



(12) **DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION**

(13) **A1**

(86) Date de dépôt PCT/PCT Filing Date: 2020/08/19
 (87) Date publication PCT/PCT Publication Date: 2021/02/25
 (85) Entrée phase nationale/National Entry: 2022/02/17
 (86) N° demande PCT/PCT Application No.: US 2020/047080
 (87) N° publication PCT/PCT Publication No.: 2021/034984
 (30) Priorités/Priorities: 2019/08/19 (US62/888,916);
 2020/01/31 (US62/968,743)

(51) Cl.Int./Int.Cl. *A61K 48/00* (2006.01),
A61P 21/00 (2006.01), *C12N 15/113* (2010.01),
C12N 15/85 (2006.01)
 (71) Demandeur/Applicant:
 DUKE UNIVERSITY, US
 (72) Inventeurs/Inventors:
 GERSBACH, CHARLES A., US;
 KWON, JENNIFER, US
 (74) Agent: SMART & BIGGAR LLP

(54) Titre : SPECIFICATION DE LIGNEE CELLULAIRES PROGENITRICES DE MYOBLASTES SQUELETTIQUES PAR DES ACTIVATEURS TRANSCRIPTIONNELS A BASE DE CRISPR/CAS9
 (54) Title: SKELETAL MYOBLAST PROGENITOR CELL LINEAGE SPECIFICATION BY CRISPR/CAS9-BASED TRANSCRIPTIONAL ACTIVATORS

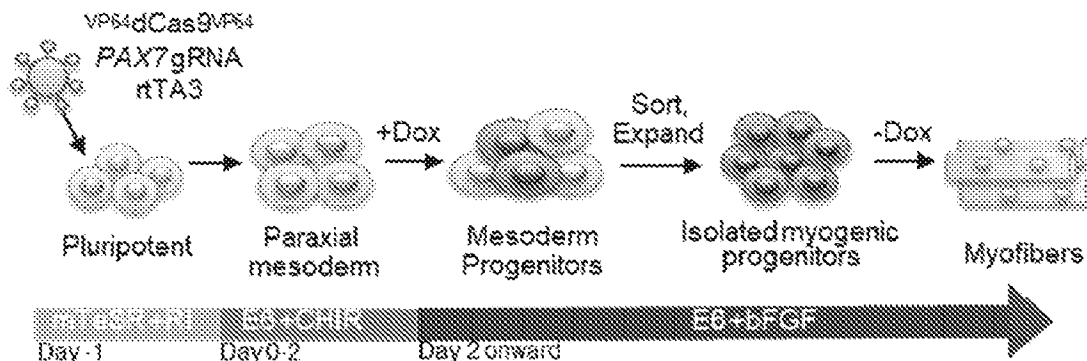


FIG. 1A

(57) **Abrégé/Abstract:**

Disclosed herein are methods and systems for increasing expression of Pax7, methods of activating endogenous myogenic transcription factor Pax7 in a cell, methods of differentiating a stem cell into a skeletal muscle progenitor cell, as well as compositions and methods for treating a subject in need of regenerative muscle progenitor cells. The compositions and methods may include a Cas9-based transcriptional activator protein and at least one guide RNA (gRNA) targeting Pax7.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(10) International Publication Number
WO 2021/034984 A3

(43) International Publication Date
25 February 2021 (25.02.2021)

(51) International Patent Classification:

A61K 48/00 (2006.01) C12N 15/113 (2010.01)
A61P 21/00 (2006.01) C12N 15/85 (2006.01)

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(21) International Application Number:

PCT/US2020/047080

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))

(22) International Filing Date:

19 August 2020 (19.08.2020)

(25) Filing Language:

English

(88) Date of publication of the international search report:

01 April 2021 (01.04.2021)

(26) Publication Language:

English

(30) Priority Data:

62/888,916 19 August 2019 (19.08.2019) US
62/968,743 31 January 2020 (31.01.2020) US

(71) Applicant: **DUKE UNIVERSITY** [US/US]; 2812 Erwin Road, Suite 306, Durham, NC 27705 (US).

(72) Inventors: **GERSBACH, Charles, A.**; 3220 Wilderness Road, Durham, NC 27712 (US). **KWON, Jennifer**; c/o Duke University, 2812 Erwin Road, Suite 306, Durham, NC 27705 (US).

(74) Agent: **COX, Julia, M.** et al.; Michael Best & Friedrich LLP, 790 N. Water Street, Suite 2500, Milwaukee, WI 53202 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, 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,



WO 2021/034984 A3

(54) Title: SKELETAL MYOBLAST PROGENITOR CELL LINEAGE SPECIFICATION BY CRISPR/CAS9-BASED TRANSCRIPTIONAL ACTIVATORS

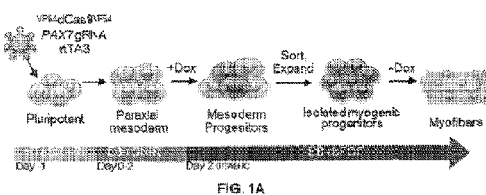


FIG. 1A

(57) Abstract: Disclosed herein are methods and systems for increasing expression of Pax7, methods of activating endogenous myogenic transcription factor Pax7 in a cell, methods of differentiating a stem cell into a skeletal muscle progenitor cell, as well as compositions and methods for treating a subject in need of regenerative muscle progenitor cells. The compositions and methods may include a Cas9-based transcriptional activator protein and at least one guide RNA (gRNA) targeting Pax7.

**SKELETAL MYOBLAST PROGENITOR CELL LINEAGE SPECIFICATION BY
CRISPR/CAS9-BASED TRANSCRIPTIONAL ACTIVATORS**

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to U.S. Provisional Patent Application No. 62/888,916, filed August 19, 2019, and U.S. Provisional Patent Application No. 62/968,743, filed January 31, 2020, each of which is incorporated herein by reference in its entirety.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

[0002] This invention was made with government support under grant 1DP2-OD008586 and 1R01DA036865 awarded by the National Institutes of Health. The government has certain rights in the invention.

FIELD

[0003] This disclosure relates to compositions and methods for increasing the expression of Pax7 in stem cells, inducing differentiation of a stem cell into a skeletal muscle progenitor cell, and using these skeletal muscle progenitor cells to regenerate damaged muscle tissue.

INTRODUCTION

[0004] Human pluripotent stem cells (hPSCs) are a promising cell source for regenerative medicine, disease modeling, and drug discovery in pathologies of muscle disease. Directed differentiation of hPSCs into skeletal muscle cells can be achieved via stepwise small molecule-based protocols or ectopic expression of transgenes. While having the benefit of being transgene-free, small molecule-based protocols tend to be relatively lengthy, inefficient, and lack the scalability required for cell therapy or drug screening applications. Transgene-based approaches rely on overexpression of key myogenic transcription factors, including Pax3, Pax7, and MyoD. These protocols are highly efficient in yielding populations of myogenic cells, and they do so more rapidly than transgene-free methods. Generation of satellite cells, such as the skeletal muscle stem cell population, is particularly appealing for myogenic cell therapies. Although satellite cells can robustly regenerate damaged muscles *in vivo*, they cannot be isolated and expanded *ex vivo* without relinquishing their stemness, resulting in loss of engraftment capabilities. As such, the generation of functional Pax7+ satellite cells from hPSCs has been attempted by pairing various differentiation protocols with exogenous Pax7 cDNA overexpression. There is a need for alternative methods for generating populations of myogenic cells.

SUMMARY

[0005] In an aspect, the disclosure relates to a guide RNA (gRNA) molecule targeting Pax7 or a promoter or regulatory element of the Pax7 gene. The gRNA may comprise a polynucleotide sequence corresponding to at least one of SEQ ID NOs: 1-8 or 69-76, or a variant thereof.

[0006] In a further aspect, the disclosure relates to a DNA targeting system for increasing expression of Pax7. The DNA targeting system may comprise at least one gRNA that binds and targets a Pax7 gene or a portion thereof. In some embodiments, the at least one gRNA comprises a polynucleotide sequence corresponding to at least one of SEQ ID NOs: 1-8 or 69-76, or a variant thereof.

[0007] In some embodiments, the DNA targeting system further includes a Clustered Regularly Interspaced Short Palindromic Repeats associated (Cas) protein or a fusion protein, wherein the fusion protein comprises two heterologous polypeptide domains, wherein the first polypeptide domain comprises a Cas protein, a zinc finger protein, or a TALE protein, and the second polypeptide domain has transcription activation activity. In some embodiments, the Cas protein comprises a *Streptococcus pyogenes* Cas9 molecule, or a variant thereof. In some embodiments, the fusion protein comprises VP64-dCas9-VP64 (^{VP64}dCas9^{VP64}). In some embodiments, the Cas protein comprises a Cas9 that recognizes a Protospacer Adjacent Motif (PAM) of NGG (SEQ ID NO: 31), NGA (SEQ ID NO: 32), NGAN (SEQ ID NO: 33), or NGNG (SEQ ID NO: 34).

[0008] Another aspect of the disclosure provides an isolated polynucleotide sequence comprising a gRNA molecule as disclosed herein.

[0009] Another aspect of the disclosure provides an isolated polynucleotide sequence encoding a DNA targeting system as disclosed herein.

[00010] Another aspect of the disclosure provides a vector comprising an isolated polynucleotide sequence as disclosed herein.

[00011] Another aspect of the disclosure provides a vector encoding a gRNA molecule as disclosed herein and a Clustered Regularly Interspaced Short Palindromic Repeats associated (Cas) protein.

[00012] Another aspect of the disclosure provides a cell comprising a gRNA as disclosed herein, a DNA targeting system as disclosed herein, an isolated polynucleotide sequence as disclosed herein, or a vector as disclosed herein, or a combination thereof.

[00013] Another aspect of the disclosure provides a pharmaceutical composition comprising a gRNA as disclosed herein, a DNA targeting system as disclosed herein, an isolated polynucleotide sequence as disclosed herein, a vector as disclosed herein, or a cell as disclosed herein, or a combination thereof.

[00014] Another aspect of the disclosure provides a method of activating endogenous myogenic transcription factor Pax7 in a cell. The method may include administering to the cell a gRNA as disclosed herein, a DNA targeting system as disclosed herein, an isolated polynucleotide sequence as disclosed herein, or a vector as disclosed herein.

[00015] Another aspect of the disclosure provides a method of differentiating a stem cell into a skeletal muscle progenitor cell. The method may include administering to the stem cell a gRNA as disclosed herein, a DNA targeting system as disclosed herein, an isolated polynucleotide sequence as disclosed herein, or a vector as disclosed herein.

[00016] In some embodiments, endogenous expression of Pax7 mRNA is increased in the skeletal muscle progenitor cell. In some embodiments, the expression of Myf5, MyoD, MyoG, or a combination thereof, is increased in the skeletal muscle progenitor cell. In some embodiments, the stem cell is induced into myogenic differentiation. In some embodiments, the skeletal muscle progenitor cell maintains Pax7 expression after at least about 6 passages.

[00017] Another aspect of the disclosure provides a method of treating a subject in need thereof. The method may include administering to the subject a cell as disclosed herein.

[00018] In some embodiments, the level of dystrophin+ fibers in the subject is increased. In some embodiments, muscle regeneration in the subject is increased.

[00019] The disclosure provides for other aspects and embodiments that will be apparent in light of the following detailed description and accompanying figures.

BRIEF DESCRIPTION OF THE DRAWINGS

[00020] **FIGS. 1A-1G. Generation of myogenic progenitors from hPSCs via VP64-dCas9-VP64-mediated activation of endogenous PAX7.** (FIG. 1A) Schematic of hPSC myogenic differentiation with small molecules and lentiviral activation of PAX7. (FIG. 1B) The lentiviral constructs used for the gRNA and inducible VP64-dCas9-VP64 and PAX7 cDNA expression. (FIG. 1C) Representative phase-contrast images showing morphological changes during the first 10 days of differentiation. Scale bar = 200 μ m. (FIG. 1D) RNA was harvested at day 0 and day 2 for qRT-PCR analysis of mesodermal markers. Results are expressed as

fold change over day 0 (mean \pm SEM, n = 3 independent replicates). (FIG. 1E) Representative FACS plot at day 14 when VP64-dCas9-VP64-2a-mCherry+ cells were sorted for expansion. (FIG. 1F) Representative immunostaining of PAX7 at 5 days post-sort. Scale bar = 100 μ m. (FIG. 1G) Growth of purified myogenic progenitors derived from iPSC differentiation during post-sort expansion phase was monitored over 2 weeks. Fold-growth over two weeks was significantly greater in VP64-dCas9-VP64-treated cells compared to PAX7 cDNA-treated cells. P value determined by one-way ANOVA followed by Tukey's post hoc test (mean \pm SEM, n = 3 independent replicates).

[00021] FIGS. 2A-2F. Characterization of myogenic progenitors derived from iPSCs via VP64-dCas9-VP64-mediated activation of endogenous PAX7 or exogenous PAX7 cDNA expression. (FIG. 2A) Relative amounts of total PAX7 mRNA was determined by qRT-PCR using primers complementary to sequences present in the gene body. (FIG. 2B) Endogenous PAX7 mRNA was detected using primers complementary to sequences in the 3' UTR of either isoforms PAX7-A or PAX7-B. (FIG. 2C) The mRNA expression levels of myogenic markers MYF5, MYOD, and MYOG during the expansion phase. (FIG. 2D) Immunofluorescence staining of early and mature myogenic markers MYF5, MYOD, and MYOG, and myosin heavy chain (MHC). (FIG. 2E) Representative FACS analysis of CD29 and CD56 surface marker expression during the expansion phase. (FIG. 2F) Mean fluorescence intensity (MFI) of CD56 staining intensity across treatments. All P values were determined by one-way ANOVA followed by Tukey's post hoc test (mean \pm SEM, n = 3 independent replicates).

[00022] FIGS. 3A-3C. Transplantation of VP64-dCas9-VP64-generated myogenic progenitors into immunodeficient mice demonstrates *in vivo* regenerative potential. (FIG. 3A) Detection of human-derived fibers in VP64-dCas9-VP64-treated cells 1 month after intramuscular injection of 5×10^5 differentiated iPSCs into NSG mice pre-injured with BaCl₂. Sections are stained with human-specific dystrophin and lamin A/C antibodies to mark donor-derived fibers and nuclei. Scale bar = 100 μ m. (FIG. 3B) Quantification of human dystrophin+ fibers in the section with highest number of dystrophin+ fibers in each muscle. *p < 0.05 determined by student's t-test compared to control (mean \pm SEM, n = 3 mice). (FIG. 3C) Identification of donor-derived satellite cells expressing PAX7 and human-specific lamin A/C, and residing adjacent to the basal lamina as indicated by laminin staining. Scale bar = 25 μ m.

[00023] FIGS. 4A-4D. Induction of endogenous PAX7 expression is sustained after multiple passages and dox withdrawal. (FIG. 4A) Representative immunostaining of PAX7 and MHC in differentiated iPSCs after 4 passages in the presence of dox. Scale bar = 200 μ m. (FIG. 4B) Representative immunostaining of PAX7 and myosin heavy chain (MHC) after

inducing differentiation by dox withdrawal for 7 days. Scale bar = 200 μ m. (FIG. 4C) Quantification of PAX7+ nuclei after 0 passages and after an average of 4 additional passages with dox or after dox withdrawal (mean \pm SEM, n = 3 independent experiments). (FIG. 4D) Representative immunostaining of the FLAG epitope for VP64-dCas9-VP64 after dox withdrawal for 7 days. Scale bar = 100 μ m.

[00024] FIGS. 5A-5D. VP64-dCas9-VP64 leads to sustained PAX7 expression and stable chromatin remodeling at target locus. (FIG. 5A) Human genomic track spanning the PAX7 TSS region depicting H3K4me3 and H3K27ac enrichment in human skeletal muscle myoblast (HSMM). Data from ENCODE (GEO:GSM733637; GEO:GSM733755). Black bars indicate ChIP-qPCR target regions. (FIG. 5B) Targeted activation of endogenous PAX7 induced significant enrichment of H3K4me3 and H3K27ac around the TSS in the presence of dox in proliferation conditions. (FIG. 5C) Enrichment of histone marks is sustained after 15 days in the absence of dox in proliferation conditions (mean \pm SEM, n = 3 independent replicates). (FIG. 5D) An N-terminal FLAG epitope tag was used to verify depletion of VP64-dCas9-VP64 after 15 days without dox, which was concomitant with sustained PAX7 protein expression.

[00025] FIGS. 6A-6E. Identification of endogenous vs. exogenous PAX7-induced global transcriptional changes. (FIG. 6A) An expression heatmap of sample-to-sample distances in the matrix using the whole gene expression profiles among the 4 groups and their replicates. (FIG. 6B) Heatmap showing differential expression of top 200 variable genes between all 4 groups after filtering genes with low read counts. The color bar indicates z-score. (FIG. 6C) Venn diagram of genes overexpressed in each group relative to gRNA only (fold-change > 2 and padj < 0.05) (FIG. 6D) GO Biological process terms of shared genes between the 3 groups derived from the Venn diagram in FIG. 4C. Term list was generated using Enrichr; P-values were computed using the Fisher exact test. (FIG. 6E) Expression profiles of select premyogenic, myogenic, and satellite cell marker genes from RNA-seq data (mean \pm SEM, n = 3 independent replicates). TPM: Transcripts Per Million.

[00026] FIGS. 7A-7C. Screening gRNAs for PAX7 activation with VP64-dCas9-VP64, related to FIGS. 1A-1G. (FIG. 7A) gRNA target sites relative to genome browser position of the human PAX7 gene. (FIG. 7B) Cells expressing VP64-dCas9-VP64 were treated for two days with CHIRON99021 and lipofected with PAX7-targeting gRNAs. Cells were harvested for qRT-PCR analysis after 6 days. gRNA 3, 4, 5 and 8 significantly upregulated PAX7 compared to mock transfection, but were not significantly different from each other. (FIG. 7C) Lentiviral transduction of gRNAs in paraxial mesoderm cells expressing P64-dCas9-VP64 and gRNAs for 1 week. gRNA 4 significantly outperformed the other gRNAs. P-values were

determined by one-way ANOVA followed by Tukey's post hoc test; $p < 0.05$ (mean \pm SEM, $n = 3$ independent replicates).

[00027] FIGS. 8A-8J. Characterization and transplantation of myogenic progenitors derived from H9 ESCs via VP64dCas9VP64-mediated activation of endogenous *PAX7* or exogenous *PAX7* cDNA expression, related to FIGS. 2A-2F and FIGS. 3A-3C. (FIG. 8A) Representative immunostaining of *PAX7* at 5 days postsort. Scale bar = 100 μm . (FIG. 8B) Growth curve of purified myogenic progenitors during post-sort expansion phase was monitored over 2 weeks. (FIG. 8C) Relative amount of total *PAX7* mRNA was determined by qRT-PCR using primers complementary to sequences present in the gene body. (FIG. 8D) Endogenous *PAX7* mRNA was detected using primers complementary to sequencing in the 3' UTR of either *PAX7-A* or *PAX7-B* isoforms. (FIG. 8E) The mRNA expression levels of myogenic markers *MYF5*, *MYOD*, and *MYOG* during the expansion phase. (FIG. 8F) Representative FACS analysis of CD29 and CD56 surface marker expression during the expansion phase. (FIG. 8G) Mean fluorescence intensity (MFI) of CD56 staining intensity across treatments. (FIG. 8H) Representative immunostaining of *PAX7* and MHC in differentiated H9 ESCs after 4 passages in the presence of dox. Scale bar = 200 μm . (FIG. 8I) Detection of human-derived fibers in VP64dCas9VP64-treated cells 1 month after intramuscular injection of 5×10^5 differentiated ESCs into NSG mice pre-injured with BaCl₂. Sections are stained with human-specific dystrophin and lamin A/C antibodies to mark donor-derived fibers and nuclei. Scale bar = 100 μm . (FIG. 8J) Identification of donor-derived satellite cells expressing *PAX7* and humanspecific lamin A/C. All P values were determined by one-way ANOVA followed by Tukey's post hoc test (mean \pm SEM, $n = 3$ independent replicates). Scale bar = 25 μm .

[00028] FIGS. 9A-9E. RNA-seq analysis, related to FIGS. 6A-6E. (FIG. 9A) Multidimensional scaling (MDS) of the top 500 differentially expressed genes. (FIG. 9B) Heatmap showing differential expression of top 50 variable genes between the 3 *PAX7*-expressing groups. The color bar indicates z-score. (FIG. 9C) Expression profile from selected genes overexpressed in response to cDNA encoding *PAX7-A* from RNA-seq (mean \pm SEM, $n = 3$ independent replicates). (FIG. 9D) GO biological process terms for genes specifically enriched in cells treated with VP64dCas9VP64 +gRNA, *PAX7-A* cDNA, or *PAX7-B* cDNA, corresponding to Venn diagram in FIG. 4C. (FIG. 9E) Additional expression profiles of known satellite cell surface markers.

DETAILED DESCRIPTION

[00029] Various DNA targeting systems and methods of use thereof are disclosed herein and may include, for example, a DNA targeting system using CRISPR/Cas, zinc fingers, or TALEs.

[00030] Advances in genome engineering technologies have established the type II clustered regularly spaced short palindromic repeat (CRISPR)/Cas9 system as a programmable transcriptional regulator capable of targeted activation or repression of endogenous genes. Mutations to the catalytic residues of the Cas9 protein results in a nuclease-null Cas9 (dCas9) that can be fused to various effector domains to exert their function on precise genomic loci defined by the guide RNA (gRNA). For example, fusion of dCas9 to the transactivation domain VP64 can potentially activate genes in their native chromosomal context when gRNAs are designed at target gene promoters. In contrast to ectopic expression of transgenes, activation of endogenous genes facilitates chromatin remodeling and induction of autonomously maintained gene networks. Targeting endogenous genes can also capture the full complexity of transcript isoforms, mRNA localization, and other effects of non-coding regulatory elements, which may be critical for proper cellular reprogramming. Cellular reprogramming may be achieved with CRISPR/Cas9-based transcriptional regulators in the context of somatic cell reprogramming as well as directed differentiation of pluripotent stem cells into various cell types. However, prior to the work detailed herein, there has not been demonstration of differentiation of hPSCs with CRISPR/Cas9-based transcriptional activators to generate cells capable of *in vivo* transplantation, engraftment, and tissue regeneration, or any attempt to generate myogenic progenitor cells via activation of the endogenous Pax7 gene.

[00031] Engineered CRISPR/Cas9-based transcriptional activators can potentially and specifically activate endogenous fate-determining genes to direct differentiation of pluripotent stem cells. As detailed herein, VP64-dCas9-VP64 was used to activate the endogenous myogenic transcription factor, Pax7, to directly reprogram human pluripotent stem cells and direct differentiation of them into skeletal muscle progenitors in both human ES and iPS cells. The functional skeletal muscle progenitor cells can be induced to differentiate *in vitro* and can also participate in regeneration of damaged muscles *in vivo* when transplanted into mice. Compared to the exogenous overexpression of Pax7 cDNA, endogenous activation results in the generation of more proliferative myogenic progenitors that can maintain Pax7 expression over multiple passages in serum-free conditions while maintaining the capacity for terminal myogenic differentiation. Transplantation of myogenic progenitors derived from endogenous activation of Pax7 into immunodeficient mice resulted in a greater number of

human dystrophin+ myofibers compared to exogenous Pax7 overexpression. The results detailed herein also reveal functional differences between myogenic progenitors generated via CRISPR-based endogenous activation of Pax7 and exogenous Pax7 cDNA overexpression. These studies demonstrate the utility of CRISPR/Cas9-based transcriptional activators for myogenic progenitor cell differentiation and their potential for cell therapy and musculoskeletal regenerative medicine. The methods of these studies may be applied using any DNA binding domain, such as a zinc finger protein or a TALE protein similarly to a Cas protein.

[00032] Described herein are systems for increasing expression of Pax7, which may include a Cas9 protein such as VP64-dCas9-VP64, and at least one guide RNA (gRNA) targeting Pax7 or a promoter or regulatory element of the Pax7 gene. Further provided herein are methods of activating endogenous myogenic transcription factor Pax7 in a cell, methods of differentiating a stem cell into a skeletal muscle progenitor cell, and methods of treating a subject in need thereof. The methods may include administering to the cell or subject the system for increasing expression of Pax7, or administering a cell transduced or transfected by the system.

1. Definitions

[00033] Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art. In case of conflict, the present document, including definitions, will control. Preferred methods and materials are described below, although methods and materials similar or equivalent to those described herein can be used in practice or testing of the present invention. All publications, patent applications, patents and other references mentioned herein are incorporated by reference in their entirety. The materials, methods, and examples disclosed herein are illustrative only and not intended to be limiting.

[00034] The terms "comprise(s)," "include(s)," "having," "has," "can," "contain(s)," and variants thereof, as used herein, are intended to be open-ended transitional phrases, terms, or words that do not preclude the possibility of additional acts or structures. The singular forms "a," "and" and "the" include plural references unless the context clearly dictates otherwise. The present disclosure also contemplates other embodiments "comprising," "consisting of" and "consisting essentially of," the embodiments or elements presented herein, whether explicitly set forth or not.

[00035] For the recitation of numeric ranges herein, each intervening number there between with the same degree of precision is explicitly contemplated. For example, for the range of 6-9, the numbers 7 and 8 are contemplated in addition to 6 and 9, and for the range 6.0-7.0, the number 6.0, 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9, and 7.0 are explicitly contemplated.

[00036] The term “about” or “approximately” as used herein as applied to one or more values of interest, refers to a value that is similar to a stated reference value. In certain aspects, the term “about” refers to a range of values that fall within 20%, 19%, 18%, 17%, 16%, 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, or less in either direction (greater than or less than) of the stated reference value unless otherwise stated or otherwise evident from the context (except where such number would exceed 100% of a possible value). Alternatively, “about” can mean within 3 or more than 3 standard deviations, per the practice in the art. Alternatively, such as with respect to biological systems or processes, the term “about” can mean within an order of magnitude, preferably within 5-fold, and more preferably within 2-fold, of a value.

[00037] “Adeno-associated virus” or “AAV” as used interchangeably herein refers to a small virus belonging to the genus Dependovirus of the Parvoviridae family that infects humans and some other primate species. AAV is not currently known to cause disease and consequently the virus causes a very mild immune response.

[00038] “Amino acid” as used herein refers to naturally occurring and non-natural synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code. Amino acids can be referred to herein by either their commonly known three-letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Amino acids include the side chain and polypeptide backbone portions.

[00039] “Binding region” as used herein refers to the region within a nuclease target region that is recognized and bound by the nuclease.

[00040] “Clustered Regularly Interspaced Short Palindromic Repeats” and “CRISPRs”, as used interchangeably herein, refers to loci containing multiple short direct repeats that are found in the genomes of approximately 40% of sequenced bacteria and 90% of sequenced archaea.

[00041] “Coding sequence” or “encoding nucleic acid” as used herein means the nucleic acids (RNA or DNA molecule) that comprise a nucleotide sequence which encodes a protein. The coding sequence can further include initiation and termination signals operably linked to regulatory elements including a promoter and polyadenylation signal capable of directing expression in the cells of an individual or mammal to which the nucleic acid is administered. The coding sequence may be codon optimize.

[00042] “Complement” or “complementary” as used herein means a nucleic acid can mean Watson-Crick (e.g., A-T/U and C-G) or Hoogsteen base pairing between nucleotides or nucleotide analogs of nucleic acid molecules. “Complementarity” refers to a property shared between two nucleic acid sequences, such that when they are aligned antiparallel to each other, the nucleotide bases at each position will be complementary.

[00043] The terms “control,” “reference level,” and “reference” are used herein interchangeably. The reference level may be a predetermined value or range, which is employed as a benchmark against which to assess the measured result. “Control group” as used herein refers to a group of control subjects. The predetermined level may be a cutoff value from a control group. The predetermined level may be an average from a control group. Cutoff values (or predetermined cutoff values) may be determined by Adaptive Index Model (AIM) methodology. Cutoff values (or predetermined cutoff values) may be determined by a receiver operating curve (ROC) analysis from biological samples of the patient group. ROC analysis, as generally known in the biological arts, is a determination of the ability of a test to discriminate one condition from another, e.g., to determine the performance of each marker in identifying a patient having CRC. A description of ROC analysis is provided in P.J. Heagerty et al. (*Biometrics* **2000**, *56*, 337-44), the disclosure of which is hereby incorporated by reference in its entirety. Alternatively, cutoff values may be determined by a quartile analysis of biological samples of a patient group. For example, a cutoff value may be determined by selecting a value that corresponds to any value in the 25th-75th percentile range, preferably a value that corresponds to the 25th percentile, the 50th percentile or the 75th percentile, and more preferably the 75th percentile. Such statistical analyses may be performed using any method known in the art and can be implemented through any number of commercially available software packages (e.g., from Analyse-it Software Ltd., Leeds, UK; StataCorp LP, College Station, TX; SAS Institute Inc., Cary, NC.). The healthy or normal levels or ranges for a target or for a protein activity may be defined in accordance with standard practice. A control may be an subject or cell without the system as detailed herein. A control may be a subject, or a sample therefrom, whose disease state is known. The subject, or sample therefrom, may be healthy, diseased,

diseased prior to treatment, diseased during treatment, or diseased after treatment, or a combination thereof.

[00044] “Fusion protein” as used herein refers to a chimeric protein created through the translation of two or more joined genes that originally coded for separate proteins. The translation of the fusion gene results in a single polypeptide with functional properties derived from each of the original separate proteins.

[00045] “Genetic construct” as used herein refers to the DNA or RNA molecules that comprise a polynucleotide that encodes a protein. The coding sequence includes initiation and termination signals operably linked to regulatory elements including a promoter and polyadenylation signal capable of directing expression in the cells of the individual to whom the nucleic acid molecule is administered. As used herein, the term “expressible form” refers to gene constructs that contain the necessary regulatory elements operable linked to a coding sequence that encodes a protein such that when present in the cell of the individual, the coding sequence will be expressed.

[00046] “Genome editing” or “gene editing” as used herein refers to changing a gene. Genome editing may include correcting or restoring a mutant gene. Genome editing may include knocking out a gene, such as a mutant gene or a normal gene. Genome editing may be used to treat disease or enhance muscle repair by changing the gene of interest.

[00047] “Identical” or “identity” as used herein in the context of two or more nucleic acids or polypeptide sequences means that the sequences have a specified percentage of residues that are the same over a specified region. The percentage may be calculated by optimally aligning the two sequences, comparing the two sequences over the specified region, determining the number of positions at which the identical residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the specified region, and multiplying the result by 100 to yield the percentage of sequence identity. In cases where the two sequences are of different lengths or the alignment produces one or more staggered ends and the specified region of comparison includes only a single sequence, the residues of single sequence are included in the denominator but not the numerator of the calculation. When comparing DNA and RNA, thymine (T) and uracil (U) may be considered equivalent. Identity may be performed manually or by using a computer sequence algorithm such as BLAST or BLAST 2.0.

[00048] “Mutant gene” or “mutated gene” as used interchangeably herein refers to a gene that has undergone a detectable mutation. A mutant gene has undergone a change, such as the loss, gain, or exchange of genetic material, which affects the normal transmission and expression of the gene. A “disrupted gene” as used herein refers to a mutant gene that has a mutation that causes a premature stop codon. The disrupted gene product is truncated relative to a full-length undisrupted gene product.

[00049] “Normal gene” as used herein refers to a gene that has not undergone a change, such as a loss, gain, or exchange of genetic material. The normal gene undergoes normal gene transmission and gene expression. For example, a normal gene may be a wild-type gene.

[00050] “Nucleic acid” or “oligonucleotide” or “polynucleotide” as used herein means at least two nucleotides covalently linked together. The depiction of a single strand also defines the sequence of the complementary strand. Thus, a polynucleotide also encompasses the complementary strand of a depicted single strand. Many variants of a polynucleotide may be used for the same purpose as a given polynucleotide. Thus, a polynucleotide also encompasses substantially identical polynucleotides and complements thereof. A single strand provides a probe that may hybridize to a target sequence under stringent hybridization conditions. Thus, a polynucleotide also encompasses a probe that hybridizes under stringent hybridization conditions. Polynucleotides may be single stranded or double stranded, or may contain portions of both double stranded and single stranded sequence. The polynucleotide can be nucleic acid, natural or synthetic, DNA, genomic DNA, cDNA, RNA, or a hybrid, where the polynucleotide can contain combinations of deoxyribo- and ribo-nucleotides, and combinations of bases including uracil, adenine, thymine, cytosine, guanine, inosine, xanthine hypoxanthine, isocytosine, and isoguanine. Polynucleotides can be obtained by chemical synthesis methods or by recombinant methods.

[00051] “Open reading frame” refers to a stretch of codons that begins with a start codon and ends at a stop codon. In eukaryotic genes with multiple exons, introns are removed, and exons are then joined together after transcription to yield the final mRNA for protein translation. An open reading frame may be a continuous stretch of codons. In some embodiments, the open reading frame only applies to spliced mRNAs, not genomic DNA, for expression of a protein.

[00052] “Operably linked” as used herein means that expression of a gene is under the control of a promoter with which it is spatially connected. A promoter may be positioned 5' (upstream) or 3' (downstream) of a gene under its control. The distance between the

promoter and a gene may be approximately the same as the distance between that promoter and the gene it controls in the gene from which the promoter is derived. As is known in the art, variation in this distance may be accommodated without loss of promoter function.

[00053] “Partially-functional” as used herein describes a protein that is encoded by a mutant gene and has less biological activity than a functional protein but more than a non-functional protein.

[00054] A “peptide” or “polypeptide” is a linked sequence of two or more amino acids linked by peptide bonds. The polypeptide can be natural, synthetic, or a modification or combination of natural and synthetic. Peptides and polypeptides include proteins such as binding proteins, receptors, and antibodies. The terms “polypeptide”, “protein,” and “peptide” are used interchangeably herein. “Primary structure” refers to the amino acid sequence of a particular peptide. “Secondary structure” refers to locally ordered, three dimensional structures within a polypeptide. These structures are commonly known as domains, e.g., enzymatic domains, extracellular domains, transmembrane domains, pore domains, and cytoplasmic tail domains. “Domains” are portions of a polypeptide that form a compact unit of the polypeptide and are typically 15 to 350 amino acids long. Exemplary domains include domains with enzymatic activity or ligand binding activity. Typical domains are made up of sections of lesser organization such as stretches of beta-sheet and alpha-helices. “Tertiary structure” refers to the complete three dimensional structure of a polypeptide monomer. “Quaternary structure” refers to the three dimensional structure formed by the noncovalent association of independent tertiary units. A “motif” is a portion of a polypeptide sequence and includes at least two amino acids. A motif may be 2 to 20, 2 to 15, or 2 to 10 amino acids in length. In some embodiments, a motif includes 3, 4, 5, 6, or 7 sequential amino acids. A domain may be comprised of a series of the same type of motif.

[00055] “Premature stop codon” or “out-of-frame stop codon” as used interchangeably herein refers to nonsense mutation in a sequence of DNA, which results in a stop codon at location not normally found in the wild-type gene. A premature stop codon may cause a protein to be truncated or shorter compared to the full-length version of the protein.

[00056] “Promoter” as used herein means a synthetic or naturally-derived molecule which is capable of conferring, activating or enhancing expression of a nucleic acid in a cell. A promoter may comprise one or more specific transcriptional regulatory sequences to further enhance expression and/or to alter the spatial expression and/or temporal expression of same. A promoter may also comprise distal enhancer or repressor elements, which may be located as much as several thousand base pairs from the start site of transcription. A

promoter may be derived from sources including viral, bacterial, fungal, plants, insects, and animals. A promoter may regulate the expression of a gene component constitutively, or differentially with respect to cell, the tissue or organ in which expression occurs or, with respect to the developmental stage at which expression occurs, or in response to external stimuli such as physiological stresses, pathogens, metal ions, or inducing agents.

Representative examples of promoters include the bacteriophage T7 promoter, bacteriophage T3 promoter, SP6 promoter, lac operator-promoter, tac promoter, SV40 late promoter, SV40 early promoter, RSV-LTR promoter, CMV IE promoter, SV40 early promoter or SV40 late promoter, human U6 (hU6) promoter, and CMV IE promoter.

[00057] The term “recombinant” when used with reference to, for example, a cell, nucleic acid, protein, or vector, indicates that the cell, nucleic acid, protein, or vector, has been modified by the introduction of a heterologous nucleic acid or protein or the alteration of a native nucleic acid or protein, or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found within the native (naturally occurring) form of the cell or express a second copy of a native gene that is otherwise normally or abnormally expressed, under expressed, or not expressed at all.

[00058] “Sample” or “test sample” as used herein can mean any sample in which the presence and/or level of a target is to be detected or determined or any sample comprising a DNA targeting system or component thereof as detailed herein. Samples may include liquids, solutions, emulsions, or suspensions. Samples may include a medical sample. Samples may include any biological fluid or tissue, such as blood, whole blood, fractions of blood such as plasma and serum, muscle, interstitial fluid, sweat, saliva, urine, tears, synovial fluid, bone marrow, cerebrospinal fluid, nasal secretions, sputum, amniotic fluid, bronchoalveolar lavage fluid, gastric lavage, emesis, fecal matter, lung tissue, peripheral blood mononuclear cells, total white blood cells, lymph node cells, spleen cells, tonsil cells, cancer cells, tumor cells, bile, digestive fluid, skin, or combinations thereof. In some embodiments, the sample comprises an aliquot. In other embodiments, the sample comprises a biological fluid. Samples can be obtained by any means known in the art. The sample can be used directly as obtained from a patient or can be pre-treated, such as by filtration, distillation, extraction, concentration, centrifugation, inactivation of interfering components, addition of reagents, and the like, to modify the character of the sample in some manner as discussed herein or otherwise as is known in the art.

[00059] “Spacers” and “spacer region” as used interchangeably herein refers to the region within a TALE or zinc finger target region that is between, but not a part of, the binding regions for two TALEs or zinc finger proteins.

[00060] "Subject" or "patient" as used herein can mean an animal that wants or is in need of the herein described compositions or methods. The subject may be a human or a non-human. The subject may be any vertebrate. The subject may be a mammal. The mammal may be a primate or a non-primate. The mammal can be a non-primate such as, for example, cow, pig, camel, llama, hedgehog, anteater, platypus, elephant, alpaca, horse, goat, rabbit, sheep, hamster, guinea pig, cat, dog, rat, and mouse. The mammal can be a primate such as a human. The mammal can be a non-human primate such as, for example, monkey, cynomolgous monkey, rhesus monkey, chimpanzee, gorilla, orangutan, and gibbon. The subject may be of any age or stage of development, such as, for example, an adult, an adolescent, or an infant. The subject may be male. The subject may be female. In some embodiments, the subject has a specific genetic marker. The subject may be undergoing other forms of treatment.

[00061] "Substantially identical" can mean that a first and second amino acid or polynucleotide sequence are at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% over a region of 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, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100 amino acids or nucleotides, respectively.

[00062] "Transcription activator-like effector" or "TALE" refers to a protein structure that recognizes and binds to a particular DNA sequence. The "TALE DNA-binding domain" refers to a DNA-binding domain that includes an array of tandem 33-35 amino acid repeats, also known as RVD modules, each of which specifically recognizes a single base pair of DNA. RVD modules may be arranged in any order to assemble an array that recognizes a defined sequence. A binding specificity of a TALE DNA-binding domain is determined by the RVD array followed by a single truncated repeat of 20 amino acids. "Repeat variable diresidue" or "RVD" refers to a pair of adjacent amino acid residues within a DNA recognition motif (also known as "RVD module"), which includes 33-35 amino acids, of a TALE DNA-binding domain. The RVD determines the nucleotide specificity of the RVD module. RVD modules may be combined to produce an RVD array. The "RVD array length" as used herein refers to the number of RVD modules that corresponds to the length of the nucleotide sequence within the TALEN target region that is recognized by a TALEN, *i.e.*, the binding region. A TALE DNA-binding domain may have 12 to 27 RVD modules, each of which contains an RVD and recognizes a single base pair of DNA. Specific RVDs have been identified that recognize each of the four possible DNA nucleotides (A, T, C, and G). Because the TALE DNA-binding domains are modular, repeats that recognize the four different DNA nucleotides may be linked together to recognize any particular DNA sequence.

These targeted DNA-binding domains may then be combined with catalytic domains to create functional enzymes, including artificial transcription factors, methyltransferases, integrases, nucleases, and recombinases.

[00063] “Target gene” as used herein refers to any nucleotide sequence encoding a known or putative gene product. The target gene may be a mutated gene involved in a genetic disease. In certain embodiments, the target gene is Pax7 or a transcription factor for Pax7 or a regulatory element for Pax7.

[00064] “Target region” as used herein refers to the region of the target gene to which the CRISPR/Cas9-based gene editing system is designed to bind.

[00065] “Transgene” as used herein refers to a gene or genetic material containing a gene sequence that has been isolated from one organism and is introduced into a different organism. This non-native segment of DNA may retain the ability to produce RNA or protein in the transgenic organism, or it may alter the normal function of the transgenic organism's genetic code. The introduction of a transgene has the potential to change the phenotype of an organism.

[00066] “Treatment” or “treating,” when referring to protection of a subject from a disease, means suppressing, repressing, ameliorating, or completely eliminating the disease. Preventing the disease involves administering a composition of the present invention to a subject prior to onset of the disease. Suppressing the disease involves administering a composition of the present invention to a subject after induction of the disease but before its clinical appearance. Repressing or ameliorating the disease involves administering a composition of the present invention to a subject after clinical appearance of the disease.

