COMPOSITIONS FOR GENE ACTIVATION COMPRISING P300 DNA BINDING DOMAINS AND ASSOCIATED METHODS

CROSS REFERENCE TO RELATED APPLICATION(S)

[0001] This application claims the benefit of U.S. Provisional Patent Application No. 63/574,735, filed April 4, 2024, and of U.S. Provisional Patent Application No.63/748,898, filed January 23, 2025, all of which applications are incorporated herein by reference in their entirety.

SEQUENCE LISTING

[0002] The application contains a Sequence Listing which has been submitted electronically in .XML format and is hereby incorporated by reference in its entirety. Said .XML copy, created on April 3, 2025, is named "153697.8003.WO00.xml" and is 146,120 bytes in size. The sequence listing contained in this .XML file is part of the specification and is hereby incorporated by reference herein in its entirety.

BACKGROUND

[0003] The histone acetyltransferase enzyme, p300, plays a crucial role in cellular processes by acetylating histones and non-histone proteins. p300 mediated histone acetylation influences chromatin structure and function, facilitates transcriptional activation of genes, and influences cell cycle progression, differentiation, and apoptosis. p300 activity is essential for the regulation of numerous biological pathways and has implications in the development and progression of various diseases, including cancer, where it may act as a co-activator for several transcription factors. Recently, p300 core domain has been fused to a catalytically inactive (dead) Cas9 (dCas9) system to facilitate gene activation. However, overexpression of p300 and/or its core domain may lead to off-target epigenetic and gene expression changes, negatively impacting its therapeutic potential. As such, more precise gene activation having reduced off target effects are needed.

SUMMARY

[0004] In some embodiments, the present technology comprises an engineered, non-naturally occurring peptide comprising or consisting of a p300 recruiter domain of double homeobox, 4 (DUX4) (p300R domain).

[0005] In some embodiments, the present technology comprises an engineered, non-naturally occurring peptide comprising or consisting of an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 100% identical to SEQ ID NO: 1.

[0006] In some embodiments, the present technology comprises an engineered, non-naturally occurring peptide comprising or consisting of a p300 recruiter domain (p300R domain) and a DNA binding domain (DBD) peptide.

[0007] In some embodiments, the present technology comprises an engineered, non-naturally occurring peptide comprising or consisting of a p300 recruiter domain (p300R domain), a DNA binding domain (DBD), and a linker peptide.

[0008] In some embodiments, the present technology comprises a composition comprising an engineered, non-naturally occurring peptide comprising or consisting of a p300 recruiter domain (p300R domain).

[0009] In some embodiments, the present technology comprises a composition comprising an engineered, non-naturally occurring peptide comprising a p300 recruiter domain (p300R domain) and a DNA binding domain (DBD) peptide.

[0010] In some embodiments, the present technology comprises a composition comprising an engineered, non-naturally occurring peptide comprising a p300 recruiter domain (p300R domain), a DNA binding domain (DBD), and a linker peptide.

[0011] In some embodiments, the present technology comprises a composition comprising an engineered, non-naturally occurring peptide comprising or consisting of an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 100% identical to SEQ ID NO: 1.

[0012] In some embodiments, the present technology comprises a composition comprising (i) a peptide comprising a DNA binding domain (DBD) fused to a p300R

domain (p300R-DBD), the p300R-DBD protein selectively binding to a locus that is altered by methylation of one or more histones and (ii) one or more guide RNA (gRNA).

[0013] In some embodiments, the DBD comprises a clustered regularly interspaced short palindromic repeats associated protein (Cas) domain, a zinc finger domain, a helix-turn-helix (HTH) domain, a leucine zipper domain, and a basic helix-loop-helix (bHLH) domain.

[0014] In some embodiments, the Cas domain comprises an inactive nuclease domain.

[0015] In some embodiments, the Cas domain comprising an inactive nuclease domain is a dead Cas (dCas) domain.

[0016] In some embodiments, the present technology comprises a composition comprising (i) a peptide comprising a dCas domain fused to a p300R domain (p300R-dCas), the p300R-dCas9 peptide selectively binding to a locus that is altered by methylation of one or more histones and (ii) a gRNA.

[0017] In some embodiments, the locus comprises a nucleotide sequence encoding a gene.

[0018] In some embodiments, the locus comprises a nucleotide sequence that is a cis regulatory element of a gene.

[0019] In some embodiments, the locus comprises a nucleotide sequence that is a trans regulatory element of a gene.

[0020] In some embodiments, the gene is LAMA1.

[0021] In some embodiments, the gene is selected from the group consisting of SMCHD1, LDRF, MATR3, IGF1, LAMA1, and PML

[0022] In some embodiments, the locus comprises or consists of a promoter.

[0023] In some embodiments, the promoter is a LAMA1 promoter.

[0024] In some embodiments, the present technology comprises a composition comprising (i) a peptide comprising a p300R-dCas peptide, the p300R-dCas9 peptide selectively binding to a locus altered by methylation of one or more histones and comprising at least a portion of a LAMA1 promoter and (ii) a gRNA.

[0025] In some embodiments, the Cas domain or the dCas domain comprises a Cas1 domain, a Cas2 domain, a Cas3 domain, Cas9 domain, a Cas10 domain, a Cas12a domain, or a Cas13 domain.

[0026] In some embodiments, the Cas domain or the dCas domain comprises an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 100% identical to the amino acid sequence of any one of SEQ ID NOS: 34 or 40-57.

[0027] In some embodiments, the gRNA specifically directs the p300R-DBD protein to the locus.

[0028] In some embodiments, the gRNA mediates selective binding of the p300R-DBD protein, the p300R-Cas peptide, or the p300R-dCas peptide to the locus.

[0029] In some embodiments, the gRNA comprises a nucleotide sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 100% identical to one or more of SEQ ID NOs: 58-80.

[0030] In some embodiments, the p300R domain comprises or consists of an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 100% identical to SEQ ID NO: 1.

[0031] In some embodiments, the p300R domain is fused to the DBD domain, the Cas domain, or the dCas domain by a peptide linker.

[0032] In some embodiments, the peptide linker comprises an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 100% identical to the amino acid sequence of any one of SEQ ID NOs: 15-33 or 81.

[0033] In some embodiments, the locus altered by the methylation of the one or more histones has a reduction in expression of the gene by at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100% relative to the control.

[0034] In some embodiments, the locus altered by the methylation of the one or more histones is repressed by the methylation.

[0035] In some embodiments, the one or more histones are selected from the group consisting of H2A, H2B, H3, and H4.

[0036] In some embodiments, the one or more histones are methylated at a lysine (K) residue or an arginine (R) residue.

[0037] In some embodiments, the lysine residue is a K4, K5, K8, K9, K14, K16, K18, K20, K23, K27, K36, K56, K119, or a K120 residue.

[0038] In some embodiments, the arginine residue is an R2, R3, R8, R17, R23, or an R26 residue.

[0039] In some embodiments, the present technology comprises a host cell comprising an engineered, non-naturally occurring peptide or a composition of the present technology.

[0040] In some embodiments, the host cell is a mammalian host cell.

[0041] In some embodiments, the mammalian host cell is a human host cell or a murine host cell.

[0042] In some embodiments, the human host cell is a myoblast.

[0043] In some embodiments, the mammalian cell is a kidney cell, a liver cell, or a muscle cell.

[0044] In some embodiments, the present technology comprises a vector comprising a nucleotide sequence encoding the engineered peptide, the peptide, or the gRNA of the present technology.

[0045] In some embodiments, the vector is a viral vector.

[0046] In some embodiments, the viral vector is a lentiviral vector.

[0047] In some embodiments, the vector comprises a detectable marker or a reporter gene.

[0048] In some embodiments, the vector comprises an antibiotic-inducible gene expression system.

[0049] In some embodiments, the antibiotic is doxycycline or tetracycline.

[0050] In some embodiments, the antibiotic-inducible gene expression system is a Tet-ON system.

[0051] In some embodiments, the composition is present in a delivery vehicle.

[0052] In some embodiments, the vector is present in a delivery vehicle.

[0053] In some embodiments, the delivery vehicle comprises a nanoparticle.

[0054] In some embodiments, the nanoparticle is a lipid nanoparticle.

[0055] In some embodiments, the present technology comprises a pharmaceutical composition comprising the composition or the vector of the present technology.

[0056] In some embodiments, the pharmaceutical composition is formulated for delivery to a subject in need thereof.

[0057] In some embodiments, the subject is mammalian.

[0058] In some embodiments, the subject is human.

[0059] In some embodiments, the present technology comprises a method of increasing a level of Laminin Subunit Alpha 1 (LAMA1), Growth Factor 1 (IGF1), or adiponectin (ADIPOQ) gene or protein expression in a subject in need thereof relative to a control, the method comprising, administering to the subject: (i) an engineered, non-naturally occurring peptide or the composition of the present technology; (ii) a host cell of the present technology; (iii) a vector of the present technology; or (iv) a pharmaceutical composition of the present technology.

[0060] In some embodiments, the level of LAMA1, IGF1, or ADIPOQ gene or protein expression is increased at least about 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 200%, 300%, 400%, 500%, 600%, 700%, 800%, 900%, or 1000%, relative to a control.

[0061] In some embodiments, the present technology comprises a method of increasing a level of LAMA1, IGF1, or ADIPOQ protein activity in a subject in need thereof relative to a control, the method comprising, administering to the subject: (i) an engineered, non-naturally occurring peptide or the composition of the present technology; (ii) a host cell of the present technology; (iii) a vector of the present technology; or (iv) a pharmaceutical composition of the present technology.

[0062] In some embodiments, the composition, the vector, the pharmaceutical composition, or the host cell is administered to the subject at a dose of at least about 0.5 ng to at least about 500 µg.

[0063] In some embodiments, the vector comprises a detectable marker or a reporter gene, and a locus that is altered by methylation of one or more histones is

altered when the reporter gene or the detectable marker is not expressed at a detectable level.

[0064] In some embodiments, the level of LAMA1, IGF1, or ADIPOQ protein activity is increased at least about 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 200%, 300%, 400%, 500%, 600%, 700%, 800%, 900%, or 1000%, relative to a control.

[0065] In some embodiments, an expression level of a gene encoding one or more proteins that increases LAMA1, IGF1, or ADIPOQ expression or activity is increased by at least about 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 200%, 300%, 400%, 500%, 600%, 700%, 800%, 900%, 1000%, or more, relative to a control.

[0066] In some embodiments, an expression level of a gene encoding one or more proteins that reduces LAMA1, IGF1, or ADIPOQ expression or activity is reduced by at least about 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100% relative to a control.

[0067] In some embodiments, expression or activity of LAMA1, IGF1, or ADIPOQ is increased compared to a control cell.

[0068] In some embodiments, a level of LAMA1, IGF1, or ADIPOQ gene or protein expression is increased at least about 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 200%, 300%, 400%, 500%, 600%, 700%, 800%, 900%, 1000%, or more, relative to a control.

[0069] In some embodiments, an expression level of a gene encoding one or more proteins that increases LAMA1, IGF1, or ADIPOQ expression or activity is increased by at least about 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 200%, 300%, 400%, 500%, 600%, 700%, 800%, 900%, 1000%, or more, relative to the control cell.

[0070] In some embodiments, an expression level of a gene encoding one or more proteins that represses LAMA1, IGF1, or ADIPOQ expression or activity is reduced by at least about 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100% relative to the control cell.

BRIEF DESCRIPTION OF THE DRAWINGS

[0071] FIG. 1 illustrates Laminin Subunit Alpha 1 (LAMA1) protein expression in myoblasts transfected with exemplary p300 recruiter (p300R)-DNA binding domain (DBD) peptides and selected guide RNAs (gRNAs) of the present technology.

[0072] FIG. 2 illustrates quantified LAMA1 gene expression in myoblasts transfected with exemplary p300R-DBD peptides and selected guide RNAs gRNAs of the present technology.

[0073] FIGS. 3A and 3B illustrate changes in Insulin-like growth factor 1 (IGF-1) and adiponectin (ADIPOQ) gene expression levels in murine hepatocytes transfected with exemplary p300R-DBD peptides and selected IGF-1 or ADIPOQ-targeting gRNAs of the present technology, compared to controls, respectively (ADIPOQ g3: GACCCAAGCTGGATTAAACC (SEQ ID NO: 68); IGF-1 g8: TTAGCAGACATTAAATACTG (SEQ ID NO: 76)).

[0074] FIG. 4 illustrates changes in IGF-1 and ADIPOQ gene expression levels in murine hepatocytes transfected with exemplary p300R-DBD peptides and selected IGF-1 or ADIPOQ-targeting gRNAs of the present technology, compared to controls (ADIPOQ g3: GACCCAAGCTGGATTAAACC (SEQ ID NO: 68); IGF-1 g8: TTAGCAGACATTAAATACTG (SEQ ID NO: 76); UT: untreated: -g: p300R-DBD transfection without gRNA; Scramble: transfection with scrambled gRNA).

[0075] FIGS. 5A and 5B illustrate changes in IGF-1 gene expression levels across three replicates in murine hepatocytes (FIG. 5A) and murine myoblast cells (FIG. 5B) when transfected with a p300R-DBD peptide of the present technology and an IGF-1-targeting gRNA, compared to controls (LAMA1 g9: GTCCGAAGGGCCTGCGCACC (SEQ ID NO: 64); IGF-1 g1: IGF1 g1: CCAGGGTAAACTAGATCGAA (SEQ ID NO: 69); IGF1 g2: ATCCATGGGGCAGCGTAAAG (SEQ ID NO: 70); IGF1 g5: GTATTCCAAATAATACCAGC (SEQ ID NO: 73); IGF1 g6: GCTCCAGTTTTTAAGAGCGA (SEQ ID NO: 74); IGF1 g7: GTGTGCTAATACACCTCACC (SEQ ID NO: 75)).

[0076] FIG. 6 illustrate changes in IGF-1 gene expression levels in murine hepatocytes and murine myoblast cells transfected with a p300R-DBD peptide of the present technology and an IGF-1-targeting gRNA combination, compared to controls

[0077] FIGS. 7A and 7B illustrate changes in IGF-1 gene expression levels in murine hepatocytes (FIG. 7A) and murine myoblasts (FIG. 7B) when transfected with a p300R-DBD peptide of the present technology and one or two IGF-1-targeting gRNAs, compared to controls.

[0078] FIGS. 8A and 8B illustrate changes in gene expression levels in murine hepatocytes (FIG. 8A) and murine myoblasts cells (FIG. 8B) transfected with a p300R-DBD peptide and a LAMA1 or IGF-1-targeting gRNA of the present technology, compared to controls across two replicates.

[0079] FIG. 9 illustrates changes in LAMA1 gene expression in murine myoblasts transfected with a p300R-DBD peptide and a LAMA1-targeting gRNA of the present technology, compared to controls.

[0080] FIGS. 10A and 10B illustrate day 1 gene expression levels of IGF-1 and LAMA1 in murine hepatocytes (FIG. 10A) and murine myoblasts (FIG. 10B) transfected with a p300R-DBD peptide and an IGF-1 or LAMA1-targeting gRNA of the present technology, compared to controls.

[0081] FIGS. 11A and 11B illustrate gene expression levels of IGF-1 and LAMA1 in murine myoblasts (FIG. 11A) and murine hepatocytes (FIG. 11B) transfected with a p300R-DBD peptide and an IGF-1 or LAMA1-targeting gRNA of the present technology, compared to controls.

[0082] FIG. 12 illustrates LAMA1 gene expression levels at day 1 (d1) and day (3) in murine myoblasts following transfection at different doses of a p300R-DBD peptide and a LAMA1-targeting gRNA of the present technology, compared to controls.

[0083] FIGS. 13A and 13B illustrate LAMA1 and Laminin Subunit Alpha 2 (LAMA2) levels in murine myoblasts transfected with a p300R domain peptide and a LAMA1-targeting gRNA of the present technology, compared to controls.

[0084] FIGS. 14A and 14B illustrate gene expression level changes in murine myoblasts transfected with a p300R-DBD peptide or an alternate DBD peptide (VP and VP-ON) and a LAMA1-targeting gRNA of the present technology, relative to controls. FIG. 16A shows Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) gene expression levels. FIG. 14B shows LAMA1 gene expression in the cells of FIG. 16A

treated with a p300 inhibitor (iP300w), normalized to GAPDH gene expression levels at day 1 (d1) and day 3 (d3).

[0085] FIGS. 15A-15D illustrate additional LAMA1 gene expression measurements for the same cells of FIGS. 14A and 14B.

[0086] FIGS. 16A and 16B illustrate LAMA1 gene expression changes in murine myoblasts transfected with a p300R-DBD peptide and a LAMA1-targeting gRNA of the present technology (g9: GTCCGAAGGGCCTGCGCACC (SEQ ID NO: 64); g10: GGGCGCGGCCAGGCAGGCGG (SEQ ID NO: 65); g11: CCCTGAATGCTCACCCCAAC (SEQ ID NO: 66); g12: GCGCCCAGGTCTGTCCTCCA (SEQ ID NO: 67)).

[0087] FIGS. 17A-17D illustrate LAMA1 gene expression changes in the same cells of FIGS. 14A and 14B treated with 0.1 μM, 1.0 μM, or 10 μM of a histone deacetylase inhibitor (HDACi) (d1: day 1; d3: day 3; d5: day 5).

[0088] FIGS. 18A and 18B illustrate LAMA1 gene expression changes in the same cells of FIGS. 14A and 14B treated with 0.1 μM, 1.0 μM, or 10 μM of a p300 inhibitor (iP300w).

DETAILED DESCRIPTION

[0089] Gene therapy may provide a promising method for correcting genetic disorders by directly modifying or activating expression of genetic sequences. However, the complexity of many diseases, characterized by intricate networks of gene interactions, necessitates a more adaptable therapeutic strategy. Traditional gene editing technologies, such as Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-nuclease-based technologies, may involve enzyme-mediated changes to DNA, which may lead to unintended off-target effects and long-term safety concerns including unwanted genetic edits. These methods often lack the ability to precisely control the duration and extent of gene activation, making it difficult to adapt treatment to the evolving needs of patients. In contrast, the present technology may include methods of controlling gene activation without altering an underlying DNA sequence by promoting euchromatin formation. These methods utilize epigenetic approaches to modulate gene expression, offering a safer and more adaptable alternative. By avoiding enzyme-mediated DNA modifications, the present technology

may reduce the risk of off-target effects with greater control over gene activation, relative to conventional technologies. As such, the present technology may comprise a safer and/or more effective method for gene activation, relative to such conventional technologies.

