

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

(19) World Intellectual Property Organization

International Bureau



(10) International Publication Number

WO 2018/071565 A1

(43) International Publication Date

19 April 2018 (19.04.2018)

(51) International Patent Classification:

A61K 35/12 (2015.01)

EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR, OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(21) International Application Number:

PCT/US2017/056178

(22) International Filing Date:

11 October 2017 (11.10.2017)

(25) Filing Language:

English

(26) Publication Language:

English

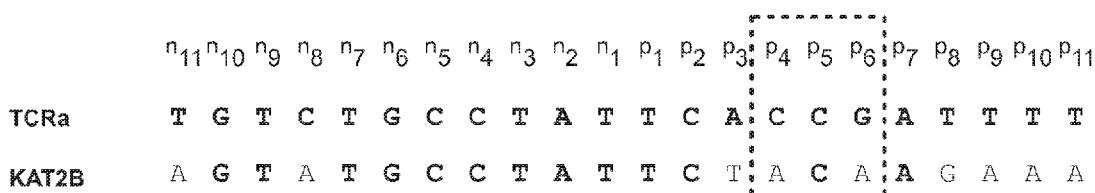
(30) Priority Data:

62/406,689 11 October 2016 (11.10.2016) US  
62/414,266 28 October 2016 (28.10.2016) US**Declarations under Rule 4.17:**

- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))
- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))

**Published:**

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))

(71) **Applicant:** BLUEBIRD BIO, INC. [US/US]; 60 Binney Street, Cambridge, Massachusetts 02142 (US).(72) **Inventors:** JARJOUR, Jordan; 11014 37th Ave SW, Seattle, Washington 98146 (US). POGSON, Mark; 3122 SE 16th St., North Bend, Washington 98045 (US).(74) **Agent:** MASHRUWALA, Mary Anne et al.; Cooley LLP, 1299 Pennsylvania Avenue, NW, Suite 700, Washington, District of Columbia 20004 (US).(81) **Designated States** (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.(84) **Designated States** (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK,(54) **Title:** TCR $\alpha$  HOMING ENDONUCLEASE VARIANTS**Figure 1**(57) **Abstract:** The present disclosure provides improved genome editing compositions and methods for editing a TCR $\alpha$  gene.

**TCR $\alpha$  HOMING ENDONUCLEASE VARIANTS****CROSS REFERENCE TO RELATED APPLICATIONS**

This application claims the benefit under 35 U.S.C. § 119(e) of U.S. Provisional Application No. 62/414,266, filed October 28, 2016, and U.S. Provisional Application No. 5 62/406,689, filed October 11, 2016, each of which is incorporated by reference herein in its entirety.

**STATEMENT REGARDING SEQUENCE LISTING**

The Sequence Listing associated with this application is provided in text format in lieu of a paper copy, and is hereby incorporated by reference into the specification. The 10 name of the text file containing the Sequence Listing is BLBD\_077\_02WO\_ST25.txt. The text file is 114 KB, was created on October 11, 2017, and is being submitted electronically via EFS-Web, concurrent with the filing of the specification.

**BACKGROUND****Technical Field**

15 The present disclosure relates to improved genome editing polypeptides and compositions. More particularly, the disclosure relates to nuclease variants and compositions, useful for editing the human T cell receptor alpha (TCR $\alpha$ ) gene.

**Description of the Related Art**

The global burden of cancer doubled between 1975 and 2000. Cancer is the second 20 leading cause of morbidity and mortality worldwide, with approximately 14.1 million new cases and 8.2 million cancer related deaths in 2012. The most common cancers are breast cancer, lung and bronchus cancer, prostate cancer, colon and rectum cancer, bladder cancer, melanoma of the skin, non-Hodgkin lymphoma, thyroid cancer, kidney and renal pelvis cancer,

endometrial cancer, leukemia, and pancreatic cancer. The number of new cancer cases is projected to rise to 22 million within the next two decades.

The immune system has a key role in detecting and combating human cancer. The majority of transformed cells are quickly detected by immune sentinels and destroyed through 5 the activation of antigen-specific T cells via clonally expressed T cell receptors (TCR).

Accordingly, cancer can be considered an immunological disorder, a failure of immune system to mount the necessary anti-tumor response to durably suppress and eliminate the disease. In order to more effectively combat cancer, certain immunotherapy interventions developed over the last few decades have specifically focused on enhancing T cell immunity. These treatments 10 have yielded only sporadic cases of disease remission, and have not had substantial overall success.

Most recently, adoptive cellular immunotherapy strategies, which are based on the isolation, modification, expansion and reinfusion of T cells, have been explored and tested in early stage clinical trials. T cells have often been the effector cells of choice for cancer 15 immunotherapy due to their selective recognition and powerful effector mechanisms. These treatments have shown mixed rates of success, but a small number of patients have experienced durable remissions, highlighting the as-yet unrealized potential for T cell-based immunotherapies.

Thus, there is a need for more effective, targeted, safer, and persistent therapies to treat 20 various forms of cancer and other immune disorders. In addition, there is a need for methods and compositions that can precisely and reproducibly disrupt endogenous TCR genes with high efficiency. Today's standards of care for most cancers fall short in some or all of these criteria.

## BRIEF SUMMARY

The present disclosure generally relates, in part, to homing endonuclease (HE) variants 25 and megaTALs that cleave a target site in exon 1 of the human TCR $\alpha$  gene, compositions comprising the HE variants and megaTALs, and methods of using the same.

In various embodiments, the present disclosure contemplates, in part, a polypeptide comprising an I-OnuI homing endonuclease (HE) variant that cleaves a target site in the human T cell receptor alpha (TCR $\alpha$ ) gene, wherein the variant comprises the amino acid substitutions:

L26I, R28D, N32R, K34N, S35E, V37N, G38R, S40R, E42S, G44R, V68K, A70T, G73S, N75R, S78M, K80R, L138M, T143N, S159P, S176A, C180H, F182G, I186K, S188V, S190G, K191T, L192A, G193K, Q195Y, Q197G, V199R, S201A, T203S, K207R, Y223S, K225R, S233R, D236E, and V238E of any one of SEQ ID NOs: 1-5, or a biologically active fragment thereof.

5 In various embodiments, the present disclosure contemplates, in part, a polypeptide comprising an I-OnuI homing endonuclease (HE) variant that cleaves a target site in the human T cell receptor alpha (TCR $\alpha$ ) gene, wherein the variant comprises the amino acid substitutions: L26I, R28D, N32R, K34N, S35E, V37N, G38R, S40R, E42S, G44R, V68K, A70T, G73S, 10 N75R, S78M, K80R, L138M, T143N, S159P, S176A, E178D, C180H, F182G, I186K, S188V, S190G, K191T, L192A, G193K, Q195Y, Q197G, V199R, S201A, T203S, K207R, Y223S, K225R, S233R, D236E, and V238E of any one of SEQ ID NOs: 1-5, or a biologically active fragment thereof.

15 In particular embodiments, the biologically active fragment lacks the 1, 2, 3, 4, 5, 6, 7, or 8 N-terminal amino acids compared to a corresponding wild type I-OnuI HE.

In certain embodiments, the biologically active fragment lacks the 4 N-terminal amino acids compared to a corresponding wild type I-OnuI HE.

In further embodiments, the biologically active fragment lacks the 8 N-terminal amino acids compared to a corresponding wild type I-OnuI HE.

20 In some embodiments, the biologically active fragment lacks the 1, 2, 3, 4, or 5 C-terminal amino acids compared to a corresponding wild type I-OnuI HE.

In additional embodiments, the biologically active fragment lacks the C-terminal amino acid compared to a corresponding wild type I-OnuI HE.

25 In particular embodiments, the biologically active fragment lacks the 2 C-terminal amino acids compared to a corresponding wild type I-OnuI HE.

In particular embodiments, the I-OnuI HE variant comprises an amino acid sequence that is at least 99% identical to the amino acid sequence set forth in any one of SEQ ID NOs: 7-8, or a biologically active fragment thereof.

30 In additional embodiments, the HE variant comprises the amino acid sequence set forth in SEQ ID NO: 7, or a biologically active fragment thereof.

In certain embodiments, the HE variant comprises the amino acid sequence set forth in SEQ ID NO: 8, or a biologically active fragment thereof.

In some embodiments, the polypeptide binds the polynucleotide sequence set forth in SEQ ID NO: 17.

5 In particular embodiments, the polypeptide further comprises a DNA binding domain.

In further embodiments, the DNA binding domain is selected from the group consisting of: a TALE DNA binding domain and a zinc finger DNA binding domain.

10 In additional embodiments, the TALE DNA binding domain comprises about 9.5 TALE repeat units, about 10.5 TALE repeat units, about 11.5 TALE repeat units, about 12.5 TALE repeat units, about 13.5 TALE repeat units, about 14.5 TALE repeat units, or about 15.5 TALE repeat units.

In some embodiments, the TALE DNA binding domain comprises 11.5 TALE repeat units and binds the polynucleotide sequence set forth in SEQ ID NO: 19.

15 In particular embodiments, the TALE DNA binding domain comprises 10.5 TALE repeat units and binds the polynucleotide sequence set forth in SEQ ID NO: 18.

In certain embodiments, the polypeptide binds and cleaves the polynucleotide sequence set forth in SEQ ID NO: 20.

In particular embodiments, the zinc finger DNA binding domain comprises 2, 3, 4, 5, 6, 7, or 8 zinc finger motifs.

20 In additional embodiments, the polypeptide further comprises a peptide linker and an end-processing enzyme or biologically active fragment thereof.

In particular embodiments, the polypeptide further comprises a viral self-cleaving 2A peptide and an end-processing enzyme or biologically active fragment thereof.

25 In some embodiments, the end-processing enzyme or biologically active fragment thereof has 5'-3' exonuclease, 5'-3' alkaline exonuclease, 3'-5' exonuclease, 5' flap endonuclease, helicase or template-independent DNA polymerase activity.

In further embodiments, the end-processing enzyme comprises Trex2 or a biologically active fragment thereof.

30 In particular embodiments, the polypeptide comprises the amino acid sequence set forth in any one of SEQ ID NOs: 10 to 12, or a biologically active fragment thereof.

In certain embodiments, the polypeptide comprises the amino acid sequence set forth in SEQ ID NO: 10, or a biologically active fragment thereof.

In additional embodiments, the polypeptide comprises the amino acid sequence set forth in SEQ ID NO: 11, or a biologically active fragment thereof.

5 In particular embodiments, the polypeptide comprises the amino acid sequence set forth in SEQ ID NO: 12, or a biologically active fragment thereof.

In various embodiments, a polynucleotide encoding a polypeptide contemplated herein is provided.

10 In particular embodiments, the present disclosure contemplates, in part, an mRNA encoding a polypeptide contemplated herein is provided.

In particular embodiments, the mRNA comprises the sequence set forth in any one of SEQ ID NOs: 16-19.

In some embodiments, a cDNA encoding a polypeptide contemplated herein is provided.

15 In certain embodiments, a vector comprising a polynucleotide encoding a polypeptide contemplated herein is provided.

In particular embodiments, a cell comprising a polypeptide contemplated herein is provided.

20 In further embodiments, a cell comprising a polynucleotide encoding a polypeptide contemplated herein is provided.

In additional embodiments, a cell comprising a vector contemplated herein is provided.

In particular embodiments, a cell edited by a polypeptide contemplated herein is provided.

In particular embodiments, the cell is a hematopoietic cell.

25 In some embodiments, the cell is a T cell.

In certain embodiments, the cell is a CD3+, CD4+, and/or CD8+ cell.

In particular embodiments, the cell is an immune effector cell.

In further embodiments, the cell is a cytotoxic T lymphocytes (CTLs), a tumor infiltrating lymphocytes (TILs), or a helper T cells.

In certain embodiments, the cell is a natural killer (NK) cell or natural killer T (NKT) cell.

In particular embodiments, the source of the cell is peripheral blood mononuclear cells, bone marrow, lymph nodes tissue, cord blood, thymus tissue, tissue from a site of infection, 5 ascites, pleural effusion, spleen tissue, or tumors.

## BRIEF DESCRIPTION OF SEVERAL VIEWS OF THE DRAWINGS

**Figure 1** shows a comparison of on- and off-target sites (SEQ ID NOS: 17 and 21) for an I-OnuI HE variant designed to target a site in exon 1 of the TCR $\alpha$  gene constant region.

**Figure 2** shows a cartoon of the strategy for grafting a set of amino acids obtained from 10 a modular HE variant, highly selective for “CCG” trinucleotide sequence (SEQ ID NOS: 17 and 67), into a first generation TCR $\alpha$  HE variant (SEQ ID NO: 6) to generate a second generation TCR $\alpha$  HE variant (TCR $\alpha$ 2.0). The TCR $\alpha$ 2.0 HE variant retains high selectivity for “CCG” at the desired substrate positions.

**Figure 3** shows the enhanced selectivity profiles for the first generation TCR $\alpha$  HE 15 variant and the TCR $\alpha$ 2.0 HE variant at substrate positions p456. Panel (A) shows yeast surface display cleavage plots for 64 possible nucleotide combinations and panel (B) shows a heat map of these activities.

**Figure 4** shows representative data of gene editing efficiency at the TCR $\alpha$  locus using 20 CD3 surface expression as a readout. A first generation TCR $\alpha$  megaTAL (A) (SEQ ID NO: 6) shows increased gene editing efficiency in human primary T cells compared to T cells treated with a TCR $\alpha$ 2.0 megaTAL (SEQ ID NO: 7).

**Figure 5** shows that introduction of the targeted mutation (E178D) in the TCR $\alpha$ 2.0 megaTAL (B) to generate the TCR $\alpha$ 2.1 megaTAL (C) (SEQ ID NO: 8) restores on-target 25 cleavage activity of the first generation TCR $\alpha$  megaTAL (A) while retaining the on-target selectivity of the TCR $\alpha$ 2.0 megaTAL.