[00067] “Variant” used herein with respect to a polynucleotide means (i) a portion or fragment of a referenced nucleotide sequence; (ii) the complement of a referenced nucleotide sequence or portion thereof; (iii) a nucleic acid that is substantially identical to a referenced nucleic acid or the complement thereof; or (iv) a nucleic acid that hybridizes under stringent conditions to the referenced nucleic acid, complement thereof, or a sequences substantially identical thereto.

[00068] “Variant” with respect to a peptide or polypeptide that differs in amino acid sequence by the insertion, deletion, or conservative substitution of amino acids, but retain at least one biological activity. Variant may also mean a protein with an amino acid sequence that is substantially identical to a referenced protein with an amino acid sequence that retains at least one biological activity. Representative examples of “biological activity”

include the ability to be bound by a specific antibody or polypeptide or to promote an immune response. Variant can mean a functional fragment thereof. Variant can also mean multiple copies of a polypeptide. The multiple copies can be in tandem or separated by a linker. A conservative substitution of an amino acid, i.e., replacing an amino acid with a different amino acid of similar properties (e.g., hydrophilicity, degree and distribution of charged regions) is recognized in the art as typically involving a minor change. These minor changes may be identified, in part, by considering the hydropathic index of amino acids, as understood in the art (Kyte et al., *J. Mol. Biol.* **1982**, *157*, 105-132). The hydropathic index of an amino acid is based on a consideration of its hydrophobicity and charge. It is known in the art that amino acids of similar hydropathic indexes may be substituted and still retain protein function. In one aspect, amino acids having hydropathic indexes of ± 2 are substituted. The hydrophilicity of amino acids may also be used to reveal substitutions that would result in proteins retaining biological function. A consideration of the hydrophilicity of amino acids in the context of a peptide permits calculation of the greatest local average hydrophilicity of that peptide. Substitutions may be performed with amino acids having hydrophilicity values within ± 2 of each other. Both the hydrophobicity index and the hydrophilicity value of amino acids are influenced by the particular side chain of that amino acid. Consistent with that observation, amino acid substitutions that are compatible with biological function are understood to depend on the relative similarity of the amino acids, and particularly the side chains of those amino acids, as revealed by the hydrophobicity, hydrophilicity, charge, size, and other properties.

[00069] "Vector" as used herein means a nucleic acid sequence containing an origin of replication. A vector may be a viral vector, bacteriophage, bacterial artificial chromosome or yeast artificial chromosome. A vector may be a DNA or RNA vector. A vector may be a self-replicating extrachromosomal vector, and preferably, is a DNA plasmid. For example, the vector may encode a Cas9 protein and at least one gRNA molecule.

[00070] "Zinc finger" as used herein refers to a protein that recognizes and binds to DNA sequences. The zinc finger domain is the most common DNA-binding motif in the human proteome. A single zinc finger contains approximately 30 amino acids, and the domain typically functions by binding 3 consecutive base pairs of DNA via interactions of a single amino acid side chain per base pair.

[00071] Unless otherwise defined herein, scientific and technical terms used in connection with the present disclosure shall have the meanings that are commonly understood by those of ordinary skill in the art. For example, any nomenclatures used in connection with, and techniques of, cell and tissue culture, molecular biology, immunology,

microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those that are well known and commonly used in the art. The meaning and scope of the terms should be clear; in the event however of any latent ambiguity, definitions provided herein take precedent over any dictionary or extrinsic definition. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular.

2. Pax7

[00072] Pax7 (paired box gene 7) is a protein that acts as a myogenic transcription factor. Pax7 may be factor in the expression of neural crest markers such as, for example, Slug, Sox9, Sox10, and HNK-1. Pax7 may be expressed in the palatal shelf of the maxilla, Meckel's cartilage, mesencephalon, nasal cavity, nasal epithelium, nasal capsule, and pons. Pax7 can bind to DNA as a heterodimer with Pax3. Pax7 may also interact with PAXBP1 and/or DAXX.

[00073] Pax7 is a transcription factor that plays a role in myogenesis through regulation of muscle precursor cells proliferation. Skeletal muscle growth and regeneration are attributed to satellite cells, which are muscle stem cells resident beneath the basal lamina that surrounds each myofibre. Quiescent satellite cells express the transcription factor Pax7, and when activated, the quiescent satellite cells may coexpress Pax7 with MyoD. Most cells may then proliferate, downregulate Pax7, and differentiate. By contrast, other cells may maintain expression of Pax7 but lose expression of MyoD, and return to a state resembling quiescence. Upon expression or activation of Pax7 in a stem cell, the stem cell may differentiate into a skeletal muscle progenitor cell. The stem cell may be, for example, an induced pluripotent stem cell (iPSC) or an embryonic stem cell (ESC). The stem cell may be induced into myogenic differentiation. In some embodiments, expression or activation of Pax7 results in expression of Myf5, MyoD, MyoG, or a combination thereof. In some embodiments, expression or activation of Pax7 results in muscle regeneration. In some embodiments, expression or activation of Pax7 results in an increase of muscle stem cells, which may contribute to dystrophin+ fibers.

3. CRISPR/Cas-based Gene Editing System

[00074] Provided herein are genetic constructs for genome editing, genomic alteration, or altering gene expression of a gene, for example, a gene encoding Pax7. The genetic constructs include at least one gRNA that targets a gene sequence. The disclosed gRNAs can be included in a CRISPR/Cas9-based gene editing system to target regions in the Pax7

gene, or a promoter or regulatory element of the Pax7 gene, causing activation of endogenous expression of Pax7.

[00075] A CRISPR/Cas-based gene editing system may be specific for the Pax7 gene, or a promoter or regulatory element of the Pax7 gene. The CRISPR/Cas-based gene editing system may be a CRISPR/Cas9-based gene editing system specific for the Pax7 gene, or a promoter or regulatory element of the Pax7 gene. "Clustered Regularly Interspaced Short Palindromic Repeats" and "CRISPRs", as used interchangeably herein, refers to loci containing multiple short direct repeats that are found in the genomes of approximately 40% of sequenced bacteria and 90% of sequenced archaea. The CRISPR system is a microbial nuclease system involved in defense against invading phages and plasmids that provides a form of acquired immunity. The CRISPR loci in microbial hosts contain a combination of CRISPR-associated (Cas) genes as well as non-coding RNA elements capable of programming the specificity of the CRISPR-mediated nucleic acid cleavage. Short segments of foreign DNA, called spacers, are incorporated into the genome between CRISPR repeats, and serve as a 'memory' of past exposures. A Cas protein, such as a Cas9 protein, forms a complex with the 3' end of the sgRNA (also referred interchangeably herein as "gRNA"), and the protein-RNA pair recognizes its genomic target by complementary base pairing between the 5' end of the sgRNA sequence and a predefined 20 bp DNA sequence, known as the protospacer. This complex is directed to homologous loci of pathogen DNA via regions encoded within the crRNA, i.e., the protospacers, and protospacer-adjacent motifs (PAMs) within the pathogen genome. The non-coding CRISPR array is transcribed and cleaved within direct repeats into short crRNAs containing individual spacer sequences, which direct Cas nucleases to the target site (protospacer). By simply exchanging the 20 bp recognition sequence of the expressed sgRNA, the Cas9 nuclease can be directed to new genomic targets. CRISPR spacers are used to recognize and silence exogenous genetic elements in a manner analogous to RNAi in eukaryotic organisms.

[00076] Three classes of CRISPR systems (Types I, II, and III effector systems) are known. The Type II effector system carries out targeted DNA double-strand break in four sequential steps, using a single effector enzyme such as Cas9, to cleave dsDNA. Compared to the Type I and Type III effector systems, which require multiple distinct effectors acting as a complex, the Type II effector system may function in alternative contexts such as eukaryotic cells. The Type II effector system consists of a long pre-crRNA, which is transcribed from the spacer-containing CRISPR locus, the Cas9 protein, and a tracrRNA, which is involved in pre-crRNA processing. The tracrRNAs hybridize to the repeat

regions separating the spacers of the pre-crRNA, thus initiating dsRNA cleavage by endogenous RNase III. This cleavage is followed by a second cleavage event within each spacer by Cas9, producing mature crRNAs that remain associated with the tracrRNA and Cas9, forming a Cas9:crRNA-tracrRNA complex.

[00077] The Cas9:crRNA-tracrRNA complex unwinds the DNA duplex and searches for sequences matching the crRNA to cleave. Target recognition occurs upon detection of complementarity between a "protospacer" sequence in the target DNA and the remaining spacer sequence in the crRNA. Cas9 mediates cleavage of target DNA if a correct protospacer-adjacent motif (PAM) is also present at the 3' end of the protospacer. For protospacer targeting, the sequence must be immediately followed by the protospacer-adjacent motif (PAM), a short sequence recognized by the Cas9 nuclease that is required for DNA cleavage. Different Type II systems have differing PAM requirements. The *Streptococcus pyogenes* CRISPR system may have the PAM sequence for this Cas9 (SpCas9) as 5'-NRG-3', where R is either A or G, and characterized the specificity of this system in human cells. A unique capability of the CRISPR/Cas9-based gene editing system is the straightforward ability to simultaneously target multiple distinct genomic loci by co-expressing a single Cas9 protein with two or more sgRNAs. For example, the *S. pyogenes* Type II system naturally prefers to use an "NGG" sequence, where "N" can be any nucleotide, but also accepts other PAM sequences, such as "NAG" in engineered systems (Hsu et al., *Nature Biotechnology* **2013** doi:10.1038/nbt.2647). Similarly, the Cas9 derived from *Neisseria meningitidis* (NmCas9) normally has a native PAM of NNNNGATT, but has activity across a variety of PAMs, including a highly degenerate NNNNGNNN PAM (Esvelt et al. *Nature Methods* **2013** doi:10.1038/nmeth.2681).

[00078] A Cas9 molecule of *S. aureus* recognizes the sequence motif NNGRR (R = A or G) (SEQ ID NO: 38) and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5, bp upstream from that sequence. In certain embodiments, a Cas9 molecule of *S. aureus* recognizes the sequence motif NNGRRN (R = A or G) (SEQ ID NO: 39) and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5, bp upstream from that sequence. In certain embodiments, a Cas9 molecule of *S. aureus* recognizes the sequence motif NNGRRT (R = A or G) (SEQ ID NO: 40) and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5, bp upstream from that sequence. In certain embodiments, a Cas9 molecule of *S. aureus* recognizes the sequence motif NNGRRV (R = A or G) (SEQ ID NO: 41) and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5, bp upstream from that sequence. In the aforementioned embodiments, N can be any

nucleotide residue, e.g., any of A, G, C, or T. Cas9 molecules can be engineered to alter the PAM specificity of the Cas9 molecule.

[00079] An engineered form of the Type II effector system of *S. pyogenes* was shown to function in human cells for genome engineering. In this system, the Cas9 protein was directed to genomic target sites by a synthetically reconstituted "guide RNA" ("gRNA", also used interchangeably herein as a chimeric single guide RNA ("sgRNA")), which is a crRNA-tracrRNA fusion that obviates the need for RNase III and crRNA processing in general. Provided herein are CRISPR/Cas9-based engineered systems for use in genome editing and treating genetic diseases. The CRISPR/Cas9-based engineered systems can be designed to target any gene, including genes involved in a genetic disease, aging, tissue regeneration, or wound healing. The CRISPR/Cas9-based gene editing systems can include a Cas9 protein or Cas9 fusion protein and at least one gRNA. In certain embodiments, the system comprises two gRNA molecules. The Cas9 fusion protein may, for example, include a domain that has a different activity than what is endogenous to Cas9, such as a transactivation domain.

[00080] The target gene (e.g., the Pax7 gene, or a regulatory element of the Pax7 gene) can be involved in differentiation of a cell or any other process in which activation of a gene can be desired, or can have a mutation such as a frameshift mutation or a nonsense mutation. In some embodiments, the target or target gene includes a regulatory element of the Pax7 gene. The CRISPR/Cas9-based gene editing system may or may not mediate off-target changes to protein-coding regions of the genome. The CRISPR/Cas9-based gene editing system may bind and recognize a target region. The targeted gene may be the Pax7 gene.

a. Cas Protein

[00081] The CRISPR/Cas-based gene editing system can include a Cas protein or a Cas fusion protein. In some embodiments, the Cas protein is a Cas12 protein (also referred to as Cpf1), such as a Cas12a protein. The Cas12 protein can be from any bacterial or archaea species, including, but not limited to, *Francisella novicida*, *Acidaminococcus* sp., *Lachnospiraceae* sp., and *Prevotella* sp. In some embodiments, the Cas protein is a Cas9 protein. Cas9 protein is an endonuclease that may cleave nucleic acid and is encoded by the CRISPR loci and is involved in the Type II CRISPR system. The Cas9 protein can be from any bacterial or archaea species, including, but not limited to, *Streptococcus pyogenes*, *Staphylococcus aureus* (*S. aureus*), *Acidovorax avenae*, *Actinobacillus pleuropneumoniae*, *Actinobacillus succinogenes*, *Actinobacillus suis*, *Actinomyces* sp., *Cytophila denitrificans*,

Aminomonas paucivorans, *Bacillus cereus*, *Bacillus smithii*, *Bacillus thuringiensis*, *Bacteroides* sp., *Blastopirellula marina*, *Bradyrhizobium* sp., *Brevibacillus laterosporus*, *Campylobacter coli*, *Campylobacter jejuni*, *Campylobacter lari*, *Candidatus Puniceispirillum*, *Clostridium cellulolyticum*, *Clostridium perfringens*, *Corynebacterium accolens*, *Corynebacterium diphtheria*, *Corynebacterium matruchotii*, *Dinoroseobacter shibae*, *Eubacterium dolichum*, *gamma proteobacterium*, *Gluconacetobacter diazotrophicus*, *Haemophilus parainfluenzae*, *Haemophilus sputorum*, *Helicobacter canadensis*, *Helicobacter cinaedi*, *Helicobacter mustelae*, *Ilyobacter polytropus*, *Kingella kingae*, *Lactobacillus crispatus*, *Listeria ivanovii*, *Listeria monocytogenes*, *Listeriaceae bacterium*, *Methylocystis* sp., *Methylosinus trichosporium*, *Mobiluncus mulieris*, *Neisseria bacilliformis*, *Neisseria cinerea*, *Neisseria flavescens*, *Neisseria lactamica*, *Neisseria* sp., *Neisseria wadsworthii*, *Nitrosomonas* sp., *Parvibaculum lavamentivorans*, *Pasteurella multocida*, *Phascolarctobacterium succinatutens*, *Ralstonia syzygii*, *Rhodopseudomonas palustris*, *Rhodovulum* sp., *Simonsiella muelleri*, *Sphingomonas* sp., *Sporolactobacillus vineae*, *Staphylococcus lugdunensis*, *Streptococcus* sp., *Subdoligranulum* sp., *Tistrella mobilis*, *Treponema* sp., or *Verminephrobacter eiseniae*. In certain embodiments, the Cas9 molecule is a *Streptococcus pyogenes* Cas9 molecule (also referred herein as "SpCas9"). In certain embodiments, the Cas9 molecule is a *Staphylococcus aureus* Cas9 molecule (also referred herein as "SaCas9").

[00082] A Cas molecule or a Cas fusion protein can interact with one or more gRNA molecules and, in concert with the gRNA molecule(s), can localize to a site which comprises a target domain, and in certain embodiments, a PAM sequence. The ability of a Cas molecule or a Cas fusion protein to recognize a PAM sequence can be determined, e.g., using a transformation assay as known in the art.

[00083] In certain embodiments, the ability of a Cas molecule or a Cas fusion protein to interact with and cleave a target nucleic acid is protospacer-adjacent motif (PAM) sequence dependent. A PAM sequence is a sequence in the target nucleic acid. In certain embodiments, cleavage of the target nucleic acid occurs upstream from the PAM sequence. Cas molecules from different bacterial species can recognize different sequence motifs (e.g., PAM sequences). In certain embodiments, a Cas12 molecule of *Francisella novicida* recognizes the sequence motif TTTN (SEQ ID NO: 56). In certain embodiments, a Cas9 molecule of *S. pyogenes* recognizes the sequence motif NGG and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5, bp upstream from that sequence. In certain embodiments, a Cas9 molecule of *S. thermophilus* recognizes the sequence motif NGGNG (SEQ ID NO: 35) and/or NNAGAAW (W = A or T) (SEQ ID NO: 36) and directs

cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5, bp upstream from these sequences. In certain embodiments, a Cas9 molecule of *S. mutans* recognizes the sequence motif NGG (SEQ ID NO: 31) and/or NAAR (R = A or G) (SEQ ID NO: 37) and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5 bp, upstream from this sequence. In certain embodiments, a Cas9 molecule of *S. aureus* recognizes the sequence motif NNGRR (R = A or G) (SEQ ID NO: 38) and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5, bp upstream from that sequence. In certain embodiments, a Cas9 molecule of *S. aureus* recognizes the sequence motif NNGRRN (R = A or G) (SEQ ID NO: 39) and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5, bp upstream from that sequence. In certain embodiments, a Cas9 molecule of *S. aureus* recognizes the sequence motif NNGRRRT (R = A or G) (SEQ ID NO: 40) and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5, bp upstream from that sequence. In certain embodiments, a Cas9 molecule of *S. aureus* recognizes the sequence motif NNGRRV (R = A or G; V = A or C or G) (SEQ ID NO: 41) and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5, bp upstream from that sequence. In the aforementioned embodiments, N can be any nucleotide residue, e.g., any of A, G, C, or T. Cas9 molecules can be engineered to alter the PAM specificity of the Cas9 molecule.

[00084] In certain embodiments, the vector encodes at least one Cas9 molecule that recognizes a Protospacer Adjacent Motif (PAM) of either NNGRRRT (SEQ ID NO: 40) or NNGRRV (SEQ ID NO: 41). In certain embodiments, the at least one Cas9 molecule is an *S. aureus* Cas9 molecule. In certain embodiments, the at least one Cas9 molecule is a mutant *S. aureus* Cas9 molecule.

[00085] The Cas protein can be mutated so that the nuclease activity is inactivated. An inactivated Cas9 protein ("iCas9", also referred to as "dCas9") with no endonuclease activity has been targeted to genes in bacteria, yeast, and human cells by gRNAs to silence gene expression through steric hindrance. Exemplary mutations with reference to the *S. pyogenes* Cas9 sequence include: D10A, E762A, H840A, N854A, N863A, and/or D986A. Exemplary mutations with reference to the *S. aureus* Cas9 sequence include D10A and N580A. In certain embodiments, the Cas9 molecule is a mutant *S. aureus* Cas9 molecule. In some embodiments, the dCas9 is a Cas9 molecule that includes at least two mutations selected from D10A, E762A, H840A, N854A, N863A, and/or D986A, with reference to the *S. pyogenes* Cas9 sequence. In some embodiments, the Cas protein is a dCas9 protein. In some embodiments, the Cas protein is a dCas12 protein.

[00086] In certain embodiments, the mutant *S. aureus* Cas9 molecule comprises a D10A mutation. The nucleotide sequence encoding this mutant *S. aureus* Cas9 is set forth in SEQ ID NO: 50.

[00087] In certain embodiments, the mutant *S. aureus* Cas9 molecule comprises a N580A mutation. The nucleotide sequence encoding this mutant *S. aureus* Cas9 molecule is set forth in SEQ ID NO: 51.

[00088] A polynucleotide encoding a Cas molecule can be a synthetic polynucleotide. For example, the synthetic polynucleotide can be chemically modified. The synthetic polynucleotide can be codon optimized, e.g., at least one non-common codon or less-common codon has been replaced by a common codon. For example, the synthetic polynucleotide can direct the synthesis of an optimized messenger mRNA, e.g., optimized for expression in a mammalian expression system, e.g., described herein.

[00089] Additionally or alternatively, a nucleic acid encoding a Cas molecule or Cas polypeptide may comprise a nuclear localization sequence (NLS). Nuclear localization sequences are known in the art. An exemplary codon optimized nucleic acid sequence encoding a Cas9 molecule of *S. pyogenes* is set forth in SEQ ID NO: 42. The corresponding amino acid sequence of an *S. pyogenes* Cas9 molecule is set forth in SEQ ID NO: 43.

[00090] Exemplary codon optimized nucleic acid sequences encoding a Cas9 molecule of *S. aureus*, and optionally containing nuclear localization sequences (NLSs), are set forth in SEQ ID NOs: 44-48, 52, and 53, which are provided below. Another exemplary codon optimized nucleic acid sequence encoding a Cas9 molecule of *S. aureus* comprises the nucleotides 1293-4451 of SEQ ID NO: 55. An amino acid sequence of an *S. aureus* Cas9 molecule is set forth in SEQ ID NO: 49. An amino acid sequence of a *Streptococcus pyogenes* Cas9 (with D10A, H849A mutations) is set forth in SEQ ID NO: 54.

b. Fusion Protein

[00091] Alternatively or additionally, the CRISPR/Cas-based gene editing system can include a fusion protein. The fusion protein can comprise two heterologous polypeptide domains, wherein the first polypeptide domain comprises a DNA binding protein such as a Cas protein, a zinc finger protein, or a TALE protein, and the second polypeptide domain has an activity such as transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, nuclease activity, nucleic acid association activity, methylase activity, or demethylase activity. The fusion protein can include a first polypeptide domain such as a Cas9 protein or a mutated Cas9 protein, fused

to a second polypeptide domain that has an activity such as transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, nuclease activity, nucleic acid association activity, methylase activity, or demethylase activity. In some embodiments, the second polypeptide domain has transcription activation activity. In some embodiments, the second polypeptide domain comprises a synthetic transcription factor. The fusion protein may include one second polypeptide domain. The fusion protein may include two of the second polypeptide domains. For example, the fusion protein may include a second polypeptide domain at the N-terminal end of the first polypeptide domain as well as a second polypeptide domain at the C-terminal end of the first polypeptide domain. In other embodiments, the fusion protein may include a single first polypeptide domain and more than one (for example, two or three) second polypeptide domains in tandem.

i) Transcription Activation Activity

[00092] The second polypeptide domain can have transcription activation activity, i.e., a transactivation domain. For example, gene expression of endogenous mammalian genes, such as human genes, can be achieved by targeting a fusion protein of a first polypeptide domain, such as dCas9 or dCas12, and a transactivation domain to mammalian promoters via combinations of gRNAs. The transactivation domain can include a VP 16 protein, multiple VP 16 proteins, such as a VP48 domain or VP64 domain, p65 domain of NF kappa B transcription activator activity, or p300. For example, the fusion protein may be dCas9-VP64. In other embodiments, the Cas9 protein may be VP64-dCas9-VP64 (SEQ ID NO: 57, encoded by SEQ ID NO: 58). In other embodiments, the fusion protein that activates transcription may be dCas9-p300. In some embodiments, p300 may comprise a polypeptide of SEQ ID NO: 59 or SEQ ID NO: 60.

ii) Transcription Repression Activity

[00093] The second polypeptide domain can have transcription repression activity. The second polypeptide domain can have a Kruppel associated box activity, such as a KRAB domain, ERF repressor domain activity, Mxil repressor domain activity, SID4X repressor domain activity, Mad-SID repressor domain activity, or TATA box binding protein activity. For example, the fusion protein may be dCas9-KRAB.

iii) Transcription Release Factor Activity

[00094] The second polypeptide domain can have transcription release factor activity. The second polypeptide domain can have eukaryotic release factor 1 (ERF1) activity or eukaryotic release factor 3 (ERF3) activity.

iv) Histone Modification Activity

[00095] The second polypeptide domain can have histone modification activity. The second polypeptide domain can have histone deacetylase, histone acetyltransferase, histone demethylase, or histone methyltransferase activity. The histone acetyltransferase may be p300 or CREB-binding protein (CBP) protein, or fragments thereof. For example, the fusion protein may be dCas9-p300. In some embodiments, p300 may comprise a polypeptide of SEQ ID NO: 59 or SEQ ID NO: 60.

v) Nuclease Activity

[00096] The second polypeptide domain can have nuclease activity that is different from the nuclease activity of the Cas9 protein. A nuclease, or a protein having nuclease activity, is an enzyme capable of cleaving the phosphodiester bonds between the nucleotide subunits of nucleic acids. Nucleases are usually further divided into endonucleases and exonucleases, although some of the enzymes may fall in both categories. Well known nucleases include deoxyribonuclease and ribonuclease.

vi) Nucleic Acid Association Activity

[00097] The second polypeptide domain can have nucleic acid association activity or nucleic acid binding protein-DNA-binding domain (DBD). A DBD is an independently folded protein domain that contains at least one motif that recognizes double- or single-stranded DNA. A DBD can recognize a specific DNA sequence (a recognition sequence) or have a general affinity to DNA. A nucleic acid association region may be selected from helix-turn-helix region, leucine zipper region, winged helix region, winged helix-turn-helix region, helix-loop-helix region, immunoglobulin fold, B3 domain, Zinc finger, HMG-box, Wor3 domain, TAL effector DNA-binding domain.

vii) Methylase Activity

[00098] The second polypeptide domain can have methylase activity, which involves transferring a methyl group to DNA, RNA, protein, small molecule, cytosine or adenine. In some embodiments, the second polypeptide domain includes a DNA methyltransferase.

viii) Demethylase Activity

[00099] The second polypeptide domain can have demethylase activity. The second polypeptide domain can include an enzyme that removes methyl (CH₃-) groups from nucleic acids, proteins (in particular histones), and other molecules. Alternatively, the second polypeptide can convert the methyl group to hydroxymethylcytosine in a mechanism for demethylating DNA. The second polypeptide can catalyze this reaction. For example, the second polypeptide that catalyzes this reaction can be Tet1.

c. gRNA

[000100] The CRISPR/Cas-based gene editing system includes at least one gRNA molecule. For example, the CRISPR/Cas-based gene editing system may include two gRNA molecules. The gRNA provides the targeting of a CRISPR/Cas-based gene editing system. The gRNA is a fusion of two noncoding RNAs: a crRNA and a tracrRNA. In some embodiments, the polynucleotide includes a crRNA, and/or a tracrRNA. The sgRNA may target any desired DNA sequence by exchanging the sequence encoding a 20 bp protospacer which confers targeting specificity through complementary base pairing with the desired DNA target. gRNA mimics the naturally occurring crRNA:tracrRNA duplex involved in the Type II Effector system. This duplex, which may include, for example, a 42-nucleotide crRNA and a 75-nucleotide tracrRNA, acts as a guide for the Cas9 to cleave the target nucleic acid. The "target region," "target sequence," or "protospacer," refers to the region of the target gene (e.g., a Pax7 gene) to which the CRISPR/Cas9-based gene editing system targets and binds. The portion of the gRNA that targets the target sequence in the genome may be referred to as the "targeting sequence" or "targeting portion" or "targeting domain." "Protospacer" or "gRNA spacer" may refer to the region of the target gene to which the CRISPR/Cas9-based gene editing system targets and binds; "protospacer" or "gRNA spacer" may also refer to the portion of the gRNA that is complementary to the targeted sequence in the genome. The gRNA may include a gRNA scaffold. A gRNA scaffold facilitates Cas9 binding to the gRNA and may facilitate endonuclease activity. The gRNA scaffold is a polynucleotide sequence that follows the portion of the gRNA corresponding to sequence that the gRNA targets. Together, the gRNA targeting portion and gRNA scaffold form one polynucleotide. The scaffold may comprise a polynucleotide sequence of SEQ ID NO: 85. The CRISPR/Cas9-based gene editing system may include at least one gRNA, wherein the gRNAs target different DNA sequences. The target DNA sequences may be overlapping. The target sequence or protospacer is followed by a PAM sequence at the 3' end of the protospacer in the genome. Different Type II systems have differing PAM requirements. For example, the *Streptococcus pyogenes* Type II system uses an "NGG"

sequence, where "N" can be any nucleotide. In some embodiments, the PAM sequence may be "NGG", where "N" can be any nucleotide. In some embodiments, the PAM sequence may be NNGRRT (SEQ ID NO: 40) or NNGRRV (SEQ ID NO: 41).

[000101] The number of gRNA molecule encoded by a genetic construct (e.g., an AAV vector) can be at least 1 gRNA, at least 2 different gRNA, at least 3 different gRNA at least 4 different gRNA, at least 5 different gRNA, at least 6 different gRNA, at least 7 different gRNA, at least 8 different gRNA, at least 9 different gRNA, at least 10 different gRNAs, at least 11 different gRNAs, at least 12 different gRNAs, at least 13 different gRNAs, at least 14 different gRNAs, at least 15 different gRNAs, at least 16 different gRNAs, at least 17 different gRNAs, at least 18 different gRNAs, at least 18 different gRNAs, at least 20 different gRNAs, at least 25 different gRNAs, at least 30 different gRNAs, at least 35 different gRNAs, at least 40 different gRNAs, at least 45 different gRNAs, or at least 50 different gRNAs. The number of gRNAs encoded by a presently disclosed vector can be between at least 1 gRNA to at least 50 different gRNAs, at least 1 gRNA to at least 45 different gRNAs, at least 1 gRNA to at least 40 different gRNAs, at least 1 gRNA to at least 35 different gRNAs, at least 1 gRNA to at least 30 different gRNAs, at least 1 gRNA to at least 25 different gRNAs, at least 1 gRNA to at least 20 different gRNAs, at least 1 gRNA to at least 16 different gRNAs, at least 1 gRNA to at least 12 different gRNAs, at least 1 gRNA to at least 8 different gRNAs, at least 1 gRNA to at least 4 different gRNAs, at least 4 gRNAs to at least 50 different gRNAs, at least 4 different gRNAs to at least 45 different gRNAs, at least 4 different gRNAs to at least 40 different gRNAs, at least 4 different gRNAs to at least 35 different gRNAs, at least 4 different gRNAs to at least 30 different gRNAs, at least 4 different gRNAs to at least 25 different gRNAs, at least 4 different gRNAs to at least 20 different gRNAs, at least 4 different gRNAs to at least 16 different gRNAs, at least 4 different gRNAs to at least 12 different gRNAs, at least 4 different gRNAs to at least 8 different gRNAs, at least 8 different gRNAs to at least 50 different gRNAs, at least 8 different gRNAs to at least 45 different gRNAs, at least 8 different gRNAs to at least 40 different gRNAs, at least 8 different gRNAs to at least 35 different gRNAs, 8 different gRNAs to at least 30 different gRNAs, at least 8 different gRNAs to at least 25 different gRNAs, 8 different gRNAs to at least 20 different gRNAs, at least 8 different gRNAs to at least 16 different gRNAs, or 8 different gRNAs to at least 12 different gRNAs. In certain embodiments, the genetic construct (e.g., an AAV vector) encodes one gRNA molecule, i.e., a first gRNA molecule, and optionally a Cas9 molecule. In certain embodiments, a first genetic construct (e.g., a first AAV vector) encodes one gRNA molecule, i.e., a first gRNA molecule, and optionally a Cas9 molecule, and a second genetic construct (e.g., a second AAV vector) encodes one gRNA molecule, i.e., a second gRNA molecule, and optionally a Cas9 molecule.

[000102] The gRNA molecule comprises a targeting domain, which is a polynucleotide sequence complementary to the target DNA sequence followed by a PAM sequence. The gRNA may comprise a "G" at the 5' end of the targeting domain or complementary polynucleotide sequence. The targeting domain of a gRNA molecule may comprise at least a 10 base pair, at least a 11 base pair, at least a 12 base pair, at least a 13 base pair, at least a 14 base pair, at least a 15 base pair, at least a 16 base pair, at least a 17 base pair, at least a 18 base pair, at least a 19 base pair, at least a 20 base pair, at least a 21 base pair, at least a 22 base pair, at least a 23 base pair, at least a 24 base pair, at least a 25 base pair, at least a 30 base pair, or at least a 35 base pair complementary polynucleotide sequence of the target DNA sequence followed by a PAM sequence. In certain embodiments, the targeting domain of a gRNA molecule has 19-25 nucleotides in length. In certain embodiments, the targeting domain of a gRNA molecule is 20 nucleotides in length. In certain embodiments, the targeting domain of a gRNA molecule is 21 nucleotides in length. In certain embodiments, the targeting domain of a gRNA molecule is 22 nucleotides in length. In certain embodiments, the targeting domain of a gRNA molecule is 23 nucleotides in length.

[000103] The gRNA may target a region within or near the Pax7 gene, or within or near a regulatory element or promoter of the Pax7 gene. In certain embodiments, the gRNA can target at least one of exons, introns, the promoter region, the enhancer region, or the transcribed region of the gene. The gRNA may target Pax7 or a promoter or regulatory element of the Pax7 gene. In some embodiments, the gRNA targets a Pax7 promoter. The gRNA may include a targeting domain that comprises a polynucleotide sequence corresponding to at least one of SEQ ID NOs: 1-8 or 69-76 or 77-84, or a complement thereof or a variant thereof, as shown in **TABLE 1**. In some embodiments, the gRNA targets a polynucleotide sequence comprising the complement of at least one of SEQ ID NOs: 1-8. In some embodiments, the gRNA is encoded by a polynucleotide sequence comprising at least one of SEQ ID NOs: 1-8. In some embodiments, the gRNA comprises a polynucleotide sequence selected from SEQ ID NOs: 69-76. In some embodiments, the gRNA binds and targets a polynucleotide comprising a sequence selected from SEQ ID NOs: 77-84, respectively, in **TABLE 4**.

| TABLE 1. gRNAs that activate endogenous Pax7. | | | |
|--|----------------------|------------------|----------------------|
| SEQ ID NO | gRNA sequence | SEQ ID NO | gRNA |
| 1 | GGCCGGGGACTCGGCGGATC | 69 | GGCCGGGGACUCGGCGGAUC |

| | | | |
|---|----------------------|----|----------------------|
| 2 | TCCCCGGCTCGACCTCGTTT | 70 | UCCCCGGCUCGACCUCGUUU |
| 3 | CCAGGGCGCAAGGGAGCGG | 71 | CCAGGGCGCAAGGGAGCGG |
| 4 | TCCTCCGCTCCCTTGCGCCC | 72 | UCCUCCGCUCCCUUGCGCCC |
| 5 | GGGGGCGCGAGTGATCAGCT | 73 | GGGGGCGCGAGUGAUCAGCU |
| 6 | CGGGTTTCAGGGCTGGACGG | 74 | CGGGUUUCAGGGCUGGACGG |
| 7 | TGGTCCGGAGAAAGAAGGCG | 75 | UGGUCCGGAGAAAGAAGGCG |
| 8 | AGCGCCAGAGCGCGAGAGCG | 76 | AGCGCCAGAGCGCGAGAGCG |

TABLE 4. Target sequences of the gRNAs that activate endogenous Pax7

| SEQ ID NO | gRNA target sequence |
|-----------|----------------------|
| 77 | GATCCGCCGAGTCCCCGGCC |
| 78 | AAACGAGGTCGAGCCGGGGA |
| 79 | CCGCTCCCTTGCGCCCTGG |
| 80 | GGGCGCAAGGGAGCGGAGGA |
| 81 | AGCTGATCACTCGCGCCCCC |
| 82 | CCGTCCAGCCCTGAAACCCG |
| 83 | CGCCTTCTTTCTCCGGACCA |
| 84 | CGCTCTCGCGCTCTGGCGCT |

[000104] Single or multiplexed gRNAs can be designed to activate expression of Pax7, thereby differentiating a stem cell into a skeletal muscle progenitor cell. Following treatment with a construct or system as detailed herein, a stem cell may be differentiated into a skeletal muscle progenitor cell. Genetically corrected stem or patient cells may be transplanted into a subject.

d. DNA Targeting System

[000105] Further provided herein are DNA targeting systems or compositions that comprise such genetic constructs. The DNA targeting compositions include at least one gRNA molecule (e.g., two gRNA molecules) that targets a gene, as described above. The at least one gRNA molecule can bind and recognize a target region.

[000106] In some embodiments, the DNA targeting composition includes a first gRNA and a second gRNA. In some embodiments, the first gRNA molecule and the second gRNA molecule comprise different targeting domains.

[000107] The DNA targeting composition may further include at least one Cas molecule or a fusion protein. In some embodiments as detailed above, the DNA targeting composition further includes at least one dCas9 protein or fusion protein. In some embodiments, the Cas9 molecule or fusion protein recognizes a PAM of either NNGRRT (SEQ ID NO: 40) or NNGRRV (SEQ ID NO: 41). In some embodiments, the DNA targeting composition includes a nucleotide sequence set forth in SEQ ID NO: 55. In certain embodiments, the vector is configured to form a first and a second double strand break in a segment within or near the Pax7 gene.

[000108] The DNA targeting composition may further comprise a donor DNA or a transgene.

4. Genetic Constructs

[000109] The DNA targeting system, or one or more components thereof, may be encoded by or comprised within a genetic construct. Genetic constructs may include polynucleotides such as vectors and plasmids. The construct may be recombinant. In some embodiments, the genetic construct comprises a promoter that is operably linked to the polynucleotide encoding at least one gRNA molecule and/or a Cas molecule or fusion protein. In some embodiments, the genetic construct comprises a promoter that is operably linked to the polynucleotide encoding at least one gRNA molecule and/or a dCas molecule or fusion protein. In some embodiments, the genetic construct comprises a promoter that is operably linked to the polynucleotide encoding at least one gRNA molecule and/or a Cas9 molecule or fusion protein. In some embodiments, the promoter is operably linked to the polynucleotide encoding a first gRNA molecule, a second gRNA molecule, and/or a Cas9 molecule or fusion protein. The genetic construct may be present in the cell as a functioning extrachromosomal molecule. The genetic construct may be a linear minichromosome including centromere, telomeres, or plasmids or cosmids. The genetic construct may be transformed or transduced into a cell. The genetic construct may be formulated into any suitable type of delivery vehicle including, for example, a viral vector, lentiviral expression, mRNA electroporation, and lipid-mediated transfection. Further provided herein is a cell transformed or transduced with a DNA targeting system or component thereof as detailed herein. The cell may be, for example, a stem cell, or a fibroblast. In some embodiments, the stem cell is a pluripotent stem cells. In some embodiments, the fibroblast is a skin fibroblast.

[000110] Further provided herein is a viral delivery system. In some embodiments, the vector is an adeno-associated virus (AAV) vector. The AAV vector is a small virus belonging to the genus Dependovirus of the Parvoviridae family that infects humans and some other primate species. AAV vectors may be used to deliver CRISPR/Cas9-based gene editing systems using various construct configurations. For example, AAV vectors may deliver Cas9 and gRNA expression cassettes on separate vectors or on the same vector. Alternatively, if the small Cas9 proteins, derived from species such as *Staphylococcus aureus* or *Neisseria meningitidis*, are used then both the Cas9 and up to two gRNA expression cassettes may be combined in a single AAV vector within the 4.7 kb packaging limit.

[000111] In some embodiments, the AAV vector is a modified AAV vector. The modified AAV vector may have enhanced cardiac and/or skeletal muscle tissue tropism. The modified AAV vector may be capable of delivering and expressing the CRISPR/Cas9-based gene editing system in the cell of a mammal. For example, the modified AAV vector may be an AAV-SASTG vector (Piacentino et al. *Human Gene Therapy* **2012**, 23, 635–646). The modified AAV vector may be based on one or more of several capsid types, including AAV1, AAV2, AAV5, AAV6, AAV8, and AAV9. The modified AAV vector may be based on AAV2 pseudotype with alternative muscle-tropic AAV capsids, such as AAV2/1, AAV2/6, AAV2/7, AAV2/8, AAV2/9, AAV2.5, and AAV/SASTG vectors that efficiently transduce skeletal muscle or cardiac muscle by systemic and local delivery (Seto et al. *Current Gene Therapy* **2012**, 12, 139-151). The modified AAV vector may be AAV2i8G9 (Shen et al. *J. Biol. Chem.* **2013**, 288, 28814-28823).

5. Pharmaceutical Compositions

[000112] Further provided herein are pharmaceutical compositions comprising the above-described genetic constructs or DNA targeting systems. The DNA targeting systems, or at least one component thereof, as detailed herein may be formulated into pharmaceutical compositions in accordance with standard techniques well known to those skilled in the pharmaceutical art. The pharmaceutical compositions can be formulated according to the mode of administration to be used. In cases where pharmaceutical compositions are injectable pharmaceutical compositions, they are sterile, pyrogen free, and particulate free. An isotonic formulation is preferably used. Generally, additives for isotonicity may include sodium chloride, dextrose, mannitol, sorbitol and lactose. In some cases, isotonic solutions such as phosphate buffered saline are preferred. Stabilizers include gelatin and albumin. In some embodiments, a vasoconstriction agent is added to the formulation.

[000113] The composition may further comprise a pharmaceutically acceptable excipient. The pharmaceutically acceptable excipient may be functional molecules as vehicles, adjuvants, carriers, or diluents. The term "pharmaceutically acceptable carrier," may be a non-toxic, inert solid, semi-solid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. Pharmaceutically acceptable carriers include, for example, diluents, lubricants, binders, disintegrants, colorants, flavors, sweeteners, antioxidants, preservatives, glidants, solvents, suspending agents, wetting agents, surfactants, emollients, propellants, humectants, powders, pH adjusting agents, and combinations thereof. The pharmaceutically acceptable excipient may be a transfection facilitating agent, which may include surface active agents, such as immune-stimulating complexes (ISCOMS), Freund's incomplete adjuvant, LPS analog including monophosphoryl lipid A, muramyl peptides, quinone analogs, vesicles such as squalene and squalene, hyaluronic acid, lipids, liposomes, calcium ions, viral proteins, polyanions, polycations, or nanoparticles, or other known transfection facilitating agents.

