up to about 100-fold.

[0100] In some examples, the degree of modulation (e.g., increase or decrease) of the expression level and/or the epigenetic level of the target gene by the nuclease as disclosed herein (e.g., having a size less than or equal to the threshold size) can persist longer than that by the control nuclease (e.g., having a size greater than the threshold size) by at least or up to about 0.1-fold, at least or up to about 0.5-fold, at least or up to about 1-fold, at least or up to about 2-fold, at least or up to about 3-fold, at least or up to about 4-fold, at least or up to about 5-fold, at least or up to about 6-fold, at least or up to about 7-fold, at least or up to about 8-fold, at least or up to about 9-fold, at least or up to about 10-fold, at least or up to about 15-fold, at least or up to about 20-fold, at least or up to about 25-fold, at least or up to about 30-fold, at least or up to about 40-fold, at least or up to about 50-fold, at least or up to about 60-fold, at least or up to about 70-fold, at least or up to about 80-fold, at least or up to about 90-fold, or at least or up to about 100-fold.

[0101] In some examples, the degree of modulation (e.g., increase or decrease) of the expression level and/or the epigenetic level of the target gene by the nuclease as disclosed herein (e.g., having a size less than or equal to the threshold size) can persist longer than (or sustained longer than) that by the control nuclease (e.g., having a size greater than the threshold size) by at least or up to about 1 cell division, at least or up to about 2 cell divisions, at least or up to about 3 cell divisions, at least or up to about 4 cell divisions, at least or up to about 5 cell divisions, at least or up to about 6 cell divisions, at least or up to about 7 cell divisions, at least or up to about 8 cell divisions, at least or up to about 9 cell divisions, at least or up to about 10 cell divisions, at least or up to about 11 cell divisions, at least or up to about 12 cell divisions, at least or up to about 13 cell divisions, at least or

up to about 14 cell divisions, at least or up to about 15 cell divisions, at least or up to about 16 cell divisions, at least or up to about 17 cell divisions, at least or up to about 18 cell divisions, at least or up to about 19 cell divisions, at least or up to about 20 cell divisions, at least or up to about 25 cell divisions, at least or up to about 30 cell divisions, at least or up to about 40 cell divisions, at least or up to about 50 cell divisions, or at least about 100 cell divisions. As disclosed herein, a cell division can be characterized by a division of a parent cell into two daughter cells with substantially the same genetic material as the parent cell.

- [0102] A heterologous gene effector can be or can comprise a sequence from a chromatin regulator (CR). Chromatin regulators include functional domains from various classes of histone and DNA modifying enzymes (e.g., DNMTs, HATs, HMTs, etc.).
- [0103] A heterologous gene effector can comprise two or more domains from chromatin regulators, e.g., located at a C-terminus, an N-terminus, or within a polypeptide sequence, in tandem or separate.
- [0104] In some embodiments, a heterologous gene effector that facilitates heterochromatin formation. Non-limiting examples of proteins that can facilitate heterochromatin formation include HP1 α , HP1 β , KAP1, KRAB, SUV39H1, and G9a.
- [0105] In some embodiments, a heterologous gene effector modulates histones through methylation. In some embodiments, a heterologous gene effector modulates histones through acetylation. In some embodiments, a heterologous gene effector modulates histones through phosphorylation. In some embodiments, a heterologous gene effector modulates histones through ADP-ribosylation. In some embodiments, a heterologous gene effector modulates histones through glycosylation. In some embodiments, a heterologous gene effector modulates histones through SUMOylation. In some embodiments, a heterologous gene effector modulates histones through ubiquitination. In some embodiments, a heterologous gene effector modulates histones by remodeling histone structure, e.g., via an ATP hydrolysis-dependent process.
- [0106] In some embodiments, a heterologous gene effector facilitates spatial positioning of proteins on or near the target polynucleotide, e.g., transcriptional repressors, transcription factors, histones, etc. In some embodiments, a heterologous gene effector is useful for manipulating the spatiotemporal organization of genomic DNA and RNA components in the nucleus and/or cytoplasm, e.g., for regulating diverse cellular functions.
- [0107] In some embodiments, a heterologous gene effector is from a histone acetyltransferase. Non-limiting examples of histone acetyltransferases include GNAT

subfamily, MYST subfamily, p300/CBP subfamily, HAT1 subfamily, GCN5, PCAF, Tip60, MOZ, MORF, MOF, HBO1, p300, CBP, HAT1, ATF-2, SRC1, and TAFII250.

- [0108]** In some embodiments, a heterologous gene effector is from a histone lysine methyltransferase. Non-limiting examples of histone lysine methyltransferases include EZH subfamily, Non-SET subfamily, Other SET subfamily, PRDM subfamily, SET1 subfamily, SET2 subfamily, SUV39 subfamily, SYMD subfamily, ASH1L, EHMT1, EHMT2, EZH1, EZH2, MLL, MLL2, MLL3, MLL4, MLL5, NSD1, NSD2, NSD3, PRDM1, PRDM10, PRDM11, PRDM12, PRDM13, PRDM14, PRDM15, PRDM16, PRDM2, PRDM4, PRDM5, PRDM6, PRDM7, PRDM8, PRDM9, SET1, SET1L, SET2L, SETD2, SETD3, SETD4, SETD5, SETD6, SETD7, SETD8, SETDB1, SETDB2, SETMAR, SUV39H1, SUV39H2, SUV420H1, SUV420H2, SYMD1, SYMD2, SYMD3, SYMD4, and SYMD5.
- [0109]** In some embodiments, a heterologous gene effector is from a component of a chromatin remodeling complex. In some embodiments, a heterologous gene effector is a component of BAF, for example, Actin, ARIDA/B, BAF155, BAF170, BAF45 A/B/C/D, BAF53 A/B, BAF57, BAF60 A/B/C, BRG1/BRM, INI1, or SS18.
- [0110]** In some embodiments, a heterologous gene effector is from a component of PBAF, for example, Actin, ARID2, BAF155, BAF170, BAF180, BAF45 A/B/C/D, BAF53 A/B, BAF57, BAF60 A/B/C, BRD7, BRG1, or INI1.
- [0111]** In some embodiments, a heterologous gene effector is from a component of an ISWI family chromatin remodeling complex, for example, ACF subfamily, RSF subfamily, CERF subfamily, CHRAC subfamily, NURF subfamily, NoRC subfamily, WICH subfamily, b-WICH subfamily, ACF1, ATPase, BPTF, CECR2, CHRAC15, CHRAC17, CSB, DEK, MYBBP1A, NM1, RBAP46/48, RHII/Gua, RSF1, SAP155, SNF2H, SNF2H/L, SNF2L, TIP5, or WSTF.
- [0112]** In some embodiments, a heterologous gene effector is from a component of a CHD family complex, for example, a NuRD complex, NuRD-like complex, or CHD complex. In some embodiments, a heterologous gene effector is from CHD1/2/6/7/8/9, CHD3/4, CHD5, GATAD2 A/B, GATAD2 B, HDAC1, HDAC2, HDAC2, MBD2/3, MTA1/2/3, MTA3, or RBAP46, RBAP46/48.
- [0113]** In some embodiments, a heterologous gene effector is from a component of an INO80 family complex, for example, from an INO80 complex, Tip60/p400 complex, SRCAP complex, AMIDA, ARP6, BAF53, BAF53, BAF53A, BRD8, DMAP1, DMAP1, EPC1/2, FLJ11730, GAS41, GAS41, IES2, IES6, ING3, INO80, INO80E, MCERS1,

MRG15, MRGBP, MRGX, NFRKB, p400, RUVBL1/2, RUVBL1/2, RUVBL1/2, SRCAP, Tip60, TRRAP, UCH37, YL-1, YY1, or ZnF-HIT1.

- [0114] A heterologous gene effector can be or can comprise a sequence from a transcriptional regulator (TR). TR gene effectors include transcriptional regulatory domains from various families of transcription factors (e.g. KRAB, p65, MED, GTFs, etc.).
- [0115] A heterologous gene effector can comprise a transcriptional activator domain. A heterologous gene effector can comprise two or more tandem transcriptional activation domains, e.g., located at a C-terminus, an N-terminus, or within a polypeptide sequence.
- [0116] Non-limiting examples of transcriptional activation domains include GAL4, herpes simplex activation domain VP16, VP64 (a Tetramer of the herpes simplex activation domain VP16), NF-KB p65 subunit, Epstein-Barr virus R transactivator (Rta). In some embodiments, such transcriptional activation domains are used as controls in methods of the disclosure. In some embodiments, such transcriptional activation domains are used as one heterologous gene effector in a complex that comprises at least one additional heterologous gene effector (e.g., a different effector).
- [0117] A heterologous gene effector can comprise a transcriptional repressor domain. A heterologous gene effector can comprise two or more transcriptional repressor domains, e.g., located at a C-terminus, an N-terminus, or within a polypeptide sequence, in tandem or separate.
- [0118] Non-limiting examples of transcriptional repressor domains include the KRAB (Kruppel-associated box) domain of Koxl, the Mad mSIN3 interaction domain (SID), and ERF repressor domain (ERD). In some embodiments, such transcriptional repressor domains are used as controls in methods of the disclosure. In some embodiments, such transcriptional repressor domains are used as one heterologous gene effector in a complex that comprises at least one additional heterologous gene effector (e.g., a different effector).
- [0119] In some embodiments, a heterologous gene effector is from a gene product that is a transcription factor.
- [0120] In some embodiments, a heterologous gene effector is from a gene product that is a hematopoietic stem cell transcription factor. Non-limiting examples of hematopoietic stem cell transcription factors include AHR, Aiolos/IKZF3, CDX4, CREB, DNMT3A, DNMT3B, EGR1, FoxO3, GATA-1, GATA-2, GATA-3, Helios, HES-1, HHEX, HIF-1

alpha/HIF1A, HMGB1/HMG-1, HMGB3, Ikaros, c-Jun, LMO2, LMO4, c-Maf, MafB, MEF2C, MYB, c-Myc, NFATC2, NFIL3/E4BP4, Nrf2, p53, PITX2, PRDM16/MEL1, Prox1, PU.1/Spi-1, RUNX1/CBFA2, SALL4, SCL/Tal1, Smad2, Smad2/3, Smad4, Smad7, Spi-B, STAT Activators, STAT Inhibitors, STAT3, STAT4, STAT5a, STAT6, and TSC22.

[0121] In some embodiments, a heterologous gene effector is from a gene product that is a mesenchymal stem cell transcription factor. Non-limiting examples of mesenchymal stem cell transcription factors include DUX4, DUX4/DUX4c, DUX4c, EBF-1, EBF-2, EBF-3, ETV5, FoxC2, FoxF1, GATA-4, GATA-6, HMGA2, c-Jun, MYF-5, Myocardin, MyoD, Myogenin, NFATC2, p53, Pax3, PDX-1/IPF1, PLZF, PRDM16/MEL1, RUNX2/CBFA1, Smad1, Smad3, Smad4, Smad5, Smad8, Smad9, Snail, SOX2, SOX9, SOX11, STAT Activators, STAT Inhibitors, STAT1, STAT3, TBX18, Twist-1, and Twist-2.

[0122] In some embodiments, a heterologous gene effector is from a gene product that is an embryonic stem cell transcription factor. Non-limiting examples of embryonic stem cell transcription factors include Brachyury, EOMES, FoxC2, FoxD3, FoxF1, FoxH1, FoxO1/FKHR, GATA-2, GATA-3, GBX2, Goosecoid, HES-1, HNF-3 alpha/FoxA1, c-Jun, KLF2, KLF4, KLF5, c-Maf, Max, MEF2C, MIXL1, MTF2, c-Myc, Nanog, NFkB/IkB Activators, NFkB/IkB Inhibitors, NFkB1, NFkB2, Oct-3/4, Otx2, p53, Pax2, Pax6, PRDM14, Rex-1/ZFP42, SALL1, SALL4, Smad1, Smad2, Smad2/3, Smad3, Smad4, Smad5, Smad8, Snail, SOX2, SOX7, SOX15, SOX17, STAT Activators, STAT Inhibitors, STAT3, SUZ12, TBX6, TCF-3/E2A, THAP11, UTF1, WDR5, WT1, ZNF206, and ZNF281.

[0123] In some embodiments, a heterologous gene effector is from a gene product that is an induced pluripotent stem cell (iPSC) transcription factor. Non-limiting examples of iPSC transcription factors include KLF2, KLF4, c-Maf, c-Myc, Nanog, Oct-3/4, p53, SOX1, SOX2, SOX3, SOX15, SOX18, and TBX18.

[0124] In some embodiments, a heterologous gene effector is from a gene product that is an epithelial stem cell transcription factor. Non-limiting examples of epithelial stem cell transcription factors include ASCL2/Mash2, CDX2, DNMT1, ELF3, Ets-1, FoxM1, FoxN1, GATA-6, Hairless, HNF-4 alpha/NR2A1, IRF6, c-Maf, MITF, Miz-1/ZBTB17, MSX1, MSX2, MYB, c-Myc, Neurogenin-3, NFATC1, NKX3.1, Nrf2, p53, p63/TP73L, Pax2, Pax3, RUNX1/CBFA2, RUNX2/CBFA1, RUNX3/CBFA3, Smad1, Smad2, Smad2/3, Smad4, Smad5, Smad7, Smad8, Snail, SOX2, SOX9, STAT Activators, STAT Inhibitors, STAT3, SUZ12, TCF-3/E2A, and TCF7/TCF1.

- [0125]** In some embodiments, a heterologous gene effector is from a gene product that is a cancer stem cell transcription factor. Non-limiting examples of cancer stem cell transcription factors include Androgen R/NR3C4, AP-2 gamma, beta-Catenin, beta-Catenin Inhibitors, Brachyury, CREB, ER alpha/NR3A1, ER beta/NR3A2, FoxM1, FoxO3, FRA-1, GLI-1, GLI-2, GLI-3, HIF-1 alpha/HIF1A, HIF-2 alpha/EPAS1, HMGA1B, c-Jun, JunB, KLF4, c-Maf, MCM2, MCM7, MITF, c-Myc, Nanog, NFkB/IkB Activators, NFkB/IkB Inhibitors, NFkB1, NKX3.1, Oct-3/4, p53, PRDM14, Snail, SOX2, SOX9, STAT Activators, STAT Inhibitors, STAT3, TAZ/WWTR1, TBX3, Twist-1, Twist-2, WT1, and ZEB1.
- [0126]** In some embodiments, a heterologous gene effector is from a gene product that is a cancer-related transcription factor. Non-limiting examples of cancer-related transcription factors include ASCL1/Mash1, ASCL2/Mash2, ATF1, ATF2, ATF4, BLIMP1/PRDM1, CDX2, CDX4, DLX5, DNMT1, E2F-1, EGR1, ELF3, Ets-1, FosB/G0S3, FoxC1, FoxC2, FoxF1, GADD153, GATA-2, HMGA2, HMGB1/HMG-1, HNF-3 alpha/FoxA1, HNF-6/ONECUT1, HSF1, ID1, ID2, JunD, KLF10, KLF12, KLF17, LMO2, MEF2C, MYCL1/L-Myc, NFkB2, Oct-1, p63/TP73L, Pax3, PITX2, Prox1, RAP80, Rex-1/ZFP42, RUNX1/CBFA2, RUNX3/CBFA3, SALL4, SCL/Tal1, Sirtuin 2/SIRT2, Smad3, Smad4, Smad5, SOX11, STAT5a/b, STAT5a, STAT5b, TCF7/TCF1, TORC1, TORC2, TRIM32, TRPS1, and TSC22.
- [0127]** In some embodiments, a heterologous gene effector is from a gene product that is an immune cell transcription factor. Non-limiting examples of immune cell transcription factors include AP-1, Bcl6, E2A, EBF, Eomes, FoxP3, GATA3, Id2, Ikaros, IRF, IRF1, IRF2, IRF3, IRF7, NFAT, NFkB, Pax5, PLZF, PU.1, ROR-gamma-T, STAT, STAT1, STAT2, STAT3, STAT4, STAT5, STAT5A, STAT5B, STAT6, T-bet, TCF7, and ThPOK.
- [0128]** In some embodiments, a heterologous gene effector is from a gene product that is a RNA polymerase related protein. In some embodiments, a heterologous gene effector is from a transcription factor with a basic domain. In some embodiments, a heterologous gene effector is from a transcription factor with a zinc-coordinated DNA binding domain. In some embodiments, a heterologous gene effector is from a transcription factor with a helix-turn-helix domain. In some embodiments, a heterologous gene effector is from a transcription factor with an alpha helical DNA binding domain. In some embodiments, a heterologous gene effector is from a transcription factor with an alpha helix exposed by beta structures. In some embodiments, a heterologous gene effector is from a transcription

factor with an immunoglobulin fold. In some embodiments, a heterologous gene effector is from a transcription factor with a with a beta-Hairpin exposed by an alpha/beta-scaffold. In some embodiments, a heterologous gene effector is from a transcription factor with a beta sheet binding to DNA. In some embodiments, a heterologous gene effector is from a transcription factor with a beta barrel DNA binding domain.

[0129] In some embodiments, a heterologous gene effector is from a gene product that is a nuclear receptor, for example, a nuclear hormone receptor. Non-limiting examples of nuclear hormone receptors include those encoded by NR0B1, NR0B2, NR1A1, NR1A2, NR1B1, NR1B2, NR1B3, NR1C1, NR1C2, NR1C3, NR1D1, NR1D2, NR1F1, NR1F2, NR1F3, NR1H4, NR1H5, NR1H3, NR1H2, NR1I1, NR1I2, NR1I3, NR2A1, NR2A2, NR2B1, NR2B2, NR2B3, NR2C1, NR2C2, NR2E1, NR2E3, NR2F1, NR2F2, NR2F6, NR3A1, NR3A2, NR3B1, NR3B2, NR3B3, NR3C4, NR3C1, NR3C2, NR3C3, NR4A1, NR4A2, NR4A3, NR5A1, NR5A2, and NR6A1.

[0130] In some embodiments, a heterologous gene effector is from a gene product that is involved in nucleosome assembly. In some embodiments, a heterologous gene effector is from a gene product that is involved in DNA metabolism. In some embodiments, a heterologous gene effector is from a gene product that is involved in nucleotide metabolism. In some embodiments, a heterologous gene effector is from a gene product that is involved in ribosome biogenesis. In some embodiments, a heterologous gene effector is from a gene product that is involved in protein folding. In some embodiments, a heterologous gene effector is from a gene product that is involved in translation. In some embodiments, a heterologous gene effector is from a gene product that is involved in signaling. In some embodiments, a heterologous gene effector is from a gene product that is involved in proteolysis. In some embodiments, a heterologous gene effector is from a gene product that is involved in negative regulation of endopeptidase activity.

[0131] In some embodiments, a heterologous gene effector or gene regulator, as used interchangeably herein, can comprise a polypeptide sequence that exhibits at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or substantially about 100% sequence identity to any of the heterologous gene effector amino acid sequences provided in **Table 3**.

Table 3. Heterologous gene effector amino acid sequences

SEQ ID NO:	Heterologous gene effector amino acid sequences
SEQ ID NO: 15	NNSQGRVTFEDVTNFTQGEWQRLNPEQRNLYRDVMLENYSNLVSVGGQGETTKPDV ILRLEQGKEPWLEEEVLGSGRAEKNNGDI
SEQ ID NO: 16	SGHPGSWEMNSVAFEDVAVNFTQEEWALLDPSQKNLYRDVMQETFRNLASIGNKGE DQSIEDQYKNSSRNLRHIISHSGNNPYGC
SEQ ID NO: 17	AAATLRTPTQGTVTTFEDVAVHFSWEEWGLLDEAQRCLYRDVMLENLALLTSLDVHHQ KQHLGEKHFRSNVGRALFVKCTCFHVSG
SEQ ID NO: 18	TTFKEAMTFKDVAVVFTEEEGLLLDAQRKLYRDVMLENFRNLLSVGHQAFHRDTFHF LREEKIWMMKTAIQREGNSGDKIQTEM
SEQ ID NO: 19	VPAETSSSGLLEEQKMMKSQGLVSFKDVAVDFTQEEWQQLDPSQRTLYRDVMLENYS HLVSMGYPVSKPDVISKLEQGEEPWIIK
SEQ ID NO: 20	MKSQGLVSFKDVAVDFTQEEWQQLDPSQRTLYRDVMLENYSHLVSMGYPVSKPDVIS KLEQGEEPWIIKGDINWIYPDEYQADG
SEQ ID NO: 21	AEGSVMFSDVSIDFSQEEWDCLDPVQRDLYRDVMLENYGNLVSMGLYTPKPQVISLLE QGKEPWMVGRELTRGLCSDLESMCETK
SEQ ID NO: 22	AAALRDPAQVPVAADLLTDHEEGYVTFEDVAVYFSQEEWRLLDDAQRLLYRNVMLEN FTLLASLGKVLTPHPSILSWARLFLFL
SEQ ID NO: 23	AAAALRDPAQVPVAADLLTDHEEGYVTFEDVAVYFSQEEWRLLDDAQRLLYRNVMLE ENFTLLASLGLASSKTHEITQLESWEEP
SEQ ID NO: 24	AAAALRDPAQVPVAADLLTDHEEGYVTFEDVAVYFSQEEWRLLDDAQRLLYRNVMLE ENFTLLASLGCWHGAEAEAEPEQIASVG
SEQ ID NO: 25	AAAALRDPAQGYVTFEDVAVYFSQEEWRLLDDAQRLLYRNVMLENFTLLASLGCWH GAEAEAEPEQIASVGLSSNIQQHQKQH
SEQ ID NO: 26	AAAALRDPAQVPVAADLLTDHEEGYVTFEDVAVYFSQEEWRLLDDAQRLLYRNVMLE NFTLLASLGVLTTPHPSILSWARLFLF
SEQ ID NO: 27	YVTFEDVAVYFSQEEWRLLDDAQRLLYRNVMLENFTLLASLGLASSKTHEITQLESWEE PFMPAWEVVTSAIPRGSWWVELREV
SEQ ID NO: 28	AAAALRDPAQGYVTFEDVAVYFSQEEWRLLDDAQRLLYRNVMLENFTLLASLGLASSK THEITQLESWEEPMPAWEVVTSAIPR
SEQ ID NO: 29	AAAALRDPAQGYVTFEDVAVYFSQEEWRLLDDAQRLLYRNVMLENFTLLASLGCWHG AEAEAEPEQIASVGLSSNIQQHQKQHC
SEQ ID NO: 30	AAAALRDPAQGYVTFEDVAVYFSQEEWRLLDDAQRLLYRNVMLENFTLLASLGLASS KTHEITQLESWEEPMPAWEVVTSAIP
SEQ ID NO: 31	AAAALRDPAQVPVAADLLTDHEEGYVTFEDVAVYFSQEEWRLLDDAQRLLYRNVMLE NFTLLASLGLASSKTHEITQLESWEEPF
SEQ ID NO: 32	VTCAHLGRRARLPAAQPSACPGTCFSQEERMAAGYLPRWSQELVTFEDVSMDFSQEE WELLEPAQKNLYREVMLENYRNVSLEA
SEQ ID NO: 33	LVTEDVSMDFSQEEWELLEPAQKNLYREVMLENYRNVSLEALKNQCTDVGIKEGPL SPAQTSQVTSLSWTGYLLFQPVASSH
SEQ ID NO: 34	KNATIVMSVRREQGSSSGEGSLSFEDVAVGFTREEWQFLDQSQKVLYKEVMLENYINL VSIGYRGTKPDSLFLKEQGEPPIAEG
SEQ ID NO: 35	SSGEGSLSFEDVAVGFTREEWQFLDQSQKVLYKEVMLENYINLVSIGYRGTKPDSLFLKE QGEPPIAEGAAHSQICPDADFL
SEQ ID NO: 36	GPLQFRDVAIEFSLEEWHCLDTAQRNLYRNVMLENYSNLVFLGITVSKPDLITCLEQGRK PLTMKRNEMIAKPSVSFLQVHSESQ
SEQ ID NO: 37	GPLQFRDVAIEFSLEEWHCLDTAQRNLYRNVMLENYSNLVFLGITVSKPDLITCLEQGRK PLTMKRNEMIAKPSVMCSHFAQDLW

SEQ ID NO: 38	APPSAPLPAQGP GKARPSRKRRPRALKFVDVAVYFSPEEWGCLRPAQRALYRDVM RETYGHLGALGCAGPKPALISWLERNTD
SEQ ID NO: 39	QTNTKDWTVTPEHVLPEQSLLTFEEVAMYFSQEEWELLDPTQKALYNDVMQENYET VISLALFVLPKPKVISCLEQGEEPWWVQV
SEQ ID NO: 40	AAATLRDPAQGGYVTFEDVAVYFSQEEWRLLDDAQRLLYRNVMLENFTLLASLGLASS KTHEITQLESWEEPFMPAWEVVTSAIL
SEQ ID NO: 41	AAATLRDPAQGGYVTFEDVAVYFSQEEWRLLDDAQRLLYRNVMLENFTLLASLGLASSK THEITQLESWEEPFMPAWEVVTSAILR
SEQ ID NO: 42	DSVAFEDVAVNFTQEEWALLDPSQKNLYREVMQETLRNLTSIGKKWNNQYIEDEHQN PRRNLRRLLIGERLSESKESHQHGEVLTQ

[0132] The heterologous polynucleotide as disclosed herein can comprise one or more guide moieties (e.g., one or more guide nucleic acid molecules) to direct a heterologous gene effector to a target gene (e.g., target endogenous gene) or a target gene regulatory sequence. A guide moiety can confer an ability to recognize and specifically bind to the target gene or the target gene regulatory sequence. The guide moiety can be configured to form a complex with the heterologous polypeptide (e.g., a guide nucleic acid forming a complex with a nuclease, such as a CRISPR/Cas protein), and the complex can be configured to exhibit specific binding to the target polypeptide sequence as disclosed herein, to modify the expression level and/or the epigenetic modification level of the target gene.

[0133] A guide moiety can comprise a guide nucleic acid. A guide moiety can comprise a nuclease and a guide nucleic acid as disclosed herein. A guide moiety can comprise a nuclease or a part thereof, for example, an endonuclease, such as a heterologous endonuclease. The nuclease can be, e.g., a DNA nuclease and/or RNA nuclease, a modified nuclease that is nuclease-deficient or has reduced nuclease activity compared to a wild-type nuclease, a derivative thereof, a variant thereof, or a fragment thereof. In some embodiments, the guide moiety has minimal nuclease activity.

[0134] Any suitable nuclease, fragment or derivative thereof can be used in a guide moiety. Suitable nucleases include, but are not limited to, CRISPR-associated (Cas) proteins or Cas nucleases including type I CRISPR-associated (Cas) polypeptides, type II CRISPR-associated (Cas) polypeptides, type III CRISPR-associated (Cas) polypeptides, type IV CRISPR-associated (Cas) polypeptides, type V CRISPR-associated (Cas) polypeptides, and type VI CRISPR-associated (Cas) polypeptides; zinc finger nucleases (ZFN); transcription activator-like effector nucleases (TALEN); meganucleases; RNA-binding proteins (RBP); CRISPR-associated RNA binding proteins; recombinases; flippases; transposases; Argonaute (Ago) proteins (e.g., prokaryotic Argonaute (pAgo), archaeal

Argonaute (aAgo), and eukaryotic Argonaute (eAgo)); any derivative thereof; any variant thereof and any fragment thereof.

[0135] In some embodiments, the guide moiety comprises a DNA nuclease such as an engineered (e.g., programmable or targetable) DNA nuclease that is nuclease-deficient. In some embodiments, the guide moiety comprises a nuclease-null DNA binding protein derived from a DNA nuclease that does not induce transcriptional activation or repression of a target DNA sequence unless it is present in a complex with one or more heterologous gene effectors of the disclosure. In some embodiments, the guide moiety comprises a nuclease-null DNA binding protein derived from a DNA nuclease that can induce transcriptional activation or repression of a target DNA sequence (e.g., which can be altered or augmented by the presence of a heterologous gene effector of the disclosure).

[0136] In some embodiments, the guide moiety comprises an RNA nuclease such as an engineered (e.g., programmable or targetable) RNA nuclease. In some embodiments, the guide moiety comprises a nuclease-null RNA binding protein derived from an RNA nuclease that does not induce transcriptional activation or repression of a target RNA sequence unless it is present in a complex with one or more heterologous gene effectors of the disclosure. In some embodiments, the guide moiety comprises a nuclease-null RNA binding protein derived from a RNA nuclease that can induce transcriptional activation or repression of a target RNA sequence (e.g., which can be altered or augmented by the presence of a heterologous gene effector of the disclosure).

[0137] In some embodiments, the guide moiety comprises a nucleic acid-guided targeting system. In some embodiments, the guide moiety comprises a DNA-guided targeting system. In some embodiments, the guide moiety comprises an RNA-guided targeting system. A guide moiety can comprise and utilize, for example, a guide nucleic acid sequence that facilitates specific binding of a CRISPR-Cas system (e.g., a nuclease deficient form thereof, such as dCas9) to a target gene (e.g., target endogenous gene) or target gene regulatory sequence. Binding specificity can be determined by use of a guide nucleic acid, such as a single guide RNA (sgRNA) or a part thereof. In some embodiments, the use of different sgRNAs allows the compositions and methods of the disclosure to be used with (e.g., targeted to) different target genes (e.g., target endogenous genes) or target gene regulatory sequences.

[0138] Prokaryotic CRISPR-Cas (Clustered regularly interspaced short palindromic repeats-CRISPR associated) systems, for example, Class II CRISPR-Cas systems such as Cas9 and Cpf1, can be repurposed as a tool for regulation of gene expression, epigenome

editing, and chromatin looping in compositions and methods of the disclosure. Nuclease-deactivated Cas (dCas) proteins complexed with heterologous gene effectors can allow for regulation of expression of target genes (e.g., target endogenous genes) adjacent to a site bound by the dCas.

- [0139] In some embodiments, the guide moiety comprises a CRISPR-associated (Cas) protein or a Cas nuclease that functions in a non-naturally occurring CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)/Cas (CRISPR-associated) system. In bacteria, this system can provide adaptive immunity against foreign DNA.
- [0140] In a wide variety of organisms including diverse mammals, animals, plants, microbes, and yeast, a CRISPR/Cas system (e.g., modified and/or unmodified) can be utilized as a genome engineering tool, or can be modified to direct specific binding of engineered proteins to target loci as disclosed herein. A CRISPR/Cas system can comprise a guide nucleic acid such as a guide RNA (gRNA) complexed with a Cas protein for targeted regulation of gene expression and/or activity or nucleic acid binding. An RNA-guided Cas protein (e.g., a Cas nuclease such as a Cas9 nuclease) can specifically bind a target polynucleotide (e.g., DNA) in a sequence-dependent manner. The Cas protein, if possessing nuclease activity, can cleave the DNA.
- [0141] In some cases, the Cas protein is mutated and/or modified to yield a nuclease deficient protein or a protein with decreased nuclease activity relative to a wild-type Cas protein. A nuclease deficient protein can retain the ability to bind DNA, but may lack or have reduced nucleic acid cleavage activity.
- [0142] In some embodiments, the guide moiety comprises a Cas protein that forms a complex with a guide nucleic acid, such as a guide RNA or a part thereof. In some embodiments, the guide moiety comprises a Cas protein that forms a complex with a single guide nucleic acid, such as a single guide RNA (sgRNA). In some embodiments, the guide moiety comprises a RNA-binding protein (RBP) optionally complexed with a guide nucleic acid, such as a guide RNA (e.g., sgRNA), which is able to form a complex with a Cas protein. In some embodiments, the guide moiety comprises a nuclease-null DNA binding protein derived from a DNA nuclease that can induce transcriptional activation or repression of a target DNA sequence. In some embodiments, the guide moiety comprises a nuclease-null RNA binding protein derived from a RNA.
- [0143] In some embodiments, a guide nucleic acid used in compositions and methods of the disclosure can be, for example, at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, or more nucleotide(s).

- [0144] In some embodiments, a guide nucleic acid used in compositions and methods of the disclosure is at most at most 10, at most 9, at most 8, at most 7, at most 6, at most 5, at most 4, at most 3, at most 2, or at most 1 nucleotide(s).
- [0145] A guide nucleic acid can be a guide RNA or a part thereof.
- [0146] Any suitable CRISPR/Cas system can be used. A CRISPR/Cas system can be referred to using a variety of naming systems. A CRISPR/Cas system can be a type I, a type II, a type III, a type IV, a type V, a type VI system, or any other suitable CRISPR/Cas system. A CRISPR/Cas system as used herein can be a Class 1, Class 2, or any other suitably classified CRISPR/Cas system. Class 1 or Class 2 determination can be based upon the genes encoding the effector module. Class 1 systems generally have a multi-subunit crRNA-effector complex, whereas Class 2 systems generally have a single protein, such as Cas9, Cpf1, C2c1, C2c2, C2c3 or a crRNA-effector complex. A Class 1 CRISPR/Cas system can use a complex of multiple Cas proteins to effect regulation. A Class 1 CRISPR/Cas system can comprise, for example, type I (e.g., I, IA, IB, IC, ID, IE, IF, IU), type III (e.g., III, IIIA, IIIB, IIIC, IIID), and type IV (e.g., IV, IVA, IVB) CRISPR/Cas type. A Class 2 CRISPR/Cas system can use a single large Cas protein to effect regulation. A Class 2 CRISPR/Cas systems can comprise, for example, type II (e.g., II, IIA, IIB) and type V CRISPR/Cas type. CRISPR systems can be complementary to each other, and/or can lend functional units in trans to facilitate CRISPR locus targeting.
- [0147] When a guide moiety comprises a Cas protein or derivative thereof, the Cas protein or derivative thereof can be a Class 1 or a Class 2 Cas protein. A Cas protein can be a type I, type II, type III, type IV, type V Cas protein, or type VI Cas protein. A Cas protein can comprise one or more domains. Non-limiting examples of domains include, guide nucleic acid recognition and/or binding domain, nuclease domains (e.g., DNase or RNase domains, RuvC, HNH), DNA binding domain, RNA binding domain, helicase domains, protein-protein interaction domains, and dimerization domains. A guide nucleic acid recognition and/or binding domain can interact with a guide nucleic acid. A nuclease domain can comprise catalytic activity for nucleic acid cleavage. A nuclease domain can lack catalytic activity to prevent nucleic acid cleavage. A Cas protein can be a chimeric Cas protein or fragment thereof that is fused to other proteins or polypeptides. A Cas protein can be a chimera of various Cas proteins, for example, comprising domains from different Cas proteins.
- [0148] Non-limiting examples of Cas proteins include c2c1, C2c2, c2c3, Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5e (CasD), CasH, Cas6e, Cas6f, Cas7, Cas8a, Cas8a1, Cas8a2,

Cas8b, Cas8c, Cas9 (Csn1 or Csx12), Cas10, Cas10d, Cas10, Cas10d, CasF, CasG, CasH, Cpf1, Csy1, Csy2, Csy3, Cse1 (CasA), Cse2 (CasB), Cse3 (CasE), Cse4 (CasC), Cse1, Cse2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, Cul966, Cas13a, Cas13b, Cas13c, Cas13d, Cas13X, Cas13Y, and homologs or modified versions thereof.

[0149] In some cases, the Cas protein as disclosed herein may not and need not be Cas9 or Cas12a. The Cas protein as disclosed herein can have a smaller size as compared to Cas9 or Cas12a. The Cas protein as disclosed herein can be derived from Un1Cas12f1 (or Cas14a1). For example, the Cas protein as disclosed herein can comprise an amino acid sequence that is at least about 50%, at least about 60%, at least about 70%, at least about 75% at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or substantially about 100% identical to the polypeptide sequence of SEQ ID NO. 43. In another example, the Cas protein as disclosed herein can comprise an amino acid sequence that is at least about 50%, at least about 60%, at least about 70%, at least about 75% at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or substantially about 100% identical to the polypeptide sequence of SEQ ID NO. 44. As disclosed herein, SEQ ID NO: 43 encodes the polypeptide sequence of Un1Cas12f1 (or Cas14a1). As disclosed herein, SEQ ID NO: 44 encodes an engineered variant of Un1Cas12f1 with reduced nuclease activity.