**Figure 6** shows the increased selectivity of the TCR $\alpha$ 2.1 megaTAL by yeast-surface-display profiling using bioinformatically derived putative off-target sites.

**Figure 7** shows the increased selectivity of the TCR $\alpha$ 2.1 megaTAL by deep sequencing of amplicons generated from genomic DNA of megaTAL treated primary T cells.

**Figure 8** shows that appending an additional TALE RVD to the TCR $\alpha$ 2.1 megaTAL to increase the TALE repeats from 10.5 to 11.5 (A) further increases on-target activity (SEQ ID NOS: 68 and 69) of the TCR $\alpha$ 2.1 megaTAL in primary T cells (B).

**Figure 9** shows that appending rationally designed minimal N-terminal TAL domains 5 derived from different *Xanthomonas* family members yields similar or, in some instances, a modest increase in activity compared to the first generation TCR $\alpha$  megaTAL containing an *X. oryzae* N-terminal TAL domain.

## BRIEF DESCRIPTION OF THE SEQUENCE IDENTIFIERS

**SEQ ID NO: 1** is an amino acid sequence of a wild type I-OnuI LAGLIDADG 10 homing endonuclease (LHE).

**SEQ ID NO: 2** is an amino acid sequence of a wild type I-OnuI LHE.

**SEQ ID NO: 3** is an amino acid sequence of a biologically active fragment of a wild-type I-OnuI LHE.

**SEQ ID NO: 4** is an amino acid sequence of a biologically active fragment of a wild-type I-OnuI LHE. 15

**SEQ ID NO: 5** is an amino acid sequence of a biologically active fragment of a wild-type I-OnuI LHE.

**SEQ ID NO: 6** is an amino acid sequence of an I-OnuI LHE variant reprogrammed to bind and cleave a target site in exon 1 of the constant region of the human TCR $\alpha$  gene.

**SEQ ID NO: 7** is an amino acid sequence of an I-OnuI LHE variant reprogrammed to bind and cleave a target site in exon 1 of the constant region of the human TCR $\alpha$  gene. 20

**SEQ ID NO: 8** is an amino acid sequence of an I-OnuI LHE variant reprogrammed to bind and cleave a target site in exon 1 of the constant region of the human TCR $\alpha$  gene.

**SEQ ID NO: 9** is an amino acid sequence of a megaTAL that binds and cleaves a 25 target site in exon 1 of the constant region of the human TCR $\alpha$  gene.

**SEQ ID NO: 10** is an amino acid sequence of a megaTAL that binds and cleaves a target site in exon 1 of the constant region of the human TCR $\alpha$  gene.

**SEQ ID NO: 11** is an amino acid sequence of a megaTAL that binds and cleaves a target site in exon 1 of the constant region of the human TCR $\alpha$  gene.

**SEQ ID NO: 12** is an amino acid sequence of a megaTAL that binds and cleaves a target site in exon 1 of the constant region of the human TCR $\alpha$  gene.

**SEQ ID NO: 13** is an mRNA sequence that encodes a megaTAL that binds and cleaves a target site in exon 1 of the constant region of the human TCR $\alpha$  gene.

5       **SEQ ID NO: 14** is an mRNA sequence that encodes a megaTAL that binds and cleaves a target site in exon 1 of the constant region of the human TCR $\alpha$  gene.

**SEQ ID NO: 15** is an mRNA sequence that encodes a megaTAL that binds and cleaves a target site in exon 1 of the constant region of the human TCR $\alpha$  gene.

10      **SEQ ID NO: 16** is an mRNA sequence that encodes a megaTAL that binds and cleaves a target site in exon 1 of the constant region of the human TCR $\alpha$  gene.

**SEQ ID NO: 17** is an I-OnuI LHE variant target site in exon 1 of the constant region of the human TCR $\alpha$  gene.

**SEQ ID NO: 18** is a 10.5 RVD TALE DNA binding domain target site in exon 1 of the constant region of the human TCR $\alpha$  gene.

15      **SEQ ID NO: 19** is an 11.5 RVD TALE DNA binding domain target site in exon 1 of the constant region of the human TCR $\alpha$  gene.

**SEQ ID NO: 20** is a megaTAL target site in exon 1 of the constant region of the human TCR $\alpha$  gene.

20      **SEQ ID NO: 21** is off-target site in the human KAT2B gene for TCR $\alpha$  nucleases that target SEQ ID NO: 17.

**SEQ ID NOs: 22-24** are amino acid sequences encoding NTD variant megaTALs.

**SEQ ID NOs: 25-27** are mRNA sequences encoding NTD variant megaTALs.

**SEQ ID NO: 28** is an amino acid sequence encoding murine Trex2.

**SEQ ID NO: 29** is an mRNA sequence encoding murine Trex2.

25      **SEQ ID NOs: 30-40** set forth the amino acid sequences of various linkers.

**SEQ ID NOs: 41-65** set forth the amino acid sequences of protease cleavage sites and self-cleaving polypeptide cleavage sites.

In the foregoing sequences, X, if present, refers to any amino acid or the absence of an amino acid.

## DETAILED DESCRIPTION

### A. OVERVIEW

The disclosure generally relates to, in part, improved genome editing compositions and methods of use thereof. Genome edited immune effector cells offer numerous advantages 5 compared to existing cell-based immunotherapies including, but not limited to, improved safety due to decreased risk of undesirable autoimmune response, precisely targeted therapy with more predictable therapeutic gene expression, increased durability in the tumor microenvironment and increased efficacy. Without wishing to be bound by any particular theory, genome editing compositions contemplated in various embodiments comprise homing 10 endonucleases and megaTALs targeting the T cell receptor alpha (TCR $\alpha$ ) gene, engineered for improved safety and increased target site specificity, selectivity and catalytic activity.

In particular embodiments, modification of one or more TCR $\alpha$  alleles ablates or substantially ablates expression of the TCR $\alpha$  allele(s), decreases expression of the TCR $\alpha$  allele(s), and/or impairs, substantially impairs, or ablates one or more functions of the TCR $\alpha$  15 allele(s) or renders the TCR $\alpha$  allele(s) non-functional. In particular embodiments, TCR $\alpha$  functions include, but are not limited to, recruiting CD3 to the cell surface, MHC dependent recognition and binding of antigen, activation of TCR $\alpha\beta$  signaling.

Genome editing compositions and methods contemplated in various embodiments comprise nuclease variants, designed to bind and cleave a target site in the human T cell 20 receptor alpha (TCR $\alpha$ ) gene. The nuclease variants contemplated in particular embodiments, can be used to introduce a double-strand break in a target polynucleotide sequence, which may be repaired by non-homologous end joining (NHEJ) in the absence of a polynucleotide template, *e.g.*, a donor repair template, or by homology directed repair (HDR), *i.e.*, homologous recombination, in the presence of a donor repair template. Nuclease variants 25 contemplated in certain embodiments, can also be designed as nickases, which generate single-stranded DNA breaks that can be repaired using the cell's base-excision-repair (BER) machinery or homologous recombination in the presence of a donor repair template. NHEJ is an error-prone process that frequently results in the formation of small insertions and deletions that disrupt gene function. Homologous recombination requires homologous DNA as a

template for repair and can be leveraged to create a limitless variety of modifications specified by the introduction of donor DNA containing the desired sequence at the target site, flanked on either side by sequences bearing homology to regions flanking the target site.

In one preferred embodiment, the genome editing compositions contemplated herein  
5 comprise a homing endonuclease variant or megaTAL that targets the human TCR $\alpha$  gene.

In various embodiments, the TCR $\alpha$  gene targeting homing endonuclease variants or megaTALs contemplated herein have increased binding site selectivity or specificity to the target site compared to existing homing endonucleases or megaTALs that target the TCR $\alpha$  gene.

10 In various embodiments, the TCR $\alpha$  gene targeting homing endonuclease variants or megaTALs contemplated herein have improved binding site selectivity or specificity to the target site while retaining high catalytic activity compared to existing homing endonucleases or megaTALs that target the TCR $\alpha$  gene.

In one preferred embodiment, the genome editing compositions contemplated herein  
15 comprise a homing endonuclease variant or megaTAL and an end-processing enzyme, *e.g.*, Trex2.

The practice of the particular embodiments will employ, unless indicated specifically to the contrary, conventional methods of chemistry, biochemistry, organic chemistry, molecular biology, microbiology, recombinant DNA techniques, genetics, immunology, and cell biology  
20 that are within the skill of the art, many of which are described below for the purpose of illustration. Such techniques are explained fully in the literature. *See e.g.*, Sambrook, *et al.*, *Molecular Cloning: A Laboratory Manual* (3rd Edition, 2001); Sambrook, *et al.*, *Molecular Cloning: A Laboratory Manual* (2nd Edition, 1989); Maniatis *et al.*, *Molecular Cloning: A Laboratory Manual* (1982); Ausubel *et al.*, *Current Protocols in Molecular Biology* (John Wiley and Sons, updated July 2008); *Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology*, Greene Pub. Associates and Wiley-Interscience; Glover, *DNA Cloning: A Practical Approach*, vol. I & II (IRL Press, Oxford, 1985); Anand, *Techniques for the Analysis of Complex Genomes*, (Academic Press, New York, 1992); *Transcription and Translation* (B. Hames & S. Higgins, Eds., 1984); Perbal, *A Practical Guide to Molecular Cloning* (1984); Harlow and Lane, *Antibodies*, (Cold Spring

Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1998) *Current Protocols in Immunology* Q. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach and W. Strober, eds., 1991); *Annual Review of Immunology*; as well as monographs in journals such as *Advances in Immunology*.

5    **B.    DEFINITIONS**

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by those of ordinary skill in the art to which the invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of particular embodiments, preferred embodiments of 10 compositions, methods and materials are described herein. For the purposes of the present disclosure, the following terms are defined below.

The articles “a,” “an,” and “the” are used herein to refer to one or to more than one (i.e., to at least one, or to one or more) of the grammatical object of the article. By way of example, “an element” means one element or one or more elements.

15    The use of the alternative (e.g., “or”) should be understood to mean either one, both, or any combination thereof of the alternatives.

The term “and/or” should be understood to mean either one, or both of the alternatives.

As used herein, the term “about” or “approximately” refers to a quantity, level, value, number, frequency, percentage, dimension, size, amount, weight or length that varies by as 20 much as 15%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2% or 1% to a reference quantity, level, value, number, frequency, percentage, dimension, size, amount, weight or length. In one embodiment, the term “about” or “approximately” refers a range of quantity, level, value, number, frequency, percentage, dimension, size, amount, weight or length  $\pm$  15%,  $\pm$  10%,  $\pm$  9%,  $\pm$  8%,  $\pm$  7%,  $\pm$  6%,  $\pm$  5%,  $\pm$  4%,  $\pm$  3%,  $\pm$  2%, or  $\pm$  1% about a reference quantity, level, 25 value, number, frequency, percentage, dimension, size, amount, weight or length.

In one embodiment, a range, e.g., 1 to 5, about 1 to 5, or about 1 to about 5, refers to each numerical value encompassed by the range. For example, in one non-limiting and merely illustrative embodiment, the range “1 to 5” is equivalent to the expression 1, 2, 3, 4, 5; or 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, or 5.0; or 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.1,

2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, 3.0, 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8, 3.9, 4.0, 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, 4.9, or 5.0.

As used herein, the term “substantially” refers to a quantity, level, value, number, frequency, percentage, dimension, size, amount, weight or length that is 80%, 85%, 90%, 91%, 5 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or higher compared to a reference quantity, level, value, number, frequency, percentage, dimension, size, amount, weight or length. In one embodiment, “substantially the same” refers to a quantity, level, value, number, frequency, percentage, dimension, size, amount, weight or length that produces an effect, *e.g.*, a physiological effect, that is approximately the same as a reference quantity, level, value, 10 number, frequency, percentage, dimension, size, amount, weight or length.

Throughout this specification, unless the context requires otherwise, the words “comprise”, “comprises” and “comprising” will be understood to imply the inclusion of a stated step or element or group of steps or elements but not the exclusion of any other step or element or group of steps or elements. By “consisting of” is meant including, and limited to, 15 whatever follows the phrase “consisting of.” Thus, the phrase “consisting of” indicates that the listed elements are required or mandatory, and that no other elements may be present. By “consisting essentially of” is meant including any elements listed after the phrase, and limited to other elements that do not interfere with or contribute to the activity or action specified in the disclosure for the listed elements. Thus, the phrase “consisting essentially of” indicates that the 20 listed elements are required or mandatory, but that no other elements are present that materially affect the activity or action of the listed elements.

Reference throughout this specification to “one embodiment,” “an embodiment,” “a particular embodiment,” “a related embodiment,” “a certain embodiment,” “an additional embodiment,” or “a further embodiment” or combinations thereof means that a particular 25 feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment. Thus, the appearances of the foregoing phrases in various places throughout this specification are not necessarily all referring to the same embodiment. Furthermore, the particular features, structures, or characteristics may be combined in any suitable manner in one or more embodiments. It is also understood that the positive recitation

of a feature in one embodiment, serves as a basis for excluding the feature in a particular embodiment.

The term “*ex vivo*” refers generally to activities that take place outside an organism, such as experimentation or measurements done in or on living tissue in an artificial environment outside the organism, preferably with minimum alteration of the natural conditions. In particular embodiments, “*ex vivo*” procedures involve living cells or tissues taken from an organism and cultured or modulated in a laboratory apparatus, usually under sterile conditions, and typically for a few hours or up to about 24 hours, but including up to 48 or 72 hours, depending on the circumstances. In certain embodiments, such tissues or cells can be collected and frozen, and later thawed for *ex vivo* treatment. Tissue culture experiments or procedures lasting longer than a few days using living cells or tissue are typically considered to be “*in vitro*,” though in certain embodiments, this term can be used interchangeably with *ex vivo*.