[000114] The transfection facilitating agent may be a polyanion, polycation, including poly-L-glutamate (LGS), or lipid. The transfection facilitating agent is poly-L-glutamate, and more preferably, the poly-L-glutamate is present in the composition for genome editing in skeletal muscle or cardiac muscle at a concentration less than 6 mg/mL. The transfection facilitating agent may also include surface active agents such as immune-stimulating complexes (ISCOMS), Freund's incomplete adjuvant, LPS analog including monophosphoryl lipid A, muramyl peptides, quinone analogs and vesicles such as squalene and squalene, and hyaluronic acid may also be used administered in conjunction with the genetic construct. In some embodiments, the DNA vector encoding the composition may also include a transfection facilitating agent such as lipids, liposomes, including lecithin liposomes or other liposomes known in the art, as a DNA-liposome mixture (see for example International Patent Publication No. W09324640), calcium ions, viral proteins, polyanions, polycations, or nanoparticles, or other known transfection facilitating agents. In some embodiments, the transfection facilitating agent is a polyanion, polycation, including poly-L-glutamate (LGS), or lipid.

6. Administration

[000115] The DNA targeting systems, or at least one component thereof, as detailed herein, or the pharmaceutical compositions comprising the same, may be administered to a subject. Such compositions can be administered in dosages and by techniques well known to those skilled in the medical arts taking into consideration such factors as the age, sex, weight, and condition of the particular subject, and the route of administration. The presently

disclosed DNA targeting systems, or at least one component thereof, genetic constructs, or compositions comprising the same, may be administered to a subject by different routes including orally, parenterally, sublingually, transdermally, rectally, transmucosally, topically, intranasal, intravaginal, via inhalation, via buccal administration, intrapleurally, intravenous, intraarterial, intraperitoneal, subcutaneous, intradermally, epidermally, intramuscular, intranasal, intrathecal, intracranial, and intraarticular or combinations thereof. In certain embodiments, the DNA targeting system, genetic construct, or composition comprising the same, is administered to a subject intramuscularly, intravenously, or a combination thereof. For veterinary use, the DNA targeting systems, genetic constructs, or compositions comprising the same may be administered as a suitably acceptable formulation in accordance with normal veterinary practice. The veterinarian may readily determine the dosing regimen and route of administration that is most appropriate for a particular animal. The DNA targeting systems, genetic constructs, or compositions comprising the same may be administered by traditional syringes, needleless injection devices, "microprojectile bombardment gone guns," or other physical methods such as electroporation ("EP"), "hydrodynamic method", or ultrasound.

[000116] The DNA targeting systems, genetic constructs, or compositions comprising the same may be delivered to a subject by several technologies including DNA injection (also referred to as DNA vaccination) with and without *in vivo* electroporation, liposome mediated, nanoparticle facilitated, recombinant vectors such as recombinant lentivirus, recombinant adenovirus, and recombinant adenovirus associated virus. The composition may be injected into the skeletal muscle or cardiac muscle. For example, the composition may be injected into the tibialis anterior muscle or tail.

[000117] In some embodiments, the DNA targeting system, genetic construct, or composition comprising the same, is administered by 1) tail vein injections (systemic) into adult mice; 2) intramuscular injections, for example, local injection into a muscle such as the TA or gastrocnemius in adult mice; 3) intraperitoneal injections into P2 mice; or 4) facial vein injection (systemic) into P2 mice. In some embodiments, the DNA targeting system, genetic construct, or composition comprising the same, is administered to a human by intravenous or intramuscular injection.

[000118] Upon delivery of the presently disclosed systems or genetic constructs as detailed herein, or at least one component thereof, or the pharmaceutical compositions comprising the same, and thereupon the vector into the cells of the subject, the transfected cells may express the gRNA molecule(s) and the Cas9 molecule or fusion protein. In some embodiments, the Cas9 is a dCas9 or fusion protein.

[000119] Any of the delivery methods and/or routes of administration detailed herein can be utilized with a myriad of cell types, for example, those cell types currently under investigation for cell-based therapies, including, but not limited to, immortalized myoblast cells, such as wild-type and patient derived lines, primate dermal fibroblasts, stem cells such as induced pluripotent stem cells, bone marrow-derived progenitors, skeletal muscle progenitors, human skeletal myoblasts from patients, CD 133+ cells, mesoangioblasts, cardiomyocytes, hepatocytes, chondrocytes, mesenchymal progenitor cells, hematopoietic stem cells, smooth muscle cells, and MyoD- or Pax7-transduced cells, or other myogenic progenitor cells. The stem cell may be a human pluripotent stem cell. The stem cell may be an induced pluripotent stem cell (iPSC). The stem cell may be an embryonic stem cell (ESC).

7. Methods

a. Methods of activating endogenous myogenic transcription factor Pax7

[000120] Provided herein are methods for activating endogenous myogenic transcription factor Pax7 in a cell. The method may include administering to the cell a DNA targeting system as detailed herein, an isolated polynucleotide sequence as detailed herein, a vector as detailed herein, a cell as detailed herein, or a combination thereof. In some embodiments, endogenous expression of Pax7 mRNA is increased in the skeletal muscle progenitor cell. In some embodiments, expression of Myf5, MyoD, MyoG, or a combination thereof, is increased in the skeletal muscle progenitor cell. In some embodiments, the stem cell is induced into myogenic differentiation. In some embodiments, the skeletal muscle progenitor cell maintains Pax7 expression after at least about 2, at least about 3, at least about 4, at least about 5, at least about 6, at least about 7, at least about 8, at least about 9, at least about 10, at least about 11, at least about 12, at least about 13, at least about 14, or at least about 15 passages.

b. Methods of differentiating a stem cell into a skeletal muscle progenitor cell

[000121] Provided herein are methods of differentiating a stem cell into a skeletal muscle progenitor cell. The method may include administering to the cell a DNA targeting system as detailed herein, an isolated polynucleotide sequence as detailed herein, a vector as detailed herein, a cell as detailed herein, or a combination thereof. In some embodiments, endogenous expression of Pax7 mRNA is increased in the skeletal muscle progenitor cell. In some embodiments, expression of Myf5, MyoD, MyoG, or a combination thereof, is increased in the skeletal muscle progenitor cell. In some embodiments, the stem cell is

induced into myogenic differentiation. In some embodiments, the skeletal muscle progenitor cell maintains Pax7 expression after at least about 2, at least about 3, at least about 4, at least about 5, at least about 6, at least about 7, at least about 8, at least about 9, at least about 10, at least about 11, at least about 12, at least about 13, at least about 14, or at least about 15 passages.

c. Methods of treating a subject

[000122] Provided herein are methods for activating endogenous myogenic transcription factor Pax7 in a cell. The method may include administering to the cell a DNA targeting system as detailed herein, an isolated polynucleotide sequence as detailed herein, a vector as detailed herein, a cell as detailed herein, or a combination thereof. In some embodiments, endogenous expression of Pax7 mRNA is increased in the subject. In some embodiments, expression of Myf5, MyoD, MyoG, or a combination thereof, is increased in the subject. In some embodiments, a cell in the subject is induced into myogenic differentiation. In some embodiments, the level of dystrophin+ fibers in the subject is increased. In some embodiments, muscle regeneration in the subject is increased.

8. Examples

Example 1

Materials and Methods

[000123] *gRNA design, transfection, and plasmid construction.* Pax7 promoter targeting gRNAs were designed using crispr.mit.edu and cloned into a gRNA vector (Addgene plasmid 41824). Candidate Pax7 gRNAs were transiently transfected with Lipofectamine 3000 on the second day of CHIRON99021-induced differentiation of H9 ESCs constitutively expressing VP64-dCas9-VP64. Cells were harvested after 6 days for qRT-PCR analysis of Pax7. For doxycycline (dox)-inducible expression of VP64-dCas9-VP64, the pLV-hUBC-VP64dCas9VP64-T2A-GFP plasmid (Addgene plasmid 59791) served as the source vector for generating the pLV-tightTRE-VP64dCas9VP64-T2A-mCherry. The Pax7 gRNA was cloned into a pLV-hU6-gRNA-PGK-rtTA3-Blast that was generated using pLV-CMV-rtTA3-Blast as the source vector (Addgene plasmid 26429). The Pax7 cDNA (DNASU plasmid HsCD00443491) was cloned into a lentiviral construct to generate pLV-tightTRE-Pax7-P2A-mCherry construct. The PAX7-A sequence was confirmed to be the same as the PAX7 sequence used in previous directed differentiation papers. The PAX7-B sequence was obtained by PCR of mRNA isolated from cells treated with VP64dCas9VP64 + gRNA and cloned into a lentiviral tightTRE-PAX7-B-P2A-mCherry construct. Sequences of the

target sequences of the gRNAs are shown in **TABLE 2**. Primers used are shown in **TABLE 3**.

| gRNA # | SEQ ID # | Protospacer Sequence (5' - 3') | Position Relative to TSS |
|--------|----------|--------------------------------|--------------------------|
| 1 | 1 | GGCCGGGGACTCGGCGGATC | -490 |
| 2 | 2 | TCCCCGGCTCGACCTCGTTT | -351 |
| 3 | 3 | CCAGGGCGCAAGGGAGCGG | -278 |
| 4 | 4 | TCCTCCGCTCCCTTGCGCCC | -282 |
| 5 | 5 | GGGGGCGCGAGTGATCAGCT | -137 |
| 6 | 6 | CGGGTTTCAGGGCTGGACGG | -70 |
| 7 | 7 | TGGTCCGGAGAAAGAAGGCG | +30 |
| 8 | 8 | AGCGCCAGAGCGCGAGAGCG | +158 |

| Target | Forward Primer (5' - 3') | Reverse Primer (5' - 3') | Cycling Condition |
|--------|---|--|----------------------------|
| GAPDH | GAAGGTGAAGGTCGGAGTC (SEQ ID NO: 9) | GAAGATGGTGATGGGATTTTC (SEQ ID NO: 10) | 95°C 5s 58°C 20s x40 |
| PAX7 | CAGCAAGCCCAGACAGGTGG (SEQ ID NO: 11) | GCACGCGGCTAATCGAACTC (SEQ ID NO: 12) | 95°C 5s 58°C 20s x40 |
| MYF5 | AATTTGGGGACGAGTTTGTG (SEQ ID NO: 13) | CATGGTGGTGGACTTCCTCT (SEQ ID NO: 14) | 95°C 5s 58°C 20s x40 |
| MYOD | AGACTGCCAGCACTTTGCTA (SEQ ID NO: 15) | GTAGCTCCATATCCTGGCGG (SEQ ID NO: 16) | 95°C 5s 58°C 20s x40 |

| | | | |
|--------------------------------------|--|--|----------------------------|
| MYOG | GGTGCCAGCGAATGC (SEQ ID NO: 17) | TGATGCTGTCCACGATGGA (SEQ ID NO: 18) | 95°C 5s 58°C 20s x40 |
| Endogenous PAX7 Isoform 1/2 (PAX7-A) | GCTACAAGGTGGTGTTCAGGGT (SEQ ID NO: 19) | GAGCCATAGTACGGAAGCAGAG (SEQ ID NO: 20) | 95°C 5s 58°C 20s x40 |
| Endogenous PAX7 Isoform 3 (PAX7-B) | TCTGGCCAAAATGTGAGCC (SEQ ID NO: 21) | GGGTCAGTTAGGGTTGGGC (SEQ ID NO: 22) | 95°C 5s 58°C 20s x40 |
| T | TGCTTCCCTGAGACCCAGTT (SEQ ID NO: 23) | GATCACTTCTTTCCTTTCATCAAG (SEQ ID NO: 24) | 95°C 5s 58°C 20s x40 |
| TBX6 | CAACCCCGCATAACCTAGT (SEQ ID NO: 25) | CGTCTCGCTCCCTCTTACAG (SEQ ID NO: 26) | 95°C 5s 58°C 20s x40 |
| MSGN1 | AACCTGCGCGAGACTTTC (SEQ ID NO: 27) | ACAGCTGGACAGGGAGAAGA (SEQ ID NO: 28) | 95°C 5s 58°C 20s x40 |
| Pax3 | CTCACCTCAGGTAATGGGAC (SEQ ID NO: 29) | CGTGGTGGTAGGTTCCAGAC (SEQ ID NO: 30) | 95°C 5s 58°C 20s x40 |
| PAX7 ChIP 1, -731bp | CGGGGCTCTGACATTACACA (SEQ ID NO: 61) | GCCAGAGTCCGCCCTATTTC (SEQ ID NO: 62) | 95°C 5s 60°C 20s x40 |
| PAX7 ChIP 2, -289bp | TATTGGTCTCCGCTCCCTT (SEQ ID NO: 63) | GTGAGCGCGATCTGATAGGT (SEQ ID NO: 64) | 95°C 5s 60°C 20s x40 |
| PAX7 ChIP 3, +562bp | TTGCCGACTTTGGATTCGTC (SEQ ID NO: 65) | TCCAAAGGGAATCCCGTGC (SEQ ID NO: 66) | 95°C 5s 60°C 20s x40 |
| PAX7 ChIP 4, +926 | CGCAGGGCTGAAATTCTGGT (SEQ ID NO: 67) | AGAGCCGAGAACTGTCAGG (SEQ ID NO: 68) | 95°C 5s 60°C 20s x40 |

[000124] **Lentiviral production.** HEK293T cells were obtained from the American Tissue Collection Center (ATCC) and purchased through the Duke University Cancer Center

Facilities and were cultured in Dulbecco's Modified Eagle's Medium (Invitrogen) supplemented with 10% FBS (Sigma) and 1% penicillin/streptomycin (Invitrogen) at 37°C with 5% CO₂. Approximately 3.5 million cells were plated per 10 cm TCPS dish. Twenty-four hours later, the cells were transfected using the calcium phosphate precipitation method with pMD2.G (Addgene #12259) and psPAX2 (Addgene #12260) second generation envelope and packaging plasmids. The medium was exchanged 12 hours post-transfection, and the viral supernatant was harvested 24 and 48 hours after this medium change. The viral supernatant was pooled and centrifuged at 500 g for 5 minutes, passed through a 0.45 µm filter, and concentrated to 20x using Lenti-X Concentrator (Clontech) in accordance with the manufacturer's protocol. Undifferentiated hPSCs were transduced with the pLV-hU6-gRNA-PGK-rTA3- Blast and cells were selected with 2 µg/mL of blasticidin (Thermo) to generate homogenous population of stably transduced cells. Just prior to differentiation, hPSCs were resuspended and plated with lentivirus encoding inducible VP64-dCas9-VP64 or Pax7 cDNA.

[000125] Cell culture. H9 ESCs (obtained from the WiCell Stem Cell Bank) and DU11 iPSCs were used for these studies. DU11 iPSCs were generated by the Duke iPSC Shared Resource Facility via episomal reprogramming of BJ fibroblasts from a healthy male newborn (ATCC cell line, CRL-2522). Stable and correct karyotype and pluripotency of the cells was confirmed. hPSCs were maintained in mTeSR (Stem Cell Technologies) and plated on tissue culture treated plates coated with ES-qualified matrigel (Corning). For differentiation, hPSCs were dissociated into single cells with Accutase (Stem Cell Technologies) and plated on matrigel coated plates at $2.3 - 3.3 \times 10^4/\text{cm}^2$ in mTeSR medium supplemented with 10 µM Y27632 (Stem Cell Technologies). The following day, mTeSR medium was replaced with E6 media supplemented with 10 µM CHIR99021 (Sigma) to initiate mesoderm differentiation. After 2 days, CHIR99021 was removed and cells were maintained in E6 media with 10 ng/mL FGF2 (Sigma) and 1 µg/mL of doxycycline (dox) (Sigma).

[000126] Fluorescence activated cell sorting and expansion of sorted cells. At day 14 after induction of differentiation, cells were dissociated with 0.25% Trypsin-EDTA (Thermo) and washed with neutralizing media (10% FBS in DMEM/F12). Cells were pelleted by centrifugation and resuspended in flow media (5% FBS in PBS). Cells were sorted for mCherry expression, pelleted, resuspended in growth media (E6 supplemented with 10 ng/mL FGF2 and 1 µg/mL dox) and plated on matrigel-coated plates. Cells were passaged every 3-4 days at ~80% confluency. Terminal differentiation was induced by withdrawing dox from the medium in 100% confluent cultures.

[000127] Flow cytometry analysis. For flow cytometry analysis of surface markers, cells were harvested during the proliferation phase at day 20 of differentiation. Cells were dissociated with 0.25% Trypsin- EDTA, washed with PBS, then resuspended in flow buffer (PBS with 5% FBS). Cells were incubated with the following conjugated antibodies at 0.25 $\mu\text{g}/10^6$ cells: IgG1-K isotype control-FITC (eBioscience 11-4714-41), CD56-FITC (eBioscience 11-0566-41), or CD29- FITC (eBioscience 11-0299-41). Cells were analyzed on SONY SH800 flow cytometer.

[000128] Cell transplantation into immunodeficient mice. All animal experiments were conducted under protocols approved by the Duke Institutional Animal Care and Use Committee. 7 week old female NOD.SCID.gamma mice (Duke CCIF Breeding Core) were used for these *in vivo* studies. Prior to intramuscular cell transplantation, mice were pre-injured with 30 μL of 1.2% BaCl₂ (Sigma). 24 hours later, MPCs from differentiated iPSCs or ESCs were injected into the tibialis anterior (TA) muscle (5×10^5 cells/ $15 \mu\text{L}$ Hank's Balanced Salt Solution). Four weeks after injection, mice were euthanized and the TA muscles were harvested.

[000129] Immunofluorescence staining of cultured cells and tissue sections. Cultured cells were plated on autoclaved glass coverslips (1 mm, Thermo) coated with matrigel for immunofluorescence staining during the proliferation phase. For differentiation, cells were grown to confluency and differentiated on 24 well tissue culture plates coated with matrigel, and immunofluorescence staining was performed directly in the well. Cells were fixed with 4% PFA for 15 min and permeabilized in blocking buffer (PBS supplemented with 3% BSA and 0.2% Triton X-100) for 1 hr at room temperature. Samples were incubated overnight at 4°C with the following antibodies: Pax7 (1:20, Developmental Studies Hybridoma Bank), Myosin Heavy Chain MF20 (1:200, DSHB), Myf5 (1:200, Santa Cruz sc-302) and MyoD 5.8A (1:200, Santa Cruz sc-32758). Samples were washed with PBS for 15 min and incubated with compatible secondary antibodies diluted 1:500 from Invitrogen and DAPI for 1 hr at room temperature. Samples were washed for 15 min with PBS and coverslips were mounted with ProLong Gold Antifade Reagent (Invitrogen) or wells were kept in PBS and imaged using conventional fluorescence microscopy. Harvested TA muscles were mounted and frozen in Optimal Cutting Temperature (OCT) compound cooled in liquid nitrogen. Serial 10 μm cryosections were collected. Cryosections were fixed with 2% PFA for 5 min and permeabilized with PBS + 0.2% Triton-X for 10 minutes. Blocking buffer (PBS supplemented with 5% goat serum, 2% BSA, and 0.1% Triton X-100) was applied for 1 hr at room temperature. Samples were incubated overnight at 4°C with a combination of the following antibodies: human-specific MANDYS106 (1:200, Sigma

MABT827), human-specific Lamin A/C (1:100, Thermo MA31000), Pax7 (1:10, Developmental Studies Hybridoma Bank), or Laminin (1:200, Sigma L9393). Samples were washed with PBS for 15 min and incubated with compatible secondary antibodies diluted 1:500 from Invitrogen and DAPI for 1 hr at room temperature. Samples were washed for 15 min with PBS and slides were mounted with ProLong Gold Antifade Reagent (Invitrogen) and imaged using conventional fluorescence microscopy.

[000130] Quantitative Reverse Transcription PCR. RNA was isolated using the RNeasy Plus RNA isolation kit (Qiagen). cDNA was synthesized with the SuperScript VILO cDNA Synthesis Kit (Invitrogen). Real-time PCR using PerfeCTa SYBR Green FastMix (Quanta Biosciences) was performed with the CFX96 Real-Time PCR Detection System (Bio-Rad). The results are expressed as fold-increase expression of the gene of interest normalized to GAPDH expression using the $\Delta\Delta C_t$ method.

[000131] Chromatin Immunoprecipitation (ChIP) qPCR. ChIP was performed using the EpiQuik ChIP Kit (EpiGentek) according to manufacturer's instructions. Soluble chromatin was immunoprecipitated with antibodies against H3K27ac and H3K4me3 (abcam), and gDNA was purified for qPCR analysis. All sequences for ChIP-qPCR primers can be found in **TABLE 3**. qPCR was performed using PerfeCTa SYBR Green FastMix (Quanta BioSciences), and the data are presented as fold change gDNA relative to negative control (gRNA only) and normalized to a region of the GAPDH locus.

[000132] RNA-Seq. RNA was extracted from freshly sorted cells at day 14 of differentiation using the Total RNA Purification Plus Micro Kit (Norgen). Library preparation and sequencing was performed by GENEWIZ on an Illumina HiSeq in the 2 x 150 bp sequencing configuration. All RNA-seq samples were first validated for consistent quality using FastQC v0.11.2 (Babraham Institute). Raw reads were trimmed to remove adapters and bases with average quality score (Q) (Phred33) of < 20 using a 4 bp sliding window (SLIDINGWINDOW:4:20) with Trimmomatic v0.32 (Bolger et al. *Bioinformatics* **2014**, *30*, 2114–2120). Trimmed reads were subsequently aligned to the primary assembly of the GRCh38 human genome using STAR v2.4.1a (Dobin et al. *Bioinformatics* **2013**, *29*, 15–21) removing alignments containing non-canonical splice junctions (--outFilterIntronMotifs RemoveNoncanonical). Aligned reads were assigned to genes in the GENCODE v19 comprehensive gene annotation (Harrow et al. *Genome Res.* **2012**, *22*, 1760–1774) using the featureCounts command in the subread package with default settings (v1.4.6-p4) (Liao et al. *Nucleic Acids Res.* **2013**, *41*, e108–e108). The subsequent counts were normalized for each replicate using the R package DESeq2 after filtering out genes that were not sufficiently quantified, and normalized values were used for analysis. Heatmaps were

generated using the pheatmap package in R software. Biological processes and pathways were generated using Enrichr (Chen et al. *BMC Bioinformatics* **2013**, *14*, 128), a web-based online tool. For estimating transcript and gene abundances, Transcript Per Million (TPMs) were computed using the rsem-calculate-expression function in the RSEM v1.2.21 package (Li and Dewey. *BMC Bioinformatics* **2011**, *12*, 323).

Example 2

Developing conditions for VP64-dCas9-VP64-Mediated Endogenous Pax7 Activation in hPSCs

[000133] During embryonic differentiation, PAX7 and its paralog PAX3 specify myogenic cells within the paraxial mesoderm. Differentiation of hPSCs into paraxial mesoderm cells can be initiated by CHIR99021, a GSK3 inhibitor (Tan et al. *Stem Cells Dev.* **2013**, *22*, 1893–1906). Two human pluripotent stem cell lines, H9 ESCs and DU11 iPSCs, were used for differentiation studies. For targeted gene activation, we used the dCas9 with the VP64 domain fused to both the N- and C-termini (VP64-dCas9-VP64), which we previously showed to be ~10-fold more potent than a single VP64 fusion. To test the efficacy of VP64-dCas9-VP64-mediated activation of *PAX7*, we designed 8 gRNAs spanning -490 to +158 base pairs relative to the transcription start site of the human *PAX7* gene (FIG. 7A). H9 ESCs stably expressing VP64-dCas9-VP64 were differentiated into paraxial mesoderm cells with addition of CHIR99021 in E6 medium for 2 days, as previously described (Shelton et al. *Stem Cell Rep.* **2014**, *3*, 516–529). Cells were transfected with the individual gRNAs and samples were harvested 6 days later for gene expression analysis using qRT-PCR. 4 out of the 8 gRNAs significantly upregulated *PAX7* compared to mock transfected cells (FIG. 7B). In a second screen, we packaged the 4 individual gRNAs that performed best in the transfection experiment into lentiviruses to achieve more stable and robust expression. Cells were harvested at 8 days post-transduction. gRNA #4 was identified as the most potent gRNA and was used for subsequent studies (FIG. 7C).

Example 3

VP64-dCas9-VP64-mediated differentiation of hPSCs into myogenic progenitor cells

[000134] Next, we tested the hypothesis that endogenous *PAX7* activation in paraxial mesoderm cells would be sufficient for generating myogenic progenitor cells (MPCs) with the potential to differentiate into myotubes in vitro (FIG. 1A). Prior to differentiation, hPSCs were transduced with a lentivirus expressing the *PAX7* promoter-targeting gRNA, a reverse tetracycline transactivator (rtTA), and a blasticidin resistance gene. Cells were selected with

blasticidin for stable expression of the vector and then transduced with an additional lentivirus encoding either doxycycline (dox)-inducible VP64-dCas9-VP64 or the *PAX7* cDNA, which also included a co-transcribed mCherry reporter gene (FIG. 1B). hPSCs were differentiated with CHIR99021 for 2 days and then maintained in E6 medium with dox and FGF2 to support MPC proliferation (FIG. 1C) (Pawlikowski et al. *Dev. Dyn.* 2017, 246, 359–367). Addition of CHIR99021 induced paraxial mesodermal differentiation, as indicated by high levels of pan-mesoderm marker Brachyury (*T*), paraxial mesoderm markers *MSGN1* and *TBX6*, and premyogenic mesoderm marker *PAX3* at the mRNA level (FIG. 1D). Transduced cells were sorted based on mCherry expression after two weeks of growth (FIG. 1E). mCherry+ cells accounted for ~20% of cells transduced with VP64-dCas9-VP64 compared to ~50% with *PAX7* cDNA transduced cells. This is likely due to the larger size of VP64-dCas9-VP64 vector compared to the *PAX7* cDNA vector (7.9 kb between LTRs vs. 4.9 kb) resulting in reduced lentiviral titers. These purified MPCs were maintained in serum-free E6 medium supplemented with dox and FGF2 and passaged when cells reached ~80% confluency. Sorted cells demonstrated high purity of *PAX7*+ cells in both the endogenous-activated cells and exogenous cDNA-expressing cells when protein expression was assessed by immunofluorescence staining 5 days after sorting (FIG. 1F and FIG. 8A). VP64-dCas9-VP64-treated iPSCs and ESCs both demonstrated notable expansion potential, averaging 85-fold and 95-fold increase in cell number, respectively, over the 2 weeks after purification. Furthermore, the growth potential of these cells outperformed the *PAX7* cDNA overexpressing cells (FIG. 1G, FIG. 8B).

Example 4

Characterization of myogenic progenitor cells derived from endogenous or exogenous *PAX7* expression.

[000135] *PAX7* mRNA levels were assessed by qRT-PCR during the proliferation phase 5 days after sorting. *PAX7* mRNA from the endogenous chromosomal locus could be discriminated from total *PAX7* mRNA, made from either the lentivirus or endogenous chromosomal locus, using distinct primer pairs. While overexpression of *PAX7* cDNA resulted in more total *PAX7* mRNA (FIG. 2A and FIG. 8C), robust detection of any endogenous *PAX7* isoform was only observed in VP64-dCas9-VP64-treated cells (FIG. 2B and FIG. 8D). The human *PAX7* gene encodes multiple isoforms of which differential sequences have been identified, but unique biological functions remain unclear. Differential transcriptional termination in either exon 8 or exon 9 yield *PAX7-A* and *PAX7-B* isoforms, respectively. The differences in the 3' ends of these transcripts allow for differential detection with unique qRT-PCR primers.

[000136] Downstream myogenic regulatory factors *MYF5*, *MYOD*, and *MYOG* were also detected at the mRNA level by qRT-PCR (FIG. 2C, FIG. 8E). At the protein level, the majority of cells in both endogenous and exogenous *PAX7*-expressing cells co-expressed the activated satellite cell marker, *MYF5* (>90%). The myoblast marker, *MYOD*, was expressed higher in cells expressing endogenous *PAX7* compared to exogenous *PAX7* cDNA, at 15.9% and 6.8%, respectively. Mature myogenic markers *MYOG* and Myosin Heavy Chain (MHC) were lowly detectable in some of the cells (FIG. 2D).

[000137] Human satellite cells co-express *PAX7* with CD29 and CD56 surface markers. At approximately 10 days after sorting, we assessed our MPCs for CD29 and CD56 expression and found 100% of cells in all groups expressed CD29, independent of *PAX7* expression. We found CD56 expression was more contingent on *PAX7* expression, with only 27.4% of cells expressing CD56 in the gRNA only group, compared to 69.2% and 87.5% of cells in the *PAX7* cDNA and VP64-dCas9-VP64-treated groups, respectively (FIG. 2E and FIG. 8F). Assessment of mean fluorescence intensity (MFI) of CD56 staining also revealed the average CD56 expression level per cell was significantly higher in the VP64-dCas9-VP64-treated group (FIG. 2F and FIG. 8G).

Example 5

Transplantation of VP64-dCas9-VP64-generated myogenic progenitors into immunodeficient mice demonstrates *in vivo* regenerative potential

[000138] We next determined if MPCs derived from VP64-dCas9-VP64-mediated *PAX7* activation possess *in vivo* regenerative potential. Cells that had been expanded and passaged 3 times post sort were transplanted into the tibialis anterior (TA) of immunodeficient NOD.SCID.gamma (NSG) mice that were pre-injured with barium chloride (BaCl₂) to create a regenerative microenvironment (Hall et al. *Sci. Transl. Med.* 2010, 2, 57ra83–57ra83). 24 hours after injury, mice were injected with 500,000 cells treated with either gRNA only, *PAX7* cDNA overexpression, or VP64-dCas9-VP64-mediated endogenous *PAX7* activation. One month after transplantation, muscles were harvested and evaluated for engraftment by immunostaining with human-specific dystrophin and lamin A/C antibodies. Human nuclei were detected by lamin A/C staining in all three conditions; however, only the endogenous *PAX7* activated group demonstrated consistent presence of human dystrophin (FIG. 3A and FIG. 8I). The number of human dystrophin+ fibers was quantified across three mice per condition by counting sections with most abundant human dystrophin+ fibers within each sample (FIG. 3B). We also investigated whether transplanted cells could seed the satellite cell niche. Immunostaining for *PAX7*, human lamin A/C, and laminin was performed

to demarcate satellite cells of human origin. PAX7 and human lamin A/C double-positive cells residing under the basal lamina were identified only in muscle transplanted with VP64dCas9VP64-activated MPCs (FIG. 3C, FIG. 8J).

Example 6

Induction of endogenous PAX7 expression is sustained after multiple passages and dox withdrawal

[000139] During expansion of sorted cells, we noticed a significant decrease in PAX7+ cells in the cDNA overexpression group after an average of 4 passages spanning an average of 32 days in three independent experiments. Although the initial number of cells expressing PAX7 protein was >90% at five days post sort, quantification of PAX7+ nuclei following approximately 4 passages after initial flow sorting revealed that only a minority of cells (35.8%) expressed PAX7 protein despite maintenance in dox during the expansion period. Conversely, a large majority (93%) of endogenously activated PAX7 cells retained PAX7 protein expression without precocious differentiation across multiple passages (FIG. 4A and FIG. 4C). As indicated by lack of MHC+ cells, depletion of PAX7+ cells in the cDNA overexpression group did not correspond to the adoption of a myogenic fate (FIG. 4A). We postulated this may be due to high levels of PAX7 protein hindering cell proliferation, allowing for cells that have silenced the promoter or contaminating cells from the sort to overtake the cell population. Consistent with this possibility, Pax7 cDNA overexpression has been previously implicated in inducing cell cycle exit without commitment to myogenic differentiation. Interestingly, a previously published study also observed this phenomenon of PAX7 loss over multiple passages when using a tet-inducible PAX7 cDNA overexpression system. That study required amending the serum-free differentiation protocol to media conditions containing highly-mitogenic 20% fetal calf serum to improve retention of PAX7 protein expression in cDNA-overexpressing cells.

[000140] Differentiation of premyogenic cells was induced by withdrawing dox when cells reached 100% confluency. Abundant MHC+ myofibers were observed in VP64-dCas9-VP64-treated cells (FIG. 4B, FIG. 8H). Interestingly, 50% of cells remained PAX7+ in these cells in which the endogenous gene had been activated even at 1 week after dox removal, in contrast the PAX7 cDNA-treated cells in which 5.2% were PAX7+ after 1 week without dox (FIG. 4C). Staining for the FLAG epitope confirmed the absence of VP64-dCas9-VP64 in differentiated cells at this time point (FIG. 4D).

Example 7

VP64-dCas9-VP64 leads to sustained PAX7 expression and stable chromatin remodeling at target locus

[000141] We hypothesized that epigenetic remodeling of the endogenous *PAX7* promoter was allowing cells to autonomously upregulate *PAX7* without the continued presence of VP64-dCas9-VP64. To investigate this, we performed chromatin immunoprecipitation (ChIP)-qPCR on cells during dox administration and at 15 days after dox withdrawal. Cells were analyzed at day 30 of differentiation for the +dox condition and then expanded and passaged 3 more times over 15 days in the absence of dox. We used ChIP-seq data generated as part of the Encyclopedia of DNA Elements (ENCODE) Project to identify histone modifications enriched at the transcriptionally active *PAX7* in human skeletal muscle myoblasts (HSMM), including H3K4me3 and H3K27ac (**FIG. 5A**). Four qPCR primers were designed to tile regions -731bp to +926bp relative to the *PAX7* transcription start site (TSS). ChIP qPCR of +dox conditions demonstrated significant enrichment of H3K4me3 and H3K27ac at the endogenous *PAX7* locus only in response to VP64-dCas9-VP64 treatment (**FIG. 5B**). Furthermore, these histone modifications were maintained for 15 days post dox withdrawal (**FIG. 5C**). To ensure that there was no leaky expression of VP64-dCas9-VP64 after dox removal, we performed a western blot for the FLAG epitope tag and were unable to detect VP64-dCas9-VP64 after 15 days of dox removal (**FIG. 5D**). Conversely, *PAX7* was still detectable by western blot in the absence of VP64-dCas9-VP64, corresponding to the ChIP-qPCR enrichment of active histone marks.

Example 8

Identification of endogenous vs. exogenous PAX7-induced global transcriptional changes

[000142] To evaluate the transcriptome-wide gene expression changes induced by endogenous activation of *PAX7* compared to exogenous cDNA overexpression, we performed RNA sequencing (RNA-seq) analysis. Differentiated cells that had been treated with either gRNA only, VP64-dCas9-VP64 with gRNA, cDNA encoding *PAX7-A* isoform, or cDNA encoding *PAX7-B* isoform were sorted for mCherry expression at day 14 and RNA was extracted for sequencing. We included *PAX7-B* because it is highly expressed in VP64-dCas9-VP64-treated cells (**FIG. 2B**), yet little is known of its relationship to *PAX7-A*. To gauge the variance between the samples, we generated a sample distance matrix of the RNA-seq data (**FIG. 6A**). This revealed distinct differences between the four treatments, and

four unique clusters were readily apparent despite the commonality of induced *PAX7* expression in three of the four groups. Multidimensional scaling (MDS) of the top 500 differentially expressed genes also showed divergent clustering of sample groups with *PAX7* cDNA overexpression contributing most to variation between transcriptomic profiles (FIG. 9A). We considered the top 200 most variable genes across the 4 groups and submitted lists of gene clusters apparent on the heat map for GO term analysis (FIG. 6B). These analyses revealed general developmental pathways including mesoderm development and WNT signaling pathway genes overexpressed in gRNA only group. Additionally, this group overexpressed genes involved in heart development such as *HAND1* and *HAND2*, which indicates slightly higher propensity of this group to differentiate into cardiac cell lineage. Consistent with this observation, CHIR99021 is also used as the initiator of differentiation of hPSCs into cardiomyocytes.

[000143] GO analyses of genes differentially expressed in the VP64-dCas9-VP64 group were strongly related to myogenesis (FIG. 6B and FIG. 9B). Genes represented in this group included embryonic myoblast marker *HOXC12*, embryonic myosin heavy chain *MYH3*, as well as other myogenic regulatory factors *MYOD* and *MYOG*.

[000144] Genes enriched genes following treatment with *PAX7-A* were associated with CNS development and NOTCH1 signaling pathways. Interestingly, one of the most differentially upregulated genes in this group was *DLK1* (FIG. 9B and FIG. 9C), which is required for normal embryonic skeletal muscle development. However, overexpression of *DLK1 in vitro* inhibits proliferation of satellite cells and induces cell cycle exit and early differentiation. Conversely, *Dlk1* knockout increases Pax7+ myogenic progenitor cell proliferation *in vitro* and enhances post-natal muscle regeneration *in vivo*. This would suggest that *DLK1* is involved in maintaining the balance between quiescence and activation of satellite cells. Furthermore, the specific upregulation of both *DLK1* and *DIO3* in these cells (FIG. 9B and FIG. 9C) suggests activity of the *DLK1-DIO3* gene cluster. This *DLK1-DIO3* locus encodes the largest mammalian megacluster of micro RNAs (miRNA), which is strongly expressed in freshly isolated satellite cells and strongly declined in proliferating satellite cells. This decline of *DLK1-DIO3* is concomitant with upregulation of muscle-specific miRNAs, including miR-1, which targets the *PAX7* 3' UTR to fine-tune its expression and control satellite cell differentiation. Thus, it is feasible that overexpression of only the *PAX7-A* isoform results in negative feedback and expression of genes and miRNAs that regulate quiescence.

[000145] Genes overexpressed specifically in response to *PAX7-B* included brain development genes *VIT* and *OTP*, as well as other *PAX* genes, *PAX2* and *PAX8*, which are

involved in kidney development. Although *PAX7* is not implicated in kidney development, CHIR99021 has been used previously to differentiate hPSCs to a kidney lineage.

[000146] Next, we compared each of the three *PAX7*-expressing groups to the gRNA only group and extracted a list of genes with greater than two-fold change and $\text{padj} < 0.05$ after filtering genes with low read counts. We compared these lists of genes and found that the 56 genes shared in all three groups were enriched for GO terms involved in skeletal muscle development (**FIG. 6C** and **FIG. 6D**). This suggests that compared to treatment with only the gRNA and 14 days of CHIR-mediated differentiation, all three groups were able to direct hPSCs into the skeletal myogenic program more effectively than the small molecule protocol alone. When individual genes are examined, however, the VP64-dCas9-VP64 group outperforms the other groups in terms of expression of pre-myogenic and myogenic genes (**FIG. 6E**). Many of the known satellite cell surface markers and genes are also more highly expressed in the VP64-dCas9-VP64 group compared to the other groups, demonstrating more specific and robust commitment to myogenesis and satellite cell differentiation (**FIG. 6E** and **FIG. 9D**).

Example 9

Discussion

[000147] Detailed herein is the utility of CRISPR/Cas9-based transcriptional activators for differentiation of hPSCs into myogenic progenitor cells via targeted activation of the endogenous *PAX7* gene. This method may serve as an alternative to the transgene overexpression model that has been previously used for myogenic progenitor cell differentiation. With a minimal small molecule differentiation protocol involving initial paraxial mesodermal differentiation with CHIR99021 and maintenance with FGF2 in serum-free media conditions, it was demonstrated that targeted activation of the endogenous *PAX7* gene generates a myogenic progenitor cell population that can be passaged at least 6 times while maintaining *PAX7* expression, differentiate readily upon dox withdrawal and subsequent loss of dCas9 activator expression, and engraft into mouse muscle to produce human dystrophin+ fibers while also occupying the satellite cell niche. It was demonstrated that targeting the endogenous *PAX7* promoter results in enrichment of H3K4me3 and H3K27ac histone modifications, which was sustained for 15 days after dox removal. Enrichment of these chromatin marks was not observed during overexpression of *PAX7* cDNA. Although *PAX7* cDNA overexpression from hPSCs has yielded various degrees of engraftment into NSG mice previously, we did not have similar positive engraftment results with *PAX7* cDNA overexpression under the conditions used here. However, the prior studies

used differentiation protocols that generate embryoid bodies, incorporate additional small molecules, or contain animal serum in the medium and thus, differ from the protocol used in this study. Detailed herein is that activation of the endogenous *PAX7* rather than exogenous *PAX7* cDNA overexpression increases the efficacy of hPSC differentiation into myogenic progenitor cells with robust growth and differentiation potential, while retaining regenerative properties following transplantation.

[000148] Prior studies using exogenous *PAX7* cDNA relied on overexpression of only the *PAX7-A* isoform. However, differential RNA cleavage and polyadenylation yields *PAX7-B*, which contains a highly conserved *paired* tail domain and is considered to be the canonical sequence. Both isoforms are expressed in human myogenic cells and orthologs of these *PAX7* protein variants are also present in mouse muscle, indicating biological significance for both isoforms. Although distinct functions of these protein variants have not been deciphered, they may play differential roles in myogenesis that may be necessary for proper satellite stem cell function and myogenic differentiation. The RNA-seq analysis demonstrated overlapping myogenic function of cells generated by VP64-dCas9-VP64 endogenous activation or *PAX7* cDNA overexpression of either isoforms; however, the VP64-dCas9-VP64 group shared more commonly upregulated genes with *PAX7-B* than *PAX7-A* (89 and 30 genes, respectively), indicating a higher degree of similarity, which is also depicted in the sample distance matrix. The dissimilarity between the overexpression of the two cDNAs indicated that they have distinct functions and can influence global gene expression in separate ways. For example, *PAX7-B* upregulates pre-myogenic genes *PAX3*, *DMRT2*, and satellite cell genes *CXCR4* and *HEY1* more effectively than *PAX7-A*. Conversely, expression of the *DLK1-DIO3* locus that is implicated in satellite cell quiescence is more robust in response to *PAX7-A* than *PAX7-B*. VP64-dCas9-VP64-mediated *PAX7* induction therefore may allow expression of both isoforms to properly induce myogenesis at levels of expression that are more likely in the physiological range. Furthermore, endogenous activation of *PAX7* may preserve the 3' UTRs, which are binding targets for the many muscle-specific miRNAs that play a role in orchestrating proper muscle development and regeneration.