[0150] SEQ ID NO: 43 (Un1Cas12f1)

```

1 MAKNTITKTL KLRIVRPYNS AEVEKIVADE KNNREKIALE KNKDKVKEAC
51 SKHLKVAAYC TTQVERNACL FCKARKLDDK FYQKLRGQFP DAVFWQEISE
101 IFRQLQKQAA EIYNQSLIEL YYEIFIKGKG IANASSVEHY LSDVCYTRAA
151 ELFKNAAIAS GLRSKIKSNF RLKELKNMKS GLPTTKSDNF PIPLVKQKGG
201 QYTGFEISNH NSDFIIKIPF GRWQVKKEID KYRPWEKFDF EQVQKSPKPI
251 SLLLSTQRRK RNKGWSKDEG TEAEIKKVMN GDYQTSYIEV KRGSKIGEKS
301 AWMLNLSIDV PKIDKGVDPS IIGGIDVGVK SPLVCAINNA FSRYSISDND
351 LFHFNKKMFA RRRILLKKNR HKRAGHGAKN KLKPITILTE KSERFRKKLI
401 ERWACEIADF FIKNKVGTQV MENLESMKRK EDSYFNIRLR GFWPYAEMQN
451 KIEFCLKQYG IEIRKVAPNN TSKTCSKCGH LNNYFNFEYR KKNKFPHPFKC
501 EKC�FKENAD YNAALNISNP KLKSTKEEP

```

[0151] SEQ ID NO: 44 (deactivated nuclease variant of Un1Cas12f1)

```

1 MAKNTITKTL KLRIVRPYNS AEVEKIVADE KNNREKIALE KNKDKVKEAC
51 SKHLKVAAYC TTQVERNACL FCKARKLDDK FYQKLRGQFP DAVFWQEISE
101 IFRQLQKQAA EIYNQSLIEL YYEIFIKGKG IANASSVEHY LSRVCYRRRAA
151 ELFKNAAIAS GLRSKIKSNF RLKELKNMKS GLPTTKSDNF PIPLVKQKGG

```

201 QYTGFEISNH NSDFI IKIPF GRWQVKKEID KYRPWEKFDF EQVQKSPKPI
 251 SLLLSTQRRK RNKGWSKDEG TEAEIKKVMN GDYQTSYIEV KRGSKICEKS
 301 AWMLNLSIDV PKIDKGVDP S IIGGIAVGVR SPLVCAINNA FSRYSISDND
 351 LFHFNKKMFA RRRILLKKNR HKRAGHGAKN KLKPITILTE KSERFRKKLI
 401 ERWACEIADF FIKNKVGT VQ MENLESMKRK EDSYFNIRLR GFWPYAEMQN
 451 KIEFKLKQYG IEIRKVAPNN TSKTCSKCGH LNNYFNFEYR KKNKFPHFKC
 501 EKC�FKENAA YNAALNISNP KLKSTKERP

[0152] A Cas protein or fragment or derivative thereof can be from any suitable organism.

Non-limiting examples include *Streptococcus pyogenes*, *Streptococcus thermophilus*, *Streptococcus sp.*, *Staphylococcus aureus*, *Nocardiosis dassonvillei*, *Streptomyces pristinae spiralis*, *Streptomyces viridochromogenes*, *Streptomyces viridochromogenes*, *Streptosporangium roseum*, *Streptosporangium roseum*, *Alicyclobacillus acidocaldarius*, *Bacillus pseudomycolides*, *Bacillus selenitireducens*, *Exiguobacterium sibiricum*, *Lactobacillus delbrueckii*, *Lactobacillus salivarius*, *Microscilla marina*, *Burkholderiales bacterium*, *Polaromonas naphthalenivorans*, *Polaromonas sp.*, *Crocospaera watsonii*, *Cyanothece sp.*, *Microcystis aeruginosa*, *Pseudomonas aeruginosa*, *Synechococcus sp.*, *Acetohalobium arabaticum*, *Ammonifex degensii*, *Caldicelulosiruptor beccii*, *Candidatus Desulforudis*, *Clostridium botulinum*, *Clostridium difficile*, *Finegoldia magna*, *Natronaerobius thermophilus*, *Pelotomaculum thermopropionicum*, *Acidithiobacillus caldus*, *Acidithiobacillus ferrooxidans*, *Allochrochromatium vinosum*, *Marinobacter sp.*, *Nitrosococcus halophilus*, *Nitrosococcus watsoni*, *Pseudoalteromonas haloplanktis*, *Ktedonobacter racemifer*, *Methanohalobium evestigatum*, *Anabaena variabilis*, *Nodularia spumigena*, *Nostoc sp.*, *Arthrospira maxima*, *Arthrospira platensis*, *Arthrospira sp.*, *Lyngbya sp.*, *Microcoleus chthonoplastes*, *Oscillatoria sp.*, *Petrotoga mobilis*, *Thermosiphon africanus*, *Acaryochloris marina*, *Leptotrichia shahii*, and *Francisella novicida*. In some aspects, the organism is *Streptococcus pyogenes* (*S. pyogenes*). In some aspects, the organism is *Staphylococcus aureus* (*S. aureus*). In some aspects, the organism is *Streptococcus thermophilus* (*S. thermophilus*).

[0153] A Cas protein can be derived from a variety of bacterial species including, but not

limited to, *Veillonella atypical*, *Fusobacterium nucleatum*, *Filifactor alocis*, *Solobacterium moorei*, *Coprococcus catus*, *Treponema denticola*, *Peptoniphilus duerdenii*, *Catenibacterium mitsuokai*, *Streptococcus mutans*, *Listeria innocua*, *Staphylococcus pseudintermedius*, *Acidaminococcus intestine*, *Olsenella uli*, *Oenococcus kitaharae*, *Bifidobacterium bifidum*, *Lactobacillus rhamnosus*, *Lactobacillus gasseri*, *Finegoldia magna*, *Mycoplasma mobile*, *Mycoplasma gallisepticum*, *Mycoplasma ovipneumoniae*, *Mycoplasma canis*, *Mycoplasma synoviae*, *Eubacterium rectale*,

Streptococcus thermophilus, *Eubacterium dolichum*, *Lactobacillus coryniformis* subsp. *Torquens*, *Ilyobacter polytropus*, *Ruminococcus albus*, *Akkermansia muciniphila*, *Acidothermus cellulolyticus*, *Bifidobacterium longum*, *Bifidobacterium dentium*, *Corynebacterium diphtheria*, *Elusimicrobium minutum*, *Nitratifractor salsuginis*, *Sphaerochaeta globus*, *Fibrobacter succinogenes* subsp. *Succinogenes*, *Bacteroides fragilis*, *Capnocytophaga ochracea*, *Rhodopseudomonas palustris*, *Prevotella micans*, *Prevotella ruminicola*, *Flavobacterium columnare*, *Aminomonas paucivorans*, *Rhodospirillum rubrum*, *Candidatus Puniceispirillum marinum*, *Verminephrobacter eiseniae*, *Ralstonia syzygii*, *Dinoroseobacter shibae*, *Azospirillum*, *Nitrobacter hamburgensis*, *Bradyrhizobium*, *Wolinella succinogenes*, *Campylobacter jejuni* subsp. *Jejuni*, *Helicobacter mustelae*, *Bacillus cereus*, *Acidovorax ebreus*, *Clostridium perfringens*, *Parvibaculum lavamentivorans*, *Roseburia intestinalis*, *Neisseria meningitidis*, *Pasteurella multocida* subsp. *Multocida*, *Sutterella wadsworthensis*, *proteobacterium*, *Legionella pneumophila*, *Parasutterella excrementihominis*, *Wolinella succinogenes*, and *Francisella novicida*.

[0154] A Cas protein as used herein can be a wildtype or a modified form of a Cas protein. A Cas protein can be an active variant, inactive variant, or fragment of a wild type or modified Cas protein. A Cas protein can comprise an amino acid change such as a deletion, insertion, substitution, variant, mutation, fusion, chimera, or any combination thereof relative to a wild-type version of the Cas protein. A Cas protein can be a polypeptide with at least about 5%, at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% sequence identity or sequence similarity to a wild type Cas protein. A Cas protein can be a polypeptide with at most about 5%, at most about 10%, at most about 20%, at most about 30%, at most about 40%, at most about 50%, at most about 60%, at most about 70%, at most about 80%, at most about 90%, or at most about 100% sequence identity and/or sequence similarity to a wild type exemplary Cas protein. Variants or fragments can comprise at least about 5%, at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% sequence

identity or sequence similarity to a wild type or modified Cas protein or a portion thereof. Variants or fragments can be targeted to a nucleic acid locus in complex with a guide nucleic acid while lacking nucleic acid cleavage activity.

[0155] A Cas protein can comprise one or more nuclease domains, such as DNase domains. For example, a Cas9 protein can comprise a RuvC-like nuclease domain and/or an HNH-like 20 nuclease domain. The in a nuclease active form of Cas9, RuvC and HNH domains can each cut a different strand of double-stranded DNA to make a double-stranded break in the DNA. A Cas protein can comprise only one nuclease domain (e.g., Cpf1 comprises RuvC domain but lacks HNH domain). In some embodiments, nuclease domains are absent. In some embodiments, nuclease domains are present but inactive or have reduced or minimal activity. In some embodiments, nuclease domains are present and active.

[0156] One or a plurality of the nuclease domains (e.g., RuvC, HNH) of a Cas protein can be deleted or mutated so that they are no longer functional or comprise reduced nuclease activity. For example, in a Cas protein comprising at least two nuclease domains (e.g., Cas9), if one of the nuclease domains is deleted or mutated, the resulting Cas protein, known as a nickase, can generate a single-strand break at a CRISPR RNA (crRNA) recognition sequence within a double-stranded DNA but not a double-strand break. Such a nickase can cleave the complementary strand or the non-complementary strand, but may not cleave both. If all of the nuclease domains of a Cas protein (e.g., both RuvC and HNH nuclease domains in a Cas9 protein; RuvC nuclease domain in a Cpf1 protein) are deleted or mutated, the resulting Cas protein can have a reduced or no ability to cleave both strands of a double-stranded DNA. An example of a mutation that can convert a Cas9 protein into a nickase is a D10A (aspartate to alanine at position 10 of Cas9) mutation in the RuvC domain of Cas9 from *S. pyogenes*. H939A (histidine to alanine at amino acid position 839) or H840A (histidine to alanine at amino acid position 840) in the HNH domain of Cas9 from *S. pyogenes* can convert the Cas9 into a nickase. An example of a mutation that can convert a Cas9 protein into a dead Cas9 is a D10A (aspartate to alanine at position 10 of Cas9) mutation in the RuvC domain and H939A (histidine to alanine at amino acid position 839) or H840A (histidine to alanine at amino acid position 840) in the HNH domain of Cas9 from *S. pyogenes*.

[0157] A nuclease dead Cas protein (e.g., one derived from any Cas protein, such as Un1Cas12f1) can comprise one or more mutations relative to a wild-type version of the protein. The mutation can result in no more than 90%, no more than 80%, no more than 70%, no more than 60%, no more than 50%, no more than 40%, no more than 30%, no

more than 20%, no more than 10%, no more than 5%, or no more than 1% of the nucleic acid-cleaving activity in one or more of the plurality of nucleic acid-cleaving domains of the wild-type Cas protein. The mutation can result in one or more of the plurality of nucleic acid-cleaving domains retaining the ability to cleave the complementary strand of the target nucleic acid but reducing its ability to cleave the non-complementary strand of the target nucleic acid. The mutation can result in one or more of the plurality of nucleic acid-cleaving domains retaining the ability to cleave the non-complementary strand of the target nucleic acid but reducing its ability to cleave the complementary strand of the target nucleic acid. The mutation can result in one or more of the plurality of nucleic acid-cleaving domains lacking the ability to cleave the complementary strand and the non-complementary strand of the target nucleic acid. The residues to be mutated in a nuclease domain can correspond to one or more catalytic residues of the nuclease. For example, residues in the wild type exemplary *S. pyogenes* Cas9 polypeptide such as Asp10, His840, Asn854 and Asn856 can be mutated to inactivate one or more of the plurality of nucleic acid-cleaving domains (e.g., nuclease domains). The residues to be mutated in a nuclease domain of a Cas protein can correspond to residues Asp10, His840, Asn854 and Asn856 in the wild type *S. pyogenes* Cas9 polypeptide, for example, as determined by sequence and/or structural alignment.

[0158] As non-limiting examples, residues D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, and/or A987 (or the corresponding mutations of any of the Cas proteins) can be mutated. For example, e.g., D 10A, G12A, G17A, E762A, H840A, N854A, N863A, H982A, H983A, A984A, and/or D986A. Mutations other than alanine substitutions can be suitable.

[0159] A D10A mutation can be combined with one or more of H840A, N854A, or N856A mutations to produce a Cas9 protein substantially lacking DNA cleavage activity (e.g., a dead Cas9 protein). A H840A mutation can be combined with one or more of D10A, N854A, or N856A mutations to produce a site-directed polypeptide substantially lacking DNA cleavage activity. An N854A mutation can be combined with one or more of H840A, D10A, or N856A mutations to produce a site-directed polypeptide substantially lacking DNA cleavage activity. A N856A mutation can be combined with one or more of H840A, N854A, or D10A mutations to produce a site-directed polypeptide substantially lacking DNA cleavage activity.

[0160] In some embodiments, a Cas protein is a Class 2 Cas protein. In some embodiments, a Cas protein is a type II Cas protein. In some embodiments, the Cas protein is a Cas9

protein, a modified version of a Cas9 protein, or derived from a Cas9 protein. For example, a Cas9 protein lacking cleavage activity. In some embodiments, the Cas9 protein is a Cas9 protein from *S. pyogenes* (e.g., SwissProt accession number Q99ZW2). In some embodiments, the Cas9 protein is a Cas9 from *S. aureus* (e.g., SwissProt accession number J7RUA5). In some embodiments, the Cas9 protein is a modified version of a Cas9 protein from *S. pyogenes* or *S. Aureus*. In some embodiments, the Cas9 protein is derived from a Cas9 protein from *S. pyogenes* or *S. Aureus*. For example, a *S. pyogenes* or *S. Aureus* Cas9 protein lacking cleavage activity.

[0161] In some embodiments, Cas9 can generally refer to a polypeptide with at least about 5%, at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, or about 100% sequence identity and/or sequence similarity to a wild type exemplary Cas9 polypeptide (e.g., Cas9 from *S. pyogenes*). In some embodiments, Cas9 can refer to a polypeptide with at most about 5%, at most about 10%, at most about 20%, at most about 30%, at most about 40%, at most about 50%, at most about 60%, at most about 70%, at most about 80%, at most about 90%, or about 100% sequence identity and/or sequence similarity to a wild type Cas9 polypeptide (e.g., from *S. pyogenes*). Cas9 can refer to the wildtype or a modified form of the Cas9 protein that can comprise an amino acid change such as a deletion, insertion, substitution, variant, mutation, fusion, chimera, or any combination thereof.

[0162] A Cas protein can comprise an amino acid sequence having at least about 5%, at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% sequence identity or sequence similarity to a nuclease domain (e.g., RuvC domain, HNH domain) of a wild-type Cas protein.

[0163] A Cas protein, variant or derivative thereof can be modified to enhance regulation of gene expression by compositions and methods of the disclosure, e.g., as part of a complex disclosed herein. A Cas protein can be modified to increase or decrease nucleic acid binding affinity, nucleic acid binding specificity, enzymatic activity, and/or binding to other factors, such as heterodimerization or oligomerization domains and induce ligands. Cas proteins can also be modified to change any other activity or property of the protein, such as stability. For example, one or more nuclease domains of the Cas protein can be

modified, deleted, or inactivated, or a Cas protein can be truncated to remove domains that are not essential for the desired function of the protein or complex. A Cas protein can be modified to modulate (e.g., enhance or reduce) the activity of the Cas protein for regulating gene expression by a complex of the disclosure that comprises a heterologous gene effector.

[0164] For example, a Cas protein can be coupled (e.g., fused, covalently coupled, or non-covalently coupled) to a heterologous gene effector (e.g., an epigenetic modification domain, a transcriptional activation domain, and/or a transcriptional repressor domain). A Cas protein can be coupled (e.g., fused, covalently coupled, or non-covalently coupled) to an oligomerization or dimerization domain as disclosed herein (e.g., a heterodimerization domain). A Cas protein can be coupled (e.g., fused, covalently coupled, or non-covalently coupled) to a heterologous polypeptide that provides increased or decreased stability. A Cas protein can be coupled (e.g., fused, covalently coupled, or non-covalently coupled) to a sequence that can facilitate degradation of the Cas protein or a complex containing the Cas protein, for example, a degron, such as an inducible degron (e.g., auxin inducible).

[0165] A Cas protein can be coupled (e.g., fused, covalently coupled, or non-covalently coupled) to any suitable number of partners, for example, at least one, at least two, at least three, at least four, or at least five, at least six, at least seven, or at least 8 partners. In some embodiments, a Cas protein of the disclosure is coupled (e.g., fused, covalently coupled, or non-covalently coupled) to at most two, at most three, at most four, at most five, at most six, at most seven, at most eight, or at most ten partners. In some embodiments, a Cas protein of the disclosure is coupled (e.g., fused, covalently coupled, or non-covalently coupled) to 1 – 5, 1 – 4, 1 – 3, 1 – 2, 2 – 5, 2 – 4, 2 – 3, 3 – 5, 3 – 4, or 4 – 5 partners. In some embodiments, a Cas protein of the disclosure is coupled (e.g., fused, covalently coupled, or non-covalently coupled) to one partner. In some embodiments, a Cas protein of the disclosure is coupled (e.g., fused, covalently coupled, or non-covalently coupled) to two partners. In some embodiments, a Cas protein of the disclosure is coupled (e.g., fused, covalently coupled, or non-covalently coupled) to three partners. In some embodiments, a Cas protein of the disclosure is coupled (e.g., fused, covalently coupled, or non-covalently coupled) to four partners. In some embodiments, a Cas protein of the disclosure is coupled (e.g., fused, covalently coupled, or non-covalently coupled) to five partners. In some embodiments, a Cas protein of the disclosure is coupled (e.g., fused, covalently coupled, or non-covalently coupled) to six partners.

[0166] A Cas protein can be a fusion protein. The fused domain or heterologous polypeptide

can be located at the N-terminus, the C-terminus, or internally within the Cas protein.

- [0167] A Cas protein can be provided in any form. For example, a Cas protein can be provided in the form of a protein, such as a Cas protein alone or complexed with a guide nucleic acid as a ribonucleoprotein. A Cas protein can be provided in a complex, for example, complexed with a guide nucleic acid and/or one or more heterologous gene effectors of the disclosure. A Cas protein can be provided in the form of a nucleic acid encoding the Cas protein, such as an RNA (e.g., messenger RNA (mRNA)), or DNA. The nucleic acid encoding the Cas protein can be codon optimized for efficient translation into protein in a particular cell or organism.
- [0168] Nucleic acids encoding Cas proteins, fragments, or derivatives thereof can be stably integrated in the genome of a cell. Nucleic acids encoding Cas proteins can be operably linked to a promoter, for example, a promoter that is constitutively or inducibly active in the cell. Nucleic acids encoding Cas proteins can be operably linked to a promoter in an expression construct. Expression constructs can include any nucleic acid constructs capable of directing expression of a gene or other nucleic acid sequence of interest (e.g., a Cas gene) and which can transfer such a nucleic acid sequence of interest to a target cell.
- [0169] In some embodiments, a Cas protein, variant or derivative thereof is a nuclease dead Cas (dCas) protein. A dead Cas protein can be a protein that lacks nucleic acid cleavage activity.
- [0170] A Cas protein can comprise a modified form of a wild type Cas protein. The modified form of the wild type Cas protein can comprise an amino acid change (e.g., deletion, insertion, or substitution) that reduces the nucleic acid-cleaving activity of the Cas protein. For example, the modified form of the Cas protein can have no more than 90%, no more than 80%, no more than 70%, no more than 60%, no more than 50%, no more than 40%, no more than 30%, no more than 20%, no more than 10%, no more than 5%, or no more than 1% of the nucleic acid-cleaving activity of the wild-type Cas protein (e.g., Cas9 from *S. pyogenes*). The modified form of Cas protein can have no substantial nucleic acid-cleaving activity. When a Cas protein is a modified form that has no substantial nucleic acid-cleaving activity, it can be referred to as enzymatically inactive, “deactivated” and/or “dead” (abbreviated by “d”). A dead Cas protein (e.g., dCas, dCas9) can bind to a target polynucleotide but may not cleave or minimally cleaves the target polynucleotide. In some aspects, a dead Cas protein is a dead Cas9 protein.
- [0171] A dCas9 polypeptide can associate with a single guide RNA (sgRNA) to activate or repress transcription of a target gene (e.g., target endogenous gene), for example, in

combination with heterologous gene effector(s) disclosed herein. sgRNAs can be introduced into cells expressing the Cas or guide moiety component of the disclosure. In some cases, such cells can contain one or more different sgRNAs that target the same target gene (e.g., target endogenous gene) or target gene regulatory sequence. In other cases, the sgRNAs target different nucleic acids in the cell (e.g., different target genes, different target gene regulatory sequences, or different sequences within the same target gene or target gene regulatory sequence).

[0172] Enzymatically inactive can refer to a nuclease that can bind to a nucleic acid sequence in a polynucleotide in a sequence-specific manner, but will not cleave a target polynucleotide or will cleave it at a substantially reduced frequency. An enzymatically inactive guide moiety can comprise an enzymatically inactive domain (e.g. nuclease domain). Enzymatically inactive can refer to no activity. Enzymatically inactive can refer to substantially no activity. Enzymatically inactive can refer to essentially no activity. Enzymatically inactive can refer to an activity no more than 1%, no more than 2%, no more than 3%, no more than 4%, no more than 5%, no more than 6%, no more than 7%, no more than 8%, no more than 9%, or no more than 10% activity compared to a comparable wild-type activity (e.g., nucleic acid cleaving activity, wild-type Cas9 activity).

[0173] In some embodiments, the guide moiety does not contain a nucleic acid-guided targeting system. For example, guide moieties can include proteins that bind to a target gene (e.g., target endogenous gene) or target gene regulatory sequence based on protein structural features, such as certain nucleases disclosed herein.

[0174] In some embodiments, a guide moiety comprises a zinc finger nuclease (ZFN) or a variant, fragment, or derivative thereof. ZFN can refer to a fusion between a cleavage domain, such as a cleavage domain of FokI, and at least one zinc finger motif (e.g., at least 2, at least 3, at least 4, or at least 5 zinc finger motifs) which can bind polynucleotides such as DNA and RNA. In some embodiments, a ZFN is used in a targeting moiety of the disclosure to bind a polynucleotide (e.g., target gene or target gene regulatory sequence), but the ZFN does not cleave or substantially does not cleave the polynucleotide, e.g., a nuclease dead ZFN. A ZFN or a variant, fragment, or derivative thereof can be fused to or associated with one of more heterologous gene effectors to form a complex of the disclosure.

[0175] The heterodimerization at certain positions in a polynucleotide of two individual ZFNs in certain orientation and spacing can lead to cleavage of the polynucleotide in

nuclease-active ZFN. For example, a ZFN binding to DNA can induce a double-strand break in the DNA. In order to allow two cleavage domains to dimerize and cleave DNA, two individual ZFNs can bind opposite strands of DNA with their C-termini at a certain distance apart. In some cases, linker sequences between the zinc finger domain and the cleavage domain can require the 5' edge of each binding site to be separated by about 5-7 base pairs. In some cases, a cleavage domain is fused to the C-terminus of each zinc finger domain.

[0176] In some embodiments, the cleavage domain of a guide moiety comprising a ZFN comprises a modified form of a wild type cleavage domain. The modified form of the cleavage domain can comprise an amino acid change (e.g., deletion, insertion, or substitution) that reduces the nucleic acid-cleaving activity of the cleavage domain. For example, the modified form of the cleavage domain can have no more than 90%, no more than 80%, no more than 70%, no more than 60%, no more than 50%, no more than 40%, no more than 30%, no more than 20%, no more than 10%, no more than 5%, or no more than 1% of the nucleic acid-cleaving activity of the corresponding wild-type cleavage domain. The modified form of the cleavage domain can have no substantial nucleic acid-cleaving activity. In some embodiments, the cleavage domain is enzymatically inactive.

[0177] In some embodiments, a guide moiety comprises a “TALEN” or “TAL-effector nuclease” or a variant, fragment, or derivative thereof. TALENs refer to engineered transcription activator-like effector nucleases that generally contain a central domain of DNA-binding tandem repeats and a cleavage domain. TALENs can be produced by fusing a TAL effector DNA binding domain to a DNA cleavage domain. In some cases, a DNA-binding tandem repeat comprises 33-35 amino acids in length and contains two hypervariable amino acid residues at positions 12 and 13 that can recognize at least one specific DNA base pair. A transcription activator-like effector (TALE) protein can be fused to a nuclease such as a wild-type or mutated FokI endonuclease or the catalytic domain of FokI. In some embodiments, a TALEN is used in a targeting moiety of the disclosure to bind a polynucleotide (e.g., target gene or target gene regulatory sequence), but the TALEN does not cleave or substantially does not cleave the polynucleotide, e.g., a nuclease dead TALEN. A TALEN or a variant, fragment, or derivative thereof can be fused to or associated with one or more heterologous gene effectors to form a complex of the disclosure.

[0178] In some embodiments, a TALEN is engineered for reduced nuclease activity. In some embodiments, the nuclease domain of a TALEN comprises a modified form of a wild type

nuclease domain. The modified form of the nuclease domain can comprise an amino acid change (e.g., deletion, insertion, or substitution) that reduces the nucleic acid-cleaving activity of the nuclease domain. For example, the modified form of the nuclease domain can have no more than 90%, no more than 80%, no more than 70%, no more than 60%, no more than 50%, no more than 40%, no more than 30%, no more than 20%, no more than 10%, no more than 5%, or no more than 1% of the nucleic acid-cleaving activity of the wild-type nuclease domain. The modified form of the nuclease domain can have no substantial nucleic acid-cleaving activity. In some embodiments, the nuclease domain is enzymatically inactive. A TALEN or a variant, fragment, or derivative thereof can be fused to or associated with one of more heterologous gene effectors to form a complex of the disclosure.

[0179] Several mutations to FokI have been made for its use in TALENs, which, for example, improve cleavage specificity or activity. Such TALENs can be engineered to bind any desired DNA sequence. TALENs can be used to generate gene modifications (e.g., nucleic acid sequence editing) by creating a double-strand break in a target DNA sequence, which in turn, undergoes NHEJ or HDR.

[0180] A TALE or a variant, fragment, or derivative thereof can be fused to or associated with one of more heterologous gene effectors to form a complex of the disclosure. In some embodiments, the transcription activator-like effector (TALE) protein is fused to a heterologous gene effector and does not comprise a nuclease. In some embodiments, a TALEN does not cleave or substantially does not cleave the polynucleotide, e.g., a nuclease dead TALE. A TALE or a variant, fragment, or derivative thereof can be fused to or associated with one of more heterologous gene effectors to form a complex of the disclosure.

[0181] In some embodiments, the complex of the transcription activator-like effector (TALE) protein and the heterologous gene effector is designed to function as a transcriptional activator. In some embodiments, the complex of the transcription activator-like effector (TALE) protein and the heterologous gene effector is designed to function as a transcriptional repressor. For example, the DNA-binding domain of the transcription activator-like effector (TALE) protein can be fused (e.g., linked) to one or more heterologous gene effectors that comprise transcriptional activation domains, or to one or more heterologous gene effectors that comprise transcriptional repression domains.

[0182] In some embodiments, a guide moiety comprises a meganuclease. Meganucleases generally refer to rare-cutting endonucleases or homing endonucleases that can be highly

sequence specific. Meganucleases can recognize DNA target sites ranging from at least 12 base pairs in length, e.g., from 12 to 40 base pairs, 12 to 50 base pairs, or 12 to 60 base pairs in length. Meganucleases can be modular DNA-binding nucleases such as any fusion protein comprising at least one catalytic domain of an endonuclease and at least one DNA binding domain or protein specifying a nucleic acid target sequence. The DNA-binding domain can contain at least one motif that recognizes single- or double-stranded DNA. A nuclease-active meganuclease can generate a double-stranded break. In some embodiments, a meganuclease is used in a targeting moiety of the disclosure to bind a polynucleotide (e.g., target gene or target gene regulatory sequence), but the meganuclease does not cleave or substantially does not cleave the polynucleotide, e.g., a nuclease dead meganuclease. A meganuclease or a variant, fragment, or derivative thereof can be fused to or associated with one of more heterologous gene effectors to form a complex of the disclosure.

[0183] The meganuclease can be monomeric or dimeric. In some embodiments, the meganuclease is naturally-occurring (found in nature) or wild-type, and in other instances, the meganuclease is non-natural, artificial, engineered, synthetic, rationally designed, or man-made. In some embodiments, the meganuclease of the present disclosure includes an I-CreI meganuclease, I-CeuI meganuclease, I-MsoI meganuclease, I-SceI meganuclease, variants thereof, derivatives thereof, and fragments thereof.

[0184] In some embodiments, the nuclease domain of a meganuclease comprises a modified form of a wild type nuclease domain. The modified form of the nuclease domain can comprise an amino acid change (e.g., deletion, insertion, or substitution) that reduces or eliminates the nucleic acid-cleaving activity of the nuclease domain. For example, the modified form of the nuclease domain can have no more than 90%, no more than 80%, no more than 70%, no more than 60%, no more than 50%, no more than 40%, no more than 30%, no more than 20%, no more than 10%, no more than 5%, or no more than 1% of the nucleic acid-cleaving activity of the wild-type nuclease domain. The modified form of the nuclease domain can have no substantial nucleic acid-cleaving activity. In some embodiments, the nuclease domain is enzymatically inactive. In some embodiments, a meganuclease can bind DNA but cannot cleave the DNA. In some embodiments, a nuclease-inactive meganuclease is fused to or associated with one or more heterologous gene effectors to generate a complex of the disclosure.

[0185] In some embodiments, the guide moiety can regulate expression and/or activity of a target gene (e.g., target endogenous gene). In some embodiments, the guide moiety can

edit the sequence of a nucleic acid (e.g., a gene and/or gene product). A nuclease-active Cas protein can edit a nucleic acid sequence by generating a double-stranded break or single-stranded break in a target polynucleotide.

[0186] In some embodiments, a guide moiety comprising a nuclease can generate a double-strand break in a target polynucleotide, such as DNA. A double-strand break in DNA can result in DNA break repair which allows for the introduction of gene modification(s) (e.g., nucleic acid editing). In some embodiments, a nuclease induces site-specific single-strand DNA breaks or nicks, thus resulting in HDR.

[0187] A double-strand break in DNA can result in DNA break repair which allows for the introduction of gene modification(s) (e.g., nucleic acid editing). DNA break repair can occur via non-homologous end joining (NHEJ) or homology-directed repair (HDR). In HDR, a donor DNA repair template or template polynucleotide that contains homology arms flanking sites of the target DNA can be provided.

[0188] In some embodiments, a guide moiety or complex comprising a nuclease does not generate a double-strand break in a target polynucleotide, such as DNA.

[0189] III. COMPLEXES

[0190] Disclosed herein, in some aspects, are one or more complexes that comprise a heterologous polypeptide and a heterologous polynucleotide. In some cases, a complex can comprise a heterologous gene effector and a guide moiety, for example, a guide nucleic acid and/or a nuclease, such as an endonuclease that lacks or substantially lacks cleavage activity.

[0191] Complexes of the disclosure can be useful, for example, for bringing one or more heterologous gene effectors into close proximity with a target gene (e.g., target endogenous gene) or target gene regulatory sequence, thereby facilitating modulation of an expression, epigenetic modification, or activity level of the target gene.

[0192] In some embodiments, a complex of the disclosure binds to DNA, e.g., genomic DNA. In some embodiments, a complex of the disclosure binds to RNA, e.g., mRNA, microRNA, siRNA, or non-coding RNA. In some embodiments, a complex of the disclosure binds to DNA and RNA.

[0193] In some embodiments, a complex can modulate (e.g., increase or decrease) expression and/or activity of a target gene (e.g., target endogenous gene) by physical obstruction of a polynucleotide sequence (e.g., a promoter, enhancer, repressor, operator, or silencer, insulator, cis-regulatory element, trans-regulatory element, epigenetic modification (e.g., DNA methylation) site, coding sequence).

- [0194] In some embodiments, a complex can modulate (e.g., increase or decrease) expression and/or activity of a target gene (e.g., target endogenous gene) by recruitment of additional factors effective to suppress or enhance expression of the target gene.
- [0195] In some embodiments, complexes of the disclosure are used for introducing epigenetic modifications to a target gene (e.g., target endogenous gene) or target gene regulatory sequence (e.g., promoter, enhancer, silencer, insulator, cis-regulatory element, trans-regulatory element, or epigenetic modification (e.g., DNA methylation) site). In some embodiments, complexes of the disclosure are used for producing three-dimensional structures, topologically associating domains, or genomic boundaries comprising a target gene or target gene regulatory sequence (e.g., distal or proximal gene from the target gene).
- [0196] In some embodiments, a complex comprises a heterologous gene effector and a guide moiety. In some embodiments, a complex comprises one heterologous gene effector and one guide moiety. In some embodiments, a complex comprises two heterologous gene effectors and one guide moiety. In some embodiments, a complex comprises three or more heterologous gene effectors and one guide moiety.
- [0197] In some embodiments, a complex comprises a heterologous gene effector and a guide nucleic acid. In some embodiments, a complex comprises one heterologous gene effector and one guide nucleic acid. In some embodiments, a complex comprises two heterologous gene effectors and one guide nucleic acid. In some embodiments, a complex comprises three or more heterologous gene effectors and one guide nucleic acid.
- [0198] Two components present in a complex can be covalently linked, for example, present in a fusion protein, or cross-linked, e.g., treated with a crosslinking agent, or joined by a peptide or non-peptide linker as disclosed herein.
- [0199] In some embodiments, two components present in a complex are part of the same fusion protein. Components can optionally be joined by a linker, such as a peptide linker or a non-peptide linker.
- [0200] In some embodiments, a guide moiety or a part thereof (e.g., nuclease, such as dCas9) is joined to a heterologous gene effector by a linker. In some embodiments the guide moiety or part thereof is further joined to a second heterologous gene effector by a second linker that is the same or different. In some embodiments, a guide moiety or a part thereof (e.g., nuclease, such as dCas9) is fused to a heterologous gene effector without a linker.
- [0201] In some embodiments, a guide moiety or a part thereof (e.g., nuclease, such as dCas9) is joined to an oligomerization domain or dimerization (e.g., heterodimerization) domain

by a linker. In some embodiments the guide moiety or part thereof is further joined to a second oligomerization domain or dimerization (e.g., heterodimerization) domain by a second linker that is the same or different. In some embodiments, a guide moiety or a part thereof (e.g., nuclease, such as dCas9) is fused to a second oligomerization domain or dimerization (e.g., heterodimerization) domain without a linker.

[0202] In some embodiments, heterologous gene effector is joined to a second heterologous gene effector by a linker. In some embodiments the heterologous gene effector is further joined to a third heterologous gene effector by a second linker that is the same or different. In some embodiments, a heterologous gene effector is fused to a second heterologous gene effector without a linker.

[0203] In some embodiments, heterologous gene effector is joined to an oligomerization domain or dimerization (e.g., heterodimerization) domain by a linker. In some embodiments the heterologous gene effector is further joined to a second oligomerization domain or dimerization (e.g., heterodimerization) domain by a second linker that is the same or different. In some embodiments, a heterologous gene effector is fused to a second oligomerization domain or dimerization (e.g., heterodimerization) domain without a linker.

[0204] Any suitable linker can be used. A flexible linker can have a sequence containing stretches of glycine and serine residues. The small size of the glycine and serine residues provides flexibility, and allows for mobility of the connected functional domains. The incorporation of serine or threonine can maintain the stability of the linker in aqueous solutions by forming hydrogen bonds with the water molecules, thereby reducing unfavorable interactions between the linker and protein moieties. Flexible linkers can also contain additional amino acids such as threonine and alanine to maintain flexibility, as well as polar amino acids such as lysine and glutamine to improve solubility. A rigid linker can have, for example, an alpha helix-structure. An alpha-helical rigid linker can act as a spacer between protein domains.

[0205] A linker sequence can be, for example, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 amino acid residues in length.

[0206] In some embodiments, a linker is at least 1, at least 2, at least 3, at least 5, at least 7, at least 9, at least 11, at least 13, at least 15, or at least 20 amino acids. In some embodiments, a linker is at most 5, at most 7, at most 9, at most 11, at most 13, at most 15, at most 20, at most 25, at most 30, at most 40, or at most 50 amino acids.

[0207] In some embodiments, non-peptide linkers are used. A non-peptide linker can be, for example a chemical linker. Two parts of a complex of the disclosure can be connected by a chemical linker. Each chemical linker of the disclosure can be alkylene, alkenylene, alkynylene, heteroalkylene, cycloalkylene, heterocycloalkylene, arylene, or heteroarylene, any of which is optionally substituted. In some embodiments, a chemical linker of the disclosure can be an ester, ether, amide, thioether, or polyethyleneglycol (PEG). In some embodiments, a linker can reverse the order of the amino acids sequence in a compound, for example, so that the amino acid sequences linked by the linker are head-to-head, rather than head-to-tail. Non-limiting examples of such linkers include diesters of dicarboxylic acids, such as oxalyl diester, malonyl diester, succinyl diester, glutaryl diester, adipyl diester, pimetyl diester, fumaryl diester, maleyl diester, phthalyl diester, isophthalyl diester, and terephthalyl diester. Non-limiting examples of such linkers include diamides of dicarboxylic acids, such as oxalyl diamide, malonyl diamide, succinyl diamide, glutaryl diamide, adipyl diamide, pimetyl diamide, fumaryl diamide, maleyl diamide, phthalyl diamide, isophthalyl diamide, and terephthalyl diamide. Non-limiting examples of such linkers include diamides of diamino linkers, such as ethylene diamine, 1,2-di(methylamino)ethane, 1,3-diaminopropane, 1,3-di(methylamino)propane, 1,4-di(methylamino)butane, 1,5-di(methylamino)pentane, 1,6-di(methylamino)hexane, and piperazine. Non-limiting examples of optional substituents include hydroxyl groups, sulfhydryl groups, halogens, amino groups, nitro groups, nitroso groups, cyano groups, azido groups, sulfoxide groups, sulfone groups, sulfonamide groups, carboxyl groups, carboxaldehyde groups, imine groups, alkyl groups, halo-alkyl groups, alkenyl groups, halo-alkenyl groups, alkynyl groups, halo-alkynyl groups, alkoxy groups, aryl groups, aryloxy groups, aralkyl groups, arylalkoxy groups, heterocyclyl groups, acyl groups, acyloxy groups, carbamate groups, amide groups, ureido groups, epoxy groups, and ester groups.