The term “*in vivo*” refers generally to activities that take place inside an organism. In one embodiment, cellular genomes are engineered, edited, or modified *in vivo*.

By “enhance” or “promote” or “increase” or “expand” or “potentiate” refers generally to the ability of a nuclease variant, genome editing composition, or genome edited cell contemplated herein to produce, elicit, or cause a greater response (*i.e.*, physiological response) compared to the response caused by either vehicle or control. A measurable response may include an increase in engineered TCR or CAR expression, increase in HR or HDR efficiency, increase in binding site selectivity, increase in binding site specificity, increase in on-target binding, increases in immune effector cell expansion, activation, persistence, and/or an increase in cancer cell death killing ability, among others apparent from the understanding in the art and the description herein. An “increased” or “enhanced” amount is typically a “statistically significant” amount, and may include an increase that is 1.1, 1.2, 1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30 or more times (*e.g.*, 500, 1000 times) (including all integers and decimal points in between and above 1, *e.g.*, 1.5, 1.6, 1.7, 1.8, *etc.*) the response produced by vehicle or control.

By “decrease” or “lower” or “lessen” or “reduce” or “abate” or “ablate” or “inhibit” or “dampen” refers generally to the ability of a nuclease variant, genome editing composition, or genome edited cell contemplated herein to produce, elicit, or cause a lesser response (*i.e.*,

physiological response) compared to the response caused by either vehicle or control. A measurable response may include a decrease in endogenous TCR expression or function, a decrease in off-target binding, a decrease in expression of biomarkers associated with immune effector cell exhaustion, and the like. A “decrease” or “reduced” amount is typically a 5 “statistically significant” amount, and may include an decrease that is 1.1, 1.2, 1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30 or more times (e.g., 500, 1000 times) (including all integers and decimal points in between and above 1, e.g., 1.5, 1.6, 1.7, 1.8, etc.) the response (reference response) produced by vehicle, or control.

10 By “maintain,” or “preserve,” or “maintenance,” or “no change,” or “no substantial change,” or “no substantial decrease” refers generally to the ability of a nuclease variant, genome editing composition, or genome edited cell contemplated herein to produce, elicit, or cause a substantially similar or comparable physiological response (i.e., downstream effects) in as compared to the response caused by either vehicle or control. A comparable response is one that is not significantly different or measurable different from the reference response.

15 The terms “specific binding affinity” or “specifically binds” or “specifically bound” or “specific binding” or “specifically targets” and describe binding of one molecule to another, e.g., DNA binding domain of a polypeptide binding to DNA, at greater binding affinity than background binding. A binding domain “specifically binds” to a target site if it binds to or associates with a target site with an affinity or  $K_a$  (i.e., an equilibrium association constant of a 20 particular binding interaction with units of  $1/M$ ) of, for example, greater than or equal to about  $10^5 M^{-1}$ . In certain embodiments, a binding domain binds to a target site with a  $K_a$  greater than or equal to about  $10^6 M^{-1}$ ,  $10^7 M^{-1}$ ,  $10^8 M^{-1}$ ,  $10^9 M^{-1}$ ,  $10^{10} M^{-1}$ ,  $10^{11} M^{-1}$ ,  $10^{12} M^{-1}$ , or  $10^{13} M^{-1}$ . “High affinity” binding domains refers to those binding domains with a  $K_a$  of at least  $10^7 M^{-1}$ , at least  $10^8 M^{-1}$ , at least  $10^9 M^{-1}$ , at least  $10^{10} M^{-1}$ , at least  $10^{11} M^{-1}$ , at least  $10^{12} M^{-1}$ , at least 25  $10^{13} M^{-1}$ , or greater.

Alternatively, affinity may be defined as an equilibrium dissociation constant ( $K_d$ ) of a particular binding interaction with units of  $M$  (e.g.,  $10^{-5} M$  to  $10^{-13} M$ , or less). Affinities of nuclease variants comprising one or more DNA binding domains for DNA target sites contemplated in particular embodiments can be readily determined using conventional

techniques, *e.g.*, yeast cell surface display, or by binding association, or displacement assays using labeled ligands.

In one embodiment, the affinity of specific binding is about 2 times greater than background binding, about 5 times greater than background binding, about 10 times greater than background binding, about 20 times greater than background binding, about 50 times greater than background binding, about 100 times greater than background binding, or about 1000 times greater than background binding or more.

The terms “selectively binds” or “selectively bound” or “selectively binding” or “selectively targets” and describe preferential binding of one molecule to a target molecule (on-target binding) in the presence of a plurality of off-target molecules. In particular embodiments, an HE or megaTAL selectively binds an on-target DNA binding site about 5, 10, 15, 20, 25, 50, 100, or 1000 times more frequently than the HE or megaTAL binds an off-target DNA target binding site.

“On-target” refers to a target site sequence.

“Off-target” refers to a sequence similar to but not identical to a target site sequence.

A “target site” or “target sequence” is a chromosomal or extrachromosomal nucleic acid sequence that defines a portion of a nucleic acid to which a binding molecule will bind and/or cleave, provided sufficient conditions for binding and/or cleavage exist. When referring to a polynucleotide sequence or SEQ ID NO. that references only one strand of a target site or target sequence, it would be understood that the target site or target sequence bound and/or cleaved by a nuclease variant is double-stranded and comprises the reference sequence and its complement. In a preferred embodiment, the target site is a sequence in a human TCR $\alpha$  gene.

A “gene,” refers to a DNA region encoding a gene product, as well as all DNA regions which regulate the production of the gene product, whether or not such regulatory sequences are adjacent to coding and/or transcribed sequences. A gene includes, but is not limited to, promoter sequences, enhancers, silencers, insulators, boundary elements, terminators, polyadenylation sequences, post-transcription response elements, translational regulatory sequences such as ribosome binding sites and internal ribosome entry sites, replication origins, matrix attachment sites, and locus control regions.

As used herein, the term “genetically engineered” or “genetically modified” refers to the chromosomal or extrachromosomal addition of extra genetic material in the form of DNA or RNA to the total genetic material in a cell. Genetic modifications may be targeted or non-targeted to a particular site in a cell’s genome. In one embodiment, genetic modification is site specific. In one embodiment, genetic modification is not site specific.

As used herein, the term “genome editing” refers to the substitution, deletion, and/or introduction of genetic material at a target site in the cell’s genome, which restores, corrects, disrupts, and/or modifies expression of a gene or gene product. Genome editing contemplated in particular embodiments comprises introducing one or more nuclease variants into a cell to generate DNA lesions at or proximal to a target site in the cell’s genome, optionally in the presence of a donor repair template.

As used herein, the term “amount” refers to “an amount effective” or “an effective amount” of a nuclease variant, genome editing composition, or genome edited cell sufficient to achieve a beneficial or desired prophylactic or therapeutic result, including clinical results.

A “prophylactically effective amount” refers to an amount of a nuclease variant, genome editing composition, or genome edited cell sufficient to achieve the desired prophylactic result. Typically but not necessarily, since a prophylactic dose is used in subjects prior to or at an earlier stage of disease, the prophylactically effective amount is less than the therapeutically effective amount.

A “therapeutically effective amount” of a nuclease variant, genome editing composition, or genome edited cell may vary according to factors such as the disease state, age, sex, and weight of the individual, and the ability to elicit a desired response in the individual. A therapeutically effective amount is also one in which any toxic or detrimental effects are outweighed by the therapeutically beneficial effects. The term “therapeutically effective amount” includes an amount that is effective to “treat” a subject (e.g., a patient). When a therapeutic amount is indicated, the precise amount of the compositions contemplated in particular embodiments, to be administered, can be determined by a physician in view of the specification and with consideration of individual differences in age, weight, tumor size, extent of infection or metastasis, and condition of the patient (subject).

### C. NUCLEASE VARIANTS

Nuclease variants contemplated in particular embodiments herein are suitable for genome editing a target site in the TCR $\alpha$  gene and comprise one or more DNA binding domains and one or more DNA cleavage domains (*e.g.*, one or more endonuclease and/or 5 exonuclease domains), and optionally, one or more linkers contemplated herein. The terms 5 “reprogrammed nuclease,” “engineered nuclease,” or “nuclease variant” are used interchangeably and refer to a nuclease comprising one or more DNA binding domains and one or more DNA cleavage domains, wherein the nuclease has been designed and/or modified from a parental or naturally occurring nuclease, to bind and cleave a double-stranded DNA 10 target sequence in exon 1 of the constant region of the human TCR $\alpha$  gene.

In particular embodiments, a nuclease variant binds and cleaves a target sequence in exon 1 of the constant region of the human TCR $\alpha$  gene, preferably at SEQ ID NO: 17 in exon 1 of the constant region of the human TCR $\alpha$  gene, and more preferably at the sequence “ATTC” in SEQ ID NO: 17 in exon 1 of the constant region of the human TCR $\alpha$  gene.

15 The nuclease variant may be designed and/or modified from a naturally occurring nuclease or from a previous nuclease variant. Nuclease variants contemplated in particular embodiments may further comprise one or more additional functional domains, *e.g.*, an end-processing enzymatic domain of an end-processing enzyme that exhibits 5'-3' exonuclease, 5'-3' alkaline exonuclease, 3'-5' exonuclease (*e.g.*, Trex2), 5' flap endonuclease, helicase, 20 template-dependent DNA polymerases or template-independent DNA polymerase activity.

Illustrative examples of nuclease variants that bind and cleave a target sequence in the TCR $\alpha$  gene include, but are not limited to homing endonuclease (meganuclease) variants and megaTALs.

#### 1. HOMING ENDONUCLEASE (MEGANUCLEASE) VARIANTS

25 In various embodiments, a homing endonuclease or meganuclease is reprogrammed to introduce a double-strand break (DSB) in a target site in a TCR $\alpha$  gene. In particular embodiments, a homing endonuclease variant introduces a double strand break in exon 1 of the constant region of the human TCR $\alpha$  gene, preferably at SEQ ID NO: 17 in exon 1 of the constant region of the human TCR $\alpha$  gene, and more preferably at the sequence “ATTC” in

SEQ ID NO: 17 in exon 1 of the constant region of the human TCR $\alpha$  gene. “Homing endonuclease” and “meganuclease” are used interchangeably and refer to naturally-occurring homing endonucleases that recognize 12-45 base-pair cleavage sites and are commonly grouped into five families based on sequence and structure motifs: LAGLIDADG, GIY-YIG, 5 HNH, His-Cys box, and PD-(D/E)XK.

A “reference homing endonuclease” or “reference meganuclease” refers to a wild type homing endonuclease or a homing endonuclease found in nature. In one embodiment, a “reference homing endonuclease” refers to a wild type homing endonuclease that has been modified to increase basal activity.

10 An “engineered homing endonuclease,” “reprogrammed homing endonuclease,” “homing endonuclease variant,” “engineered meganuclease,” “reprogrammed meganuclease,” or “meganuclease variant” refers to a homing endonuclease comprising one or more DNA binding domains and one or more DNA cleavage domains, wherein the homing endonuclease has been designed and/or modified from a parental or naturally occurring homing 15 endonuclease, to bind and cleave a DNA target sequence in a TCR $\alpha$  gene. The homing endonuclease variant may be designed and/or modified from a naturally occurring homing endonuclease or from another homing endonuclease variant. Homing endonuclease variants contemplated in particular embodiments may further comprise one or more additional functional domains, *e.g.*, an end-processing enzymatic domain of an end-processing enzyme 20 that exhibits 5'-3' exonuclease, 5'-3' alkaline exonuclease, 3'-5' exonuclease (*e.g.*, Trex2), 5' flap endonuclease, helicase, template dependent DNA polymerase or template-independent DNA polymerase activity.

In particular embodiments, the homing endonuclease variants contemplated herein comprise increased selectivity to a target site compared to existing homing endonucleases, *e.g.*,

25 SEQ ID NO: 6. In particular embodiments, the homing endonucleases contemplated herein comprise increased selectivity to a target site while retaining catalytic activity compared to existing homing endonucleases, *e.g.*, SEQ ID NO: 6.

Homing endonuclease (HE) variants do not exist in nature and can be obtained by recombinant DNA technology or by random mutagenesis. HE variants may be obtained by 30 making one or more amino acid alterations, *e.g.*, mutating, substituting, adding, or deleting one

or more amino acids, in a naturally occurring HE or HE variant. In particular embodiments, a HE variant comprises one or more amino acid alterations to the DNA recognition interface.

HE variants contemplated in particular embodiments may further comprise one or more linkers and/or additional functional domains, *e.g.*, an end-processing enzymatic domain of an 5 end-processing enzyme that exhibits 5'-3' exonuclease, 5'-3' alkaline exonuclease, 3'-5' exonuclease (*e.g.*, Trex2), 5' flap endonuclease, helicase, template-dependent DNA polymerase or template-independent DNA polymerase activity. In particular embodiments, HE variants are introduced into a T cell with an end-processing enzyme that exhibits 5'-3' exonuclease, 5'-3' alkaline exonuclease, 3'-5' exonuclease (*e.g.*, Trex2), 5' flap endonuclease, 10 helicase, template-dependent DNA polymerase or template-independent DNA polymerase activity. The HE variant and 3' processing enzyme may be introduced separately, *e.g.*, in different vectors or separate mRNAs, or together, *e.g.*, as a fusion protein, or in a polycistronic construct separated by a viral self-cleaving peptide or an IRES element.