[000149] Although conditional expression of *PAX7* in hPSCs via lentiviral transduction may be the most promising approach for generating a homogenous population of engraftable MPCs, integration-free reprogramming may ultimately be used for avoiding undesired consequences of genomic integration of viral vectors. VP64-dCas9-VP64 has been demonstrated to rapidly remodel the epigenetic signature of target loci when gRNAs were transiently delivered to achieve neuronal differentiation. It is demonstrated herein that

epigenetic signatures were stably maintained in the absence of VP64-dCas9-VP64. Transient delivery of these targeted transcriptional activators via transfection, electroporation, or nonviral nanoparticle delivery of mRNA/gRNA or purified ribonucleoprotein complexes may offer an alternative to integration-prone methods.

[000150] The expansive CRISPR genome engineering toolbox offers many possibilities to manipulate cell fates to improve our understanding of the molecular differences between myoblasts, satellite cells, and MPCs generated from hPSCs. Forced transitioning of cell fate may rely on stochastic factors that have remained largely elusive, but generally include activation of endogenous networks to generate a stable new identity while also opposing epigenetic memory of the old identity. Further investigation of tissue-specific progenitor cell differentiation from pluripotent cells may unveil fundamental guidelines that may inform a revised model for the generation of a well-defined population of cells capable of repopulating the progenitor cell niche long term.

[000151] The results detailed herein introduced a novel method for differentiation and expansion of myogenic progenitors from hPSCs by deterministic editing of transcriptional regulation with new genome engineering tools, which may enable new disease modeling and cell therapy in disorders of skeletal muscle regeneration.

[000152] The foregoing description of the specific aspects will so fully reveal the general nature of the invention that others can, by applying knowledge within the skill of the art, readily modify and/or adapt for various applications such specific aspects, without undue experimentation, without departing from the general concept of the present disclosure. Therefore, such adaptations and modifications are intended to be within the meaning and range of equivalents of the disclosed aspects, based on the teaching and guidance presented herein. It is to be understood that the phraseology or terminology herein is for the purpose of description and not of limitation, such that the terminology or phraseology of the present specification is to be interpreted by the skilled artisan in light of the teachings and guidance.

[000153] The breadth and scope of the present disclosure should not be limited by any of the above-described exemplary aspects, but should be defined only in accordance with the following claims and their equivalents.

[000154] All publications, patents, patent applications, and/or other documents cited in this application are incorporated by reference in their entirety for all purposes to the same extent

as if each individual publication, patent, patent application, and/or other document were individually indicated to be incorporated by reference for all purposes.

[000155] For reasons of completeness, various aspects of the invention are set out in the following numbered clauses:

[000156] Clause 1. A guide RNA (gRNA) molecule targeting Pax7, the gRNA comprising a polynucleotide sequence corresponding to at least one of SEQ ID NOs: 1-8 or 69-76, or a variant thereof.

[000157] Clause 2. The gRNA of clause 1, wherein the gRNA comprises a crRNA, a tracrRNA, or a combination thereof.

[000158] Clause 3. A DNA targeting system for increasing expression of Pax7, the DNA targeting system comprising at least one gRNA that binds and targets a Pax7 gene, a regulatory region of a Pax7 gene, a promoter region of a Pax7 gene, or a portion thereof.

[000159] Clause 4. The DNA targeting system of clause 3, wherein the at least one gRNA comprises a polynucleotide sequence corresponding to at least one of SEQ ID NOs: 1-8 or 69-76, or a variant thereof.

[000160] Clause 5. The DNA targeting system of clause 3 or 4, wherein the gRNA comprises a crRNA, a tracrRNA, or a combination thereof.

[000161] Clause 6. The DNA targeting system of any one of clauses 3-5, further comprising a Clustered Regularly Interspaced Short Palindromic Repeats associated (Cas) protein or a fusion protein, wherein the fusion protein comprises two heterologous polypeptide domains, wherein the first polypeptide domain comprises a Cas protein, a zinc finger protein, or a TALE protein, and the second polypeptide domain has transcription activation activity.

[000162] Clause 7. The DNA targeting system of clause 6, wherein the Cas protein comprises a *Streptococcus pyogenes* Cas9 molecule, or a variant thereof.

[000163] Clause 8. The DNA targeting system of clause 6, wherein the fusion protein comprises VP64-dCas9-VP64.

[000164] Clause 9. The DNA targeting system of clause 6, wherein the Cas protein comprises a Cas9 that recognizes a Protospacer Adjacent Motif (PAM) of NGG (SEQ ID NO: 31), NGA (SEQ ID NO: 32), NGAN (SEQ ID NO: 33), or NGNG (SEQ ID NO: 34).

[000165] Clause 10. An isolated polynucleotide sequence comprising the gRNA molecule of clause 1 or 2.

[000166] Clause 11. An isolated polynucleotide sequence encoding the DNA targeting system of any one of clauses 3-9.

[000167] Clause 12. A vector comprising the isolated polynucleotide sequence of clause 10 or 11.

[000168] Clause 13. A vector encoding the gRNA molecule of clause 1 or 2 and a Clustered Regularly Interspaced Short Palindromic Repeats associated (Cas) protein.

[000169] Clause 14. A cell comprising the gRNA of clause 1 or 2, the DNA targeting system of any one of clauses 3-9, the isolated polynucleotide sequence of clause 10 or 11, or the vector of clause 12 or 13, or a combination thereof.

[000170] Clause 15. A pharmaceutical composition comprising the gRNA of clause 1 or 2, the DNA targeting system of any one of clauses 3-9, the isolated polynucleotide sequence of clause 10 or 11, the vector of clause 12 or 13, or the cell of clause 14, or a combination thereof.

[000171] Clause 16. A method of activating endogenous myogenic transcription factor Pax7 in a cell, the method comprising administering to the cell the gRNA of clause 1 or 2, the DNA targeting system of any one of clauses 3-9, the isolated polynucleotide sequence of clause 10 or 11, or the vector of clause 12 or 13.

[000172] Clause 17. A method of differentiating a stem cell into a skeletal muscle progenitor cell, the method comprising administering to the stem cell the gRNA of clause 1 or 2, the DNA targeting system of any one of clauses 3-9, the isolated polynucleotide sequence of clause 10 or 11, or the vector of clause 12 or 13.

[000173] Clause 18. The method of clause 17, wherein endogenous expression of Pax7 mRNA is increased in the skeletal muscle progenitor cell.

[000174] Clause 19. The method of any one of clauses 17-18, wherein the expression of Myf5, MyoD, MyoG, or a combination thereof, is increased in the skeletal muscle progenitor cell.

[000175] Clause 20. The method of any one of clauses 17-19, wherein the stem cell is induced into myogenic differentiation.

[000176] Clause 21. The method of any one of clauses 17-20, wherein the skeletal muscle progenitor cell maintains Pax7 expression after at least about 6 passages.

[000177] Clause 22. A method of treating a subject in need thereof, the method comprising administering to the subject the cell of clause 14.

[000178] Clause 23. The method of clause 22, wherein the level of dystrophin+ fibers in the subject is increased.

[000179] Clause 24. The method of clause 22, wherein muscle regeneration in the subject is increased.

SEQUENCES

| SEQ ID NO | gRNA sequence | SEQ ID NO | gRNA |
|-----------|----------------------|-----------|-----------------------|
| 1 | ggccggggactcggcggatc | 69 | ggccggggacucggcggauc |
| 2 | tccccggctcgacctcgttt | 70 | uccccggcucgaccucguuu |
| 3 | ccagggcgcaagggagcgg | 71 | ccagggcgcaagggagcgg |
| 4 | tcttcgctcccttgcgccc | 72 | uccuccgcucccuugcgccc |
| 5 | gggggcgcgagtgatcagct | 73 | gggggcgcgagugaucagcu |
| 6 | cgggtttcagggctggacgg | 74 | cggguuucagggcuggacgg |
| 7 | tggtcgggagaaagaaggcg | 75 | ugguccgggagaaagaaggcg |
| 8 | agcgccagagcgcgagagcg | 76 | agcgccagagcgcgagagcg |

| SEQ ID NO | gRNA target sequence |
|-----------|----------------------|
| 77 | GATCCGCCGAGTCCCCGGCC |
| 78 | AAACGAGGTCGAGCCGGGGA |
| 79 | CCGCTCCCTTGCGCCCTGG |
| 80 | GGGCGCAAGGGAGCGGAGGA |
| 81 | AGCTGATCACTCGCGCCCCC |
| 82 | CCGTCCAGCCCTGAAACCCG |

| | |
|----|----------------------|
| 83 | CGCCTTCTTTCTCCGGACCA |
| 84 | CGCTCTCGCGCTCTGGCGCT |

| Target | Forward Primer (5' - 3') | Reverse Primer (5' - 3') |
|-----------------------------------|--|--|
| GAPDH | gaaggtgaaggtcggagtc (SEQ ID NO: 9) | gaagatggtgatgggatttc (SEQ ID NO: 10) |
| PAX7 | cagcaagcccagacaggtgg (SEQ ID NO: 11) | gcacgoggctaatacgaactc (SEQ ID NO: 12) |
| MYF5 | aatttggggacgagtttgtg (SEQ ID NO: 13) | catggtggtggacttcctct (SEQ ID NO: 14) |
| MYOD | agactgccagcactttgcta (SEQ ID NO: 15) | gtagctccatatacctggcgg (SEQ ID NO: 16) |
| MYOG | ggtgccagcgaatgc (SEQ ID NO: 17) | tgatgctgtccacgatgga (SEQ ID NO: 18) |
| Endogenous PAX7 Isoform 1/2 | gctacaaggtggtgtcagggt (SEQ ID NO: 19) | gagccatagtacggaagcagag (SEQ ID NO: 20) |
| Endogenous PAX7 Isoform 3 | tctggccaaaaatgtgagcct (SEQ ID NO: 21) | gggtcagttagggttgggc (SEQ ID NO: 22) |
| T | tgcttcctgagaccagtt (SEQ ID NO: 23) | gatcacttctttcctttgcatcaag (SEQ ID NO: 24) |
| TBX6 | caaccocgcatacacctagt (SEQ ID NO: 25) | cgtctcgctccctcttacag (SEQ ID NO: 26) |
| MSGN1 | aacctgcgcgagactttcc (SEQ ID NO: 27) | acagctggacagggagaaga (SEQ ID NO: 28) |
| Pax3 | ctcacctcaggtaatgggact (SEQ ID NO: 29) | cgtggtggtaggttccagac (SEQ ID NO: 30) |
| PAX7 ChIP 1, -731bp | cggggtcttgacattacaca (SEQ ID NO: 61) | gccagagtccgcccattttc (SEQ ID NO: 62) |
| PAX7 ChIP 2, -289bp | tattggtcctccgctccctt (SEQ ID NO: 63) | gtgagcgcgatctgataggt (SEQ ID NO: 64) |
| PAX7 ChIP 3, +562bp | ttgccgactttggattcgtc (SEQ ID NO: 65) | tccaaagggaaatcccgtgc (SEQ ID NO: 66) |
| PAX7 ChIP 4, +926 | cgcagggtgaaattctggt (SEQ ID NO: 67) | agagccgagaaaactgtcagg (SEQ ID NO: 68) |

SEQ ID NO: 31

ngg

SEQ ID NO: 32

nga

SEQ ID NO: 33

ngan

SEQ ID NO: 34

ngng

SEQ ID NO: 35

nggng

SEQ ID NO: 36

nnagaaw (W = A or T)

SEQ ID NO: 37

naar (R = A or G)

SEQ ID NO: 38

nngrrr (R = A or G; N can be any nucleotide residue, e.g., any of A, G, C, or T)

SEQ ID NO: 39

nngrrn (R = A or G; N can be any nucleotide residue, e.g., any of A, G, C, or T)

SEQ ID NO: 40

nngrrt (R = A or G; N can be any nucleotide residue, e.g., any of A, G, C, or T)

SEQ ID NO: 41

nngrrv (R = A or G; N can be any nucleotide residue, e.g., any of A, G, C, or T)

SEQ ID NO: 42

codon optimized polynucleotide encoding *S. pyogenes* Cas9

```

atggataaaa agtacagcat cgggctggac atcgggtacaa actcagtggg gtgggcccgtg
attacggacg agtacaaggt accctccaaa aaatttaaag tgctgggtaa cacggacaga
cactctataa agaaaaatct tattggagcc ttgctgttcg actcaggcga gacagccgaa
gccacaaggt tgaagcggac cgccaggagg cggtatacca ggagaaagaa ccgcatatgc
tacctgcaag aaatcttcag taacgagatg gcaaagggtg acgatagctt tttocatcgc
ctggaagaat cctttcttgt tgaggaagac aagaagcacg aacggcacc cctctttggc
aatattgtcg acgaagtggc atatcacgaa aagtaccgga ctatctacca cctcaggaag
aagctggtgg actctaccga taaggcggac ctcagactta tttatttggc actcgcccac
atgattaaat ttagaggaca tttcttgatc gagggcgacc tgaaccggga caacagtgac
gtcgataagc tgttcatcca acttgtgcag acctacaatc aactgttcga agaaaacct
ataaatgctt caggagtoga cgctaaagca atcctgtccg cgcgcctctc aaaatctaga
agacttgaga atctgattgc tcagttgccc ggggaaaaga aaaatggatt gtttggaac
ctgatcgccc tcagtctcgg actgacccca aatttcaaaa gtaacttcga cctggccgaa
gacgctaagc tccagctgtc caaggacaca tacgatgacg acctcgacaa tctgctggcc
cagattgggg atcagtacgc cgatctcttt ttggcagcaa agaacctgtc cgacgccatc
ctgttgagcg atatcttgag agtgaacacc gaaattacta aagcaccctc tagcgcatct
atgatcaagc ggtacgacga gcatcatcag gatctgacce tgctgaaggc tottgtgagg
caacagctcc ccgaaaaata caaggaatc ttctttgacc agagcaaaaa cggctacgct

```

ggctatatag atgggtggggc cagtcaggag gaattctata aattcatcaa gccattctc
 gagaaaatgg acggcacaga ggagttgctg gtcaaactta acagggagga cctgctgogg
 aagcagcggg cctttgacaa cgggtctatc ccccaccaga ttcattctggg cgaactgcac
 gcaatcctga ggaggcagga ggatttttat cctttctta aagataaccg cgagaaaata
 gaaaagattc ttacattcag gatcccgtac tacgtgggac ctctcgcccg gggcaattca
 cggtttgctt ggatgacaag gaagtcagag gagactatta caccttgtaa cttcgaagaa
 gtgggtggaca aggggtgatc tgcccagtct ttcattcgagc ggatgacaaa ttttgacaag
 aacctcccta atgagaaggt gctgcccata cttctctgc tctacgagta ctttaccgtc
 tacaatgaac tgactaaagt caagtagctc accgagggaa tgaggaagcc ggcattcctt
 agtggagAAC agaagaaggc gattgtagac ctgttgttca agaccaacag gaaggtgact
 gtgaagcaac ttaaagaaga ctactttaag aagatcgaat gttttgacag tgtggaaatt
 tcaggggttg aagaccgctt caatgctca ttggggactt accatgatct tctcaagatc
 ataaaggaca aagacttctt ggacaacgaa gaaaatgagg atattctcga agacatcgtc
 ctcaacctga cctgttcga agacagggaa atgatagaag agcgttgaa aacctatgcc
 cacctcttcg acgataaaagt tatgaagcag ctgaagcgcg ggagatacac aggatgggga
 agattgtcaa ggaagctgat caatggaatt agggataaac agagtggcaa gaccatactg
 gatttctca aatctgatgg ctctgccaat aggaacttca tgcaactgat tcaogatgac
 tctcttacct tcaaggagga cattcaaaag gctcaggtga gggggcaggg agactccctt
 catgaacaca tcgcgaattt ggacaggttc cccgctatta aaaagggcat ccttcaaact
 gtcaaggtgg tggatgaatt ggtcaaggtA atgggcagac ataagccaga aaatattgtg
 atcgagatgg cccgcgaaaa ccagaccaca cagaagggcc agaaaaatag tagagagcgg
 atgaagagga tcgaggaggg catcaaagag ctgggatctc agattctcaa agaacacccc
 gttagaaaaca cacagctgca gaacgaaaaa ttgtacttgt actatctgca gaacggcaga
 gacatgtacg tcgaccaaga acttgatatt aatagactgt ccgactatga cgtagaccat
 atcgtgcccc agtcccttct gaaggacgac tccattgata acaaagtctt gacaagaagc
 gacaagaaca ggggtaaaag tgataatgtg cctagcaggg aggtggtgaa aaaaatgaag
 aactactggc gacagctgct taatgcaaag ctcttacac aacggaagtt cgataatctg
 acgaaagcag agagaggtgg ctgtctgag ttggacaagg cagggtttat taagcggcag
 ctgggtgaaa ctaggcagat cacaagcac gtggcgcaga ttttgacag ccggatgaac
 acaaaatacg acgaaaatga taaactgata cgagaggtca aagttatcac gctgaaaagc
 aagctggtgt ccgattttct gaaagacttc cagttctaca aagttcgcga gattaataac
 taccatcatg ctcacgatgc gtacctgaac gctgttctcg ggaaccgctt gataaagaag
 taccocaaagc tgyaatccga gttcgtatac ggggattaca aagtgtacga tgtgaggaaa
 atgatagcca agtccgagca ggagattgga aaggccaag ctaagtactt cttttattct
 aacatcatga atttttttaa gacggaaatt accctggcca acggagagat cagaaaagcgg
 ccccttatag agacaaatgg tgaacaggt gaaatcgtct gggataaggg cagggatttc
 gctactgtga ggaaggtgct gagtatgcca caggtaaata tcgtgaaaaa aaccgaagta
 cagaccggag gattttccaa ggaaagcatt ttgcctaaaa gaaactcaga caagctcctc
 gcccgcaaga aagattgga ccctaagaaa tacgggggat ttgactcacc caccgtagcc
 tattctgtgc tgggtgtagc taaggtgaa aaaggaaagt ctaagaagct gaagtcogtg
 aaggaaactct tgggaatcac tatcatgaa agatcctct ttgaaaagaa cctatogat
 ttcttgaggt ctaagggtta caaggaggtc aagaaagacc tcatcattaa actgocaaaa
 tactctctct tcgagctgga aatggcagg aagagaatgt tggccagcc cggagagctg
 caaaagggaa acgagcttgc tctgccctcc aaatatgtta attttctcta tctcgcttc
 cactatgaaa agctgaaagg gtctcccgaa gataacgagc agaagcagct gttcgtcgaa
 cagcacaagc actatctgga tgaaataatc gaacaaataa gcgagttcag caaaaggggtt
 atcctggcgg atgctaattt ggacaaagta ctgtctgctt ataacaagca cgggataag
 cotattaggg aacaagccga gaatataatt cacctcttta cactcaogaa tctcggagcc
 cccgccct tcaaatactt tgatacgaact atcgaccgga aacggtatac cagtaccaaa
 gaggtctctg atgccacct catccaccag tcaattactg gcctgtacga
 aacacggatcgacctctc aactgggagg cgactag

SEQ ID NO: 43

Amino acid sequence of *Streptococcus pyogenes* Cas9

MDKKYSIGLDIGTNSVGVAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
 RRRYTRRKNRI CYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIY
 HLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINAS
 GVDAKAILLSARLSKSRRLLENLIAQLPGEKKNLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKDPTD
 DDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVR
 QQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFKIPILEKMDGTEELLVKNLREDLLRKQRTFDNG
 SIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPIYVVGPLARGNSRFAMTRKSEETITPW
 NFEVVDKGSASQSFIERMTNFDKNLPNEKVLPHKSLLEYEFTVYNELTKVKYVTEGMRKPAFLSSEQ
 KKAIVDLEFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEEN
 EDILEDIVLTLTLFEDREMI EERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLLINGIRDKQSGKTI L
 DFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGGDSLHEHIANLAGSPAIIKGILOTVKVVDELV
 KVMGRHKPENIVIEMARENQTTQKQKNSRERMKRIEEGKELGSQILKEHPVENTQLQNEKLYLYYL
 QNGRDMYVDQELDNRLSDYDVDHIVPQSFLKDDSIDNKVLRSDKNRSGKSDNVPSEEVVKKMKNYWR
 QLLNAKLITQRKFDNLTKAERGGSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI RE
 VKVITLKSCLVSDFRKDFQFYKVIENNYHHHADAYLNAVVGTAIIKYPKLESEFVYGDYKVVYDVRK
 MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL S
 MPQVNI VKKTEVQTTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGSKK
 LKSVKELLGITIMERSSEFKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
 ELALPSKYVNFYLLASHYEKLGSPEDNEQQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL S
 AYNKHRDKPIREQAENIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRI
 DLSQLGGD

SEQ ID NO: 44

codon optimized nucleic acid sequence encoding *S. aureus* Cas9

atgaaaagga actacattct ggggctggac atcgggatta caagcgtggg gtatgggatt
 attgactatg aaacaagggga cgtgatcgac gcaggcgtca gactggtcaa ggaggccaac
 gtggaaaaaca atgagggacg gagaagcaag aggggagcca ggcgcctgaa acgacggaga
 aggcacagaa tccagaggggt gaagaaactg ctgttcgatt acaacctgct gaccgaccat
 tctgagctga gtggaattaa tccttatgaa gccaggggtga aaggcctgag tcagaagctg
 tcagaggaag agttttccgc agctctgctg cacctggcta agcgcctgagg agtgcataac
 gtcaatgagg tggaaagagga caccggcaac gagctgtcta caaaggaaca gatctcacgc
 aatagcaaaag ctctggaaga gaagtatgtc gcagagctgc agctggaacg gctgaagaaa
 gatggcagagg tgagaggggtc aattaatagg ttcaagacaa gcgactacgt caaagaagcc
 aagcagctgc tgaagtgca gaaggcttac caccagctgg atcagagctt catcgatact
 tataatcgacc tgctggagac tcggagaacc tactatgagg gaccaggaga agggagcccc
 ttccgatgga aagacatcaa ggaatggtac gagatgctga tgggacattg cacctatfff
 ccagaagagc tgagaagcgt caagtacgct tataacgcag atctgtacaa cgccctgaat
 gacctgaaca acctggtcat caccagggat gaaaacgaga aactggaata ctatgagaag
 ttccagatca tcgaaaacgt gttaagcag aagaaaagc ctacactgaa acagattgct
 aaggagatcc tggtaacga agaggacatc aagggtacc ggggtgacaag cactggaaaa
 ccagagttca ccaatctgaa agtgtatcac gatattaagg acatcacagc acggaaagaa
 atcattgaga acgcccgaact gctggatcag attgctaaga tcctgactat ctaccagagc
 tccgaggaca tccaggaaga gctgactaac ctgaacagcg agctgaccca ggaagagatc
 gaacagatta gtaatctgaa ggggtacacc ggaacacaca acctgtccct gaaagctatc
 aatctgattc tggatgagct gtggcataca aacgacaatc agattgcaat ctttaaccgg
 ctgaagctgg tccccaaaaa ggtggacctg agtcagcaga aagagatccc aaccacactg
 gtggacgatt tcattctgtc acccgtggtc aagcggagct tcatccagag catcaaagtg
 atcaacgcca tcatcaagaa gtacggcctg cccaatgata tcattatcga gctggctagg
 gagaagaaca gcaaggacgc acagaagatg atcaatgaga tgcagaaacg aaaccggcag
 accaatgaac gcattgaaga gattatccga actaccggga aagagaacgc aaagtacctg
 attgaaaaaa tcaagctgca cgatatgcag gagggaaagt gtctgtatc tctggaggcc
 tccccctgg aggacctgct gaacaatcca ttcaactacg aggtcgatca tattatcccc
 agaagcgtgt ccttcgacaa ttcctttaac aacaaggtgc tggtaagca ggaagagaac
 totaaaaagg gcaataggac tcctttccag tacctgtcta gttcagatc caagatctct
 tacgaaaact ttaaaaagca cattctgaat ctggccaaag gaaagggccg catcagcaag

```

accaaaaagg agtacctgct ggaagagcgg gacatcaaca gattctccgt ccagaaggat
tttattaacc ggaatctggt ggacacaaga tacgctactc gcggcctgat gaatctgctg
cgatcctatt tccgggtgaa caatctggat gtgaaagtca agtccatcaa cggcgggttc
acatcttttc tgaggcgcaa atggaagttt aaaaaggagc gcaacaaagg gtacaagcac
catgccgaag atgctctgat tatcgcaaat gccgacttca tctttaagga gtggaaaaag
ctggacaaaag ccaagaaaag gatggagaac cagatgttcg aagagaagca ggccgaatct
atgcccgaaa tcgagacaga acaggagtac aaggagattt tcatcactcc tcaccagatc
aagcatatca aggatttcaa ggactacaag tactctcacc gggtgataa aaagcccaac
agagagctga tcaatgacac cctgtatagt acaagaaaag acgataaggg gaataccctg
attgtgaaca atctgaacgg actgtacgac aaagataatg acaagctgaa aaagctgac
aacaaaagtc ccgagaagct gctgatgtac caccatgac ctcagacata tcagaaactg
aagctgatta tggagcagta cggcgacgag aagaaccac tgtataagta ctatgaagag
actgggaact acctgaccaa gtatagcaaa aaggataatg gccccgtgat caagaagatc
aagtactatg ggaacaagct gaatgcccat ctggacatca cagacgatta ccctaacagt
cgcaacaagg tggccaagct gtcactgaag ccatacagat tcgatgtcta tctggacaac
ggcgtgtata aatttgtgac tgtcaagaat ctggatgtca tcaaaaagga gaactactat
gaagtgaata gcaagtgcta cgaagaggct aaaaagctga aaaagattag caaccaggca
gagttcatcg cctcctttta caacaacgac ctgattaaga tcaatggcga actgtatagg
gtcatcgggg tgaacaatga tetgctgaac cgcattgaag tgaatatgat tgacatcact
taccgagagt atctggaaaa catgaatgat aagcgcccc ctcgaattat caaaaatt
gcctctaaga ctcagagtat caaaaagtac tcaaccgaca ttctgggaaa cctgtatgag
gtgaagagca aaaagcacc tcagattatc aaaaagggc

```

SEQ ID NO: 45

codon optimized nucleic acid sequence encoding *S. aureus* Cas9

```

atgaagcggg actacatcct gggcctggac atcggcatca ccagcgtggg ctacggcacc
atcgactacg agacacggga cgtgatcgat gccggcgtgc ggctgttcaa agaggccaac
gtggaaaaaca acgagggcag gcggagcaag agaggcgcca gaaggctgaa gcggcggagg
cggcatagaa tccagagagt gaagaagctg ctgttcgact acaacctgct gaccgaccac
agcgagctga gcggcatcaa cccctacgag gccagagtga agggcctgag ccagaagctg
agcgagggaag agttctctgc cgcctgctg cacctggcca agagaagagg cgtgcacaac
gtgaacgagg tggaaagagg caccggcaac gagctgtcca ccaaagagca gatcagccgg
aacagcaagg ccctggaaga gaaatacgtg gccgaactgc agctggaacg gctgaagaaa
gacggcgaaag tgcggggcag catcaacaga ttcaagacca gcgactacgt gaaagaagcc
aaacagctgc tgaagggtgca gaaggcctac caccagctgg accagagctt catcgacacc
tacatcgacc tgctggaaac ccggcggacc tactatgagg gacctggcga gggcagcccc
ttcggctgga aggacatcaa agaatggtac gagatgctga tgggccactg cacctacttc
cccgaggaac tgcggagcgt gaagtacgcc tacaacgccg acctgtaca cyccctgaac
gacctgaaca atctcgtgat caccagggac gagaacgaga agctggaata ttacgagaag
ttccagatca tcgagaacgt gttcaagcag aagaagaagc ccacctgaa gcagatcgcc
aaagaaatcc tcgtgaacga agaggatatt aagggtaca gagtgaccag caccggcaag
cccgagttca ccaacctgaa ggtgtaccac gacatcaagg acattaccgc ccggaaagag
attattgaga acgccgagct gctggatcag attgccaaga tccctgacct ctaccagagc
agcgaggaca tccaggaaga actgaccaat ctgaactccg agctgaccga ggaagagatc
gagcagatct ctaatctgaa gggctatacc ggcaccaca acctgagcct gaaggccatc
aacctgatcc tggacgagct gtggcacacc aacgacaacc agatcgctat cttcaaccgg
ctgaagctgg tgcccaagaa ggtggacctg tcccagcaga aagagatccc caccaccctg
gtggacgact tcatcctgag ccccgctcgtg aagagaagct tcatccagag catcaaagtg
atcaacgcca tcatcaagaa gtacggcctg cccaacgaca tcattatcga gctggccccg
gagaagaact ccaaggacgc ccagaaaatg atcaacgaga tgcagaagcg gaaccggcag
accaacgagc ggatcgagga aatcatccgg accaccggca aagagaacgc caagtacctg
atcgagaaga tcaagctgca cgacatgcag gaaggcaagt gcctgtacag cctggaagcc
atccctctgg aagatctgct gaacaacccc ttcaactatg aggtggacca catcatcccc
agaagcgtgt ccttcgacaa cagcttcaac aacaaggtgc tcgtgaagca ggaagaaaac
agcaagaag gcaaccggac ccattccag tacctgagca gcagcgacag caagatcagc

```



```

tacgaaacct tcaagaagca catcctgaat ctggccaagg gcaagggcag aatcagcaag
accaagaaaag agtatctgct ggaagaacgg gacatcaaca ggttctccgt gcagaaagac
ttcatcaacc ggaacctggt ggataccaga tacgccacca gaggcctgat gaacctgctg
cggagctact tcagagtga caacctggac gtgaaagtga agtccatcaa tggcggttc
accagctttc tgcgcgga gtggaagttt aagaaagagc ggaacaaggg gtacaagcac
cacgccgagg acgacctgat cattgccaac gccgatttca tcttcaaaga gtggaagaaa
ctggacaagg ccaaaaaagt gatggaaaac cagatgttcg aggaaaagca ggccgagagc
atgcccgaga tcgaaaccga gcaggagtac aaagagatct tcatcaccoc ccaccagatc
aagcacatta aggacttcaa ggactacaag tacagccacc ggggtggacaa gaagcctaata
agagagctga ttaacgacac cctgtactcc acccggaagg acgacaaggg caacacccctg
atcgtgaaca atctgaacgg cctgtacgac aaggacaatg acaagctgaa aaagctgatc
aacaagagcc ccgaaaagct gctgatgtac caccacgacc ccagaccta ccagaaactg
aagctgatta tggaaacagta cggcgacgag aagaatcccc tgtacaagta ctacgaggaa
accgggaact acctgaccaa gtactccaaa aaggacaacg gccccgtgat caagaagatt
aagtattacg gcaacaaaact gaacgccccat ctggacatca ccgacgacta cccaacagc
agaaacaagg tcgtgaagct gtcctgaag cctacagat tcgacgtgta cctggacaat
ggcgtgtaca agttcgtgac cgtgaagaat ctggatgtga tcaaaaaaga aaactactac
gaagtgaata gcaagtgcta tgaggaagct aagaagctga agaagatcag caaccaggcc
gagtttatcg cctcctteta caacaacgat ctgatcaaga tcaacggcga gctgtataga
gtgatcggcg tgaacaacga cctgctgaac cggatcgaag tgaacatgat cgacatcacc
taccgagagt acctggaaaa catgaacgac aagaggcccc ccaggatcat taagacaatc
gctccaaga ccagagcat taagaagtac agcacagaca ttctgggcaa cctgtatgaa
gtgaaatcta agaagcacc tcagatcatc aaaaagggc

```

SEQ ID NO: 46

codon optimized nucleic acid sequence encoding *S. aureus* Cas9

```

atgaagcgca actacatcct cggactggac atcggcatta cctccgtggg atacggcatc
atcgattacg aaactagggg tgtgatcgac gctggagtca ggctgttcaa agaggcgaac
gtggagaaca acgagggggcg gcgctcaaag agggggggccc gccggctgaa gcgcccgcgc
agacatagaa tccagcgcgt gaagaagctg ctgttcgact acaaccttct gaccgaccac
tccgaacttt ccggcatcaa cccatatgag gctagagtga agggattgtc caaaagctg
tccgaggaag agttctccgc cgcgttgcct cacctcgcca agcgcagggg agtgcacaat
gtgaacgaag tggaaagaaga taccggaaac gagctgtcca ccaaggagca gatcagccgg
aactccaagg ccctggaaga gaaatcctg gcggaactgc aactggagcg gctgaagaaa
gacggagaag tgcgcggtc gatcaaccgc ttcaagacct cggactacgt gaaggaggcc
aagcagctcc tgaagtgc aaggcctat caccaacttg accagtcctt tatcgatacc
tacatcgatc tgctcgagac tcggcggact tactacgagg gtccagggga gggctcccca
tttggttggg aggatattaa ggagtggtag gaaatgctga tgggacactg cacatacttc
cctgaggagc tcgaggagct gaaatcgc tacaacgcag acctgtacaa cgcgctgaac
gacctgaaca atctcgtgat caccggggac gagaacgaaa agctcgagta ttacgaaaag
ttccagatta ttgagaacgt gttcaaacag aagaagaagc cgacactgaa gcagattgcc
aaggaaatcc tcgtgaacga agaggacatc aagggtatc gactgacctc aacgggaaaag
ccggagttea ccaatctgaa ggtctaccac gatcaaaag acattaccgc ccggaaggag
atcattgaga acgcgagct gttggaccag attgcgaaga ttctgacct ctaccaatcc
tccgaggata ttcaggaaga actcaccaac ctcaacagcg aactgacca ggaggagata
gagcaaatct ccaacctgaa gggctacacc ggaactcata acctgacct gaaggccatc
aacttgatcc tggacgagct gtggcacacc aacgataacc agatcgctat tttcaatcgg
ctgaagctgg tccccaaaga agtggacctc tcacaacaaa aggagatccc tactaccctt
gtggacgatt tcattctgtc ccccggtggtc aagagaagct tcatacagtc aatcaaagtg
atcaatgcca ttatcaagaa atacggtctg cccaacgaca ttatcattga gctcgccgc
gagaagaact cgaaggacgc ccagaagatg attaacgaaa tgcagaagag gaaccgacag
actaacgaac ggatcgaaga aatcatccgg accaccggga aggaaaacgc gaagtacctg
atcgaaaaaga tcaagctcca tgacatgcag gaaggaaagt gtctgtactc gctggaggcc
attccgctgg aggacttgc gaacaaccct ttaactacg aagtggatca tatcattccg
aggagcgtgt cattcgacaa ttccttcaac aacaaggtcc tcgtgaagca ggaggaaaac

```

tcgaagaagg gaaaccgcac gccgttccag tacctgagca gcagcgactc caagatttcc
 tacgaaacct tcaagaagca catcctcaac ctggcaaagg ggaaggggtcg catctccaag
 accaagaagg aatatctgct ggaagaaaga gacatcaaca gattctccgt gcaaaaggac
 ttcatacaacc gcaacctcgt ggatactaga tacgctactc ggggtctgat gaacctctg
 agaagctact ttagagttaa caatctggac gtgaaggta agtcgattaa cggaggttcc
 acctccttcc tgcggcgcaa gtggaagttc aagaaggaac ggaacaaggg ctacaagcac
 cacgccgagg acgacctgat cattgccaac gccgacttca tcttcaaaga atggaagaaa
 cttgacaagg ctaagaaggt catggaaaac cagatgttcg aagaaaagca ggccgagttc
 atgcctgaaa tcgagactga acaggagtac aaggaaatct ttattacgcc acaccagatc
 aaacacatca aggatttcaa ggattacaag tactcacatc gcgtggacaa aaagccgaac
 agggaaactga tcaacgcac cctctactcc acccggaagg atgacaaagg gaatacctc
 atcgtcaaca accttaacgg cctgtacgac aaggacaacg ataagctgaa gaagctcatt
 aacaagtcgc ccgaaaagtt gctgatgtac caccacgacc ctacagacta ccagaagctc
 aagctgatca tggagcagta tggggacgag aaaaaccctg tgtacaagta ctacgaagaa
 actgggaatt atctgactaa gtactccaag aaagataacg gccccgtgat taagaagatt
 aagtactacg gcaacaagct gaacgcccat ctggacatca ccgatgacta ccctaattcc
 cgcaacaagg tcgtcaagct gagcctcaag ccctaccggt ttgatgtgta ccttgacaat
 ggagtgtaca agttcgtgac tgtgaagaac cttgacgtga tcaagaagga gaactactac
 gaagtcaact ccaagtgcta cgaggaagca aagaagttga agaagatctc gaaccaggcc
 gagttcattg cctccttcta taacaacgac ctgattaaga tcaacggcga actgtaccgc
 gtcattggcg tgaacaacga tctcctgaac cgcacgaag tgaacatgat cgacatcact
 taccgggaat acctggagaa tatgaacgac aagcggccgc cccggatcat taagactatc
 gcctcaaaga ccagtcgat caagaagtac agcaccgaca tcctgggcaa cctgtacgag
 gtcaaatcga agaagcacc ccagatcatc aagaagga

SEQ ID NO: 47

codon optimized nucleic acid sequence encoding *S. aureus* Cas9

atggccccaaagaagaagcgggaaggtcgggtatccacggagtcccagcagccaagoggaaactacatcct
 gggcctggacatcggcatcaccagcgtgggctacggcatcatcgactacgagacacgggacgtgatcg
 atgccggtgctggcgtgttcaaagaggccaacgtggaaaaaacgagggcaggcggagcaagagaggc
 gccagaaggctgaagcggcggagggcggcatagaatccagagagtgaaagaagctgctgttcgactaca
 cctgctgaccgaccacagcagctgagcggcatcaacccctacgagggccagagtgaaagggcctgagcc
 agaagctgagcaggaagagttctctgcccctgctgcaacctggccaagagaagagggcgtgcacaac
 gtgaacgaggtggaagaggacaccggcaacgagctgtccaccagagagcagatcagccggaaacagcaa
 ggccctggaagagaaatacgtggccgaactgacgtggaaacggctgaagaaagacggcgaagtgccggg
 gcagcatcaacagattcaagaccagcgaactacgtgaaagaagccaaacagctgctgaaggtgcagaag
 gcctaccaccagctggaccagagcttcatcgacacctacatcgacctgctggaaccggcggacctta
 ctatgagggacctggcagggcagccccttcggctggaaggacatcaaagaatggtacgagatgctga
 tgggccaactgcacctacttccccgaggaaactgcccggcgtgaagtacgctacaacgccgacctgtac
 aacgccctgaacgacctgaacaatctcgtgatcaccagggacgagaaacgagaagctggaatattacga
 gaagttccagatcatcgagaacgtgttcaagcagaagaagaagcccaacctgaagcagatcgccaaag
 aaatcctcgtgaaacgaagaggatattaagggtacagagtgaccagcaccggcaagcccaggttccacc
 aacctgaaggtgtaccaagcatcaaggacattaccgcccggaaagagattattgagaacgcccagct
 gctggatcagattgccaaagtcctgacctctaccagagcagcagagcatccaggaagaactgacca
 atctgaactccgagctgaccaggaagagatcgagcagatctctaactctgaagggctataccggcacc
 cacaacctgagcctgaaggccatcaacctgatcctggacgagctgtggcacaccaacgacaaccagat
 cgctatcttcaaccggctgaagctggtgcccagaaggtggacctgtcccagcagaaagagatcccca
 ccacctggtggaacgacttcatcctgagccccgtcgtgaagagaagcttcatccagagcatcaaaagtg
 atcaacgccatcatcaagaagtacggcctgcccacgacatcattatcgagctggccccggagaagaa
 ctccaaggacgccagaaaaatgatcaacgagatgcagaagcggaaacggcagaccacagagcggatcg
 aggaaatcatccggaccaccggcaagagaaacgccaagtacctgatcgagaagatcaagctgcacgac
 atgcaggaaggcaagtgctgtacagcctggaagccatccctctggaagatctgctgaacaacccctt
 caactatgaggtggaaccatcatcccagaagcgtgtccttcgacaacagcttcaacaaacaggtgc
 tcgtgaagcaggaagaaaacagcaagaagggcaaccggacccattccagtaacctgagcagcagcgac
 agcaagatcagctacgaaacctcaagaagcactcctgaatctggccaagggcaagggcagaatcag

caagaccaagaaagagatctctgctggaagaacgggacatcaacaggttctccgtgcagaaagacttca
tcaaccggaaacctggtggataccagatacggcaccagaggcctgatgaacctgctgcgaggctacttc
agagtgaacaacctggacgtgaaagtgaagtccatcaatggcggcttcaccagctttctgcgggcggaa
gtggaagtttaagaaagagcggaaacaaggggtacaagcaccacgcccaggacgcctgatcattgcca
acgcccgatctcatcttcaaagagtggaaagaaactggacaaggccaaaaaagtgatggaaaaaccagatg
ttcgaggaaaggcaggccgagagcatgcccagatcgaaaaccgagcaggagtaaaaagagatcttcat
caccaccaccagatcaagcacattaaggacttcaaggactacaagtacagccaccgggtggacaaga
agcctaataagagagctgattaacgacacctgtactccaccgggaaggacgacaagggcaaacacctg
atcgtgaacaatctgaacggcctgtacgacaaggacaatgacaagctgaaaaagctgatcaacaagag
ccccgaaaagctgctgatgtaccaccacgacccccagacctaccagaaactgaagctgattatggaac
agtacggcgacgagaagaatccccgtacaagtactacgaggaaaccgggaactacctgaccaagtac
tccaaaaaggacaacggccccctgatcaagaagattaagattacggcaacaaactgaacgcccactct
ggacatcaccgacgactaccccaacagcagaaaacaaggtcgtgaagctgtccctgaagccctacagat
tcgacgtgtacctggacaatggcgtgtacaagttcgtgaccgtgaagaatctggatgtgatcaaaaaa
gaaaactactacgaagtgaatagcaagtgctatgaggaagctaagaagctgaagaagatcagcaacca
ggccgagtttatcgctcctctctacaacaacgatctgatcaagatcaacggcgagctgtatagagtga
tcggcgtgaaacacgacctgctgaaccggatcgaagtgaacatgatcgacatcacctaaccgaggtac
ctggaaaaacatgaacgacaagaggccccccaggatcattaagacaatcgccctccaagaccagagcat
taagaagtacagcacagacattctgggcaacctgtatgaagtgaatctaagaagcacacctcagatca
tcaaaaagggcaaaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaag

SEQ ID NO: 48

codon optimized nucleic acid sequence encoding *S. aureus* Cas9

accggtgcca ccatgtacc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc
aaggtcgaag cgtccatgaa aaggaactac attctggggc tggacatcgg gattacaagc
gtgggggatg ggattattga ctatgaaaca agggacgtga tcgacgcagg cgtcagactg
ttcaaggagg ccaacgtgga aaacaatgag ggacggagaa gcaagagggg agccaggcgc
ctgaaacgac ggagaaggca cagaatccag aggggtgaaga aactgctgtt cgattacaac
ctgctgaccg accattctga gctgagtgga attaatcctt atgaagccag ggtgaaaggc
ctgagtcaga agctgtcaga ggaagagttt tccgcagctc tgctgcacct ggctaagcgc
cgaggagtgc ataacgtcaa tgaggtggaa gaggacaccg gcaacgagct gtctacaaag
gaacagatct cacgcaatag caaagctctg gaagagaagt atgtcgcaga gctgcagctg
gaacggctga agaaagatgg cgaggtgaga gggtcatta ataggttcaa gacaagcgcac
tacgtcaaaag aagccaagca gctgctgaaa gtgcagaagg cttaccacca gctggatcag
agcttcatcg atacttatat cgacctgctg gagactcggg gaacctacta tgagggacca
ggagaaggga gcccttcgg atggaaagac atcaaggaat ggtacgagat gctgatggga
cattgcaacct atttccaga agagctgaga agcgtcaagt acgcttataa cgcagatct
tacaacgccc tgaatgacct gaacaacctg gtcatcacca gggatgaaaa cgagaaactg
gaatactatg agaagttcca gatcatcgaa aacgtgttta agcagaagaa aaagcctaca
ctgaaacaga ttgctaagga gatcctggtc aacgaagagg acatcaaggg ctaccgggtg
acaagcactg gaaaaccaga gttcaccat ctgaaagtgt atcacgatat taaggacatc
acagcacgga aagaaatcat tgagaacgcc gaactgctgg atcagattgc taagatcctg
acctatctaac agagctccga ggacatccag gaagagctga ctaacctgaa cagcagctg
accaggaag agatcgaaca gattagtaat ctgaaggggt acaccggaac acacaacctg
tccctgaaaag ctatcaatct gattctggat gagctgtggc atacaaacga caatcagatt
gcaatcttta accggctgaa gctggtecca aaaaagggtg acctgagtca gcagaaagag
atcccaacca cactggtgga cgatttcatt ctgtcaccgg tggtaagcg gagcttcac
cagagcatca aagtgatcaa cgccatcctc aagaagtacg gcctgcccga tgatatcatt
atcgagctgg ctagggagaa gaacagcaag gacgcacaga agatgatcaa tgagatgcag
aaacgaaacc ggacagacca tgaacgcatt gaagagatta tccgaactac cgggaaagag
aacgcaaagt acctgattga aaaaatcaag ctgcacgata tgcaggaggg aaagtgtctg
tattctctgg aggccatccc cctggaggac ctgctgaaca atccattcaa ctacgaggtc
gatcatatta tcccagaag cgtgtccttc gacaattcct ttaacaacaa ggtgctggtc
aagcaggaag agaactctaa aaagggcaat aggactcctt tccagtacct gtctagttca
gattccaaga tctcttacga aacctttaa aagcacatc tgaatctggc caaaggaaaag

ggccgcatca gcaagaccaa aaaggagtac ctgctggaag agcgggacat caacagatcc
 tccgtccaga aggattttat taaccggaat ctgggtggaca caagatacgc tactcgcggc
 ctgatgaatc tgctgcgatc ctatttccgg gtgaacaatc tggatgtgaa agtcaagtcc
 atcaacggcg gggtcacatc ttttctgagg cgcaaagga agtttaaaaa ggagcgcaac
 aaagggatca agcaccatgc cgaagatgct ctgattatcg caaatgccga cttcatcttt
 aaggagtgga aaaagctgga caaagccaag aaagtgatgg agaaccagat gttcgaagag
 aagcaggccg aatctatgcc cgaaatcgag acagaacagg agtacaagga gattttcatc
 actcctcacc agatcaagca tatcaaggat ttcaaggact acaagtactc tcaccgggtg
 gataaaaagc ccaacagaga gctgatcaat gacaccctgt atagtacaag aaaagacgat
 aaggggaata ccttgattgt gaacaatctg aacggactgt acgacaaaga taatgacaag
 ctgaaaaagc tgatcaacaa aagtcccgag aagctgctga tgtaccacca tgatcctcag
 acatatacaga aactgaagct gattatggag cagtacggcg acgagaagaa cccactgtat
 aagtactatg aagagactgg gaactacctg accaagtata gcaaaaagga taatggcccc
 gtgatcaaga agatcaagta ctatgggaac aagctgaatg cccatctgga catcacagac
 gattacccta acagtcgcaa caagggtggtc aagctgtcac tgaagccata cagattcgat
 gtctatctgg acaacggcgt gtataaattt gtgactgtca agaactctgga tgcatacaaa
 aaggagaact actatgaagt gaatagcaag tgctacgaag aggctaaaaa gctgaaaaag
 attagcaacc aggcagagtt catcgctcc ttttacaaca acgacctgat taagatcaat
 ggogaactgt atagggatc cgggggtgaac aatgatctgc tgaaccgat tgaagtgaat
 atgattgaca tcacttaccg agagtatctg gaaaacatga atgataagcg cccccctcga
 attatcaaaa caattgcctc taagactcag agtatcaaaa agtactcaac cgacattctg
 ggaaaacctgt atgaggtgaa gagcaaaaag caccctcaga ttatcaaaa gggctaagaa
 ttc

SEQ ID NO: 49

Amino acid sequence of *Staphylococcus aureus* Cas9

MKRNYI LGLDIGITSVGYGI IDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVK
 KLLFDYNLLTDHSELGINPYEARVKGLS QKLS EEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKE
 QISRNSKALEEKYVAELQLERLKKDGEVRGS INRFKTS DYVKEAKQLLKVQKAYHQLDQS FIDTYIDL
 LETRRTY YEGPGEPS PFGWKDI KEWY EMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVI TRDEN
 EKLEYEYKFPQI IENVFKQKKKPTLKQ IAKEI LVNEEDI KGYRVTSTGKPEFTNLKVYHDI KDITARKE
 I IENABELLDQIAK I LTIYQS SEDIQEELTNLNS ELTQEEI EQI SNLKGYTGTNLSL KAINLI LDELW
 HTNDNQIAI FNRLKLV PPKVDLSQQKEI PTTLVDDF I LSPVVKRSFI QSIKVINAI IKKYGLPNDI I I
 ELAREKNSKDAQKMINEMQKRNRQTNERIEE I I RTTGKENAKYLI EKI KLHDMQEGKCLYSLEAI PLE
 DLLNPFNYEVDHII PRSVSFDNS FNNKVLVKQEENS KKGNRTPFQYLS SSSDSKISYETFKKHILNLA
 KGKGRISKTKEYLLEERDINRF SVQKDFINRN LVDTRYATRGLMNL LRSYFRVNNLDVKVKS INGGF
 TSFLRPKWKFKKER NKGYKHHAEDALI IANADFI FKWKKLDKAKKVMENQMFEEKQAESMPEIETE Q
 EYKEI F I T P H Q I K H I K D F K D Y K Y S H R V D K K P N R E L I N D T L Y S T R K D D K G N T L I V N N L N G L Y D K D N D K L
 KKLINKSPEKLLMYHHPQTYQKLKLIMEQY GDEKNFLYKYEEETGNYLTKYSKKDNGPVIKKIKYYG
 NKLNAHLDI TDDYPNSRNKVVKLS LKPYRFDVYLDNGVYKFVTVKNLDVIKKENYEVNSKCYEEAKK
 LKKI SNQAEFIAS FYNNDLI KINGEL YRVIGVNNDLLNRI EVNMIDIT YREYLENMNDKRPPRI IKTI
 ASKTQSIKKYSTDILGNLYEVKSKKHPQII KKG

SEQ ID NO: 50

Nucleic acid sequence encoding D10A mutant of *S. aureus* Cas9

atgaaaagga actacattct ggggctggcc atcgggatta caagcgtggg gtatgggatt
 attgactatg aaacaagggg cgtgatcgac gcaggcgtca gactgttcaa ggaggccaac
 gtgaaaaaca atgagggacg gagaagcaag aggggagcca ggcgcctgaa acgacggaga
 aggcacagaa tccagagggg gaagaaactg ctgttcgatt acaacctgct gaccgacct
 tctgagctga gtggaattaa tccttatgaa gccaggggtga aaggcctgag tcagaagctg
 tcagaggaag agttttccgc agctctgctg cacctggcta agcgcggagg agtgcataac
 gtcaatgagg tggaaagagg caccggcaac gagctgtcta caaaggaaca gatctcacgc
 aatagcaaag ctctggaaga gaagtatgtc gcagagctgc agctggaacg gctgaagaaa
 gatggcgagg tgagaggggtc aattaatagg ttcaagacaa gcgactacgt caaagaagcc
 aagcagctgc tgaaagtgca gaaggcttac caccagctgg atcagagctt catcgatact

tatatcgacc tgctggagac tccgagaacc tactatgagg gaccaggaga agggagcccc
 ttccgatgga aagacatcaa ggaatggtag gagatgctga tgggacattg cacctatfff
 ccagaagagc tgagaagcgt caagtacgct tataacgcag atctgtacaa cgcctgaat
 gacctgaaca acctggtcat caccagggat gaaaacgaga aactggaata ctatgagaag
 ttccagatca tcgaaaacgt gtttaagcag aagaaaaagc ctacactgaa acagattgct
 aaggagatcc tgggtcaacga agaggacatc aagggctacc ggggtgacaag cactggaaaa
 ccagagttca ccaatctgaa agtgtatcac gatattaagg acatcacagc acggaaagaa
 atcattgaga acgccgaact gctggatcag attgctaaga tcctgactat ctaccagagc
 tccgaggaca tccaggaaga gctgactaac ctgaacagcg agctgaccca ggaagagatc
 gaacagatta gtaatctgaa ggggtacacc ggaacacaca acctgtccct gaaagctatc
 aatctgattc tggatgagct gtggcataca aacgacaatc agattgcaat ctttaaccgg
 ctgaagctgg tcccaaaaaa ggtggacctg agtcagcaga aagagatccc aaccacactg
 gtggacgatt tcattctgtc acccgtggtc aagcggagct tcatccagag catcaaagtg
 atcaacgcca tcatcaagaa gtacggcctg cccaatgata tcattatcga gctggctagg
 gagaagaaca gcaaggacgc acagaagatg atcaatgaga tgcagaaacg aaaccggcag
 accaatgaac gcattgaaga gattatccga actaccggga aagagaacgc aaagtacctg
 attgaaaaaa tcaagctgca cgatatgcag gagggaaagt gtctgtatcc totggaggcc
 atccccctgg aggacctgct gaacaatcca ttcaactacg aggtcgatca tattatcccc
 agaagcgtgt ccttcgacaa ttcctttaac aacaaggtgc tgggtcaagca ggaagagaac
 tctaaaaagg gcaataggac tctttccag tacctgtcta gttcagatcc caagatctct
 tacgaaaact ttaaaaaagca cattctgaat ctggccaaag gaaagggccg catcagcaag
 accaaaaagg agtacctgct ggaagagcgg gacatcaaca gattctccgt ccagaaggat
 tttattaacc ggaatctggt ggacacaaga tacgctactc gcggcctgat gaatctgtg
 cgatcctatt tccgggtgaa caatctggat gtgaaagtca agtccatcaa cggcgggttc
 acatcttttc tgagggcycaa atggaagttt aaaaaggagc gcaacaaagg gtacaagcac
 catgccgaag atgctctgat tatcgcaaat gccgacttca tctttaagga gtggaaaaag
 ctggacaaaag ccaagaaagt gatggagaac cagatgttcg aagagaagca ggccgaatct
 atgcccgaaa tcgagacaga acaggagtac aaggagattt tcatcactcc taccagatc
 aagcatatca aggatttcaa ggactacaag tactctcacc ggggtggataa aaagcccaac
 agagagctga tcaatgacac cctgtatagt acaagaaaag acgataaggg gaataccctg
 attgtgaaca atctgaacgg actgtacgac aaagataatg acaagctgaa aaagctgatc
 acaaaaagtc ccgagaagct gctgatgtac caccatgatc ctacagacata tcagaaactg
 aagctgatta tggagcagta cggcgacgag aagaaccac tgtataagta ctatgaagag
 actgggaact acctgaccaa gtatagcaaa aaggataatg gccccgtgat caagaagatc
 aagtactatg ggaacaagct gaatgcccat ctggacatca cagacgatta ccctaacagt
 cgcaacaagg tgggtcaagct gtcaactgaag ccatacagat tcgatgtcta tctggacaac
 ggcgtgtata aatthgtgac tgtcaagaat ctggatgtca tcaaaaagga gaactactat
 gaagtgaata gcaagtgcta cgaagaggct aaaaagctga aaaagattag caaccaggca
 gagttcatcg cctcctttta caacaacgac ctgattaaga tcaatggcga actgtatagg
 gtcacogggg tgaacaatga tctgtctgac cgcattgaag tgaatatgat tgacatcact
 taccgagagt atctggaaaa catgaatgat aagcgccccc ctcgaattat caaaaatt
 gcctctaaga ctcagagtat caaaaagtac tcaaccgaca ttctgggaaa cctgtatgag
 gtgaagagca aaaagcacc tcagattatc aaaaagggc

SEQ ID NO: 51

Nucleic acid sequence encoding N580A mutant of *S. aureus* Cas9

atgaaaagga actacattct ggggctggac atcgggatta caagcgtggg gtatgggatt
 attgactatg aaacaagggg cgtgatcgac gcaggcgtca gactgttcaa ggaggccaac
 gtggaaaaaca atgagggagc gagaagcaag aggggagcca ggcgcctgaa acgacggaga
 aggcacagaa tccagagggg gaagaaactg ctgttcgatt acaacctgct gaccgacctat
 totgagctga gtggaattaa tcttatgaa gccagggtga aaggcctgag tcagaagctg
 tcagaggaag agttttccgc agctctgctg cacctggcta agcgcggagg agtgcataac
 gtcaatgagg tgyaagagga caccggcaac gagctgtcta caaaggaaca gatctcacgc
 aatagcaaaag ctctggaaga gaagtatgtc gcagagctgc agctggaacg gctgaagaaa
 gatggcgagg tgagagggtc aattaatagg ttcaagacaa gcgactacgt caaagaagcc

aagcagctgc tgaagtgca gaaggcttac caccagctgg atcagagctt catcgatact
tatatcgacc tgctggagac tgggagaacc tactatgagg gaccaggaga agggagcccc
ttcggatgga aagacatcaa ggaatggtac gagatgctga tgggacattg cacctatctt
ccagaagagc tgagaagcgt caagtacgct tataacgcag atctgtacaa cgcctgaat
gacctgaaca acctggtcat caccaggat gaaaacgaga aactggaata ctatgagaag
ttccagatca tcgaaaaagt gtttaagcag aagaaaaagc ctacactgaa acagattgct
aaggagatcc tgggtcaacga agaggacatc aagggctacc ggggtgacaag cactggaaaa
ccagagttca ccaatctgaa agtgtatcac gatattaagg acatcacagc acggaaagaa
atcattgaga acgcccgaact gctggatcag attgctaaga tcctgactat ctaccagagc
tccgaggaca tccaggaaga gctgactaac ctgaacagcg agctgaccca ggaagagatc
gaacagatta gtaatctgaa ggggtacacc ggaacacaca acctgtccct gaaagctatc
aatctgattc tggatgagct gtggcataca aacgacaatc agattgcaat ctttaaccgg
ctgaagctgg tcccaaaaaa ggtggacctg agtcagcaga aagagatccc aaccaactg
gtggacgatt tcaatctgtc acctgtggtc aagcggagct tcatccagag catcaaagtg
atcaacgcca tcatcaagaa gtacggcctg cccaatgata tcattatcga gctggctagg
gagaagaaca gcaaggacgc acagaagatg atcaatgaga tgcagaaacg aaaccggcag
accaatgaac gcattgaaga gattatccga actaccggga aagagaacgc aaagtacctg
attgaaaaaa tcaagctgca cgatatgcag gagggaaagt gtctgtatc tctggaggcc
atccccctgg aggacctgct gaacaatcca ttcaactacg aggtcgatca tattatcccc
agaagcgtgt ccttcgacaa ttcctttaac aacaaggtgc tgggtcaagca ggaagaggcc
tctaaaaaag gcaataggac tctttccag tacctgtcta gttcagatc caagatctct
tacgaaaact ttaaaaaagc cattctgaat ctggccaaag gaaagggccg ctagcaag
accaaaaagc agtacctgct ggaagagcgg gacatcaaca gattctccgt ccagaaggat
tttattaacc ggaatctggt ggacacaaga tacgctactc gcggcctgat gaatctgctg
cgatcctatt tccgggtgaa caatctggat gtgaaagtca agtccatcaa cggcgggttc
acatcttttc tgagggcycaa atggaagttt aaaaaggagc gcaacaaagg gtacaagcac
catgccgaag atgctctgat tatcgcaaat gccgacttca tctttaagga gtggaaaaag
ctggacaaaag ccaagaaaagt gatggagaac cagatgttcg aagagaagca ggccgaatct
atgcccgaaa tcgagacaga acaggagtac aaggagattt tcatcactcc taccagatc
aagcatatca aggatttcaa ggactacaag tactctcacc ggggtggataa aaagcccaac
agagagctga tcaatgacac cctgtatagt acaagaaaag acgataaggg gaatacctctg
attgtgaaca atctgaacgg actgtacgac aaagataatg acaagctgaa aaagctgac
aacaaaagtc ccgagaagct gctgatgtac caccatgatc ctcagacata tcagaaactg
aagctgatta tggagcagta cggcgacgag aagaaccac tgtataagta ctatgaagag
actgggaaact acctgaccaa gtatagcaaa aaggataatg gccccgtgat caagaagatc
aagtactatg ggaacaagct gaatgcccat ctggacatca cagacgatta cctaacagt
cgcaacaagg tgggtcaagct gtcaactgaag ccatacagat tcgatgtcta tctggacaac
ggcgtgtata aatttgtgac tgtcaagaat ctggatgtca tcaaaaagga gaactactat
gaagtgaata gcaagtgcta cgaagaggct aaaaagctga aaaagattag caaccaggca
gagttcatcg cctcctttta caacaacgac ctgattaaga tcaatggcga actgtatagg
gtcatcgggg tgaacaatga tctgctgaac cgcattgaag tgaatatgat tgacatcact
taccgagagt atctggaaaa catgaatgat aagcggcccc ctcgaattat caaaaatatt
gcctctaaga ctcagagtat caaaaagtac tcaaccgaca ttctgggaaa cctgtatgag
gtgaagagca aaaagcacc tcagattatc aaaaagggc

SEQ ID NO: 52

codon optimized nucleic acid sequence encoding *S. aureus* Cas9

atggccccaagaagaagcgggaaggtcgggtatccacggagtcccagcagccaagcgggaactacatcct
gggcctggacatcggcatcaccagcgtgggctacggcatcatcgactacgagacacgggacgtgatcg
atgccggcgtgcggctgttcaaaagaggccaacgtggaaaacaacgagggcagggcggagcaagagagggc
gccagaaggctgaagcggcggagggcggcatagaatccagagagtgaaagaagctgctgttccgactaaa
cctgctgaccgaccacagcagctgagcggcatcaaccctacgagggccagagtgaaagggcctgagcc
agaagctgagcagyaagagttctctgcccctgctgcacctggccaagagaagagggcgtgcacaac
gtgaacgaggtggaagaggacaccggcaacgagctgtccaccaaagagcagatcagccggaaacagcaa
ggccctggaagagaaatacgtggccgaactgcagctggaacggctgaagaaagacggcgaagtgccggg

gcagcatcaacagattcaagaccagcgcactacgtgaaagaagccaaaacagctgctgaagggtgcagaag
gctaccaccagctggaccagagcttcatcgacacctacatcgacctgctggaaacccggcggacctg
ctatgagggacctggcagggcagccccttcggctggaaggacatcaaagaatggtagcagatgctga
tgggccaactgcacctacttccccgaggaactgcggagcgtgaagtacgcctacaacgccgacctgtac
aacgccctgaacgacctgaacaatctcgtgatcaccagggacgagaacgagaagctggaatattacga
gaagttccagatcatcgagaacgtgttcaagcagaagaagaagcccaccctgaagcagatcgccaaag
aaatcctcgtgaaacgaaggatattaagggtacagagtgaccagcaccggcaagcccaggttcacc
aacctgaagggtgtaccacgacatcaaggacattaccgcccggaaagagattattgagaacgccgagct
gctggatcagattgccaaagatcctgaccatctaccagagcagcggagacatccaggaagaactgacca
atctgaactccgagctgaccaggaagagatcgagcagatctctaactctgaagggctataccggcacc
cacaacctgagcctgaaggccatcaacctgatcctggacgagctgtggcacaccaacgacaaccagat
cgctatcttcaacccggctgaagctgggcccagaaggtggacctgtcccagcagaagagatcccca
ccacctgggtggaagacttcatcctgagccccgtcgtgaagagaagcttcatccagagcatcaaagtg
atcaacgccatcatcaagaagtacggcctgcccacgacatcattatcgagctggccccgcgagaagaa
ctccaaaggaagccagaaaaatgatcaacgagatgcagaagcggaaaccggcagaccacgagcggatcg
aggaaatcatccggaccaccggcaagagaaagcgaagtaacctgatcgagaagatcaagctgcacgac
atgcaggaaggcaagtgccctgtacagcctggaagccatcccctctggaagatctgctgaaacaacccctt
caactatgaggtggaccacatcatcccagaagcgtgtccttcgacaacagcttcaacaaacagggtgc
tcgtgaagcaggaagaaaaacagcaagaagggcaaccggacccttccagtagcctgagcagcagcgac
agcaagatcagctacgaaaccttcaagaagcaccatcctgaatctggccaagggcaagggcagaatcag
caagaccaagaaaagatctcgtggaagaacgggacatcaacaggttctcctgagcagaagacttca
tcaacccggaacctgggtgataccagatacggcaccagagggcctgatgaacctgctgaggagctacttc
agagtgaacaaacctggagcgtgaaagtgaagtcctcaatggcggcttccaccagcttctcggcgga
gtggaagtttaagaaaagagcggaaacaggggtaacaagcaccacgcccaggagcgcctgatcattgcca
acgcccgatcttcatcttcaagagtggaagaaactggacaaggccaaaaaagtgatggaaaaaccagatg
ttcgaggaaaaagcaggccgagagcatgcccagatcgaaaaccgagcaggagtaaaaagagatcttcat
caccccccaccagatcaagcacattaaggacttcaaggactacaagtagcagccaccgggtggacaaga
agcctaataagagagctgattaacgacacctgtactccaccgggaaggacgacaagggcaacacctg
atcgtgaaacatctgaacggcctgtacgacaaggacaatgacaagctgaaaaagctgatcaacaagag
ccccgaaaagctgctgatgtaccaccagacccccagacctaccagaaactgaagctgattatggaac
agtagcggcgacgagaagaatccccgtacaagtagtagcagggaaaccgggaaactacctgaccaagtag
tccaaaaaggacaacggccccgtgatcaagaagattaagtagtagcggcaacaaactgaacgcccactct
ggacatcaccgacgactaccccaacagcagaaaacagggtcgtgaagctgtccctgaagcctacagat
tcgacgtgtacctgacaatggcgtgtacaagttcgtgaccgtgaagaatctggatgtgatcaaaaa
gaaaactactacgaagtgaatagcaagtgctatgaggaagctaagaagctgaagaagatcagcaacca
ggccgagtttatcgctccttctacaacaacgatctgatcaagatcaacggcgagctgtatagagtgat
tcggcgtgaaacaacgacctgctgaaccggatcgaaagtgaacatgatcgacatcacctaccgagtag
ctggaaaaacatgaacgacaagagggccccccaggatcattaagacaatcgccctccaagaccagagcat
taagaagtacagcacagacattctgggcaacctgtatgaagtgaatctaagaagcaccctcagatca
tcaaaaagggcaaaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaag

SEQ ID NO: 53**codon optimized nucleic acid sequence encoding *S. aureus* Cas9**

aagcggaaactacatcctggcctggacatcggcatcaccagcgtgggctacggcctcatcgactacga
gacaacgggacgtgatcgatgcccggctgcccgtgttcaaaagaggccaacgtggaaaaacaacgagggca
ggcggagcaagagagggcgcagaaaggctgaagcggcggaggcggcatagaatccagagagtgagaag
ctgctgttcgactacaacctgctgaccgaccacagcagctgagcggcatcaacccctacgagggccag
agtgaagggcctgagccagaagctgagcaggaagagttctctgcccctgctgcacctggccaaga
gaagagggcgtgcacaacgtgaacgaggtggaagaggacaccggcaacgagctgtccaccaagagcag
atcagccggaacagcaaggccctggaagagaaatacgtggccgaactgcagctggaaacggctgaagaa
agacggcgaagtgcggggcagcatcaacagattcaagaccagcagctacgtgaaagaagccaaacagc
tgctgaaggtgcagaaggcctaccaccagctggaccagagcttcatcgacacctacatcgacctgctg
gaaaccggcggacctactatgagggacctggcagggcagcccccttcggctggaaggacatcaaga
atggtagcagatgctgatggccaactgcacctctccccgaggaactgaggagcgtgaagtagcct
acaacgccgacctgtacaacgacctgaacacatctcgtgatcaccagggacgagaacgag

aagctggaatattacgagaagttccagatcatcgagaacgtgttcaagcagaagaagaagcccacct
 gaagcagatcgccaaagaaatcctcgtgaacgaagaggatattaagggctacagagtgaccagcaccg
 gcaagcccagagttaccaacctgaaggtgtaccacgacatcaaggacattaccgcccggaaagagatt
 attgagaacgcccagctgctggatcagattgccaagatcctgaccatctaccagagcagcaggacat
 ccaggaagaactgaccaatctgaactccgagctgaccaggaagagatcgagcagatctctaactga
 agggctataccggcaccacaacctgagcctgaaggccatcaacctgatcctggacgagctgtggcac
 accaacgacaaccagatcgctatcttcaaccggctgaagctgggtgcccaagaagggtggacctgtcca
 gcagaaagagatccccaccacctggaggcagacttcatcctgagccccgtcgtgaagagaagcttca
 tccagagcatcaaagtgatcaacgccatcatcaagaagtaaggcctgcccaacgacatcattatcgag
 ctggcccgcgagaagaactccaaggacgcccagaaaatgatcaacgagatgcagaagcggaaaccggca
 gaccaacgagcggatcgaggaatcatccggaccaccggcaaaagagaacgccaagtaacctgatcgaga
 agatcaagctgcaagacatgcaggaaggcaagtgcctgtacagcctggaagccatccctctggaagat
 ctgctgaacaaaccttcaactatgaggtggaccacatcatccccagaagcgtgtccttcgacaaacag
 cttcaacaaacaggtgctcgtgaagcaggaagaaacagcaagaagggaaccggacccccattccagt
 acctgagcagcagcagcaagatcagctacgaaaccttcaagaagcacatcctgaatctggccaag
 ggcaagggcagaaatcagcaagaccaagaagagatctgctggaagaacgggacatcaacaggttctc
 cgtgcagaaaagacttcatcaaccggaacctgggtgataccagatacgccaccagaggcctgatgaacc
 tgctgcggagctacttcagagtgaacaacctggacgtgaaagtgaagtccatcaatggcggcttccc
 agctttctgcggcggaaagtggaaagtttaagaaagagcggaaacaaggggtacaagcaccacgcccaggga
 cgccctgatcattgccaacgcccatttcatcttcaaagagtggaaagaaactggacaaggccaaaaaag
 tgatggaaaaccagatggttcgaggaaaagcaggccgagagcatgcccgagatcgaaaaccgagcaggag
 taaaaagagatcttcatccccccaccagatcaagcacaattaaggacttcaaggactcaagtaacag
 ccaccgggtggacaagaagcctaataagagagctgattaacgacaccctgtactccaccgggaagcag
 acaagggcaaacacctgatcgtgaacaatctgaacggcctgtacgacaaggacaatgacaagctgaaa
 aagctgatcaacaagacccccgaaaagctgctgatgtaccaccacgacccccagacctaccagaaact
 gaagctgattatggaacagtacggcagcagagaagaatcccctgtacaagtaactacgaggaaaaccggga
 actacctgaccaagtaactccaaaaggacaacggccccctgatcaagaagattaagtattacggcaac
 aaactgaacgcccactctggacatcaccgacgactaccccaacagcagaaaacaaggtcgtgaagctgtc
 cctgaagccctacagattcagcgtgtacctggacaatggcgtgtacaagttcgtgaccgtgaagaatc
 tggatgtgatcaaaaagaaaactactacgaagtgaatagcaagtgctatgaggaagctaagaagctg
 aagaagatcagcaaccaggccgagtttatcgccctccttctacaacaacgatctgatcaagatcaacgg
 cgagctgtatagagtgatcggcgtgacaacgacctgctgaaccggatcgaagtgaacatgatcgaca
 tcacctaccgagtagctggaaaaacatgaacgacaagaggccccccaggatcattaaga caatcgcc
 tccaagaccagagcattaagaagtaagcagacagacattctgggcaacctgtatgaagtgaatctaa
 gaagcaccctcagatcatcaaaaaggc

SEQ ID NO: 54

Streptococcus pyogenes Cas9 (with D10A, H849A)

MDKKYSIGLAI GTNSVGVAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
 RRRYTRRKNRICYLQEIFSNEMARVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIY
 HLRKKLVSDTDKADLRILIYLA LAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINAS
 GVDAKAILSARLSKSRLENLIAQLPGEKKNLFGNLIASLGLTPNFKSNFDLAEADAKLQLSKDITYD
 DDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLKALVR
 QQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFKPILEKMDGTEELLVKLNRDILLRKQRTFDNG
 SIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRI PYYVGPLARGNSRFAMTRKSEETITPW
 NFEVVVDKGASASQSFIERMTNFDKNLPNEKVLPHKSLLEYEYFTVYNELTKVKYVTEGMRKPAFLSSEQ
 KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEEN
 EDILEDIVLTLTLFEDREMI EERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLLINGIRDKQSGKTI L
 DFLKSDGFANRFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDELV
 KVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIELGSQLKEHPVENTQLQNEKLYLYYL
 QNGRDMYVDQELDINRLSDYDVAIVPQSFLKDDSIDNKVLRSDKNRSGSDNVPSEEVVKKMKNYWR
 QLLNAKLITQRKFDNLTKAERGGLELKDAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI RE
 VKVITLKS KLVSDFRKDFQFYKVIENNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVDYVRK
 MIAKSEQEIGKATARYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL S
 MPQVNI VVKTEVQTGGFSSKESILPKRNSDKLIARKKDWDPKKGFFDSPTVAYSVLVVAKVEKGSKK

LKSVKELLGITIMERSSSFENKPIDFLEAKGYKEVKKDLIIKLPKYSLEFELENGRKRMLASAGELQKGN
 ELALPSKYVNFYLYLASHYEKLGKSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVI LADANLQKVL
 AYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRI
 DLSQLGGD

SEQ ID NO: 55

Vector (pDO242) encoding codon optimized nucleic acid sequence encoding *S. aureus*

Cas9

ctaaattgtaagcgttaatatTTTTgttaaaattcgcgttaaattTTTTgttaaatcagctcattTTTTta
 accaataggccgaaatcggcaaaatcccttataaatcaaaagaatagaccgagataggggttgagtgtt
 gttccagtttggaacaagagtcactattaaagaacgtggactccaacgtcaaagggcgaaaaaccgt
 ctatcagggcgatggccactacgtgaaccatcacctaatcaagttTTTTggggctcgaggtgcccgt
 aagcactaaatcggaaacctaaagggagccccgatttagagcttgacggggaagccggcgaaacgtg
 gcgagaaaggaaggaagaaagcgaagggagcggcgctagggcgctggcaagtgtagcgggtcaacgt
 gcgcgtaaccaccacaccccgccgcttaatgcgcgcgctacagggcgctccattcgcattcaggc
 tgccgcaactggtgggaagggcgatcgggtgcccctcttcgctattaacgacgctggcgaaaggggga
 tgtgctgcaagggcattaagtgggtaacgccagggTTTTccagtcacgacggttgtaaacgacggc
 cagtgagcgcgcgtaatacgaactcactatagggcgaattgggtacCttaaattctagtaactatgcaTg
 cgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagccata
 tatggagttccgcgttacataaactacggtaaatggcccgcctggctgaccgcccacga cccccgc
 cattgacgtcaataatgacgtatggtcccatagtaacgccaatagggactttccattgacgtcaatgg
 gtggagtatttacggtaaaactgcccacttggcagtaacatcaagtgtatcatatgccaaagtacgcccc
 tattgacgtcaatgacggtaaatggcccgcctggcattatgccagtaacatgaccttatgggactttc
 ctacttggcagtaacatcactacgtattatgctacgctattaccatgggtgatgcggttttggcagtaac
 aatggcgctggatagcggtttgactcacgggatttccaagtctccacccattgacgtcaatgggag
 tttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaa
 tgggcggtaggcgtgtacggtgggaggtctataaagcagagctctctggctaaactaccggtgccacc
 ATGAAAAGGAACTACATTTCTGGGGCTGGACATCGGGATTACAAGCGTGGGGTATGGGATTATTGACTA
 TGAAACAAGGGACGTGATCGACGCAGGCGTCAGACTGTTCAAGGAGGCCAACGTGGAAAACAATGAGG
 GACGGAGAAGCAAGAGGGGAGCCAGGCGCTGAAACGACGGAGAAGGCACAGAATCCAGAGGGTGAAG
 AAATGCTGTTTCGATTACAACCTGCTGACCGACCATTCTGAGCTGAGTGGAAATTAATCCTTATGAAGC
 CAGGGTAAAAGCCTGAGTCAGAAGCTGTCAGAGGAAGAGTTTTCCGCAGCTCTGCTGCACCTGGCTA
 AGCGCCGAGGAGTGCATAACGTCAATGAGGTGGAAGAGGACACCGGCAACGAGCTGTCTACAAAGGAA
 CAGATCTCACGCAATAGCAAAGCTCTGGAAGAGAAGTATGTGCGCAGAGCTGCAGCTGGAACGGCTGAA
 GAAAGATGGCGAGGTGAGAGGGTCAATTAATAGGTTCAAGACAAGCGACTACGTCAAAGAAGCCAAGC
 AGCTGCTGAAAGTGCAGAAGGCTTACCACCAGCTGGATCAGAGCTTCATCGATACTTATATCGACCTG
 CTGGAGACTCGGAGAACCTACTATGAGGGACCAGGAGAAGGGAGCCCCCTTCGGATGGAAAGACATCAA
 GGAATGGTACGAGATGCTGATGGGACATTCACCTATTTTCCAGAAGAGCTGAGAAGCGTCAAGTACG
 CTTATAACGCAGATCTGTACAACGCCCTGAATGACCTGAACAACCTGGTCAACAGGGGATGAAAAC
 GAGAACTGGAATACTATGAGAAGTCCAGATCATCGAAAACGTGTTAAGCAGAAGAAAAAGCCTAC
 ACTGAAAACAGATTGCTAAGGAGATCCTGGTCAACGAAGAGGACATCAAGGGCTACCGGGTACAAGCA
 ATCATTGAGAACGCCGAACCTGCTGGATCAGATTGCTAAGATCCTGACTATCTACCAGAGCTCCGAGGA
 CATCCAGGAAGAGCTGACTAACCTGAACAGCGAGCTGACCCAGGAAGAGATCGAACAGATTAGTAATC
 TGAAGGGGTACACCGGAACACACAACCTGTCCCTGAAAGCTATCAATCTGATTCTGGATGAGCTGTGG
 CATACAAACGACAATCAGATTGCAATCTTTAACCGGCTGAAGCTGGTCCCAAAAAGGTGGACCTGAG
 TCAGCAGAAAAGAGATCCCAACCACACTGGTGGACGATTTTCATTCTGTACCCCGTGGTCAAGCGGAGCT
 TCATCCAGAGCATCAAAGTGATCAACGCCATCATCAAGAAGTACGGCCTGCCAATGATATCATTATC
 GAGCTGGCTAGGGAGAAGAACAGCAAGGACGCACAGAAGATGATCAATGAGATGCAGAAAACGAAACCG
 GCAGACCAATGAACGCATTGAAGAGATTATCCGAACTACCGGGAAGAGAACGCAAAGTACCTGATTG
 AAAAAATCAAGCTGCACGATATGCAGGAGGGAAAGTGTCTGTATTCTCTGGAGGCCATCCCCCTGGAG
 GACCTGCTGAACAATCCATTCAACTACGAGGTCGATCATATTATCCCAGAAGCGTGTCTCTCGACAA
 TTCCTTTAACAACAAGGTGCTGGTCAAGCAGGAAGAGAAGCTCTAAAAGGGCAATAGGACTCCTTTCC
 AGTACCTGTCTAGTTCCAGATTCCAAGATCTCTTACGAAACCTTTAAAAGCACATTCTGAATCTGGCC
 AAAGGAAAAGGGCCGCATCAGCAAGACCAAAAAGGAGTACCTGCTGGAAGAGCGGGACATCAACAGATT