[0208] Two components present in a complex can be non-covalently coupled, for example, by ionic bonds, hydrogen bonds, interactions mediated by oligomerization or dimerization domains disclosed herein, etc.

[0209] In some embodiments, a guide moiety or a part thereof (e.g., nuclease, such as dCas9) is joined to a heterologous gene effector by non-covalent coupling. In some embodiments the guide moiety or part thereof is further joined to a second heterologous gene effector by non-covalent coupling. In some embodiments the guide moiety or part thereof is joined to a first heterologous gene effector covalently (e.g., as a fusion protein, optionally with a

linker), and the guide moiety or part thereof is further joined to a second heterologous gene effector by non-covalent coupling.

[0210] In some embodiments, a guide moiety or a part thereof (e.g., nuclease, such as dCas9) is joined to an oligomerization domain or dimerization (e.g., heterodimerization) domain by non-covalent coupling. In some embodiments the guide moiety or part thereof is further joined to a second oligomerization domain or dimerization (e.g., heterodimerization) domain by non-covalent coupling. In some embodiments, a guide moiety or a part thereof (e.g., nuclease, such as dCas9) is fused to a first oligomerization domain or dimerization (e.g., heterodimerization) domain by covalent coupling (e.g., fused, optionally by a linker) and is joined to a second oligomerization domain or dimerization (e.g., heterodimerization) domain by non-covalent coupling.

[0211] In some embodiments, a first component of a guide moiety (e.g., a guide nucleic acid) is joined to a second component of the guide moiety (e.g., nuclease) non-covalently. In some embodiments, a first component of a guide moiety (e.g., a guide nucleic acid) is joined to a second component of the guide moiety (e.g., nuclease) covalently.

[0212] Any combination of covalent and non-covalent coupling can be used in a complex of the disclosure, for example, one or more heterologous gene effectors can be fused to a guide moiety non-covalently, and one or more oligomerization domains can be bound to a component of the complex (e.g., nuclease) covalently.

[0213] In some embodiments, a polypeptide providing increased or decreased stability is fused to or otherwise associated with a component of a complex of the disclosure, e.g., a guide moiety or a heterologous gene effector. The fused polypeptide can be located at the N-terminus, the C-terminus, or internally within the fusion protein.

[0214] In some embodiments, one or more components of a complex of the disclosure is fused to a domain that directs desirable sub-cellular localization, for example, a nuclear localization signal or a protein for targeting to the inner nuclear membrane, outer nuclear membrane, Cajal body, nuclear speckle, nuclear pore complex, PML body, nucleolus, P granule, GW body, stress granule, sponge body, endoplasmic reticulum, mitochondria, etc.

[0215] In some embodiments, a complex of the disclosure comprises a first protein linked to a first oligomerization (e.g., dimerization) domain, and a second protein linked to a second oligomerization (e.g., dimerization) domain. In some embodiments, an oligomerization domain or a dimerization domain can comprise a peptide interaction domain, for example, systems utilizing sgRNA2.0, SAM, SunTag, RAB, FLAG-biotin, or

inducible oligomerization (e.g., dimerization) systems disclosed herein.

[0216] DELIVERY

[0217] One or more genes encoding any of the heterologous polypeptide (e.g., the heterologous gene effectors) and any additional molecule operatively coupled thereto (e.g., the heterologous polynucleotide, such as one or more guide nucleic acid molecules), as disclosed herein, can be integrated into a genome of the cell, in which the aberrant expression of a target gene is to be modified. Alternatively, the one or more genes may not and need not be integrated into the genome of the cell.

[0218] Any of the heterologous polypeptide (e.g., the heterologous gene effectors) and any additional molecule operatively coupled thereto (e.g., the heterologous polynucleotide, such as one or more guide nucleic acid molecules) can be introduced (e.g., delivered, expressed, etc.) to a cell by various methods, e.g., viral and non-viral delivery methods. Viral vector delivery systems can include DNA and RNA viruses, which can have either episomal or integrated genomes after delivery to the cell. Non-viral vector delivery systems can include DNA plasmids, RNA (e.g. a transcript of a vector described herein), naked nucleic acid, and nucleic acid complexed with a delivery vehicle, such as a liposome.

[0219] RNA or DNA viral based systems can be used to target specific cells and traffic the viral payload to the nucleus of the cell. Viral vectors can be used to treat cells in vitro, and the modified cells can optionally be administered (ex vivo). Alternatively, viral vectors can be administered directly (in vivo) to the subject. Viral based systems can include retroviral, lentivirus, adenoviral, adeno-associated and herpes simplex virus vectors for gene transfer. Integration in the host genome can occur with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, which can result in long term expression of the inserted transgene.

[0220] Non-limiting examples of viral vectors that can be utilized to deliver the heterologous polypeptide and/or heterologous polynucleotide (or one or more genes encoding thereof) can include, but are not limited to, retroviral vectors, lentiviral vectors, adenovirus vectors, poxvirus vectors, herpesvirus vectors, adeno-associated virus (AAV) vectors. Non-limiting examples of AAV vectors can include AAV1, AAV10, AAV106.1/hu.37, AAV11, AAV114.3/hu.40, AAV12, AAV127.2/hu.41, AAV127.5/hu.42, AAV128.1/hu.43, AAV128.3/hu.44, AAV130.4/hu.48, AAV145.1/hu.53, AAV145.5/hu.54, AAV145.6/hu.55, AAV16.12/hu. 11, AAV16.3, AAV16.8/hu. 10, AAV161.10/hu.60, AAV161.6/hu.61, AAV1-7/rh.48, AAV1-8/rh.49, AAV2, AAV2.5T,

AAV2-15/rh.62, AAV223.1, AAV223.2, AAV223.4, AAV223.5, AAV223.6, AAV223.7, AAV2-3/rh.61, AAV24.1, AAV2-4/rh.50, AAV2-5/rh.51, AAV27.3, AAV29.3/bb.1, AAV29.5/bb.2, AAV2G9, AAV-2-pre-miRNA-101, AAV3, AAV3.1/hu.6, AAV3.1/hu.9, AAV3-11/rh.53, AAV3-3, AAV33.12/hu. 17, AAV33.4/hu. 15, AAV33.8/hu. 16, AAV3-9/rh.52, AAV3a, AAV3b, AAV4, AAV4-19/rh.55, AAV42.12, AAV42-10, AAV42-11, AAV42-12, AAV42-13, AAV42-15, AAV42-1b, AAV42-2, AAV42-3a, AAV42-3b, AAV42-4, AAV42-5a, AAV42-5b, AAV42-6b, AAV42-8, AAV42-aa, AAV43-1, AAV43-12, AAV43-20, AAV43-21, AAV43-23, AAV43-25, AAV43-5, AAV4-4, AAV44.1, AAV44.2, AAV44.5, AAV46.2/hu.28, AAV46.6/hu.29, AAV4-8/r11.64, AAV4-8/rh.64, AAV4-9/rh.54, AAV5, AAV52.1/hu.20, AAV52/hu. 19, AAV5-22/rh.58, AAV5-3/rh.57, AAV54.1/hu.21, AAV54.2/hu.22, AAV54.4R/hu.27, AAV54.5/hu.23, AAV54.7/hu.24, AAV58.2/hu.25, AAV6, AAV6.1, AAV6.1.2, AAV6.2, AAV7, AAV7.2, AAV7.3/hu.7, AAV8, AAV-8b, AAV-8h, AAV9, AAV9.11, AAV9.13, AAV9.16, AAV9.24, AAV9.45, AAV9.47, AAV9.61, AAV9.68, AAV9.84, AAV9.9, AAVA3.3, AAVA3.4, AAVA3.5, AAVA3.7, AAV-b, AAVC1, AAVC2, AAVC5, AAVCh.5, AAVCh.5R1, AAVcy.2, AAVcy.3, AAVcy.4, AAVcy.5, AAVCy.5R1, AAVCy.5R2, AAVCy.5R3, AAVCy.5R4, AAVcy.6, AAV-DJ, AAV-DJ8, AAVF3, AAVF5, AAV-h, AAVH-1/hu.1, AAVH2, AAVH-5/hu.3, AAVH6, AAVhE1.1, AAVhEr1.14, AAVhEr1.16, AAVhEr1.18, AAVhEr1.23, AAVhEr1.35, AAVhEr1.36, AAVhEr1.5, AAVhEr1.7, AAVhEr1.8, AAVhEr2.16, AAVhEr2.29, AAVhEr2.30, AAVhEr2.31, AAVhEr2.36, AAVhEr2.4, AAVhEr3.1, AAVhu.1, AAVhu.10, AAVhu.11, AAVhu.11, AAVhu.12, AAVhu.13, AAVhu.14/9, AAVhu.15, AAVhu.16, AAVhu.17, AAVhu.18, AAVhu.19, AAVhu.2, AAVhu.20, AAVhu.21, AAVhu.22, AAVhu.23.2, AAVhu.24, AAVhu.25, AAVhu.27, AAVhu.28, AAVhu.29, AAVhu.29R, AAVhu.3, AAVhu.31, AAVhu.32, AAVhu.34, AAVhu.35, AAVhu.37, AAVhu.39, AAVhu.4, AAVhu.40, AAVhu.41, AAVhu.42, AAVhu.43, AAVhu.44, AAVhu.44R1, AAVhu.44R2, AAVhu.44R3, AAVhu.45, AAVhu.46, AAVhu.47, AAVhu.48, AAVhu.48R1, AAVhu.48R2, AAVhu.48R3, AAVhu.49, AAVhu.5, AAVhu.51, AAVhu.52, AAVhu.53, AAVhu.54, AAVhu.55, AAVhu.56, AAVhu.57, AAVhu.58, AAVhu.6, AAVhu.60, AAVhu.61, AAVhu.63, AAVhu.64, AAVhu.66, AAVhu.67, AAVhu.7, AAVhu.8, AAVhu.9, AAVhu.t 19, AAVLG-10/rh.40, AAVLG-4/rh.38, AAVLG-9/hu.39, AAVLG-9/hu.39, AAV-LKO1, AAV-LKO2, AAVLK03, AAV-LKO3, AAV-LKO4, AAV-LKO5, AAV-LKO6, AAV-LKO7, AAV-LKO8, AAV-LKO9, AAV-LK10, AAV-LK11, AAV-LK12, AAV-LK13, AAV-LK14, AAV-LK15, AAV-LK17,

AAV-LK18, AAV-LK19, AAVN721-8/rh.43, AAV-PAEC, AAV-PAEC11, AAV-PAEC12, AAV-PAEC2, AAV-PAEC4, AAV-PAEC6, AAV-PAEC7, AAV-PAEC8, AAVpi.1, AAVpi.2, AAVpi.3, AAVrh.10, AAVrh.12, AAVrh.13, AAVrh.13R, AAVrh.14, AAVrh.17, AAVrh.18, AAVrh.19, AAVrh.2, AAVrh.20, AAVrh.21, AAVrh.22, AAVrh.23, AAVrh.24, AAVrh.25, AAVrh.2R, AAVrh.31, AAVrh.32, AAVrh.33, AAVrh.34, AAVrh.35, AAVrh.36, AAVrh.37, AAVrh.37R2, AAVrh.38, AAVrh.39, AAVrh.40, AAVrh.43, AAVrh.44, AAVrh.45, AAVrh.46, AAVrh.47, AAVrh.48, AAVrh.48, AAVrh.48.1, AAVrh.48.1.2, AAVrh.48.2, AAVrh.49, AAVrh.50, AAVrh.51, AAVrh.52, AAVrh.53, AAVrh.54, AAVrh.55, AAVrh.56, AAVrh.57, AAVrh.58, AAVrh.59, AAVrh.60, AAVrh.61, AAVrh.62, AAVrh.64, AAVrh.64R1, AAVrh.64R2, AAVrh.65, AAVrh.67, AAVrh.68, AAVrh.69, AAVrh.70, AAVrh.72, AAVrh.73, AAVrh.74, AAVrh.8, AAVrh.8R, AAVrh8R, AAVrh8R A586R mutant, AAVrh8R R533A mutant, BAAV, BNP61 AAV, BNP62 AAV, BNP63 AAV, bovine AAV, caprine AAV, Japanese AAV 10, true type AAV (ttAAV), UPENN AAV 10, AAV-LK16, AAV, AAV Shuffle 100-1, AAV Shuffle 100-2, AAV Shuffle 100-3, AAV Shuffle 100-7, AAV Shuffle 10-2, AAV Shuffle 10-6, AAV Shuffle 10-8, AAV SM 100-10, AAV SM 100-3, AAV SM 10-1, AAV SM 10-2, and AAV SM 10-8. For example, AAVrh.74 can be used as a viral vector to deliver a polynucleotide sequence encoding the heterologous polypeptide and the heterologous polynucleotide (e.g., Cas protein-gene effector fusion and one or more guide nucleic acid molecules).

[0221] Methods of non-viral delivery of nucleic acids can include lipofection, nucleofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, polycation or lipid:nucleic acid conjugates, lipid nanoparticles (LNPs), naked DNA, artificial virions, and agent-enhanced uptake of DNA. Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides can be used.

[0222] Any of the compositions disclosed herein (or one or more genes encoding any portion of the compositions), such as the heterologous gene effector(s) and/or the guide nucleic acid molecule(s), can be administered by any suitable administration route, including but not limited to, parenteral (e.g., intravenous, intratumoral, subcutaneous, intramuscular, intracerebral, intracerebroventricular, intra-articular, intraperitoneal, or intracranial), intranasal, buccal, sublingual, oral, or rectal administration routes. In some instances, the pharmaceutical composition is formulated for parenteral (e.g., intravenous, intratumoral, subcutaneous, intramuscular, intracerebral, intracerebroventricular, intra-articular, intraperitoneal, or intracranial) administration.

[0223] TARGET GENE(S)

- [0224]** The disclosure provides compositions, methods, and systems for modulating expression of target genes (e.g., target endogenous genes). For example, disclosed herein are complexes that comprise a guide moiety and one or more heterologous gene effectors that can increase or decrease an activity or expression level of a target gene.
- [0225]** In some embodiments, a target gene or regulatory sequence thereof is endogenous to a subject, for example, present in the subject's genome. In some embodiments, a target gene or regulatory sequence thereof is not part of an engineered reporter system.
- [0226]** In some embodiments, a target gene is exogenous to a host subject, for example, a pathogen target gene or an exogenous gene expressed as a result of a therapeutic intervention, such as a gene therapy and/or cell therapy. In some embodiments, a target gene is an exogenous reporter gene. In some embodiments, a target gene is an exogenous synthetic gene.
- [0227]** In some embodiments, a target gene (e.g., target endogenous gene) is a gene that is over-expressed or under-expressed in a disease or condition. In some embodiments, a target gene is a gene that is over-expressed or under-expressed in a heritable genetic disease.
- [0228]** In some embodiments, a target gene (e.g., target endogenous gene) is a gene that is over-expressed or under-expressed in an autoimmune disease. In some embodiments, a target gene is a gene that is over-expressed or under-expressed in Acute disseminated encephalomyelitis, Acute motor axonal neuropathy, Addison's disease, Adiposis dolorosa, Adult-onset Still's disease, Alopecia areata, Ankylosing Spondylitis, Anti-Glomerular Basement Membrane nephritis, Anti-neutrophil cytoplasmic antibody-associated vasculitis, Anti-N-Methyl-D-Aspartate Receptor Encephalitis, Antiphospholipid syndrome, Antisynthetase syndrome, Aplastic anemia, Autoimmune Angioedema, Autoimmune Encephalitis, Autoimmune enteropathy, Autoimmune hemolytic anemia, Autoimmune hepatitis, Autoimmune inner ear disease, Autoimmune lymphoproliferative syndrome, Autoimmune neutropenia, Autoimmune oophoritis, Autoimmune orchitis, Autoimmune pancreatitis, Autoimmune polyendocrine syndrome, Autoimmune polyendocrine syndrome type 2, Autoimmune polyendocrine syndrome type 3, Autoimmune progesterone dermatitis, Autoimmune retinopathy, Autoimmune thrombocytopenic purpura, Autoimmune thyroiditis, Autoimmune urticaria, Autoimmune uveitis, Balo concentric sclerosis, Behçet's disease, Bickerstaff's encephalitis, Bullous pemphigoid, Celiac disease, Chronic fatigue syndrome, Chronic inflammatory

demyelinating polyneuropathy, Churg-Strauss syndrome, Cicatricial pemphigoid, Cogan syndrome, Cold agglutinin disease, Complex regional pain syndrome, CREST syndrome, Crohn's disease, Dermatitis herpetiformis, Dermatomyositis, Diabetes mellitus type 1, Discoid lupus erythematosus, Endometriosis, Enthesitis, Enthesitis-related arthritis, Eosinophilic esophagitis, Eosinophilic fasciitis, Epidermolysis bullosa acquisita, Erythema nodosum, Essential mixed cryoglobulinemia, Evans syndrome, Felty syndrome, Fibromyalgia, Gastritis, Gestational pemphigoid, Giant cell arteritis, Goodpasture syndrome, Graves' disease, Graves ophthalmopathy, Guillain–Barré syndrome, Hashimoto's Encephalopathy, Hashimoto Thyroiditis, Henoch-Schonlein purpura, Hidradenitis suppurativa, Idiopathic dilated cardiomyopathy, Idiopathic inflammatory demyelinating diseases, IgA nephropathy, IgG4-related systemic disease, Inclusion body myositis, Inflammatory Bowel Disease (IBD), Intermediate uveitis, Interstitial cystitis, Juvenile Arthritis, Kawasaki's disease, Lambert-Eaton myasthenic syndrome, Leukocytoclastic vasculitis, Lichen planus, Lichen sclerosus, Ligneous conjunctivitis, Linear IgA disease, Lupus nephritis, Lupus vasculitis, Lyme disease, Ménière's disease, Microscopic colitis, Microscopic polyangiitis, Mixed connective tissue disease, Mooren's ulcer, Morphea, Mucha-Habermann disease, Multiple sclerosis, Myasthenia gravis, Myocarditis, Myositis, Neuromyelitis optica, Neuromyotonia, Opsoclonus myoclonus syndrome, Optic neuritis, Ord's thyroiditis, Palindromic rheumatism, Paraneoplastic cerebellar degeneration, Parry Romberg syndrome, Parsonage-Turner syndrome, Pediatric Autoimmune Neuropsychiatric Disorder Associated with Streptococcus, Pemphigus vulgaris, Pernicious anemia, Pityriasis lichenoides et varioliformis acuta, POEMS syndrome, Polyarteritis nodosa, Polymyalgia rheumatica, Polymyositis, Postmyocardial infarction syndrome, Postpericardiotomy syndrome, Primary biliary cirrhosis, Primary immunodeficiency, Primary sclerosing cholangitis, Progressive inflammatory neuropathy, Psoriasis, Psoriatic arthritis, Pure red cell aplasia, Pyoderma gangrenosum, Raynaud's phenomenon, Reactive arthritis, Relapsing polychondritis, Restless leg syndrome, Retroperitoneal fibrosis, Rheumatic fever, Rheumatoid arthritis, Rheumatoid vasculitis, Sarcoidosis, Schnitzler syndrome, Scleroderma, Sjogren's syndrome, Stiff person syndrome, Subacute bacterial endocarditis, Susac's syndrome, Sydenham chorea, Sympathetic ophthalmia, Systemic Lupus Erythematosus, Systemic scleroderma, Thrombocytopenia, Tolosa-Hunt syndrome, Transverse myelitis, Ulcerative colitis, Undifferentiated connective tissue disease, Urticaria, Urticarial vasculitis, Vasculitis, or Vitiligo.

[0229] In some embodiments, a target gene (e.g., target endogenous gene) is a gene that is over-expressed or under-expressed in a cancer, for example, acute leukemia, astrocytomas, biliary cancer (cholangiocarcinoma), bone cancer, breast cancer, brain stem glioma, bronchioloalveolar cell lung cancer, cancer of the adrenal gland, cancer of the anal region, cancer of the bladder, cancer of the endocrine system, cancer of the esophagus, cancer of the head or neck, cancer of the kidney, cancer of the parathyroid gland, cancer of the penis, cancer of the pleural/peritoneal membranes, cancer of the salivary gland, cancer of the small intestine, cancer of the thyroid gland, cancer of the ureter, cancer of the urethra, carcinoma of the cervix, carcinoma of the endometrium, carcinoma of the fallopian tubes, carcinoma of the renal pelvis, carcinoma of the vagina, carcinoma of the vulva, cervical cancer, chronic leukemia, colon cancer, colorectal cancer, cutaneous melanoma, ependymoma, epidermoid tumors, Ewings sarcoma, gastric cancer, glioblastoma, glioblastoma multiforme, glioma, hematologic malignancies, hepatocellular (liver) carcinoma, hepatoma, Hodgkin's Disease, intraocular melanoma, Kaposi sarcoma, lung cancer, lymphomas, medulloblastoma, melanoma, meningioma, mesothelioma, multiple myeloma, muscle cancer, neoplasms of the central nervous system (CNS), neuronal cancer, small cell lung cancer, non-small cell lung cancer, osteosarcoma, ovarian cancer, pancreatic cancer, pediatric malignancies, pituitary adenoma, prostate cancer, rectal cancer, renal cell carcinoma, sarcoma of soft tissue, schwannoma, skin cancer, spinal axis tumors, squamous cell carcinomas, stomach cancer, synovial sarcoma, testicular cancer, uterine cancer, or tumors and their metastases, including refractory versions of any of the above cancers, or a combination thereof.

[0230] In some embodiments, a target gene (e.g., target endogenous gene) is a differentiation-associated gene, for example, SSEA1, SSEA3/4, SSEA5, TRA1-60/81, TRA1-85, TRA2-54, GCTM-2, TG343, TG30, CD9, CD29, CD133/prominin, CD140a, CD56, CD73, CD90, CD105, OCT4, NANOG, SOX2, CD30, CD50, AHR, Aiolos/IKZF3, CDX4, CREB, DNMT3A, DNMT3B, EGR1, FoxO3, GATA-1, GATA-2, GATA-3, Helios, HES-1, HHEX, HIF-1 alpha/HIF1A, HMGB1/HMG-1, HMGB3, Ikaros, c-Jun, LMO2, LMO4, c-Maf, MafB, MEF2C, MYB, c-Myc, NFATC2, NFIL3/E4BP4, Nrf2, p53, PITX2, PRDM16/MEL1, Prox1, PU.1/Spi-1, RUNX1/CBFA2, SALL4, SCL/Tal1, Smad2, Smad2/3, Smad4, Smad7, Spi-B, STAT Activators, STAT Inhibitors, STAT3, STAT4, STAT5a, STAT6, TSC22, DUX4, DUX4/DUX4c, DUX4c, EBF-1, EBF-2, EBF-3, ETV5, FoxC2, FoxF1, GATA-4, GATA-6, HMGA2, c-Jun, MYF-5, Myocardin, MyoD, Myogenin, NFATC2, p53, Pax3, PDX-1/IPF1, PLZF,

PRDM16/MEL1, RUNX2/CBFA1, Smad1, Smad3, Smad4, Smad5, Smad8, Smad9, Snail, SOX2, SOX9, SOX11, STAT Activators, STAT Inhibitors, STAT1, STAT3, TBX18, Twist-1, Twist-2, Brachyury, EOMES, FoxC2, FoxD3, FoxF1, FoxH1, FoxO1/FKHR, GATA-2, GATA-3, GBX2, Goosecoid, HES-1, HNF-3 alpha/FoxA1, c-Jun, KLF2, KLF4, KLF5, c-Maf, Max, MEF2C, MIXL1, MTF2, c-Myc, Nanog, NFkB/IkB Activators, NFkB/IkB Inhibitors, NFkB1, NFkB2, Oct-3/4, Otx2, p53, Pax2, Pax6, PRDM14, Rex-1/ZFP42, SALL1, SALL4, Smad1, Smad2, Smad2/3, Smad3, Smad4, Smad5, Smad8, Snail, SOX2, SOX7, SOX15, SOX17, STAT Activators, STAT Inhibitors, STAT3, SUZ12, TBX6, TCF-3/E2A, THAP11, UTF1, WDR5, WT1, ZNF206, ZNF281, KLF2, KLF4, c-Maf, c-Myc, Nanog, Oct-3/4, p53, SOX1, SOX2, SOX3, SOX15, SOX18, TBX18, ASCL2/Mash2, CDX2, DNMT1, ELF3, Ets-1, FoxM1, FoxN1, GATA-6, Hairless, HNF-4 alpha/NR2A1, IRF6, c-Maf, MITF, Miz-1/ZBTB17, MSX1, MSX2, MYB, c-Myc, Neurogenin-3, NFATC1, NKX3.1, Nrf2, p53, p63/TP73L, Pax2, Pax3, RUNX1/CBFA2, RUNX2/CBFA1, RUNX3/CBFA3, Smad1, Smad2, Smad2/3, Smad4, Smad5, Smad7, Smad8, Snail, SOX2, SOX9, STAT Activators, STAT Inhibitors, STAT3, SUZ12, TCF-3/E2A, TCF7/TCF1, Androgen R/NR3C4, AP-2 gamma, beta-Catenin, beta-Catenin Inhibitors, Brachyury, CREB, ER alpha/NR3A1, ER beta/NR3A2, FoxM1, FoxO3, FRA-1, GLI-1, GLI-2, GLI-3, HIF-1 alpha/HIF1A, HIF-2 alpha/EPAS1, HMGA1B, c-Jun, JunB, KLF4, c-Maf, MCM2, MCM7, MITF, c-Myc, Nanog, NFkB/IkB Activators, NFkB/IkB Inhibitors, NFkB1, NKX3.1, Oct-3/4, p53, PRDM14, Snail, SOX2, SOX9, STAT Activators, STAT Inhibitors, STAT3, TAZ/WWTR1, TBX3, Twist-1, Twist-2, WT1, or ZEB1.

[0231] In some embodiments, modulation of the expression level and/or epigenetic level (e.g., methylation level) of the target gene in the target cell (e.g., muscle cell) can effect modification (e.g., upregulation or downregulation) of a downstream gene (e.g., one or more downstream genes) of the target gene. In some cases, the target gene can be encoded by the D4Z4 repeat array (e.g., target gene being DUX4), and the downstream genes that are in turn modified in their expressions (e.g., downregulated) can include, but are not limited to, ZSCAN4, LEUTX, MBD3L2, TRIM48, TRIM43, DEFB103, ZFN217, RNASEL, EIF2AK2, BMP2, SP1 P21, MYC, MURF1, ATROGIN1, CRYM, PRAMEF1, RFPL2, KHDC1, SPRYD5, TPRX1, HSPA2, FGFR3, SLC2A14, ID2, PVRL3, SFRS2B, THOC4, ZNHIT6, DBR1, TFIP11, FBXO33, USP29, TRIM23, SLC34A2, CSAG3, and/or PNMA6B.

[0232] In some embodiments, modulation of the expression level and/or epigenetic level

(e.g., methylation level) of the target gene in the target cell can effect apoptosis of the target cell (e.g., muscle cell). In some cases, such modulation of the target gene can reduce stress in the target cell. For example, the modulation of the target gene (e.g., DUX4) can effect downregulation of one or more stress-related markers in the target cell. Non-limiting examples of the one or more stress-related markers can include ACTH, glucocorticoid receptor, CRHR-1/2, POMC, prolactin, arginine vasopressin receptor V1a, superoxide dismutase 1, superoxide dismutase 2, peroxiredoxin-3, CCR5, iNOS, eNOS, heme oxygenase-2, cyclooxygenase-2, HSP27, HSP40, HSP60, HSP70, HSP70i, HSP90, HSP110, GRP78/BIP, AIF, annexin II, annexin IV, caspase 1, caspase 2, caspase 3, caspase 6, cytokeratin, E-cadherin, and/or Annexin V, caspase 5, caspase 7, caspase 8, caspase 9, caspase 10, BAD, BAX, BAK, BCL2, BID, PARP-1, NOXA, PUMA, RIPK3, RIPK1, FADD, APAF1, DFF40, DFF45, ROCK. The one or more stress-related markers as disclosed herein can be an apoptotic marker.

[0233] In some embodiments, a heterologous gene effector is from a gene product that is a hematopoietic stem cell transcription factor. In some embodiments, a target gene is a mesenchymal stem cell transcription factor. In some embodiments, a target gene is an embryonic stem cell transcription factor. In some embodiments, a target gene is an induced pluripotent stem cell (iPSC) transcription factor. In some embodiments, a target gene is an epithelial stem cell transcription factor. In some embodiments, a target gene is a cancer stem cell transcription factor.

[0234] In some embodiments, a target gene is an age-related gene. In some embodiments, a target gene is a senescence-associated protein. In some embodiments, a target gene is a drug target.

[0235] In some embodiments, a target gene (e.g., target endogenous gene) is a cancer-related gene. Non-limiting examples of cancer-related genes include A1CF, ABI1, ABL1, ABL2, ACKR3, ACSL3, ACSL6, ACVR1, ACVR2A, AFDN, AFF1, AFF3, AFF4, AKAP9, AKT1, AKT2, AKT3, ALDH2, ALK, AMER1, ANK1, APC, APOBEC3B, AR, ARAF, ARHGAP26, ARHGAP5, ARHGEF10, ARHGEF10L, ARHGEF12, ARID1A, ARID1B, ARID2, ARNT, ASPSCR1, ASXL1, ASXL2, ATF1, ATIC, ATM, ATP1A1, ATP2B3, ATR, ATRX, AXIN1, AXIN2, B2M, BAP1, BARD1, BAX, BAZ1A, BCL10, BCL11A, BCL11B, BCL2, BCL2L12, BCL3, BCL6, BCL7A, BCL9, BCL9L, BCLAF1, BCOR, BCORL1, BCR, BIRC3, BIRC6, BLM, BMP5, BMPR1A, BRAF, BRCA1, BRCA2, BRD3, BRD4, BRIP1, BTG1, BTK, BUB1B, C15orf65, CACNA1D, CALR, CAMTA1, CANT1, CARD11, CARS, CASP3, CASP8, CASP9, CBFA2T3, CBFB, CBL, CBLB,

CBLC, CCDC6, CCNB1IP1, CCNC, CCND1, CCND2, CCND3, CCNE1, CCR4, CCR7, CD209, CD274, CD28, CD74, CD79A, CD79B, CDC73, CDH1, CDH10, CDH11, CDH17, CDK12, CDK4, CDK6, CDKN1A, CDKN1B, CDKN2A, CDKN2C, CDX2, CEBPA, CEP89, CHCHD7, CHD2, CHD4, CHEK2, CHIC2, CHST11, CIC, CIITA, CLIP1, CLP1, CLTC, CLTCL1, CNBD1, CNBP, CNOT3, CNTNAP2, CNTRL, COL1A1, COL2A1, COL3A1, COX6C, CPEB3, CREB1, CREB3L1, CREB3L2, CREBBP, CRLF2, CRNKL1, CRTCL1, CRTCL3, CSF1R, CSF3R, CSMD3, CTCF, CTNNA2, CTNNB1, CTNND1, CTNND2, CUL3, CUX1, CXCR4, CYLD, CYP2C8, CYSLTR2, DAXX, DCAF12L2, DCC, DCTN1, DDB2, DDIT3, DDR2, DDX10, DDX3X, DDX5, DDX6, DEK, DGCR8, DICER1, DNAJB1, DNMT2, DNMT3A, DROSHA, DUX4L1, EBF1, ECT2L, EED, EGFR, EIF1AX, EIF3E, EIF4A2, ELF3, ELF4, ELK4, ELL, ELN, EML4, EP300, EPAS1, EPHA3, EPHA7, EPS15, ERBB2, ERBB3, ERBB4, ERC1, ERCC2, ERCC3, ERCC4, ERCC5, ERG, ESR1, ETNK1, ETV1, ETV4, ETV5, ETV6, EWSR1, EXT1, EXT2, EZH2, EZR, FAM131B, FAM135B, FAM47C, FANCA, FANCC, FANCD2, FANCE, FANCF, FANCG, FAS, FAT1, FAT3, FAT4, FBLN2, FBXO11, FBXW7, FCGR2B, FCRL4, FEN1, FES, FEV, FGFR1, FGFR1OP, FGFR2, FGFR3, FGFR4, FH, FHIT, FIP1L1, FKBP9, FLCN, FLI1, FLNA, FLT3, FLT4, FNBP1, FOXA1, FOXL2, FOXO1, FOXO3, FOXO4, FOXP1, FOXR1, FSTL3, FUBP1, FUS, GAS7, GATA1, GATA2, GATA3, GLI1, GMPS, GNA11, GNAQ, GNAS, GOLGA5, GOPC, GPC3, GPC5, GPHN, GRIN2A, GRM3, H3F3A, H3F3B, HERPUD1, HEY1, HIF1A, HIP1, HIST1H3B, HIST1H4I, HLA-A, HLF, HMGA1, HMGA2, HMGN2P46, HNF1A, HNRNPA2B1, HOOK3, HOXA11, HOXA13, HOXA9, HOXC11, HOXC13, HOXD11, HOXD13, HRAS, HSP90AA1, HSP90AB1, ID3, IDH1, IDH2, IGF2BP2, IGH, IGK, IGL, IKBKB, IKZF1, IL2, IL21R, IL6ST, IL7R, IRF4, IRS4, ISX, ITGAV, ITK, JAK1, JAK2, JAK3, JAZF1, JUN, KAT6A, KAT6B, KAT7, KCNJ5, KDM5A, KDM5C, KDM6A, KDR, KDSR, KEAP1, KIAA1549, KIF5B, KIT, KLF4, KLF6, KLK2, KMT2A, KMT2C, KMT2D, KNL1, KNSTRN, KRAS, KTN1, LARP4B, LASP1, LATS1, LATS2, LCK, LCP1, LEF1, LEPROTL1, LHFPL6, LIFR, LMNA, LMO1, LMO2, LPP, LRIG3, LRP1B, LSM14A, LYL1, LZTR1, MACC1, MAF, MAFB, MALAT1, MALT1, MAML2, MAP2K1, MAP2K2, MAP2K4, MAP3K1, MAP3K13, MAPK1, MAX, MB21D2, MDM2, MDM4, MDS2, MECOM, MED12, MEN1, MET, MGMT, MITF, MLF1, MLH1, MLLT1, MLLT10, MLLT11, MLLT3, MLLT6, MN1, MNX1, MPL, MRTFA, MSH2, MSH6, MSI2, MSN, MTCP1, MTOR, MUC1, MUC16, MUC4, MUTYH, MYB, MYC, MYCL, MYCN, MYD88, MYH11,

MYH9, MYO5A, MYOD1, N4BP2, NAB2, NACA, NBEA, NBN, NCKIPSD, NCOA1, NCOA2, NCOA4, NCOR1, NCOR2, NDRG1, NF1, NF2, NFATC2, NFE2L2, NFIB, NFKB2, NFKBIE, NIN, NKX2-1, NONO, NOTCH1, NOTCH2, NPM1, NR4A3, NRAS, NRG1, NSD1, NSD2, NSD3, NT5C2, NTHL1, NTRK1, NTRK3, NUMA1, NUP214, NUP98, NUTM1, NUTM2B, NUTM2D, OLIG2, OMD, P2RY8, PABPC1, PAFAH1B2, PALB2, PATZ1, PAX3, PAX5, PAX7, PAX8, PBRM1, PBX1, PCBP1, PCM1, PDCD1LG2, PDE4DIP, PDGFB, PDGFRA, PDGFRB, PER1, PHF6, PHOX2B, PICALM, PIK3CA, PIK3CB, PIK3R1, PIM1, PLAG1, PLCG1, PML, PMS1, PMS2, POLD1, POLE, POLG, POLQ, POT1, POU2AF1, POU5F1, PPARG, PPFIBP1, PPM1D, PPP2R1A, PPP6C, PRCC, PRDM1, PRDM16, PRDM2, PREX2, PRF1, PRKACA, PRKAR1A, PRKCB, PRPF40B, PRRX1, PSIP1, PTCH1, PTEN, PTK6, PTPN11, PTPN13, PTPN6, PTPRB, PTPRC, PTPRD, PTPRK, PTPRT, PWWP2A, QKI, RABEP1, RAC1, RAD17, RAD21, RAD51B, RAF1, RALGDS, RANBP2, RAP1GDS1, RARA, RB1, RBM10, RBM15, RECQL4, REL, RET, RFWD3, RGPD3, RGS7, RHOA, RHOH, RMI2, RNF213, RNF43, ROBO2, ROS1, RPL10, RPL22, RPL5, RPN1, RSPO2, RSPO3, RUNX1, RUNX1T1, S100A7, SALL4, SBDS, SDC4, SDHA, SDHAF2, SDHB, SDHC, SDHD, 44444, 44445, 44448, SET, SETBP1, SETD1B, SETD2, SETDB1, SF3B1, SFPQ, SFRP4, SGK1, SH2B3, SH3GL1, SHTN1, SIRPA, SIX1, SIX2, SKI, SLC34A2, SLC45A3, SMAD2, SMAD3, SMAD4, SMARCA4, SMARCB1, SMARCD1, SMARCE1, SMC1A, SMO, SND1, SNX29, SOCS1, SOX2, SOX21, SPECC1, SPEN, SPOP, SRC, SRGAP3, SRSF2, SRSF3, SS18, SS18L1, SSX1, SSX2, SSX4, STAG1, STAG2, STAT3, STAT5B, STAT6, STIL, STK11, STRN, SUFU, SUZ12, SYK, TAF15, TAL1, TAL2, TBL1XR1, TBX3, TCEA1, TCF12, TCF3, TCF7L2, TCL1A, TEC, TENT5C, TERT, Tet1, Tet2, TFE3, TFEB, TFG, TFPT, TFRC, TGFB2, THRAP3, TLX1, TLX3, TMEM127, TMPRSS2, TNC, TNFAIP3, TNFRSF14, TNFRSF17, TOP1, TP53, TP63, TPM3, TPM4, TPR, TRA, TRAF7, TRB, TRD, TRIM24, TRIM27, TRIM33, TRIP11, TRRAP, TSC1, TSC2, TSHR, U2AF1, UBR5, USP44, USP6, USP8, VAV1, VHL, VTI1A, WAS, WDCP, WIF1, WNK2, WRN, WT1, WWTR1, XPA, XPC, XPO1, YWHAE, ZBTB16, ZCCHC8, ZEB1, ZFH3, ZMYM2, ZMYM3, ZNF331, ZNF384, ZNF429, ZNF479, ZNF521, ZNRF3, and ZRSR2.