[0090] While the present technology is capable of being embodied in various forms, the description below of several embodiments is made with the understanding that the present technology is to be considered as an exemplification of the technology and is not intended to limit the present technology to the specific embodiments illustrated. Headings are provided for convenience only and are not to be construed to limit the present technology in any manner. Embodiments illustrated under any heading may be combined with embodiments illustrated under any other heading.

[0091] The use of numerical values in the various quantitative values specified in this application, unless expressly indicated otherwise, are stated as approximations as though the minimum and maximum values within the stated ranges were both preceded by the word "about." It is to be understood, although not always explicitly stated, that all numerical designations are preceded by the term "about." It is to be understood that such range format is used for convenience and brevity and should be understood flexibly to include numerical values explicitly specified as limits of a range, but also to include all individual numerical values or sub-ranges encompassed within that range, as if each numerical value and sub-range is explicitly specified. For example, a ratio in the range of about 1 to about 200 should be understood to include the explicitly recited limits of about 1 and about 200, but also to include individual ratios, such as about 2, about 3, and about 4, and sub-ranges, such as about 10 to about 50, about 20 to about 100, and so forth. It also is to be understood, although not always explicitly stated, that the reagents described herein are merely exemplary and that equivalents of such are known in the art.

[0092] To the extent any materials incorporated by reference herein conflict with the present technology, the present technology controls.

Definitions

[0093] The term “about” means a quantity, level, value, number, frequency, percentage, dimension, size, amount, weight or length that varies by acceptable levels

in the art. Typically, such variation may be as much 10% above and below a reference quantity, level, value, number, frequency, percentage, dimension, size, amount, weight or length and such variation may be influenced by standard applicable measurement practices. When the term "about" is used in conjunction with a numerical range, it modifies that range by extending the boundaries above and below the numerical values set forth.

[0094] "Amino acids" are molecules containing an amine group, a carboxylic acid group, and a side-chain that is specific to each amino acid. The key elements of an amino acid are carbon, hydrogen, oxygen, and nitrogen and have the generic formula H_2N—CHR—COOH, wherein R represents a side chain group. The various α-amino acids differ in the side-chain moiety that is attached to the α-carbon. The "amino acids" of the present technology include the known naturally occurring protein amino acids, which are referred to by both their common three letter abbreviation and single letter abbreviation. As set forth above, the term "amino acid" also includes stereoisomers and modifications of naturally occurring protein amino acids, non-protein amino acids, post-translationally modified amino acids, enzymatically synthesized amino acids, derivatized amino acids, constructs or structures designed to mimic amino acids, and the like.

[0095] Amino acids, including stereoisomers and modifications of naturally occurring amino acids, protein amino acids, non-protein amino acids, post-translationally modified amino acids, enzymatically synthesized amino acids, derivatized amino acids, constructs, or structures designed to mimic amino acids (peptide mimetics), and the like, including all of the foregoing, are sometimes referred to herein as "residues."

[0096] "Non-naturally occurring amino acids," as used herein, may refer to an amino acid that is a non-proteinogenic or unnatural amino acid. This may comprise synthetic or modified forms of amino acids that are not naturally present in biological systems and may be chemically synthesized or comprise a modified version of a naturally occurring amino acid. Nonlimiting examples of non-naturally occurring amino acids include Norleucine, Homocysteine, Beta-Alanine, Azetidine-2-carboxylic acid (Aze), Diaminopimelic acid (DAP), Alpha-Aminoisobutyric acid (Aib), 2-Aminoisobutyric acid (Aib), 4-Thiaproline.

[0097] As used herein, "conservative amino acid substitution" means a given amino acid can be replaced by a residue having similar physiochemical characteristics, e.g., substituting one aliphatic residue for another (such as lie, Val, Leu, or Ala for one another), or substitution of one polar residue for another (such as between Lys and Arg; Glu and Asp; or Gin and Asn). Other such conservative substitutions, e.g., substitutions of entire regions having similar hydrophobicity characteristics, are known. Polypeptides comprising conservative amino acid substitutions can be tested in any one of the assays described herein to confirm that a desired activity, e.g., antigen-binding activity and specificity of a native or reference polypeptide is retained. Amino acids can be grouped according to similarities in the properties of their side chains (in A. L. Lehninger, in Biochemistry, second ed., pp. 73-75, Worth Publishers, New York (1975)): (1) non-polar: Ala (A), Val (V), Leu (L), He (I), Pro (P), Phe (F), Trp (W), Met (M); (2) uncharged polar: Gly (G), Ser (S), Thr (T), Cys (C), Tyr (Y), Asn (N), Gin (Q); (3) acidic: Asp (D), Glu (E); (4) basic: Lys (K), Arg (R), His (H). Alternatively, naturally occurring residues can be divided into groups based on common side-chain properties: (1) hydrophobic: Norleucine, Met, Ala, Val, Leu, lie; (2) neutral hydrophilic: Cys, Ser, Thr, Asn, Gin; (3) acidic: Asp, Glu; (4) basic: His, Lys, Arg; (5) residues that influence chain orientation: Gly, Pro; (6) aromatic: Trp, Tyr, Phe. Non-conservative substitutions will entail exchanging a member of one of these classes for another class. Particular conservative substitutions include, for example; Ala into Gly or into Ser; Arg into Lys; Asn into Gin or into H is; Asp into Glu; Cys into Ser; Gin into Asn; Glu into Asp; Gly into Ala or into Pro; His into Asn or into Gin; He into Leu or into Val; Leu into He or into Val; Lys into Arg, into Gin or into Glu; Met into Leu, into Tyr or into He; Phe into Met, into Leu or into Tyr; Ser into Thr; Thr into Ser; Trp into Tyr; Tyr into Trp; and/or Phe into Val, into He or into Leu.

[0098] The term "expression vector" includes vectors that operatively link a nucleic acid coding region or gene to any control sequences capable of effecting expression of the gene product. "Control sequences" operably linked to the nucleic acid sequences of the disclosure are nucleic acid sequences capable of effecting the expression of the nucleic acid molecules. The control sequences need not be contiguous with the nucleic acid sequences, so long as they function to direct the expression thereof. Thus, for example, intervening untranslated yet transcribed sequences can be present between a promoter sequence and the nucleic acid sequences and the promoter sequence can

still be considered "operably linked" to the coding sequence. Other such control sequences include, but are not limited to, polyadenylation signals, termination signals, and ribosome binding sites. Such expression vectors can be of any type, including but not limited plasmid and viral -based expression vectors. The control sequence used to drive expression of the disclosed nucleic acid sequences in a mammalian system may be constitutive (driven by any of a variety of promoters, including but not limited to, CMV, SV40, RSV, actin, EF) or inducible (driven by any of a number of inducible promoters including, but not limited to, tetracycline, ecdysone, steroid-responsive). The expression vector must be replicable in the host organisms either as an episome or by integration into host chromosomal DNA. In various embodiments, the expression vector may comprise a plasmid, viral-based vector, or any other suitable expression vector.

[0099] The term "subject" refers to a mammalian subject, preferably a human. A "subject in need thereof" may refer to a subject who has been diagnosed with a disease, or is at an elevated risk of developing a disease. The phrases "subject" and "patient" are used interchangeably herein.

[0100] A "therapeutically effective amount" as used herein is an amount that produces a desired effect in a subject for treating a disease. In some embodiments, the therapeutically effective amount is an amount that yields maximum therapeutic effect. In other embodiments, the therapeutically effective amount yields a therapeutic effect that is less than the maximum therapeutic effect. For example, a therapeutically effective amount may be an amount that produces a therapeutic effect while avoiding one or more side effects associated with a dosage that yields maximum therapeutic effect. A therapeutically effective amount for a particular composition will vary based on a variety of factors, including, but not limited to, the characteristics of the therapeutic composition (e.g., activity, pharmacokinetics, pharmacodynamics, and bioavailability), the physiological condition of the subject (e.g., age, body weight, sex, disease type and stage, medical history, general physical condition, responsiveness to a given dosage, and other present medications), the nature of any pharmaceutically acceptable carriers, excipients, and preservatives in the composition, and the route of administration. One skilled in the clinical and pharmacological arts will be able to determine a therapeutically effective amount through routine experimentation, namely by monitoring a subject's response to administration of a composition, a vector, or a pharmaceutical composition containing the same, and adjusting the dosage accordingly. For additional guidance,

see, *e.g.*, Remington, *The Science and Practice of Pharmacy,* 22nd Edition, Pharmaceutical Press, London, 2012, and Goodman & Gilman, *The Pharmacological Basis of Therapeutics*, 12th Edition, McGraw-Hill, New York, NY, 2011, the entire disclosures of which are incorporated by reference herein.

[0101] A "clinically effective amount," "clinically effective concentration," or "clinically effective dose" refers to a concentration or dose of a composition containing the same that is shown to be effective in clinical trials or is predicted to be effective based on early phase or pre-clinical trials. In some embodiments, a "clinically effective amount" is the same as a "therapeutically effective amount." In some embodiments, a "clinically effective amount" is higher or lower than a "therapeutically effective amount." For additional guidance, see Remington: The Science and Practice of Pharmacy, 21st Edition, Univ. of Sciences in Philadelphia (USIP), Lippincott Williams & Wilkins, Philadelphia, PA, 2005, the entire disclosure of which is incorporated by reference herein.

[0102] The term “p300 recruiter” (i.e., “p300R”) refers to a peptide that binds to and/or recruits a p300 protein. This includes the C-terminal p300 binding domain of a Double Homebox 4 (DUX4) protein. The DUX4 domain of p300R may specifically recruit and/or bind at least a portion of a p300 protein and/or may comprise an intrinsically disordered domain. The intrinsically disordered domain may recruit and/or bind to at least a portion of a p300 protein, but may also recruit other non-p300 proteins.

[0103] The term “p300R-DBD” refers to a peptide comprising or consisting of a p300 recruiter domain of DUX4 (p300R domain) that is fused to a DNA binding domain (DBD), forming a fusion peptide. The p300R domain comprises or consists of a C-terminal p300 binding domain of a DUX4 protein, and specifically recruits and binds to p300. The p300R-DBD peptide selectively binds to target loci in DNA, which may be activated after the recruitment of endogenous p300 to the p300R. The p300R-DBD as used herein may also be referred to “engineered peptides,” “non-naturally occurring engineered peptides,” “p300R peptides,” and “peptides.”

[0104] The term "gRNA" (guide RNA) refers to an RNA molecule designed to guide a DNA-binding enzyme to a specific genomic location.

[0105] The term "dCas" refers to a modified version of a Cas nuclease enzyme or at least a portion thereof that is catalytically inactive. The dCas may bind to a target

DNA sequence guided by a gRNA, without introducing a double-strand break or another nuclease activity. The dCas may be fused with other peptide domains for modulating gene expression without altering the DNA sequence.

[0106] "Promoting" of euchromatin as used herein refers to the initiation or increase of euchromatin formation, relative to baseline or a control.

[0107] The term "control" refers to any cell or subject used as a basis for comparison. A control subject includes, but is not limited to, the subject of the present technology at baseline or a subject that has not been administered a composition of the present technology. The control may also comprise a system, composition, formulation, method, or lipid nanoparticle that lacks one or more features of the present technology.

[0108] While the present technology is capable of being embodied in various forms, the description below of several embodiments is made with the understanding that the present technology is to be considered as an exemplification of the technology and is not intended to limit the present technology to the specific embodiments illustrated. Headings are provided for convenience only and are not to be construed to limit the present technology in any manner. Embodiments illustrated under any heading may be combined with embodiments illustrated under any other heading.

Compositions

[0109] The present technology comprises engineered, non-naturally occurring peptides having p300R domains and compositions thereof. The p300R domains may be fused to DNA binding domains (i.e., "p300R-DBD") that selectively recruit p300 proteins, including native p300 proteins (i.e., the endogenous p300 proteins expressed in an engineered cell, a host cell or a subject of the present technology), to the site at which the p300R domain is present. Recruitment of p300 promotes acetylation of proteins (e.g., histone or non-histone proteins) at or proximal to the p300R domain site. This acetylation promotes euchromatin formation, thereby activating and/or increasing expression of genes whose expression may be altered by post translational modifications of the proteins. The euchromatin formation and/or activation or increase in gene expression may occur in less than about 1 second, 15 seconds, 30 seconds, 1 minute, 2 minutes, 5 minutes, 10 minutes, 15 minutes, 30 minutes, 1 hour, 2 hours, 3 hours, 4 hours, 5 hours, 6 hours, 12 hours, or 1 day.

[0110] Rather than expressing an exogenous p300 construct to acetylate histones, the p300R peptides of the present technology recruit native p300 to the p300R domain. Without intending to be bound by any particular theory, the p300R peptides of the present technology may reduce the risk of off-target epigenetic or gene expression changes that are otherwise observed with exogeneous p300 expression thereby circumventing use of an overexpression system to deliver exogenous p300. In this way, the p300R peptides of the present technology may reduce toxicity and off-target p300 acetylation activity due to the increase in p300 expression, relative to baseline.

[0111] In some embodiments, the p300R peptides are fused to another peptide comprising a DNA binding domain (DBD), referred to herein as "p300R-DBD proteins". The DBD selectively binds specific regions of the genome, guiding the p300R domains to these regions. This further reduces off-target acetylation events that are observed with the expression of exogenous p300 constructs, such as the acetylation of non-histone proteins.

[0112] Nonlimiting examples of p300R domains and DBD domains of p300R-DBD proteins are further detailed under the "*p300R Domains*" and "*DNA Binding Domains*" headers.

[0113] The present technology further includes compositions and pharmaceutical compositions comprising a p300R-DBD protein, a vector encoding a p300R-DBD protein and, optionally, a gRNA, a host cell comprising the vector, the composition, and/or the pharmaceutical composition, and associated methods of making and using the same.

[0114] The peptides present technology (e.g., p300R-DBD proteins) may comprise one or more features of the peptides disclosed in WO2020185850A1, such as the DBD peptides disclosed therein, incorporated by reference in its entirety herein.

p300R Domains

[0115] The p300R domains of the present technology specifically bind to and/or recruit p300 proteins. The p300R domains may be comprised of an amino acid sequence of a p300 binding domain. Nonlimiting examples of proteins having p300 binding domains useful with the present technology include DUX4 (NCBI Accession(s): AUA60624.1, AUA60623.1, AUA60622.1, AUA60621.1, AUA60620.1, AUA60367.1),

NUT (NCBI Accession(s): NP_001271221.2, NP_786883.2, NP_001271222.2), p53 (NCBI Accession: AAC12971.1), SP1 (NCBI Accession(s): P08047.3, CBM42955.1, NP_612482.2, NP_003100.1, NP_001238754.1), MEF2 (NCBI Accession(s): Q02078.1, AAB25838.1, CAA48517.1), CITED2 (NCBI Accession(s): NP_006070.2, NP_001161860.1, NP_001161861.2, Q99967.2), RelA (NCBI Accession(s): NP_068810.3, NP_001138610.1, NP_001230913.1, NP_001230914.1, NP_001391586.1), and HIF1α (NCBI Accession(s): Q16665.1, NP_001230013.1, NP_851397.1, NP_001521.1).

[0116] In some embodiments, the p300R domains comprise or consist of at least a portion of the C-terminal domain of the DUX4 p300 binding domain. In some embodiments, the p300R domains comprise or consist of at least a portion of the p300 binding or p300 recruitment domains of a protein selected from the group consisting of NUT, p53, SP1, MEF2, CITED2, RelA, and HIF1α.

[0117] In some embodiments, the p300R domains of the present technology may comprise or consist of the amino acid sequence MSARQGQMQGIPAPSQALQEPAPWSALPCGLLLDELLASPEFLQQAQPLLETEAPG ELEASEEAASLEAPLSEEEYRALLEEL (SEQ ID NO: 1).

[0118] In some embodiments, the p300R domain comprises or consists of an amino acid sequence about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence of SEQ ID NO: 1.

[0119] In some embodiments, the p300R domain comprises or consists of an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence of SEQ ID NO: 1.

[0120] In some embodiments, the p300R domain comprises or consists of an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence of SEQ ID NO: 1.

[0121] Any p300R domain amino acid sequence of the present technology may comprise one or more substitution variants. Substitution variants involve the

replacement of one or more amino acids with the same number of amino acids and making conservative amino acid substitutions. For example, an amino acid may be substituted with an alternative amino acid having similar properties, for example, another basic amino acid, another acidic amino acid, another neutral amino acid, another charged amino acid, another hydrophilic amino acid, another hydrophobic amino acid, another polar amino acid, another aromatic amino acid or another aliphatic amino acid. Such substitutions may be conservative amino acid substitutions.

[0122] Any p300R domain amino acid sequence of the present technology may be modified to comprise one or more non-naturally occurring amino acids. In some embodiments, the p300R domains comprise or consist of an amino acid sequence having one or more non-naturally occurring amino acids. In some embodiments, the p300R domains comprise or consist of an amino acid sequence having two or more non-naturally occurring amino acids. In some embodiments, the modification of the amino acid sequences to comprise non-naturally occurring amino acids is a conservative substitution.

[0123] The p300R domains may be shorter than the p300R domain amino acid sequences of the present technology that described above. For example, the p300R domain may comprise an amino acid sequence about 1%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, or 80% shorter or reduced by about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, or 75 amino acids relative to the p300R domain amino acid sequences described above.

DNA Binding Domains

[0124] The p300R domains of the present technology may be fused to a DBD, forming the p300R-DBD protein (i.e., a fusion peptide). The DBDs bind specific genomic regions, thereby guiding the p300R domain to a precise site. DBDs may comprise a clustered regularly interspaced short palindromic repeats associated protein (Cas) domain. Non-limiting examples of Cas domains include a Cas1 domain, a Cas2 domain, a Cas3 domain, Cas9 domain, a Cas10 domain, a Cas12a domain (e.g., ErCas12a domain), or a Cas13 domain. Other DBDs may comprise a zinc finger domain, a helix-turn-helix (HTH) domain, a leucine zipper domain, and a basic helix-

loop-helix (bHLH) domain. In some embodiments, the Cas domain comprises a responsive portion which interacts with a gRNA that selectively binds to the locus.