A “DNA recognition interface” refers to the HE amino acid residues that interact with 15 nucleic acid target bases as well as those residues that are adjacent. For each HE, the DNA recognition interface comprises an extensive network of side chain-to-side chain and side chain-to-DNA contacts, most of which is necessarily unique to recognize a particular nucleic acid target sequence. Thus, the amino acid sequence of the DNA recognition interface corresponding to a particular nucleic acid sequence varies significantly and is a feature of any 20 natural or HE variant. By way of non-limiting example, a HE variant contemplated in particular embodiments may be derived by constructing libraries of HE variants in which one or more amino acid residues localized in the DNA recognition interface of the natural HE (or a previously generated HE variant) are varied. The libraries may be screened for target cleavage activity against each predicted TCR $\alpha$  target site using cleavage assays (*see e.g.*, Jarjour *et al.*, 25 2009. *Nuc. Acids Res.* **37**(20): 6871-6880).

LAGLIDADG homing endonucleases (LHE) are the most well studied family of 30 homing endonucleases, are primarily encoded in archaea and in organellar DNA in green algae and fungi, and display the highest overall DNA recognition specificity. LHEs comprise one or two LAGLIDADG catalytic motifs per protein chain and function as homodimers or single chain monomers, respectively. Structural studies of LAGLIDADG proteins identified a highly

conserved core structure (Stoddard 2005), characterized by an  $\alpha\beta\beta\alpha\beta\beta\alpha$  fold, with the LAGLIDADG motif belonging to the first helix of this fold. The highly efficient and specific cleavage of LHE's represent a protein scaffold to derive novel, highly specific and selective endonucleases. However, engineering LHEs to bind and cleave a non-natural or non-canonical 5 target site requires selection of the appropriate LHE scaffold, examination of the target locus, selection of putative target sites, and extensive alteration of the LHE to alter its DNA contact points and cleavage specificity, at up to two-thirds of the base-pair positions in a target site.

Illustrative examples of LHEs include, but are not limited to I-AabMI, I-AaeMI, I-AnI, I-ApaMI, I-CapIII, I-CapIV, I-CkaMI, I-CpaMI, I-CpaMII, I-CpaMIII, I-CpaMIV, I-CpaMV, 10 I-CpaV, I-CraMI, I-EjeMI, I-GpeMI, I-GpiI, I-GzeMI, I-GzeMII, I-GzeMIII, I-HjeMI, I-LtrII, I-LtrI, I-LtrWI, I-MpeMI, I-MveMI, I-NcrII, I-NcrI, I-NcrMI, I-OheMI, I-OnuI, I-OsoMI, I-OsoMII, I-OsoMIII, I-OsoMIV, I-PanMI, I-PanMII, I-PanMIII, I-PnoMI, I-ScuMI, I-SmaMI, I-SscMI, and I-Vdi141I.

In one embodiment, the reprogrammed LHE or LHE variant is an I-OnuI variant. *See* 15 *e.g.*, SEQ ID NOs: 7-8.

In one embodiment, reprogrammed I-OnuI LHEs or I-OnuI variants targeting the TCR $\alpha$  gene were generated from a natural I-OnuI or biologically active fragment thereof (SEQ ID NOs: 1-5). In a preferred embodiment, reprogrammed I-OnuI LHEs or I-OnuI variants targeting the human TCR $\alpha$  gene were generated from an existing I-OnuI variant.

20 In one embodiment, the I-OnuI LHE that binds and cleaves a human TCR $\alpha$  gene comprises at least 99% sequence identity with the DNA recognition interface of an I-OnuI LHE variant as set forth in SEQ ID NO: 7 or SEQ ID NO: 8, and biologically active fragments thereof.

In a particular embodiment, an I-OnuI LHE variant that binds and cleaves a human 25 TCR $\alpha$  gene comprises one or more amino acid substitutions or modifications in the DNA recognition interface of an I-OnuI as set forth in SEQ ID NO: 7 or SEQ ID NO: 8, biologically active fragments thereof, and/or further variants thereof.

In certain embodiments, the HE variant cleaves a TCR $\alpha$  exon 1 constant region target site and comprises following amino acid substitutions: L26I, R28D, N32R, K34N, S35E, 30 V37N, G38R, S40R, E42S, G44R, V68K, A70T, G73S, N75R, S78M, K80R, L138M, T143N,

S159P, S176A, C180H, F182G, I186K, S188V, S190G, K191T, L192A, G193K, Q195Y, Q197G, V199R, S201A, T203S, K207R, Y223S, K225R, S233R, D236E, and V238E of any one of SEQ ID NOs: 1-5, or a biologically active fragment thereof.

5 In certain embodiments, the HE variant cleaves a TCR $\alpha$  exon 1 constant region target site and comprises following amino acid substitutions: L26I, R28D, N32R, K34N, S35E, V37N, G38R, S40R, E42S, G44R, V68K, A70T, G73S, N75R, S78M, K80R, L138M, T143N, S159P, S176A, E178D, C180H, F182G, I186K, S188V, S190G, K191T, L192A, G193K, Q195Y, Q197G, V199R, S201A, T203S, K207R, Y223S, K225R, S233R, D236E, and V238E of any one of SEQ ID NOs: 1-5, or a biologically active fragment thereof.

10 In particular embodiments, an I-OnuI LHE variant that binds and cleaves a human TCR $\alpha$  gene comprises an amino acid sequence that is at least 99% identical to the amino acid sequence set forth in any one of SEQ ID NO: 7 or SEQ ID NO: 8, or a biologically active fragment thereof.

15 In particular embodiments, an I-OnuI LHE variant comprises an amino acid sequence set forth in any one of SEQ ID NOs: 7-8, or a biologically active fragment thereof.

In particular embodiments, an I-OnuI LHE variant comprises an amino acid sequence set forth in SEQ ID NO: 7, or a biologically active fragment thereof.

In particular embodiments, an I-OnuI LHE variant comprises an amino acid sequence set forth in SEQ ID NO: 8, or a biologically active fragment thereof.

20 **2. MEGATALS**

In various embodiments, a megaTAL comprising a homing endonuclease variant is reprogrammed to introduce a double-strand break (DSB) in a target site in exon1 of the constant region of the human TCR $\alpha$  gene. In particular embodiments, a megaTAL introduces a DSB in in exon 1 of the constant region of the human TCR $\alpha$  gene, preferably at SEQ ID NO: 17 in exon 1 of the constant region of the human TCR $\alpha$  gene, and more preferably at the sequence “ATTC” in SEQ ID NO: 17 in exon 1 of the constant region of the human TCR $\alpha$  gene. A “megaTAL” refers to a polypeptide comprising a TALE DNA binding domain and a homing endonuclease variant that binds and cleaves a DNA target sequence in a TCR $\alpha$  gene, and optionally comprises one or more linkers and/or additional functional domains, *e.g.*, an end-processing enzymatic domain of an end-processing enzyme that exhibits 5'-3'

exonuclease, 5'-3' alkaline exonuclease, 3'-5' exonuclease (e.g., Trex2), 5' flap endonuclease, helicase or template-independent DNA polymerase activity.

In particular embodiments, a megaTAL can be introduced into a cell along with an end-processing enzyme that exhibits 5'-3' exonuclease, 5'-3' alkaline exonuclease, 3'-5' exonuclease (e.g., Trex2), 5' flap endonuclease, helicase, template-dependent DNA polymerase, or template-independent DNA polymerase activity. The megaTAL and 3' processing enzyme may be introduced separately, e.g., in different vectors or separate mRNAs, or together, e.g., as a fusion protein, or in a polycistronic construct separated by a viral self-cleaving peptide or an IRES element.

A "TALE DNA binding domain" is the DNA binding portion of transcription activator-like effectors (TALE or TAL-effectors), which mimics plant transcriptional activators to manipulate the plant transcriptome (see e.g., Kay *et al.*, 2007. *Science* 318:648-651). TALE DNA binding domains contemplated in particular embodiments are engineered *de novo* or from naturally occurring TALEs, e.g., AvrBs3 from *Xanthomonas campestris* pv. *vesicatoria*, *Xanthomonas gardneri*, *Xanthomonas translucens*, *Xanthomonas axonopodis*, *Xanthomonas perforans*, *Xanthomonas alfalfa*, *Xanthomonas citri*, *Xanthomonas euvesicatoria*, and *Xanthomonas oryzae* and brg11 and hpx17 from *Ralstonia solanacearum*. Illustrative examples of TALE proteins for deriving and designing DNA binding domains are disclosed in U.S. Patent No. 9,017,967, and references cited therein, all of which are incorporated herein by reference in their entireties.

In particular embodiments, a megaTAL comprises a TALE DNA binding domain comprising one or more repeat units that are involved in binding of the TALE DNA binding domain to its corresponding target DNA sequence. A single "repeat unit" (also referred to as a "repeat") is typically 33-35 amino acids in length. Each TALE DNA binding domain repeat unit includes 1 or 2 DNA-binding residues making up the Repeat Variable Di-Residue (RVD), typically at positions 12 and/or 13 of the repeat. The natural (canonical) code for DNA recognition of these TALE DNA binding domains has been determined such that an HD sequence at positions 12 and 13 leads to a binding to cytosine (C), NG binds to T, NI to A, NN binds to G or A, and NG binds to T. In certain embodiments, non-canonical (atypical) RVDs are contemplated.

Illustrative examples of non-canonical RVDs suitable for use in particular megaTALs contemplated in particular embodiments include, but are not limited to HH, KH, NH, NK, NQ, RH, RN, SS, NN, SN, KN for recognition of guanine (G); NI, KI, RI, HI, SI for recognition of adenine (A); NG, HG, KG, RG for recognition of thymine (T); RD, SD, HD, ND, KD, YG for 5 recognition of cytosine (C); NV, HN for recognition of A or G; and H\*, HA, KA, N\*, NA, NC, NS, RA, S\* for recognition of A or T or G or C, wherein (\*) means that the amino acid at position 13 is absent. Additional illustrative examples of RVDs suitable for use in particular megaTALs contemplated in particular embodiments further include those disclosed in U.S. Patent No. 8,614,092, which is incorporated herein by reference in its entirety.

10 In particular embodiments, a megaTAL contemplated herein comprises a TALE DNA binding domain comprising 3 to 30 repeat units. In certain embodiments, a megaTAL comprises 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 TALE DNA binding domain repeat units. In a preferred embodiment, a megaTAL contemplated herein comprises a TALE DNA binding domain comprising 5-15 repeat units, more preferably 7-15 repeat units, more preferably 9-15 repeat units, and more 15 preferably 9, 10, 11, 12, 13, 14, or 15 repeat units.

10 In particular embodiments, a megaTAL contemplated herein comprises a TALE DNA binding domain comprising 3 to 30 repeat units and an additional single truncated TALE repeat unit comprising 20 amino acids located at the C-terminus of a set of TALE repeat units, *i.e.*, an additional C-terminal half-TALE DNA binding domain repeat unit (amino acids -20 to -1 of the C-cap disclosed elsewhere herein, *infra*). Thus, in particular embodiments, a megaTAL contemplated herein comprises a TALE DNA binding domain comprising 3.5 to 30.5 repeat units. In certain embodiments, a megaTAL comprises 3.5, 4.5, 5.5, 6.5, 7.5, 8.5, 9.5, 10.5, 11.5, 12.5, 13.5, 14.5, 15.5, 16.5, 17.5, 18.5, 19.5, 20.5, 21.5, 22.5, 23.5, 24.5, 25.5, 26.5, 27.5, 28.5, 29.5, or 30.5 TALE DNA binding domain repeat units. In a preferred embodiment, a megaTAL contemplated herein comprises a TALE DNA binding domain comprising 5.5-15.5 repeat units, more preferably 7.5-15.5 repeat units, more preferably 9.5-15.5 repeat units, and more 25 preferably 9.5, 10.5, 11.5, 12.5, 13.5, 14.5, or 15.5 repeat units.

10 In particular embodiments, a megaTAL comprises a TAL effector architecture comprising an “N-terminal domain (NTD)” polypeptide, one or more TALE repeat 30

domains/units, a “C-terminal domain (CTD)” polypeptide, and a homing endonuclease variant. In some embodiments, the NTD, TALE repeats, and/or CTD domains are from the same species. In other embodiments, one or more of the NTD, TALE repeats, and/or CTD domains are from different species.

5 As used herein, the term “N-terminal domain (NTD)” polypeptide refers to the sequence that flanks the N-terminal portion or fragment of a naturally occurring TALE DNA binding domain. The NTD sequence, if present, may be of any length as long as the TALE DNA binding domain repeat units retain the ability to bind DNA. In particular embodiments, the NTD polypeptide comprises at least 120 to at least 140 or more amino acids N-terminal to 10 the TALE DNA binding domain (0 is amino acid 1 of the most N-terminal repeat unit). In particular embodiments, the NTD polypeptide comprises at least about 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, or at least 140 amino acids N-terminal to the TALE DNA binding domain. In one embodiment, a megaTAL contemplated herein comprises an NTD polypeptide of at least about amino acids +1 to +122 15 to at least about +1 to +137 of a *Xanthomonas* TALE protein (0 is amino acid 1 of the most N-terminal repeat unit). In particular embodiments, the NTD polypeptide comprises at least about 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, or 137 amino acids N-terminal to the TALE DNA binding domain of a *Xanthomonas* TALE protein. In one embodiment, a megaTAL contemplated herein comprises an NTD polypeptide of at least 20 amino acids +1 to +121 of a *Ralstonia* TALE protein (0 is amino acid 1 of the most N-terminal repeat unit). In particular embodiments, the NTD polypeptide comprises at least about 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, or 137 amino acids N-terminal to the TALE DNA binding domain of a *Ralstonia* TALE protein.