[0236] In some embodiments, a target gene (e.g., target endogenous gene) is an immune cell-related gene, for example, a cytokine, cytokine receptor, chemokine, chemokine receptor, co-inhibitory immune receptor, co-stimulatory immune receptor, immune cell transcription factor, etc.

[0237] In some embodiments, a target gene (e.g., target endogenous gene) is a cytokine, for example, 4-1BBL, APRIL, CD153, CD154, CD178, CD70, G-CSF, GITRL, GM-CSF, IFN- α , IFN- β , IFN- γ , IL-1RA, IL-1 α , IL-1 β , IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IL-16, IL-17, IL-18, IL-20, IL-23, LIF, LIGHT, LT- β , M-CSF, MSP, OSM, OX40L, SCF, TALL-1, TGF- β , TGF- β 1, TGF- β 2, TGF- β 3, TNF- α , TNF- β , TRAIL, TRANCE, or TWEAK.

[0238] In some embodiments, a target gene (e.g., target endogenous gene) is a cytokine receptor, for example, A common gamma chain receptor, a common beta chain receptor, an interferon receptor, a TNF family receptor, a TGF-B receptor, Apo3, BCMA, CD114, CD115, CD116, CD117, CD118, CD120, CD120a, CD120b, CD121, CD121a, CD121b, CD122, CD123, CD124, CD126, CD127, CD130, CD131, CD132, CD212, CD213, CD213a1, CD213a13, CD213a2, CD25, CD27, CD30, CD4, CD40, CD95 (Fas), CDw119, CDw121b, CDw125, CDw131, CDw136, CDw137 (41BB), CDw210, CDw217, GTR, HVEM, IL-11R, IL-11Ra, IL-14R, IL-15R, IL-15Ra, IL-18R, IL-18R α , IL-18R β , IL-20R, IL-20R α , IL-20R β , IL-9R, LIFR, LT β R, OPG, OSMR, OX40, RANK, TACI, TGF- β R1, TGF- β R2, TGF- β R3, TRAILR1, TRAILR2, TRAILR3, or TRAILR4.

[0239] In some embodiments, a target gene (e.g., target endogenous gene) is a chemokine, for example, ACT-2, AMAC-a, ATAC, ATAC, BLC, CCL1, CCL11, CCL13, CCL14, CCL15, CCL16, CCL17, CCL18, CCL19, CCL2, CCL20, CCL21, CCL22, CCL23, CCL24, CCL25, CCL26, CCL27, CCL3, CCL4, CCL5, CCL7, CCL8, CKb-6, CKb-8, CTACK, CX3CL1, CXCL1, CXCL10, CXCL11, CXCL12, CXCL13, CXCL14, CXCL2, CXCL3, CXCL4, CXCL5, CXCL6, CXCL7, CXCL8, CXCL9, DC-CK1, ELC, ENA-78, eotaxin, eotaxin-2, eotaxin-3, Eskine, exodus-1, exodus-2, exodus-3, fractalkine, GCP-2, GROa, GROb, GROg, HCC-1, HCC-2, HCC-4, I-309, IL-8, ILC, IP-10, I-TAC, LAG-1, LARC, LCC-1, LD78 α , LEC, Lkn-1, LMC, lymphoactin, lymphoactin b, MCAF, MCP-1, MCP-2, MCP-3, MCP-4, MDC, MDNCF, MGSA-a, MGSA-b, MGSA-g, Mig, MIP-1d, MIP-1 α , MIP-1 β , MIP-2a, MIP-2b, MIP-3, MIP-3 α , MIP-3 β , MIP-4, MIP-4a, MIP-5, MPIF-1, MPIF-2, NAF, NAP-1, NAP-2, oncostatin, PARC, PF4, PPBP, RANTES, SCM-1a, SCM-1b, SDF-1 α / β -, SLC, STCP-1, TARC, TECK, XCL1, or XCL2.

[0240] In some embodiments, a target gene (e.g., target endogenous gene) is a chemokine receptor, for example, CCR1, CCR2, CCR3, CCR4, CCR5, CCR6, CCR7, CCR8, CCR9, CCR10, CX3CR1, CXCR1, CXCR2, CXCR3, CXCR4, CXCR5, XCR1, or XCR1.

[0241] In some embodiments, a target gene (e.g., target endogenous gene) is an activating NK receptor, for example, CD100 (SEMA4D), CD16 (Fc γ R11A), CD160 (BY55),

CD244 (2B4, SLAMF4), CD27, CD94– NKG2C, CD94– NKG2E, CD94–NKG2H, CD96, CRTAM, DAP12, DNAM1 (CD226), KIR2DL4, KIR2DS1, KIR2DS2, KIR2DS3, KIR2DS4, KIR2DS5, KIR3DS1, Ly49, NCR, NKG2D (KLRK1, CD314), NKp30 (NCR3), NKp44 (NCR2), NKp46 (NCR1), NKp80 (KLRF1, CLEC5C), NTB-A (SLAMF6), PSGL1, or SLAMF7 (CRACC, CS1, CD319).

[0242] In some embodiments, a target gene (e.g., target endogenous gene) is an inhibitory NK receptor, for example, CD161 (NKR-P1A, NK1.1), CD94– NKG2A, CD96, CEACAM1, KIR2DL1, KIR2DL2, KIR2DL3, KIR2DL4, KIR2DL5A, KIR2DL5B, KIR3DL1, KIR3DL2, KIR3DL3, KLRG1, LAIR1, LIR1 (ILT2, LILRB1), Ly49a, Ly49b, NKR-P1A (KLRB1), SIGLEC-10, SIGLEC-11, SIGLEC-14, SIGLEC-16, SIGLEC-3 (CD33), SIGLEC-5 (CD170), SIGLEC-6 (CD327), SIGLEC-7 (CD328), SIGLEC-8, SIGLEC-9 (CD329), SIGLEC-E, SIGLEC-F, SIGLEC-G, SIGLEC-H, or TIGIT.

[0243] In some embodiments, a target gene (e.g., target endogenous gene) is a co-inhibitory immune receptor, for example, 2B4, B7-1, BTLA, CD160, CTLA-4, DR6, Fas, LAG3, LAIR1, Ly108, PD-1, PD-L1, PD1H, TIGIT, TIM1, TIM2, or TIM3.

[0244] In some embodiments, a target gene (e.g., target endogenous gene) is co-stimulatory immune receptor, for example, 2B4, 4-1BB, CD2, CD4, CD8, CD21, CD27, CD28, CD30, CD40, CD84, CD226, CD355, CRACC, DcR3, DR3, GITR, HVEM, ICOS, Ly9, Ly108, LIGHT, LT β R, OX40, SLAM, TIM1, or TIM2.

[0245] In some embodiments, a target gene (e.g., target endogenous gene) is itself a gene effector, such as any of the gene effectors disclosed herein (e.g., a transcription factor disclosed herein).

[0246] In some embodiments, a target gene (e.g., target endogenous gene) is an immune cell transcription factor, for example, AP-1, Bcl6, E2A, EBF, Eomes, FoxP3, GATA3, Id2, Ikaros, IRF, IRF1, IRF2, IRF3, IRF7, NFAT, NFkB, Pax5, PLZF, PU.1, ROR- γ -T, STAT, STAT1, STAT2, STAT3, STAT4, STAT5, STAT5A, STAT5B, STAT6, T-bet, TCF7, or ThPOK.

[0247] In some embodiments, a target gene is a kinase, for example, a tyrosine kinase, or serine/threonine kinase. In some embodiments, a target gene is a phosphatase, for example, a tyrosine phosphatase, or serine/threonine phosphatase.

[0248] In some embodiments, a target gene is a receptor. In some embodiments, a target gene is an ion channel. In some embodiments, a target gene is a GPCR. In some embodiments, a target gene is a receptor tyrosine kinase. In some embodiments, a target gene is a ribosomal protein. In some embodiments, a target gene is a membrane protein. In

some embodiments, a target gene is a cytoplasmic protein. In some embodiments, a target gene is a nuclear protein. In some embodiments, a target gene is a mitochondrial protein. In some embodiments, a target gene is a ubiquitin ligase. In some embodiments, a target gene is a methyltransferase. In some embodiments, a target gene is a glycosyltransferase. In some embodiments, a target gene is a hydrolase.

[0249] In some embodiments, CD45 is a target gene used in compositions and methods of the disclosure (e.g., for gene expression activation screens). In some embodiments, CD45 is not used as a target gene. Compositions and methods disclosed herein to identify complexes that modulate CD45 expression can similarly be modified and adapted to other target genes (e.g., target endogenous genes), including those disclosed herein.

[0250] In some embodiments, CD71 is a target gene used in compositions and methods of the disclosure (e.g., for gene expression reduction screens). In some embodiments, CD71 is not used as a target gene. Compositions and methods disclosed herein to identify complexes that modulate CD71 expression can similarly be modified and adapted to other target genes (e.g., target endogenous genes), including those disclosed herein.

[0251] CELLS

[0252] Compositions, methods, and systems of the disclosure can be applied to cells of various types, and populations thereof. For example, a complex of the disclosure can be used to elicit changes in the expression, epigenetic modification, or activity level of a target gene (e.g., target endogenous gene) in cells of a particular type, or populations thereof. Methods of the disclosure can be used to identify complexes that are capable of eliciting changes in the expression or activity of target genes (e.g., target endogenous genes) in cells of a particular type, or populations thereof.

[0253] In some embodiments, a complex or a heterologous gene effector identified by methods of the disclosure effects a desirable change in expression of a target gene (e.g., target endogenous gene) that is specific to a particular cell type. In some embodiments, a complex or a heterologous gene effector identified by methods of the disclosure effects a desirable change in expression of a target gene (e.g., target endogenous gene) that is applicable to two or more cell types. In some embodiments, a complex or a heterologous gene effector identified by methods of the disclosure effects a desirable change in expression of a target gene (e.g., target endogenous gene) that is applicable to three or more cell types. In some embodiments, a complex or a heterologous gene effector identified by methods of the disclosure effects a desirable change in expression of a target gene (e.g., target endogenous gene) that is applicable to a class of cell types, for example,

cell types with overlapping functional roles, that are present in similar tissues, or that are from the same or similar differentiation lineages, e.g., stem cells, immune cells, T cells, T effector cells, etc. In some embodiments, a complex or a heterologous gene effector identified by methods of the disclosure effects a desirable change in expression of a target gene (e.g., target endogenous gene) that is broadly applicable to a wide variety of cell types, for example, elicits an expression level of a target gene that is above or below a certain threshold for multiple target cell types when introduced to the cells using suitable methods.

[0254] In some embodiments, a composition, complex, system, or method of the disclosure is used to effect a change in the expression, epigenetic modification, or activity level of a target gene in a primary cell. In some embodiments, a composition, complex, system, or method of the disclosure is used to effect a change in the expression, epigenetic modification, or activity level of a target gene in a cell line. In some embodiments, a composition, complex, system, or method of the disclosure is used to effect a change in the expression, epigenetic modification, or activity level of a target gene in an immortalized cell.

[0255] In some embodiments, a composition, complex, system, or method of the disclosure is used to effect a change in the expression, epigenetic modification, or activity level of a target gene (e.g., target endogenous gene) in a mammalian cell, for example, a human cell, non-human primate cell, non-rodent mammal cell, non-human mammal cell, swine cell, lagomorph cell, canine cell, etc. In some embodiments, a composition, complex, system, or method of the disclosure is used to effect a change in the expression, epigenetic modification, or activity level of a target gene in a plant cell, an avian cell, a reptilian cell, a bacterial cell, or an archaeal cell.

[0256] A composition, complex, system, or method of the disclosure can be used to effect a change in the expression, epigenetic modification, or activity level of a target gene (e.g., target endogenous gene) in a human cell.

[0257] A composition, complex, system, or method of the disclosure can be used to effect a change in the expression, epigenetic modification, or activity level of a target gene (e.g., target endogenous gene) in a stem cell.

[0258] A composition, complex, system, or method of the disclosure can be used to effect a change in the expression, epigenetic modification, or activity level of a target gene (e.g., target endogenous gene) in a differentiated cell.

[0259] A composition, complex, system, or method of the disclosure can be used to effect a

change in the expression, epigenetic modification, or activity level of a target gene (e.g., target endogenous gene) in a disease-associated cell.

[0260] A composition, complex, system, or method of the disclosure can be used to effect a change in the expression, epigenetic modification, or activity level of a target gene (e.g., target endogenous gene) in a cancer cell.

[0261] A composition, complex, system, or method of the disclosure can be used to effect a change in the expression, epigenetic modification, or activity level of a target gene (e.g., target endogenous gene) in a non-cancer cell.

[0262] A composition, complex, system, or method of the disclosure can be used to effect a change in the expression, epigenetic modification, or activity level of a target gene (e.g., target endogenous gene) in a lymphoid cell, such as a B cell, a T cell (Cytotoxic T cell, Natural Killer T cell, Regulatory T cell, T helper cell), Natural killer cell, cytokine induced killer (CIK) cells (see e.g. US20080241194); myeloid cells, such as granulocytes (Basophil granulocyte, Eosinophil granulocyte, Neutrophil granulocyte/Hypersegmented neutrophil), Monocyte/Macrophage, Red blood cell, Reticulocyte, Mast cell, Thrombocyte/Megakaryocyte, Dendritic cell; cells from the endocrine system, including thyroid (Thyroid epithelial cell, Parafollicular cell), parathyroid (Parathyroid chief cell, Oxyphil cell), adrenal (Chromaffin cell), pineal (Pinealocyte) cells; cells of the nervous system, including glial cells (Astrocyte, Microglia), Magnocellular neurosecretory cell, Stellate cell, Boettcher cell, and pituitary (Gonadotrope, Corticotrope, Thyrotrope, Somatotrope, Lactotroph); cells of the Respiratory system, including Pneumocyte (Type I pneumocyte, Type II pneumocyte), Clara cell, Goblet cell, Dust cell; cells of the circulatory system, including Myocardocyte, Pericyte; cells of the digestive system, including stomach (Gastric chief cell, Parietal cell), Goblet cell, Paneth cell, G cells, D cells, ECL cells, I cells, K cells, S cells; enteroendocrine cells, including enterochromaffin cell, APUD cell, liver cells (e.g., Hepatocyte, or Kupffer cell), Cartilage/bone/muscle; bone cells, including Osteoblast, Osteocyte, Osteoclast, teeth cells, (Cementoblast, Ameloblast); cartilage cells, including Chondroblast, Chondrocyte; skin cells, including Trichocyte, Keratinocyte, Melanocyte (Nevus cell); muscle cells, including Myocyte; urinary system cells, including Podocyte, Juxtaglomerular cell, Intraglomerular mesangial cell/Extraglomerular mesangial cell, Kidney proximal tubule brush border cell, Macula densa cell; reproductive system cells, including Spermatozoon, Sertoli cell, Leydig cell, Ovum; and other cells, including Adipocyte, Fibroblast, Tendon cell, Epidermal keratinocyte, Epidermal basal cell, Keratinocyte of fingernails and toenails, Nail bed basal

cell, Medullary hair shaft cell, Cortical hair shaft cell, Cuticular hair shaft cell, Cuticular hair root sheath cell, Hair root sheath cell of Huxley's layer, Hair root sheath cell of Henle's layer, External hair root sheath cell, Hair matrix cell, Wet stratified barrier epithelial cells, Surface epithelial cell of stratified squamous epithelium of cornea, tongue, oral cavity, esophagus, anal canal, distal urethra and vagina, basal cell of epithelia of cornea, tongue, oral cavity, esophagus, anal canal, distal urethra and vagina, Urinary epithelium cell, Exocrine secretory epithelial cells, Salivary gland mucous cell, Salivary gland serous cell, Von Ebner's gland cell in tongue, Mammary gland cell, Lacrimal gland cell, Ceruminous gland cell in ear, Eccrine sweat gland dark cell, Eccrine sweat gland clear cell, Apocrine sweat gland cell, Gland of Moll cell in eyelid, Sebaceous gland cell, Bowman's gland cell in nose, Brunner's gland cell in duodenum, Seminal vesicle cell, Prostate gland cell, Bulbourethral gland cell, Bartholin's gland cell, Gland of Littre cell, Uterus endometrium cell, Isolated goblet cell of respiratory and digestive tracts, Stomach lining mucous cell, Gastric gland zymogenic cell, Gastric gland oxyntic cell, Pancreatic acinar cell, Paneth cell of small intestine, Type II pneumocyte of lung, Clara cell of lung, Hormone secreting cells, Anterior pituitary cells, Somatotropes, Lactotropes, Thyrotropes, Gonadotropes, Corticotropes, Intermediate pituitary cell, Magnocellular neurosecretory cells, Gut and respiratory tract cells, Thyroid gland cells, thyroid epithelial cell, parafollicular cell, Parathyroid gland cells, Parathyroid chief cell, Oxyphil cell, Adrenal gland cells, chromaffin cells, Leydig cell of testes, Theca interna cell of ovarian follicle, Corpus luteum cell of ruptured ovarian follicle, Granulosa lutein cells, Theca lutein cells, Juxtaglomerular cell, Macula densa cell of kidney, Metabolism and storage cells, Barrier function cells (e.g., Lung, Gut, Exocrine Glands and Urogenital Tract), Kidney, Type I pneumocyte, Pancreatic duct cell (centroacinar cell), Nonstriated duct cell (of sweat gland, salivary gland, mammary gland, etc.), Duct cell (of seminal vesicle, prostate gland, etc.), Epithelial cells lining closed internal body cavities, Ciliated cells with propulsive function, Extracellular matrix secretion cells, Contractile cells; Skeletal muscle cells, stem cell, Heart muscle cells, Blood and immune system cells, Erythrocyte, Megakaryocyte, Monocyte, Connective tissue macrophage (various types), Epidermal Langerhans cell, Osteoclast, Dendritic cell, Microglial cell, Neutrophil granulocyte, Eosinophil granulocyte, Basophil granulocyte, Mast cell, Helper T cell, Suppressor T cell, Cytotoxic T cell, Natural Killer T cell, B cell, Natural killer cell, Reticulocyte, Stem cells and committed progenitors for the blood and immune system (various types), Pluripotent stem cells, Totipotent stem cells, Induced pluripotent stem cells, adult stem cells, Sensory

transducer cells, neurons, Autonomic neuron cells, Sense organ and peripheral neuron supporting cells, Central nervous system neurons and glial cells, Lens cells, Pigment cells, Melanocyte, Retinal pigmented epithelial cell, Germ cells, Oogonium/Oocyte, Spermatid, Spermatocyte, Spermatogonium cell, Spermatozoon, Nurse cells, Ovarian follicle cell, Sertoli cell, Thymus epithelial cell, Interstitial cells, Interstitial kidney cells, common myeloid progenitors, common lymphoid progenitors, and stem cells that are differentiated into or are to be differentiated into any cell type disclosed herein.

[0263] A composition, complex, system, or method of the disclosure can be used to effect a change in the expression, epigenetic modification, or activity level of a target gene (e.g., target endogenous gene) in a stem cell, for example, an isolated stem cell (e.g., an ESC) or an induced stem cell (e.g., an iPSC).

[0264] A composition, complex, system, or method of the disclosure can be used to effect a change in the expression, epigenetic modification, or activity level of a target gene (e.g., target endogenous gene) in a hematopoietic stem cell, for example, a hematopoietic stem cell from a subject, for example, from bone marrow, or peripheral blood (e.g., a mobilized peripheral blood apheresis product, for example, mobilized by administration of GCSF, GM-CSF, mozobil, or a combination thereof).

[0265] In some cases, pluripotency of stem cells (e.g., ESCs or iPSCs) can be determined, in part, by assessing pluripotency characteristics of the cells. Pluripotency characteristics can include, but are not limited to: pluripotent stem cell morphology; the potential for unlimited self-renewal; expression of pluripotent stem cell markers including, but not limited to SSEA1, SSEA3/4, SSEA5, TRA1-60/81, TRA1-85, TRA2-54, GCTM-2, TG343, TG30, CD9, CD29, CD133/prominin, CD140a, CD56, CD73, CD90, CD105, OCT4, NANOG, SOX2, CD30 and/or CD50; ability to differentiate to all three somatic lineages (ectoderm, mesoderm and endoderm); ability to form teratomas comprising the three somatic lineages; and/or (vi) formation of embryoid bodies comprising cells from the three somatic lineages.

[0266] A composition, complex, system, or method of the disclosure can be used to effect a change in the expression, epigenetic modification, or activity level of a target gene (e.g., target endogenous gene) in an immune cell, for example, lymphocytes, T cells, CD4+ T cells, CD8+ T cells, alpha-beta T cells, gamma-delta T cells, T regulatory cells (Tregs), cytotoxic T lymphocytes, Th1 cells, Th2 cells, Th17 cells, Th9 cells, naïve T cells, memory T cells, effector T cells, effector-memory T cells (TEM), central memory T cells (TCM), resident memory T cells (TRM), follicular helper T cells (TFH), Natural killer T

cells (NKTs), tumor-infiltrating lymphocytes (TILs), Natural killer cells (NKs), Innate Lymphoid Cells (ILCs), ILC1 cells, ILC2 cells, ILC3 cells, lymphoid tissue inducer (LTi) cells, B cells, B1 cells, B1a cells, B1b cells, B2 cells, plasma cells, B regulatory cells, memory B cells, marginal zone B cells, follicular B cells, germinal center B cells, antigen presenting cells (APCs), monocytes, macrophages, M1 macrophages, M2 macrophages, tissue-associated macrophages, dendritic cells, plasmacytoid dendritic cells, neutrophils, mast cells, basophils, eosinophils, common myeloid progenitors, common lymphoid progenitors, or any combination thereof.

[0267] A composition, complex, system, or method of the disclosure can be used to effect a change in the expression, epigenetic modification, or activity level of an engineered cell that is used to manufacture a biologic, for example, an antibody or other protein-based therapeutic.

EXAMPLES

Example 1: Regulation of DUX4 expression in a target cell population

[0268] A population of lymphoblasts was used as an example target cell population for regulating DUX4 expression level by the compositions and methods disclosed herein. The population of lymphoblasts were contacted by (i) the heterologous actuator moiety coupled to a gene regulator and (ii) a guide RNA (see **Table 1**) designed to direct the heterologous actuator moiety to a target polynucleotide sequence between two CpG islands within a D4Z4 repeat array that encodes DUX4 in the population of lymphoblasts. A number of the guide RNAs was able to allow the heterologous actuator moiety coupled to the gene regulator to complex with its respective target polynucleotide sequence in the population of lymphoblasts, and yield between about 0.2-fold and about 0.8-fold reduction in the DUX4 expression level (**Figure. 2**).

Table 1. Guide RNA molecules used in **FIG. 2**.

SEQ ID NO	Name	Sequence
SEQ ID NO. 1	DUX4_1 gRNA	CGCGGGGAGGGTGCTGTCCG
SEQ ID NO. 2	DUX4_2 gRNA	CCATCGCGGTGAGCCCCGGC
SEQ ID NO. 3	DUX4_3 gRNA	GGGCGTCGCCGTTGCCGGA
SEQ ID NO. 4	DUX4_4 gRNA	GAATGGCGGTGAGCCCCCT
SEQ ID NO. 5	DUX4_5 gRNA	CGGCTCTCCGGACCTCTCCA
SEQ ID NO. 6	DUX4_6 gRNA	GACCCAGGGCGTCGAGGCCT
SEQ ID NO. 7	DUX4_7 gRNA	TGACGGCGGTCCGCTTTCGC
SEQ ID NO. 8	DUX4_10 gRNA	TCCAGGCATCGCCGCCCGGG

SEQ ID NO. 9	DUX4_11 gRNA	CGGGACGGTCTCGCACACGC
SEQ ID NO. 10	DUX4_12 gRNA	CCTTTACAAGGGCGGCTGGC
SEQ ID NO. 11	DUX4_13 gRNA	CTCTCTGGGCTCCACGCGT
SEQ ID NO. 12	DUX4_14 gRNA	GTGCGAGACCGTCCCGGCAA
SEQ ID NO. 13	DUX4_15 gRNA	TCTCCCTGCTGCCGACGCGT
SEQ ID NO. 14	DUX4_91 gRNA	GTACGGGTTCCGCTCAAAGC

Example 2: Validation of *in vitro* FSHD model

[0269] In order to develop an *in vitro* FSHD model, two immortalized patient-derived FSHD skeletal myoblasts (SkM) cells, 12ABIC (12A) and 15ABIC (15A), were expanded in complete growth media. Upon confluence, the complete growth media was changed to differentiation media conditions for either 2 days or 7 days (D2 or D7 in **FIG. 3A**). The experiment included a negative control which was undifferentiated, proliferating myoblasts (UD in **FIG. 3A**). After differentiation, total RNA was extracted from the myoblasts and qRT-PCR was performed using TaqMan probes for DUX4 and DUX4-target genes ZSAC4, LEUTZ, MBD3L2, TRIM48, and TRIM43. GAPDH was included as an internal reference control for the qRT-pCR measurements and the double delta Ct method was used to calculate the gene fold change. The 12ABIC and 15ABIC cells showed increased DUX4 and DUX4-target gene expression consistent with FSHD presentation in patients.

[0270] The 12ABIC and 15ABIC cells were then tested for whether they also showed increased apoptosis consistent with the FSHD phenotype in patients. The 12ABIC and 15ABIC cells, as well as two corresponding healthy control cells 12UBIC and 15VBIC, respectively, were grown and then differentiated for 2 days after staining for an apoptosis marker, Caspase 3. The assay also included DAPI staining as a positive control. The cells were then imaged and analyzed using CellXpress PICO Imager. As shown in **FIG. 3B**, the 12ABIC and 15ABIC cells had increased apoptosis levels compared to their healthy sibling control myoblasts, 12UBIC and 15VBIC, at day 2 of differentiation. The 12ABIC, 15ABIC, 12UBIC, and 15VBIC cells were grown, differentiated, stained, imaged, and analyzed until day 7 of differentiation using the CellXpress PICO Imager. The percent of apoptotic cells for each cell type was plotted on day 0, 1, 2, and 7 of differentiation (**FIG. 3C**). 12ABIC and 15ABIC cells had a higher level of apoptosis at day 2 and day 7 compared to their corresponding healthy controls. This increase in apoptosis during differentiation was consistent with the *in vivo* phenotype of FSHD.

[0271] To ensure that the 12ABIC and 15ABIC cells had similar myoblast differentiation compared to their healthy sibling controls, all four cell types were immunostained for

Myosin Heavy Chain (MYHC), which was a late muscle gene used as a marker for muscle cell differentiation (**FIG. 3D**). The immunostaining assay also included DAPI staining as a positive control. In addition, the four cell types were assayed for the expression of MYHC, MYOG, and MyoMaker (MYMK) via qRT-PCR (**FIG. 3E**). GAPDH was included in the qRT-PCR as an internal control for the qRT-PCR measurements. MYOG was an essential myogenic regulatory factor that regulates skeletal muscle differentiation and MYMK was a late muscle gene used as a marker for muscle cell differentiation. The results of both the immunostaining and qRT-PCR experiments showed that the 12ABIC and 15ABIC cells had similar differentiation to their corresponding healthy sibling controls.

[0272] Overall, the results of these validation experiments showed that 12ABIC and 15ABIC cells presented the *in vivo* phenotypes of FSHD myoblasts and thus were an appropriate *in vitro* model for FSHD.

Example 3: Targeting of DUX4 for downregulation of expression

[0273] In order to target DUX4 for downregulation, numerous gRNAs were designed that spanned over the entire DZ4Z region locus, which included regions coding for long non-coding RNA, DBET, in the 5' end of the D4Z4 locus. The DZ4Z region locus was known to upregulate DUX4 gene expression upon deletion of the repeat units and DBET lncRNA has been shown to positively regulate the expression of DUX4 from the DZ4Z locus. The gRNAs were designed using the ChopChop CRISPR guide design tool. When designing the gRNAs, Hg38 human genome assembly was used with TTTR as the PAM sequence requirement. The map of the gRNAs designed to the DZ4Z locus is shown in **FIG. 4**.

[0274] Once the gRNAs were designed, the different gRNAs were tested with a Cas12f variant construct coupled with a KRAB modulator. 12ABIC cells stably expressing the Cas12f-KRAB effector-modulator were generated after lentiviral transduction of 12ABIC myoblasts. The design of the Cas12f-KRAB effector-modulator vector is shown in **FIG. 5**. The vector included a muscle-specific promoter (CK8e) to drive the expression of the Cas12f variant effector and the KRAB and DNMT3L domains. A human U6g promoter was included to drive the expression of the sgRNA spacer sequence with scaffold driven by RNA polymerase III. The vector additionally included a modified WPRE and polyadenylation regulatory sequences. The Cas12f-KRAB effector-modulator was labeled with mCherry, so after transduction, mCherry⁺ cells were sorted for enrichment. Following enrichment, annealed crRNA:trcrRNA constructs for 78 guides were

nucleofected in the Cas12f-KRAB effector-modulator-expressing 12ABIC cells. After differentiation of the myoblasts for 7 cells, the cells were assayed for expression of DUX4 (**FIG. 6A** and **6B**) and MDB3L2 (**FIG. 6B**) using Quantigene assay probes. The relative expression of DUX4 was normalized to expression of control gene HPRT1. The experiments showed that the different gRNAs were capable of downregulating expression of DUX4 and MDB3L2 in cells expressing the Cas12f-KRAB effector-modulator. In addition, the downregulation of DUX4 and MDB3L2 by the different gRNAs were positively correlated (**FIG. 6B**).

[0275] From the initial screen, six of the gRNAs were further tested. The six gRNAs were transfected into immortalized patient-derived FSHD myoblasts, along with one of two different Cas12f-KRAB effector-modulators. The two different Cas12f-KRAB effector-modulators included one of two different DNMT3L domains (e.g., DNMT3L-K1a or DNMT3L-K1b). Following transfection, the cells were differentiated and expression of DUX4 and DUX4-target genes, MBD3L2, TRIM48, and MYOG, were assayed using qRT-PCR after 17 days (**FIG. 7A**) and 18 days (**FIG. 7B**) post-transfection to measure for persistence of DUX4 repression. MYOG was included as a positive control to ensure that the differentiation ability of DUX4 sgRNA transfected cells was similar to control sgRNA transfected myoblasts. Overall, it was found that the Cas12f-KRAB-DNMT3L modulators resulted in persistent repression of DUX4 and DUX4-target genes.

[0276] In addition to testing the expression of levels of DUX4 and DUX4-target genes in patient-derived myoblasts treated with the Cas12f-KRAB-DNMT3L modulator, the cells were also tested for apoptosis levels of treated cells. The treated cells were stained for apoptotic marker, Caspase-3, 2 days after differentiation. Following staining, the apoptotic-positive cells were counted using a high content imager and the percent positive cells were calculated based on total nuclei stained by DAPI (blue cells). The cells treated with the Cas12f-KRAB-DNMT3L modulator showed a decrease in apoptosis compared to cells transfected with a control sgRNA (**FIGs. 8A** and **8B**).

Example 4: Ex Vivo FSHD model Establishment and Validation

[0277] Immortalized healthy sibling control cells and FSHD skeletal myoblasts can be thawed and expanded for *ex vivo* 3D studies. The skeletal myoblasts can be split onto 2D surfaces and then engineered into 3D Mantarray tissues as per established Curi Bio lab protocols as described in Fayazi, M., "Passive-Stretch Induced Skeletal Muscle Injury Platform for Duchenne Muscular Dystrophy Modeling," *Archives of Physical Medicine*

and Rehabilitation, volume 103, issue 3, March 2022, page e26, which is hereby incorporated in its entirety by reference. Briefly, 3D-skeletal myoblasts tissues can be cultured for 7 days to allow for compaction, and then additionally cultured for 14 days. Functional measurements can be taken three-times a week during culture to assess contractile force over a period and stimulated to assess the phenotypic differences in mechanical force, tetanic force, and fatigue (**FIG. 9**). Upon establishment of the model, the patient-derived FSHD skeletal myoblasts can be used to test the efficacy of the control and the Cas12f effector-modulator AAV targeting DZ4Z locus for the rescue in 3D tissue morphology, gene expression profile, and mechanical forces assessments.

[0278] In some cases, a 3D-skeletal myoblast tissue that is treated with the system, composition or method as disclosed herein (e.g., to modulate expression level or epigenetic level of a gene encoded by a D4Z4 repeat array, such as DUX4) can be characterized by exhibiting (i) enhanced mechanical force (e.g., a maximum mechanical force, an average mechanical force over a period of time), (ii) enhanced tetanic force (e.g., force indicative of a sustained muscle contraction evoked when the motor nerve that innervates a skeletal muscle emits action potentials at a very high rate), and/or (iii) reduced fatigue (e.g., as measured via a contraction against a fixed, immovable object (a static test or isometric measurement), or via a dynamic muscular contraction at a controlled velocity (repeated contractions or isokinetic assessment)), as compared to a control 3D-skeletal myoblast tissue (e.g., which is not treated with the system, composition or method as disclosed herein).

[0279] In some cases, the 3D-skeletal myoblast tissue that is treated with the system, composition or method as disclosed herein can be characterized by exhibiting a mechanical force that is greater than that in the control 3D-skeletal myoblast tissue, by at least or up to about 1%, at least or up to about 5%, at least or up to about 10%, at least or up to about 15%, at least or up to about 20%, at least or up to about 30%, at least or up to about 40%, at least or up to about 50%, at least or up to about 60%, at least or up to about 70%, at least or up to about 80%, at least or up to about 90%, at least or up to about 95%, at least or up to about 100%, at least or up to about 120%, at least or up to about 150%, at least or up to about 200%, at least or up to about 300%, at least or up to about 400%, or at least or up to about 500%.

[0280] In some cases, the 3D-skeletal myoblast tissue that is treated with the system, composition or method as disclosed herein can be characterized by exhibiting a tetanic force that is greater than that in the control 3D-skeletal myoblast tissue, by at least or up to

about 1%, at least or up to about 5%, at least or up to about 10%, at least or up to about 15%, at least or up to about 20%, at least or up to about 30%, at least or up to about 40%, at least or up to about 50%, at least or up to about 60%, at least or up to about 70%, at least or up to about 80%, at least or up to about 90%, at least or up to about 95%, at least or up to about 100%, at least or up to about 120%, at least or up to about 150%, at least or up to about 200%, at least or up to about 300%, at least or up to about 400%, or at least or up to about 500%.

[0281] In some cases, the 3D-skeletal myoblast tissue that is treated with the system, composition or method as disclosed herein can be characterized by exhibiting fatigue that is less than that in the control 3D-skeletal myoblast tissue, by at least or up to about 1%, at least or up to about 5%, at least or up to about 10%, at least or up to about 15%, at least or up to about 20%, at least or up to about 30%, at least or up to about 40%, at least or up to about 50%, at least or up to about 60%, at least or up to about 70%, at least or up to about 80%, at least or up to about 90%, at least or up to about 95%, at least or up to about 99%, or at least or up to about 100%.