[0125] The Cas domain may comprise an inactive nuclease domain. In some embodiments, the inactive nuclease domain is a dead Cas (dCas) domain.

[0126] In some embodiments, the Cas domain comprises or consists of an amino acid sequence about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 100% identical to the amino acid sequence of any one of SEQ ID NOS: 34 or 40-57 (Table 1).

[0127] In some embodiments, the Cas domain comprises or consists of an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 100% identical to the amino acid sequence of any one of SEQ ID NOS: 34 or 40-57.

[0128] In some embodiments, the Cas domain comprises or consists of an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 100% identical to the amino acid sequence of any one of SEQ ID NOS: 34 or 40-57.

Table 1: Cas domain amino acid sequences

Name	Amino Acid Sequence	SEQ ID NO
Cas9	MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKK NLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMA KVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDK LFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDL DNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGG ASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGN SRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLP NEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAI VDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT	34

Name	Amino Acid Sequence	SEQ ID NO
	YHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSD GFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSP AIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKN SRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRD MYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGK SDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVI TLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKY PKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFF KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQ VNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDS PTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFL EAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQI SEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGG D	
Streptococcus thermophiles Cas 9	MTKPYSIGLDIGTNSVGWAVTTDNYKVPSKKMKVLGNTSKKYIKK NLLGVLLFDSGITAEGRRLKRTARRRYTRRRNRILYLQEIFSTEMA TLDDAFFQRLDDSFLVPDDKRDSKYPIFGNLVEEKAYHDEFPTIY HLRKYLADSTKKADLRLVYLALAHMIKYRGHFLIEGEFNSKNNDIQ KNFQDFLDTYNAIFESDLSLENSKQLEEIVKDKISKLEKKDRILKLF PGEKNSGIFSEFLKLIVGNQADFRKCFNLDEKASLHFSKESYDED LETLLGYIGDDYSDVFLKAKKLYDAILLSGFLTVTDNETEAPLSSA MIKRYNEHKEDLALLKEYIRNISLKTYNEVFKDDTKNGYAGYIDGK TNQEDFYVYLKKLLAEFEGADYFLEKIDREDFLRKQRTFDNGSIP YQIHLQEMRAILDKQAKFYPFLAKNKERIEKILTFRIPYYVGPLARG NSDFAWSIRKRNEKITPWNFEDVIDKESSAEAFINRMTSFDLYLP EEKVLPKHSLLYETFNVYNELTKVRFIAESMRDYQFLDSKQKKDI VRLYFKDKRKVTDKDIIEYLHAIYGYDGIELKGIEKQFNSSLSTYHD	40

Name	Amino Acid Sequence	SEQ ID NO
	LLNIINDKEFLDDSSNEAIIEEIIHTLTIFEDREMIKQRLSKFENIFDK SVLKKLSRRHYTGWGKLSAKLINGIRDEKSGNTILDYLIDDGISNR NFMQLIHDDALSFKKKIQKAQIIGDEDKGNIKEVVKSLPGSPAIKK GILQSIKIVDELVKVMGGRKPESIVVEMARENQYTNQGKSNSQQ RLKRLEKSLKELGSKILKENIPAKLSKIDNNALQNDRLYLYYLQNG KDMYTGDDLDIDRLSNYDIDHIIPQAFLKDNSIDNKVLVSSASNRG KSDDVPSLEVVKKRKTFWYQLLKSKLISQRKFDNLTKAERGGLSP EDKAGFIQRQLVETRQITKHVARLLDEKFNNKKDENNRAVRTVKII TLKSTLVSQFRKDFELYKVREINDFHHAHDAYLNAVVASALLKKY PKLEPEFVYGDYPKYNSFRERKSATEKVYFYSNIMNIFKKSISLAD GRVIERPLIEVNEETGESVWNKESDLATVRRVLSYPQVNVVKKVE EQNHGLDRGKPKGLFNANLSSKPKPNSNENLVGAKEYLDPKKY GGYAGISNSFTVLVKGTIEKGAKKKITNVLEFQGISILDRINYRKDK LNFLLEKGYKDIELIIELPKYSLFELSDGSRRMLASILSTNNKRGEI HKGNQIFLSQKFVKLLYHAKRISNTINENHRKYVENHKKEFEELFY YILEFNENYVGAKKNGKLLNSAFQSWQNHSIDELCSSFIGPTGSE RKGLFELTSRGSAADFEFLGVKIPRYRDYTPSSLLKDATLIHQSVT GLYETRIDLAKLGEG	
A0A1L8RGR8 Enterococcus canis	MKQNKELVNIGFDIGIASVGWSVVSKQSGKILETGVSIFPSGTASK NEERRSFRQARRLLRRRKNRISDLKILLEENGFRIAKLNQLVTPYE LRVRGLNEQLSKEELSVALLHLVKRRGISYSLEDSEGEGDNQTSY KQSVSINQKLLKEKTPGEIQLERLEKYGKIRGQVKDLQEENAAVL MNVFPNTAYVREAELILLKQKEYYSEITDNFIKEATALISRKREYFV GPGSEKSRTDYGIYRTDGTKLDNLFEILIGKDKIFPNEFRAAGNSY TAQLYNLLNDLNNLKIKTLEDGKLTKDQKLSIIEELKTTTKKVNMM QLIKKIAKAEESDISGYRIDRNDKPEIHSMAIFYKVRKKFLEQEIDIN DWPIDFLDILGRVLTLNTENGEIRRSLTELKKDYIFLDETLIELIINSK DSFKLTSNQKWHRFSLKTMQLLIPELLNSSKEQMTILTELGLLHE NKQDYSNKTKIDVKNLTENIYNPVVRKSVKQAMDIFNSLFKKYPNI AYLVVEMPRDEAEDEVEQKKQAQKFQKENEAEKEKSLKEFQELA GVSDSQLENQIYKRRKLRMKIRLWYQQLGKCPYSGKTIAAEDLF	41

Name	Amino Acid Sequence	SEQ ID NO
	WTDHLFEIDHVIPLSISYDDGQNNKVLCYSEMNQEKGQKTPYGF MQSGKGQGFSALQAMLKSNSRMSGAKKRNLLFTEDINDIEVRKR FIARNLVDTRYASRIVLNELQQFTRSKQLDTKVTVIRGKFTSKLRE TWRLNKSRETHHHHAVDATIIAVSPMLKLWERNAEIIPMKVNENV VDIKTGEILTDKVYQEEMYQLPYASLLEDIAVMENKIKFHHQVDKK MNRKVSDATIYATRSAKVGKDKEPQNYVLGKIKDIYDTKEYENFK KIYDKDKSKFLMQQLDPMTFEKLEKVLKEYPDFEEVQQDNGRVK RIPISPFELYRREKGPITKFAKRNNGPAIKSVKYYDSKMGSAIDITP QTAKNKKVVLQSLKPWRTDVYFNQETKEYEIMGIKYSDMQYLNG NYGITNERYKEIQREEGVADNSEFMMSLYRGDRIKVIDTNSDESV ELLFGSRTIPTKKGYVELKPIEKTKFDSKEIVSFYGQVTPNGQFVK KFTRNGYRLLKVNTNILGNPYYISKEGINPRNILDTGFKG	
J7RUA5 Staphylococcus aureus Cas9	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEG RRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEA RVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTK EQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEA KQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWK DIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITR DENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVT STGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQ EELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTND NQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKV INAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEI IRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEV DHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKIS YETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNL VDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKF KKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEE KQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNR ELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKL LMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKK	42

Name	Amino Acid Sequence	SEQ ID NO
	DNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDV YLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAE FIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENM NDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	
G3ECR1 Streptococcus thermophilusICas9	MLFNKCIIISINLDFSNKEKCMTKPYSIGLDIGTNSVGWAVITDNYK VPSKKMKVLGNTSKKYIKKNLLGVLLFDSGITAEGRRLKRTARRR YTRRRNRILYLQEIFSTEMATLDDAFFQRLDDSFLVPDDKRDSKY PIFGNLVEEKVYHDEFPTIYHLRKYLADSTKKADLRLVYLALAHMI KYRGHFLIEGEFNSKNNDIQKNFQDFLDTYNAIFESDLSLENSKQL EEIVKDKISKLEKKDRILKLFPGEKNSGIFSEFLKLIVGNQADFRKC FNLDEKASLHFSKESYDEDLETLLGYIGDDYSDVFLKAKKLYDAIL LSGFLTVTDNETEAPLSSAMIKRYNEHKEDLALLKEYIRNISLKTY NEVFKDDTKNGYAGYIDGKTNQEDFYVYLKNLLAEFEGADYFLE KIDREDFLRKQRTFDNGSIPYQIHLQEMRAILDKQAKFYPFLAKNK ERIEKILTFRIPYYVGPLARGNSDFAWSIRKRNEKITPWNFEDVID KESSAEAFINRMTSFDLYLPEEKVLPKHSLLYETFNVYNELTKVRF IAESMRDYQFLDSKQKKDIVRLYFKDKRKVTDKDIIEYLHAIYGYD GIELKGIEKQFNSSLSTYHDLLNIINDKEFLDDSSNEAIIEEIIHTLTIF EDREMIKQRLSKFENIFDKSVLKKLSRRHYTGWGKLSAKLINGIR DEKSGNTILDYLIDDGISNRNFMQLIHDDALSFKKKIQKAQIIGDED KGNIKEVVKSLPGSPAIKKGILQSIKIVDELVKVMGGRKPESIVVEM ARENQYTNQGKSNSQQRLKRLEKSLKELGSKILKENIPAKLSKID NNALQNDRLYLYYLQNGKDMYTGDDLDIDRLSNYDIDHIIPQAFLK DNSIDNKVLVSSASNRGKSDDFPSLEVVKKRKTFWYQLLKSKLIS QRKFDNLTKAERGGLLPEDKAGFIQRQLVETRQITKHVARLLDEK FNNKKDENNRAVRTVKIITLKSTLVSQFRKDFELYKVREINDFHHA HDAYLNAVIASALLKKYPKLEPEFVYGDYPKYNSFRERKSATEKV YFYSNIMNIFKKSISLADGRVIERPLIEVNEETGESVWNKESDLAT VRRVLSYPQVNVVKKVEEQNHGLDRGKPKGLFNANLSSKPKPN SNENLVGAKEYLDPKKYGGYAGISNSFAVLVKGTIEKGAKKKITN VLEFQGISILDRINYRKDKLNFLLEKGYKDIELIIELPKYSLFELSDG	43

Name	Amino Acid Sequence	SEQ ID NO
	SRRMLASILSTNNKRGEIHKGNQIFLSQKFVKLLYHAKRISNTINE NHRKYVENHKKEFEELFYYILEFNENYVGAKKNGKLLNSAFQSW QNHSIDELCSSFIGPTGSERKGLFELTSRGSAADFEFLGVKIPRY RDYTPSSLLKDATLIHQSVTGLYETRIDLAKLGEG	
Q99ZW2 Streptococcus pyogenes Cas 9	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKK NLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMA KVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDK LFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDL DNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGG ASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGN SRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLP NEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAI VDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT YHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSD GFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSP AIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKN SRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRD MYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGK SDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVI TLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKY PKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFF KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQ VNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDS PTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFL EAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA	44

Name	Amino Acid Sequence	SEQ ID NO
	LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQI SEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGG D	
C9X1G5 Neisseria meningitidis Cas9	MAAFKPNSINYILGLDIGIASVGWAMVEIDEEENPIRLIDLGVRVFE RAEVPKTGDSLAMARRLARSVRRLTRRRAHRLLRTRRLLKREGV LQAANFDENGLIKSLPNTPWQLRAAALDRKLTPLEWSAVLLHLIK HRGYLSQRKNEGETADKELGALLKGVAGNAHALQTGDFRTPAEL ALNKFEKESGHIRNQRSDYSHTFSRKDLQAELILLFEKQKEFGNP HVSGGLKEGIETLLMTQRPALSGDAVQKMLGHCTFEPAEPKAAK NTYTAERFIWLTKLNNLRILEQGSERPLTDTERATLMDEPYRKSK LTYAQARKLLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRA LEKEGLKDKKSPLNLSPELQDEIGTAFSLFKTDEDITGRLKDRIQP EILEALLKHISFDKFVQISLKALRRIVPLMEQGKRYDEACAEIYGDH YGKKNTEEKIYLPPIPADEIRNPVVLRALSQARKVINGVVRRYGSP ARIHIETAREVGKSFKDRKEIEKRQEENRKDREKAAAKFREYFPN FVGEPKSKDILKLRLYEQQHGKCLYSGKEINLGRLNEKGYVEIDH ALPFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSRE WQEFKARVETSRFPRSKKQRILLQKFDEDGFKERNLNDTRYVNR FLCQFVADRMRLTGKGKKRVFASNGQITNLLRGFWGLRKVRAE NDRHHALDAVVVACSTVAMQQKITRFVRYKEMNAFDGKTIDKET GEVLHQKTHFPQPWEFFAQEVMIRVFGKPDGKPEFEEADTLEKL RTLLAEKLSSRPEAVHEYVTPLFVSRAPNRKMSGQGHMETVKSA KRLDEGVSVLRVPLTQLKLKDLEKMVNREREPKLYEALKARLEAH KDDPAKAFAEPFYKYDKAGNRTQQVKAVRVEQVQKTGVWVRN HNGIADNATMVRVDVFEKGDKYYLVPIYSWQVAKGILPDRAVVQ GKDEEDWQLIDDSFNFKFSLHPNDLVEVITKKARMFGYFASCHR GTGNINIRIHDLDHKIGKNGILEGIGVKTALSFQKYQIDELGKEIRP CRLKKRPPVR	45
A0Q5Y3	MNFKILPIAIDLGVKNTGVFSAFYQKGTSLERLDNKNGKVYELSKD SYTLLMNNRTARRHQRRGIDRKQLVKRLFKLIWTEQLNLEWDKD	46

Name	Amino Acid Sequence	SEQ ID NO
Francisella novicida Cas9	TQQAISFLFNRRGFSFITDGYSPEYLNIVPEQVKAILMDIFDDYNG EDDLDSYLKLATEQESKISEIYNKLMQKILEFKLMKLCTDIKDDKVS TKTLKEITSYEFELLADYLANYSESLKTQKFSYTDKQGNLKELSYY HHDKYNIQEFLKRHATINDRILDTLLTDDLDIWNFNFEKFDFDKNE EKLQNQEDKDHIQAHLHHFVFAVNKIKSEMASGGRHRSQYFQEI TNVLDENNHQEGYLKNFCENLHNKKYSNLSVKNLVNLIGNLSNLE LKPLRKYFNDKIHAKADHWDEQKFTETYCHWILGEWRVGVKDQ DKKDGAKYSYKDLCNELKQKVTKAGLVDFLLELDPCRTIPPYLDN NNRKPPKCQSLILNPKFLDNQYPNWQQYLQELKKLQSIQNYLDS FETDLKVLKSSKDQPYFVEYKSSNQQIASGQRDYKDLDARILQFI FDRVKASDELLLNEIYFQAKKLKQKASSELEKLESSKKLDEVIANS QLSQILKSQHTNGIFEQGTFLHLVCKYYKQRQRARDSRLYIMPEY RYDKKLHKYNNTGRFDDDNQLLTYCNHKPRQKRYQLLNDLAGV LQVSPNFLKDKIGSDDDLFISKWLVEHIRGFKKACEDSLKIQKDNR GLLNHKINIARNTKGKCEKEIFNLICKIEGSEDKKGNYKHGLAYEL GVLLFGEPNEASKPEFDRKIKKFNSIYSFAQIQQIAFAERKGNANT CAVCSADNAHRMQQIKITEPVEDNKDKIILSAKAQRLPAIPTRIVD GAVKKMATILAKNIVDDNWQNIKQVLSAKHQLHIPIITESNAFEFEP ALADVKGKSLKDRRKKALERISPENIFKDKNNRIKEFAKGISAYSG ANLTDGDFDGAKEELDHIIPRSHKKYGTLNDEANLICVTRGDNKN KGNRIFCLRDLADNYKLKQFETTDDLEIEKKIADTIWDANKKDFKF GNYRSFINLTPQEQKAFRHALFLADENPIKQAVIRAINNRNRTFVN GTQRYFAEVLANNIYLRAKKENLNTDKISFDYFGIPTIGNGRGIAEI RQLYEKVDSDIQAYAKGDKPQASYSHLIDAMLAFCIAADEHRNDG SIGLEIDKNYSLYPLDKNTGEVFTKDIFSQIKITDNEFSDKKLVRKK AIEGFNTHRQMTRDGIYAENYLPILIHKELNEVRKGYTWKNSEEIK IFKGKKYDIQQLNNLVYCLKFVDKPISIDIQISTLEELRNILTTNNIAA TAEYYYINLKTQKLHEYYIENYNTALGYKKYSKEMEFLRSLAYRSE RVKIKSIDDVKQVLDKDSNFIIGKITLPFKKEWQRLYREWQNTTIK DDYEFLKSFFNVKSITKLHKKVRKDFSLPISTNEGKFLVKRKTWD NNFIYQILNDSDSRADGTKPFIPAFDISKNEIVEAIIDSFTSKNIFWL	