As used herein, the term “C-terminal domain (CTD)” polypeptide refers to the 25 sequence that flanks the C-terminal portion or fragment of a naturally occurring TALE DNA binding domain. The CTD sequence, if present, may be of any length as long as the TALE DNA binding domain repeat units retain the ability to bind DNA. In particular embodiments, the CTD polypeptide comprises at least 20 to at least 85 or more amino acids C-terminal to the last full repeat of the TALE DNA binding domain (the first 20 amino acids are the half-repeat 30 unit C-terminal to the last C-terminal full repeat unit). In particular embodiments, the CTD

polypeptide comprises at least about 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 443, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, or at least 85 amino acids C-terminal to the last full repeat of the TALE DNA binding domain.

5 In one embodiment, a megaTAL contemplated herein comprises a CTD polypeptide of at least about amino acids -20 to -1 of a *Xanthomonas* TALE protein (-20 is amino acid 1 of a half-repeat unit C-terminal to the last C-terminal full repeat unit). In particular embodiments, the CTD polypeptide comprises at least about 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 amino acids C-terminal to the last full repeat of the TALE DNA binding domain of a *Xanthomonas* TALE protein. In one embodiment, a megaTAL contemplated herein 10 comprises a CTD polypeptide of at least about amino acids -20 to -1 of a *Ralstonia* TALE protein (-20 is amino acid 1 of a half-repeat unit C-terminal to the last C-terminal full repeat unit). In particular embodiments, the CTD polypeptide comprises at least about 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 amino acids C-terminal to the last full repeat 15 of the TALE DNA binding domain of a *Ralstonia* TALE protein.

In particular embodiments, a megaTAL contemplated herein, comprises a fusion polypeptide comprising a TALE DNA binding domain engineered to bind a target sequence, a homing endonuclease reprogrammed to bind and cleave a target sequence, and optionally an NTD and/or CTD polypeptide, optionally joined to each other with one or more linker 20 polypeptides contemplated elsewhere herein. Without wishing to be bound by any particular theory, it is contemplated that a megaTAL comprising TALE DNA binding domain, and optionally an NTD and/or CTD polypeptide is fused to a linker polypeptide which is further fused to a homing endonuclease variant. Thus, the TALE DNA binding domain binds a DNA target sequence that is within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 nucleotides 25 away from the target sequence bound by the DNA binding domain of the homing endonuclease variant. In this way, the megaTALs contemplated herein, increase the specificity, selectivity, and efficiency of genome editing.

In one embodiment, a megaTAL comprises a homing endonuclease variant and a TALE DNA binding domain that binds a nucleotide sequence that is within about 4, 5, or 6

nucleotides, preferably, 5 or 6 nucleotides upstream of the binding site of the reprogrammed homing endonuclease.

In one embodiment, a megaTAL comprises a homing endonuclease variant and a TALE DNA binding domain that binds the nucleotide sequence set forth in SEQ ID NO: 18, 5 which is 6 nucleotides upstream of the nucleotide sequence bound and cleaved by the homing endonuclease variant (SEQ ID NO: 17). In preferred embodiments, the megaTAL target sequence is SEQ ID NO: 18.

In one embodiment, a megaTAL comprises a homing endonuclease variant and a TALE DNA binding domain that binds the nucleotide sequence set forth in SEQ ID NO: 19, 10 which is 5 nucleotides upstream of the nucleotide sequence bound and cleaved by the homing endonuclease variant (SEQ ID NO: 17) . In preferred embodiments, the megaTAL target sequence is SEQ ID NO: 19.

In particular embodiments, a megaTAL contemplated herein, comprises an NTD of about 122 amino acids to 137 amino acids, about 9.5, about 10.5, about 11.5, about 12.5, about 15 13.5, about 14.5, or about 15.5 binding repeat units, a CTD of about 20 amino acids to about 85 amino acids, and an I-OnuI LHE variant. In particular embodiments, any one of, two of, or all of the NTD, DNA binding domain, and CTD can be designed from the same species or different species, in any suitable combination.

In particular embodiments, a megaTAL contemplated herein, comprises the amino acid 20 sequence set forth in any one of SEQ ID NOs: 10-12.

In particular embodiments, a megaTAL contemplated herein, comprises the amino acid sequence set forth in SEQ ID NO: 10.

In particular embodiments, a megaTAL contemplated herein, comprises the amino acid sequence set forth in SEQ ID NO: 11.

25 In particular embodiments, a megaTAL contemplated herein, comprises the amino acid sequence set forth in SEQ ID NO: 12.

In certain embodiments, a megaTAL comprises a TALE DNA binding domain and an I-OnuI LHE variant binds and cleaves the nucleotide sequence set forth in SEQ ID NO: 17 or SEQ ID NO: 20.

### 3. END-PROCESSING ENZYMES

Genome editing compositions and methods contemplated in particular embodiments comprise editing cellular genomes using a nuclease variant and one or more copies of an end-processing enzyme. In particular embodiments, a single polynucleotide encodes a homing endonuclease variant and an end-processing enzyme, separated by a linker, a self-cleaving peptide sequence, *e.g.*, 2A sequence, or by an IRES sequence. In particular embodiments, genome editing compositions comprise a polynucleotide encoding a nuclease variant and a separate polynucleotide encoding an end-processing enzyme. In particular embodiments, genome editing compositions comprise a polynucleotide encoding a homing endonuclease variant end-processing enzyme single polypeptide fusion in addition to a tandem copy of the end-processing enzyme separated by a self-cleaving peptide.

The term “end-processing enzyme” refers to an enzyme that modifies the exposed ends of a polynucleotide chain. The polynucleotide may be double-stranded DNA (dsDNA), single-stranded DNA (ssDNA), RNA, double-stranded hybrids of DNA and RNA, and synthetic DNA (for example, containing bases other than A, C, G, and T). An end-processing enzyme may modify exposed polynucleotide chain ends by adding one or more nucleotides, removing one or more nucleotides, removing or modifying a phosphate group and/or removing or modifying a hydroxyl group. An end-processing enzyme may modify ends at endonuclease cut sites or at ends generated by other chemical or mechanical means, such as shearing (for example by passing through fine-gauge needle, heating, sonicating, mini bead tumbling, and nebulizing), ionizing radiation, ultraviolet radiation, oxygen radicals, chemical hydrolysis and chemotherapy agents.

In particular embodiments, genome editing compositions and methods contemplated in particular embodiments comprise editing cellular genomes using a homing endonuclease variant or megaTAL and a DNA end-processing enzyme.

The term “DNA end-processing enzyme” refers to an enzyme that modifies the exposed ends of DNA. A DNA end-processing enzyme may modify blunt ends or staggered ends (ends with 5' or 3' overhangs). A DNA end-processing enzyme may modify single stranded or double stranded DNA. A DNA end-processing enzyme may modify ends at endonuclease cut sites or at ends generated by other chemical or mechanical means, such as

shearing (for example by passing through fine-gauge needle, heating, sonicating, mini bead tumbling, and nebulizing), ionizing radiation, ultraviolet radiation, oxygen radicals, chemical hydrolysis and chemotherapy agents. DNA end-processing enzyme may modify exposed DNA ends by adding one or more nucleotides, removing one or more nucleotides, removing or 5 modifying a phosphate group and/or removing or modifying a hydroxyl group.

Illustrative examples of DNA end-processing enzymes suitable for use in particular embodiments contemplated herein include, but are not limited to: 5'-3' exonucleases, 5'-3' alkaline exonucleases, 3'-5' exonucleases, 5' flap endonucleases, helicases, phosphatases, hydrolases and template-independent DNA polymerases.

10 Additional illustrative examples of DNA end-processing enzymes suitable for use in particular embodiments contemplated herein include, but are not limited to, Trex2, Trex1, Trex1 without transmembrane domain, Apollo, Artemis, DNA2, Exo1, ExoT, ExoIII, Fen1, Fan1, MreII, Rad2, Rad9, TdT (terminal deoxynucleotidyl transferase), PNKP, RecE, RecJ, RecQ, Lambda exonuclease, Sox, Vaccinia DNA polymerase, exonuclease I, exonuclease III, 15 exonuclease VII, NDK1, NDK5, NDK7, NDK8, WRN, T7-exonuclease Gene 6, avian myeloblastosis virus integration protein (IN), Bloom, Antarctic Phophatase, Alkaline Phosphatase, Poly nucleotide Kinase (PNK), ApeI, Mung Bean nuclease, Hex1, TTRAP (TDP2), Sgs1, Sae2, CUP, Pol mu, Pol lambda, MUS81, EME1, EME2, SLX1, SLX4 and UL-12.

20 In particular embodiments, genome editing compositions and methods for editing cellular genomes contemplated herein comprise polypeptides comprising a homing endonuclease variant or megaTAL and an exonuclease. The term “exonuclease” refers to enzymes or domains that cleave phosphodiester bonds at the end of a polynucleotide chain via a hydrolyzing reaction that breaks phosphodiester bonds at either the 3' or 5' end.

25 Illustrative examples of exonucleases suitable for use in particular embodiments contemplated herein include, but are not limited to: hExoI, Yeast ExoI, *E. coli* ExoI, hTREX2, mouse TREX2, rat TREX2, hTREX1, mouse TREX1, rat TREX1, and Rat TREX1.

30 In particular embodiments, the DNA end-processing enzyme is a 3' to 5' exonuclease, preferably Trex 1 or Trex2, more preferably Trex2, and even more preferably human or mouse Trex2.

**D. POLYPEPTIDES**

Various polypeptides are contemplated herein, including, but not limited to, homing endonuclease variants, megaTALs, and fusion polypeptides. In preferred embodiments, a polypeptide comprises the amino acid sequence set forth in SEQ ID NOs: 1-12 and 22-24.

5 “Polypeptide,” “polypeptide fragment,” “peptide” and “protein” are used interchangeably, unless specified to the contrary, and according to conventional meaning, *i.e.*, as a sequence of amino acids. In one embodiment, a “polypeptide” includes fusion polypeptides and other variants. Polypeptides can be prepared using any of a variety of well-known recombinant and/or synthetic techniques. Polypeptides are not limited to a specific length, *e.g.*, they may 10 comprise a full length protein sequence, a fragment of a full length protein, or a fusion protein, and may include post-translational modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations and the like, as well as other modifications known in the art, both naturally occurring and non-naturally occurring.

An “isolated protein,” “isolated peptide,” or “isolated polypeptide” and the like, as used 15 herein, refer to *in vitro* synthesis, isolation, and/or purification of a peptide or polypeptide molecule from a cellular environment, and from association with other components of the cell, *i.e.*, it is not significantly associated with *in vivo* substances.

20 Illustrative examples of polypeptides contemplated in particular embodiments include, but are not limited to homing endonuclease variants, megaTALs, end-processing nucleases, fusion polypeptides and variants thereof.

25 Polypeptides include “polypeptide variants.” Polypeptide variants may differ from a naturally occurring polypeptide in one or more amino acid substitutions, deletions, additions and/or insertions. Such variants may be naturally occurring or may be synthetically generated, for example, by modifying one or more amino acids of the above polypeptide sequences. For example, in particular embodiments, it may be desirable to improve the biological properties of a homing endonuclease, megaTAL or the like that binds and cleaves a target site in the human TCR $\alpha$  gene by introducing one or more substitutions, deletions, additions and/or insertions into the polypeptide. In particular embodiments, polypeptides include polypeptides having at least about 65%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 30 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%

amino acid identity to any of the reference sequences contemplated herein, typically where the variant maintains at least one biological activity of the reference sequence.

Polypeptides variants include biologically active “polypeptide fragments.” Illustrative examples of biologically active polypeptide fragments include DNA binding domains,

5 nuclease domains, and the like. As used herein, the term “biologically active fragment” or “minimal biologically active fragment” refers to a polypeptide fragment that retains at least 100%, at least 90%, at least 80%, at least 70%, at least 60%, at least 50%, at least 40%, at least 30%, at least 20%, at least 10%, or at least 5% of the naturally occurring polypeptide activity.

In preferred embodiments, the biological activity is binding affinity and/or cleavage activity for 10 a target sequence. In certain embodiments, a polypeptide fragment can comprise an amino acid chain at least 5 to about 1700 amino acids long. It will be appreciated that in certain embodiments, fragments are at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 150, 200, 250, 300, 350, 400, 450, 15 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700 or more amino acids long. In particular embodiments, a polypeptide comprises a biologically active fragment of a homing endonuclease variant. In particular embodiments, the polypeptides set forth herein may comprise one or more amino acids denoted as “X.” “X” if present in an amino acid SEQ ID NO, refers to any amino acid. One or more “X” residues 20 may be present at the N- and C-terminus of an amino acid sequence set forth in particular SEQ ID NOs contemplated herein. If the “X” amino acids are not present the remaining amino acid sequence set forth in a SEQ ID NO may be considered a biologically active fragment.

In particular embodiments, a polypeptide comprises a biologically active fragment of a homing endonuclease variant, *e.g.*, SEQ ID NOs: 6-7, or a megaTAL (SEQ ID NOs: 10-12).

25 The biologically active fragment may comprise an N-terminal truncation and/or C-terminal truncation. In a particular embodiment, a biologically active fragment lacks or comprises a deletion of the 1, 2, 3, 4, 5, 6, 7, or 8 N-terminal amino acids of a homing endonuclease variant compared to a corresponding wild type homing endonuclease sequence, more preferably a deletion of the 4 N-terminal amino acids of a homing endonuclease variant compared to a 30 corresponding wild type homing endonuclease sequence. In a particular embodiment, a

biologically active fragment lacks or comprises a deletion of the 1, 2, 3, 4, or 5 C-terminal amino acids of a homing endonuclease variant compared to a corresponding wild type homing endonuclease sequence, more preferably a deletion of the 2 C-terminal amino acids of a homing endonuclease variant compared to a corresponding wild type homing endonuclease sequence. In a particular preferred embodiment, a biologically active fragment lacks or comprises a deletion of the 4 N-terminal amino acids and 2 C-terminal amino acids of a homing endonuclease variant compared to a corresponding wild type homing endonuclease sequence.