Example 5: *In vivo* FSHD model for DUX4 targeting

[0282] On the morning of designated day -7, mice can be anesthetized using 90-200 mg/mg of ketamine and 10 mg/kg of xylazine administered intraperitoneally. The hind-limb of the mouse can be subjected to X-irradiation at 25 Gy at 2.2 Gy/minute over 11-12 minutes. Six days later, 60 uL of 0.3 mg/kg cardiotoxin can be administered along the length of the TA muscle to promote degradation. One day later, 2×10^6 human myoblast cells in 60 uL can be administered along the TA muscle. Isoflurane anesthesia can be used for subsequent cardiotoxin and human myoblast administration. One day after administration of the myoblast cells, the Cas12f modulator vector of the previous examples and a control AAVrh74 vector can be administered via retroorbital injection. At day 4 and day 21, the animals can be euthanized. The TA muscles and other major organs (e.g., heart, lung, liver) can be harvested. The harvested TA muscles can be sectioned, fixed, and H&E stained. The remaining organs can be processed, and total RNA/DNA can be extracted to perform gene expression experiments using qRT-PCR. The gene expression experiments can measure the expression of DUX4 and DUX4-target genes to determine the level of DUX4 repression. The gene expression experiments can measure the expression of one or more downstream genes of DUX4, such as, for example, ZSCAN4, LEUTX, MBD3L2, TRIM48, and/or TRIM43. The gene expression

experiments may also examine the enrichment of human myoblasts in the mouse as well as the AAV tropism to specific tissues. The workflow of the experiment is shown in **FIG. 10**.

Table 2. Guide RNA molecules for binding a target polynucleotide sequence for modifying expression level or epigenetic level of a gene (e.g., DUX4) encoded by a D4Z4 repeat array in a target cell (e.g., a muscle cell).

SEQ ID NO	Name	Sequence
45	DUX4_gRNA	TTTAAAGAGATCTGGGGATCTATA
46	DUX4_gRNA	TTTAAAGAATGGGAAAATTACGGG
47	DUX4_gRNA	TTTAACTTGGAAACACAGCGAAGT
48	DUX4_gRNA	TTTGCCTGTGAGTTCGAATGCACT
49	DUX4_gRNA	TTTGATGAAGTCTGGCTTACAGCC
50	DUX4_gRNA	TTTGCATATCTGATGGAGAACTTA
51	DUX4_gRNA	TTTGGGAATGTGTTTGTGAAGCAC
52	DUX4_gRNA	TTTGAATATACTGTGGTCATCTCT
53	DUX4_gRNA	TTTGCACTGGAGCAGAGATGACCA
54	DUX4_gRNA	TTTATTCTACTCTGCAATCCCCTA
55	DUX4_gRNA	TTTAAGATTCTGGGAGGGAGAGAA
56	DUX4_gRNA	TTTATAAATCTATTGTGCCTCAAG
57	DUX4_gRNA	TTTGATGAGTGCTGTATAGATCCC
58	DUX4_gRNA	TTTGCTACAGCACTAGTGAACTG
59	DUX4_gRNA	TTTACTGAGCCAGTCTTTAAATGC
60	DUX4_gRNA	TTTATAAAAAATGGCATGACAAGG
61	DUX4_gRNA	TTTATCAAAAAGCCAAACATTTCA
62	DUX4_gRNA	TTTGCTCACTGAGAATGCATAAGA
63	DUX4_gRNA	TTTAAGTTCTCCATCAGATATGCA
64	DUX4_gRNA	TTTATAAATTCACACAGAGACAC
65	DUX4_gRNA	TTTGAAATCTGGAAAGTTCTTAGC
66	DUX4_gRNA	TTTGTTGATATTTTGCTCATTCTG
67	DUX4_gRNA	TTTATGTTTTCTTCCAATGGGGA
68	DUX4_gRNA	TTTGTGAAGCACCTAGAATCTATA
69	DUX4_gRNA	TTTAAAGACTGGCTCAGTAAAGGG
70	DUX4_gRNA	TTTAATTCTCTCCTGAAGGAGATA
71	DUX4_gRNA	TTTAAATTTCACTCAGTTGTCTCT
72	DUX4_gRNA	TTTGTGTCTGCTGAGAAGAAAGAT
73	DUX4_gRNA	TTTGATAAATTGTCTAATGACTAG
74	DUX4_gRNA	TTTATTTTTTCACCCAGAACAGTA
75	DUX4_gRNA	TTTGCTGTTGTGTCTCTGTAGTGA
76	DUX4_gRNA	TTTAAAATATTAGTTTCCAGGACT
77	DUX4_gRNA	TTTAAATGCTAGATTTGATGAGTG
78	DUX4_gRNA	TTTAAAAGACTCTATCTCTGAATG
79	DUX4_gRNA	TTTATGTTCTCACAAGATTCTGGG
80	DUX4_gRNA	TTTATGCCATTTCTCCCTCTATT
81	DUX4_gRNA	TTTGGCATTGCTTTGGGGATCTG

82	DUX4_gRNA	TTTAACAAATAAAGATTTTTGCAT
83	DUX4_gRNA	TTTAAAAATAACGAATGAGCAAAA
84	DUX4_gRNA	TTTGAAATACAGTATTCCCAGAT
85	DUX4_gRNA	TTTGGGGATCTGGGAAAATCTGTG
86	DUX4_gRNA	TTTGGCTTTTTGATAAATTGTCTA
87	DUX4_gRNA	TTTGCTCATTCGTTATTTTAAAT
88	DUX4_gRNA	TTTATAAACTCAGTTATTATATT
89	DUX4_gRNA	TTTAAAAACCAACAGAAATCATA
90	DUX4_gRNA	TTTGTGAATAATATATGTTCAATT
91	DUX4_gRNA	TTTATCTCTTTGTTGATATTTTGC
92	DUX4_gRNA	TTTAGATTCTATTGTATATTTTCT
93	DUX4_gRNA	TTTAAAAAGAATAGAGGGAGAAAA
94	DUX4_gRNA	TTTATTTTAATCTTTGAAAGTCTT
95	DUX4_gRNA	TTTATTTGTTAAAATTCAGTTTCT
96	DUX4_gRNA	TTTGTTAAAATTCAGTTTCTGAAT
97	DUX4_gRNA	TTTAATCTTTGAAAGTCTTTATTT
98	DUX4_gRNA	TTTACAAGGGCGGCTGGCTGGGTG
99	DUX4_gRNA	TTTGTCCCGGAGGAAACCGCCAC
100	DUX4_gRNA	TTTGCCCTCCGCAAGGCGGCCTGT
101	DUX4_gRNA	TTTGGTTTCCGCGTGGCTTTGCCC
102	DUX4_gRNA	TTTAAAAAAAAAAATCACAAGGCA
103	DUX4_gRNA	TTTGAAAGTCTTATTTTTTTCTA
104	DUX4_gRNA	TTTACAAGGGCGGCTGGCTGGCTG
105	DUX4_gRNA	TTTAAAAATAGTTTTATCTCTTT
106	DUX4_gRNA	TTTATTTTTTTCTAATTTTTGAAA
107	DUX4_gRNA	TTTGCCCGGGTGCGGAGGCCAGCG
108	DUX4_gRNA	TTTAGGACGCGGGGTTGGGACGGG
109	DUX4_gRNA	TTTGCCCGGGTGCGGAGGCCACCG
110	DUX4_gRNA	TTTGGAACCTGGCAAGGAGAGCGA
111	DUX4_gRNA	TTTGCGGGCAGCCGCCTGGGCTGT
112	DUX4_gRNA	TTTGGCTCGGGGTCCAAACGAGTC
113	DUX4_gRNA	TTTGAGCGGAACCCGTACCCGGGC
114	DUX4_gRNA	TTTGGACCCCGAGCCAAAGCGAGG
115	DUX4_gRNA	TTTGCTCCCGGAGCTCTGCGGGCA
116	DUX4_gRNA	TTTACTCCCGGAGCTCTGCGGGCA
117	DUX4_gRNA	TTTGGTTTCAGAATGAGAGGTCAC
118	DUX4_gRNA	TTTACAAGAGAAAAACAAAAAACC
119	DUX4_gRNA	TTTGAGAAGGATCGCTTTCAGGC
120	DUX4_gRNA	TTTGTTTTCTCTGTAAATTTTT
121	DUX4_gRNA	TTTAATAGGGTTTTTTGTTTTTCT
122	DUX4_gRNA	TTTGTCCGGAGGAAACCGCCCACT
123	DUX4_gRNA	TTTGACCGCCAGGAGCTCCGCGCT
124	DUX4_gRNA	TTTACATGAGGTTCTACTACATAC
125	DUX4_gRNA	TTTGGATTCTGGGTTTCAAGTTAAGA
126	DUX4_gRNA	TTTAGGGTTAGGGTAGTGTAATA
127	DUX4_gRNA	TTTGGCTTATAGGGGCTTTGTGAG
128	DUX4_gRNA	TTTGTGGTAAAGAGTTGTGATTCT

129	DUX4_gRNA	TTTGGCCTACAGGGGGCTTTGTGA
130	DUX4_gRNA	TTTATTCACTAAATACAAATCACA
131	DUX4_gRNA	TTTATCAGTGTAATTATTAGTCAT
132	DUX4_gRNA	TTTGCAGAGATATGTCACAATCCC
133	DUX4_gRNA	TTTGGTCTAGTTTTATCAACAGAG
134	DUX4_gRNA	TTTGTCCAGTATGCTGCGGGTTGT
135	DUX4_gRNA	TTTGTTCCTGCAATATGTCACAA
136	DUX4_gRNA	TTTACACTACCCTAACCTAAACC
137	DUX4_gRNA	TTTGGAACGTAGGATGTTTCCATT
138	DUX4_gRNA	TTTATGACTAATAATTACACTGAT
139	DUX4_gRNA	TTTAGCAGGAACACACTACCTTTC
140	DUX4_gRNA	TTTATCAACAGAGCTAGTATTTAC
141	DUX4_gRNA	TTTAGCCTCTGCCTACAGGAGGCA
142	DUX4_gRNA	TTTACATCTCCTGAGTGAGCATTG
143	DUX4_gRNA	TTTGTCCATGATTTAGCAGGAACA
144	DUX4_gRNA	TTTGTTCATGAGAGATGTGGCAGGA
145	DUX4_gRNA	TTTGGGGGACGTGCTCCTTCTGCA
146	DUX4_gRNA	TTTAAGATGAAGCCCCTTGCTCC
147	DUX4_gRNA	TTTACCACAAACAACACAGCTTCA
148	DUX4_gRNA	TTTGTTGTGTGTGAATGAGAACA
149	DUX4_gRNA	TTTAAGGAAAATATGCTAATTTTA
150	DUX4_gRNA	TTTAGATTCATATGGGAATACTGA
151	DUX4_gRNA	TTTATATTACACTATTACTTAATA
152	DUX4_gRNA	TTTAGCTGAGGGAGATTGAGTGAC
153	DUX4_gRNA	TTTAGAATGCTACCTATTGCCTTC
154	DUX4_gRNA	TTTGCTCCTTCCTTAAGGATGTCT
155	DUX4_gRNA	TTTGTCTTCAAAGAATGGCCTTGG
156	DUX4_gRNA	TTTGGAGTTTAAAATTAGCATATT
157	DUX4_gRNA	TTTAGCTTTCTGGAACCTGGTATG
158	DUX4_gRNA	TTTACTGATCAACCAGATGATGTA
159	DUX4_gRNA	TTTAAAATTAGCATATTTTCCTTA
160	DUX4_gRNA	TTTGGTGATATATGACACAGAGAT
161	DUX4_gRNA	TTTGATTCTGATGTAAGAAATGAT
162	DUX4_gRNA	TTTGCATCTCTGTGTCATATATCA
163	DUX4_gRNA	TTTAAACTCCAAATACTTATGAAT
164	DUX4_gRNA	TTTGAAGACAAACATGTCTTAATA
165	DUX4_gRNA	TTTGCCTAGACAGCGTCGGAAGGT
166	DUX4_gRNA	TTTATTATTAGTAATAATGTGAAA
167	DUX4_gRNA	TTTGTATTTAGTGAATAAAAAACAA
168	DUX4_gRNA	TTTGTTTTTATTCACTAAATACAA
169	DUX4_gRNA	TTTAGATCACCTAGGTGATCAGTG
170	DUX4_gRNA	TTTGTTCTTATTTTAAGGAAAATA
171	DUX4_gRNA	TTTAGTGAATAAAAAACAAACAAAA
172	DUX4_gRNA	TTTGTTTGTTTTTATTCACTAAAT
173	DUX4_gRNA	TTTAGGCAGATCCTAGAAAAGAGT
174	DUX4_gRNA	TTTGCATCTTTTGTGTGATGAGTG
175	DUX4_gRNA	TTTAATATATCTCTGAACTAATCA

176	DUX4_gRNA	TTTGTCTAGGCTCTGCTTACTTGG
177	DUX4_gRNA	TTTAGGGTTAGGGTTAGGGTTATG
178	DUX4_gRNA	TTTGACATATGTCTGCACTGATGA
179	DUX4_gRNA	TTTGCCCGCTTCCTGGCTAGACCT
180	DUX4_gRNA	TTTGTCTAGGCTCTGGCTACACAG
181	DUX4_gRNA	TTTGTGTGATGAGTGCAGAGATAT
182	DUX4_gRNA	TTTAGGGTTAGGGTTAGGGTTAGG
183	DUX4_gRNA	TTTGTCTAGGCTCTGCCTACATAG
184	DUX4_gRNA	TTTAACATATCTCTACACTGATCA
185	DUX4_gRNA	TTTGCCTATGGGGGCAATGTGACA
186	DUX4_gRNA	TTTGAGATATCTCTGCACTGATCA
187	DUX4_gRNA	TTTGTGATATATATTTCCACTGCT
188	DUX4_gRNA	TTTGAGCAGTGGAAATATATATCA
189	DUX4_gRNA	TTTACATAACTTCGGTGATCAGTG
190	DUX4_gRNA	TTTAGGCACAGCTTAGACAAGCGT
191	DUX4_gRNA	TTTGTCAAGGATATGGCTACAGGG
192	DUX4_gRNA	TTTGTGAGATATCTCTGCACTGAT
193	DUX4_gRNA	TTTGTGACATACTTCTGTACTGAT
194	DUX4_gRNA	TTTGCTCTGATCACCCAGGTGATG
195	DUX4_gRNA	TTTGTTGATCAGTTCAGAGATGTG
196	DUX4_gRNA	TTTGTCTACGCTCTGCCTATGGGG
197	DUX4_gRNA	TTTGTCTACAGGGGGCTTTGTGAT
198	DUX4_gRNA	TTTGTGAAATTCCTGTAGGCAG
199	DUX4_gRNA	TTTGTGACATACCTTTGCTCTGAT
200	DUX4_gRNA	TTTAGGCAGAGCTTAGACTAGAGT
201	DUX4_gRNA	TTTGTCTAGGCTCTGTCTACGGGG
202	DUX4_gRNA	TTTGACATATCTCTGCACTGTAA
203	DUX4_gRNA	TTTGGCAGAGCCTAGACAAGGGTT
204	DUX4_gRNA	TTTGCCTACAGAAGGCTTTGTGAC
205	DUX4_gRNA	TTTGTCTACAGGGGGCTTTGTGAC
206	DUX4_gRNA	TTTGACACAATGCCCCCATAGACA
207	DUX4_gRNA	TTTGTGACTTCTCTCTGCACTGAT
208	DUX4_gRNA	TTTGTGACATAACTCTGCACTAAT
209	DUX4_gRNA	TTTGACATAGCTCTGCACAGATCA
210	DUX4_gRNA	TTTAAGCAGAGCCTAGACAATAGT
211	DUX4_gRNA	TTTGTGACATATCTTTGCACTGAT
212	DUX4_gRNA	TTTGCCTACAGGGGACATTGTGAC
213	DUX4_gRNA	TTTGTCTAGGCTCTGCCTAAGGGG
214	DUX4_gRNA	TTTGTGATCAGTTCAGGGATATGT
215	DUX4_gRNA	TTTGCACTGATCACCCAGGAGATG
216	DUX4_gRNA	TTTGTGACACATCTCTGCACTGAT
217	DUX4_gRNA	TTTGTGACATATCCCTGCAATGAT
218	DUX4_gRNA	TTTATAAGCACTGCCTACAGGGAA
219	DUX4_gRNA	TTTGTGACATATTTCTGCACTGAT
220	DUX4_gRNA	TTTGTGACATATCACTGCACTGAT
221	DUX4_gRNA	TTTGACATATCTCTGCACTGATCA
222	DUX4_gRNA	TTTGTGACATATCTATGCACTGAT

223	DUX4_gRNA	TTTATGACTTATCCCTGCACTGAT
224	DUX4_gRNA	TTTGTGACATATCTCTGCACAGAT
225	DUX4_gRNA	TTTGTGACATATCTCTGCACTGAT
226	DUX4_gRNA	TTTGAGATGGAGTCTTGCTCTGTT
227	DUX4_gRNA	TTTGTATTTTTAGTAGAGACGAGG
228	DUX4_gRNA	TTTGCTATCAAAAGCTTGGGTCAA
229	DUX4_gRNA	TTTAAAAGAACTTGCAGGTGTTT
230	DUX4_gRNA	TTTAAAAGATCTCTGTGGCCAGGC
231	DUX4_gRNA	TTTGAAGTATAGATACAGCAGAAG
232	DUX4_gRNA	TTTAAAATATTTAACATTTAGCCC
233	DUX4_gRNA	TTTGATAGCAAAAGGTAGAAAAGA
234	DUX4_gRNA	TTTGGAGAATGAAAACGTGCAGTA
235	DUX4_gRNA	TTTGAAGGAGATCTCACAAACAGG
236	DUX4_gRNA	TTTAGAAAGGAAACAGGCTGGAAA
237	DUX4_gRNA	TTTGCTGCAAAATAAATACACGCT
238	DUX4_gRNA	TTTATCATCTATCTATCTACCTCC
239	DUX4_gRNA	TTTGTTTGAAGTATAGATACAGCA
240	DUX4_gRNA	TTTGCAACTGTGGGTTTTCCAGCC
241	DUX4_gRNA	TTTAAAACAAGAACTCTGTAGGA
242	DUX4_gRNA	TTTGACCTTAAATCTGTGAAATC
243	DUX4_gRNA	TTTGTCGCTTCAAACACCGCAAGT
244	DUX4_gRNA	TTTACATCCATGATTTTTCACTGT
245	DUX4_gRNA	TTTGTCATCTCACCTCTCCAGAT
246	DUX4_gRNA	TTTGAAGGTTAGAACTAGTGGTCT
247	DUX4_gRNA	TTTAAGGTGCAAAAAGTCACTGGG
248	DUX4_gRNA	TTTACTTACAAACGCAGACTGTGT
249	DUX4_gRNA	TTTGTCACAGCAGCCTTTGTCGCT
250	DUX4_gRNA	TTTGAAGCGACAAAGGCTGCTGTG
251	DUX4_gRNA	TTTGTTTCCACATTACAGAGTGGG
252	DUX4_gRNA	TTTGCCACACCTTGTTTTAGAAAG
253	DUX4_gRNA	TTTAATCTCAGTGACAGGGGAACA
254	DUX4_gRNA	TTTAAAAGAATTATATCAACCTTT
255	DUX4_gRNA	TTTGCCTCATCTTTGTTTGAACATA
256	DUX4_gRNA	TTTATGATGTGATGGAGAATTCTT
257	DUX4_gRNA	TTTACTAATCTGCTTATTACCCAC
258	DUX4_gRNA	TTTGTGAGATCTCCTTCAAATACT
259	DUX4_gRNA	TTTGTGCATTGTCTGTTACTGTGT
260	DUX4_gRNA	TTTGCAGCAAACATTTACATCCAT
261	DUX4_gRNA	TTTACTATCTGTCTTTTCTACCTT
262	DUX4_gRNA	TTTATGCCTGGCCTTTCCATCCTT
263	DUX4_gRNA	TTTGGTTTTTGAAGGTTAGAACTA
264	DUX4_gRNA	TTTAAATGGTAGAGTGAACATACA
265	DUX4_gRNA	TTTGCTTTGCAACTGTGGGTTTTTC
266	DUX4_gRNA	TTTAGAAAGCTCAGGTTTATGATG
267	DUX4_gRNA	TTTAAAAAAAATTCCTTTCACTG
268	DUX4_gRNA	TTTAAATTTTCCAAGCCATATGGT
269	DUX4_gRNA	TTTGGTTCATGAAATTTTCAGTTT

270	DUX4_gRNA	TTTGTTTTAATCTCAGTGACAGGG
271	DUX4_gRNA	TTTAAAACAAGAGTCAGCAAATAT
272	DUX4_gRNA	TTTGTTGAGAAAAAATGAGTTGGA
273	DUX4_gRNA	TTTATTTTGCAGCAAACATTTACA
274	DUX4_gRNA	TTTGCTGACTCTTGTTTTAAAAGA
275	DUX4_gRNA	TTTGTTGTTTGCTTTTCTGTGGGG
276	DUX4_gRNA	TTTGCTTTTCTGTGGGGTTTTTGT
277	DUX4_gRNA	TTTAACATTTAGCCCTTGCAGAA
278	DUX4_gRNA	TTTGACTTCTTCCTCTGTTTTTG
279	DUX4_gRNA	TTTGAAAATAATTAAGCAATATCT
280	DUX4_gRNA	TTTAGCCCTTGCAGAAAATATTT
281	DUX4_gRNA	TTTAAAAAAGCAAACAACAACAA
282	DUX4_gRNA	TTTGCTTTTTTAAAAAAAATTCC
283	DUX4_gRNA	TTTGTAAGTAAAGTTGTAATGGGA
284	DUX4_gRNA	TTTGTTTGTTGTTTGCTTTTCTGT
285	DUX4_gRNA	TTTGCAGAAAATATTTGCTGACTC
286	DUX4_gRNA	TTTGGGATGCCGAGGCTAGCCGAT
287	DUX4_gRNA	TTTGTTGTTGTTGTTTGCTTTTTT
288	DUX4_gRNA	TTTGTTTGTTGTTGTTTGCTTTT
289	DUX4_gRNA	TTTAGTAGAGACGAGGTTTCACTG
290	DUX4_gRNA	TCTAATGACTAGATTCTTTCTCTC
291	DUX4_gRNA	TCTGGCCTTGCCGTGACGTTTAA
292	DUX4_gRNA	TCTATAGCCTGGACTATTGCTGTC
293	DUX4_gRNA	TCTGTTATAGTTTGTCTACTGAG
294	DUX4_gRNA	TCTGGAAACAGTTTGCACTGGAGC
295	DUX4_gRNA	TCTGGGGATCTATACAGCACTCAT
296	DUX4_gRNA	TCTATATGACTCCATCATGTCCTT
297	DUX4_gRNA	TCTGGCTTACAGCCTGTCCACTGC
298	DUX4_gRNA	TCTGCAATCCCCTAAGGCTTTTTT
299	DUX4_gRNA	TCTGGACTTCGCTGTGTTTCCAAG
300	DUX4_gRNA	TCTGCTCCAGTGCAAAGTGTTC
301	DUX4_gRNA	TCTAGTCATTAGACAATTTATCAA
302	DUX4_gRNA	TCTGTGCACACTTCTGGAGACCCT
303	DUX4_gRNA	TCTATACAGCACTCATCAAATCTA
304	DUX4_gRNA	TCTACTCTGCAATCCCCTAAGGCT
305	DUX4_gRNA	TCTGGGAGGATTTTGCCTGTGAGT
306	DUX4_gRNA	TCTGGAAAGTTCTTAGCATCCCCG
307	DUX4_gRNA	TCTAGCATTTAAAGACTGGCTCAG
308	DUX4_gRNA	TCTATTTTCTTGCTGTAACAGAG
309	DUX4_gRNA	TCTAGTAGTTAACAACCTCAGCTT
310	DUX4_gRNA	TCTGAATGTATTGGTCTACTTGAT
311	DUX4_gRNA	TCTGCTCCATTGTTTGCTGTTGTG
312	DUX4_gRNA	TCTATCTCTGAATGTATTGGTCTA
313	DUX4_gRNA	TCTGTTACAGCAAGGAAAATAGAA
314	DUX4_gRNA	TCTGCTGAGAAGAAAGATGAGTGT
315	DUX4_gRNA	TCTGATGGAGAACTTAAAATAATC
316	DUX4_gRNA	TCTGGAGACCCTTGTCATGCCATT

317	DUX4_gRNA	TCTGACTTGAGGCACAATAGATTT
318	DUX4_gRNA	TCTAGGTGCTTCACAAACACATTC
319	DUX4_gRNA	TCTGTGGTATTGCAGTTTCACTAG
320	DUX4_gRNA	TCTGTAGTGAATTTATAAACTCA
321	DUX4_gRNA	TCTATTGTATATTTTCTTCCCCAG
322	DUX4_gRNA	TCTGCCTTCTCTGTGTGCCTTGTG
323	DUX4_gRNA	TCTACTGTTTTAATTCTCTCCTGA
324	DUX4_gRNA	TCTGGGAGGGAGAGAAAAAGCCTT
325	DUX4_gRNA	TCTGGGTGAAAAAATAAACTGCAG
326	DUX4_gRNA	TCTGGGAAAATCTGTGCACACTTC
327	DUX4_gRNA	TCTATTGTGCCTCAAGTCAGAAGT
328	DUX4_gRNA	TCTGACTAGTTTGGCATTGCTTTT
329	DUX4_gRNA	TCTGGGAAATACTGTATTTCAAAA
330	DUX4_gRNA	TCTATTCTTTTTAAAAGACTCTAT
331	DUX4_gRNA	TCTGAAATAATGTTTATGCCATTT
332	DUX4_gRNA	TCTGTCCGGCCCCACCACCACCAC
333	DUX4_gRNA	TCTGCACCAATGAAAAAAAAAATTT
334	DUX4_gRNA	TCTAAAATACATTGAGAAAAAATT
335	DUX4_gRNA	TCTATGATTTCTGTTGGGTTTTTA
336	DUX4_gRNA	TCTAATTTTTGAAATACAGTATTT
337	DUX4_gRNA	TCTGAATATTTATGTTTTTCTTCC
338	DUX4_gRNA	TCTGTTGGGTTTTTAAAAATAGTT
339	DUX4_gRNA	TCTGTGTGCCTTGTGATTTTTTTT
340	DUX4_gRNA	TCTACTTGATGGTGTCCAGTAAGT
341	DUX4_gRNA	TCTGTGAACCGCGCGGGTGAAGAC
342	DUX4_gRNA	TCTGTGAACCGCGCGGGTGAAAC
343	DUX4_gRNA	TCTGGCGGGCCGCGTCTCCCGGGC
344	DUX4_gRNA	TCTAGGTCTCCCGTTCCTCTCTCC
345	DUX4_gRNA	TCTACGTGGAAATGAACGAGAGCC
346	DUX4_gRNA	TCTGTCTTTCCCTCCGTTCCTCCC
347	DUX4_gRNA	TCTGCCCCGCTTCCCTCCCGCCTG
348	DUX4_gRNA	TCTGCCCCTGCCGCGCGGAGGCGG
349	DUX4_gRNA	TCTGCGCCCCCGCGCCACCGTCGC
350	DUX4_gRNA	TCTGGGCTCCACGCGTCGGCAGC
351	DUX4_gRNA	TCTGGCCAGCTCCTCCCGGGCGGC
352	DUX4_gRNA	TCTGCCCCTGCCGCGCGGAGGCGT
353	DUX4_gRNA	TCTGCCCCGCTCCGTCCGTGAAAT
354	DUX4_gRNA	TCTGCCGTCGCGGCCTGGCTGGGC
355	DUX4_gRNA	TCTAGGAGAGGTTGCGCCTGCTGC
356	DUX4_gRNA	TCTGCAGCAGGCGCAACCTCTCCT
357	DUX4_gRNA	TCTGCGTTCGCGGCCAGGCGCTC
358	DUX4_gRNA	TCTAGGCCCGGTGAGAGACTCCAC
359	DUX4_gRNA	TCTGGTCTTCTACGTGGAAATGAA
360	DUX4_gRNA	TCTGCAGTGTGGCCGTTTGAAC
361	DUX4_gRNA	TCTAGGTCTAGGCCCGGTGAGAGA
362	DUX4_gRNA	TCTGCACTCCCCTGCGGCCTGCTG
363	DUX4_gRNA	TCTGGGGTCTCGCTCTGGTCTTCT

364	DUX4_gRNA	TCTGGTGGCGATGCCCCGGGTACGG
365	DUX4_gRNA	TCTGGGATCCGGTGACGGCGGTCC
366	DUX4_gRNA	TCTGCTGGAGGAGCTTTAGGACGC
367	DUX4_gRNA	TCTGCCGGCGCGGCCTGGCTGGGC
368	DUX4_gRNA	TCTGGGATCCCCGGGATGCCAGG
369	DUX4_gRNA	TCTGCCCGGGCTGCTCCACAGCC
370	DUX4_gRNA	TCTGAATCCTGGACTCCGGGAGGC
371	DUX4_gRNA	TCTGCGGGCACCCGGAACATGCA
372	DUX4_gRNA	TCTGGACCCTGGGCTCCGGAATGC
373	DUX4_gRNA	TCTGGTTTCAGAATCGAAGGGCCA
374	DUX4_gRNA	TCTGAAACCAAATCTGGACCCTGG
375	DUX4_gRNA	TCTGAAACCAGATCTGAATCCTGG
376	DUX4_gRNA	TCTGTCTCTCCCTCCGTTCTCCC
377	DUX4_gRNA	TCTGGGCTCCACGCATCGGCAGC
378	DUX4_gRNA	TCTGGTTTCAGAATTGAAGGGCCA
379	DUX4_gRNA	TCTGTTCTCATTACACACAACA
380	DUX4_gRNA	TCTAGGCAAACCTGGATTAGAGTT
381	DUX4_gRNA	TCTAAACCTTGTATGGGCTTTGCC
382	DUX4_gRNA	TCTACGGCAGCTTTGACATATGTC
383	DUX4_gRNA	TCTAATCCAGGTTTGCCTAGACAG
384	DUX4_gRNA	TCTGGCTGAATGTCTCCCCCACC
385	DUX4_gRNA	TCTACACTCTGTCTACGGCAGCTT
386	DUX4_gRNA	TCTGTCTACGGCAGCTTTGACATA
387	DUX4_gRNA	TCTAGTCTTTTCTATGTGGGTTT
388	DUX4_gRNA	TCTACTATGGAGTTCTGAAACACA
389	DUX4_gRNA	TCTGTCTTTGCCGCTTCTGGCT
390	DUX4_gRNA	TCTAGGTTCACTCTACTATGGAGT
391	DUX4_gRNA	TCTAGGCTTTGGCCTACAGGGGGC
392	DUX4_gRNA	TCTGCAGCCTGTAGCTCCTGGGGA
393	DUX4_gRNA	TCTATCACAGTGCCCCCATAGGCA
394	DUX4_gRNA	TCTGGCTTCATTTTGGGGGACGTG
395	DUX4_gRNA	TCTATAGGATCCACAGGGAGGGGG
396	DUX4_gRNA	TCTGACACATCTCTGAACTGATCA
397	DUX4_gRNA	TCTATGTTCTTCACTGCCTCATAC
398	DUX4_gRNA	TCTGGGCGATCAGTGCAGAGAGAA
399	DUX4_gRNA	TCTGTCTGTCTTTGCCGCTTCCT
400	DUX4_gRNA	TCTGTCTGGCTTCATTTTGGGGGA
401	DUX4_gRNA	TCTGTGGACAGTTTCTCCTCATGG
402	DUX4_gRNA	TCTGTTGATAAACTAGACCAAAA
403	DUX4_gRNA	TCTAAACACTACTCTGCTATTAGT
404	DUX4_gRNA	TCTGTGTTCACTATTCCCATATGA
405	DUX4_gRNA	TCTACGGGGGCATTGTGACATATC
406	DUX4_gRNA	TCTGTGCAGAGATATGTCACAAAG
407	DUX4_gRNA	TCTGAAATTGTCATGCAGTGA CTC
408	DUX4_gRNA	TCTAAGCCTCGTGGTTAGTGGGGA
409	DUX4_gRNA	TCTGTAGATCTCTGCCATT CATAA
410	DUX4_gRNA	TCTGCCTAAGCTTGAGTGAGTCAC

411	DUX4_gRNA	TCTGCCTCTCAAGAAATTCCTGCC
412	DUX4_gRNA	TCTGATGTAAGAAATGATGCTCAC
413	DUX4_gRNA	TCTGCCATTCTAAGTATTTGGAG
414	DUX4_gRNA	TCTGAACCTAGACAGGAGTTACAT
415	DUX4_gRNA	TCTAGTTTTATCAACAGAGCTAGT
416	DUX4_gRNA	TCTGTTTAGAATGCTACCTATTGC
417	DUX4_gRNA	TCTATCCAGAAGGCAATAGGTAGC
418	DUX4_gRNA	TCTATATCCAGCCTCATCTATTTT
419	DUX4_gRNA	TCTACTACATACCAGGTTCCAGAA
420	DUX4_gRNA	TCTGGAACCTGGTATGTAGTAGAA
421	DUX4_gRNA	TCTGGATAGAATCACAACTCTTTA
422	DUX4_gRNA	TCTAAACAGAGATCCTTTTTTTTT
423	DUX4_gRNA	TCTGCATCTCCCAGAGCCAGCCTG
424	DUX4_gRNA	TCTGGGAAGCTGACAATCCATCAG
425	DUX4_gRNA	TCTGGGAGATGCAGAAGGAGCACG
426	DUX4_gRNA	TCTACAGATTTGATTCTGATGTAA
427	DUX4_gRNA	TCTGTGTCATATATCACCAAATCT
428	DUX4_gRNA	TCTGAAACACATCTGCACTGATCA
429	DUX4_gRNA	TCTACAGGGGATATTGTGACATAT
430	DUX4_gRNA	TCTATTTCTGCTCCTCCTCCTTAT
431	DUX4_gRNA	TCTGCTCCTCCTCCTTATTTTCCT
432	DUX4_gRNA	TCTAGGTGATGTAACCTTTGTCCA
433	DUX4_gRNA	TCTGGTTGATCAGTAAAGAGATAT
434	DUX4_gRNA	TCTGCTATTAGTAGCTGTGTGACC
435	DUX4_gRNA	TCTGATCACCCAGGTGATGTAACCT
436	DUX4_gRNA	TCTAGGTGATGTAACCTTTGTCTA
437	DUX4_gRNA	TCTGTAGGCAGAGCCTAGACAAGA
438	DUX4_gRNA	TCTACTGGGAGCATTGTGACATAT
439	DUX4_gRNA	TCTAGCCAGGAAGCGGGCAAAGAC
440	DUX4_gRNA	TCTGGGTGATCTTTGCAGAGATAT
441	DUX4_gRNA	TCTAGACAAGAGTTACATCTCCTG
442	DUX4_gRNA	TCTGACATATCTCTGCACTGATCA
443	DUX4_gRNA	TCTGGGTCATCAGTGCAGACATAT
444	DUX4_gRNA	TCTATACTCTGCCTGCAGGGACAT
445	DUX4_gRNA	TCTGCAATGATCACTCATGTGATG
446	DUX4_gRNA	TCTACCCTCTGCCTACAGGGGGCG
447	DUX4_gRNA	TCTGTGCCCTTGTTCTTCCGTGAA
448	DUX4_gRNA	TCTGCACTCATCACACAAAAGATG
449	DUX4_gRNA	TCTAAGCTCTGCCTACAGGGGCGT
450	DUX4_gRNA	TCTGCTTACTTGGGGGATTGTGAC
451	DUX4_gRNA	TCTGCCTGCAGGGACATTTTGAGA
452	DUX4_gRNA	TCTGTCTACTGGGAGCATTGTGAC
453	DUX4_gRNA	TCTAGGCTCTGCTTACTTGGGGGA
454	DUX4_gRNA	TCTAGGCTCCGCCCACAGGGGGCA
455	DUX4_gRNA	TCTACACGAGAATTTTAACATATC
456	DUX4_gRNA	TCTAGGCTCTGTCTACTGGGAGCA
457	DUX4_gRNA	TCTGTCTACACGAGAATTTTAACA