Name	Amino Acid Sequence	SEQ ID NO
	PKNIELQKVDNKNIFAIDTSKWFEVETPSDLRDIGIATIQYKIDNNS RPKVRVKLDYVIDDDSKINYFMNHSLLKSRYPDKVLEILKQSTIIEF ESSGFNKTIKEMLGMKLAGIYNETSNN	
Q73QW6 Treponema denticola Cas9	MKKEIKDYFLGLDVGTGSVGWAVTDTDYKLLKANRKDLWGMRC FETAETAEVRRLHRGARRRIERRKKRIKLLQELFSQEIAKTDEGFF QRMKESPFYAEDKTILQENTLFNDKDFADKTYHKAYPTINHLIKA WIENKVKPDPRLLYLACHNIIKKRGHFLFEGDFDSENQFDTSIQAL FEYLREDMEVDIDADSQKVKEILKDSSLKNSEKQSRLNKILGLKPS DKQKKAITNLISGNKINFADLYDNPDLKDAEKNSISFSKDDFDALS DDLASILGDSFELLLKAKAVYNCSVLSKVIGDEQYLSFAKVKIYEK HKTDLTKLKNVIKKHFPKDYKKVFGYNKNEKNNNNYSGYVGVCK TKSKKLIINNSVNQEDFYKFLKTILSAKSEIKEVNDILTEIETGTFLP KQISKSNAEIPYQLRKMELEKILSNAEKHFSFLKQKDEKGLSHSEK IIMLLTFKIPYYIGPINDNHKKFFPDRCWVVKKEKSPSGKTTPWNF FDHIDKEKTAEAFITSRTNFCTYLVGESVLPKSSLLYSEYTVLNEIN NLQIIIDGKNICDIKLKQKIYEDLFKKYKKITQKQISTFIKHEGICNKT DEVIILGIDKECTSSLKSYIELKNIFGKQVDEISTKNMLEEIIRWATIY DEGEGKTILKTKIKAEYGKYCSDEQIKKILNLKFSGWGRLSRKFLE TVTSEMPGFSEPVNIITAMRETQNNLMELLSSEFTFTENIKKINSG FEDAEKQFSYDGLVKPLFLSPSVKKMLWQTLKLVKEISHITQAPP KKIFIEMAKGAELEPARTKTRLKILQDLYNNCKNDADAFSSEIKDL SGKIENEDNLRLRSDKLYLYYTQLGKCMYCGKPIEIGHVFDTSNY DIDHIYPQSKIKDDSISNRVLVCSSCNKNKEDKYPLKSEIQSKQRG FWNFLQRNNFISLEKLNRLTRATPISDDETAKFIARQLVETRQATK VAAKVLEKMFPETKIVYSKAETVSMFRNKFDIVKCREINDFHHAH DAYLNIVVGNVYNTKFTNNPWNFIKEKRDNPKIADTYNYYKVFDY DVKRNNITAWEKGKTIITVKDMLKRNTPIYTRQAACKKGELFNQTI MKKGLGQHPLKKEGPFSNISKYGGYNKVSAAYYTLIEYEEKGNKI RSLETIPLYLVKDIQKDQDVLKSYLTDLLGKKEFKILVPKIKINSLLKI NGFPCHITGKTNDSFLLRPAVQFCCSNNEVLYFKKIIRFSEIRSQR EKIGKTISPYEDLSFRSYIKENLWKKTKNDEIGEKEFYDLLQKKNL	47

Name	Amino Acid Sequence	SEQ ID NO
	EIYDMLLTKHKDTIYKKRPNSATIDILVKGKEKFKSLIIENQFEVILEI LKLFSATRNVSDLQHIGGSKYSGVAKIGNKISSLDNCILIYQSITGIF EKRIDLLKV	
U2UMQ6 Acidaminococcus C&sl2& (Cpfl)	MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARND HYKELKPIIDRIYKTYADQCLQLVQLDWENLSAAIDSYRKEKTEET RNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAELF NGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSAE DISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGI FVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLN EVLNLAIQKNDETAHIIASLPHRFIPLFKQILSDRNTLSFILEEFKSD EEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHIFISHKKLETI SSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINL QEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTLKKQEEKEIL KSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLS FYNKARNYATKKPYSVEKFKLNFQMPTLASGWDVNKEKNNGAIL FVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMYYDYFPD AAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNN PEKEPKKFQTAYAKKTGDQKGYREALCKWIDFTRDFLSKYTKTT SIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVET GKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLN GQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQKTPIPDTLYQEL YDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFFFH VPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVI DSTGKILEQRSLNTIQQFDYQKKLDNREKERVAARQAWSVVGTIK DLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKRTGIAEKAV YQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKM GTQSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEG FDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNE TQFDAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGI VFRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNAATGEDYI NSPVRDLNGVCFDSRFQNPEWPM	48

Name	Amino Acid Sequence	SEQ ID NO
A0A182D WE3 Lachnospiraceae Casl2a (Cpfl)	AASKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRAE DYKGVKKLLDRYYLSFINDVLHSIKLKNLNNYISLFRKKTRTEKEN KELENLEINLRKEIAKAFKGAAGYKSLFKKDIIETILPEAADDKDEIA LVNSFNGFTTAFTGFFDNRENMFSEEAKSTSIAFRCINENLTRYIS NMDIFEKVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQ EGIDVYNAIIGGFVTESGEKIKGLNEYINLYNAKTKQALPKFKPLYK QVLSDRESLSFYGEGYTSDEEVLEVFRNTLNKNSEIFSSIKKLEKL FKNFDEYSSAGIFVKNGPAISTISKDIFGEWNLIRDKWNAEYDDIH LKKKAVVTEKYEDDRRKSFKKIGSFSLEQLQEYADADLSVVEKLK EIIIQKVDEIYKVYGSSEKLFDADFVLEKSLKKNDAVVAIMKDLLDS VKSFENYIKAFFGEGKETNRDESFYGDFVLAYDILLKVDHIYDAIR NYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGS KYYLAIMDKKYAKCLQKIDKDDVNGNYEKINYKLLPGPNKMLPKV FFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNLNDCHKLIDFFKD SISRYPKWSNAYDFNFSETEKYKDIAGFYREVEEQGYKVSFESA SKKEVDKLVEEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDE NNHGQIRLSGGAELFMRRASLKKEELVVHPANSPIANKNPDNPK KTTTLSYDVYKDKRFSEDQYELHIPIAINKCPKNIFKINTEVRVLLK HDDNPYVIGIDRGERNLLYIVVVDGKGNIVEQYSLNEIINNFNGIRI KTDYHSLLDKKEKERFEARQNWTSIENIKELKAGYISQVVHKICEL VEKYDAVIALEDLNSGFKNSRVKVEKQVYQKFEKMLIDKLNYMVD KKSNPCATGGALKGYQITNKFESFKSMSTQNGFIFYIPAWLTSKID PSTGFVNLLKTKYTSIADSKKFISSFDRIMYVPEEDLFEFALDYKN FSRTDADYIKKWKLYSYGNRIRIFAAAKKNNVFAWEEVCLTSAYK ELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSI TGRTDVDFLISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNI ARKVLWAIGQFKKAEDEKLDKVKIAISNKEWLEYAQTSVK	49
C7NBY4 Leptotrichia buccalis	MKVTKVGGISHKKYTSEGRLVKSESEENRTDERLSALLNMRLDM YIKNPSSTETKENQKRIGKLKKFFSNKMVYLKDNTLSLKNGKKENI DREYSETDILESDVRDKKNFAVLKKIYLNENVNSEELEVFRNDIKK KLNKINSLKYSFEKNKANYQKINENNIEKVEGKSKRNIIYDYYRES	50

Name	Amino Acid Sequence	SEQ ID NO
Casl3a (C2c2)	AKRDAYVSNVKEAFDKLYKEEDIAKLVLEIENLTKLEKYKIREFYH EIIGRKNDKENFAKIIYEEIQNVNNMKELIEKVPDMSELKKSQVFYK YYLDKEELNDKNIKYAFCHFVEIEMSQLLKNYVYKRLSNISNDKIK RIFEYQNLKKLIENKLLNKLDTYVRNCGKYNYYLQDGEIATSDFIA RNRQNEAFLRNIIGVSSVAYFSLRNILETENENDITGRMRGKTVK NNKGEEKYVSGEVDKIYNENKKNEVKENLKMFYSYDFNMDNKN EIEDFFANIDEAISSIRHGIVHFNLELEGKDIFAFKNIAPSEISKKMF QNEINEKKLKLKIFRQLNSANVFRYLEKYKILNYLKRTRFEFVNKNI PFVPSFTKLYSRIDDLKNSLGIYWKTPKTNDDNKTKEIIDAQIYLLK NIYYGEFLNYFMSNNGNFFEISKEIIELNKNDKRNLKTGFYKLQKF EDIQEKIPKEYLANIQSLYMINAGNQDEEEKDTYIDFIQKIFLKGFM TYLANNGRLSLIYIGSDEETNTSLAEKKQEFDKFLKKYEQNNNIKI PYEINEFLREIKLGNILKYTERLNMFYLILKLLNHKELTNLKGSLEKY QSANKEEAFSDQLELINLLNLDNNRVTEDFELEADEIGKFLDFNG NKVKDNKELKKFDTNKIYFDGENIIKHRAFYNIKKYGMLNLLEKIAD KAGYKISIEELKKYSNKKNEIEKNHKMQENLHRKYARPRKDEKFT DEDYESYKQAIENIEEYTHLKNKVEFNELNLLQGLLLRILHRLVGY TSIWERDLRFRLKGEFPENQYIEEIFNFENKKNVKYKGGQIVEKYI KFYKELHQNDEVKINKYSSANIKVLKQEKKDLYIRNYIAHFNYIPHA EISLLEVLENLRKLLSYDRKLKNAVMKSVVDILKEYGFVATFKIGA DKKIGIQTLESEKIVHLKNLKKKKLMTDRNSEELCKLVKIMFEYKM EEKKSEN	51
PODOC6 Leptotrichia shahii Casl3a (C2c2)	MGNLFGHKRWYEVRDKKDFKIKRKVKVKRNYDGNKYILNINENN NKEKIDNNKFIRKYINYKKNDNILKEFTRKFHAGNILFKLKGKEGIIR IENNDDFLETEEVVLYIEAYGKSEKLKALGITKKKIIDEAIRQGITKD DKKIEIKRQENEEEIEIDIRDEYTNKTLNDCSIILRIIENDELETKKSIY EIFKNINMSLYKIIEKIIENETEKVFENRYYEEHLREKLLKDDKIDVIL TNFMEIREKIKSNLEILGFVKFYLNVGGDKKKSKNKKMLVEKILNIN VDLTVEDIADFVIKELEFWNITKRIEKVKKVNNEFLEKRRNRTYIKS YVLLDKHEKFKIERENKKDKIVKFFVENIKNNSIKEKIEKILAEFKID ELIKKLEKELKKGNCDTEIFGIFKKHYKVNFDSKKFSKKSDEEKEL	51

Name	Amino Acid Sequence	SEQ ID NO
	YKIIYRYLKGRIEKILVNEQKVRLKKMEKIEIEKILNESILSEKILKRV KQYTLEHIMYLGKLRHNDIDMTTVNTDDFSRLHAKEELDLELITFF ASTNMELNKIFSRENINNDENIDFFGGDREKNYVLDKKILNSKIKII RDLDFIDNKNNITNNFIRKFTKIGTNERNRILHAISKERDLQGTQDD YNKVINIIQNLKISDEEVSKALNLDVVFKDKKNIITKINDIKISEENNN DIKYLPSFSKVLPEILNLYRNNPKNEPFDTIETEKIVLNALIYVNKEL YKKLILEDDLEENESKNIFLQELKKTLGNIDEIDENIIENYYKNAQIS ASKGNNKAIKKYQKKVIECYIGYLRKNYEELFDFSDFKMNIQEIKK QIKDINDNKTYERITVKTSDKTIVINDDFEYIISIFALLNSNAVINKIRN RFFATSVWLNTSEYQNIIDILDEIMQLNTLRNECITENWNLNLEEFI QKMKEIEKDFDDFKIQTKKEIFNNYYEDIKNNILTEFKDDINGCDVL EKKLEKIVIFDDETKFEIDKKSNILQDEQRKLSNINKKDLKKKVDQY IKDKDQEIKSKILCRIIFNSDFLKKYKKEIDNLIEDMESENENKFQEI YYPKERKNELYIYKKNLFLNIGNPNFDKIYGLISNDIKMADAKFLFN IDGKNIRKNKISEIDAILKNLNDKLNGYSKEYKEKYIKKLKENDDFF AKNIQNKNYKSFEKDYNRVSEYKKIRDLVEFNYLNKIESYLIDINW KLAIQMARFERDMHYIVNGLRELGIIKLSGYNTGISRAYPKRNGSD GFYTTTAYYKFFDEESYKKFEKICYGFGIDLSENSEINKPENESIR NYISHFYIVRNPFADYSIAEQIDRVSNLLSYSTRYNNSTYASVFEV FKKDVNLDYDELKKKFKLIGNNDILERLMKPKKVSVLELESYNSDY IKNLIIELLTKIENTNDTL	
Q0P897 Campylobacter jejuni Cas9	MARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTGESLAL PRRLARSARKRLARRKARLNHLKHLIANEFKLNYEDYQSFDESLA KAYKGSLISPYELRFRALNELLSKQDFARVILHIAKRRGYDDIKNS DDKEKGAILKAIKQNEEKLANYQSVGEYLYKEYFQKFKENSKEFT NVRNKKESYERCIAQSFLKDELKLIFKKQREFGFSFSKKFEEEVLS VAFYKRALKDFSHLVGNCSFFTDEKRAPKNSPLAFMFVALTRIINL LNNLKNTEGILYTKDDLNALLNEVLKNGTLTYKQTKKLLGLSDDYE FKGEKGTYFIEFKKYKEFIKALGEHNLSQDDLNEIAKDITLIKDEIKL KKALAKYDLNQNQIDSLSKLEFKDHLNISFKALKLVTPLMLEGKKY DEACNELNLKVAINEDKKDFLPAFNETYYKDEVTNPVVLRAIKEY	52

Name	Amino Acid Sequence	SEQ ID NO
	RKVLNALLKKYGKVHKINIELAREVGKNHSQRAKIEKEQNENYKA KKDAELECEKLGLKINSKNILKLRLFKEQKEFCAYSGEKIKISDLQD EKMLEIDHIYPYSRSFDDSYMNKVLVFTKQNQEKLNQTPFEAFGN DSAKWQKIEVLAKNLPTKKQKRILDKNYKDKEQKNFKDRNLNDT RYIARLVLNYTKDYLDFLPLSDDENTKLNDTQKGSKVHVEAKSGM LTSALRHTWGFSAKDRNNHLHHAIDAVIIAYANNSIVKAFSDFKKE QESNSAELYAKKISELDYKNKRKFFEPFSGFRQKVLDKIDEIFVSK PERKKPSGALHEETFRKEEEFYQSYGGKEGVLKALELGKIRKVN GKIVKNGDMFRVDIFKHKKTNKFYAVPIYTMDFALKVLPNKAVAR SKKGEIKDWILMDENYEFCFSLYKDSLILIQTKDMQEPEFVYYNAF TSSTVSLIVSKHDNKFETLSKNQKILFKNANEKEVIAKSIGIQNLKV FEKYIVSALGEVTKAEFRQREDFKK	
A0A386IRG9 Staphylococcus aureus dCas9	MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKK NLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMA KVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDK LFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDL DNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGG ASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGN SRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLP NEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAI VDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT YHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSD GFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSP AIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKN SRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRD MYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGK	53

Name	Amino Acid Sequence	SEQ ID NO
	SDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVI TLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKY PKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFF KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQ VNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDS PTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFL EAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQI SEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGG D	
P14727 Xanthomonas euvesicatoria	MDPIRSRTPSPARELLPGPQPDGVQPTADRGVSPPAGGPLDGL PARRTMSRTRLPSPPAPSPAFSAGSFSDLLRQFDPSLFNTSLFD SLPPFGAHHTEAATGEWDEVQSGLRAADAPPPTMRVAVTAARP PRAKPAPRRRAAQPSDASPAAQVDLRTLGYSQQQQEKIKPKVR STVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALP EATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKI AKRGGVTAVEAVHAWRNALTGAPLNLTPEQVVAIASHDGGKQAL ETVQRLLPVLCQAHGLTPQQVVAIASNGGGKQALETVQRLLPVL CQAHGLTPQQVVAIASNSGGKQALETVQRLLPVLCQAHGLTPEQ VVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGK QALETVQALLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLP VLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQAHGLTPE QVVAIASHDGGKQALETVQRLLPVLCQAHGLTPEQVVAIASHDG GKQALETVQRLLPVLCQAHGLTPQQVVAIASNGGGKQALETVQR LLPVLCQAHGLTPEQVVAIASNSGGKQALETVQALLPVLCQAHGL TPEQVVAIASNSGGKQALETVQRLLPVLCQAHGLTPEQVVAIASH DGGKQALETVQRLLPVLCQAHGLTPEQVVAIASHDGGKQALETV QRLLPVLCQAHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQA HGLTPQQVVAIASNGGGRPALETVQRLLPVLCQAHGLTPEQVVAI	54

Name	Amino Acid Sequence	SEQ ID NO
	ASHDGGKQALETVQRLLPVLCQAHGLTPQQVVAIASNGGGRPAL ESIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLPHAP ALIKRTNRRIPERTSHRVADHAQVVRVLGFFQCHSHPAQAFDDA MTQFGMSRHGLLQLFRRVGVTELEARSGTLPPASQRWDRILQA SGMKRAKPSPTSTQTPDQASLHAFADSLERDLDAPSPMHEGDQ TRASSRKRSRSDRAVTGPSAQQSFEVRVPEQRDALHLPLSWRV KRPRTSIGGGLPDPGTPTAADLAASSTVMREQDEDPFAGAADDF PAFNEEELAWLMELLPQ	
B2SU53 Xanthomonas oryzae pv. Oryzae	MDPIRSRTPSPARELLPGPQPDRVQPTADRGGAPPAGGPLDGL PARRTMSRTRLPSPPAPSPAFSAGSFSDLLRQFDPSLLDTSLLDS MPAVGTPHTAAAPAECDEVQSGLRAADDPPPTVRVAVTAARPP RAKPAPRRRAAQPSDASPAAQVDLRTLGYSQQQQEKIKPKVGS TVAQHHEALVGHGFTHAHIVALSRHPAALGTVAVKYQDMIAALPE ATHEDIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLVKIA KRGGVTAVEAVHASRNALTGAPLNLTPAQVVAIASNNGGKQALE TVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETMQRLLPVLC QAHGLPPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQV VAIASHGGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGK QALETVQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQRLL PVLCQAHGLTPDQVVAIASNGGKQALETVQRLLPVLCQAHGLTP DQVVAIASHDGGKQALETVQRLLPVLCQTHGLTPAQVVAIASHD GGKQALETVQQLLPVLCQAHGLTPDQVVAIASNIGGKQALATVQ RLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQAH GLTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLTQVQVVAI ASNIGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQAL ETVQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQRLLPVL CQAHGLTQEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQ VVAIASNGGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGGK QALETVQRLLPVLCQDHGLTLAQVVAIASNIGGKQALETVQRLLP VLCQAHGLTQDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTP DQVVAIASNIGGKQALETVQRLLPVLCQDHGLTLDQVVAIASNGG	55