CTCCGTCAGAAAGGATTTTATTAAACCAGGAATCTGGTGGACACAAGATACGCTACTCGCGGCCCTGATGA
ATCTGCTGCGATCCTATTTCCGGGTGAACAATCTGGATGTGAAAGTCAAGTCCATCAACGGCGGGTTTC
ACATCTTTTCTGAGGCGCAAATGGAAGTTTAAAAAGGAGCGCAACAAAGGGTACAAGCACCATGCCGA
AGATGCTCTGATTATCGCAAATGCCGACTTCATCTTTAAGGAGTGGAAAAAGCTGGACAAAGCCAAGA
AAGTGATGGAGAACCAGATGTTCGAAGAGAAGCAGGCCGAATCTATGCCCGAAATCGAGACAGAACAG
GAGTACAAGGAGATTTTTCATCCTCCTCACCAGATCAAGCATATCAAGGATTTCAAGGACTACAAGTA
CTCTCACCGGGTGGATAAAAAGCCCAACAGAGAGCTGATCAATGACACCCTGTATAGTACAAGAAAAG
ACGATAAGGGGAATACCCTGATGTGAACAATCTGAACGGACTGTACGACAAAAGATAATGACAAGCTG
AAAAAGCTGATCAACAAAAGTCCCAGAAAGCTGCTGATGTACCACCATGATCCTCAGACATATCAGAA
ACTGAAGCTGATTATGGAGCAGTACGGCGACGAGAAGAACCCTACTGTATAAGTACTATGAAGAGACTG
GGAATACCTGACCAAGTATAGCAAAAAGGATAATGGCCCCGTGATCAAGAAGATCAAGTACTATGGG
AACAGCTGAATGCCCATCTGGACATCACAGACGATTACCCTAACAGTCGCAACAAGGTGGTCAAGCT
GTCACTGAAGCCATACAGATTCGATGTCTATCTGGACAACGGCGTGTATAAATTTGTGACTGTCAAGA
ATCTGGATGTCATCAAAAAGGAGAACTACTATGAAGTGAATAGCAAGTGTACGAAGAGGGCTAAAAAG
CTGAAAAAGATTAGCAACCAGGCAGAGTTCATCGCCTCCTTTTACAACAACGACCTGATTAAGATCAA
TGGCGAACTGTATAGGGTCATCGGGGTGAACAATGATCTGCTGAACCGCATTGAAGTGAATATGATTG
ACATCACTTACCAGAGATATCTGAAAAACATGAATGATAAGCGCCCCCTCGAATTATCAAAAACATTT
GCCTTAAGACTCAGAGTATCAAAAAGTACTCAACCGACATTTCTGGGAAACCTGTATGAGGTGAAGAG
CAAAAAGCACCTTCAGATTATCAAAAAGGGCagcggaggcaagcgtcctgctgctactaagaaagctg
gtcaagctaagaaaaagaaaggatcctaccatacagatggtccagattacgcttaagaattcctagag
ctcgtgatcagcctcgactgtgctctctagttgcccagccatctgttgtttgcccctccccgctgct
tctgagtaggtgtcattctattctgggggtgggggtggggcaggacagcaagggggaggatggggaag
agaatagcaggcatgctggggaggtagcggcggcCCgcggtggagctccagcttttctcccttagt
gagggtaattgcggccttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctc
acaattccacacacaacatacagagccggaagcataaagtgtaaagcctgggggtgctaatgagtgagcta
actcacattaattgogtggctcactgcccgtttccagtcgggaaacctgtcgtgcccagctgcatt
aatgaatcggccaacgcgcggggagagggcgttttgcgtattgggcgctcttccgcttctcgtcact
gactcgtcgctcggctcgttcggctgcccgcagcggatcagctcactcaaaggcggtaatacgggt
atccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaacc
gtaaaaaggccgcttgcgtggcgttttccataggctccgccccctgacgagcatcaaaaaatcga
cgtcagagtcagagtggtggcgaaccggacaggactataaagataccaggcgtttccccctggaagctc
cctcgtgcgctctcctgttcgaccctgcccgttacccgatacctgtccgctttctcccttcgggaa
gctggcgctttctcatagctcacgctgtaggtatctcagttcgggtgtaggtcgttcgctccaagctg
ggctgtgtgcaagcaacccccgttcagcccagccgctgccccttatccggtaactatcgtcttgagtc
caacccggtaagacacagacttatcgccactggcagcagccactggtaacaggattagcagagcggaggt
atgtaggcgggtgctacagagttcctgaagtgggtggcctaactacggctacactagaaggacagtattt
ggatctcgcgtctcgtgaagccagttacctcggaaaaagagttggtagctcttgatccggcaaaaca
aaccaccgctggtagcgggtggttttttggtttgaagcagcagattacgcgcagaaaaaaaggatctc
aagaagatcctttgatcttttctacggggctgacgctcagtggaacgaaaactcagtttaagggatt
ttggtcatgagattatcaaaaaggatcttcacctagatccttttaaatataaaaatgaagttttaaatc
aatctaaagtataatagtaaaacttggctcagcttaccatgcttaatacagtgagggcaccctatct
cagcgtctgtctatctcgttcacatagttgctgactccccgtcgtgtagataaactacgatacgg
gagggcttaaccatctggccccagtgctgcaatgataccgcgagacccacgctcacgggctccagattt
atcagcaataaaaccagccagccggaaggccgagcgcagaagtggctcctgcaactttatccgcctcca
tccagctctattaattggttgcgggaagctagagtaagtagttcggcagttaatagtttggcacaagtt
gttgcattgtctacaggcatcgtgggtgtcacgctcgtcgtttggatggcttcattcagctccgggtc
ccaacgatcaaggcgagttacatgatccccatgttgtgcaaaaaagcggttagctccttcggctcctc
cgatcgttgtcagaagtaagttggccgcagtggtatcactcatggttatggcagcactgcataattct
cttactgtcatgccatccgtaagatgcttttctgtgactgggtgagtaactcaaccaagtcattctgaga
atagtgatgocggcagccagttgctcttgcggcgctcaatacgggataataccgcgccaacatagca
gaactttaaaagtgtcatcattggaaaagcttctcggggcgaaaaactctcaaggatcttacogctg
ttgagatccagttcgatgtaaccactcgtgcacccactgatcttcagcatcttttactttaccag
cgtttctgggtgagcaaaaaaggaaggcaaaaatgcccgaaaaaagggaataagggcgacacggaaat
gttgaaatactcactcttctcttttcaatattattgaagcatttatcagggttattgtctcatgagc

ggatacatatttgatgtatttagaaaaataaacaataggggttccgcgacatttccccgaaaagt
gccac

SEQ ID NO: 56

tttn (N can be any nucleotide residue, e.g., any of A, G, C, or T)

SEQ ID NO: 57

VP64-dCas9-VP64 protein

RADALDDFDLMLGSDALDDFDLMLGSDALDDFDLDMVNPKKKRKVGRGMDKKY
SIGLAI GTNSVGVAVITDEYKVPSSKFKVLGNTRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT
RRKNRI CYLQEIFSNEMAKVDDSFHRLSESLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKK
LVDSTDKADLRLLYLALAHMIKFRGHFLIEGDLNPDNSVDKLFIQLVQTYNQLFEENPINASGVDAK
AII SARLSKSRRENLI AQLPGEKKNLFGNLI ALSGLTPNFKSNFDLAEDAKLQLSKD TYDDDLN
LLAQIGDQYADLF LAAKNLSDAI LLSDI LRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKNLREDLLRKQRTFDNGSI PHQ
IHLGELHAILRRQEDFY PFLKDNREKIEKILTFRI PYYVGPLARGNSRFAMWTRKSEETITPWNFEV
VDKGASAQSFIEPMTNFDKNLPNEKVLPKHSLLEYEFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIV
DLLEKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILE
DIVLTLTLFEDREMIEERLKYAHLFDKVMKQLKRRRYTGWGRLSRKLLINGIRDKQSGKTI LDFLKS
DGFANRNFMLIHDDSLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGR
HKPENIVIMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRD
MYVDQELDINRLSDYDVAIVPQSFLKDDSDKNVLTNRSDKNRGSNDNVPSEEVVKKMKNYWRQLLNA
KLITQRKFDNLTKAERGGLELDKAGFIKQVLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVIT
LKSKLVSDFRKDFQFYKVIENNYHHAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVVYDVRKMI AKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVN
IVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLSVK
ELLGITIMERSSEFEKNPIDFLEAKGYKEVKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALP
SKYVNFYLYLASHYEKLGKSPEDNEQQLFVEQHKHYLDEII EQISEFSKRVI LADANL DKVLSAYNKH
RDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQL
GGDSRADPKKKRQVASRADALDDFDLMLGSDALDDFDLMLGSDALDDFDLMLGSDALDDFDLML
I

SEQ ID NO: 58

VP64-dCas9-VP64 DNA

cgggctgacgcattggacgattttgatctggatagctgggaagtgacgccctcgatgattttgacct
tgacatgcttggttcggatgcccttgatgactttgacctcgacatgctcggcagtgacgcccttgatg
atctcgacctggacatggttaaccccaagaagaagaggaaggtgggccggaatggacaagaagtac
tccattgggctcgccatcgccacaaacagcgtcggctgggccgctcattacggacgagtagaagggtgcc
gagcaaaaaattcaaagttctgggcaataccgatcgccacagcataaagaagaacctcattggcgccc
tctgttcgactccggggaaaccgcccgaagccacgcggtcaaaagaacagcacggcgagatatacc
cgcagaaagaatcggatctgctacctgcaggagatcttttagtaatgagatggctaagggtggatgactc
tttcttccataggctggaggagtcttttgggtggaggaggataaaaagcacgagcgccaccaaatct
ttggcaatatcgtggacgaggtggcgtaaccatgaaaagtacccaaccatataatcatctgaggaagaag
ctttagacagtagataaggctgacttgcggttgatctatctcgcgctggcgcatatgatcaaatt
tcggggacacttctcatcgagggggacctgaaccagacaacagcagatgtcgacaaactctttatcc
aactggttcagacttacaatcagcttttcgaagagaaccgatcaacgcacccggagtgcagcggcaaa
gcaatcctgagcgtaggctgtccaaatcccggcgctcgaaaacctcatcgacagctccctgggga
gaagaagaacggcctgtttggtaatcttatcgccctgtcactcgggctgacccccaaactttaaatcta
acttcgacctggccgaagatgccaagcttcaactgagcaaaagacacctacgatgatgatctcgacaat
ctgctggccagatcgccgaccagtagcagaccttttttggcggaagaacctgtcagacgccat
ctgctgagtgatattctgagtagaacacggagatcaccaaagctccgctgagcgtatgatgatca
agcgtatgatgagcaccaccaagacttgactttgctgaaggcccttctcagacagcaactgctgag
aagtacaaggaaaattttctcgatcagctctaaaaatggctacgcccggatacattgacggcgggagcaag
ccaggaggaattttacaaatttattaagccatcttggaaaaaatggacggcaccgaggagctgctgg
taaagcttaacagagaagatctgctgcaaacagcgcactttcgacaatggaagcatccccaccag

attcaacctggggcgaactgcacgctatcctcaggcggcaagaggatttctaccctttttgaaagataa
 cagggaaaagattgagaaaaatcctcacatttcggataccctactatgtaggccccctcgccccgggaa
 attccagattcgcgtggatgactcgcgaaatcagaagagaccatcactccctggaacttcgaggaagtc
 gtggataaaggggctctgcccagtccttcacgaaaggatgactaaactttgataaaaaatctgcctaa
 cgaaaagggtgcttctaaacactctctgctgtacgagtaacttcacagtttataacgagctcaccaagg
 tcaaaatcgtcacagaaggatgagaaagccagcattcctgtctggagagcagaagaaagctatcgtg
 gacctcctcttcaagacgaaacgggaaagtaccgtgaaacagctcaaagaagactatctcaaaaagat
 tgaatgtttcgactctgttgaaatcagcggagtgaggatcgcttcaacgcacccctgggaaagctatc
 acgatctcctgaaaaatcattaaagacaaggacttctggacaatgaggagaacgaggacattcttgag
 gacattgtcctcacccttaagttgtttgaagatagggagatgattgaagaacgcttgaaaacttacgc
 tcatctcttcgacgacaaaagtcataaagacagctcaagaggcgcgatatacaggatggggggcgctgt
 caagaaaactgatcaatgggatccgagacaagcagagtggaagacaatcctggattttcttaagtcc
 gatggattttgccaacgggaaacttcagcagttgatccatgatgactctctcacccttaaggaggacat
 ccagaaaagcacaagtctctggccagggggacagctcttcacgagcacatcgctaactctgacaggtagcc
 cagctatcaaaaaagggaatactgcagaccgttaaggctcgtggatgaaactcgtcaaaagtaatgggaaagg
 cataagcccggagaatctcgttatcgagatggcccggagagaaccaaactaccagaagggaagagaaga
 cagtagggaaaaggatgaagaggattgaagagggtataaaaagaactgggggtcccaaatccttaaggaa
 acccagttgaaaacacccagcttcagaatgagaagctctacctgtactacctgcagaaacggcagggac
 atgtacgtggatcaggaactygacatcaatcggctctccgactacgacgctggatgccaatcgtgcccc
 gtctttctcaaaagatgattctattgataataaagtgttgacaagatccgataaaaaatagagggaaga
 gtgataacgtcccctcagaagaagtgtcagaagaaatgaaaaattatggcggcagctgctgacacgccc
 aactgatcacacaaacggaaagttcgataactcagactaaggctgaacgaggtggcctgtctgagttgga
 taaagccggcttcatcaaaaaggcagcttctgttgagacacgacagatcaccagcagctggcccaattc
 tcgattcacgcatgaacaccaagtcagatgaaaaatgacaaactgattcgagaggtgaaagttact
 ctgaagtctaaagctggtctcagatttcagaaaggactttcagttttataaggtgagagagatcaacaa
 ttaccaccatgcgcatgatgctacctgaatgcagtggttaggcactgcacttatcaaaaaatataccca
 agcttgaatctgaatttgtttacgggagactataaagtgtacgatgttaggaaaatgatcgcaaaagtct
 gagcaggaaaataggcaaggccaccgctaagtacttcttttacagcaatattatgaatttttcaagac
 cgagattacactggccaatggagagattcggaaagcgaccacttatcgaaaacaaacgggagaaacaggag
 aaatcgtgtgggacaagggtagggattctcgacagctccggaaggctcctgtccatgcccaggtgaa
 atcgttataaaagaccgaagtacagaccggaggcttctccaaggaaagtatcctcccgaaaaaggaaacag
 cgacaagctgatcgacgcaaaaaagattgggaccccaagaaatacggcggatttcgattctcctacag
 tcgcttacagtgactggtgtggccaaagtgggagaaagggaagtctaaaaaaccaaaagcgtcaag
 gaaactgctgggcatcaaatcatggagcgatcaagcttcgaaaaaaaaccccatcgactttctcgaggc
 gaaaggatataaaagggtcaaaaaagacctcatcattaagcttcccaagtaactctctctttgagcttg
 aaaaacggccggaaaacgaatgctcgtcagtcggggcgagctgcagaaaggtaacgagctggcactgccc
 tctaaatagcttaatttcttctgtatctggccagccactatgaaaagctcaaagggtctcccgaagataa
 tgagcagaagcagctgttcgtggaacaacacaaacactaccttgatgagatcatcgagcaaaataagcg
 aattctccaaaagagtgatcctcgcgacgctaacctcgataagggtgctttctgcttacaataagcac
 agggataagccatcaggggagcaggcagaaaacattatccacttgtttactctgaccaacttggggcgc
 gctgcagccttcaagtaacttcgacaccaccatagacagaaagcgggtacacctctacaaaaggaggtcc
 tggacgcccactgattcatcagtcattacggggctctatgaaacaagaatcgacctctctcagctc
 ggtggagacagcagggctgaccccaagaagaagaggagggtggctagccgcgacgctggaacga
 tttcgatctcgacatgctgggttctgatgccctcgatgactttgacctggatagttgggaaagcagc
 cattggatgactttgatctggacatgctcggctccgatgctctggacgatttcgatctcgatagttta
 atc

SEQ ID NO: 59

Human p300 (with L553M mutation) protein

MAENVVEPFPSPAKRPKLSSPALSASASDGTDFGSLFDLEHDLPELINSTELGLTNGGDINQLQTSI
 GMVQDAASKHKQLSELLRSGSSPNLNMVGVPQVMASQAQQSSPGLGLINSMVKS PMTQAGLTS PNM
 GMGTS GPNQGP TQSTGMMNS PVNQ PAMGMNTGMNAGMNPGLAAGNGQGIMPNOVMNGSI GAGRGRQN
 MQYFNP GMGSAGNLLTEPLQQGSPQMGQTGLRGPQFLKMGMMNPNPYGSPYTQNPQQIGASGLGL
 QIQTKTVLSNNLSPFAMDKAVPGGGMPNMGGQPAPQVQPPGLVTPVAQGMGSGAHTADPEKRKLIQQ
 QLVLHHAHKCQRREQANGEVRQCENLPHCRTMKNVLNHMTHCQSGKSCQVAHCASSRQII SHWKNCTR

HDCPVCFLPLKNAGDKRNQQPIILTGAPVGLGNPSSLVGVGQQSAPNLSSTVVSQIDPSSIERAYAALGLPYQ
 VNQMPTQPCVQAKNQQNQQPQGS PQGMRPMSNMSASFMGVNNGGVGVQTPSLLSDSMLHSAINSONPMM
 SENASVPSMGPMPTAAQPSSTTGI RKQWHEDITQDLRNHLVHKLQVQAI FPTPDPAALKDRRMMENLVAYA
 RKVEGDMYESANNRAEYHLLAEKI YKI QKELEEKRRRLQKQNMPLPNAAGMVPVSMNPGPNMGQPQP
 GMTSNGPLEPDPSPMIRGSVFNQMMFRITPQSGLNQFGQMSMAQPPIVPRQTPPLQHHGQLAQPGALNPP
 MGYGFRMQQPSNQGFQFLPQTQFQPSQGMNVTNIP LAPSSGQAPVSQAQMSSSSSCPVNSPIMPPGSGQSH
 IHCPQLPQPALHQNSPSPVPSRTPTPHHTPPSIGAQPPATTIPAFVPTPPAMP PGPQSQUALHPPPRQ
 TPTPPTTQLPQQVQPSLPAAPSADQPOQQPRSQQSTAASVPTPTAPLLPPQPATPLSQPAVSI EGQVS
 NPPSTSSTEVNSQAI AEKQPSQEVKMEAKMEVDQPEPADTQPEDISESKVEDCKMESTETEERSTELK
 TEIKEEEDQPSSTSATQSSPAPGQSKKKI FKPEELRQALMPTLEALYRQDPESLFFRQPVDPQLLGI PD
 YFDIVKSPMDLSTIKRKLDTGQYQEPWQYVDDIWLWMFNNAWLYNRKTSRVYKYCSKLSSEVFEQEIDPV
 MQSLGYCCGRKLEFSPQTLCCYQKQLCTIPRATYYSYQNRHYHFCEKCFNEIQGESVSLGDDPSQPQT
 TINKEQFSKRKNDTLDPELFVECTECGRKMHQICVLHHEI IWPAGFVCDGCLKKSARTRKENKFSKR
 LPSTRLGTFLENRVNDFLRRQNHPESGEVTVRVVHASDKTVEVKPGMKARFVDSGEMAESFPYRTKAL
 FAFEEIDGVDLCFFGMHVQEYGSDCPPPNQRRVYISYLDVSVHFFRPKCLRTAVYHEILIGYLEYVKKL
 GYTTGHIWACPPSEGDDYIFHCHFPDQKIPKPKRLQEWYKMLDKAVSERIVHDYKDI FKQATEDRLT
 SAKELPYFEGDFWPNVLEESI KELEQEEEEERKREENTSNESTDVTKGDSKNAKKNNKTSKNKSSLS
 RGNKKKPGMPNVSNLDSQKLYATMEKHKEVFFVIRLIAGPAANSLPPIVDPDPLIPCDLMDGRDAFLT
 LARDKHLEFSSLRRAQWSTMCMLEVELHTQSDRFVYTCNECKHHVETRWHTVCEYDLCITCYNTKN
 HDHKMEKLGGLDDESNNQQAATQSPGDSRRLSIQRCIQSLVHACQCRNANCSLPSQCKMKRVVQHT
 KGCKRKTNGGCPICKQLIALCCYHAKHCQENKCPVPFCLNIKQKLRQQQLQHRLOQAQMLRRRMAQ
 RTGVRVGGQGLPSPATPPTPTGQQPTTPTQPTSQPPTPPNSMPPYLPRTQAAGPVSQGKAAGQ
 VTFPTFPQTAQPLPGPPFAAVEMAMQIQRAAETQRQMAHVQIQFRPIQHQMPPMTPMAPMGMPPEPM
 TRGFSGHLEFGMGPTGMQQQPPWSQGGLPQPQQQLQSGMFRPAMMSVAQHGGQPLNMAPQPGLGQVGISP
 LKPGTVSQQALQNLRLRSPSSPLQQQQVLSILHANPQLLAAFIKQRAAKYANSNPQPIFGQPMPQ
 GQFGLQPPTMPGQQGVHNSPAMQNMNFMQAGVQRAGLPQQQPQQQLQPPMGGMSPOAQOMNMNHTMP
 SQFRDILRRQMMQQQQQQGAGPGIGPGMANHNQFQQPQGVGYPPQQQQRMQHMQMOMQGNMGQIGQ
 LPQALGAEAGASLQAYQQRLLQQQMGSPVQPNPMSPOQHMLPNQAQSPHLOGQOIPNSLSNQVRSPOP
 VPSRPPQSPPHSSPSPRMQPQPSPHHVSQOTS SPHPGLVAAQANPMEQGHFASPDQNSMLSQLASNP
 GMANLHGASATDLGLSTDNSDLNSNLSQSTLDIH

SEQ ID NO: 60

Human p300 Core Effector protein (aa 1048-1664 of SEQ ID NO: 59)

IFKPEELRQALMPTLEALYRQDPESLFFRQPVDPQLLGI PDYFDIVKSPMDLSTIKRKLDTGQYQEPW
 QYVDDIWLWMFNNAWLYNRKTSRVYKYCSKLSSEVFEQEIDPVMQSLGYCCGRKLEFSPQTLCCYQKQLC
 TIPRATYYSYQNRHYHFCEKCFNEIQGESVSLGDDPSQPQTINKEQFSKRKNDTLDPELFVECTECG
 RKMHQICVLHHEI IWPAGFVCDGCLKKSARTRKENKFSKRLPSTRLGTFLENRVNDFLRRQNHPESG
 EVTVRVHASDKTVEVKPGMKARFVDSGEMAESFPYRTKALFAFEEIDGVDLCFFGMHVQEYGSDCPP
 PNQRRVYISYLDVSVHFFRPKCLRTAVYHEILIGYLEYVKKLGYTTGHIWACPPSEGDDYIFHCHFPDQ
 KIPKPKRLQEWYKMLDKAVSERIVHDYKDI FKQATEDRLTSAKELPYFEGDFWPNVLEESI KELEQE
 EEEKREENTSNESTDVTKGDSKNAKKNNKTSKNKSSLSRGNKKKPGMPNVSNLDSQKLYATMEKH
 KEVFFVIRLIAGPAANSLPPIVDPDPLIPCDLMDGRDAFLT LARDKHLEFSSLRRAQWSTMCMLEVELH
 TQSQD

SEQ ID NO: 85

Polynucleotide sequence of a gRNA scaffold

gttttagagctagaaatagcaagttaaaataaggctagtcggttatcaacttgaaaaagtgg
 caccgagtcggtgcttttttt

CLAIMS

1. A guide RNA (gRNA) molecule targeting Pax7, the gRNA comprising a polynucleotide sequence corresponding to at least one of SEQ ID NOs: 1-8 or 69-76, or a variant thereof.
2. The gRNA of claim 1, wherein the gRNA comprises a crRNA, a tracrRNA, or a combination thereof.
3. A DNA targeting system for increasing expression of Pax7, the DNA targeting system comprising at least one gRNA that binds and targets a Pax7 gene, a regulatory region of a Pax7 gene, a promoter region of a Pax7 gene, or a portion thereof.
4. The DNA targeting system of claim 3, wherein the at least one gRNA comprises a polynucleotide sequence corresponding to at least one of SEQ ID NOs: 1-8 or 69-76, or a variant thereof.
5. The DNA targeting system of claim 3 or 4, wherein the gRNA comprises a crRNA, a tracrRNA, or a combination thereof.
6. The DNA targeting system of any one of claims 3-5, further comprising a Clustered Regularly Interspaced Short Palindromic Repeats associated (Cas) protein or a fusion protein,
wherein the fusion protein comprises two heterologous polypeptide domains, wherein the first polypeptide domain comprises a Cas protein, a zinc finger protein, or a TALE protein, and the second polypeptide domain has transcription activation activity.
7. The DNA targeting system of claim 6, wherein the Cas protein comprises a *Streptococcus pyogenes* Cas9 molecule, or a variant thereof.
8. The DNA targeting system of claim 6, wherein the fusion protein comprises VP64-dCas9-VP64.
9. The DNA targeting system of claim 6, wherein the Cas protein comprises a Cas9 that recognizes a Protospacer Adjacent Motif (PAM) of NGG (SEQ ID NO: 31), NGA (SEQ ID NO: 32), NGAN (SEQ ID NO: 33), or NGNG (SEQ ID NO: 34).
10. An isolated polynucleotide sequence comprising the gRNA molecule of claim 1 or 2.

11. An isolated polynucleotide sequence encoding the DNA targeting system of any one of claims 3-9.
12. A vector comprising the isolated polynucleotide sequence of claim 10 or 11.
13. A vector encoding the gRNA molecule of claim 1 or 2 and a Clustered Regularly Interspaced Short Palindromic Repeats associated (Cas) protein.
14. A cell comprising the gRNA of claim 1 or 2, the DNA targeting system of any one of claims 3-9, the isolated polynucleotide sequence of claim 10 or 11, or the vector of claim 12 or 13, or a combination thereof.
15. A pharmaceutical composition comprising the gRNA of claim 1 or 2, the DNA targeting system of any one of claims 3-9, the isolated polynucleotide sequence of claim 10 or 11, the vector of claim 12 or 13, or the cell of claim 14, or a combination thereof.
16. A method of activating endogenous myogenic transcription factor Pax7 in a cell, the method comprising administering to the cell the gRNA of claim 1 or 2, the DNA targeting system of any one of claims 3-9, the isolated polynucleotide sequence of claim 10 or 11, or the vector of claim 12 or 13.
17. A method of differentiating a stem cell into a skeletal muscle progenitor cell, the method comprising administering to the stem cell the gRNA of claim 1 or 2, the DNA targeting system of any one of claims 3-9, the isolated polynucleotide sequence of claim 10 or 11, or the vector of claim 12 or 13.
18. The method of claim 17, wherein endogenous expression of Pax7 mRNA is increased in the skeletal muscle progenitor cell.
19. The method of any one of claims 17-18, wherein the expression of Myf5, MyoD, MyoG, or a combination thereof, is increased in the skeletal muscle progenitor cell.
20. The method of any one of claims 17-19, wherein the stem cell is induced into myogenic differentiation.

21. The method of any one of claims 17-20, wherein the skeletal muscle progenitor cell maintains Pax7 expression after at least about 6 passages.

22. A method of treating a subject in need thereof, the method comprising administering to the subject the cell of claim 14.