458	DUX4_gRNA	TCTGCCTACTGGGGCGTAGTGACA
459	DUX4_gRNA	TCTAGACTCGACCTACAGGGGCTT
460	DUX4_gRNA	TCTAGGCTCTGCTTACGGGGGTAT
461	DUX4_gRNA	TCTAGGCCCTGCCTACAAGGGAAT
462	DUX4_gRNA	TCTGCCTACAGGGGACATTGTGAC
463	DUX4_gRNA	TCTAGGTTCAGACTACAGGAGCGT
464	DUX4_gRNA	TCTGTACGGATCACCTGGGTTATG
465	DUX4_gRNA	TCTGCCTATGGGGGCATTGCAACA
466	DUX4_gRNA	TCTGCCTACAGAGGGCATTGTGGC
467	DUX4_gRNA	TCTGTCACAATGCCCTGTAGGCA
468	DUX4_gRNA	TCTAGGTGATGTAACCTTGCTTA
469	DUX4_gRNA	TCTGCTTACGGGGGTATTGTGACA
470	DUX4_gRNA	TCTGCACTGATCAGCCCAGGGAGG
471	DUX4_gRNA	TCTGCCTACATGGGCATTCTGACA
472	DUX4_gRNA	TCTGTGTATGGGGGCTTTCTGACA
473	DUX4_gRNA	TCTGCCTATGGGGGCACTGTGATA
474	DUX4_gRNA	TCTGCCTACTGGAGCATTGTGACA
475	DUX4_gRNA	TCTAAGCTGTGCCTAAAGGGGAAT
476	DUX4_gRNA	TCTAGGCTCTGTCTACACGAGAAT
477	DUX4_gRNA	TCTGCCTACATAGGCATTGTGACA
478	DUX4_gRNA	TCTGTCTACGGGGGCATTGTGACA
479	DUX4_gRNA	TCTGCCTACTGGGGGCATTGTTAC
480	DUX4_gRNA	TCTAGGCTCTGCCTACTGGCGGCA
481	DUX4_gRNA	TCTAAGACAAGCGTCCATCACCTG
482	DUX4_gRNA	TCTGCAAAGATCACCCAGATGATG
483	DUX4_gRNA	TCTGCCCTTATGACCCAGGTGATG
484	DUX4_gRNA	TCTATGCACTGATCTCTGAGGTGA
485	DUX4_gRNA	TCTGAGGTGATTCAACTCTTGTCT
486	DUX4_gRNA	TCTGCCTACTGGCGGCATTGTCAC
487	DUX4_gRNA	TCTAGGCTCTGCCTATGGGGGCAC
488	DUX4_gRNA	TCTGCACTGATAACCTAGGTGATG
489	DUX4_gRNA	TCTAGGCTCTGCCTACATAGGCAT
490	DUX4_gRNA	TCTAGGCAGAGTATAGAGAAGAGT
491	DUX4_gRNA	TCTAGGCTCTGCCTACAGAGGGCA
492	DUX4_gRNA	TCTGAACATAATCATCCAGGAGATG
493	DUX4_gRNA	TCTGCCTACAGAGGGCGTTGTGAC
494	DUX4_gRNA	TCTAGGCCCCACCTACAGGGGGTA
495	DUX4_gRNA	TCTACAGGGGGCTTTGTGATATAT
496	DUX4_gRNA	TCTGCCTACAGGGGGCGTTGTGAA
497	DUX4_gRNA	TCTGCCTACAGGAGGCATTGTGAC
498	DUX4_gRNA	TCTATATCTGCCTACTGGCGGCAT
499	DUX4_gRNA	TCTGCACTGATCACCTGAGGAGG
500	DUX4_gRNA	TCTGCCTAAGGGGGCATTGTGACG
501	DUX4_gRNA	TCTAAGCTCTGCCTACAGGAGCTT
502	DUX4_gRNA	TCTAGGCTCTGCCTACACGGGAAT
503	DUX4_gRNA	TCTGCCTACAGGGGCATTGTGACG
504	DUX4_gRNA	TCTGCCTATGGGGGCATTGCGACA

496	DUX4_gRNA	TCTACGCTCTGCCTATGGGGGCAT
497	DUX4_gRNA	TCTGCCTACGGGGCATAGTGACAT
498	DUX4_gRNA	TCTAGGCTCTGTGTATGGGGGCTT
499	DUX4_gRNA	TCTGCCTATGGGGGCTTTGTGACA
500	DUX4_gRNA	TCTGTGTATGGGGGCTTTGTGACA
501	DUX4_gRNA	TCTAGGCTCTGCCAAAAGGGGGCA
502	DUX4_gRNA	TCTGCTTACAGGGTGCTTTGTGAC
503	DUX4_gRNA	TCTACGCTCTGCCTACAGGAGGCT
504	DUX4_gRNA	TCTAGTCTCTGCCTACAGAGGGCG
505	DUX4_gRNA	TCTGGCTACACAGCATTGTGACAT
506	DUX4_gRNA	TCTAGGCTCTGGCTACACAGCATT
507	DUX4_gRNA	TCTGCCTATAGGGGGCCTTGTGAC
508	DUX4_gRNA	TCTAGGCTCTGCCTACTGGAGCAT
509	DUX4_gRNA	TCTGCCTAGAGGGGGATTGTGAC
510	DUX4_gRNA	TCTGCCTACAGGGGCATTGTGATA
511	DUX4_gRNA	TCTGTAGGCAGAGACTAGAAAAGA
512	DUX4_gRNA	TCTGCCTACAGGGGCATTGCGATG
513	DUX4_gRNA	TCTGCCTGCAGGGGCATTGTGAAA
514	DUX4_gRNA	TCTGCCTACATGGGCATTGTGACA
515	DUX4_gRNA	TCTGCCTACAGGGGGTATTGTGAA
516	DUX4_gRNA	TCTAGGCTCTGCCTATGGGGGCTT
517	DUX4_gRNA	TCTGCCAAAAGGGGGCATTGTGAC
518	DUX4_gRNA	TCTACAGGGATTTTTGTGACATAT
519	DUX4_gRNA	TCTATACTCTGCCTAGAGGGGGAT
520	DUX4_gRNA	TCTATGGGGGCATTGTGTCAAATA
521	DUX4_gRNA	TCTGAACTGATCAACAAAGTGATG
522	DUX4_gRNA	TCTAGGCTCTGCCTACAGGGGGCA
523	DUX4_gRNA	TCTGCCTACTGGAGACATTGTGAC
524	DUX4_gRNA	TCTGCCTACTGGCGGCATTGTGGC
525	DUX4_gRNA	TCTGCTTACAGGGGGCATTGTGAC
526	DUX4_gRNA	TCTGCCTATAGGGGCATTGTGACA
527	DUX4_gRNA	TCTAGGATCTGCCTAAAGGGACTT
528	DUX4_gRNA	TCTGTCTACAGGGATTTTTGTGAC
529	DUX4_gRNA	TCTAGGCTCTGCCTACTGGGGGCA
530	DUX4_gRNA	TCTGCACAGATCATCTAGGTGATG
531	DUX4_gRNA	TCTACAGGGGGCTTTGTGACATAT
532	DUX4_gRNA	TCTAGGCTCTGTCTACGGGGGCAT
533	DUX4_gRNA	TCTAGGCTCTGTCTACAGGGATTT
534	DUX4_gRNA	TCTGTGCAGAGCTATGTCAAACG
535	DUX4_gRNA	TCTGCACTGATGACCCAGATGATG
536	DUX4_gRNA	TCTGCTTACAGGGGGTATTGTGAC
537	DUX4_gRNA	TCTGCCTACACGGGAATTCTCACA
538	DUX4_gRNA	TCTGCCTACAGGGGCGTTTTGACA
539	DUX4_gRNA	TCTGCACTAATCATCCAGGTGATG
540	DUX4_gRNA	TCTATGCTCTGCCTACAGGGGGCA
541	DUX4_gRNA	TCTAGGCACTGCCTACAGGGGACA
542	DUX4_gRNA	TCTGTGCTCTGCCTACAGGGGACA

543	DUX4_gRNA	TCTGTCTATGGGGGCATTGTGTCA
544	DUX4_gRNA	TCTGCACTGTTAACCGAGGTGATG
545	DUX4_gRNA	TCTGCCTACAGGGGGCATTGTGAA
546	DUX4_gRNA	TCTGCAGTGATCACGCAGGTGATG
547	DUX4_gRNA	TCTGCCTAAAGGGACTTTGTGACA
548	DUX4_gRNA	TCTGCCTACAGGAGGCTTTATGAC
549	DUX4_gRNA	TCTGCCTATGGGGGCATAGTGACA
550	DUX4_gRNA	TCTGTAGGCAAAGCCCATACAAGG
551	DUX4_gRNA	TCTGCACTGATCACCTAGGTCATA
552	DUX4_gRNA	TCTGCCTACAGGGGGCTTGTGACA
553	DUX4_gRNA	TCTAGGCTCTGCCTACTGGAGACA
554	DUX4_gRNA	TCTGCCTACAGGGGGCATTGTGAC
555	DUX4_gRNA	TCTGCACTGATCCCCAAGGTGATG
556	DUX4_gRNA	TCTGCACTGATCAACTAGGTGATG
557	DUX4_gRNA	TCTAGGCTCTGCTTACAGGGGGTA
558	DUX4_gRNA	TCTGCTTAAAGGGGCCTTGTGACA
559	DUX4_gRNA	TCTGAACTGATCAACCAAGTGATG
560	DUX4_gRNA	TCTGCCTAAAGGGGCATTGTGACA
561	DUX4_gRNA	TCTGCCTACTGGGGACATTGTGAC
562	DUX4_gRNA	TCTGCCTACAGGGGCGTTTTGACA
563	DUX4_gRNA	TCTGCACTGATCCCGAGGTGATCC
564	DUX4_gRNA	TCTGCCTACAGTGGCATTGTGACA
565	DUX4_gRNA	TCTGCAATGATCACCCAGGTGATG
566	DUX4_gRNA	TCTGCCCTGATCACCCAGGTGATG
567	DUX4_gRNA	TCTGCCTACAGGGGCATTGCAATG
568	DUX4_gRNA	TCTAGGCTGTGCCACAGGGGGAT
569	DUX4_gRNA	TCTAGGCTCTGCCTACAGGGGCTT
570	DUX4_gRNA	TCTAGGCTCTGCTTAAAGGGGCCT
571	DUX4_gRNA	TCTGCACTGATCACTCAGGTGATG
572	DUX4_gRNA	TCTGCCTATGGGGGCATTGTGACA
573	DUX4_gRNA	TCTAAGCTCTGCCTAAAGGGGCAT
574	DUX4_gRNA	TCTGTCACAATGCCCTTTAGGCA
575	DUX4_gRNA	TCTAGGCTCTGCCTAAGGGGGCAT
576	DUX4_gRNA	TCTGCACTGATAACCCAGGTGATG
577	DUX4_gRNA	TCTGCACTGATCATCTAGGTGATG
578	DUX4_gRNA	TCTGCCTACAGGGGAATTGTGAGA
579	DUX4_gRNA	TCTAAGCTCTGCCTACAGGGGCAT
580	DUX4_gRNA	TCTGCCTACAGGGTGCTTTGTGAC
581	DUX4_gRNA	TCTAGTCTAAGCTCTGCCTAAAGG
582	DUX4_gRNA	TCTGCACTGATCACCGAAGTTATG
583	DUX4_gRNA	TCTGCACTGATCTCCCAGGTGCTG
584	DUX4_gRNA	TCTGGGATTTGTCTACAGGGGGCT
585	DUX4_gRNA	TCTGCCTACAGGGGCTTTGTGACA
586	DUX4_gRNA	TCTGCCTACAGGAGCTTTGTGACA
587	DUX4_gRNA	TCTGCACTGATCACCCAGGAGACG
588	DUX4_gRNA	TCTGCCTACAGGGGCATTGTGACA
589	DUX4_gRNA	TCTAGGCTCTGCCTACAGGGGGCT

590	DUX4_gRNA	TCTGCACTGATCACCTAGGTCATG
591	DUX4_gRNA	TCTGCACTGATCACTTAGGTGATG
592	DUX4_gRNA	TCTAGGATCTGCCTACAGGGGGTA
593	DUX4_gRNA	TCTGCACTGATCGCCCAGATGATG
594	DUX4_gRNA	TCTAGGATCTGCCTACAGGGTGCT
595	DUX4_gRNA	TCTGCACTGATCACCCAAGTAATG
596	DUX4_gRNA	TCTAGGCTCTGCCTACAGTGGCAT
597	DUX4_gRNA	TCTAGGCTCTGCCTACAGGGGCGT
598	DUX4_gRNA	TCTGGAGTAGCTGGGACTACAGTC
599	DUX4_gRNA	TCTGGGATCTGCTTACAGGGGGCA
600	DUX4_gRNA	TCTAGGATCTGCTTACAGGGTGCT
601	DUX4_gRNA	TCTGCACTGATCACCTTGGTGATG
602	DUX4_gRNA	TCTGCACTGATCACCCAGGTGACT
602	DUX4_gRNA	TCTGGGCTCTGCCTACAGGGGCAT
604	DUX4_gRNA	TCTAGGCTCTGCCTACAGGGGCAT
605	DUX4_gRNA	TCTGTTGCCCCGGGCTGGAATGCAG
606	DUX4_gRNA	TCTGCACTGATCACCTAGGTGATG
607	DUX4_gRNA	TCTGTACTGATCACCCAGGTGATG
608	DUX4_gRNA	TCTGGGCTTTGTCTACAGGGGGCT
609	DUX4_gRNA	TCTACACTGATCACCTAAGTGATG
610	DUX4_gRNA	TCTACACTGATCACACAGGTGATG
611	DUX4_gRNA	TCTGCACTGATCACCTAAGTGATG
612	DUX4_gRNA	TCTGCACTGATCACCCAGGTGAAG
613	DUX4_gRNA	TCTGCACAGATCACCCAGGTGATG
614	DUX4_gRNA	TCTGCACTGATCACCGAGGTGATG
615	DUX4_gRNA	TCTGCACTGATCACCCAGGTGGTG
617	DUX4_gRNA	TCTGCACTGATCACCCAGGGGATG
618	DUX4_gRNA	TCTGCACTGATCACCCAGGTAATG
619	DUX4_gRNA	TCTGCACTGATCACCCAGGTGATA
620	DUX4_gRNA	TCTGCACTGATCAACCAGGTGATG
621	DUX4_gRNA	TCTGCACTGATCACCCAGGTCATG
622	DUX4_gRNA	TCTGCACTGATCACCCAGGCGATG
623	DUX4_gRNA	TCTGCACTGATCACCCAAGTGATG
624	DUX4_gRNA	TCTACTAAAAATACAAAAAATTA
625	DUX4_gRNA	TCTGCACTGATCACCCAGGTGATG
626	DUX4_gRNA	TCTACACTGATCACCCAGGTGATG
627	DUX4_gRNA	TCTACCTCAGATGAGATATTGCTT
628	DUX4_gRNA	TCTGTCTCGGAATGAAATGAATTC
629	DUX4_gRNA	TCTATAGTTCAAACAAAGATGAGG
630	DUX4_gRNA	TCTGAGGTAGAATGTTTCTAGTGG
631	DUX4_gRNA	TCTGCTGTATCTATAGTTCAAACA
632	DUX4_gRNA	TCTATACTGCTTGACCCAAGCTTT
633	DUX4_gRNA	TCTAGCGTGTATTTATTTTCAGC
634	DUX4_gRNA	TCTGAAACGTGGTATCTGGAGAGG
635	DUX4_gRNA	TCTACCTTTTGCTATCAAAAGCTT
636	DUX4_gRNA	TCTAGGAACAGTAAGAGGACCTTG
637	DUX4_gRNA	TCTGGAGAATTCATTTCAATCCGA

638	DUX4_gRNA	TCTGCTTATTACCCACTCTGTAAT
639	DUX4_gRNA	TCTGAGGGAGAAAACTAATCTTT
640	DUX4_gRNA	TCTGTTACTGTGTGCAAGGTGAAG
641	DUX4_gRNA	TCTAGTGGTTGTGTTCTGAGGGAG
642	DUX4_gRNA	TCTGCTTTTGGTTCATGAAATTTT
643	DUX4_gRNA	TCTGGAGAGGTGAGATGGACAAAG
644	DUX4_gRNA	TCTGGGTCACAGCTATATTAGAGC
645	DUX4_gRNA	TCTGTTTCTAGCGTGTATTTATTT
646	DUX4_gRNA	TCTAATATAGCTGTGACCCAGATG
647	DUX4_gRNA	TCTGCTTCACTTCAATAACAGCCT
648	DUX4_gRNA	TCTGGAACAGCTATGTACTTTCTT
649	DUX4_gRNA	TCTGAAATCCTTTTATGCCTGGCC
650	DUX4_gRNA	TCTGTAATGTGGAAACAAATTATT
651	DUX4_gRNA	TCTGACACAGTCTGCGTTTGTAAG
652	DUX4_gRNA	TCTGGGATTCTTCTGCTGGAAAAA
653	DUX4_gRNA	TCTAAGAAGTCTGGGATTCTTCTG
654	DUX4_gRNA	TCTACCATTTAAAACAAGAACTCT
655	DUX4_gRNA	TCTGCTGGAAAAATAAGTTTGTTG
656	DUX4_gRNA	TCTGTGAAATCCTCATGTTTTCTT
657	DUX4_gRNA	TCTAAAGTATATTACTCTGCTTTT
658	DUX4_gRNA	TCTAACCTTCAAAAACCAAACCTG
659	DUX4_gRNA	TCTGGCTACTTTTATGGTATAATG
660	DUX4_gRNA	TCTATCTGTTTACTATCTGTCTTT
661	DUX4_gRNA	TCTGTTTACTATCTGTCTTTTCTA
662	DUX4_gRNA	TCTAAAACAAGGTGTGGCAAACCTA
663	DUX4_gRNA	TCTGTTTTCTGGAACAGCTATGTA
664	DUX4_gRNA	TCTGTCTTTTCTACCTTTTGCTAT
665	DUX4_gRNA	TCTAGTTTTGCCTCATCTTTGTTT
666	DUX4_gRNA	TCTGCGTTTGTAAGTAAAGTTGTA
667	DUX4_gRNA	TCTGCAAAGGGCTAAATGTTAAAT
668	DUX4_gRNA	TCTATCTATCTGTTTACTATCTGT
669	DUX4_gRNA	TCTGTGGGGTTTTTGTGTTGTTG
670	DUX4_gRNA	TCTACCTCCTATCATCTATCTATC
671	DUX4_gRNA	TCTATCTACCTCCTATCATCTATC
672	DUX4_gRNA	TCTATCTATCTACCTCCTATCATC
673	DUX4_gRNA	TCTGTGGCCAGGCGTGGTGGCTCA
674	DUX4_gRNA	TCTGTTTTTTGTTTGTTGTTGTT
675	DUX4_gRNA	TCTGTAATCCCAGCACTTTGGGAT
676	DUX4_gRNA	TCGAACTCACAGGCAAAATCCTCC
677	DUX4_gRNA	TCGGTATCCCCCTTTACTGAGCCA
678	DUX4_gRNA	TCGAATGCACTTTAAGATTCTGGG
679	DUX4_gRNA	TCGGGTCTTACCCGCGCGGTTCA
680	DUX4_gRNA	TCGGGTGGTTCGGGGCAGGGCCGT
681	DUX4_gRNA	TCGGGTTTTTACCCGCGCGGTTCA
682	DUX4_gRNA	TCGGCCTCGCGCCGCGTTGCAGGG
683	DUX4_gRNA	TCGGGTTGCCGTGGGTCTTCACC
684	DUX4_gRNA	TCGGCATGGCCAGCCTTTCGGGGG

685	DUX4_gRNA	TCGGCAGCAGGGAGAAACCAGCCT
686	DUX4_gRNA	TCGGGTGGTTCGGGGCAGGGCGGT
687	DUX4_gRNA	TCGGCCTCCGGGAGTAGCGGGACC
688	DUX4_gRNA	TCGGAAGAGGCCCTCGCTGGAA
689	DUX4_gRNA	TCGAGGCCTGGGGCCGGCCGGCGG
690	DUX4_gRNA	TCGGGTTGCCGTCGGGTTTTCACC
691	DUX4_gRNA	TCGGGGGCCGGAGAGACGTGAGCA
692	DUX4_gRNA	TCGGGGGCCGGCTCTCCGGACCTC
692	DUX4_gRNA	TCGGTGGCCTCCGCACCCGGGCAA
694	DUX4_gRNA	TCGACGCCCTGGGTCCCTTCCGGG
695	DUX4_gRNA	TCGGACAGCACCTCCCCGCGGAA
696	DUX4_gRNA	TCGGGAGGGCCATCGCGGTGAGCC
697	DUX4_gRNA	TCGGGGCAGGGCCGTGGCCTCTCT
698	DUX4_gRNA	TCGGGGTCCAAACGAGTCTCCGTC
699	DUX4_gRNA	TCGGCCCTGGCCCGGGAGACGCGG
700	DUX4_gRNA	TCGGCATTCCGGAGCCCAGGGTCC
701	DUX4_gRNA	TCGGAGGAGCAGGGCGGTCTGGGA
702	DUX4_gRNA	TCGGGGCAGGGCGGTGGCCTCTCT
703	DUX4_gRNA	TCGAAGGGCCAGGCACCCGGGACA
704	DUX4_gRNA	TCGATTCTGAAACCAGATCTGAAT
705	DUX4_gRNA	TCGGAAGGTGGGGGGAGACATTCA
706	DUX4_gRNA	TCGAGTCTAGACAAGAGTTACATC
707	DUX4_gRNA	TCGGGTTCAGGTTAAGAGTTAGGG
708	DUX4_gRNA	TCGGTGATCAGTGCAGAGATACGT
709	DUX4_gRNA	TCGACAAATCTCTGCACTGATCAC
710	DUX4_gRNA	TCGAAATTCCTGTAGGCAGTGCT
711	DUX4_gRNA	TCGGTGATCAGTGCAGATGTGTTT
712	DUX4_gRNA	TCGACCTACAGGGGCTTTGTGACA
713	DUX4_gRNA	TCGGTGATCAATGCAGCGATATGT
714	DUX4_gRNA	TCGGTTAACAGTGCAGAGATATGT
715	DUX4_gRNA	TCGGGATCAGTGCAGAGATATGTC
716	DUX4_gRNA	TCGGAATGAAATGAATTCTCCAGA
717	DUX4_gRNA	TCGGCTAGCCTCGGCATCCCAAAG
718	DUX4_gRNA	TCGGCATCCCAAAGTGCTGGGATT

[0283] It shall be understood that different aspects of the invention can be appreciated individually, collectively, or in combination with each other. Various aspects of the invention described herein may be applied to any of the particular applications disclosed herein. The compositions of matter disclosed herein in the composition section of the present disclosure may be utilized in the method section including methods of use and production disclosed herein, or vice versa.

[0284] While preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. It is not intended that the invention be limited by the

specific examples provided within the specification. While the invention has been described with reference to the aforementioned specification, the descriptions and illustrations of the embodiments herein are not meant to be construed in a limiting sense. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. Furthermore, it shall be understood that all aspects of the invention are not limited to the specific depictions, configurations or relative proportions set forth herein which depend upon a variety of conditions and variables. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is therefore contemplated that the invention shall also cover any such alternatives, modifications, variations or equivalents. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

CLAIMS

WHAT IS CLAIMED IS:

1. A system for regulating aberrant expression of a target gene in a muscle cell, comprising:
a heterologous polypeptide comprising a nuclease, wherein the nuclease has a length that is less than or equal to about 900 amino acids; and
a guide nucleic acid molecule configured to form a complex with the heterologous polypeptide, wherein the guide nucleic acid molecule exhibits specific binding to a target polynucleotide sequence at or adjacent to, a D4Z4 repeat array in the muscle cell,
wherein, upon formation of the complex, the complex is capable of binding the target polynucleotide sequence, to effect modification of an expression level and/or a methylation level of the target gene in the muscle cell, wherein the target gene is within the D4Z4 repeat array.
2. The system of claim 1, wherein upon formation of the complex, the modified expression level and/or methylation level of the target gene in the muscle cell is sustained for at least about 2 days.
3. The system of claim 2, wherein the modified expression level and/or methylation level of the target gene is sustained for at least about 3 days, 4 days, 5 days, 6 days, 1 week, 2 weeks, 2 weeks, 4 weeks, or 2 months.
4. The system of claim 2, wherein the modified expression level and/or methylation level of the target gene is sustained for at least about 17 days.
5. The system of claim 2, wherein the modified expression level and/or methylation level of the target gene is sustained for at least about 18 days.
6. The system of claim 1, wherein the muscle cell is in a subject having or is suspected of having facioscapulohumeral muscular dystrophy (FSHD).
7. The system of claim 1, wherein the target gene is Dux4.
8. The system of claim 1, wherein the nuclease has a length that is less than or equal to about 800 amino acids.
9. The system of claim 1, wherein the nuclease has a length that is less than or equal to about 750 amino acids.
10. The system of claim 1, wherein the nuclease is Un1Cas12f1 or a modified variant thereof.
11. The system of claim 1, wherein the nuclease comprises an amino acid sequence that is at least about 80%, at least about 90%, at least about 95%, or at least about 99% identical to the polypeptide sequence of SEQ ID NO: 43.

12. The system of claim 1, wherein the nuclease comprises an amino acid sequence that is at least about 80%, at least about 90%, at least about 95%, or at least about 99% identical to the polypeptide sequence of SEQ ID NO: 44.
13. The system of claim 1, wherein the heterologous polypeptide further comprises a transcriptional regulator.
14. The system of claim 13, wherein the transcriptional regulator comprises at least one methyltransferases.
15. The system of claim 14, wherein the transcriptional regulator comprises at least one DNA Methyltransferases (DNMT).
16. The system of claim 15, wherein the transcriptional regulator comprises DNMT-A or DNMT-L.
17. The system of claim 14, wherein the transcriptional regulator comprises (i) DNMT-A or DNMT-L and (ii) KRAB or a variant of KRAB.
18. The system of claim 14, wherein the transcriptional regulator comprises (i) DNMT-L and (ii) KRAB or a variant of KRAB.
19. The system of claim 14, wherein the transcriptional regulator comprises DNMT-A, DNMT-L, and KRAB or a variant of KRAB.
20. The system of claim 13, wherein the transcriptional regulator comprises DNMT-L or KRAB or variant of KRAB.
21. The system of claim 13, wherein the transcriptional regulator comprises KRAB or a variant of KRAB.
22. The system of claim 13, wherein the transcriptional regulator comprises a plurality of different transcriptional regulators.
23. The system of claim 1, wherein the modification of the expression level and/or the methylation level of the target gene effects downregulation of a downstream gene of the target gene, wherein the downstream gene comprises one or more members selected from the group consisting of ZSCAN4, LEUTX, MBD3L2, TRIM48, and TRIM43.
24. The system of claim 1, wherein the modification of the expression level and/or the methylation level of the target gene effects downregulation of an apoptosis marker in the muscle cell.
25. The system of claim 24, wherein the apoptosis marker comprises Caspase 3.
26. The system of claim 1, wherein the complex effects the modification of the expression level of the target gene in the muscle gene.

27. The system of claim 26, wherein the modification of the expression level results in downregulation of the target gene.
28. The system of claim 1, wherein the complex effects the modification of the methylation level of the target gene in the muscle gene.
29. The system of claim 28, wherein the modification of the methylation level results in downregulation of the target gene.
30. The system of claim 1, wherein the nuclease is a deactivated nuclease.
31. A composition comprising the system of any one of the preceding claims.
32. A viral vector comprising the system of any one of the preceding claims.
33. The viral vector of claim 32, wherein the viral vector comprises an adeno-associated virus (AAVs), a retrovirus, a lentivirus, a poxvirus, or an adenovirus.
34. The viral vector of claim 33, wherein the AAV comprises a AAV serotype RH74 AAV.
35. A method for regulating aberrant expression of a target gene in a muscle cell, comprising:
 - (a) contacting the muscle cell with a complex comprising (i) a heterologous polypeptide comprising a nuclease, wherein the nuclease has a length that is less than or equal to about 900 amino acids and (ii) a guide nucleic acid molecule exhibiting specific binding to a target polynucleotide sequence at or adjacent to, a D4Z4 repeat array in the muscle cell; and
 - (b) upon the contacting, binding the target gene with the complex to effect modification of an expression level and/or a methylation level of the target gene in the muscle cell, wherein the target gene is within the D4Z4 repeat array.
36. The method of claim 35, wherein upon formation of the complex, the modified expression level and/or methylation level of the target gene in the muscle cell is sustained for at least about 2 days.
37. The method of claim 36, wherein the modified expression level and/or methylation level of the target gene is sustained for at least about 3 days, 4 days, 5 days, 6 days, 1 week, 2 weeks, 2 weeks, 4 weeks, or 2 months.
38. The method of claim 36, wherein the modified expression level and/or methylation level of the target gene is sustained for at least about 17 days.
39. The method of claim 36, wherein the modified expression level and/or methylation level of the target gene is sustained for at least about 18 days.
40. The method of claim 35, wherein the contacting comprises injecting a composition comprising the complex to a subject in need thereof, wherein the subject has or is suspected of having facioscapulohumeral muscular dystrophy (FSHD).
41. The method of claim 35, wherein the target gene is Dux4.

42. The method of claim 35, wherein the nuclease has a length that is less than or equal to about 800 amino acids.
43. The method of claim 35, wherein the nuclease has a length that is less than or equal to about 750 amino acids.
44. The method of claim 35, wherein the nuclease is UnlCas12f1 or a modified variant thereof.
45. The method of claim 35, wherein the nuclease comprises an amino acid sequence that is at least about 80%, at least about 90%, at least about 95%, or at least about 99% identical to the polypeptide sequence of SEQ ID NO: 43.
46. The method of claim 35, wherein the nuclease comprises an amino acid sequence that is at least about 80%, at least about 90%, at least about 95%, or at least about 99% identical to the polypeptide sequence of SEQ ID NO: 44.
47. The method of claim 35, wherein the heterologous polypeptide further comprises a transcriptional regulator.
48. The method of claim 47, wherein the transcriptional regulator comprises at least one methyltransferases.
49. The method of claim 48, wherein the transcriptional regulator comprises at least one DNA Methyltransferases (DNMT).
50. The method of claim 49, wherein the transcriptional regulator comprises DNMT-A or DNMT-L.
51. The method of claim 49, wherein the transcriptional regulator comprises (i) DNMT-A or DNMT-L and (ii) KRAB or a variant of KRAB.
52. The method of claim 49, wherein the transcriptional regulator comprises (i) DNMT-L and (ii) KRAB or a variant of KRAB.
53. The method of claim 49, wherein the transcriptional regulator comprises DNMT-A, DNMT-L, and KRAB or variant of KRAB.
54. The method of claim 47, wherein the transcriptional regulator comprises DNMT-L or KRAB or a variant of KRAB.
55. The method of claim 47, wherein the transcriptional regulator comprises KRAB or a variant of KRAB.
56. The method of claim 47, wherein the transcriptional regulator comprises a plurality of different transcriptional regulators.
57. The method of claim 35, wherein the modification of the expression level and/or the methylation level of the target gene effects downregulation of a downstream gene of the target

gene, wherein the downstream gene comprises one or more members selected from the group consisting of ZSCAN4, LEUTX, MBD3L2, TRIM48, and TRIM43.

58. The method of claim 35, wherein the modification of the expression level and/or the methylation level of the target gene effects downregulation of an apoptosis marker in the muscle cell.

59. The method of claim 58, wherein the apoptosis marker comprises Caspase 3.

60. The method of claim 35, wherein the complex effects the modification of the expression level of the target gene in the muscle gene.

61. The method of claim 35, wherein the modification of the expression level results in downregulation of the target gene.

62. The method of claim 35, wherein the complex effects the modification of the methylation level of the target gene in the muscle gene.

63. The method of claim 62, wherein the modification of the methylation level results in downregulation of the target gene.

64. The method of claim 35, wherein the nuclease is a deactivated nuclease.

65. A system for regulating aberrant expression of a target gene in a muscle cell, the system comprising:

a heterologous actuator moiety coupled to a gene regulator, wherein the heterologous actuator moiety is capable of forming a complex with the target gene in the muscle cell, and wherein the gene regulator is capable of modifying an expression level and/or a methylation level of the target gene in the muscle cell,

wherein, upon formation of the complex, the modified expression level and/or methylation level of the target gene in the muscle cell is sustained for at least about 2 days.

66. The system of claim 65, wherein the sustained modified expression level and/or the methylation level of the target gene is characterized by maintaining at least about 70%, 75%, 80%, 85%, 90%, 95%, 99%, or 100% of the modified expression level and/or methylation level of the target gene.

67. The system of any one of the preceding claims, wherein the modified expression level and/or methylation level of the target gene is sustained for at least about 3 days, 4 days, 5 days, 6 days, 1 week, 2 weeks, 3 weeks, 4 weeks, or 2 months.

68. The system of any one of the preceding claims, wherein the modified expression level of the target gene is a decreased expression level of the target gene.

69. The system of any one of the preceding claims, wherein the modified methylation level of the target gene is an increased degree of methylation of the target gene.

70. The system of any one of the preceding claims, wherein the gene regulator comprises an epigenetic regulator.
71. The system of any one of the preceding claims, wherein the epigenetic regulator comprises a chromatin modifier.
72. The system of any one of the preceding claims, wherein the epigenetic regulator comprises at least one methyltransferases.
73. The system of any one of the preceding claims, wherein the epigenetic regulator comprises at least one DNA Methyltransferases (DNMT).
74. The system of any one of the preceding claims, wherein the epigenetic regulator comprises DNMT-A or DNMT-L.
75. The system of any one of the preceding claims, wherein the epigenetic regulator comprises (i) DNMT-A or DNMT-L and (ii) KRAB or a variant of KRAB.
76. The system of any one of the preceding claims, wherein the epigenetic regulator comprises DNMT-A, DNMT-L, and KRAB or a variant of KRAB.
77. The system of any one of the preceding claims, wherein the epigenetic regular comprises of KRAB or a variant of KRAB.
78. The system of any one of the preceding claims, wherein the gene regulator comprises a plurality of different gene regulators.
79. The system of any one of the preceding claims, further comprising a guide nucleic acid molecule capable of directing the heterologous actuator moiety to the target gene, to form the complex.
80. The system of any one of the preceding claims, wherein the heterologous actuator moiety is capable of forming a complex with a first portion of the muscle-regulating gene, and wherein the system further comprises an additional heterologous actuator moiety coupled to an additional gene regulator, wherein the additional heterologous actuator moiety is capable of forming a complex with a second portion of the muscle-regulating gene.
81. The system of any one of the preceding claims, further comprising an additional guide nucleic acid molecule capable of directing the additional heterologous actuator moiety to the second portion of the muscle-regulating gene.
82. The system of any one of the preceding claims, wherein the target gene is a transcription factor.
83. The system of any one of the preceding claims, wherein the target gene is within a D4Z4 repeat array.
84. The system of any one of the preceding claims, wherein the target gene encodes DUX4.

85. The system of any one of the preceding claims, wherein the target gene is not C9orf72.
86. The system of any one of the preceding claims, wherein the muscle cell is a skeletal muscle cell.
87. The system of any one of the preceding claims, wherein the heterologous actuator moiety or the additional heterologous actuator moiety comprises an endonuclease.
88. The system of any one of the preceding claims, wherein the heterologous actuator moiety or the additional heterologous actuator moiety comprises a CRISPR-Cas protein.
89. The system of any one of the preceding claims, wherein the heterologous actuator moiety or the additional heterologous actuator moiety comprises a dCas protein.
90. The system of any one of the preceding claims, wherein the guide nucleic acid molecule or the additional guide nucleic acid molecule comprises a guide RNA molecule.
91. A method for regulating aberrant expression of a target gene in a muscle cell, the method comprising:
- (a) contacting the muscle cell with a heterologous actuator moiety coupled to a gene regulator, wherein the heterologous actuator moiety is capable of forming a complex with the target gene in the muscle cell, and wherein the gene regulator is capable of modifying an expression level and/or a methylation level of the target gene in the muscle cell; and
 - (b) upon formation of the complex, sustaining the modified expression level and/or methylation level of the target gene in the muscle cell for at least about 2 days.
92. The method of claim 91, wherein the sustained modified expression level and/or the methylation level of the target gene is characterized by maintaining at least about 70%, 75%, 80%, 85%, 90%, 95%, 99%, or 100% of the modified expression level and/or methylation level of the target gene.
93. The method of any one of the preceding claims, wherein the modified expression level and/or methylation level of the target gene is sustained for at least about 3 days, 4 days, 5 days, 6 days, 1 week, 2 weeks, 3 weeks, 4 weeks, or 2 months.
94. The method of any one of the preceding claims, wherein the modified expression level of the target gene is a decreased expression level of the target gene.
95. The method of any one of the preceding claims, wherein the modified methylation level of the target gene is an increased degree of methylation of the target gene.
96. The method of any one of the preceding claims, wherein the gene regulator comprises an epigenetic regulator.
97. The method of any one of the preceding claims, wherein the epigenetic regulator comprises a chromatin modifier.

98. The method of any one of the preceding claims, wherein the epigenetic regulator comprises at least one methyltransferases.
99. The method of any one of the preceding claims, wherein the epigenetic regulator comprises at least one DNA Methyltransferases (DNMT).
100. The method of any one of the preceding claims, wherein the epigenetic regulator comprises DNMT-A or DNMT-L.
101. The method of any one of the preceding claims, wherein the epigenetic regulator comprises (i) DNMT-A or DNMT-L and (ii) KRAB or a variant of KRAB.
102. The method of any one of the preceding claims, wherein the epigenetic regulator comprises DNMT-A, DNMT-L, and KRAB or a variant of KRAB.
103. The method of any one of the preceding claims, wherein the epigenetic regulator comprises KRAB or a variant of KRAB.
103. The method of any one of the preceding claims, wherein the gene regulator comprises a plurality of different gene regulators.
104. The method of any one of the preceding claims, further comprising contacting the muscle cell with a guide nucleic acid molecule capable of directing the heterologous actuator moiety to the target gene, to form the complex.
105. The method of any one of the preceding claims, wherein the heterologous actuator moiety is capable of forming a complex with a first portion of the muscle-regulating gene, and wherein the method further comprises contacting the muscle cell with an additional heterologous actuator moiety coupled to an additional gene regulator, wherein the additional heterologous actuator moiety is capable of forming a complex with a second portion of the muscle-regulating gene.
106. The method of any one of the preceding claims, further comprising contacting the muscle cell with an additional guide nucleic acid molecule capable of directing the additional heterologous actuator moiety to the second portion of the muscle-regulating gene.
107. The method of any one of the preceding claims, wherein the target gene is a transcription factor.
108. The method of any one of the preceding claims, wherein the target gene is within a D4Z4 repeat array.
109. The method of any one of the preceding claims, wherein the target gene encodes DUX4.
110. The method of any one of the preceding claims, wherein the target gene is not C9orf72.
111. The method of any one of the preceding claims, wherein the muscle cell is a skeletal muscle cell.