Name	Amino Acid Sequence	SEQ ID NO
	KQALETVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETVQRL LPVLCQDHGLTPNQVVAIASNGGKQALESIVAQLSRPDPALAALT NDHLVALACLGGRPAMDAVKKGLPHAPELIRRVNRRIGERTSHR VADYAQVVRVLEFFQCHSHPAYAFDEAMTQFGMSRNGLVQLFR RVGVTELEARGGTLPPASQRWDRILQASGMKRAKPSPTSAQTP DQASLHAFADSLERDLDAPSPMHEGDQTGASSRKRSRSDRAVT GPSAQHSFEVRVPEQRDALHLPLSWRVKRPRTRIGGGLPDPGT PIAADLAASSTVMWEQDAAPFAGAADDFPAFNEEELAWLMELLP QSGSVGGTI	
A0A158RFF2 Gremmeniella abietina	INPWFLTGFIDGEGCFRISVTKINRAIDWRVQLFFQINLHEKDRALL ESIKDYLKVGKIHISGKNLVQYRIQTFDELTILIKHLKEYPLVSKKRA DFELFNTAHKLIKNNEHLNKEGINKLVSLKASLNLGLSESLKLAFP NVISATRLTDFTVNIPDPHWLSGFASAEGCFMVGIAKSSASSTGY QVYLTFILTQHVRDENLMKCLVDYFNWGRLARKRNVYEYQVSKF SDVEKLLSFFDKYPILGEKAKDLQDFCSVSDLMKSKTHLTEEGVA KIRKIKEGMNRG	56
Q94AD9 Arabidopsis thaliana	MRTPMSDTQHVQSSLVSIRSSDKIEDAFRKMKVNETGVEELNPY PDRPGERDCQFYLRTGLCGYGSSCRYNHPTHLPQDVAYYKEEL PERIGQPDCEYFLKTGACKYGPTCKYHHPKDRNGAQPVMFNVIG LPMRLGEKPCPYYLRTGTCRFGVACKFHHPQPDNGHSTAYGMS SFPAADLRYASGLTMMSTYGTLPRPQVPQSYVPILVSPSQGFLP PQGWAPYMAASNSMYNVKNQPYYSGSSASMAMAVALNRGLSE SSDQPECRFFMNTGTCKYGDDCKYSHPGVRISQPPPSLINPFVL PARPGQPACGNFRSYGFCKFGPNCKFDHPMLPYPGLTMATSLP TPFASPVTTHQRISPTPNRSDSKSLSNGKPDVKKESSETEKPDN GEVQDLSEDASSP	57

[0129] Any DBD amino acid sequence of the present technology may be modified to comprise one or more non-naturally occurring amino acids. In some embodiments, the DBDs comprise or consist of an amino acid sequence having one or more non-naturally occurring amino acids. In some embodiments, the DBDs comprise or consist

of an amino acid sequence having two or more non-naturally occurring amino acids. In some embodiments, the modification of the amino acid sequences to comprise non-naturally occurring amino acids is a conservative substitution.

[0130] In some embodiments, the p300R-DBD protein comprises or consists of the amino acid sequence MSARQGQMQGIPAPSQALQEPAPWSALPCGLLLDELLASPEFLQQAQPLLETEAPG ELEASEEAASLEAPLSEEEYRALLEELSGGGGSRGGGSGGGGSGGGGSGGGGSG GGGMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDS GETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPG EKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKE IFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDN GSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMT RKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNEL TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG VEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHL FDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHD DSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPE NIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYY LQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITK HVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFF KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG FSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVK ELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQ KGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKR VILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTST KEVLDATLIHQSITGLYETRIDLSQLGGDAYPYDVPDYASLGSGSPKKKRKVEDPKKK RKVDGIGSGSNGSSGSATNFSLLKQAGDVEENPGPMVSKGEEDNMAIIKEFMRFKV HMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFMYGSKAY

VKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNF PSDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLKDGGHYDAEVKTTYKAK KPVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGRHSTGGMDELYK (SEQ ID NO: 9).

[0131] In some embodiments, the p300R-DBD protein comprises or consists of an amino acid sequence about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence of SEQ ID NO: 9.

[0132] In some embodiments, the p300R-DBD protein comprises or consists of an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence of SEQ ID NO: 9.

[0133] In some embodiments, the p300R-DBD protein comprises or consists of amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence of SEQ ID NO: 9.

[0134] Any fusion peptide amino acid sequence of the present technology may be modified to comprise one or more non-naturally occurring amino acids. In some embodiments, the fusion peptides comprise or consist of an amino acid sequence having one or more non-naturally occurring amino acids. In some embodiments, the fusion peptides comprise or consist of an amino acid sequence having two or more non-naturally occurring amino acids. In some embodiments, the modification of the amino acid sequences to comprise non-naturally occurring amino acids is a conservative substitution.

Additional Peptide Features

[0135] Localization domains: In some embodiments, the peptides of the present technology, such as the p300R-DBD proteins, further comprise a localization domain. Any suitable localization domain can be used, including but not limited to any nuclear localization domain.

[0136] The localization domain may comprise an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%,

97%, 98%, 99%, or 100% identical to the amino acid sequence of YPYDVPDYASLGSGSPKKKRKVEDPKKKRKVDGIGSGSNGSSGSATNFSLLKQAGDVEENPGP (SEQ ID NO: 35), wherein the underlined residues are optional. In some embodiments, the optional residues are present. In some embodiments, the optional residues are absent, in whole or in part.

[0137] The localization domain may comprise an amino acid sequence about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence of YPYDVPDYASLGSGSPKKKRKVEDPKKKRKVDGIGSGSNGSSGSATNFSLLKQAGDVEENPGP (SEQ ID NO: 35), wherein the underlined residues are optional. In some embodiments, the optional residues are present. In some embodiments, the optional residues are absent, in whole or in part.

[0138] The localization domain may comprise an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence of YPYDVPDYASLGSGSPKKKRKVEDPKKKRKVDGIGSGSNGSSGSATNFSLLKQAGDVEENPGP (SEQ ID NO: 35), wherein the underlined residues are optional. In some embodiments, the optional residues are present. In some embodiments, the optional residues are absent, in whole or in part.

[0139] Detectable domains: In some embodiments, the p300R-DBD proteins of the present technology further comprise a detectable domain. The detectable domain may comprise an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence of SEQ ID NO: 36, 37, or 38 (Table 2).

[0140] Detectable domains: In some embodiments, the p300R-DBD proteins of the present technology further comprise a detectable domain. The detectable domain may comprise an amino acid sequence about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence of SEQ ID NO: 36, 37, or 38.

[0141] Detectable domains: In some embodiments, the p300R-DBD proteins of the present technology further comprise a detectable domain. The detectable domain may comprise an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%,

80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence of SEQ ID NO: 36, 37, or 38.

Table 2: Detectable Domain Amino Acid Sequences

Name	Amino Acid Sequence	SEQ ID NO
Detectable Domain A	MVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPY EGTQTAKLKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDY LKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLR GTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKL KDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVE QYERAEGRHSTGGMDELYK	36
Detectable Domain B	MINEIKKNAQERMDETVEQLKNELSKVRTGGGGTEERRLELAK QVVFAANRALIRVRTIALEAAWRLRMLGSDKEVNKRDISQALEEI EKLTKVAAKKIKEVLEAKIKELREVMAVNSGGGGSRGGGSGGG GSGGGGSGGGGSGGGGMDKKYSIGLAIGTNSVGWAVITDEYK VPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRR YTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHER HPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI KFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVD AKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFK SNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEE LLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFL KDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNF EEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYN ELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKE DYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYT GWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDE LVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS	37

Name	Amino Acid Sequence	SEQ ID NO
	DYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLV ETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRK DFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGD YKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIR KRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVK KDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFL YLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL ADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	
Detectable Domain C	MINEIKKNAQERMDETVEQLKNELSKVRTGGGGTEERRLELAK QVV**F**AANRAL**I**RVRTIALEAAWRLRMLGSDKEWKRDISQALEEIE KLTKVAAKKIKEVLEAKIKELREVMAWSGGGGSRGGGSGGGGS GGGGSGGGGSGGGGMDKKYSIGLAIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKF RGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAK AILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSN FDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI LLSDILRVTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEK YKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLV KLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDN REKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEW DKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTK VKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFK KIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDIL EDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFK	38

Name	Amino Acid Sequence	SEQ ID NO
	EDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKWDELVKV MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDV DAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEWKKMKNY WRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF YKVREINNYHHAHDAYLNAWGTALIKKYPKLESEFVYGDYKVYD VRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE TNGETGEIVWDKGRDFATVRKVLSMPQWIVKKTEVQTGGFSKE SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLWAKVEKGK SKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK YSLFELENGRKRMLASAGELQKGNELALPSKYWFLYLASHYEKL KGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRY TSTKEVLDATLIHQSITGLYETRIDLSQLGGDYPYDVPDYASLGS GSPKKKRKVEDPKKKRKVDGIGSGSNGSSGSATNFSLLKQAGD VEENPGPMVSKGEEDNMAIIKEFMRFKVHMEGSWGHEFEIEGE GEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFMYGSKAYVKH PADIPDYLKLSFPEGFKWERVMNFEDGGWTVTQDSSLQDGEFI YKVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEI KQRLKLKDGGHYDAEVKTTYKAKKPVQLPGAYNWIKLDITSHNE DYTIVEQYERAEGRHSTGGMDELYK	

[0142] The DBD may be fused to p300R domain by a peptide linker. The peptide linkers of the present technology may vary from 2 to 31 amino acids of any primary sequence in length and do not impose any constraints on the conformation or interactions of the linked partners. In some embodiments, the linkers vary from about 2-30, 2-29, 2-28, 2-27, 2-26, 2-25, 2-24, 2-23, 2-22, 2-21, 2-20, 2-19, 2-18, 2-17, 2-16, 2-15, 2- 14, 2-13, 2-12, 2-11, 2-10, 2-9, 2-8, 2-20, 2-6, 2-5, 2-4, 2-3, 3-31, 3-30, 3-29, 3-28, 3-27, 3-26, 3-25, 3-24, 3-23, 3-22, 3-21, 3-20, 3-19, 3-18, 3-17, 3-16, 3-15, 3-14, 3-13, 3-12, 3-11, 3-10, 3-9, 3-8, 3-7, 3-6, 3-5, 3-4, 4-31, 4-30, 4-29, 4-28, 4-27, 4-26, 4-

25, 4-24, 4-23, 4-22, 4-21, 4- 20, 4-19, 4-18, 4-17, 4-16, 4-15, 4-14, 4-13, 4-12, 4-11, 4-10, 4-9, 4-8, 4-7, 4-6, 4-5, 5-31, 5- 30, 5-29, 5-28, 5-27, 5-26, 5-25, 5-24, 5-23, 5-22, 5-21, 5-20, 5-19, 5-18, 5-17, 5-16, 5-15, 5- 14, 5-13, 5-12, 5-11, 5-10, 5-9, 5-8, 5-7, 5-6, 6-31, 6-30, 6-29, 6-28, 6-27, 6-26, 6-25, 6-24, 6-23, 6-22, 6-21, 6-20, 6-19, 6-18, 6-17, 6-16, 6-15, 6-14, 6-13, 6-12, 6-11, 6-10, 6-9, 6-8, 6- 7, 7-31, 7-30, 7-29, 7-28, 7-27, 7-26, 7-25, 7-24, 7-23, 7-22, 7-21, 7-20, 7-19, 7-18, 7-17, 7- 16, 7-15, 7-14, 7-13, 7-12, 7-11, 7-10, 7-9, 7-8, 8-31, 8-30, 8-29, 8-28, 8-27, 8-26, 8-25, 8-24, 8-23, 8-22, 8-21, 8-20, 8-19, 8-18, 8-17, 8-16, 8-15, 8-14, 8-13, 8-12, 8-11, 8-10, 8-9, 9-31, 9-30, 9-29, 9-28, 9-27, 9-26, 9-25, 9-24, 9-23, 9-22, 9-21, 9-20, 9-19, 9-18, 9-17, 9-16, 9-15, 9-14, 9-13, 9-12, 9-11, 9-10, 10-31, 10-30, 10-29, 10-28, 10-27, 10-26, 10-25, 10-24, 10-23, 10-22, 10-21, 10-20, 10-19, 10-18, 10-17, 10-16, 10-15, 10-14, 10-13, 10-12, 10-11, 11-31, 11-30, 11-29, 11-28, 11-27, 11-26, 11-25, 11-24, 11-23, 11-22, 11-21, 11-20, 11-19, 11-18, 11-17, 11-16, 11-15, 11-14, 11-13, 11-12, 12-31, 12-30, 12-29, 12-28, 12-27, 12-26, 12-25,12-24, 12-23, 12-22, 12-21, 12-20, 12-19, 12-18, 12-17, 12-16, 12-15, 12-14, 12-13, 13-31, 13-30, 13-29, 13-28, 13-27, 13-26, 13-25, 13-24, 13-23, 13-22, 13-21, 13-20, 13-19, 13-18, 13-17, 13-16, 13-15, 13-14, 14-31, 14-30, 14-29, 14-28, 14-27, 14-26, 14-25, 14-24, 14-23, 14-22, 14-21, 14-20, 14-19, 14-18, 14-17, 14-16, 14-15, 15-31, 15-30, 15-29, 15-28, 15-27, 15-26, 15-25, 15-24, 15-23, 15-22, 15-21, 15-20, 15-19, 15-18, 15-17, 15-16, 16-31, 16-30, 16-29, 16-28, 16-27, 16-26, 16-25, 16-24, 16-23, 16-22, 16-21, 16-20, 16-19, 16-18, 16-17, 17-31, 17-30, 17-29, 17-28, 17-27, 17-26, 17-25, 17-24, 17-23, 17-22, 17-21, 17-20, 17-19, 17-18, 18-31, 18-30, 18-29, 18-28, 18-27, 18-26, 18-25, 18-24, 18-23, 18-22, 18-21, 18-20, 18-19, 19-31, 19-30, 19-29, 19-28, 19-27, 19-26, 19-25, 19-24, 19-23, 19-22, 19-21, 19-20, 20-31, 20-30, 20-29, 20-28, 20-27, 20-26, 20-25, 20-24, 20-23, 20-22, 20-21, 21-31, 21-30, 21-29, 21-28, 21-27, 21-26, 21-25, 21-24, 21-23, 21-22, 22-31, 22-30, 22-29, 22-28, 22-27, 22-26, 22-25, 22-24, 22-23, 23-31, 23-30, 23-29, 23-28, 23-27, 23-26, 23-25, 23-24, 24-31, 24-30, 24-29, 24-28, 24-27, 24-26, 24-25, 25-31, 25-30, 25-29, 25-28, 25-27, 25-26, 26-31, 26-30, 26-29, 26-28, 26-27, 27-31, 27-30, 27-29, 27-28, 28-31, 28-30, 28-29, 29-31, 29-30, or 30-31 amino acids of any primary sequence in length. The peptide linkers may be designed as appropriate for an intended use.

[0143] The peptide linkers may comprise one or more of a Gly-rich linkers (e.g., a flexible linker connecting various domains in a single protein without interfering with the function of each domain; a linker forming stable covalently linked dimers; a linker to

connect two independent domains that create a ligand-binding site or recognition sequence), a Serine linker (e.g., a coiled structure linker); a coiled structure linker comprising a Gln, Arg, Glu, Ser, and/or Pro amino acids; a rigid space linker comprising one or more of a Pro, Arg, Phe, Thr, Glu, and/or Gln residues; a linker comprising a flexible Gly-rich regions that may can generate loops connecting domains; or a linker comprising a Thr, Ser, Gly, and/or Ala residue.

[0144] In some embodiments, the peptide linker comprises or consists of an amino acid sequence listed in Table 3.

[0145] In some embodiments, the peptide linker comprises or consists of an amino acid sequence about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 100% identical to the amino acid sequence of any one of SEQ ID NOs: 15-33 or 81 (Table 3).

[0146] In some embodiments, the peptide linker comprises or consists of an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 100% identical to the amino acid sequence of any one of SEQ ID NOs: 15-33 or 81.

[0147] In some embodiments, the peptide linker comprises or consists of an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 100% identical to the amino acid sequence of any one of SEQ ID NOs: 15-33 or 81.

[0148] Any peptide linker amino acid sequence of the present technology may be modified to comprise one or more non-naturally occurring amino acids. In some embodiments, the peptide linkers comprise or consist of an amino acid sequence having one or more non-naturally occurring amino acids. In some embodiments, the peptide linkers comprise or consist of an amino acid sequence having two or more non-naturally occurring amino acids.

Table 3: Peptide Linker Amino Acid Sequences

Name	Amino Acid	SEQ ID NO
Linker A	$(GGS)_{1-5}$	15
Linker B	GGGGSLVPRGSGGGGS	16
Linker C	GSGSGS	17

Name	Amino Acid	SEQ ID NO
Linker D	GS_8	18
Linker E	GGSGGHMGSGG	19
Linker F	GGSGGSGGSGG	20
Linker G	GGSGG	21
Linker H	GGSGGGGG	22
Linker I	GSGSGSGS	23
Linker J	GSGSGSGSGSGSGSGSGSGSGSGSGSGSG	24
Linker K	GGGSEGGGSEGGGSEGGG	25
Linker L	AAGAATAA	26
Linker M	GGGGG	27
Linker N	GGSSG	28
Linker O	GSGGGTGGGSG	29
Linker P	GT	30
Linker Q	GSGSGSGSGGSG	31
Linker R	GSGGSGSGGSGGSG	32
Linker S	SGGGGSRGGGSGGGG**SGGGGSGGGGSGGGG**	33*
Linker T	SGGGG	81

[0149] *One or more of the underlined residues in SEQ ID NO: 33 are optional.

gRNAs

[0150] The gRNAs of the present technology may comprise a guanine nucleotide at the 5′ end of its polynucleotide sequence. The guanine nucleotide in 5′ is may be present when the gRNA is expressed under the control of the U6 promoter. The gRNAs may be of varying lengths.