5 In a particular embodiment, an I-OnuI variant comprises a deletion of 1, 2, 3, 4, 5, 6, 7, or 8 the following N-terminal amino acids: M, A, Y, M, S, R, R, E; and/or a deletion of the following 1, 2, 3, 4, or 5 C-terminal amino acids: R, G, S, F, V.

In a particular embodiment, an I-OnuI variant comprises a deletion or substitution of 1, 2, 3, 4, 5, 6, 7, or 8 the following N-terminal amino acids: M, A, Y, M, S, R, R, E; and/or a deletion or substitution of the following 1, 2, 3, 4, or 5 C-terminal amino acids: R, G, S, F, V.

15 In a particular embodiment, an I-OnuI variant comprises a deletion of 1, 2, 3, 4, 5, 6, 7, or 8 the following N-terminal amino acids: M, A, Y, M, S, R, R, E; and/or a deletion of the following 1 or 2 C-terminal amino acids: F, V.

In a particular embodiment, an I-OnuI variant comprises a deletion or substitution of 1, 2, 3, 4, 5, 6, 7, or 8 the following N-terminal amino acids: M, A, Y, M, S, R, R, E; and/or a deletion or substitution of the following 1 or 2 C-terminal amino acids: F, V.

As noted above, polypeptides may be altered in various ways including amino acid substitutions, deletions, truncations, and insertions. Methods for such manipulations are generally known in the art. For example, amino acid sequence variants of a reference polypeptide can be prepared by mutations in the DNA. Methods for mutagenesis and 25 nucleotide sequence alterations are well known in the art. *See*, for example, Kunkel (1985, *Proc. Natl. Acad. Sci. USA.* 82: 488-492), Kunkel *et al.*, (1987, *Methods in Enzymol.*, 154: 367-382), U.S. Pat. No. 4,873,192, Watson, J. D. *et al.*, (*Molecular Biology of the Gene*, Fourth Edition, Benjamin/Cummings, Menlo Park, Calif., 1987) and the references cited therein. Guidance as to appropriate amino acid substitutions that do not affect biological activity of the

protein of interest may be found in the model of Dayhoff *et al.*, (1978) *Atlas of Protein Sequence and Structure* (*Natl. Biomed. Res. Found.*, Washington, D.C.).

In certain embodiments, a variant will contain one or more conservative substitutions. A “conservative substitution” is one in which an amino acid is substituted for another amino acid that has similar properties, such that one skilled in the art of peptide chemistry would expect the secondary structure and hydropathic nature of the polypeptide to be substantially unchanged. Modifications may be made in the structure of the polynucleotides and polypeptides contemplated in particular embodiments, polypeptides include polypeptides having at least about and still obtain a functional molecule that encodes a variant or derivative 10 polypeptide with desirable characteristics. When it is desired to alter the amino acid sequence of a polypeptide to create an equivalent, or even an improved, variant polypeptide, one skilled in the art, for example, can change one or more of the codons of the encoding DNA sequence, *e.g.*, according to Table 1.

**TABLE 1- Amino Acid Codons**

Amino Acids	One letter code	Three letter code	Codons					
Alanine	A	Ala	GCA	GCC	GCG	GCU		
Cysteine	C	Cys	UGC	UGU				
Aspartic acid	D	Asp	GAC	GAU				
Glutamic acid	E	Glu	GAA	GAG				
Phenylalanine	F	Phe	UUC	UUU				
Glycine	G	Gly	GGA	GGC	GGG	GGU		
Histidine	H	His	CAC	CAU				
Isoleucine	I	Iso	AUA	AUC	AUU			
Lysine	K	Lys	AAA	AAG				
Leucine	L	Leu	UUA	UUG	CUA	CUC	CUG	CUU
Methionine	M	Met	AUG					
Asparagine	N	Asn	AAC	AAU				
Proline	P	Pro	CCA	CCC	CCG	CCU		
Glutamine	Q	Gln	CAA	CAG				

Arginine	R	Arg	AGA	AGG	CGA	CGC	CGG	CGU
Serine	S	Ser	AGC	AGU	UCA	UCC	UCG	UCU
Threonine	T	Thr	ACA	ACC	ACG	ACU		
Valine	V	Val	GUA	GUC	GUG	GUU		
Tryptophan	W	Trp	UGG					
Tyrosine	Y	Tyr	UAC	UAU				

Guidance in determining which amino acid residues can be substituted, inserted, or deleted without abolishing biological activity can be found using computer programs well known in the art, such as DNASTAR, DNA Strider, Geneious, Mac Vector, or Vector NTI software. Preferably, amino acid changes in the protein variants disclosed herein are

5 conservative amino acid changes, *i.e.*, substitutions of similarly charged or uncharged amino acids. A conservative amino acid change involves substitution of one of a family of amino acids which are related in their side chains. Naturally occurring amino acids are generally divided into four families: acidic (aspartate, glutamate), basic (lysine, arginine, histidine), non-polar (alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), and 10 uncharged polar (glycine, asparagine, glutamine, cysteine, serine, threonine, tyrosine) amino acids. Phenylalanine, tryptophan, and tyrosine are sometimes classified jointly as aromatic amino acids. In a peptide or protein, suitable conservative substitutions of amino acids are known to those of skill in this art and generally can be made without altering a biological activity of a resulting molecule. Those of skill in this art recognize that, in general, single 15 amino acid substitutions in non-essential regions of a polypeptide do not substantially alter biological activity (see, *e.g.*, Watson *et al. Molecular Biology of the Gene*, 4th Edition, 1987, The Benjamin/Cummings Pub. Co., p.224).

In one embodiment, where expression of two or more polypeptides is desired, the polynucleotide sequences encoding them can be separated by an IRES sequence as disclosed 20 elsewhere herein.

Polypeptides contemplated in particular embodiments include fusion polypeptides. In particular embodiments, fusion polypeptides and polynucleotides encoding fusion polypeptides are provided. Fusion polypeptides and fusion proteins refer to a polypeptide having at least two, three, four, five, six, seven, eight, nine, or ten polypeptide segments.

In another embodiment, two or more polypeptides can be expressed as a fusion protein that comprises one or more self-cleaving polypeptide sequences as disclosed elsewhere herein.

In one embodiment, a fusion protein contemplated herein comprises one or more DNA binding domains and one or more nucleases, and one or more linker and/or self-cleaving 5 polypeptides.

In one embodiment, a fusion protein contemplated herein comprises nuclease variant; a linker or self-cleaving peptide; and an end-processing enzyme including but not limited to a 5'-3' exonuclease, a 5'-3' alkaline exonuclease, and a 3'-5' exonuclease (e.g., Trex2).

Fusion polypeptides can comprise one or more polypeptide domains or segments 10 including, but are not limited to signal peptides, cell permeable peptide domains (CPP), DNA binding domains, nuclease domains, *etc.*, epitope tags (e.g., maltose binding protein ("MBP"), glutathione S transferase (GST), HIS6, MYC, FLAG, V5, VSV-G, and HA), polypeptide linkers, and polypeptide cleavage signals. Fusion polypeptides are typically linked C-terminus to N-terminus, although they can also be linked C-terminus to C-terminus, N-terminus to N- 15 terminus, or N-terminus to C-terminus. In particular embodiments, the polypeptides of the fusion protein can be in any order. Fusion polypeptides or fusion proteins can also include conservatively modified variants, polymorphic variants, alleles, mutants, subsequences, and interspecies homologs, so long as the desired activity of the fusion polypeptide is preserved. Fusion polypeptides may be produced by chemical synthetic methods or by chemical linkage 20 between the two moieties or may generally be prepared using other standard techniques. Ligated DNA sequences comprising the fusion polypeptide are operably linked to suitable transcriptional or translational control elements as disclosed elsewhere herein.

Fusion polypeptides may optionally comprises a linker that can be used to link the one or more polypeptides or domains within a polypeptide. A peptide linker sequence may be 25 employed to separate any two or more polypeptide components by a distance sufficient to ensure that each polypeptide folds into its appropriate secondary and tertiary structures so as to allow the polypeptide domains to exert their desired functions. Such a peptide linker sequence is incorporated into the fusion polypeptide using standard techniques in the art. Suitable peptide linker sequences may be chosen based on the following factors: (1) their ability to 30 adopt a flexible extended conformation; (2) their inability to adopt a secondary structure that

could interact with functional epitopes on the first and second polypeptides; and (3) the lack of hydrophobic or charged residues that might react with the polypeptide functional epitopes. Preferred peptide linker sequences contain Gly, Asn and Ser residues. Other near neutral amino acids, such as Thr and Ala may also be used in the linker sequence. Amino acid 5 sequences which may be usefully employed as linkers include those disclosed in Maratea *et al.*, *Gene* 40:39-46, 1985; Murphy *et al.*, *Proc. Natl. Acad. Sci. USA* 83:8258-8262, 1986; U.S. Patent No. 4,935,233 and U.S. Patent No. 4,751,180. Linker sequences are not required when a particular fusion polypeptide segment contains non-essential N-terminal amino acid regions that can be used to separate the functional domains and prevent steric interference. Preferred 10 linkers are typically flexible amino acid subsequences which are synthesized as part of a recombinant fusion protein. Linker polypeptides can be between 1 and 200 amino acids in length, between 1 and 100 amino acids in length, or between 1 and 50 amino acids in length, including all integer values in between.

Exemplary linkers include, but are not limited to the following amino acid sequences: 15 glycine polymers (G)<sub>n</sub>; glycine-serine polymers (G<sub>1-5</sub>S<sub>1-5</sub>)<sub>n</sub>, where n is an integer of at least one, two, three, four, or five; glycine-alanine polymers; alanine-serine polymers; GGG (SEQ ID NO: 30); DGGGS (SEQ ID NO: 31); TGEKP (SEQ ID NO: 32) (see e.g., Liu *et al.*, *PNAS* 5525-5530 (1997)); GGRR (SEQ ID NO: 33) (Pomerantz *et al.* 1995, *supra*); (GGGGS)<sub>n</sub> 20 wherein n = 1, 2, 3, 4 or 5 (SEQ ID NO: 34) (Kim *et al.*, *PNAS* 93, 1156-1160 (1996.); EGKSSGSGSESKVD (SEQ ID NO: 35) (Chaudhary *et al.*, 1990, *Proc. Natl. Acad. Sci. U.S.A.* 87:1066-1070); KESGSVSSEQLAQFRSLD (SEQ ID NO: 36) (Bird *et al.*, 1988, *Science* 242:423-426), GGRRGGGS (SEQ ID NO: 37); LRQRDGERP (SEQ ID NO: 38); LRQKDGGGSERP (SEQ ID NO: 39); LRQKD(GGGS)<sub>2</sub>ERP (SEQ ID NO: 40).

Alternatively, flexible linkers can be rationally designed using a computer program capable of 25 modeling both DNA-binding sites and the peptides themselves (Desjarlais & Berg, *PNAS* 90:2256-2260 (1993), *PNAS* 91:11099-11103 (1994) or by phage display methods.

Fusion polypeptides may further comprise a polypeptide cleavage signal between each 30 of the polypeptide domains described herein or between an endogenous open reading frame and a polypeptide encoded by a donor repair template. In addition, a polypeptide cleavage site can be put into any linker peptide sequence. Exemplary polypeptide cleavage signals include

polypeptide cleavage recognition sites such as protease cleavage sites, nuclease cleavage sites (e.g., rare restriction enzyme recognition sites, self-cleaving ribozyme recognition sites), and self-cleaving viral oligopeptides (see deFelipe and Ryan, 2004. *Traffic*, 5(8); 616-26).

Suitable protease cleavages sites and self-cleaving peptides are known to the skilled person (see, e.g., in Ryan *et al.*, 1997. *J. Gener. Virol.* 78, 699-722; Scymczak *et al.* (2004) *Nature Biotech.* 5, 589-594). Exemplary protease cleavage sites include, but are not limited to the cleavage sites of potyvirus NIa proteases (e.g., tobacco etch virus protease), potyvirus HC proteases, potyvirus P1 (P35) proteases, byovirus NIa proteases, byovirus RNA-2-encoded proteases, aphthovirus L proteases, enterovirus 2A proteases, rhinovirus 2A proteases, picorna 10 3C proteases, comovirus 24K proteases, nepovirus 24K proteases, RTSV (rice tungro spherical virus) 3C-like protease, PYVF (parsnip yellow fleck virus) 3C-like protease, heparin, thrombin, factor Xa and enterokinase. Due to its high cleavage stringency, TEV (tobacco etch virus) protease cleavage sites are preferred in one embodiment, e.g., EXXYXQ(G/S) (SEQ ID NO: 41), for example, ENLYFQG (SEQ ID NO: 42) and ENLYFQS (SEQ ID NO: 43), 15 wherein X represents any amino acid (cleavage by TEV occurs between Q and G or Q and S).

In certain embodiments, the self-cleaving polypeptide site comprises a 2A or 2A-like site, sequence or domain (Donnelly *et al.*, 2001. *J. Gen. Virol.* 82:1027-1041). In a particular embodiment, the viral 2A peptide is an aphthovirus 2A peptide, a potyvirus 2A peptide, or a cardiovirus 2A peptide.

20 In one embodiment, the viral 2A peptide is selected from the group consisting of: a foot-and-mouth disease virus (FMDV) 2A peptide, an equine rhinitis A virus (ERAV) 2A peptide, a Thosea asigna virus (TaV) 2A peptide, a porcine teschovirus-1 (PTV-1) 2A peptide, a Theilovirus 2A peptide, and an encephalomyocarditis virus 2A peptide.