23. The method of claim 22, wherein the level of dystrophin+ fibers in the subject is increased.

24. The method of claim 22 or 23, wherein muscle regeneration in the subject is increased.

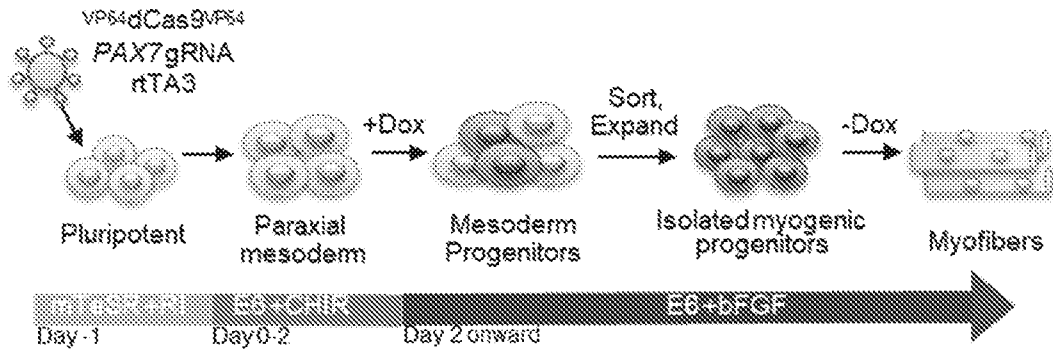


FIG. 1A

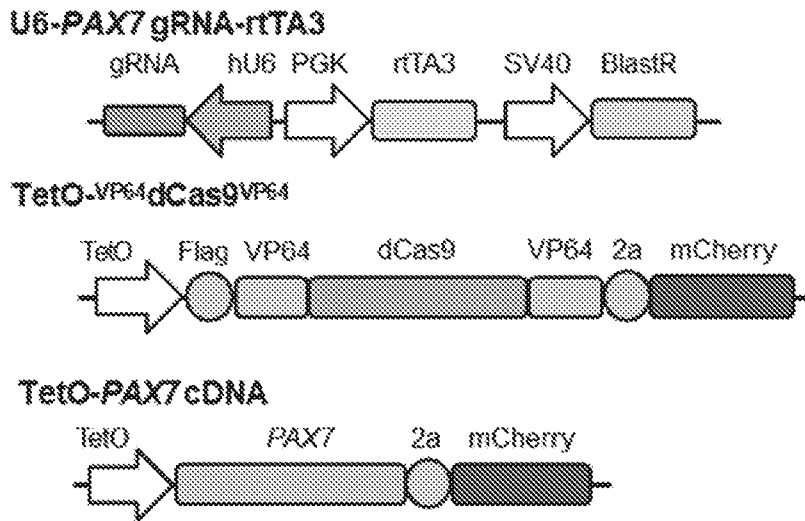


FIG. 1B

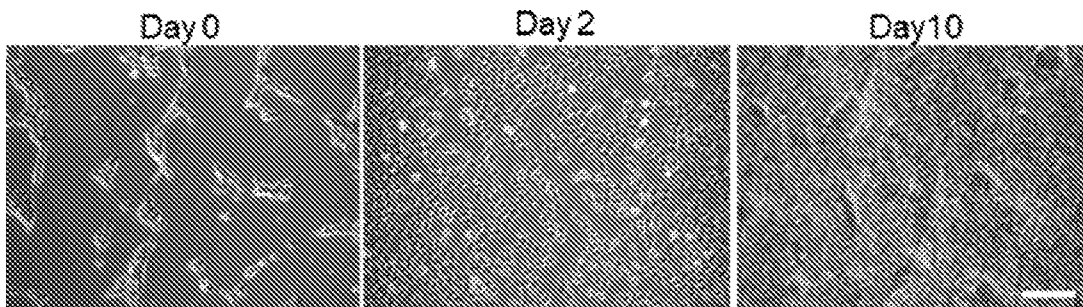


FIG. 1C

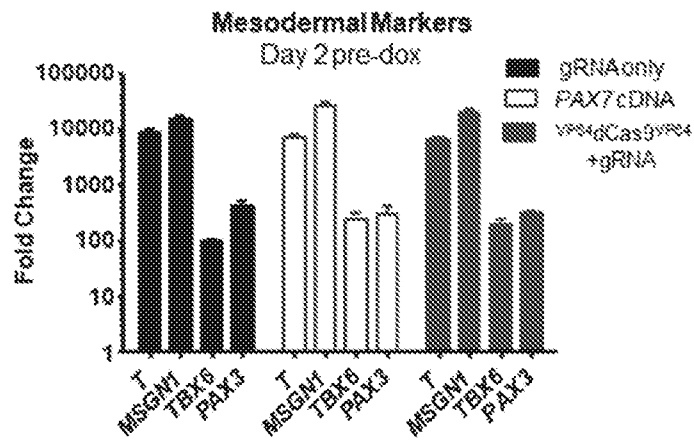


FIG. 1D

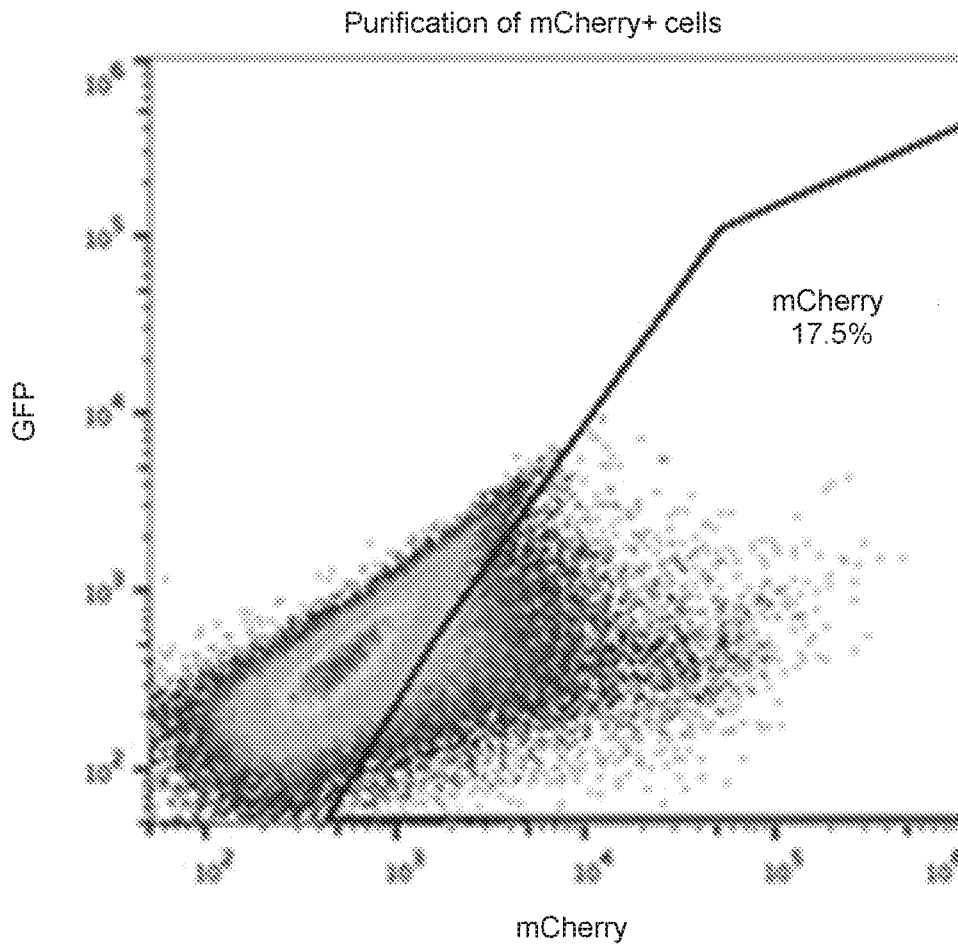


FIG. 1E

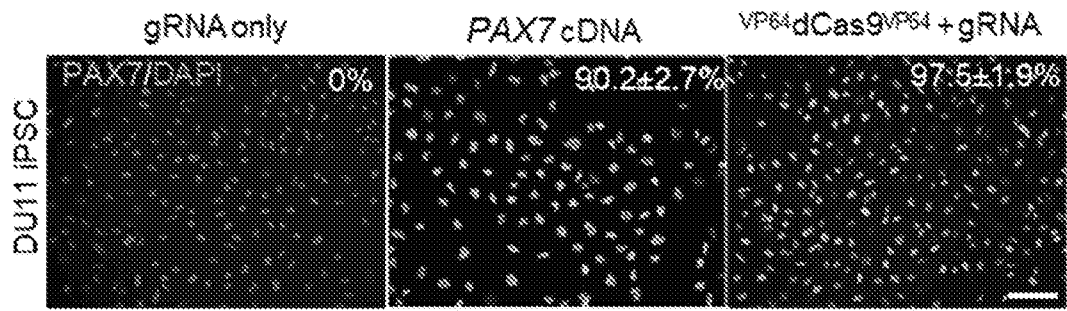


FIG. 1F

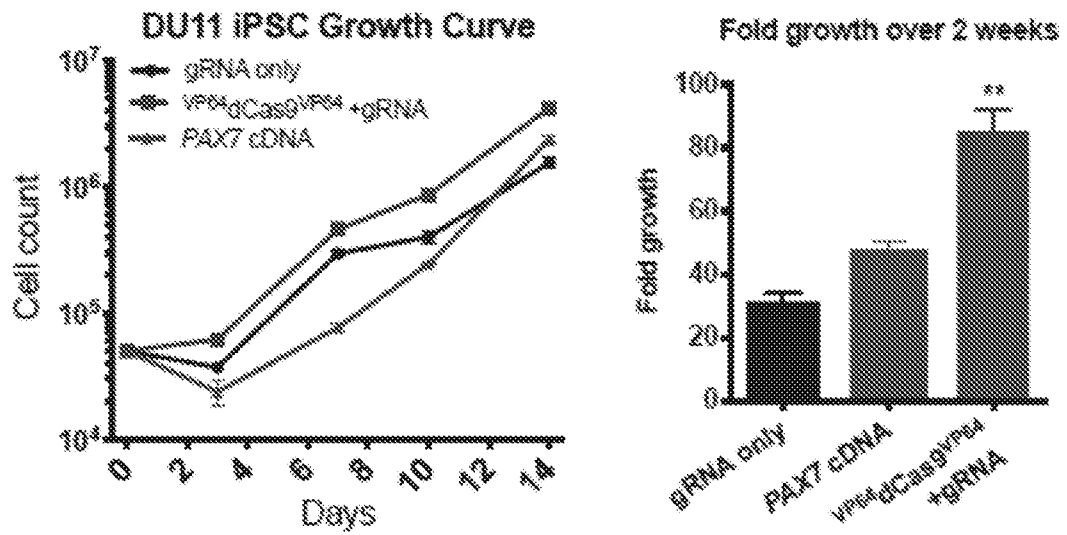


FIG. 1G

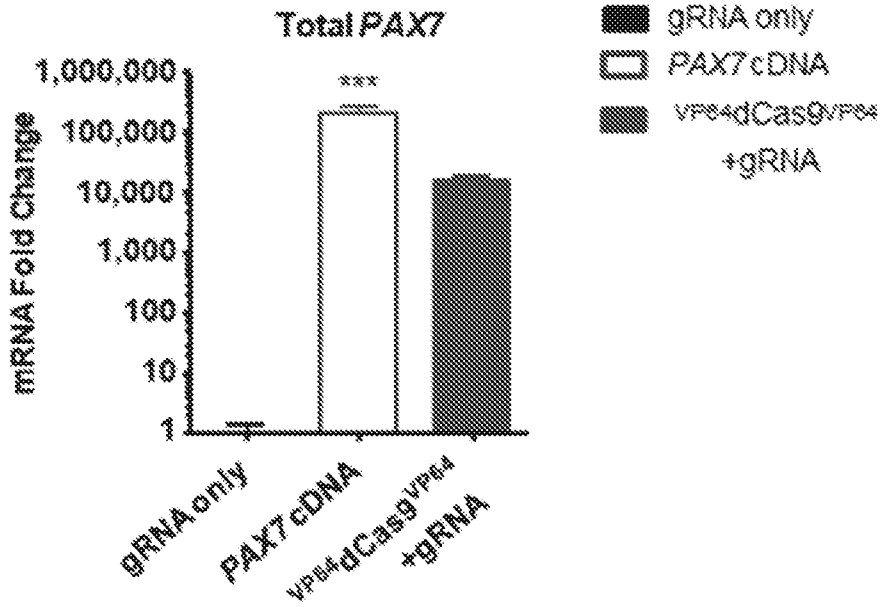


FIG. 2A

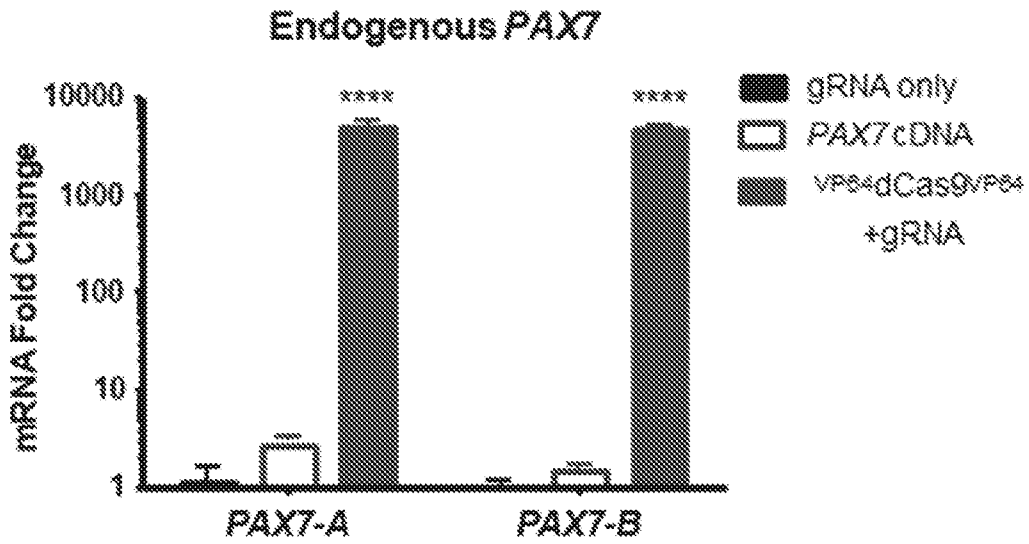


FIG. 2B

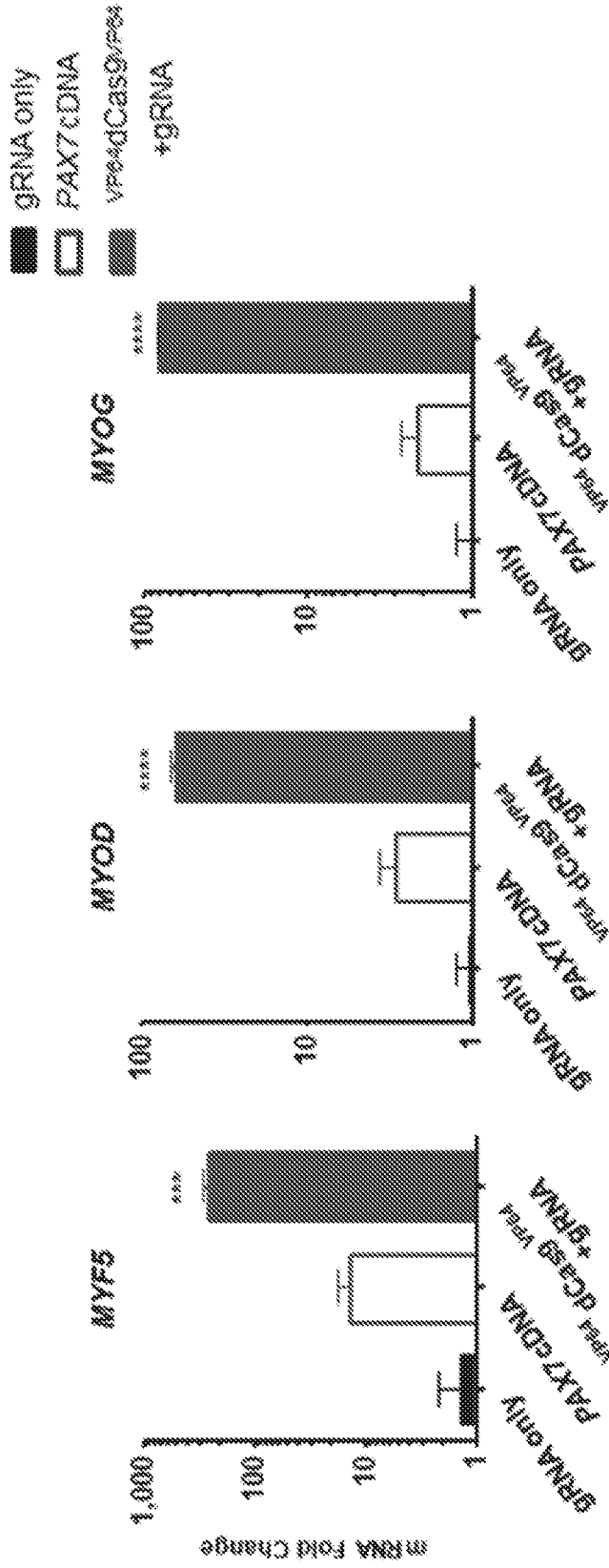


FIG. 2C

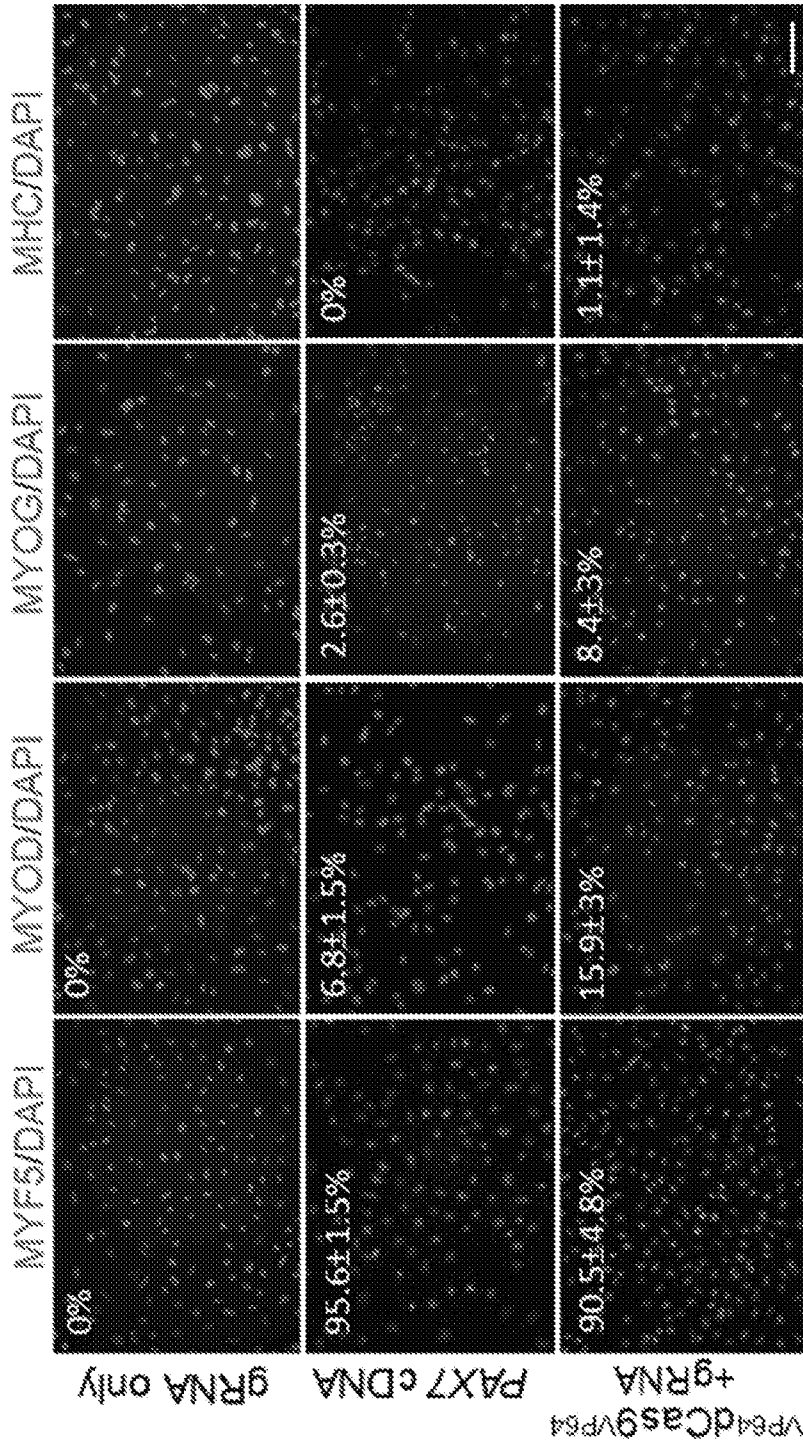


FIG. 2D

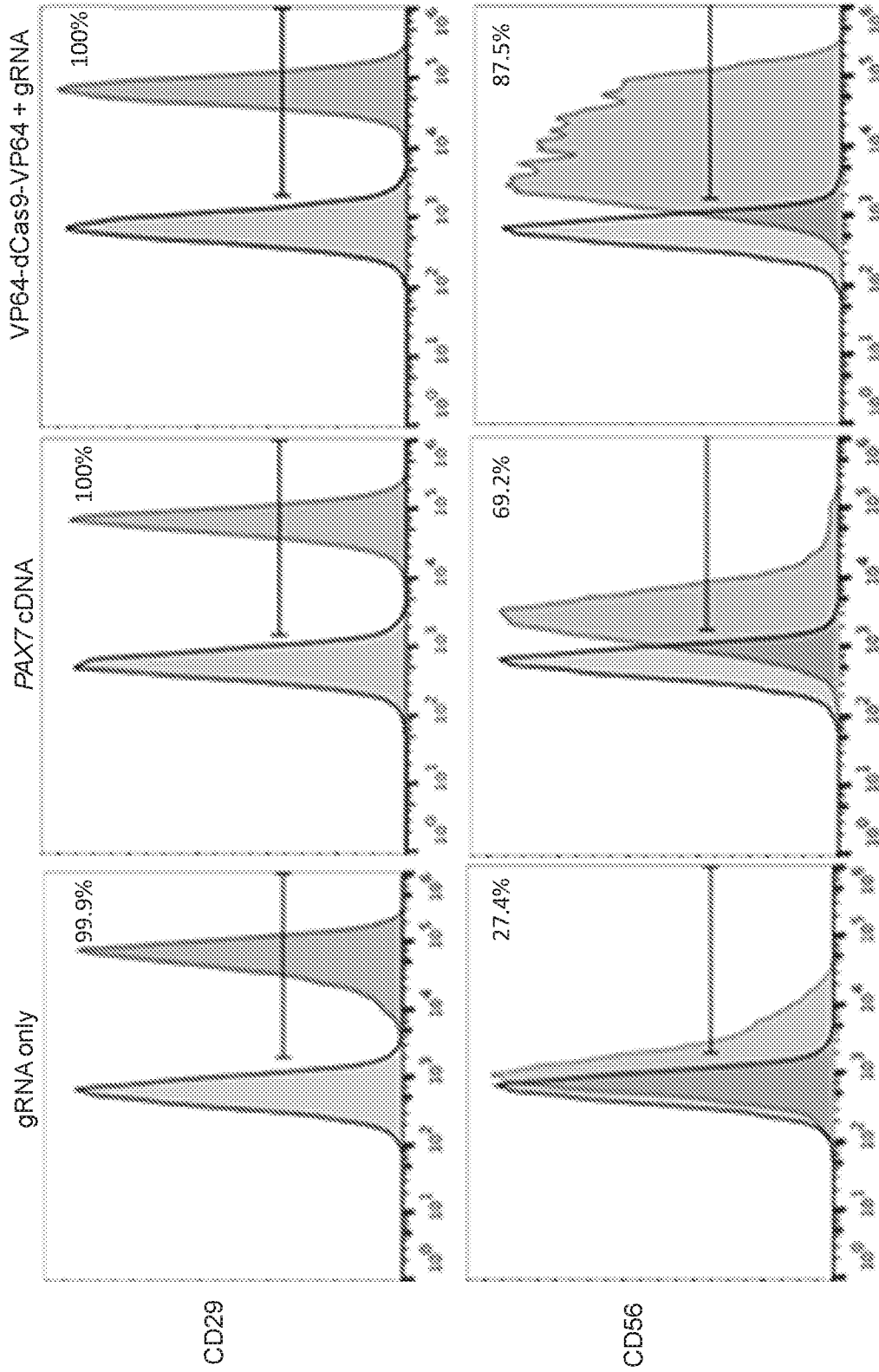


FIG. 2E

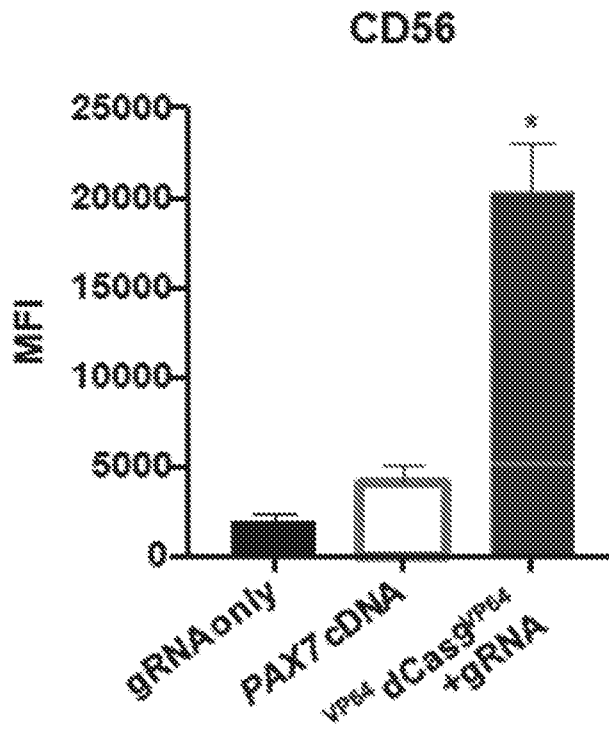


FIG. 2F



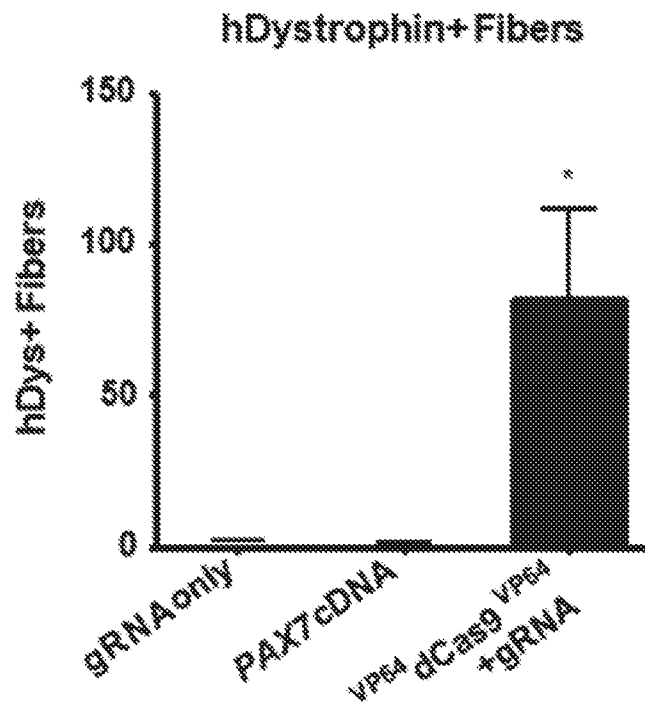


FIG. 3B

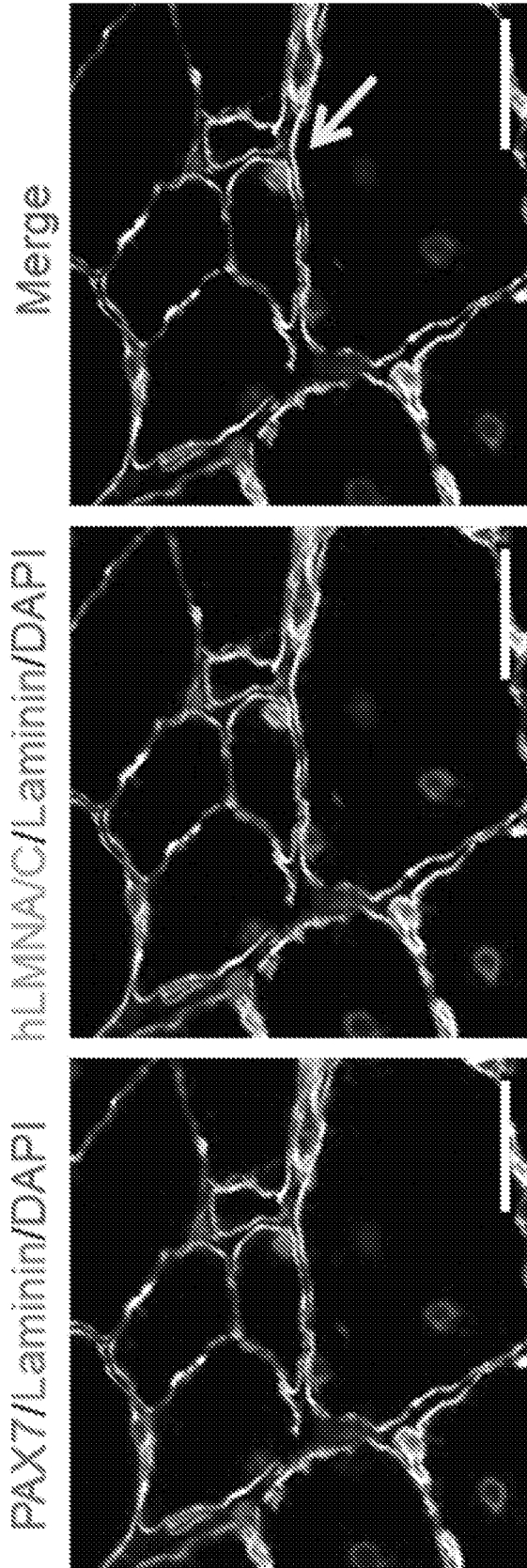


FIG. 3C

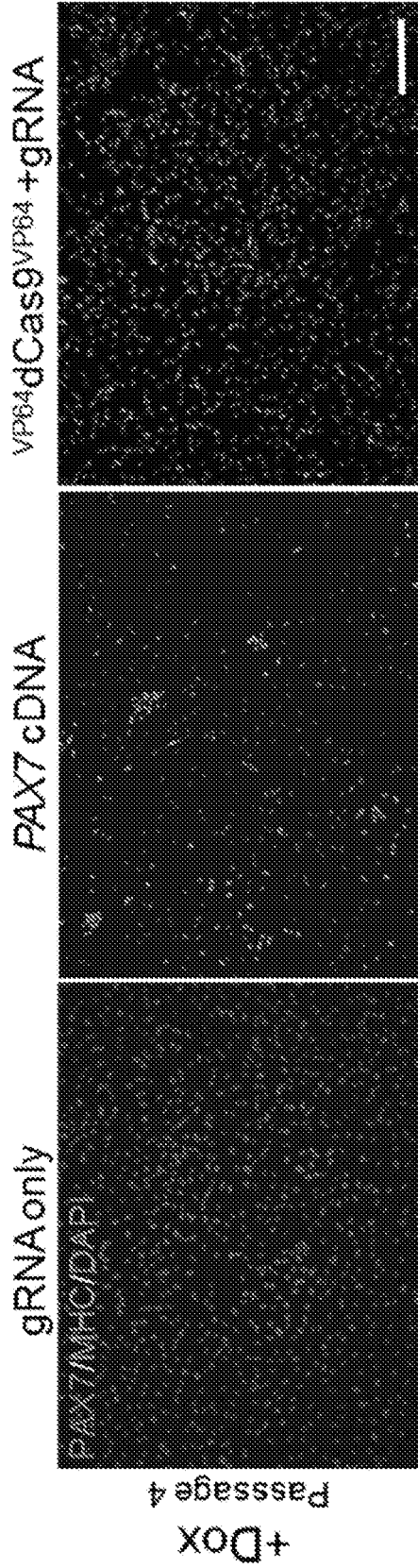


FIG. 4A

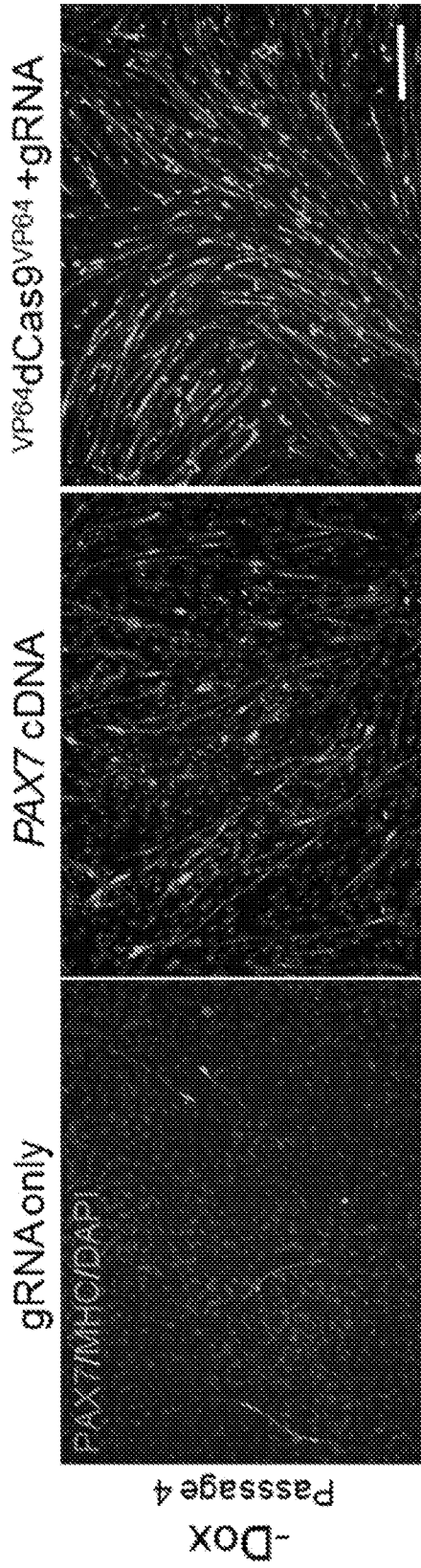


FIG. 4B

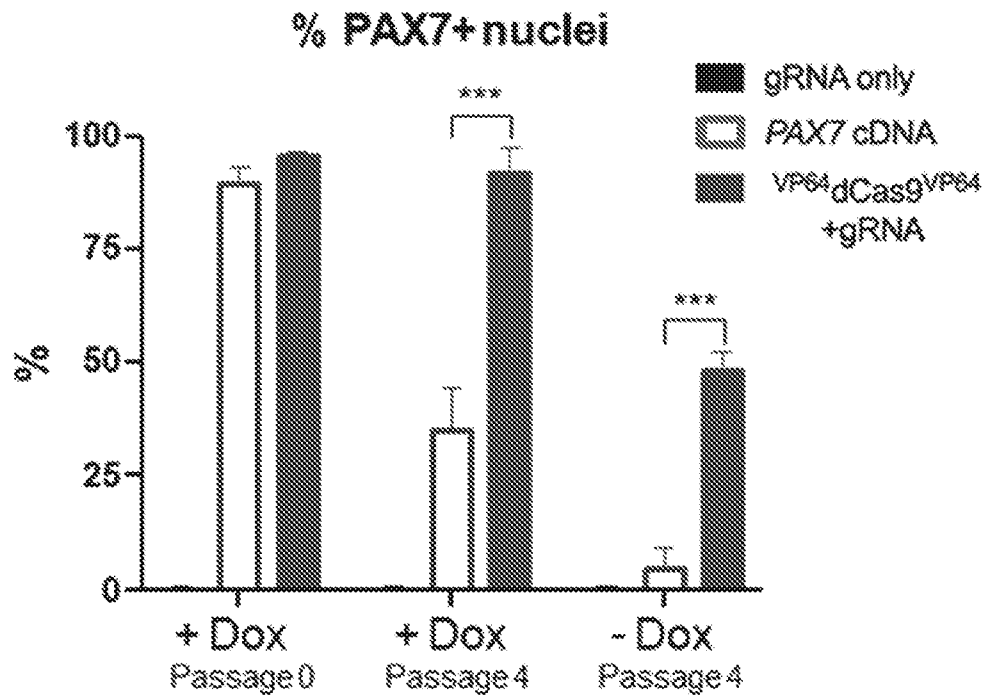


FIG. 4C

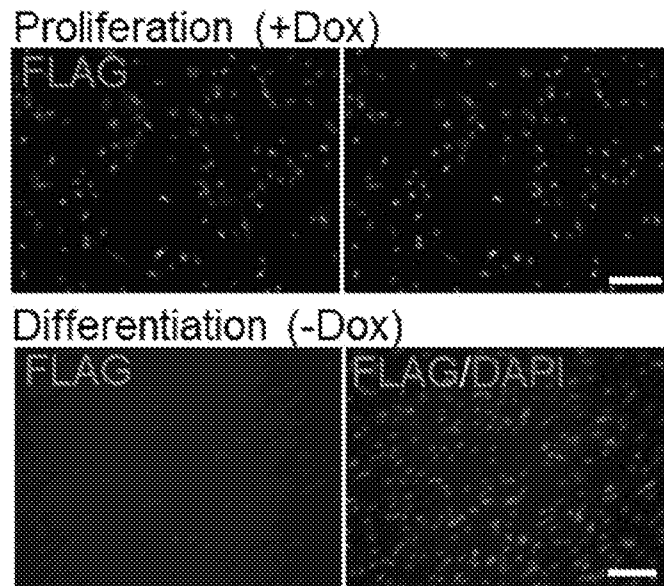


FIG. 4D

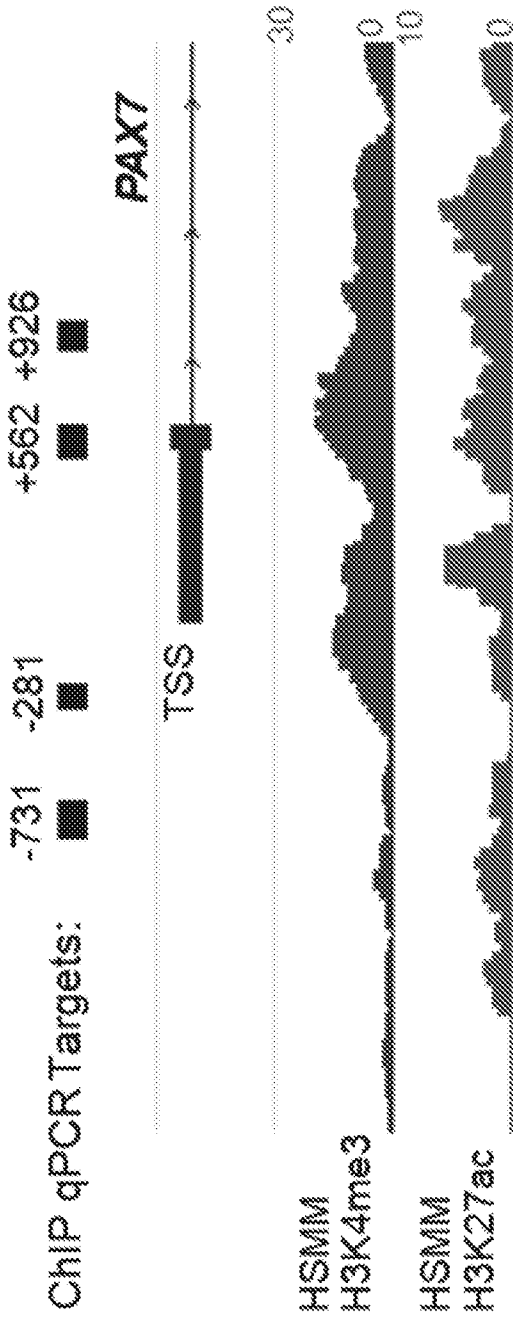


FIG. 5A

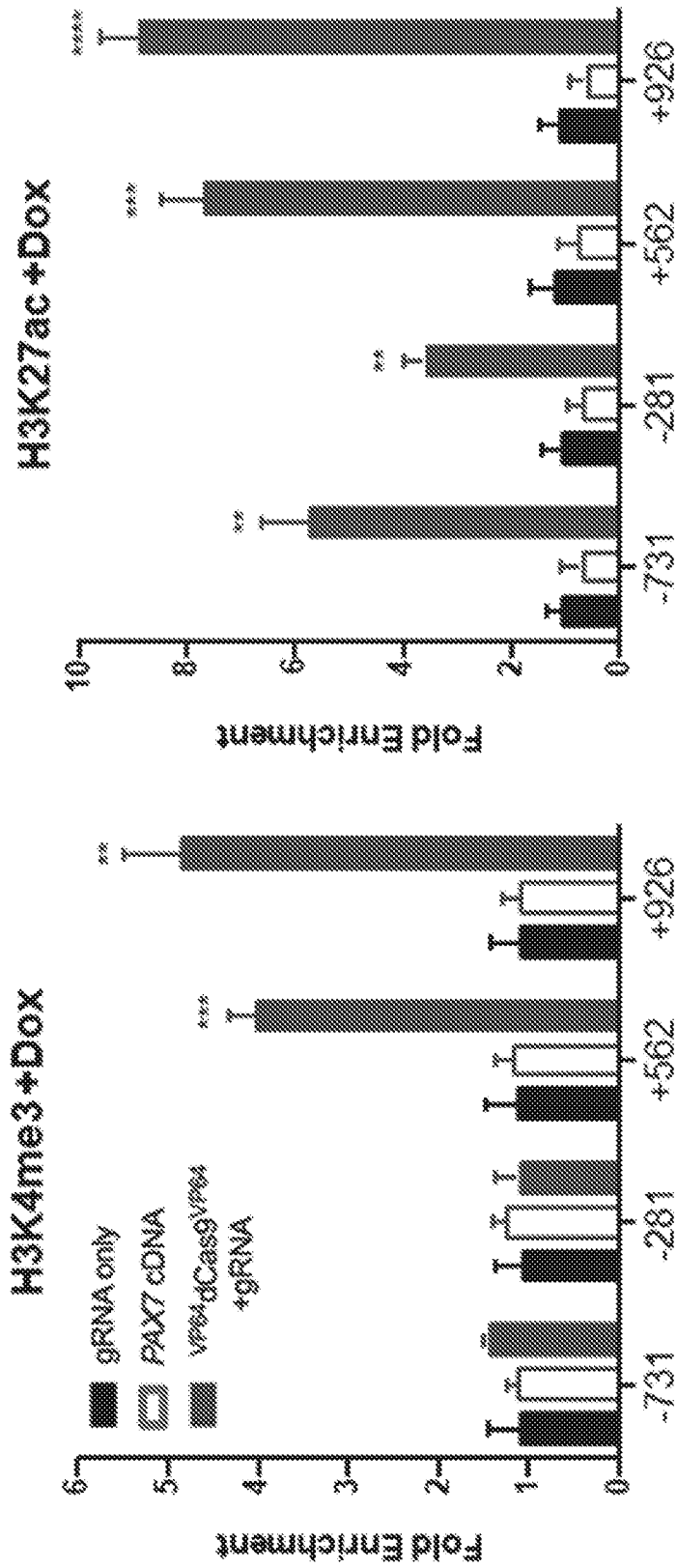


FIG. 5B

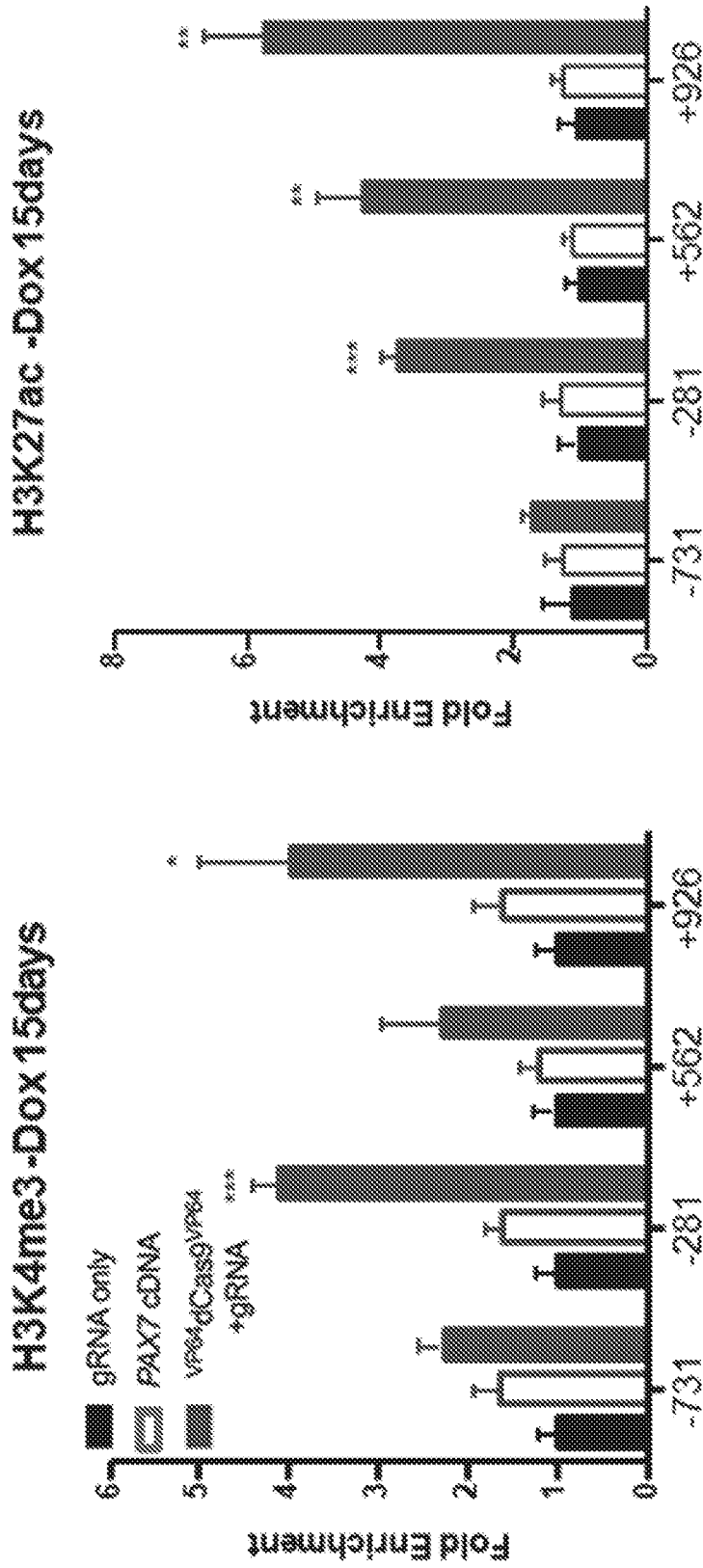


FIG. 5C