[0151] The gRNA may comprise a gRNA guide sequence of about 10 nucleotides (nts), about 11 nts, about a 12 nts, about a 13 nts, about a 14 nts, about a 15 nts, about a 16 nts, about a 17 nts, about a 18 nts, about a 19 nts, about a 20 nts, about a 21 nts, about a 22 nts, about a 23 nts, about a 24 nts, about a 25 nts, about a 30 nts, or about a 35 nts of a target sequence in the gene target. In embodiments, the "gRNA guide sequence" or "gRNA target sequence" may be least 10 nucleotides long, for example,

about 10-40 nts long (e.g., 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39 or 40 nts long).

[0152] The gRNA may comprise a gRNA guide sequence of at least 10 nucleotides (nts), at least 11 nts, at least a 12 nts, at least a 13 nts, at least a 14 nts, at least a 15 nts, at least a 16 nts, at least a 17 nts, at least a 18 nts, at least a 19 nts, at least a 20 nts, at least a 21 nts, at least a 22 nts, at least a 23 nts, at least a 24 nts, at least a 25 nts, at least a 30 nts, or at least a 35 nts of a target sequence in the gene target. In embodiments, the "gRNA guide sequence" or "gRNA target sequence" may be least 10 nucleotides long, for example, at least 10-40 nts long (e.g., 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39 or 40 nts long).

[0153] The gRNA may comprise a gRNA guide sequence of at least about 10 nucleotides (nts), at least about 11 nts, at least about a 12 nts, at least about a 13 nts, at least about a 14 nts, at least about a 15 nts, at least about a 16 nts, at least about a 17 nts, at least about a 18 nts, at least about a 19 nts, at least about a 20 nts, at least about a 21 nts, at least about a 22 nts, at least about a 23 nts, at least about a 24 nts, at least about a 25 nts, at least about a 30 nts, or at least about a 35 nts of a target sequence in the gene target. In embodiments, the "gRNA guide sequence" or "gRNA target sequence" may be least 10 nucleotides long, for example, at least about 10-40 nts long (e.g., 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39 or 40 nts long).

[0154] In some embodiments, gRNA guide sequence is between 17-30, 17-22, 10-40, 10-30, 12-30, 15-30, 18-30, or 10-22 nucleotides long.

[0155] The number of gRNAs administered to or expressed in a cell of the present technology may be about 1 gRNA, about 2 gRNAs, about 3 gRNAs about 4 gRNAs, about 5 gRNAs, about 6 gRNAs, about 7 gRNAs, about 8 gRNAs, about 9 gRNAs, about 10 gRNAs, about 11 gRNAs, about 12 gRNAs, about 13 gRNAs, about 14 gRNAs, about 15 gRNAs, about 16 gRNAs, about 17 gRNAs, or about 18 gRNAs.

[0156] The number of gRNAs administered to or expressed in a cell of the present technology may be at least 1 gRNA, at least 2 gRNAs, at least 3 gRNAs at least 4 gRNAs, at least 5 gRNAs, at least 6 gRNAs, at least 7 gRNAs, at least 8 gRNAs, at least 9 gRNAs, at least 10 gRNAs, at least 11 gRNAs, at least 12 gRNAs, at least 13

gRNAs, at least 14 gRNAs, at least 15 gRNAs, at least 16 gRNAs, at least 17 gRNAs, or at least 18 gRNAs.

[0157] The number of gRNAs administered to or expressed in a cell of the present technology may be at least about 1 gRNA, at least about 2 gRNAs, at least about 3 gRNAs at least about 4 gRNAs, at least about 5 gRNAs, at least about 6 gRNAs, at least about 7 gRNAs, at least about 8 gRNAs, at least about 9 gRNAs, at least about 10 gRNAs, at least about 11 gRNAs, at least about 12 gRNAs, at least about 13 gRNAs, at least about 14 gRNAs, at least about 15 gRNAs, at least about 16 gRNAs, at least about 17 gRNAs, or at least about 18 gRNAs.

[0158] In some embodiments, when at least a complete match (e.g., a 100% complementary match) between the gRNA guide sequence and the DNA sequence on the target locus does not occur, a mismatch between a gRNA guide sequence and target sequence on the gene sequence of interest is also permitted in the event it still allows hybridization of the gRNA with the complementary strand of the gRNA target polynucleotide sequence on the targeted gene.

[0159] Any gRNA guide sequence may be selected in the target gene, as long as it allows introducing at the proper location, the desired modification(s). Accordingly, the gRNA guide sequence or target sequence of the present invention may be in coding or non-coding regions of the target gene.

[0160] In some embodiments, the loci is selected from the group consisting of a Laminin Subunit Alpha 1 (LAMA1) loci, a Structural Maintenance of Chromosomes flexible Hinge Domain Containing 1 (SMCHD1) loci, a (Ligand Dependent Nuclear Receptor Interacting Factor 1) LRIF1 loci, a (Matrin 3) MATR3 loci, an Insulin-like Growth Factor 1 (IGF1) loci, a promyelocytic leukemia protein (PML) loci, and/or an adiponectin (ADIPOQ) loci.

[0161] The loci may comprise a nucleotide sequence encoding a gene, a nucleotide sequence that is a cis regulatory element of a gene, a nucleotide sequence that is a trans regulatory element of a gene, and/or a promoter region. Nonlimiting examples of genes include LAMA1 (NCBI GENE ID: 284217; NCBI Protein Accession(s): P25391.2, AAH62318.1, AAH39051.1), SMCHD1 (NCBI GENE ID: 23347; NCBI Protein Accession: NP_056110.2), LRIF1 (NCBI GENE ID: 55791; NCBI Protein Accession(s): NP_001006946.1, NP_060842.3, Q5T3J3.1, XP_054193607.1,

XP_054193607.1), MATR3 (NCBI GENE ID: 9782; NCBI Protein Accession: P43243.2), IGF1 (NCBI GENE ID: 3479; NCBI Protein Accession: CAG46659.1), PML (NCBI GENE ID: 5371; NCBI Protein Accession: AAB19601.2), and ADIPOQ (NCBI GENE ID: 9370; NCBI Protein Accession: AAH54496.1).

[0162] The gRNAs of the present technology may be used to target one or more genes. Genes may be targeted to increase a gene expression level thereof, relative to a control. In some embodiments, the gRNA or at least a nucleotide thereof, binds at or near a transcription start site (TSS) of a gene within and/or associated with the loci. In some embodiments, the gRNA or at least the nucleotide thereof binds at least within at least within 100 nucleotides, 200 nucleotides, 300 nucleotides, 400 nucleotides, 500 nucleotides, 600 nucleotides, 700 nucleotides, 800 nucleotides, 900 nucleotides, 1000 nucleotides, 1100 nucleotides, 1200 nucleotides, 1300 nucleotides, 1400 nucleotides, 1500 nucleotides, 1600 nucleotides, 11700 nucleotides, 1800 nucleotides, 1900 nucleotides, 2000 nucleotides, 3000 nucleotides, 4000 nucleotides, or 5000 nucleotides of the TSS of the gene within and/or associated with the loci.

[0163] In some embodiments, the gRNA sequence comprises a nucleotide sequence listed in Table 4.

[0164] In some embodiments, the gRNA sequence comprises a nucleotide sequence about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 100% identical to any one of SEQ ID NOs: 58-80.

[0165] In some embodiments, the gRNA sequence comprises a nucleotide sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 100% identical to any one of SEQ ID NOs: 58-80.

[0166] In some embodiments, the gRNA sequence comprises a nucleotide sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 100% identical to any one of SEQ ID NOs: 58-80.

Table 4: gRNA sequences

gRNA	Nucleotide Sequence	SEQ ID NO
gRNA 1	GCGTTATGAATACGTTTGCG	58
gRNA 2	CCCTCTTACATTGAGTCGTA	59

gRNA 3	GATATAGGGTCCCGCAAGAA	60
gRNA 4	GCCCTTTAAACAGGTACGGG	61
gRNA 5	GCGAAATTATCACACACACG	62
gRNA 6	GAAAATTTCACAACAGCACA	63
LAMA1 gRNA9	GTCCGAAGGGCCTGCGCACC	64
LAMA1 gRNA10	GGGCGCGGCCAGGCAGGCGG	65
LAMA1 gRNA11	CCCTGAATGCTCACCCCAAC	66
LAMA1 gRNA12	GCGCCCAGGTCTGTCCTCCA	67
ADIPOQ gRNA 3	GACCCAAGCTGGATTAAACC	68
IGF1 gRNA1	CCAGGGTAAACTAGATCGAA	69
IGF1 gRNA2	ATCCATGGGGCAGCGTAAAG	70
IGF1 gRNA3	TCAGATACTCAACAGACGTC	71
IGF1 gRNA4	AAAGCACTAACTTACACGAA	72
IGF1 gRNA5	GTATTCCAAATAATACCAGC	73
IGF1 gRNA6	GCTCCAGTTTTTTAAGAGCGA	74
IGF1 gRNA7	GTGTGCTAATACACCTCACC	75
IGF1 gRNA8	TTAGCAGACATTAAATACTG	76
ADIPOQ gRNA 1	TAAGATGTGAGTCCGCCGAG	77
ADIPOQ gRNA 2	ATGGGATCCGGTCTAGCAAG	78
ADIPOQ gRNA 4	GGAGCAAGGGGCCCACTCAT	79
ADIPOQ gRNA 5	CCTTTCCCATAAGTTTGTCC	80

Vectors

[0167] In some embodiments, the present technology comprises vectors comprising the nucleotide sequence of one or more gRNAs, one or more peptides, or one or more protein domains of the present technology (e.g., a p300R-DBD protein, a

p300R domain, a DBD). In some embodiments, the present technology comprises vectors comprising a gRNA sequence of the present technology.

[0168] In some embodiments, the gRNA nucleotide sequence is operatively linked to a suitable control sequence. In one embodiment, the sequence encoding the gRNA is within 50-100 base pairs of a TATA box. In one such embodiment, the TATA box is 5′ to the gRNA encoding sequence; in another embodiment, the TATA box is 3′ to the gRNA encoding sequence. As described in the examples that follow, no sequence or structural gRNA requirements were found with regard to TATA box proximity, and no functional PAM specificity was found with regard to TATA box region.

[0169] The vector may be a viral vector (e.g., a lentiviral vector).

[0170] The vectors of the present technology may comprise a detectable marker and/or a reporter gene.

[0171] In some embodiments, the vector comprises an antibiotic-inducible gene expression system. Nonlimiting examples of antibiotics for use in the antibiotic-inducible gene expression system include doxycycline and tetracycline. In some embodiments, the antibiotic-inducible gene expression system is a Tet-ON system.

[0172] In some embodiments, an engineered, non-naturally occurring peptide or at least a portion thereof, or a gRNA is present in the composition in an amount of about 5 ng, 10 ng, 100 ng, 500 ng, 1000 ng, 2000 ng, 25 µg, 27 µg, 30 µg, 50 µg, 100 µg, or 500 µg.

[0173] In some embodiments, an engineered, non-naturally occurring peptide or at least a portion thereof, or a gRNA is present in the composition in an amount of at least 5 ng, 10 ng, 100 ng, 500 ng, 1000 ng, 2000 ng, 25 µg, 27 µg, 30 µg, 50 µg, 100 µg, or 500 µg.

[0174] In some embodiments, an engineered, non-naturally occurring peptide or at least a portion thereof, or a gRNA is present in the composition in an amount of at least about 5 ng, 10 ng, 100 ng, 500 ng, 1000 ng, 2000 ng, 25 µg, 27 µg, 30 µg, 50 µg, 100 µg, or 500 µg.

Host cells

[0175] The compositions or the vectors of the present technology may be present in a host cell. In some embodiments, the host cell is a mammalian host cell (e.g., a

human host cell or a murine host cell). The host cell may be a myoblast. In some embodiments, the host cell is a kidney cell, a muscle cell, a liver cell.

Delivery Vehicles

[0176] The compositions (e.g., peptides and/or gRNAs) vectors, or host cells may be present in a delivery vehicle. In some embodiments, the delivery vehicle comprises a nanoparticle. Nonlimiting examples of nanoparticles include lipid nanoparticles and gold nanoparticles.

Scaffolds

[0177] In some embodiments, the composition or the p300R-DBD protein is carried by or otherwise bound to a scaffold, including but not limited to a nanoparticle, virus-like particle (VLP), or other polypeptide scaffold. In embodiments where the composition is a p300R-DBD protein (i.e., a fusion protein), the fusion protein may be further covalently linked to be expressed as part of a polypeptide scaffold. Alternatively, the composition or fusion protein may be linked to the scaffold via any suitable means, as will be apparent to those of skill in the art based on the teachings herein. Any suitable nanoparticle, VLP, or other polypeptide scaffold may be used as deemed appropriate for an intended use.

[0178] In some embodiments, the present technology comprises nucleic acids encoding the polypeptide of any embodiment or combination of embodiments of the disclosure. The nucleic acid sequence may comprise single stranded or double stranded RNA or DNA in genomic or cDNA form, or DNA-RNA hybrids, each of which may include chemically or biochemically modified, non-natural, or derivatized nucleotide bases. Such nucleic acid sequences may comprise additional sequences useful for promoting expression and/or purification of the encoded polypeptide, including but not limited to polyA sequences, modified Kozak sequences, and sequences encoding epitope tags, export signals, and secretory signals, nuclear localization signals, and plasma membrane localization signals. It will be apparent to those of skill in the art, based on the teachings herein, what nucleic acid sequences will encode the polypeptides of the disclosure.

Pharmaceutical Compositions, Formulations, and Administration

[0179] In some embodiments, the compositions, vectors, and/or delivery vehicles of the present technology are present in a pharmaceutical composition. The pharmaceutical composition may comprise one or more pharmaceutically acceptable carriers and/or excipients such as, for example, buffers, binders, excipients, stabilizers, lubricants, oils, adjuvants, and antioxidants. For example, when the pharmaceutical composition is formulated for subcutaneous administration, the one or more pharmaceutically acceptable carriers and/or excipients may comprise water. The pharmaceutical composition may comprise a therapeutically effective amount and/or a clinically effective amount of the compositions, vectors, and/or delivery vehicles of the present technology. In some embodiments, two or more dosage units of the pharmaceutical composition comprise a therapeutically effective amount and/or a clinically effective amount of the compositions and/or vectors of the present technology.

[0180] In some embodiments, the pharmaceutical compositions of the present technology are formulated for delivery to a subject in need thereof. In some embodiments, the pharmaceutical compositions are formulated for intraperitoneal, intravenous, parenteral, subcutaneous, intramuscular, intracerebroventricular, or oral administration to the subject.

Associated Methods

[0181] The compositions, vectors, and pharmaceutical compositions of the present technology may be used to increase a gene or protein expression in a subject in need thereof. This may be accomplished by (1) promoting euchromatin formation at or near a loci which comprises the gene or a gene encoding the protein of interest, or (2) promoting euchromatin formation at or near a loci that comprises a gene encoding an expressional activator (e.g., a transcription factor) of the gene or protein of interest. In some embodiments, the subject has muscular dystrophy (e.g., Merosin-deficient congenital muscular dystrophy (MDC1A) or facioscapulohumeral muscular dystrophy (FSHD)). In some embodiments, the subject has cancer.

[0182] In some embodiments, the present technology comprises a method of increasing a level of LAMA1 gene or protein expression in a subject in need thereof relative to a control, the method comprising administering to the subject an engineered,

non-naturally occurring peptide, a composition, a vector, a pharmaceutical composition, or a host cell of the present technology. In some embodiments, the subject has a LAMA2 mutation that reduces LAMA2 activity and/or gene expression or protein expression relative to a control. The methods may comprise (i) altering a locus altered by the methylation of one or more histones of the present technology, thereby increasing a level of LAMA1, SMCHD1, LRIF1, MATR3, IGF1, or PML or (ii) increasing the level of LAMA1 gene or protein expression and/or activity, where (i) or (ii) may occur without the use of an enzyme or a nucleotide sequence encoding an enzyme in the composition, the vector, the pharmaceutical composition, or the host cell.

[0183] In some embodiments, the present technology comprises a method of increasing a level of LAMA1, SMCHD1, LRIF1, MATR3, PML, IGF1, or ADIPOQ gene or protein expression in a subject in need thereof, relative to a control, the method comprising administering to the subject an engineered, non-naturally occurring peptide, a composition, a vector, a pharmaceutical composition, or a host cell of the present technology.

[0184] In some embodiments, the engineered, non-naturally occurring peptide, the gRNA, the composition, the vector, the pharmaceutical composition, or the host cell is administered to the subject at a dose of about 0.5 ng to about 500 µg. For example, a dose of about 5 ng, 10 ng, 100 ng, 500 ng, 1000 ng, 2000 ng, 25 µg, 27 µg, 30 µg, 50 µg, 100 µg, or 500 µg.

[0185] In some embodiments, the engineered, non-naturally occurring peptide, the gRNA, the composition, the vector, the pharmaceutical composition, or the host cell is administered to the subject at a dose of at least 0.5 ng to at least 500 µg. For example, a dose of at least 5 ng, 10 ng, 100 ng, 500 ng, 1000 ng, 2000 ng, 25 µg, 27 µg, 30 µg, 50 µg, 100 µg, or 500 µg.

[0186] In some embodiments, the engineered, non-naturally occurring peptide, the gRNA, the composition, the vector, the pharmaceutical composition, or the host cell is administered to the subject at a dose of at least about 0.5 ng to at least about 500 µg. For example, a dose of at least 5 ng, 10 ng, 100 ng, 500 ng, 1000 ng, 2000 ng, 25 µg, 27 µg, 30 µg, 50 µg, 100 µg, or 500 µg.

[0187] In some embodiments, the level of LAMA1, SMCHD1, LRIF1, MATR3, PML, IGF1, or ADIPOQ gene or protein expression is increased by about 5%, 10%,

15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 200%, 300%, 400%, or 500% relative to a control.