Illustrative examples of 2A sites are provided in Table 2.

25 **TABLE 2:** Exemplary 2A sites include the following sequences:

SEQ ID NO: 44	GSGATNFSLLKQAGDVEENPGP
SEQ ID NO: 45	ATNFSLLKQAGDVEENPGP
SEQ ID NO: 46	LLKQAGDVEENPGP
SEQ ID NO: 47	GSGEGRGSLLTCGDVEENPGP

SEQ ID NO: 48	EGRGSLLTCGDVEENPGP
SEQ ID NO: 49	LLTCGDVEENPGP
SEQ ID NO: 50	GSGQCTNYALLKLAGDVESNPGP
SEQ ID NO: 51	QCTNYALLKLAGDVESNPGP
SEQ ID NO: 52	LLKLAGDVESNPGP
SEQ ID NO: 53	GSGVKQTLNFDLLKLAGDVESNPGP
SEQ ID NO: 54	VKQTLNFDLLKLAGDVESNPGP
SEQ ID NO: 55	LLKLAGDVESNPGP
SEQ ID NO: 56	LLNFDLLKLAGDVESNPGP
SEQ ID NO: 57	TLNFDLLKLAGDVESNPGP
SEQ ID NO: 58	LLKLAGDVESNPGP
SEQ ID NO: 59	NFDLLKLAGDVESNPGP
SEQ ID NO: 60	QLLNFDLLKLAGDVESNPGP
SEQ ID NO: 61	APVKQTLNFDLLKLAGDVESNPGP
SEQ ID NO: 62	VTELLYRMKRAETYCPRPLLAIHPTEARHKQKIVAPVKQT
SEQ ID NO: 63	LNFDLLKLAGDVESNPGP
SEQ ID NO: 64	LLAIHPTEARHKQKIVAPVKQTLNFDLLKLAGDVESNPGP
SEQ ID NO: 65	EARHKQKIVAPVKQTLNFDLLKLAGDVESNPGP

#### E. POLYNUCLEOTIDES

In particular embodiments, polynucleotides encoding one or more homing endonuclease variants, megaTALs, end-processing enzymes, and fusion polypeptides contemplated herein are provided. As used herein, the terms “polynucleotide” or “nucleic acid” refer to deoxyribonucleic acid (DNA), ribonucleic acid (RNA) and DNA/RNA hybrids. Polynucleotides may be single-stranded or double-stranded and either recombinant, synthetic, or isolated. Polynucleotides include, but are not limited to: pre-messenger RNA (pre-mRNA), messenger RNA (mRNA), RNA, short interfering RNA (siRNA), short hairpin RNA (shRNA), microRNA (miRNA), ribozymes, genomic RNA (gRNA), plus strand RNA (RNA(+)), minus strand RNA (RNA(-)), tracrRNA, crRNA, single guide RNA (sgRNA),

synthetic RNA, synthetic mRNA, genomic DNA (gDNA), PCR amplified DNA, complementary DNA (cDNA), synthetic DNA, or recombinant DNA. Polynucleotides refer to a polymeric form of nucleotides of at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 40, at least 50, at least 100, at least 200, at least 300, at least 400, at least 500, 5 at least 1000, at least 5000, at least 10000, or at least 15000 or more nucleotides in length, either ribonucleotides or deoxyribonucleotides or a modified form of either type of nucleotide, as well as all intermediate lengths. It will be readily understood that "intermediate lengths," in this context, means any length between the quoted values, such as 6, 7, 8, 9, *etc.*, 101, 102, 103, *etc.*; 151, 152, 153, *etc.*; 201, 202, 203, *etc.* In particular embodiments, polynucleotides or 10 variants have at least or about 50%, 55%, 60%, 65%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to a reference sequence.

In particular embodiments, polynucleotides may be codon-optimized. As used herein, the term "codon-optimized" refers to substituting codons in a polynucleotide encoding a 15 polypeptide in order to increase the expression, stability and/or activity of the polypeptide. Factors that influence codon optimization include, but are not limited to one or more of: (i) variation of codon biases between two or more organisms or genes or synthetically constructed bias tables, (ii) variation in the degree of codon bias within an organism, gene, or set of genes, (iii) systematic variation of codons including context, (iv) variation of codons according to 20 their decoding tRNAs, (v) variation of codons according to GC %, either overall or in one position of the triplet, (vi) variation in degree of similarity to a reference sequence for example a naturally occurring sequence, (vii) variation in the codon frequency cutoff, (viii) structural properties of mRNAs transcribed from the DNA sequence, (ix) prior knowledge about the function of the DNA sequences upon which design of the codon substitution set is to be based, 25 (x) systematic variation of codon sets for each amino acid, and/or (xi) isolated removal of spurious translation initiation sites.

As used herein the term "nucleotide" refers to a heterocyclic nitrogenous base in N-glycosidic linkage with a phosphorylated sugar. Nucleotides are understood to include natural bases, and a wide variety of art-recognized modified bases. Such bases are generally located at 30 the 1' position of a nucleotide sugar moiety. Nucleotides generally comprise a base, sugar and

a phosphate group. In ribonucleic acid (RNA), the sugar is a ribose, and in deoxyribonucleic acid (DNA) the sugar is a deoxyribose, *i.e.*, a sugar lacking a hydroxyl group that is present in ribose. Exemplary natural nitrogenous bases include the purines, adenosine (A) and guanidine (G), and the pyrimidines, cytidine (C) and thymidine (T) (or in the context of RNA, uracil (U)).

5 The C-1 atom of deoxyribose is bonded to N-1 of a pyrimidine or N-9 of a purine. Nucleotides are usually mono, di- or triphosphates. The nucleotides can be unmodified or modified at the sugar, phosphate and/or base moiety, (also referred to interchangeably as nucleotide analogs, nucleotide derivatives, modified nucleotides, non-natural nucleotides, and non-standard nucleotides; see for example, WO 92/07065 and WO 93/15187). Examples of modified 10 nucleic acid bases are summarized by Limbach *et al.*, (1994, *Nucleic Acids Res.* 22, 2183-2196).

A nucleotide may also be regarded as a phosphate ester of a nucleoside, with esterification occurring on the hydroxyl group attached to C-5 of the sugar. As used herein, the term “nucleoside” refers to a heterocyclic nitrogenous base in N-glycosidic linkage with a 15 sugar. Nucleosides are recognized in the art to include natural bases, and also to include well known modified bases. Such bases are generally located at the 1' position of a nucleoside sugar moiety. Nucleosides generally comprise a base and sugar group. The nucleosides can be unmodified or modified at the sugar, and/or base moiety, (also referred to interchangeably as nucleoside analogs, nucleoside derivatives, modified nucleosides, non-natural nucleosides, or 20 non-standard nucleosides). As also noted above, examples of modified nucleic acid bases are summarized by Limbach *et al.*, (1994, *Nucleic Acids Res.* 22, 2183-2196).

Illustrative examples of polynucleotides include, but are not limited to polynucleotides encoding SEQ ID NOs: 1-12 and 22-24, and polynucleotide sequences set forth in SEQ ID NOs: 13-16 and 25-27.

25 In various illustrative embodiments, polynucleotides contemplated herein include, but are not limited to polynucleotides encoding homing endonuclease variants, megaTALs, end-processing enzymes, fusion polypeptides, and expression vectors, viral vectors, and transfer plasmids comprising polynucleotides contemplated herein.

As used herein, the terms “polynucleotide variant” and “variant” and the like refer to 30 polynucleotides displaying substantial sequence identity with a reference polynucleotide

sequence or polynucleotides that hybridize with a reference sequence under stringent conditions that are defined hereinafter. These terms also encompass polynucleotides that are distinguished from a reference polynucleotide by the addition, deletion, substitution, or modification of at least one nucleotide. Accordingly, the terms “polynucleotide variant” and 5 “variant” include polynucleotides in which one or more nucleotides have been added or deleted, or modified, or replaced with different nucleotides. In this regard, it is well understood in the art that certain alterations inclusive of mutations, additions, deletions and substitutions can be made to a reference polynucleotide whereby the altered polynucleotide retains the biological function or activity of the reference polynucleotide.

10 In one embodiment, a polynucleotide comprises a nucleotide sequence that hybridizes to a target nucleic acid sequence under stringent conditions. To hybridize under “stringent conditions” describes hybridization protocols in which nucleotide sequences at least 60% identical to each other remain hybridized. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (Tm) for the specific sequence at a defined 15 ionic strength and pH. The Tm is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Since the target sequences are generally present at excess, at Tm, 50% of the probes are occupied at equilibrium.

20 The recitations “sequence identity” or, for example, comprising a “sequence 50% identical to,” as used herein, refer to the extent that sequences are identical on a nucleotide-by-nucleotide basis or an amino acid-by-amino acid basis over a window of comparison. Thus, a “percentage of sequence identity” may be calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (e.g., A, T, C, G, I) or the identical amino acid residue (e.g., Ala, 25 Pro, Ser, Thr, Gly, Val, Leu, Ile, Phe, Tyr, Trp, Lys, Arg, His, Asp, Glu, Asn, Gln, Cys and Met) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. Included are nucleotides and polypeptides having at least about 50%, 55%, 60%, 65%, 70%, 30 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to any of the

reference sequences described herein, typically where the polypeptide variant maintains at least one biological activity of the reference polypeptide.

Terms used to describe sequence relationships between two or more polynucleotides or polypeptides include “reference sequence,” “comparison window,” “sequence identity,” “percentage of sequence identity,” and “substantial identity”. A “reference sequence” is at least 12 but frequently 15 to 18 and often at least 25 monomer units, inclusive of nucleotides and amino acid residues, in length. Because two polynucleotides may each comprise (1) a sequence (*i.e.*, only a portion of the complete polynucleotide sequence) that is similar between the two polynucleotides, and (2) a sequence that is divergent between the two polynucleotides, sequence comparisons between two (or more) polynucleotides are typically performed by comparing sequences of the two polynucleotides over a “comparison window” to identify and compare local regions of sequence similarity. A “comparison window” refers to a conceptual segment of at least 6 contiguous positions, usually about 50 to about 100, more usually about 100 to about 150 in which a sequence is compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. The comparison window may comprise additions or deletions (*i.e.*, gaps) of about 20% or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Optimal alignment of sequences for aligning a comparison window may be conducted by computerized implementations of algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Drive Madison, WI, USA) or by inspection and the best alignment (*i.e.*, resulting in the highest percentage homology over the comparison window) generated by any of the various methods selected. Reference also may be made to the BLAST family of programs as for example disclosed by Altschul *et al.*, 1997, *Nucl. Acids Res.* 25:3389. A detailed discussion of sequence analysis can be found in Unit 19.3 of Ausubel *et al.*, *Current Protocols in Molecular Biology*, John Wiley & Sons Inc., 1994-1998, Chapter 15.

An “isolated polynucleotide,” as used herein, refers to a polynucleotide that has been purified from the sequences which flank it in a naturally-occurring state, *e.g.*, a DNA fragment that has been removed from the sequences that are normally adjacent to the fragment. In particular embodiments, an “isolated polynucleotide” refers to a complementary DNA

(cDNA), a recombinant polynucleotide, a synthetic polynucleotide, or other polynucleotide that does not exist in nature and that has been made by the hand of man.

In various embodiments, a polynucleotide comprises an mRNA encoding a polypeptide contemplated herein including, but not limited to, a homing endonuclease variant, a megaTAL, 5 and an end-processing enzyme. In certain embodiments, the mRNA comprises a cap, one or more nucleotides, and a poly(A) tail.

As used herein, the terms “5’ cap” or “5’ cap structure” or “5’ cap moiety” refer to a chemical modification, which has been incorporated at the 5’ end of an mRNA. The 5’ cap is involved in nuclear export, mRNA stability, and translation.

10 In particular embodiments, a mRNA contemplated herein comprises a 5’ cap comprising a 5’-PPP-5’-triphosphate linkage between a terminal guanosine cap residue and the 5’-terminal transcribed sense nucleotide of the mRNA molecule. This 5’-guanylate cap may then be methylated to generate an N7-methyl-guanylate residue.

15 Illustrative examples of 5’ cap suitable for use in particular embodiments of the mRNA polynucleotides contemplated herein include, but are not limited to: unmethylated 5’ cap analogs, e.g., G(5’)PPP(5’)G, G(5’)PPP(5’)C, G(5’)PPP(5’)A; methylated 5’ cap analogs, e.g., m<sup>7</sup>G(5’)PPP(5’)G, m<sup>7</sup>G(5’)PPP(5’)C, and m<sup>7</sup>G(5’)PPP(5’)A; dimethylated 5’ cap analogs, e.g., m<sup>2,7</sup>G(5’)PPP(5’)G, m<sup>2,7</sup>G(5’)PPP(5’)C, and m<sup>2,7</sup>G(5’)PPP(5’)A; trimethylated 5’ cap analogs, e.g., m<sup>2,2,7</sup>G(5’)PPP(5’)G, m<sup>2,2,7</sup>G(5’)PPP(5’)C, and m<sup>2,2,7</sup>G(5’)PPP(5’)A; dimethylated 20 symmetrical 5’ cap analogs, e.g., m<sup>7</sup>G(5’)PPPM<sup>7</sup>(5’)G, m<sup>7</sup>G(5’)PPPM<sup>7</sup>(5’)C, and m<sup>7</sup>G(5’)PPPM<sup>7</sup>(5’)A; and anti-reverse 5’ cap analogs, e.g. Anti-Reverse Cap Analog (ARCA) cap, designated 3’O-Me-m<sup>7</sup>G(5’)PPP(5’)G, 2’O-Me-m<sup>7</sup>G(5’)PPP(5’)G, 2’O-Me- 25 m<sup>7</sup>G(5’)PPP(5’)C, 2’O-Me-m<sup>7</sup>G(5’)PPP(5’)A, m<sup>7</sup>2’d(5’)PPP(5’)G, m<sup>7</sup>2’d(5’)PPP(5’)C, m<sup>7</sup>2’d(5’)PPP(5’)A, 3’O-Me-m<sup>7</sup>G(5’)PPP(5’)C, 3’O-Me-m<sup>7</sup>G(5’)PPP(5’)A, m<sup>7</sup>3’d(5’)PPP(5’)G, m<sup>7</sup>3’d(5’)PPP(5’)C, m<sup>7</sup>3’d(5’)PPP(5’)A and their tetraphosphate derivatives) (see, e.g., Jemielity *et al.*, RNA, 9: 1108-1122 (2003)).