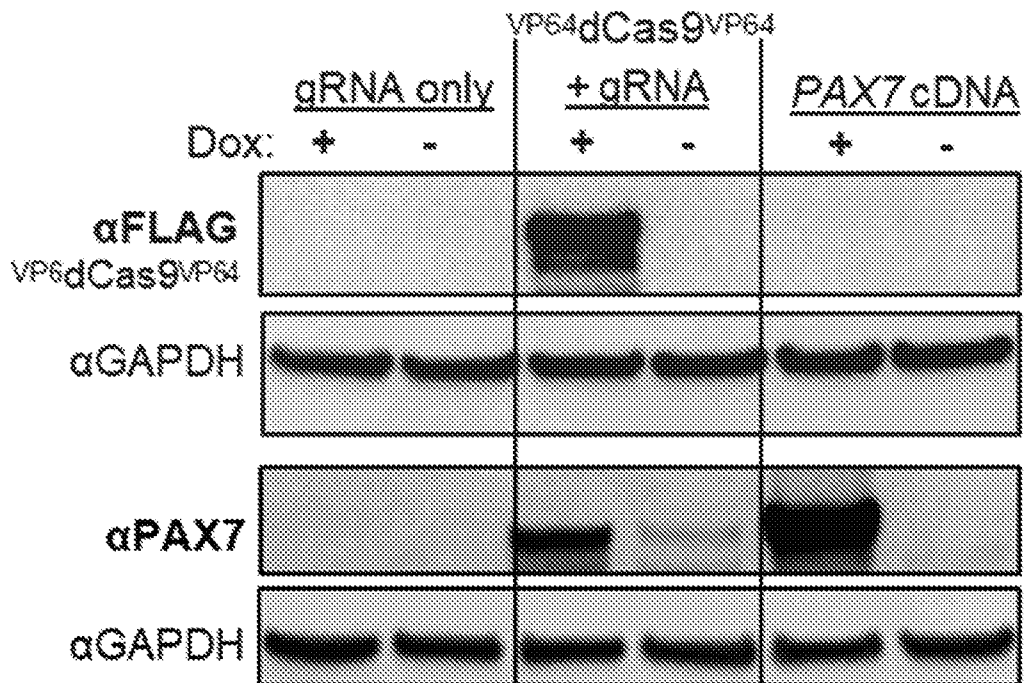


FIG. 5D

Sample Distance Matrix

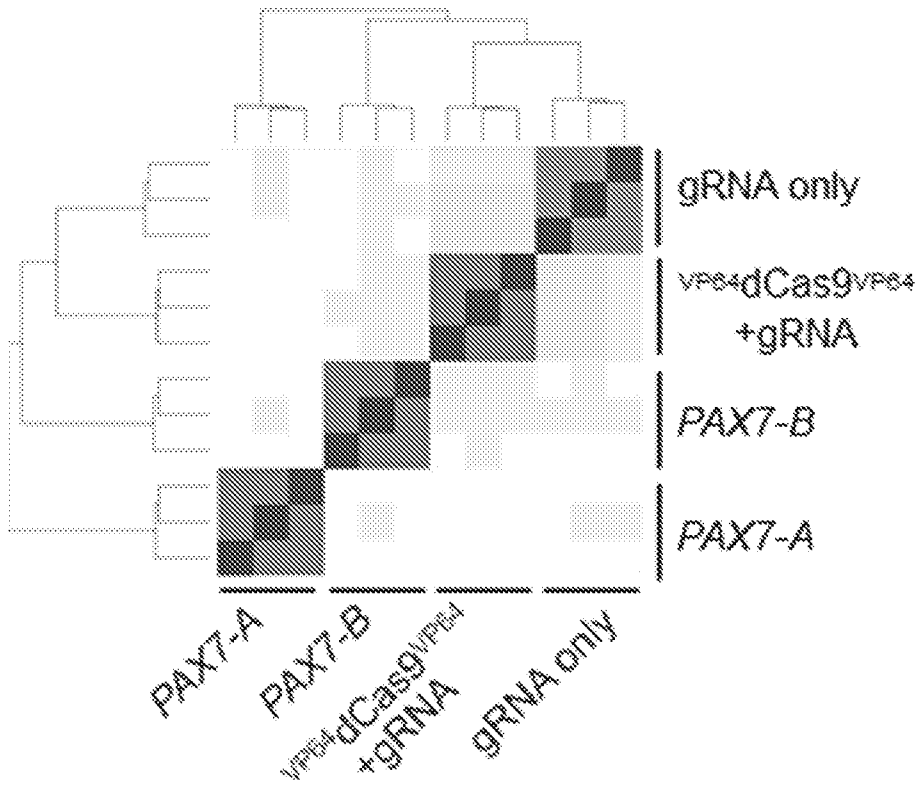


FIG. 6A

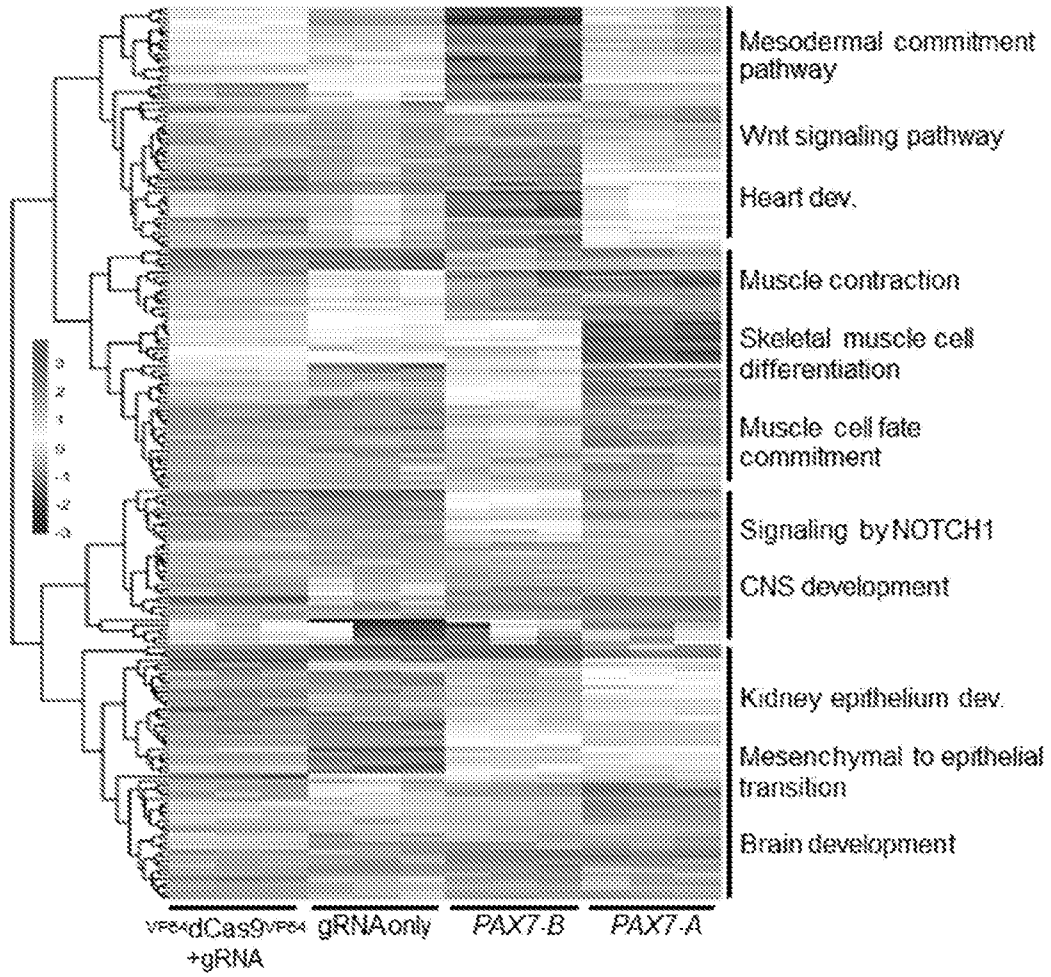


FIG. 6B

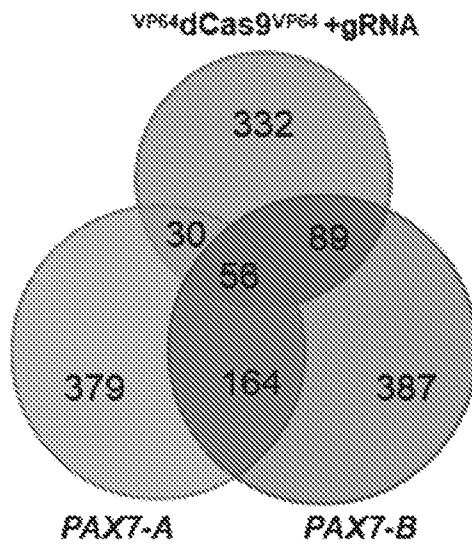


FIG. 6C

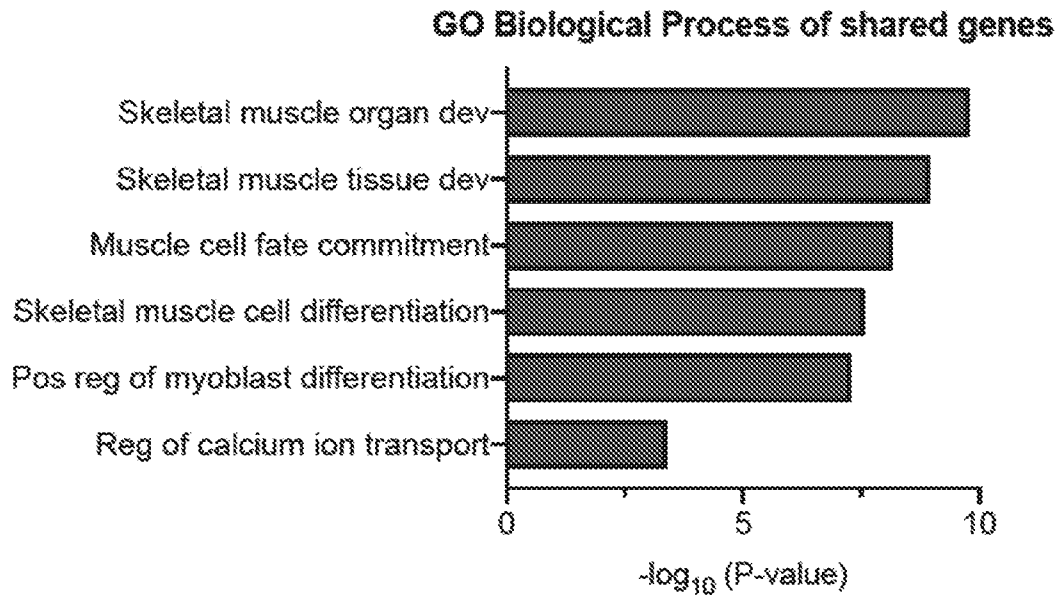


FIG. 6D

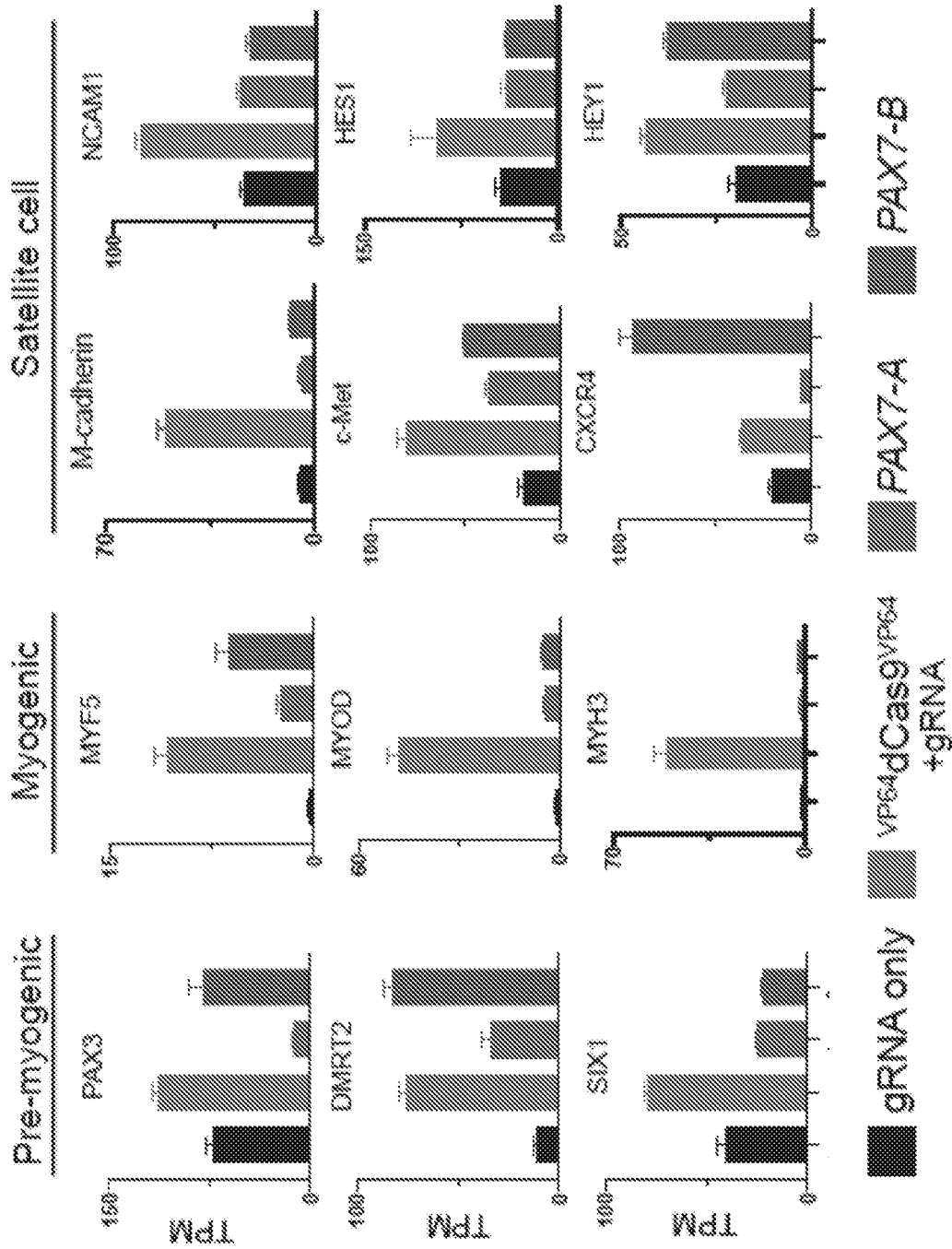


FIG. 6E

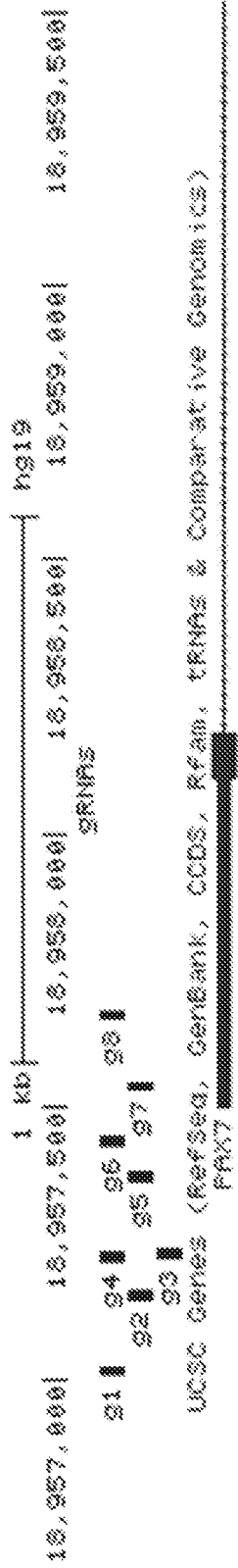


FIG. 7A

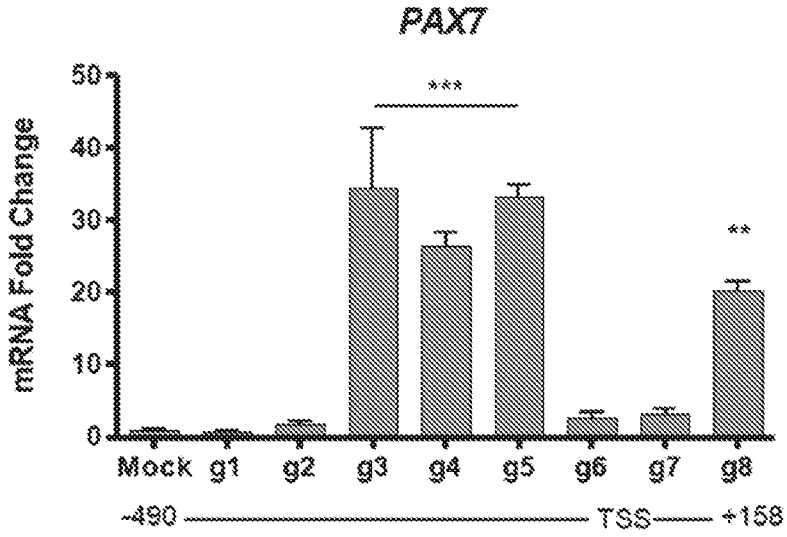


FIG. 7B

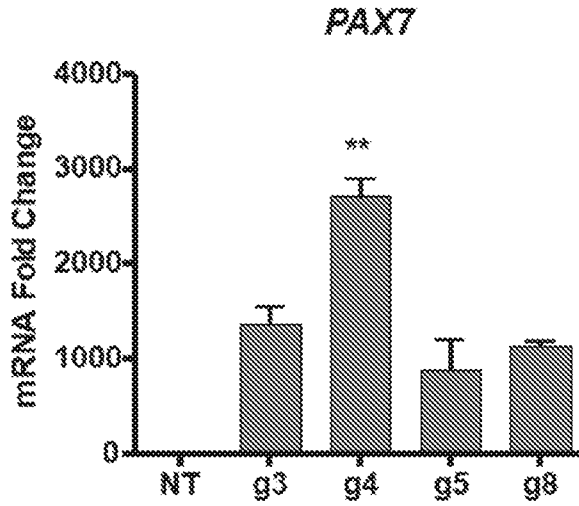


FIG. 7C

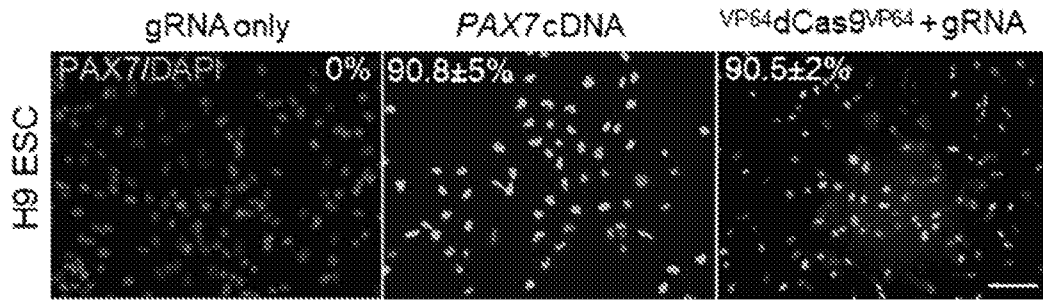


FIG. 8A

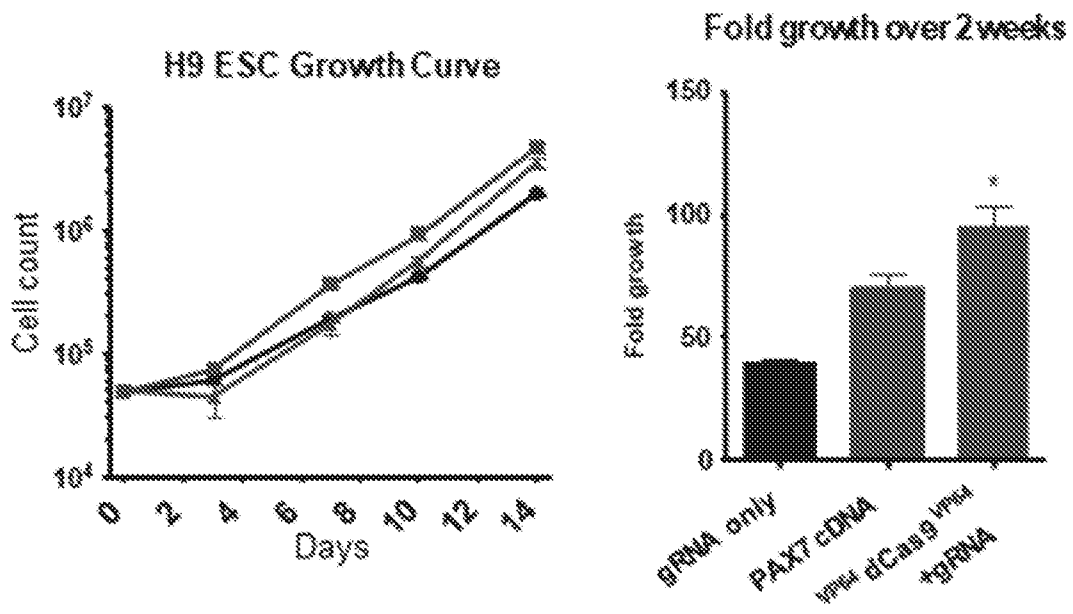


FIG. 8B

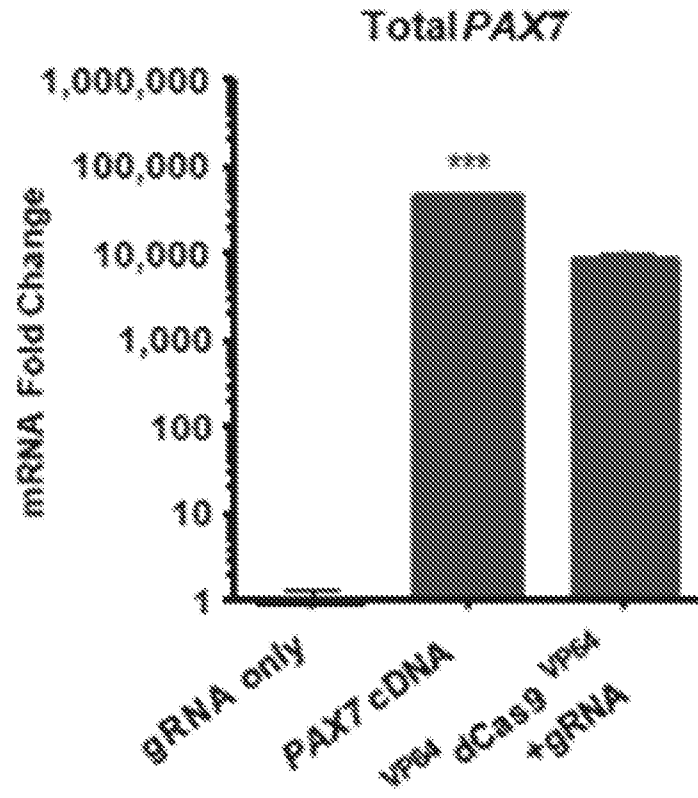


FIG. 8C

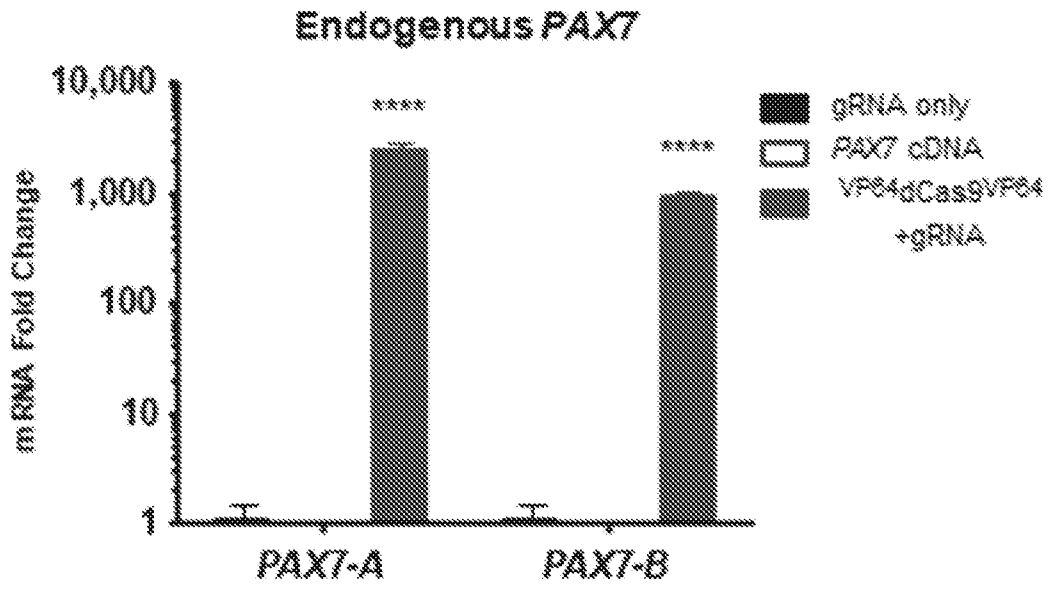


FIG. 8D

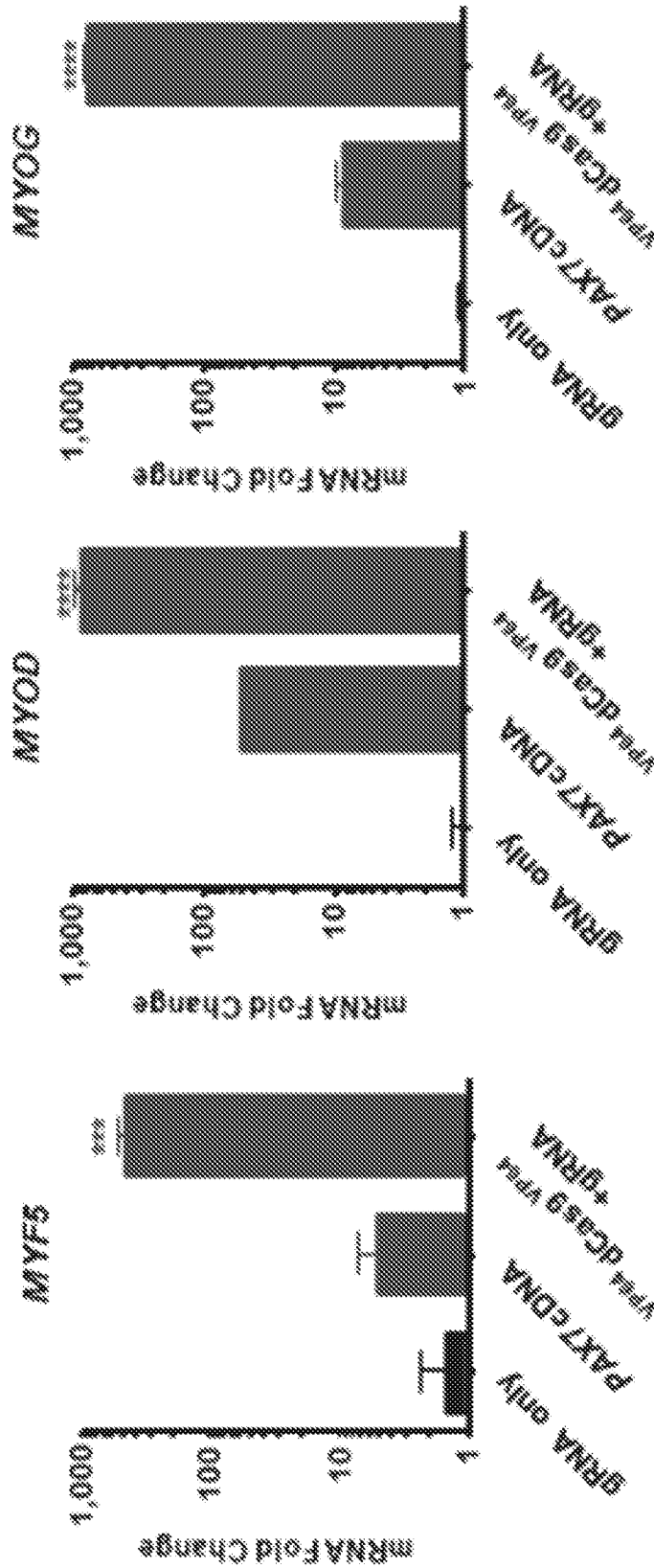


FIG. 8E

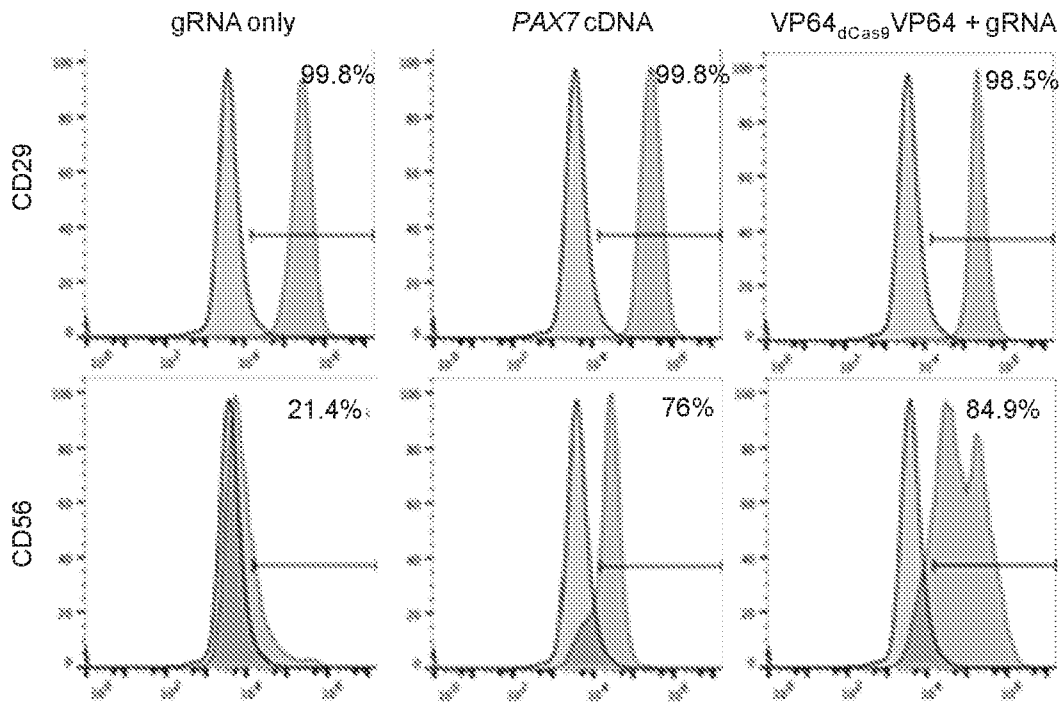


FIG. 8F

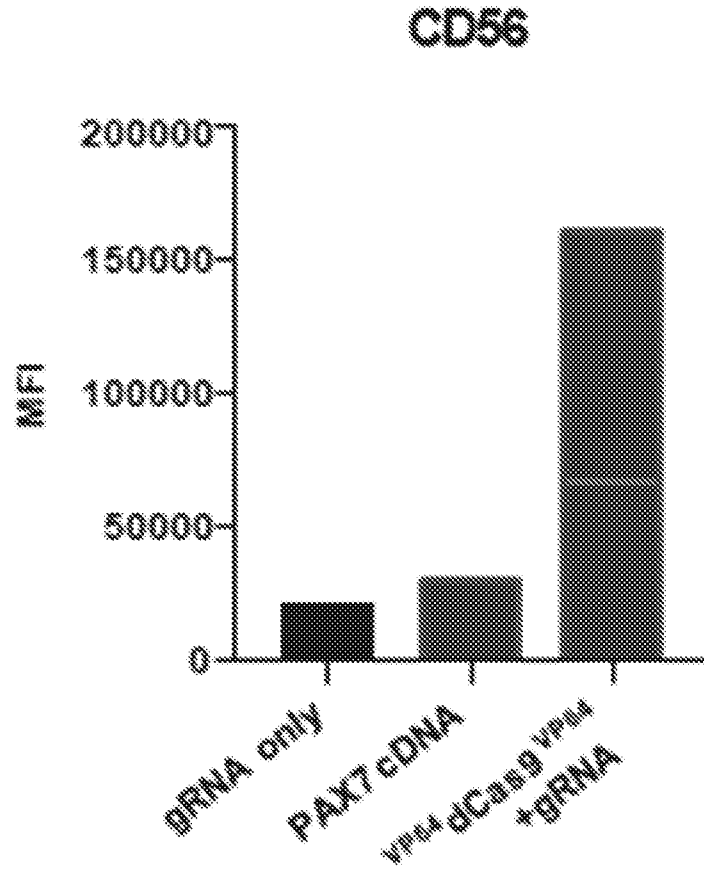


FIG. 8G

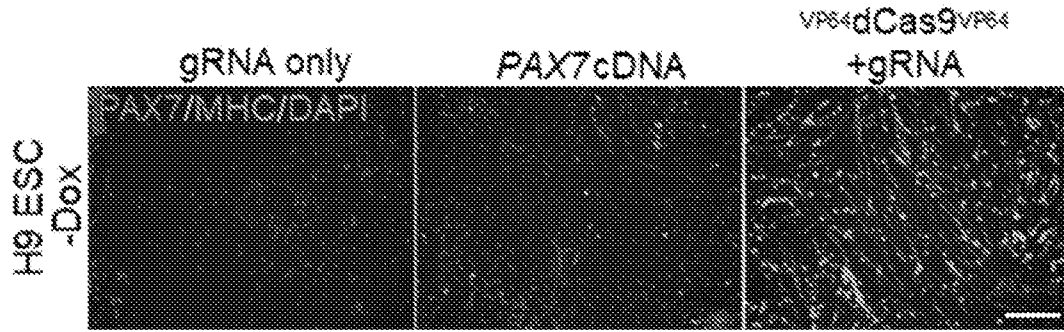


FIG. 8H

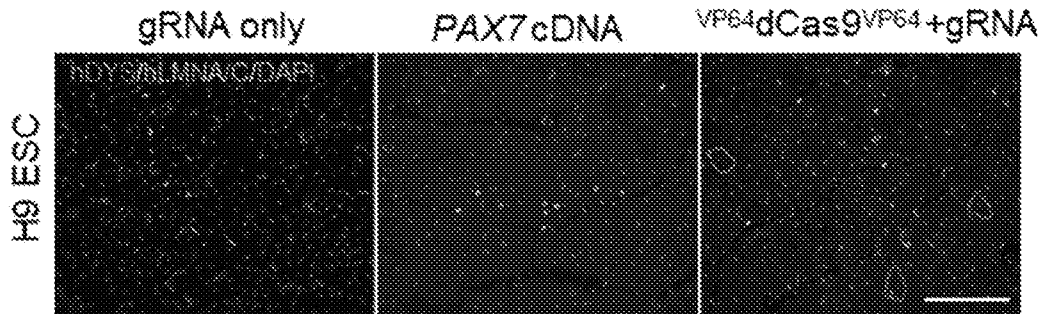


FIG. 8I

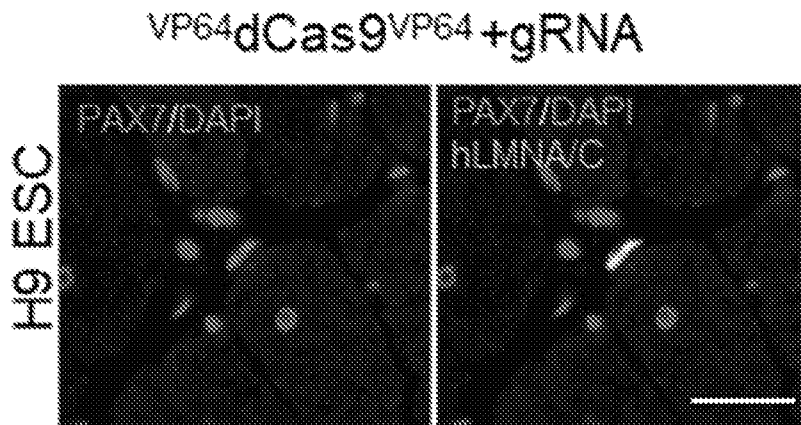


FIG. 8J

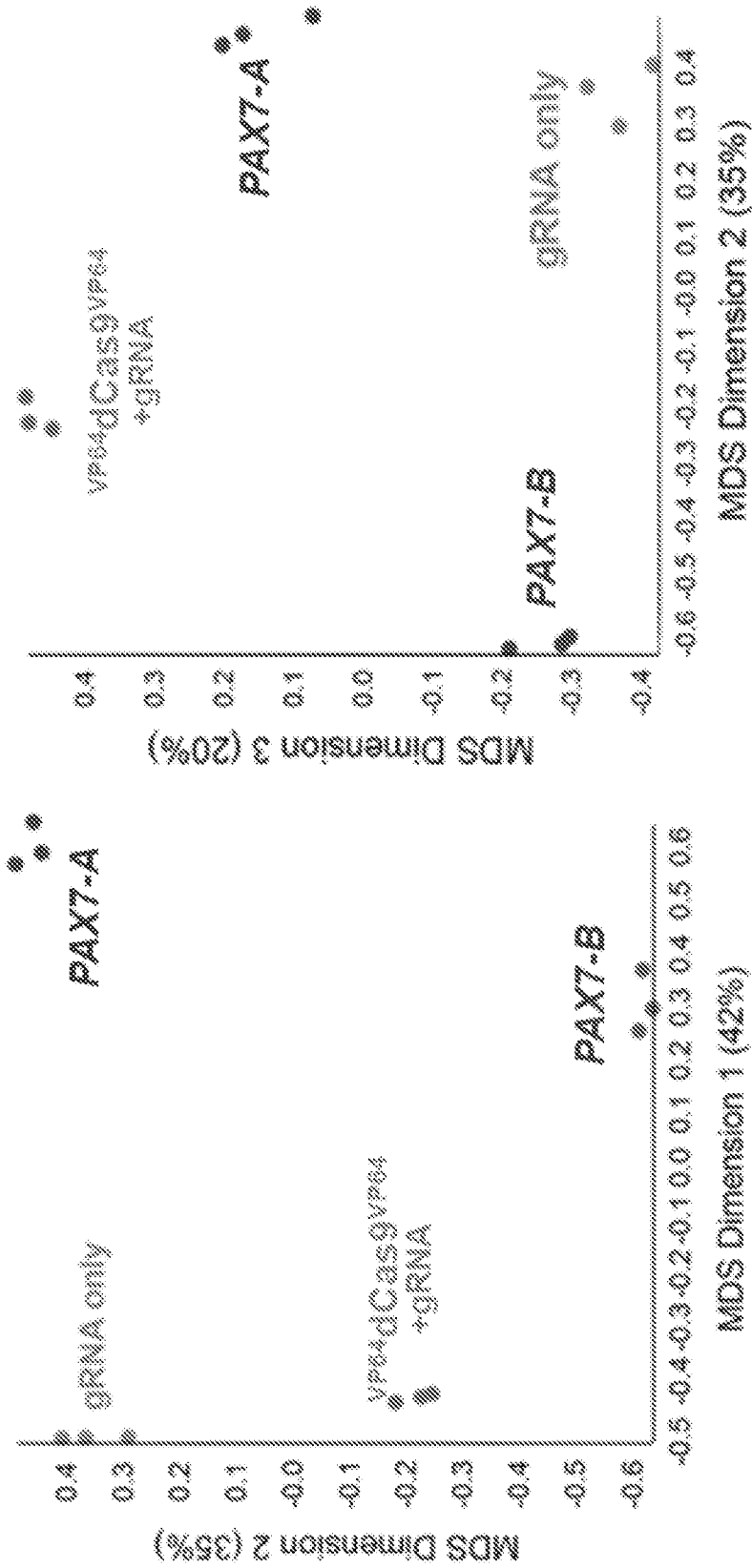


FIG. 9A

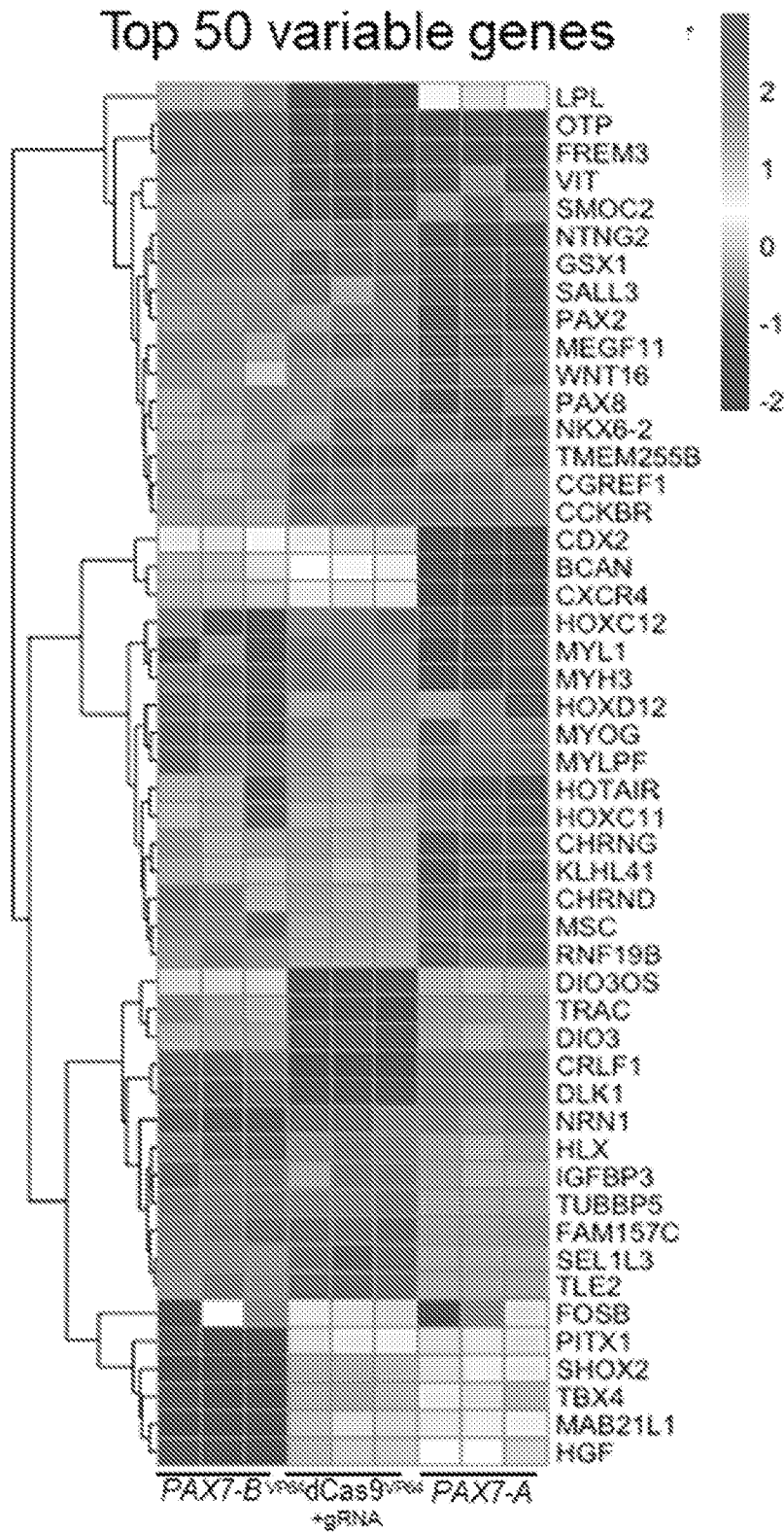


FIG. 9B

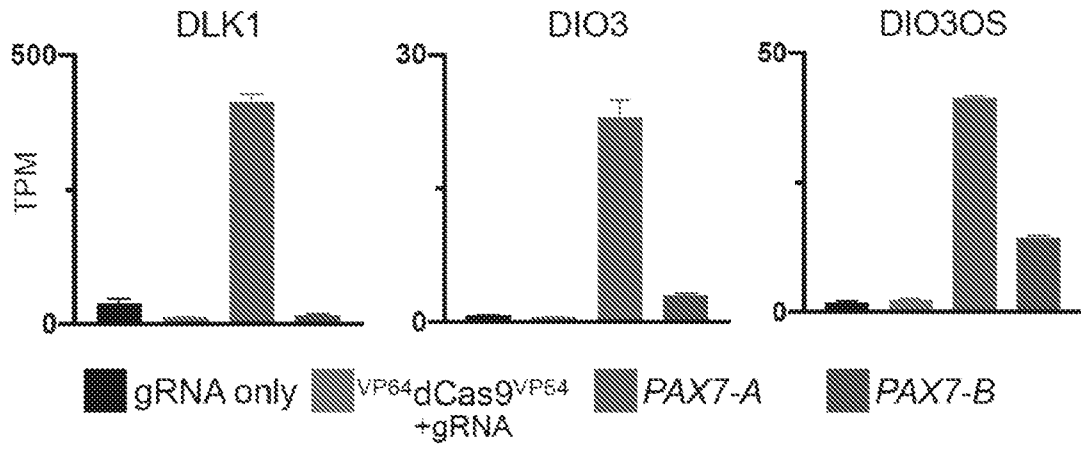


FIG. 9C

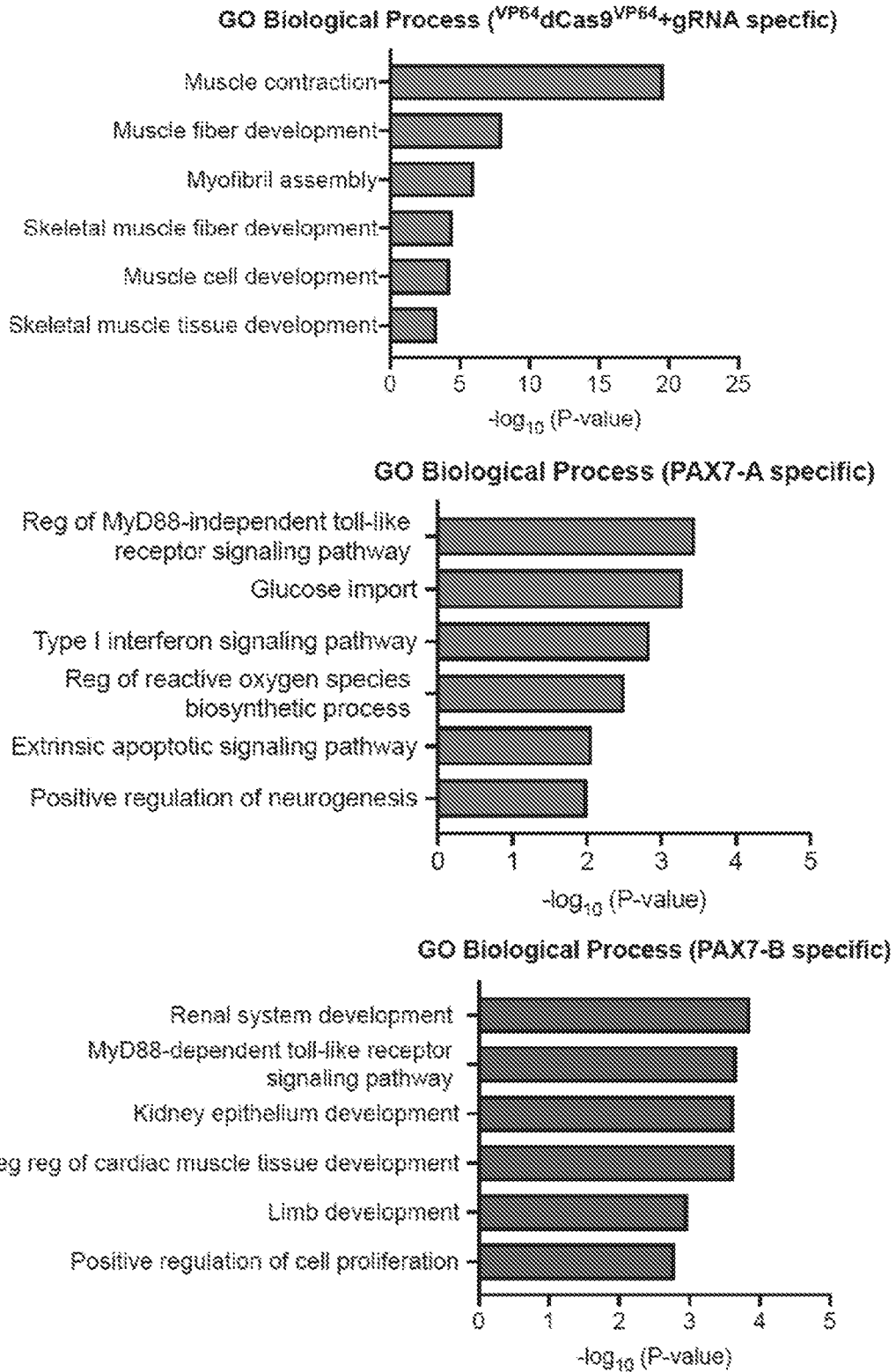


FIG. 9D

Known satellite cell surface markers

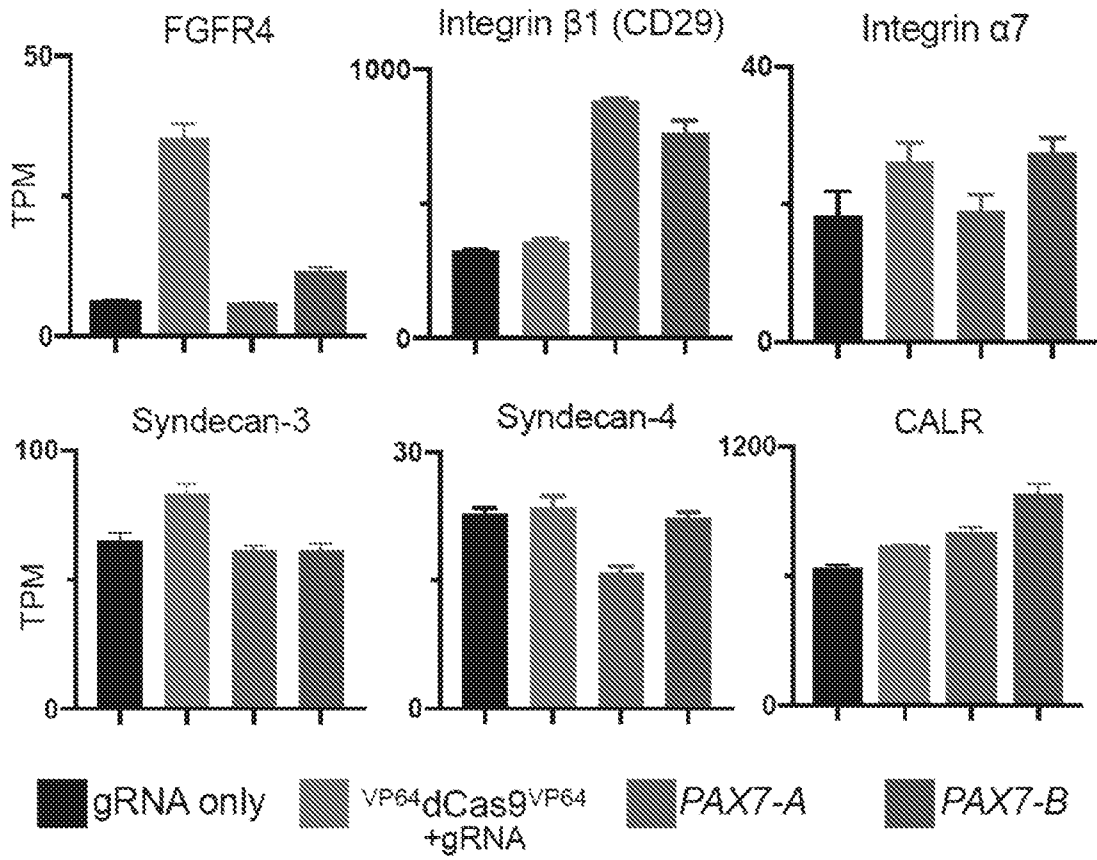


FIG. 9E

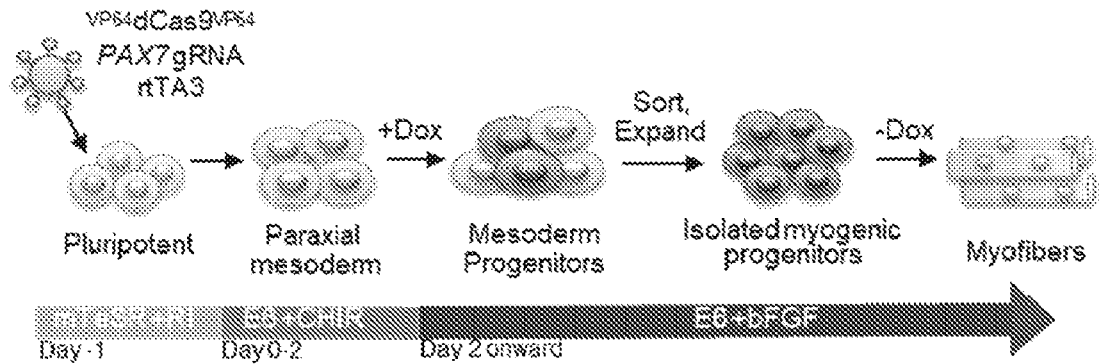


FIG. 1A