[0188] In some embodiments, the level of LAMA1, SMCHD1, LRIF1, MATR3, PML, IGF1, or ADIPOQ gene or protein expression is increased by at least 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 200%, 300%, 400%, or 500% relative to a control.

[0189] In some embodiments, the level of LAMA1, SMCHD1, LRIF1, MATR3, PML, IGF1, or ADIPOQ gene or protein expression is increased by at least about 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 200%, 300%, 400%, or 500% relative to a control.

[0190] In some embodiments, the methods target a locus that is altered by methylation of one or more histones of the present technology, thereby increasing a level of LAMA1, SMCHD1, LRIF1, MATR3, PML, IGF1, or ADIPOQ gene or protein expression and/or activity. In some embodiments, the alteration of the locus and/or reduction in the level of LAMA1, SMCHD1, LRIF1, MATR3, PML, IGF1, or ADIPOQ gene or protein expression or activity occurs in a mammalian cell (e.g., a human cell or a murine cell). The mammalian cell may be a myoblast. In some embodiments, the mammalian cell is a kidney cell, a muscle cell, or a liver cell.

[0191] When the methods of the present technology comprise administering a vector having a reporter gene and/or a detectable marker, the methods may target a locus that is altered by methylation of one or more histones of the present technology, thereby increasing a level of LAMA1, SMCHD1, LRIF1, MATR3, PML, IGF1, or ADIPOQ gene or protein expression and/or activity, when the reporter gene or the detectable marker is not expressed at a detectable level.

[0192] In some embodiments, the alteration of the locus is maintained for at least about 30 minutes, 1 hour, 2 hours, 3 hours, 4 hours, 6 hours, 12 hours, 1 day, 2 days, 3 days, 4 days, 5 days, 6 days, 1 week, 2 weeks, 3 weeks, 4 weeks, 6 weeks, 2 months, 3 months, 4 months, 5 months, or 6 months.

[0193] In some embodiments, the alteration of the locus is maintained for about 30 minutes, 1 hour, 2 hours, 3 hours, 4 hours, 6 hours, 12 hours, 1 day, 2 days, 3 days,

4 days, 5 days, 6 days, 1 week, 2 weeks, 3 weeks, 4 weeks, 6 weeks, 2 months, 3 months, 4 months, 5 months, or 6 months.

[0194] In some embodiments, the alteration of the locus is maintained for at least 30 minutes, 1 hour, 2 hours, 3 hours, 4 hours, 6 hours, 12 hours, 1 day, 2 days, 3 days, 4 days, 5 days, 6 days, 1 week, 2 weeks, 3 weeks, 4 weeks, 6 weeks, 2 months, 3 months, 4 months, 5 months, or 6 months.

[0195] In some embodiments, the increase in LAMA1, SMCHD1, LRIF1, MATR3, PML, IGF1, or ADIPOQ, gene expression, protein expression, and/or protein activity is maintained for at least about 30 minutes, 1 hour, 2 hours, 3 hours, 4 hours, 6 hours, 12 hours, 1 day, 2 days, 3 days, 4 days, 5 days, 6 days, 1 week, 2 weeks, 3 weeks, 4 weeks, 6 weeks, 2 months, 3 months, 4 months, 5 months, or 6 months.

[0196] In some embodiments, the increase in LAMA1, SMCHD1, LRIF1, MATR3, PML, IGF1, or ADIPOQ, gene expression, protein expression, and/or protein activity is maintained for about 30 minutes, 1 hour, 2 hours, 3 hours, 4 hours, 6 hours, 12 hours, 1 day, 2 days, 3 days, 4 days, 5 days, 6 days, 1 week, 2 weeks, 3 weeks, 4 weeks, 6 weeks, 2 months, 3 months, 4 months, 5 months, or 6 months.

[0197] In some embodiments, the increase in LAMA1, SMCHD1, LRIF1, MATR3, PML, IGF1, or ADIPOQ, gene expression, protein expression, and/or protein activity is maintained for at least 30 minutes, 1 hour, 2 hours, 3 hours, 4 hours, 6 hours, 12 hours, 1 day, 2 days, 3 days, 4 days, 5 days, 6 days, 1 week, 2 weeks, 3 weeks, 4 weeks, 6 weeks, 2 months, 3 months, 4 months, 5 months, or 6 months.

EXAMPLES

[0198] The following examples are intended to illustrate various embodiments of the present technology. As such, the specific embodiments discussed are not to be construed as limitations on the scope of the present technology. It will be apparent to one skilled in the art that various equivalents, changes, and modifications may be made without departing from the scope of present technology, and it is understood that such equivalent embodiments, are to be included herein. Further, all references cited herein are hereby incorporated by reference in their entirety, as if fully set forth herein.

Example 1: Increasing LAMA1 Expression

[0199] This example assessed the efficacy of p300R domain peptides cotransfected with gRNAs targeting the LAMA1 promoter.

[0200] The C terminal p300 recruiter domain of DUX4 (p300R) was fused to dCas9 with a 30 amino acid residue 6x5 (SGGGG) (SEQ ID NO: 81) linker. The catalytically inactive cassette of dCas9 had an NLS-2A-mCherry for peptide cleavage. The p300RdCas9-mCherry open reading frame (ORF) was 5343 base pairs (bp) with a molecular weight of 206kDa.

[0201] The expression vectors comprising nucleotide sequences encoding p300R-dCas9 were referred to as p300R-linker-dCas9-NLS-2A-mCherry (p300RdCas9). Vectors comprised a gRNA designed to guide the p300R-dCas9 peptides to the LAMA1 promoter or a p300 nucleotide sequence as a control. All vectors were transfected into C2C12 myoblasts and cells that were not transfected served as an additional control.

[0202] The gRNAs assessed included gRNA 1 (SEQ ID NO: 58), gRNA 2 (SEQ ID NO: 59), gRNA 3 (SEQ ID NO: 60), gRNA 4 (SEQ ID NO: 61), gRNA 5 (SEQ ID NO: 62), and gRNA 6 (SEQ ID NO: 63). Distances of each gRNA binding from the LAMA1 transcription start site are shown in Table 5.

Table 5: Distance from LAMA1 Transcription Start Site

gRNA	distance from TSS
gRNA 1	681
gRNA 2	1535
gRNA 3	442
gRNA 4	930
gRNA 5	1118
gRNA 6	1735

[0203] After transfections, cells were incubated for about 24 hours. To assess effects on LAMA1 expression, LAMA1 protein levels were assessed by immunofluorescence staining (FIG. 1), and LAMA1 gene expression levels were

quantified by quantitative PCR (FIG. 2). Negative controls included no transfection and transfection using a p300 peptide only.

[0204] Use of gRNAs 1-6 in combination with the p300R-dCas9 peptides increased protein levels of LAMA1 (FIG. 1.), and use of gRNAs 5 and 6 resulted in the highest LAMA1 gene expression levels.

[0205] These experiments will be repeated using gRNAs targeting SMCHD1, LDRF, MATR3, IGF1, PML, or ADIPOQ loci in place of the LAMA1 gRNAs. SMCHD1, LDRF, MATR3, IGF1, PML, and ADIPOQ protein levels and gene expression changes will be assessed.

Example 2: Assessing Gene Expression in Hepatocytes and Myoblasts

[0206] To assess the systems and methods of the present technology in hepatocytes, alpha mouse liver 12 (AML12) cells were transfected with a p300R-DBD peptide of the present technology and a gRNA selected from an ADIPOQ-targeting gRNA: GACCCAAGCTGGATTAAACC (SEQ ID NO: 68) (ADIPOQ gRNA 3) and an IGF-1-targeting gRNA: TTAGCAGACATTAAATACTG (SEQ ID NO: 76) (IGF1 gRNA 8). This demonstrated that the compositions in accordance with the present technology successfully increased ADIPOQ and IGF1 gene expression levels, relative to untreated cells (UT), cells administered peptide without a gRNA (-g), and scrambled gRNA controls (scramble) (FIGS. 3A-FIG. 4).

IGF-1 gene expression levels were next assessed across three replicates in murine hepatocytes (AML12 cells) (FIG. 5A) and murine myoblast (C2C12 cells) cells (FIG. 5B) when transfected with a p300R-DBD peptide of the present technology and an IGF-1-targeting gRNA, compared to controls (LAMA1 g9: GTCCGAAGGGCCTGCGCACC (SEQ ID NO: 64); IGF-1 g1: IGF1 g1: CCAGGGTAAACTAGATCGAA (SEQ ID NO: 69); IGF1 g2: ATCCATGGGGCAGCGTAAAG (SEQ ID NO: 70); IGF1 g5: GTATTCCAAATAATACCAGC (SEQ ID NO: 73); IGF1 g6: GCTCCAGTTTTTAAGAGCGA (SEQ ID NO: 74); IGF1 g7: GTGTGCTAATACACCTCACC (SEQ ID NO: 75)). This demonstrated that the compositions in accordance with the present technology successfully increased IGF1 gene expression levels in both cell types, relative to untreated cells (UT), cells administered peptide without a gRNA (-gRNA), scrambled gRNA controls (scramble), and cells administered a non-IGF1-targeting gRNA (LAMA1 g9). This was repeated with

twice the gRNA amount, demonstrating that doubled gRNA content increases IGF1 gene expression relative to lower amounts with some gRNAs of the present technology (FIGS. 7A and 7B).

Example 3: Assessing Impact of Multiple gRNAs

[0207] To assess impact of multiple gRNAs in the methods of the present technology, AML12 and C2C12 cells were each transfected with a p300R-DBD peptide and one of the following IGF-1-targeting gRNA combinations: GCTCCAGTTTTTAAGAGCGA (SEQ ID NO: 74) (IGF1 gRNA 6; "g6") alone; TTAGCAGACATTAAATACTG (SEQ ID NO: 76) (IGF1 gRNA 8; "g8") alone; CCAGGGTAAACTAGATCGAA (SEQ ID NO: 69) (IGF1 gRNA 1; "g1") in combination with ATCCATGGGGCAGCGTAAAG (SEQ ID NO: 70) (IGF1 gRNA 2; "g2"); g1 in combination with g6; g1 in combination with g8; g2 in combination with g6; g2 in combination with g8; g6 in combination with g8; g1 in combination with g2andn g6; g1 in combination with g2 and g8; g1 in combination with g6 and g8; g2 in combination with g6 and g8; and g1 in combination with g2, g6, and g8. While each of these successfully increased IGF1 gene expression in both AML12 cells and C2C12 cells, some gRNA combinations increased LAMA1 gene expression more than individual gRNA use (FIG. 6).

Example 4: Assessing Gene Expression Between Cell Types

[0208] To further compare results of Example 2 and 3 of the present technology between cells, AML12 cells and C2C12 cells (murine myocytes) were each transfected with a p300R-DBD peptide of the present technology in combination with an IGF1 gRNA (e.g., g1, g2, g3, g4, g5, g6, g7, or g8 of Example 3) or a LAMA1 gRNA (GTCCGAAGGGCCTGCGCACC (SEQ ID NO: 64); LAMA1 gRNA9). All gene expression levels were assessed relative to GAPDH (FIG. 7E) and were assessed using two different analysis machines (Replicate 1: Applied Biosystems™ 7300 Real Time PCR System and Replicate 2: Bio-Rad™ CFX384 Touch Real-Time PCR Detection System). This demonstrated that IGF1 increased in both cell types and LAMA1 was most increased in C2C12s (FIGS. 8A and 8B, FIG. 9). Trends were replicated with different machine types.

[0209] This analysis was repeated with an additional screen, which assessed LAMA1 and IGF1 gene expression in both AML12 and C2C12 cells after day 1 and day

3 of transfection. Similarly, IGF1 was increased in both cell types and LAMA1 was increased in the assessed C2C12 cells (FIGS. 10A and 10B and FIGS. 11A and 11B).

Example 5: Assessing Composition Mechanisms

[0210] To further assess mechanisms of the compositions of the present technology, LAMA1 and LAMA2 gene expression levels at day 1 (d1) and day 3 (d3) were measured in C2C12 cells transfected with a p300R-DBD peptide (3R-ON) and LAMA1-targeting gRNA, GTCCGAAGGGCCTGCGCACC (SEQ ID NO: 64) (LAMA1 gRNA9; "g9") at different p300R-DBD or gRNA doses (FIGS. 12-13B) or an alternate domain peptide (VP-ON) comprising a domain encoded by the nucleotide sequence GACGCATTGGACGATTTTGATCTGGATATGCTGGGAAGTGACGCCCTCGATGAT TTTGACCTTGACATGCTTGGTTCGGATGCCCTTGATGACTTTGACCTCGACATGC TCGGCAGTGACGCCCTTGATGATTTTGACCTGGACATGCTG (SEQ ID NO: 82), and the same LAMA1-targeting gRNA in cells treated with a p300 inhibitor (iP300w), normalized to GAPDH gene expression levels (FIGS. 14A-15C). These results showed that the p300R-DBD peptides may be distinct from other potential gene activators due to the iP300w response.

[0211] Next, C2C12 cells were transfected with p300R-DBD peptide and a gRNA. gRNAs used included GTCCGAAGGGCCTGCGCACC (SEQ ID NO: 64) (LAMA1 gRNA9; "g9"); GGGCGCGGCCAGGCAGGCGG (SEQ ID NO: 65) (LAMA1 gRNA10; "g10"); CCCTGAATGCTCACCCCAAC (SEQ ID NO: 66) (LAMA1 gRNA11; "g11"); and GCGCCCAGGTCTGTCCTCCA (SEQ ID NO: 67) (LAMA1 gRNA12; "g12"). Cells were assessed for LAMA1 gene expression levels, normalized to GAPDH expression (FIGS. 16A and 16B). This demonstrated successful LAMA1 gene expression activation in murine myoblasts transfected with the compositions of the present technology.

[0212] LAMA1 levels were next measured in C2C12 cells transfected with p300R peptide and g9 over 5 days (FIGS. 17A-17D). LAMA1 gene expression changes were assessed also assessed the same C2C12 protocol of FIGS. 14A-15C, but cells were treated with 0.1 µM, 1.0 µM, or 10 µM of a histone deacetylase inhibitor (HDACi) (FIGS. 17A-17D) or with 0.1 µM, 1.0 µM, or 10 µM of a p300 inhibitor (iP300w) (FIGS. 18A and 18B). This demonstrated that LAMA1 gene expression levels trended similarly regardless of whether an inhibitor was used.

Additional Embodiments

[0213] Various embodiments of the present technology are set forth below in paragraphs [0214] to [0280]:

[0214] 1. An engineered, non-naturally occurring peptide comprising or consisting of a p300 recruiter domain of double homeobox, 4 (DUX4) (p300R domain).

[0215] 2. An engineered, non-naturally occurring peptide comprising or consisting of an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 100% identical to SEQ ID NO: 1.

[0216] 3. An engineered, non-naturally occurring peptide comprising or consisting of a p300 recruiter domain (p300R domain) and a DNA binding domain (DBD) peptide.

[0217] 4. An engineered, non-naturally occurring peptide comprising or consisting of a p300 recruiter domain (p300R domain), a DNA binding domain (DBD), and a linker peptide.

[0218] 5. A composition comprising an engineered, non-naturally occurring peptide comprising or consisting of a p300 recruiter domain (p300R domain).

[0219] 6. A composition comprising an engineered, non-naturally occurring peptide comprising a p300 recruiter domain (p300R domain) and a DNA binding domain (DBD) peptide.

[0220] 7. A composition comprising an engineered, non-naturally occurring peptide comprising a p300 recruiter domain (p300R domain), a DNA binding domain (DBD), and a linker peptide.

[0221] 8. A composition comprising an engineered, non-naturally occurring peptide comprising or consisting of an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 100% identical to SEQ ID NO: 1.

[0222] 9. A composition comprising (i) a peptide comprising a DNA binding domain (DBD) fused to a p300R domain (p300R-DBD), the p300R-DBD protein selectively binding to a locus that is altered by methylation of one or more histones and (ii) one or more guide RNA (gRNA).

[0223] 10. The engineered peptide of embodiments 3 or 4 or the composition of embodiment 2, wherein the DBD comprises a clustered regularly interspaced short palindromic repeats associated protein (Cas) domain, a zinc finger domain, a helix-turn-helix (HTH) domain, a leucine zipper domain, and a basic helix-loop-helix (bHLH) domain.

[0224] 11. The engineered, non-naturally occurring peptide or the composition of embodiment 10, wherein the Cas domain comprises an inactive nuclease domain.

[0225] 12. The engineered, non-naturally occurring peptide or the composition of embodiment 11, wherein the Cas domain comprising an inactive nuclease domain is a dead Cas (dCas) domain.

[0226] 13. A composition comprising (i) a peptide comprising a dCas domain fused to a p300R domain (p300R-dCas), the p300R-dCas9 peptide selectively binding to a locus that is altered by methylation of one or more histones and (ii) a gRNA.

[0227] 14. The composition of any one of embodiments 5-13, wherein the locus comprises a nucleotide sequence encoding a gene.

[0228] 15. The composition of any one of embodiments 5-13, wherein the locus comprises a nucleotide sequence that is a cis regulatory element of a gene.

[0229] 16. The composition of any one of embodiments 5-13, wherein the locus comprises a nucleotide sequence that is a trans regulatory element of a gene.

[0230] 17. The composition of any one of embodiments 14-16, wherein the gene is LAMA1.

[0231] 18. The composition of any one of embodiments 14-16, wherein the gene is selected from the group consisting of SMCHD1, LDRF, MATR3, IGF1, LAMA1, and PML

[0232] 19. The composition of any one of embodiments 5-15, wherein the locus comprises or consists of a promoter.

[0233] 20. The composition of embodiment 19, wherein the promoter is a LAMA1 promoter.

[0234] 21. A composition comprising (i) a peptide comprising a p300R-dCas peptide, the p300R-dCas9 peptide selectively binding to a locus altered by methylation of one or more histones and comprising at least a portion of a LAMA1 promoter and (ii) a gRNA.

[0235] 22. The engineered, non-naturally occurring peptide or the composition of any one of embodiments 10-21, wherein the Cas domain or the dCas domain comprises a Cas1 domain, a Cas2 domain, a Cas3 domain, Cas9 domain, a Cas10 domain, a Cas12a domain, or a Cas13 domain.