112. The method of any one of the preceding claims, wherein the heterologous actuator moiety or the additional heterologous actuator moiety comprises an endonuclease.
113. The method of any one of the preceding claims, wherein the heterologous actuator moiety or the additional heterologous actuator moiety comprises a CRISPR-Cas protein.
114. The method of any one of the preceding claims, wherein the heterologous actuator moiety or the additional heterologous actuator moiety comprises a dCas protein.
115. The method of any one of the preceding claims, wherein the guide nucleic acid molecule or the additional guide nucleic acid molecule comprises a guide RNA molecule.

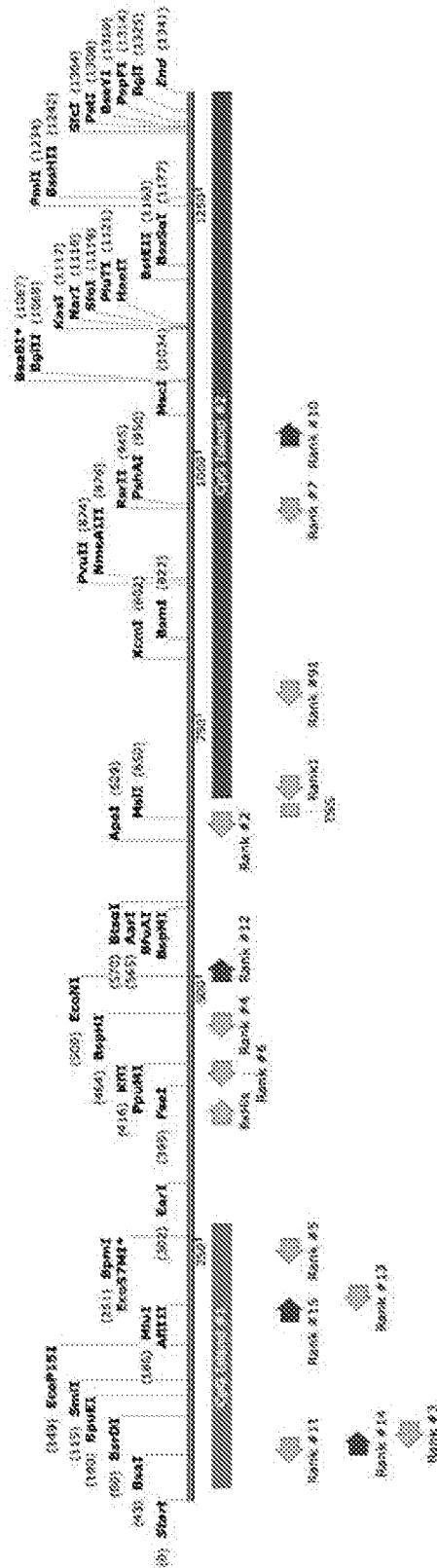


FIG. 1

DUX4 Repression by dCas9KAL

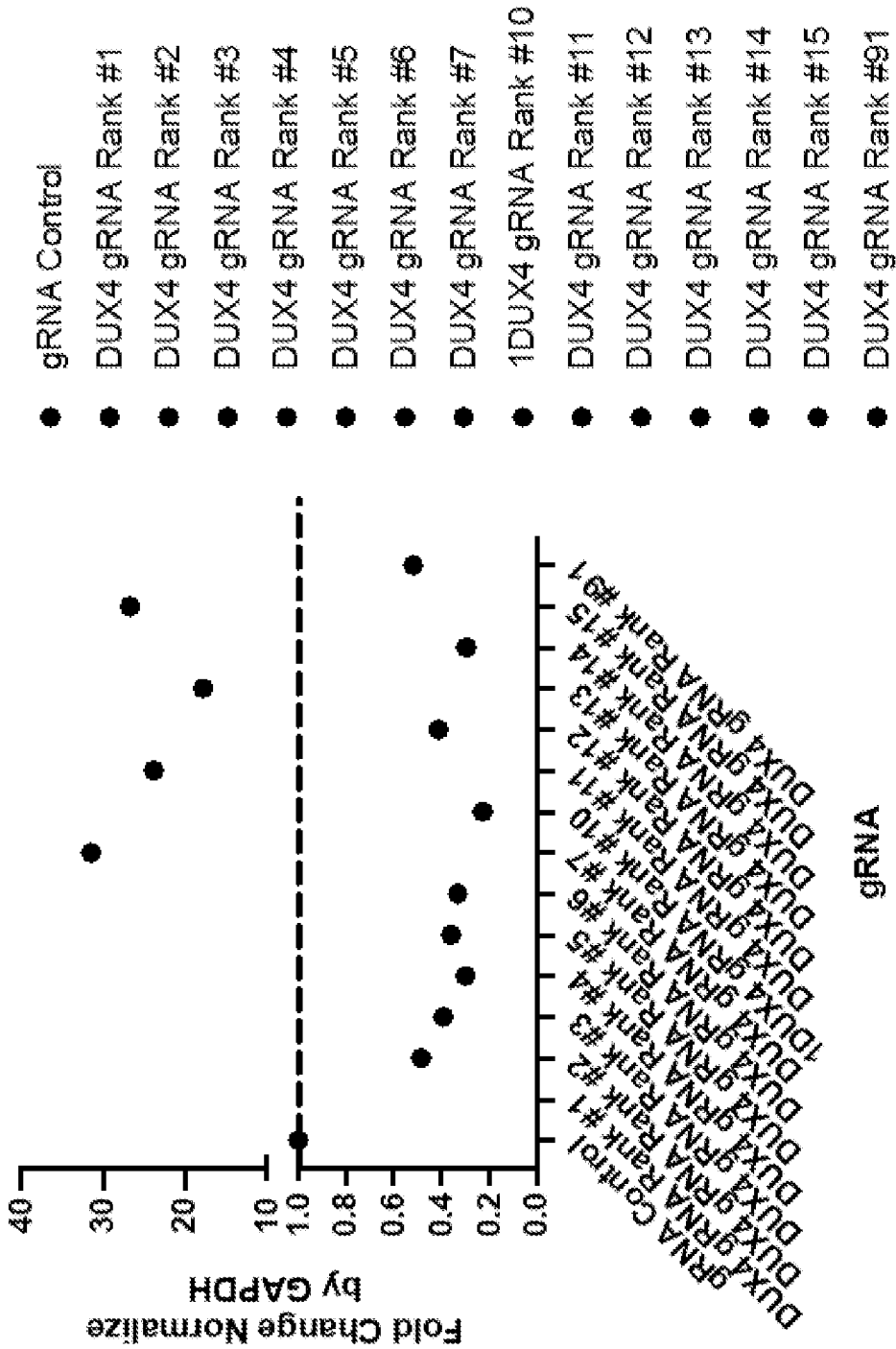


FIG. 2

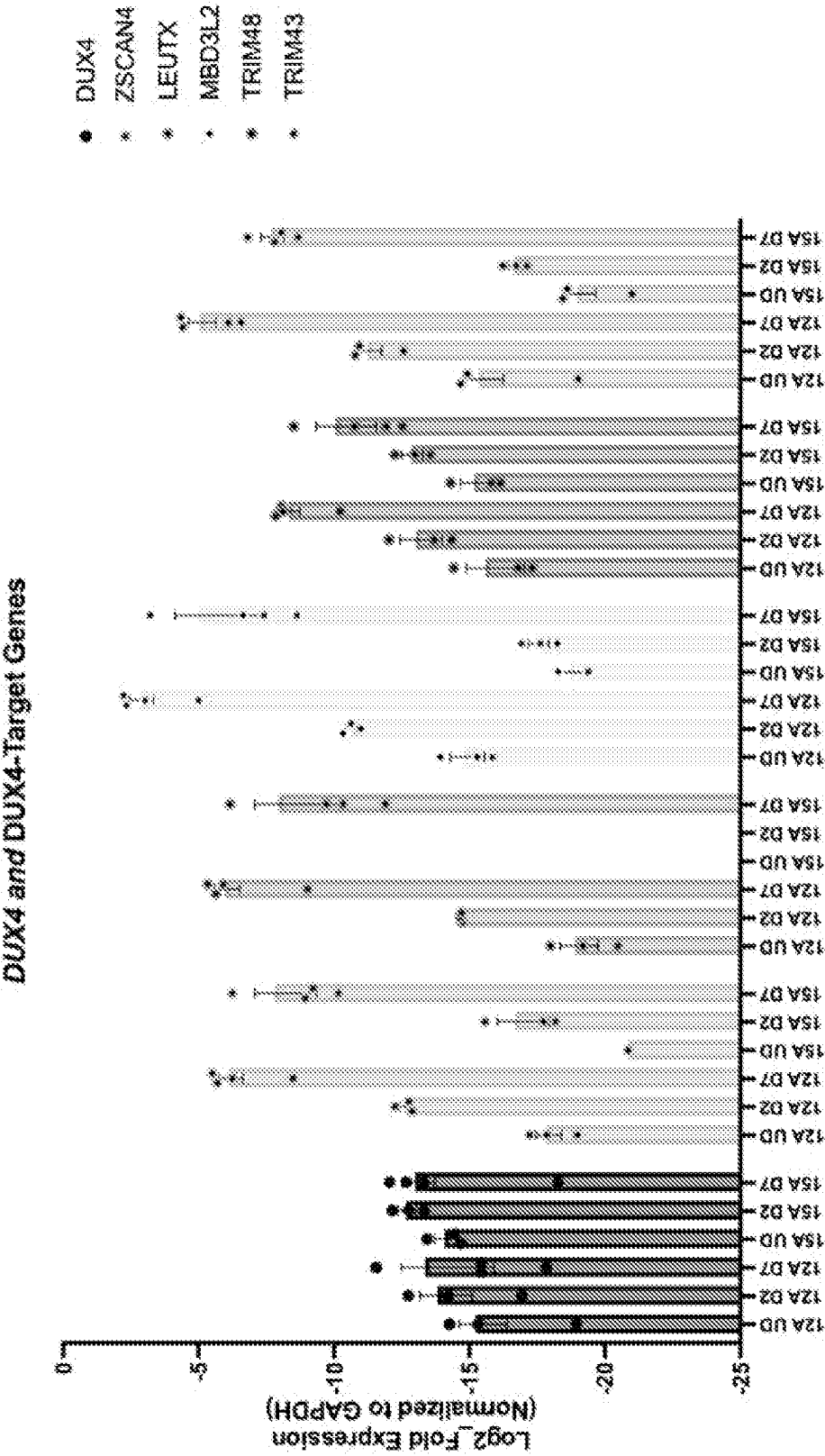


FIG. 3A

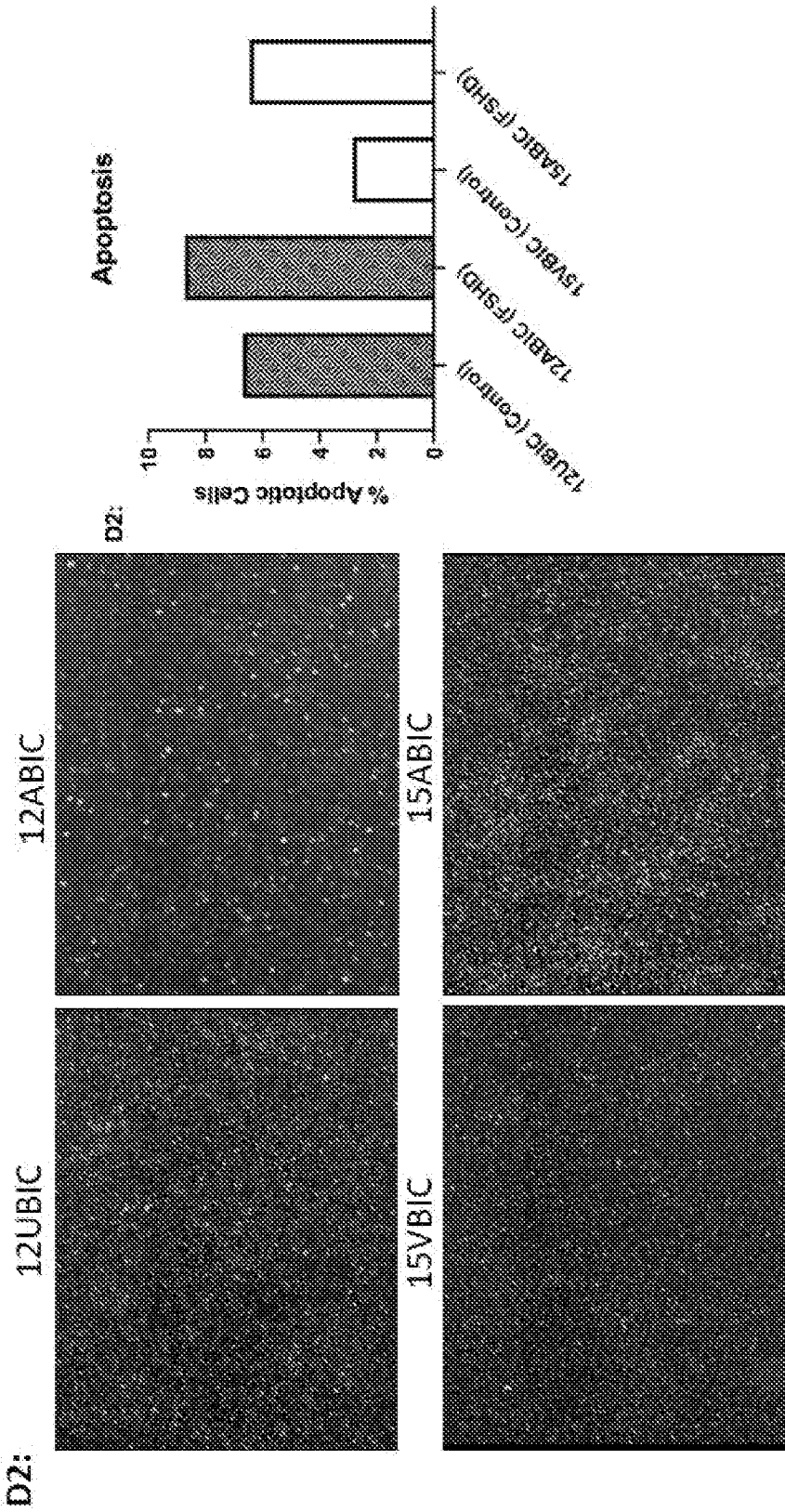


FIG. 3B

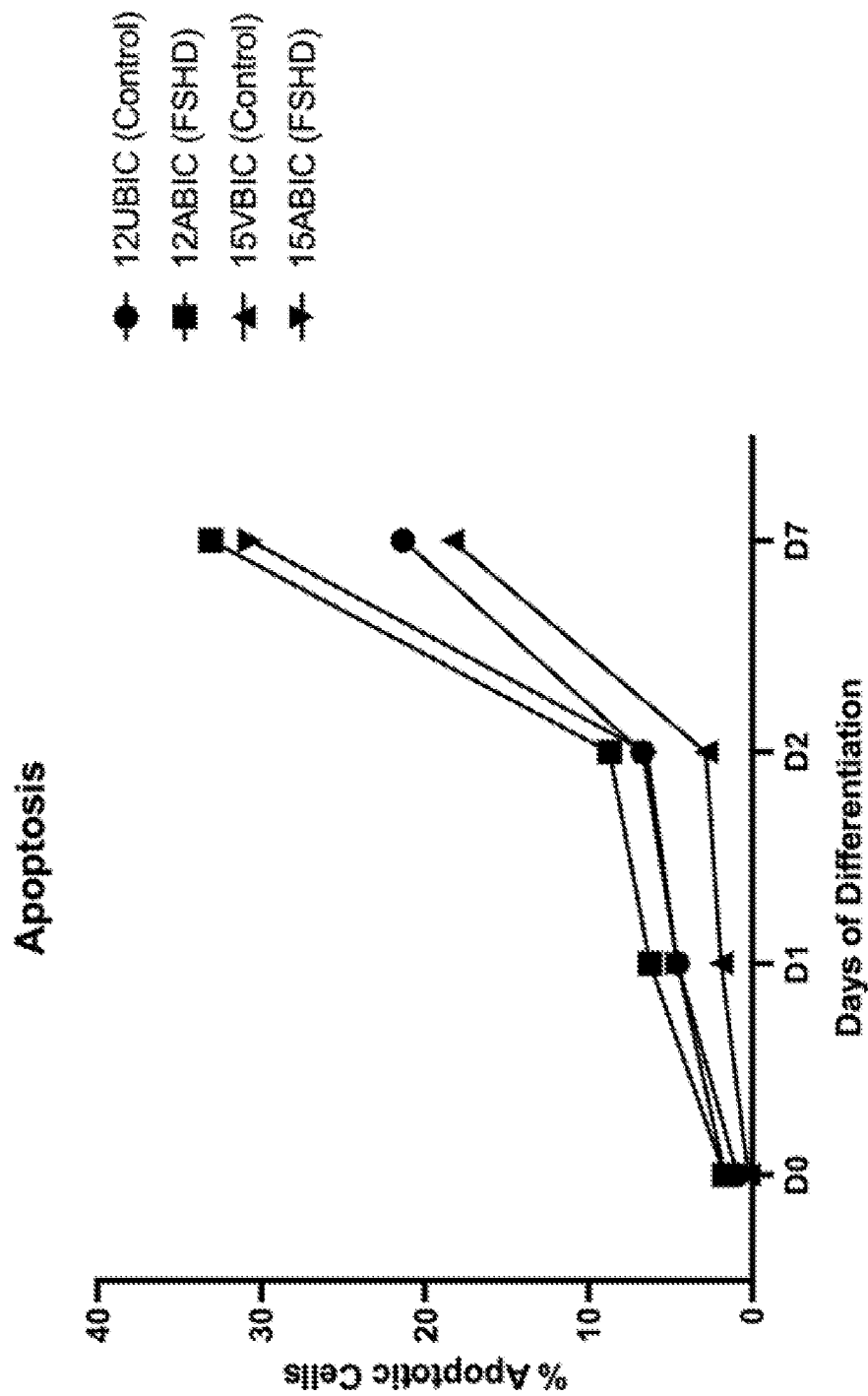


FIG. 3C

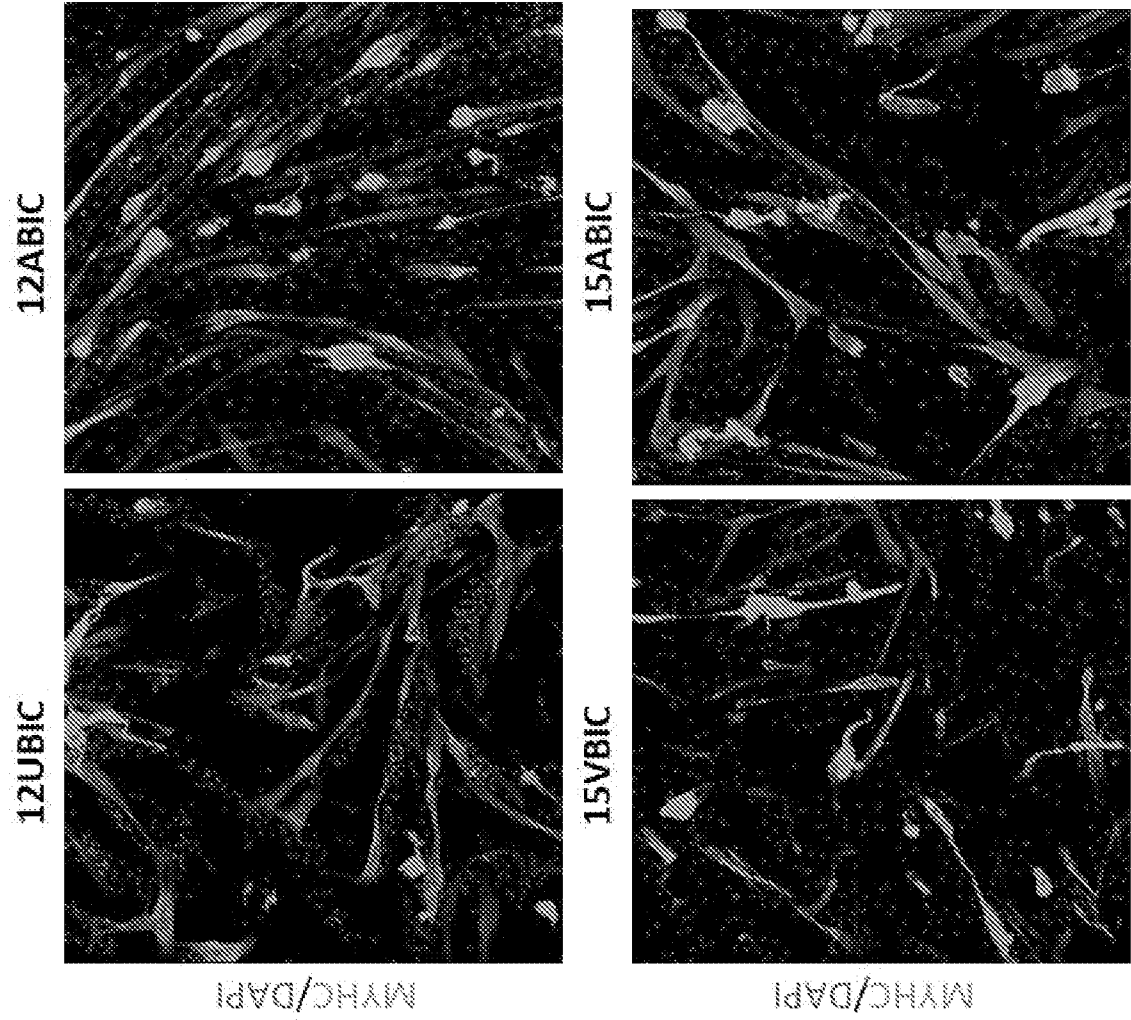


FIG. 3D

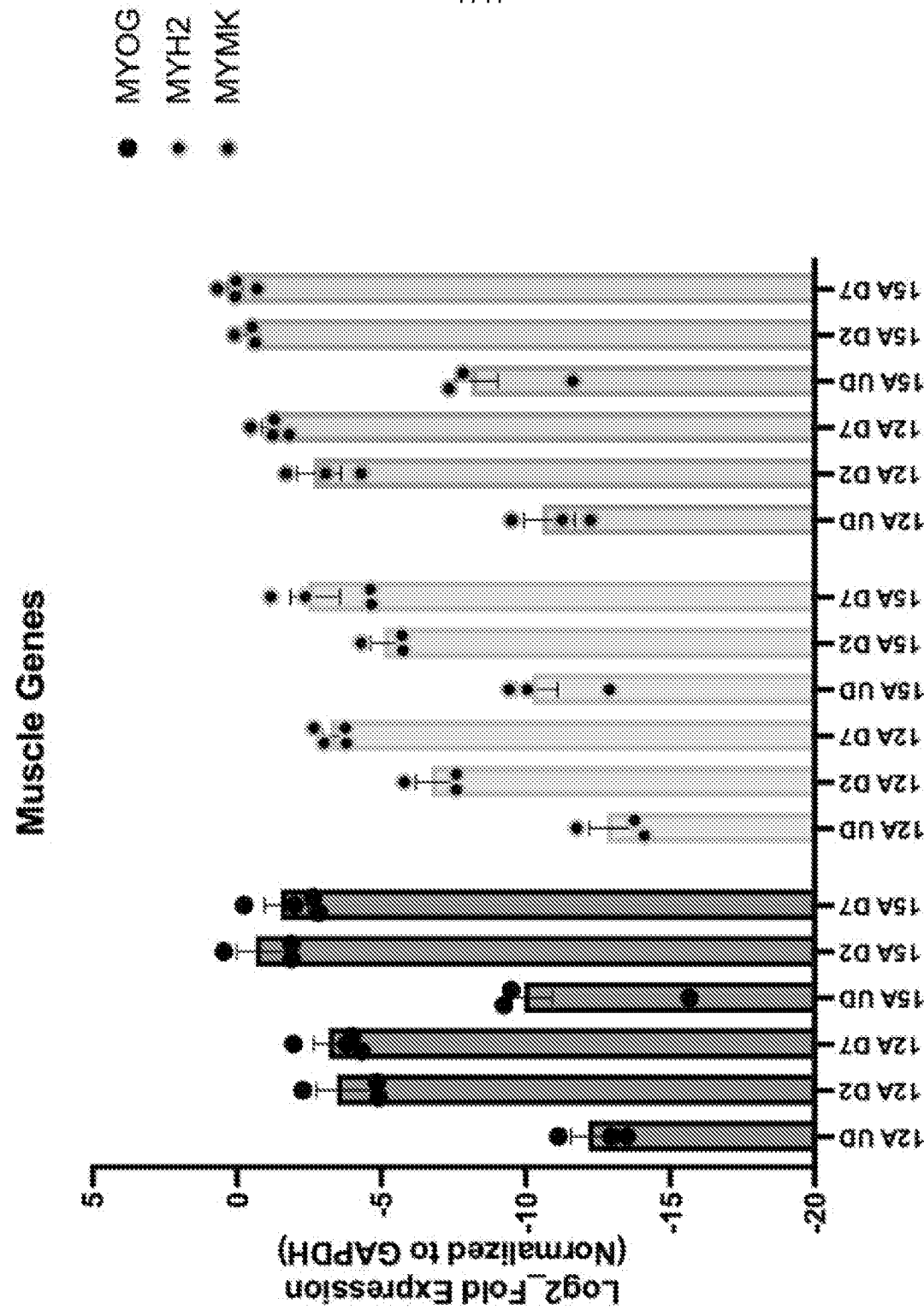


FIG. 3E

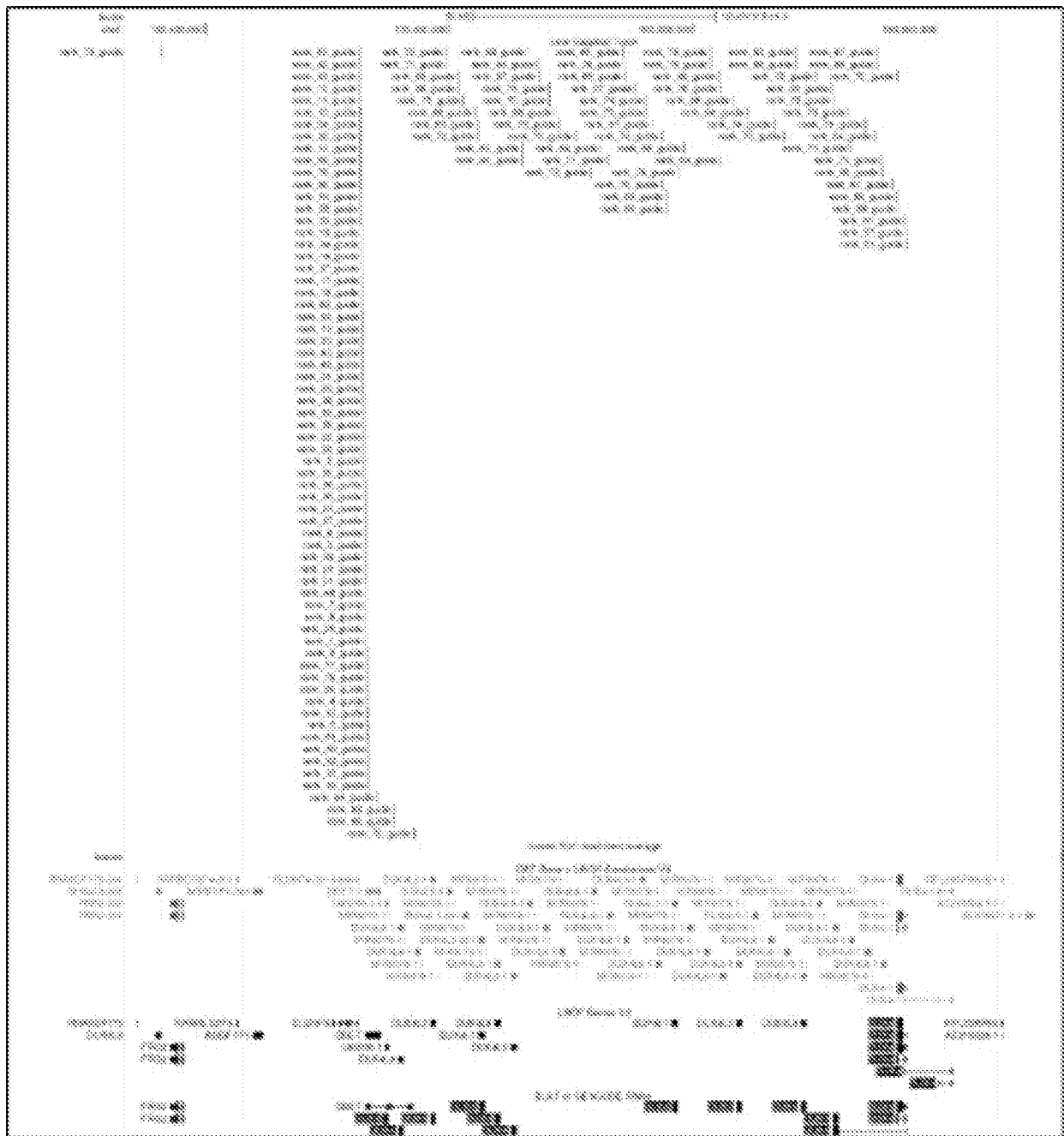


FIG. 4

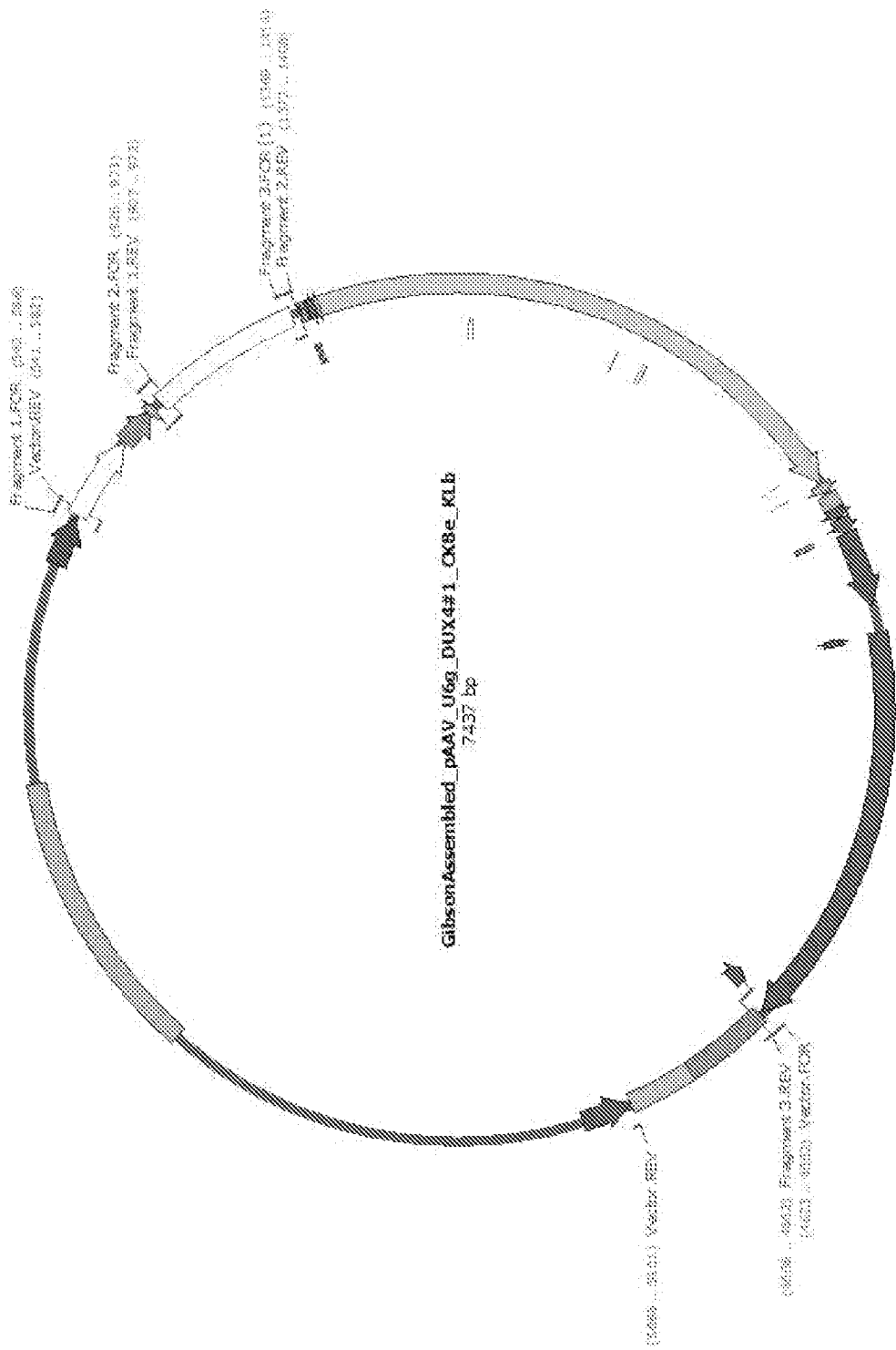


FIG. 5

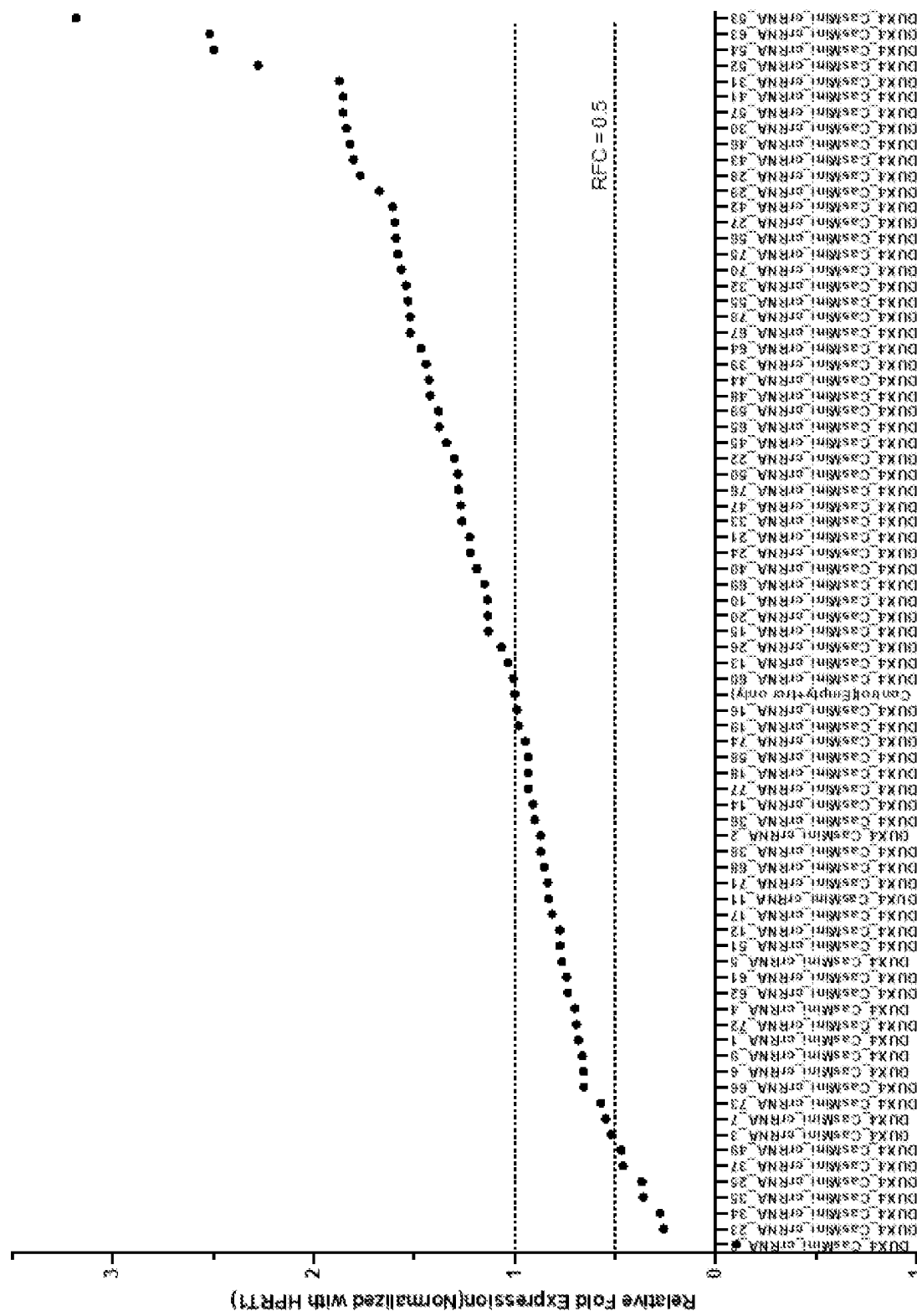


FIG. 6A

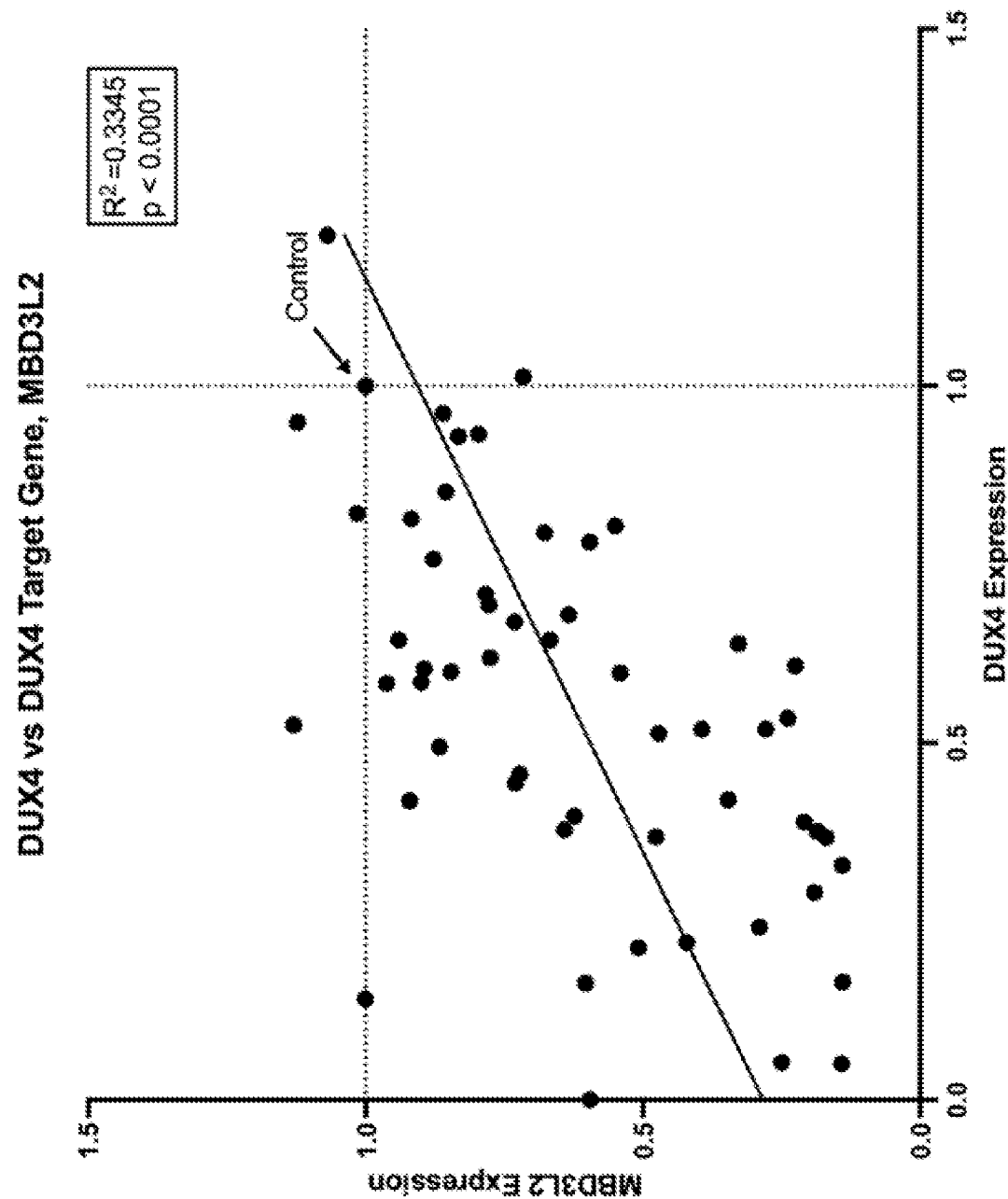


FIG. 6B

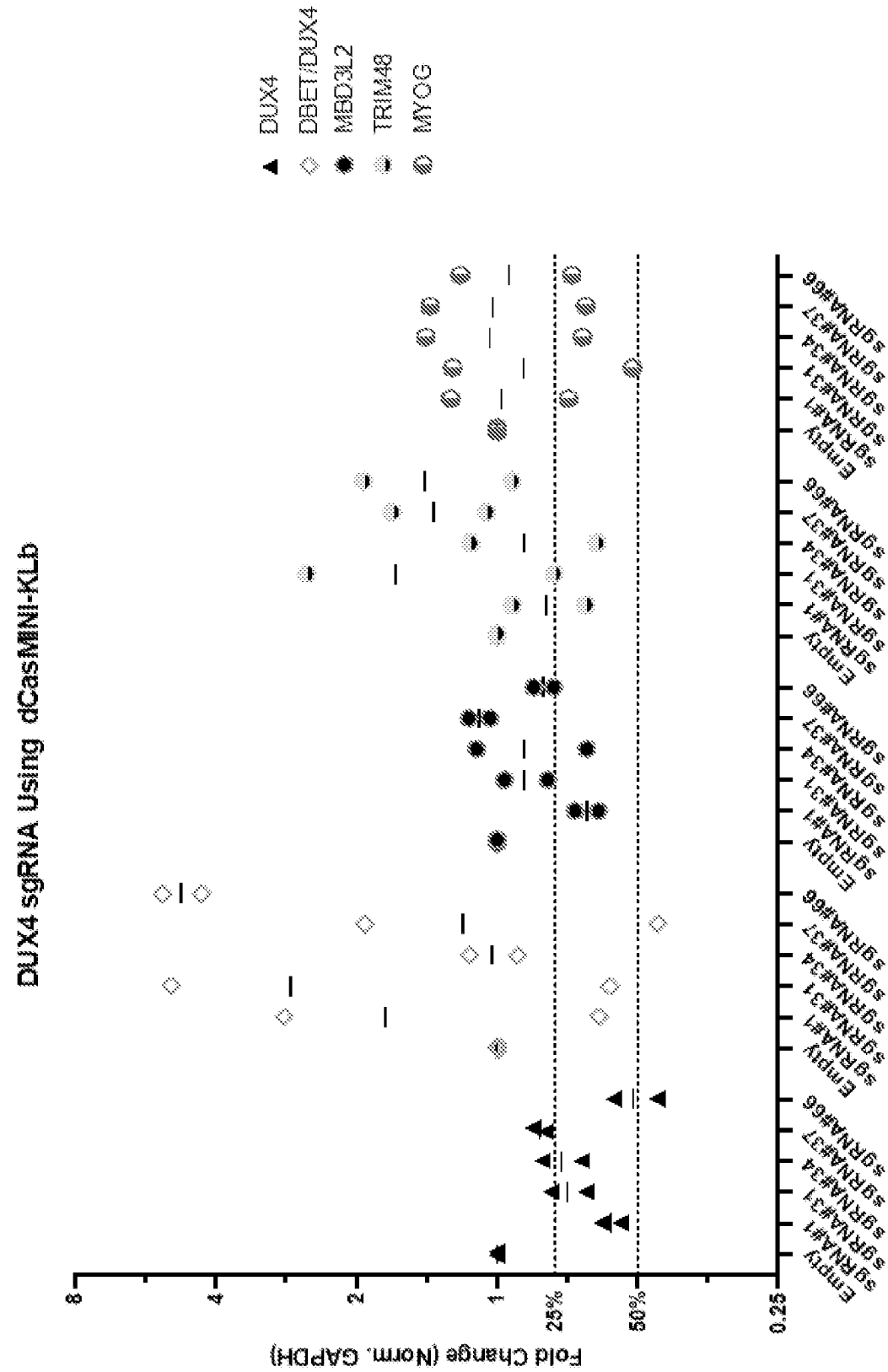


FIG. 7A

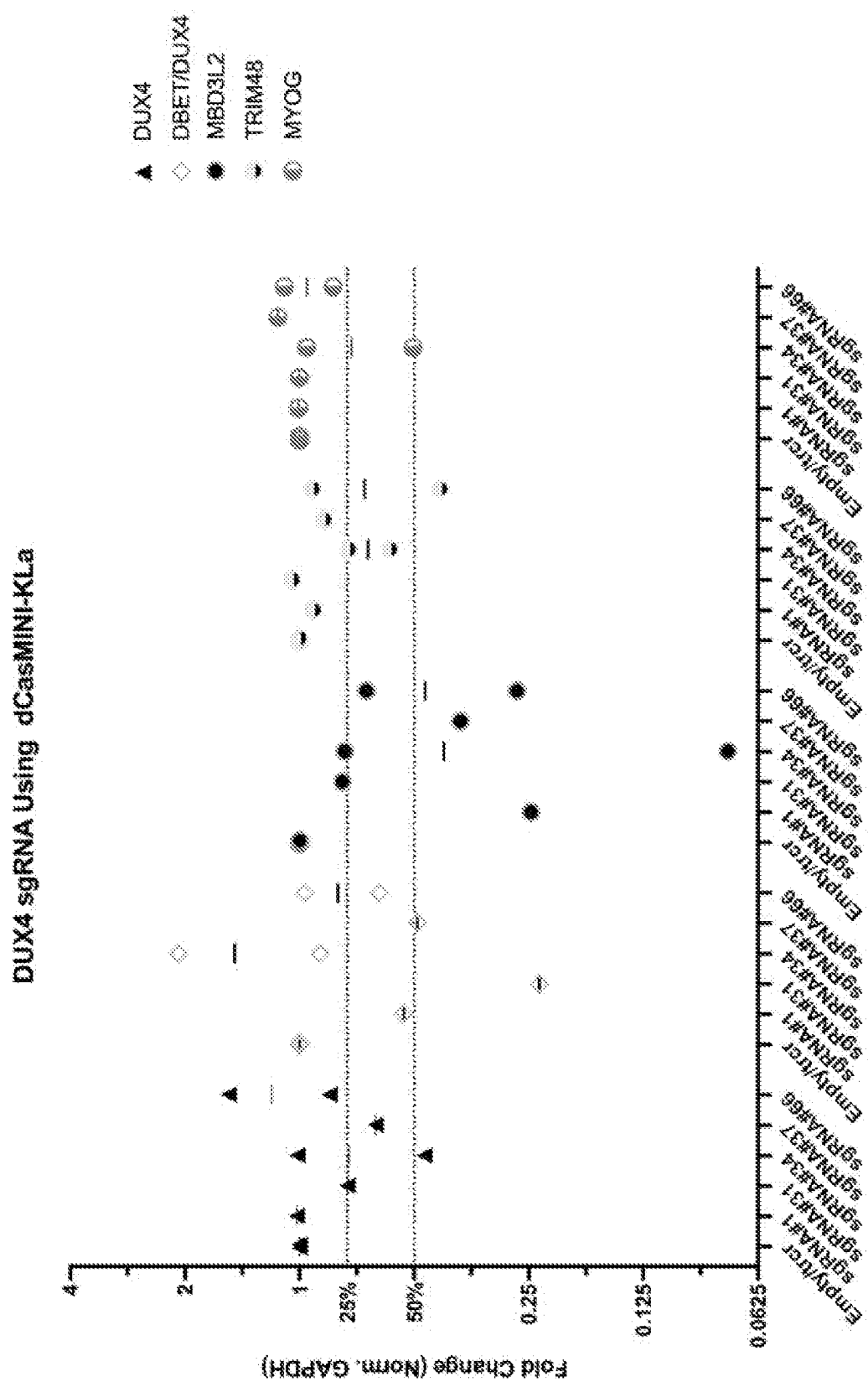


FIG. 7B

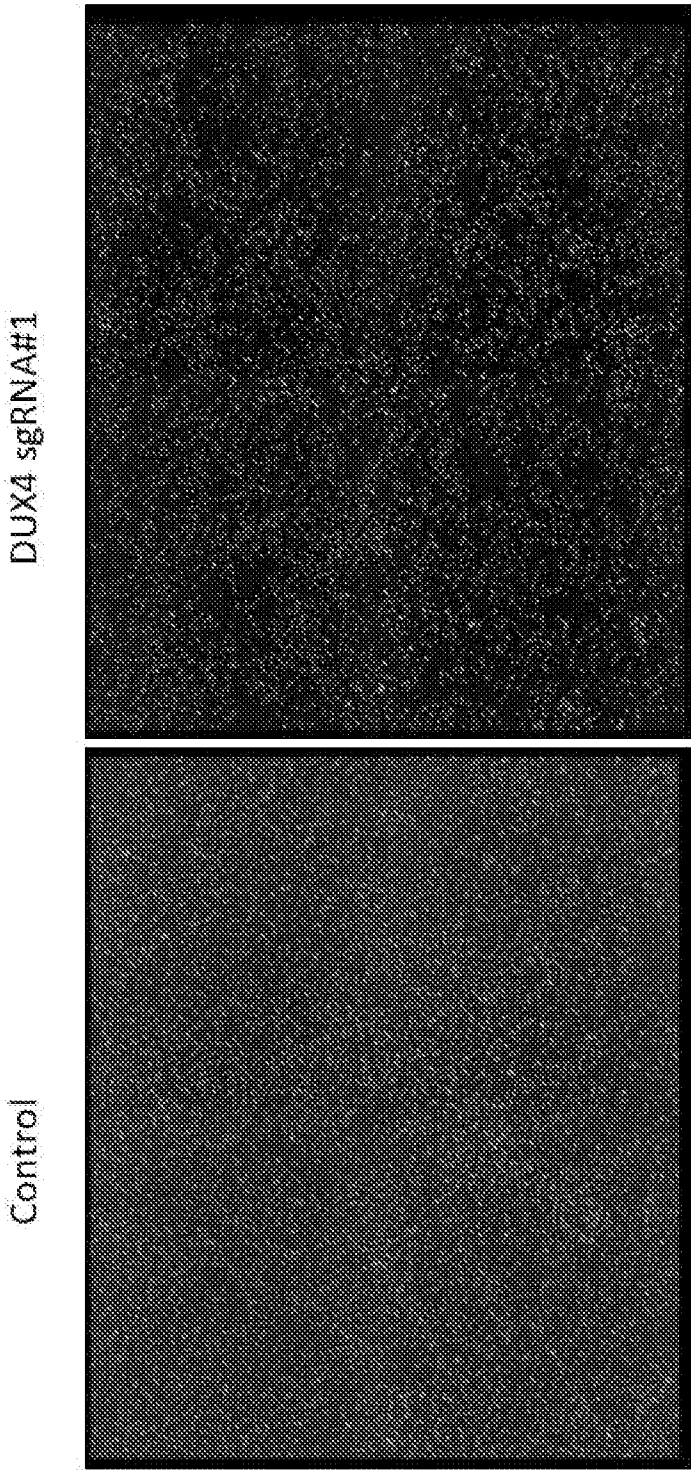


FIG. 8A

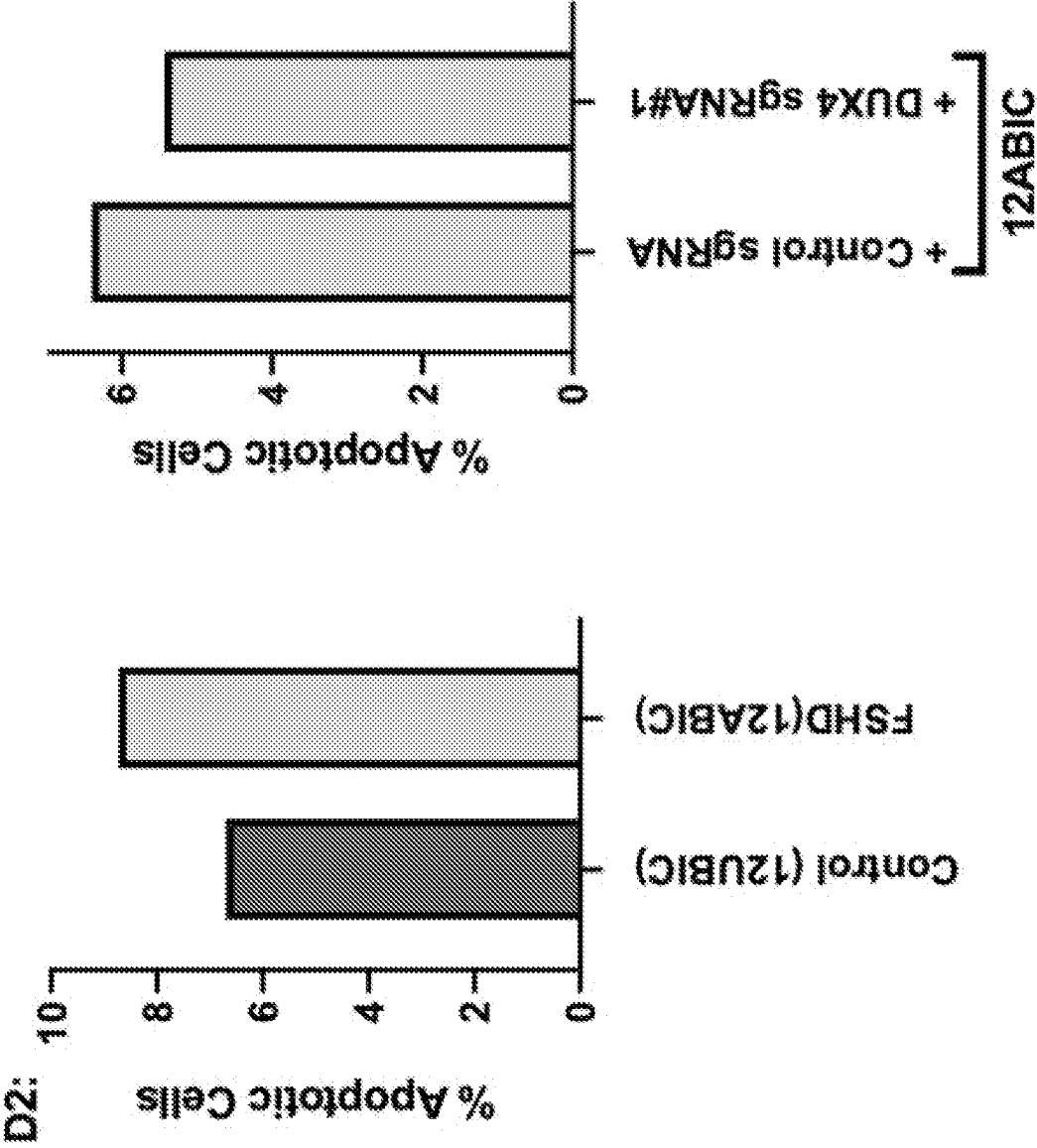


FIG. 8B

Name	Drug	Replicates	Assay Timeline				
			Differentiation	Twitch	Tetanus	Fatigue	Termination
FSHD 12ABIC	Control AAV	12	Day 0	Day 5, 12, 19, 26	Day 6, 13, 20, 27	Day 7, 14, 21, 28	Day 30
FSHD 12ABIC	Cas12f effector-modulator AAV	12					

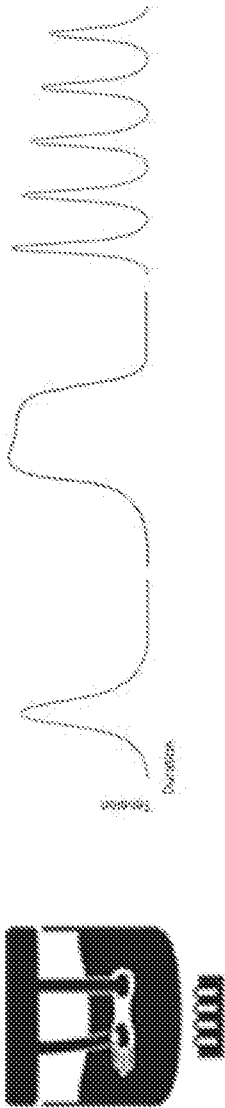


FIG. 9

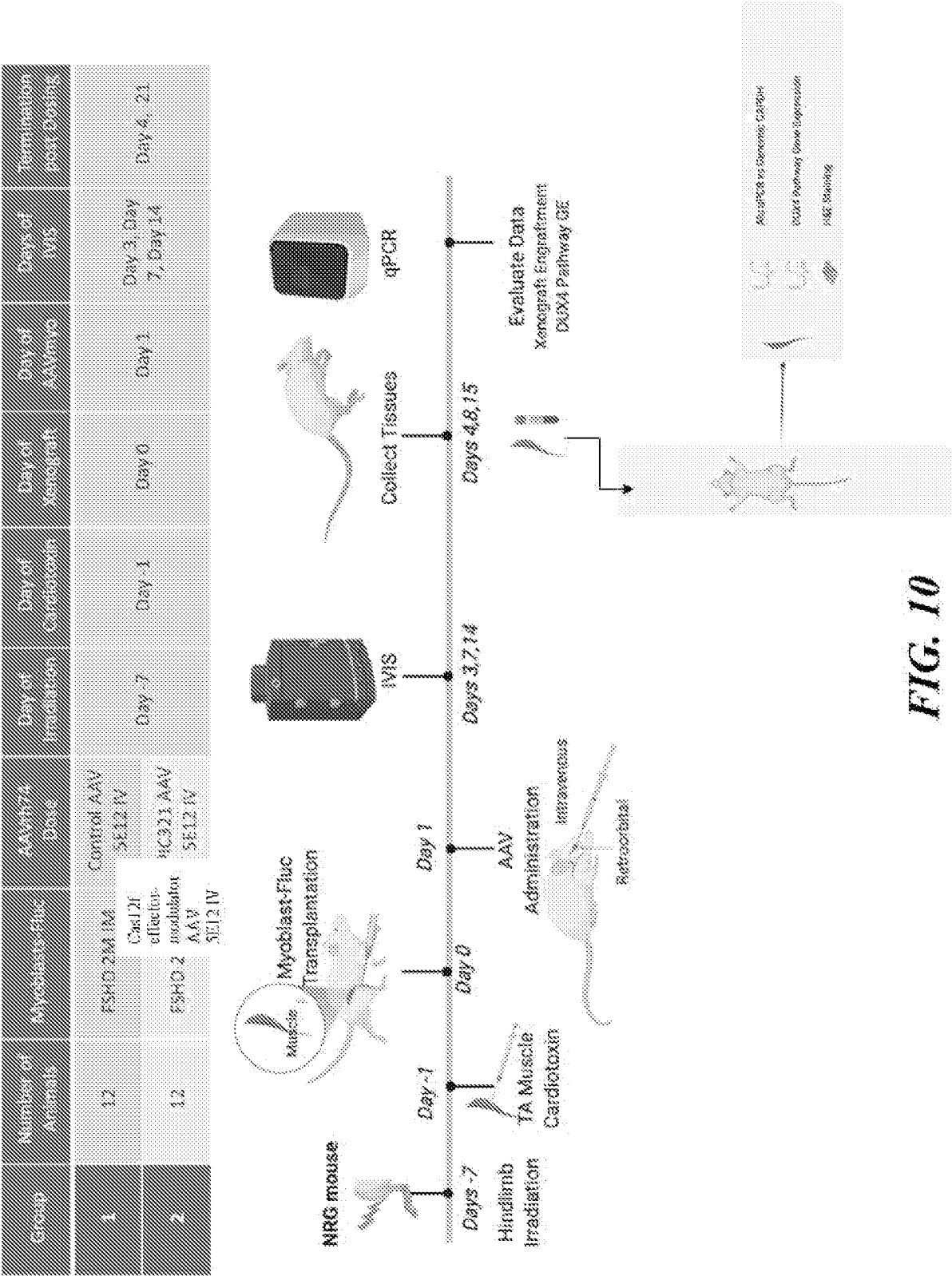


FIG. 10

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 22/33797

A. CLASSIFICATION OF SUBJECT MATTER

IPC - INV. A61K 38/46, A61K 48/00, C12N 15/00; ADD. A61P 21/00 (2022.01)

CPC - INV. C07K 14/4702, C12N 15/113, A61K 38/465; ADD. C12N 2310/20, C12N 15/00, C12N 9/22

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History document

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

See Search History document

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

See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 2020/142479 A1 (RESEARCH INSTITUTE AT NATIONWIDE CHILDREN'S HOSPITAL) 09 July 2020 (09.07.2020) para [0008]-[0009]; [0021]-[0027]; [0036]-[0038]; [0049]; [0056]; [0066]-[0070]; [0091]; [0097]; [0127]; [0163]-[0166]; Fig. 11	1-10, 23-27, 30-34
Y	XIAO et al., Structural basis for substrate recognition and cleavage by the dimerization-dependent CRISPR-Cas12f nuclease. Nucleic Acids Research, Volume 49, Issue 7, 19 April 2021, Pages 4120-4128. full document, especially abstract; p. 4122, col 2, para 3-4	1-10, 23-27, 30-34
X, P	WO 2022/133062 A1 (EPICRISPR BIOTECHNOLOGIES INC.) 23 June 2022 (23.06.2022) full document	1-10, 23-27, 30-34

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

* Special categories of cited documents:

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

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search

12 October 2022

Date of mailing of the international search report

NOV 07 2022

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Kari Rodriguez

Telephone No. PCT Helpdesk: 571-272-4300

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 22/33797

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

a. ☐ forming part of the international application as filed:

☐ in the form of an Annex C/ST.25 text file.

☐ on paper or in the form of an image file.

b. ☐ furnished together with the international application under PCT Rule 13ter.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

☐ in the form of an Annex C/ST.25 text file (Rule 13ter.1(a)).

☐ on paper or in the form of an image file (Rule 13ter.1(b) and Administrative Instructions, Section 713).

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

Pursuant to PCT Rule 13ter.1(a) to (d) applicant failed to timely pay the late sequence furnishing fees. Therefore, the international search was not carried out on the basis of any sequence listing.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 22/33797

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

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

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 11-12, 45-46
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
claims 11-12, 45-46 are held unsearchable because they require a claimed SEQ ID NO that is not supported by an approved sequence listing submission from the applicant. Pursuant to PCT Rule 13ter.1(a) to (d) applicant failed to timely pay the late sequence furnishing fees. Therefore, the claims requiring a sequence are held as unsearchable.

3. ☒ Claims Nos.: 68-90, 94-115
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

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

This International Searching Authority found multiple inventions in this international application, as follows:
This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

--continued on first extra sheet--

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-10, 23-27, 30-34, limited to modification of an expression level; a nuclease, and a guide nucleic acid molecule

Remark on Protest

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

--continued from: Box III Unity of invention is lacking--

Group I+, claims 1-10, 13-34, 65-67, directed to a system for regulating aberrant expression of a target gene in a muscle cell comprising one or more heterologous actuator moieties and one or more gene regulators, or a composition or viral vector comprising the same. The system will be searched to the extent that the modification effected by the system encompass a modification of an expression level, the one or more heterologous actuator moieties encompasses a nuclease, and the one or more gene regulators encompass a guide nucleic acid molecule. The first named invention was determined based on modification of an expression level being the first claimed type of modification (claims 1, 26), a nuclease being the first claimed heterologous actuator moiety (claim 1 (also see claims 70-73, 87-88)), and a guide nucleic acid molecule being the first claimed gene regulator (claim 1). This first named invention has been selected based on the guidance set forth in section 10.54 of the PCT International Search and Preliminary Examination Guidelines. It is believed that claims 1-10, 23-27, 30-34 encompass this first named invention, and thus these claims will be searched without fee to the extent that the system encompasses modification of an expression level, a nuclease, and a guide nucleic acid molecule. Additional modification(s), heterologous actuator moiety(s), and/or gene regulator(s) will be searched upon the payment of additional fees. Applicants must specify the claims that encompass any additionally elected modification(s), heterologous actuator moiety(s), and/or gene regulator(s). Applicants must further indicate, if applicable, the claims which encompass the first named invention, if different than what was indicated above for this group. Failure to clearly identify how any paid additional invention fees are to be applied to the "+" group(s) will result in only the first claimed invention to be searched. An exemplary election would be a system comprising a modification of an expression level and/or methylation level, a nuclease and/or a methyltransferase, and a guide nucleic acid molecule, (claims 1-10, 13-16, 20, 23-34).

[note, claims requiring one or more transcriptional regulators in addition to or alternatively to a nuclease (i.e. claims 13-22) are not included in the first embodiment of Group I+, as these are additional heterologous actuator moieties; and claims requiring KRAB or a variant of KRAB (i.e. claims 17-19, 21-22) are not included in the first election of Group I+, as these are additional non-nuclease and non-methyltransferase actuator moieties, as KRAB is a ZFP]

Group II+, claims 35-44, 47-64, 91-93, directed to a method for regulating aberrant expression of a target gene in a muscle cell. Group II+ will be searched upon payment of additional fees. The method may be searched, for example, to encompass a modification of an expression level, where the one or more heterologous actuator moieties encompass a nuclease, and the one or more gene regulators encompass a guide nucleic acid molecule for an additional fee and election as such. It is believed that claims 35-44, 57-61, 64, 91-93 read on this exemplary invention. Additional modification(s), heterologous actuator moiety(s), and/or gene regulator(s) will be searched upon the payment of additional fees. Applicants must specify the claims that encompass any additionally elected modification(s), heterologous actuator moiety(s), and/or gene regulator(s). Failure to clearly identify how any paid additional invention fees are to be applied to the "+" group(s) will result in only the first claimed invention to be searched. Another exemplary election would be a method comprising a modification of an expression level and/or methylation level, a nuclease and/or a methyltransferase, and a guide nucleic acid molecule (claims 35-44, 47-50, 54, 57-64, 91-93).