[0236] 23. The engineered, non-naturally occurring peptide or the composition of any one of embodiments 10-22, wherein the Cas domain or the dCas domain comprises an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 100% identical to the amino acid sequence of any one of SEQ ID NOS: 34 or 40-57.

[0237] 24. The composition of any one of embodiments 5-23, wherein the gRNA specifically directs the p300R-DBD protein to the locus.

[0238] 25. The composition of embodiment 24, wherein the gRNA mediates selective binding of the p300R-DBD protein, the p300R-Cas peptide, or the p300R-dCas peptide to the locus.

[0239] 26. The composition of any one of embodiments 5-25, wherein the gRNA comprises a nucleotide sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 100% identical to one or more of SEQ ID NOs: 58-80.

[0240] 27. The engineered, non-naturally occurring peptide or the composition of any one of embodiments 1-26, wherein the p300R domain comprises or consists of an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 100% identical to SEQ ID NO: 1.

[0241] 28. The engineered, non-naturally occurring peptide or the composition of any one of embodiments 3-27, wherein the p300R domain is fused to the DBD domain, the Cas domain, or the dCas domain by a peptide linker.

[0242] 29. The engineered, non-naturally occurring peptide or the composition of embodiment 28, wherein the peptide linker comprises an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 100% identical to the amino acid sequence of any one of SEQ ID NOs: 15-33 or 81.

[0243] 30. The composition of any one of embodiments 5-18, wherein the locus altered by the methylation of the one or more histones has a reduction in expression of the gene by at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100% relative to the control.

[0244] 31. The composition of any one of embodiments 5-30, wherein the locus altered by the methylation of the one or more histones is repressed by the methylation.

[0245] 32. The composition of any one of embodiments 5-31, wherein the one or more histones are selected from the group consisting of H2A, H2B, H3, and H4.

[0246] 33. The composition of any one of embodiments 5-32, wherein the one or more histones are methylated at a lysine (K) residue or an arginine (R) residue.

[0247] 34. The composition of embodiment 33, wherein the lysine residue is a K4, K5, K8, K9, K14, K16, K18, K20, K23, K27, K36, K56, K119, or a K120 residue.

[0248] 35. The composition of embodiment 33, wherein the arginine residue is an R2, R3, R8, R17, R23, or an R26 residue.

[0249] 36. A host cell comprising The engineered, non-naturally occurring peptide or the composition of any one of embodiments 1-35.

[0250] 37. The host cell of embodiment 36, wherein the host cell is a mammalian host cell.

[0251] 38. The host cell of embodiment 37, wherein the mammalian host cell is a human host cell or a murine host cell.

[0252] 39. The host cell of embodiment 38, wherein the human host cell is a myoblast.

[0253] 40. The composition of embodiment 38, wherein the mammalian cell is a kidney cell, a liver cell, or a muscle cell.

[0254] 41. A vector comprising a nucleotide sequence encoding the engineered peptide, the peptide, or the gRNA of any one of embodiments 1-35.

[0255] 42. The vector of embodiment 41, wherein the vector is a viral vector.

[0256] 43. The vector of embodiment 42, wherein the viral vector is a lentiviral vector.

[0257] 44. The vector of embodiment 42 or 43, wherein the vector comprises a detectable marker or a reporter gene.

[0258] 45. The vector of any one of embodiments 41-44, wherein the vector comprises an antibiotic-inducible gene expression system.

[0259] 46. The vector of embodiment 45, wherein the antibiotic is doxycycline or tetracycline.

[0260] 47. The vector of embodiment 45 or 46, wherein the antibiotic-inducible gene expression system is a Tet-ON system.

[0261] 48. The composition of any one of embodiments 5-35, wherein the composition is present in a delivery vehicle.

[0262] 49. The vector of any one of embodiments 41-47, wherein the vector is present in a delivery vehicle.

[0263] 50. The composition of embodiment 48 or the vector of embodiment 40, wherein the delivery vehicle comprises a nanoparticle.

[0264] 51. The composition or the vector of embodiment 50, wherein the nanoparticle is a lipid nanoparticle.

[0265] 52. A pharmaceutical composition comprising the composition or the vector of any one of embodiments 48-51.

[0266] 53. The pharmaceutical composition of embodiment 52, wherein the pharmaceutical composition is formulated for delivery to a subject in need thereof.

[0267] 54. The pharmaceutical composition of embodiment 53, wherein the subject is mammalian.

[0268] 55. The pharmaceutical composition of embodiment 54, wherein the subject is human.

[0269] 56. A method of increasing a level of Laminin Subunit Alpha 1 (LAMA1), Growth Factor 1 (IGF1), or adiponectin (ADIPOQ) gene or protein expression in a subject in need thereof relative to a control, the method comprising, administering to the subject:

(i) the engineered, non-naturally occurring peptide or the composition of any one of embodiments 1-35;

(ii) the host cell of any one of embodiments 36-39;

(iii) the vector of any one of embodiments 41-47; or

(iv) the pharmaceutical composition of any one of embodiments 52-55.

[0270] 57. The method of embodiment 56, wherein the level of LAMA1, IGF1, or ADIPOQ gene or protein expression is increased at least about 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 200%, 300%, 400%, 500%, 600%, 700%, 800%, 900%, or 1000%, relative to a control.

[0271] 58. A method of increasing a level of LAMA1, IGF1, or ADIPOQ protein activity in a subject in need thereof relative to a control, the method comprising, administering to the subject:

(i) the engineered, non-naturally occurring peptide or the composition of any one of embodiments 1-35;

(ii) the host cell of any one of embodiments 36-39;

(iii) the vector of any one of embodiments 41-47; or

(iv) the pharmaceutical composition of any one of embodiments 52-55.

[0272] 59. The method of any one of embodiments 56-58, wherein the composition, the vector, the pharmaceutical composition, or the host cell is administered to the subject at a dose of at least about 0.5 ng to at least about 500 µg.

[0273] 60. The method of any one of embodiments 56-59, wherein the vector comprises a detectable marker or a reporter gene, and a locus that is altered by methylation of one or more histones is altered when the reporter gene or the detectable marker is not expressed at a detectable level.

[0274] 61. The method of any one of embodiments 58-60, wherein the level of LAMA1, IGF1, or ADIPOQ protein activity is increased at least about 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 200%, 300%, 400%, 500%, 600%, 700%, 800%, 900%, or 1000%, relative to a control.

[0275] 62. The method of any one of embodiments 58-61, wherein an expression level of a gene encoding one or more proteins that increases LAMA1, IGF1, or ADIPOQ expression or activity is increased by at least about 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 200%, 300%, 400%, 500%, 600%, 700%, 800%, 900%, 1000%, or more, relative to a control.

[0276] 63. The method of any one of embodiments 58-61, wherein an expression level of a gene encoding one or more proteins that reduces LAMA1, IGF1, or ADIPOQ expression or activity is reduced by at least about 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100% relative to a control.

[0277] 64. The host cell of any one of embodiments 36-39, wherein expression or activity of LAMA1, IGF1, or ADIPOQ is increased compared to a control cell.

[0278] 65. The host cell of embodiment 64, wherein a level of LAMA1, IGF1, or ADIPOQ gene or protein expression is increased at least about 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 200%, 300%, 400%, 500%, 600%, 700%, 800%, 900%, 1000%, or more, relative to a control.

[0279] 66. The host cell of embodiment 64 or 65, wherein an expression level of a gene encoding one or more proteins that increases LAMA1, IGF1, or ADIPOQ expression or activity is increased by at least about 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 200%, 300%, 400%, 500%, 600%, 700%, 800%, 900%, 1000%, or more, relative to the control cell.

[0280] 67. The host cell of embodiment 64 or 65, wherein an expression level of a gene encoding one or more proteins that represses LAMA1, IGF1, or ADIPOQ expression or activity is reduced by at least about 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100% relative to the control cell.

CLAIMS

1. An engineered, non-naturally occurring peptide comprising or consisting of a p300 recruiter domain of double homeobox, 4 (DUX4) (p300R domain).

2. An engineered, non-naturally occurring peptide comprising or consisting of an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 100% identical to SEQ ID NO: 1.

3. An engineered, non-naturally occurring peptide comprising or consisting of a p300 recruiter domain (p300R domain) and a DNA binding domain (DBD) peptide.

4. An engineered, non-naturally occurring peptide comprising or consisting of a p300 recruiter domain (p300R domain), a DNA binding domain (DBD), and a linker peptide.

5. A composition comprising an engineered, non-naturally occurring peptide comprising or consisting of a p300 recruiter domain (p300R domain).

6. A composition comprising an engineered, non-naturally occurring peptide comprising a p300 recruiter domain (p300R domain) and a DNA binding domain (DBD) peptide.

7. A composition comprising an engineered, non-naturally occurring peptide comprising a p300 recruiter domain (p300R domain), a DNA binding domain (DBD), and a linker peptide.

8. A composition comprising an engineered, non-naturally occurring peptide comprising or consisting of an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 100% identical to SEQ ID NO: 1.

9. A composition comprising (i) a peptide comprising a DNA binding domain (DBD) fused to a p300R domain (p300R-DBD), the p300R-DBD protein selectively binding to a locus that is altered by methylation of one or more histones and (ii) one or more guide RNA (gRNA).

10. The engineered peptide of claims 3 or 4 or the composition of claim 2, wherein the DBD comprises a clustered regularly interspaced short palindromic repeats associated protein (Cas) domain, a zinc finger domain, a helix-turn-helix (HTH) domain, a leucine zipper domain, and a basic helix-loop-helix (bHLH) domain.

11. The engineered, non-naturally occurring peptide or the composition of claim 10, wherein the Cas domain comprises an inactive nuclease domain.

12. The engineered, non-naturally occurring peptide or the composition of claim 11, wherein the Cas domain comprising an inactive nuclease domain is a dead Cas (dCas) domain.

13. A composition comprising (i) a peptide comprising a dCas domain fused to a p300R domain (p300R-dCas), the p300R-dCas9 peptide selectively binding to a locus that is altered by methylation of one or more histones and (ii) a gRNA.

14. The composition of any one of claims 5-13, wherein the locus comprises a nucleotide sequence encoding a gene.

15. The composition of any one of claims 5-13, wherein the locus comprises a nucleotide sequence that is a cis regulatory element of a gene.

16. The composition of any one of claims 5-13, wherein the locus comprises a nucleotide sequence that is a trans regulatory element of a gene.

17. The composition of any one of claims 14-16, wherein the gene is LAMA1.

18. The composition of any one of claims 14-16, wherein the gene is selected from the group consisting of SMCHD1, LDRF, MATR3, IGF1, LAMA1, and PML

19. The composition of any one of claims 5-15, wherein the locus comprises or consists of a promoter.

20. The composition of claim 19, wherein the promoter is a LAMA1 promoter.

21. A composition comprising (i) a peptide comprising a p300R-dCas peptide, the p300R-dCas9 peptide selectively binding to a locus altered by methylation of one or more histones and comprising at least a portion of a LAMA1 promoter and (ii) a gRNA.

22. The engineered, non-naturally occurring peptide or the composition of any one of claims 10-21, wherein the Cas domain or the dCas domain comprises a Cas1 domain, a Cas2 domain, a Cas3 domain, Cas9 domain, a Cas10 domain, a Cas12a domain, or a Cas13 domain.

23. The engineered, non-naturally occurring peptide or the composition of any one of claims 10-22, wherein the Cas domain or the dCas domain comprises an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 100% identical to the amino acid sequence of any one of SEQ ID NOS: 34 or 40-57.

24. The composition of any one of claims 5-23, wherein the gRNA specifically directs the p300R-DBD protein to the locus.

25. The composition of claim 24, wherein the gRNA mediates selective binding of the p300R-DBD protein, the p300R-Cas peptide, or the p300R-dCas peptide to the locus.

26. The composition of any one of claims 5-25, wherein the gRNA comprises a nucleotide sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 100% identical to one or more of SEQ ID NOs: 58-80.

27. The engineered, non-naturally occurring peptide or the composition of any one of claims 1-26, wherein the p300R domain comprises or consists of an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 100% identical to SEQ ID NO: 1.

28. The engineered, non-naturally occurring peptide or the composition of any one of claims 3-27, wherein the p300R domain is fused to the DBD domain, the Cas domain, or the dCas domain by a peptide linker.

29. The engineered, non-naturally occurring peptide or the composition of claim 28, wherein the peptide linker comprises an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 100% identical to the amino acid sequence of any one of SEQ ID NOs: 15-33 or 81.

30. The composition of any one of claims 5-18, wherein the locus altered by the methylation of the one or more histones has a reduction in expression of the gene by at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100% relative to the control.

31. The composition of any one of claims 5-30, wherein the locus altered by the methylation of the one or more histones is repressed by the methylation.

32. The composition of any one of claims 5-31, wherein the one or more histones are selected from the group consisting of H2A, H2B, H3, and H4.

33. The composition of any one of claims 5-32, wherein the one or more histones are methylated at a lysine (K) residue or an arginine (R) residue.

34. The composition of claim 33, wherein the lysine residue is a K4, K5, K8, K9, K14, K16, K18, K20, K23, K27, K36, K56, K119, or a K120 residue.

35. The composition of claim 33, wherein the arginine residue is an R2, R3, R8, R17, R23, or an R26 residue.

36. A host cell comprising The engineered, non-naturally occurring peptide or the composition of any one of claims 1-35.

37. The host cell of claim 36, wherein the host cell is a mammalian host cell.

38. The host cell of claim 37, wherein the mammalian host cell is a human host cell or a murine host cell.

39. The host cell of claim 38, wherein the human host cell is a myoblast.

40. The composition of claim 38, wherein the mammalian cell is a kidney cell, a liver cell, or a muscle cell.

41. A vector comprising a nucleotide sequence encoding the engineered peptide, the peptide, or the gRNA of any one of claims 1-35.

42. The vector of claim 41, wherein the vector is a viral vector.

43. The vector of claim 42, wherein the viral vector is a lentiviral vector.

44. The vector of claim 42 or 43, wherein the vector comprises a detectable marker or a reporter gene.

45. The vector of any one of claims 41-44, wherein the vector comprises an antibiotic-inducible gene expression system.

46. The vector of claim 45, wherein the antibiotic is doxycycline or tetracycline.

47. The vector of claim 45 or 46, wherein the antibiotic-inducible gene expression system is a Tet-ON system.

48. The composition of any one of claims 5-35, wherein the composition is present in a delivery vehicle.

49. The vector of any one of claims 41-47, wherein the vector is present in a delivery vehicle.

50. The composition of claim 48 or the vector of claim 40, wherein the delivery vehicle comprises a nanoparticle.

51. The composition or the vector of claim 50, wherein the nanoparticle is a lipid nanoparticle.

52. A pharmaceutical composition comprising the composition or the vector of any one of claims 48-51.

53. The pharmaceutical composition of claim 52, wherein the pharmaceutical composition is formulated for delivery to a subject in need thereof.

54. The pharmaceutical composition of claim 53, wherein the subject is mammalian.

55. The pharmaceutical composition of claim 54, wherein the subject is human.

56. A method of increasing a level of Laminin Subunit Alpha 1 (LAMA1), Growth Factor 1 (IGF1), or adiponectin (ADIPOQ) gene or protein expression in a subject in need thereof relative to a control, the method comprising, administering to the subject:

(i) the engineered, non-naturally occurring peptide or the composition of any one of claims 1-35;

(ii) the host cell of any one of claims 36-39;

(iii) the vector of any one of claims 41-47; or

(iv) the pharmaceutical composition of any one of claims 52-55.

57. The method of claim 56, wherein the level of LAMA1, IGF1, or ADIPOQ gene or protein expression is increased at least about 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 200%, 300%, 400%, 500%, 600%, 700%, 800%, 900%, or 1000%, relative to a control.

58. A method of increasing a level of LAMA1, IGF1, or ADIPOQ protein activity in a subject in need thereof relative to a control, the method comprising, administering to the subject:

(i) the engineered, non-naturally occurring peptide or the composition of any one of claims 1-35;

(ii) the host cell of any one of claims 36-39;

(iii) the vector of any one of claims 41-47; or

(iv) the pharmaceutical composition of any one of claims 52-55.

59. The method of any one of claims 56-58, wherein the composition, the vector, the pharmaceutical composition, or the host cell is administered to the subject at a dose of at least about 0.5 ng to at least about 500 µg.

60. The method of any one of claims 56-59, wherein the vector comprises a detectable marker or a reporter gene, and a locus that is altered by methylation of one or more histones is altered when the reporter gene or the detectable marker is not expressed at a detectable level.

61. The method of any one of claims 58-60, wherein the level of LAMA1, IGF1, or ADIPOQ protein activity is increased at least about 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 200%, 300%, 400%, 500%, 600%, 700%, 800%, 900%, or 1000%, relative to a control.

62. The method of any one of claims 58-61, wherein an expression level of a gene encoding one or more proteins that increases LAMA1, IGF1, or ADIPOQ expression or activity is increased by at least about 5%, 10%, 15%, 20%, 30%, 40%,

50%, 60%, 70%, 80%, 90%, 100%, 200%, 300%, 400%, 500%, 600%, 700%, 800%, 900%, 1000%, or more, relative to a control.

63. The method of any one of claims 58-61, wherein an expression level of a gene encoding one or more proteins that reduces LAMA1, IGF1, or ADIPOQ expression or activity is reduced by at least about 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100% relative to a control.

64. The host cell of any one of claims 36-39, wherein expression or activity of LAMA1, IGF1, or ADIPOQ is increased compared to a control cell.

65. The host cell of claim 64, wherein a level of LAMA1, IGF1, or ADIPOQ gene or protein expression is increased at least about 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 200%, 300%, 400%, 500%, 600%, 700%, 800%, 900%, 1000%, or more, relative to a control.

66. The host cell of claim 64 or 65, wherein an expression level of a gene encoding one or more proteins that increases LAMA1, IGF1, or ADIPOQ expression or activity is increased by at least about 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 200%, 300%, 400%, 500%, 600%, 700%, 800%, 900%, 1000%, or more, relative to the control cell.

67. The host cell of claim 64 or 65, wherein an expression level of a gene encoding one or more proteins that represses LAMA1, IGF1, or ADIPOQ expression or activity is reduced by at least about 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100% relative to the control cell.