The inventions listed as Groups I+ and II+ do not relate to a single special technical feature under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Special technical features

Group I+ has the special technical feature of a composition comprising or consisting of a heterologous actuator moiety coupled to a gene regulator, that is not required by Group II+.

Group II+ has the special technical feature of a method comprising contacting a muscle cell with a heterologous actuator moiety coupled to a gene regulator, and upon the contacting, binding the target gene with the complex to effect modification of an expression level and/or a methylation level of the target gene in the muscle cell, that is not required by Group I+.

The inventions of Group I+ and Group II+ each include the special technical feature of a different type of modification, a different heterologous actuator moiety, and/or a different gene regulator, and is considered a distinct technical feature.

Common technical features

The inventions of Groups I+ and II+ share the common technical feature of regulation of aberrant expression of a target gene in a muscle cell, comprising:
a heterologous polypeptide comprising a nuclease, wherein the nuclease has a length that is less than or equal to about 900 amino acids; and
a guide nucleic acid molecule configured to form a complex with the heterologous polypeptide, wherein the guide nucleic acid molecule exhibits specific binding to a target polynucleotide sequence at or adjacent to, a D4Z4 repeat array in the muscle cell;
wherein, upon formation of the complex, there is modified expression level and/or methylation level of the target gene in the muscle cell.

The inventions of Group I+ further share the common technical feature of a system for regulating aberrant expression of a target gene in a muscle cell comprising one or more heterologous actuator moieties and one or more gene regulators, or a composition or viral vector comprising the same (claims 1, 31, 32, 65).

The inventions of Group II+ further share the common technical feature of a method for regulating aberrant expression of a target gene in a muscle cell, comprising: contacting the muscle cell with a heterologous actuator moiety coupled to a gene regulator, wherein the heterologous actuator moiety is capable of forming a complex with the target gene in the muscle cell, and wherein the gene regulator is capable of modifying an expression level and/or a methylation level of the target gene in the muscle cell; and (b) upon formation of the complex, sustaining the modified expression level and/or methylation level of the target gene in the muscle cell for at least about 2 days (claims 35, 91).

The feature shared by Groups I+ and II+ and the features shared by the inventions listed as Group I+ and II+ are taught or made obvious by WO 2020/142479 A1 to Research Institute at Nationwide Children's Hospital (hereinafter "NCH").

--continued on next sheet--

--continued from previous sheet--

NCH teaches (claim 1) a system for regulating aberrant expression of a target gene in a muscle cell (para [0009]; [0091]), comprising: a heterologous polypeptide comprising a nuclease (para [0097] "a recombinant gene editing complex comprising a nucleic acid comprising...a Cas13b direct repeat sequence, which is delivered in conjunction with Cas13 (e.g., Cas13b)"); and a guide nucleic acid molecule configured to form a complex with the heterologous polypeptide, wherein the guide nucleic acid molecule exhibits specific binding to a target polynucleotide sequence at or adjacent to, a D4Z4 repeat array in the muscle cell (para [0097] "a recombinant gene editing complex comprising a nucleic acid comprising various gRNA nucleotide sequences described herein attached to a Cas13b direct repeat sequence, which is delivered in conjunction with Cas13 (e.g., Cas13b) to edit the DUX4 gene"; [0050] "the DUX4 gene is embedded within identical D4Z4 DNA repeats"; [0091] "methods of transducing muscle cells and muscle tissues directed"), wherein, upon formation of the complex, the complex is capable of binding the target polynucleotide sequence, to effect modification of an expression level and/or a methylation level of the target gene in the muscle cell (para [0023] "wherein the target gene is within the D4Z4 repeat array a recombinant gene editing complex...wherein binding of the complex to the target nucleic acid sequence results in inhibition of DUX4 gene expression"; [0091]).

NCH does not specifically teach the nuclease has a length that is less than or equal to about 900 amino acids. However, NCH does teach the nuclease may be of numerous types, and may be an active fragment thereof (para [0023]; [0049]; [0056] "In some aspects, the nucleic acid encoding the Cas13b enzyme comprises...at least about 70%...identity to the nucleotide sequence set forth in SEQ ID NO: 36, or a biologically active fragment thereof"). It would have been obvious to one of ordinary skill in the art to have determined and applied a functional fragment of a known nuclease that is less than or equal to about 900 amino acids, during the ordinary course of experimentation, in order to optimize transfection and expression of the nuclease in cells.

NCH teaches (claim 31) a composition comprising the system (para [0009] "compositions and viral vectors comprising the nucleic acids which are designed to inhibit DUX4 expression"), and (claim 32) a viral vector comprising the system (para [0009] "compositions and viral vectors comprising the nucleic acids which are designed to inhibit DUX4 expression").

NCH teaches (claim 35) a method for regulating aberrant expression of a target gene in a muscle cell (para [0009] "methods for using these products for inhibiting and/or interfering with expression of a DUX4 gene in a cell"; [0091] "methods of transducing muscle cells and muscle tissues directed"), comprising:

(a) contacting the muscle cell with a complex (para [0018] "a method of inhibiting and/or interfering with expression of a DUX4 gene in a cell comprising contacting the cell with an adeno- associated virus or a composition"; [0091]) comprising (i) a heterologous polypeptide comprising a nuclease (para [0009] "a recombinant gene editing complex comprising at least one nucleic acid comprising a nucleotide sequence encoding Cas13 or a Cas13 ortholog or variant") and (ii) a guide nucleic acid molecule exhibiting specific binding to a target polynucleotide sequence at or adjacent to, a D4Z4 repeat array in the muscle cell (para [0009] "a nucleotide sequence encoding a guide RNA (gRNA) that specifically hybridizes to a target nucleic acid sequence encoding a DUX4"); and
(b) upon the contacting, binding the target gene with the complex to effect modification of an expression level of the target gene in the muscle cell for at least about 2 days (para [0009] "wherein binding of the complex to the target nucleic acid sequence results in inhibition of DUX4 gene expression"; [0091]; [0036] "Fig. 11 shows the reduction of DUX4 mRNA and protein (as indicated by relative expression of WFDC3) three weeks post-injection with DUX4 (le9) after co-injection of sAAV6-Cas13b and scAAV6-gRNA1 (5el0) into the mouse TA"; [0173]), wherein the target gene is within the D4Z4 repeat array (para [0050] "the DUX4 gene is embedded within identical D4Z4 DNA repeats").

As the technical features were known in the art at the time of the invention, they cannot be considered special technical features that would otherwise unify the groups.

Therefore, Groups I+ and II+ inventions lack unity under PCT Rule 13 because they do not share the same or corresponding special technical feature.

**continuation of number 4 above: claims 68-90, 94-115 are held unsearchable because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Further, claims 11-12, 45-46 are held unsearchable because they require a claimed SEQ ID NO that is not supported by an approved sequence listing submission from the applicant. Pursuant to PCT Rule 13ter.1(a) to (d) applicant failed to timely pay the late sequence furnishing fees. Therefore, the claims requiring a sequence are held as unsearchable.

NOTE, claim 32 which depends from "the system of any one of the preceding claims", as drafted, is objected to, because claim 31 is not directed to a system. Claim 32 is reconstrued to depend from "the system of any one of claims 1-30".

NOTE, claim 67 which depends from "the system of any one of the preceding claims", as drafted, is objected to, because claims 31-64 are not directed to a system. Claim 67 is reconstrued to depend from "the system of any one of claims 65-66".

NOTE, claim 93 which depends from "the method of any one of the preceding claims", as drafted, is objected to, because claims 1-34, 65-90 are not directed to a method. Claim 93 is reconstrued to depend from "the method of any one of claims 91-92".