In particular embodiments, mRNAs comprise a 5’ cap that is a 7-methyl guanylate (“m<sup>7</sup>G”) linked via a triphosphate bridge to the 5’-end of the first transcribed nucleotide, resulting in m<sup>7</sup>G(5’)PPP(5’)N, where N is any nucleoside.

In some embodiments, mRNAs comprise a 5' cap wherein the cap is a Cap0 structure (Cap0 structures lack a 2'-O-methyl residue of the ribose attached to bases 1 and 2), a Cap1 structure (Cap1 structures have a 2'-O-methyl residue at base 2), or a Cap2 structure (Cap2 structures have a 2'-O-methyl residue attached to both bases 2 and 3).

5 In one embodiment, an mRNA comprises a m<sup>7</sup>G(5')ppp(5')G cap.

In one embodiment, an mRNA comprises an ARCA cap.

In particular embodiments, an mRNA contemplated herein comprises one or more modified nucleosides.

In one embodiment, an mRNA comprises one or more modified nucleosides selected 10 from the group consisting of: pseudouridine, pyridin-4-one ribonucleoside, 5-aza-uridine, 2-thio-5-aza-uridine, 2-thiouridine, 4-thio-pseudouridine, 2-thio-pseudouridine, 5-hydroxyuridine, 3-methyluridine, 5-carboxymethyl-uridine, 1-carboxymethyl-pseudouridine, 5-propynyl-uridine, 1-propynyl-pseudouridine, 5-taurinomethyluridine, 1-taurinomethyl-pseudouridine, 5-taurinomethyl-2-thio-uridine, 1-taurinomethyl-4-thio-uridine, 5-methyl-uridine, 1-methyl-pseudouridine, 4-thio-1-methyl-pseudouridine, 2-thio-1-methyl-pseudouridine, 1-methyl-1-deaza-pseudouridine, 2-thio-1-methyl-1-deaza-pseudouridine, dihydrouridine, dihydropseudouridine, 2-thio-dihhydrouridine, 2-thio-dihydropseudouridine, 2-methoxyuridine, 2-methoxy-4-thio-uridine, 4-methoxy-pseudouridine, 4-methoxy-2-thio-pseudouridine, 5-aza-cytidine, pseudoisocytidine, 3-methyl-cytidine, N4-acetylcytidine, 5-formylcytidine, N4-methylcytidine, 5-hydroxymethylcytidine, 1-methyl-pseudoisocytidine, pyrrolo-cytidine, pyrrolo-pseudoisocytidine, 2-thio-cytidine, 2-thio-5-methyl-cytidine, 4-thio-pseudoisocytidine, 4-thio-1-methyl-pseudoisocytidine, 4-thio-1-methyl-1-deaza-pseudoisocytidine, 1-methyl-1-deaza-pseudoisocytidine, zebularine, 5-aza-zebularine, 5-methyl-zebularine, 5-aza-2-thio-zebularine, 2-thio-zebularine, 2-methoxy-cytidine, 2-methoxy- 25 5-methyl-cytidine, 4-methoxy-pseudoisocytidine, 4-methoxy-1-methyl-pseudoisocytidine, 2-aminopurine, 2,6-diaminopurine, 7-deaza-adenine, 7-deaza-8-aza-adenine, 7-deaza-2-aminopurine, 7-deaza-8-aza-2-aminopurine, 7-deaza-2,6-diaminopurine, 7-deaza-8-aza-2,6-diaminopurine, 1-methyladenosine, N6-methyladenosine, N6-isopentenyladenosine, N6-(cis-hydroxyisopentenyl)adenosine, 2-methylthio-N6-(cis-hydroxyisopentenyl) adenosine, N6-glycylcarbamoyladenine, N6-threonylcarbamoyladenine, 2-methylthio-N6-threonyl

carbamoyladenosine, N6,N6-dimethyladenosine, 7-methyladenine, 2-methylthio-adenine, 2-methoxy-adenine, inosine, 1-methyl-inosine, wyo sine, wybutosine, 7-deaza-guanosine, 7-deaza-8-aza-guanosine, 6-thio-guanosine, 6-thio-7-deaza-guanosine, 6-thio-7-deaza-8-aza-guanosine, 7-methyl-guanosine, 6-thio-7-methyl-guanosine, 7-methylinosine, 6-methoxy-5 guanosine, 1-methylguanosine, N2-methylguanosine, N2,N2-dimethylguanosine, 8-oxo-guanosine, 7-methyl-8-oxo-guanosine, 1-methyl-6-thio-guanosine, N2-methyl-6-thio-guanosine, and N2,N2-dimethyl-6-thio-guanosine.

In one embodiment, an mRNA comprises one or more modified nucleosides selected from the group consisting of: pseudouridine, pyridin-4-one ribonucleoside, 5-aza-uridine, 2-thio-5-aza-uridine, 2-thiouridine, 4-thio-pseudouridine, 2-thio-pseudouridine, 5-hydroxyuridine, 3-methyluridine, 5-carboxymethyl-uridine, 1-carboxymethyl-pseudouridine, 5-propynyl-uridine, 1-propynyl-pseudouridine, 5-taurinomethyluridine, 1-taurinomethyl-pseudouridine, 5-taurinomethyl-2-thio-uridine, 1-taurinomethyl-4-thio-uridine, 5-methyl-uridine, 1-methyl-pseudouridine, 4-thio-1-methyl-pseudouridine, 2-thio-1-methyl-15 pseudouridine, 1-methyl-1-deaza-pseudouridine, 2-thio-1-methyl-1-deaza-pseudouridine, dihydrouridine, dihydropseudouridine, 2-thio-dihhydrouridine, 2-thio-dihydropseudouridine, 2-methoxyuridine, 2-methoxy-4-thio-uridine, 4-methoxy-pseudouridine, and 4-methoxy-2-thio-pseudouridine.

In one embodiment, an mRNA comprises one or more modified nucleosides selected from the group consisting of: 5-aza-cytidine, pseudoisocytidine, 3-methyl-cytidine, N4-acetylcytidine, 5-formylcytidine, N4-methylcytidine, 5-hydroxymethylcytidine, 1-methyl-pseudoisocytidine, pyrrolo-cytidine, pyrrolo-pseudoisocytidine, 2-thio-cytidine, 2-thio-5-methyl-cytidine, 4-thio-pseudoisocytidine, 4-thio-1-methyl-pseudoisocytidine, 4-thio-1-methyl-1-deaza-pseudoisocytidine, 1-methyl-1-deaza-pseudoisocytidine, zebularine, 5-aza-25 zebularine, 5-methyl-zebularine, 5-aza-2-thio-zebularine, 2-thio-zebularine, 2-methoxy-cytidine, 2-methoxy-5-methyl-cytidine, 4-methoxy-pseudoisocytidine, and 4-methoxy-1-methyl-pseudoisocytidine.

In one embodiment, an mRNA comprises one or more modified nucleosides selected from the group consisting of: 2-aminopurine, 2,6-diaminopurine, 7-deaza-adenine, 7-deaza-8-30 aza-adenine, 7-deaza-2-aminopurine, 7-deaza-8-aza-2-aminopurine, 7-deaza-2,6-

diaminopurine, 7-deaza-8-aza-2,6-diaminopurine, 1-methyladenosine, N6-methyladenosine, N6-isopentenyladenosine, N6-(*cis*-hydroxyisopentenyl)adenosine, 2-methylthio-N6-(*cis*-hydroxyisopentenyl) adenosine, N6-glycylcarbamoyladenosine, N6-threonylcarbamoyladenosine, 2-methylthio-N6-threonyl carbamoyladenosine, N6,N6-  
5 dimethyladenosine, 7-methyladenine, 2-methylthio-adenine, and 2-methoxy-adenine.

In one embodiment, an mRNA comprises one or more modified nucleosides selected from the group consisting of: inosine, 1-methyl-inosine, wyosine, wybudosine, 7-deaza-guanosine, 7-deaza-8-aza-guanosine, 6-thio-guanosine, 6-thio-7-deaza-guanosine, 6-thio-7-deaza-8-aza-guanosine, 7-methyl-guanosine, 6-thio-7-methyl-guanosine, 7-methylinosine, 6-  
10 methoxy-guanosine, 1-methylguanosine, N2-methylguanosine, N2,N2-dimethylguanosine, 8-oxo-guanosine, 7-methyl-8-oxo-guanosine, 1-methyl-6-thio-guanosine, N2-methyl-6-thio-guanosine, and N2,N2-dimethyl-6-thio-guanosine.

In one embodiment, an mRNA comprises one or more pseudouridines, one or more 5-methyl-cytosines, and/or one or more 5-methyl-cytidines.

15 In one embodiment, an mRNA comprises one or more pseudouridines.

In one embodiment, an mRNA comprises one or more 5-methyl-cytidines.

In one embodiment, an mRNA comprises one or more 5-methyl-cytosines.

In particular embodiments, an mRNA contemplated herein comprises a poly(A) tail to help protect the mRNA from exonuclease degradation, stabilize the mRNA, and facilitate  
20 translation. In certain embodiments, an mRNA comprises a 3' poly(A) tail structure.

In particular embodiments, the length of the poly(A) tail is at least about 10, 25, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, or at least about 500 or more adenine nucleotides or any intervening number of adenine nucleotides. In particular embodiments, the length of the poly(A) tail is at least about 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 25 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 202, 203, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 30 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251,

252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, or 275 or more adenine nucleotides.

In particular embodiments, the length of the poly(A) tail is about 10 to about 500 adenine nucleotides, about 50 to about 500 adenine nucleotides, about 100 to about 500 adenine nucleotides, about 150 to about 500 adenine nucleotides, about 200 to about 500 adenine nucleotides, about 250 to about 500 adenine nucleotides, about 300 to about 500 adenine nucleotides, about 50 to about 450 adenine nucleotides, about 50 to about 400 adenine nucleotides, about 50 to about 350 adenine nucleotides, about 100 to about 500 adenine nucleotides, about 100 to about 450 adenine nucleotides, about 100 to about 400 adenine nucleotides, about 100 to about 350 adenine nucleotides, about 100 to about 300 adenine nucleotides, about 150 to about 500 adenine nucleotides, about 150 to about 450 adenine nucleotides, about 150 to about 400 adenine nucleotides, about 150 to about 350 adenine nucleotides, about 150 to about 300 adenine nucleotides, about 150 to about 250 adenine nucleotides, about 150 to about 200 adenine nucleotides, about 200 to about 500 adenine nucleotides, about 200 to about 450 adenine nucleotides, about 200 to about 400 adenine nucleotides, about 200 to about 350 adenine nucleotides, about 200 to about 300 adenine nucleotides, about 250 to about 500 adenine nucleotides, about 250 to about 450 adenine nucleotides, about 250 to about 400 adenine nucleotides, about 250 to about 350 adenine nucleotides, or about 250 to about 300 adenine nucleotides or any intervening range of adenine nucleotides.

Terms that describe the orientation of polynucleotides include: 5' (normally the end of the polynucleotide having a free phosphate group) and 3' (normally the end of the polynucleotide having a free hydroxyl (OH) group). Polynucleotide sequences can be annotated in the 5' to 3' orientation or the 3' to 5' orientation. For DNA and mRNA, the 5' to 3' strand is designated the "sense," "plus," or "coding" strand because its sequence is identical to the sequence of the pre-messenger (pre-mRNA) [except for uracil (U) in RNA, instead of thymine (T) in DNA]. For DNA and mRNA, the complementary 3' to 5' strand which is the strand transcribed by the RNA polymerase is designated as "template," "antisense," "minus," or "non-coding" strand. As used herein, the term "reverse orientation" refers to a 5' to 3'

sequence written in the 3' to 5' orientation or a 3' to 5' sequence written in the 5' to 3' orientation.

The terms "complementary" and "complementarity" refer to polynucleotides (*i.e.*, a sequence of nucleotides) related by the base-pairing rules. For example, the complementary 5 strand of the DNA sequence 5' A G T C A T G 3' is 3' T C A G T A C 5'. The latter sequence is often written as the reverse complement with the 5' end on the left and the 3' end on the right, 5' C A T G A C T 3'. A sequence that is equal to its reverse complement is said to be a palindromic sequence. Complementarity can be "partial," in which only some of the nucleic acids' bases are matched according to the base pairing rules. Or, there can be "complete" or 10 "total" complementarity between the nucleic acids.

Polynucleotides can be prepared, manipulated, expressed and/or delivered using any of a variety of well-established techniques known and available in the art. In order to express a desired polypeptide, a nucleotide sequence encoding the polypeptide, can be inserted into appropriate vector. A desired polypeptide can also be expressed by delivering an mRNA 15 encoding the polypeptide into the cell.

Illustrative examples of vectors include, but are not limited to plasmid, autonomously replicating sequences, and transposable elements, *e.g.*, Sleeping Beauty, PiggyBac.

Additional illustrative examples of vectors include, without limitation, plasmids, phagemids, cosmids, artificial chromosomes such as yeast artificial chromosome (YAC), 20 bacterial artificial chromosome (BAC), or P1-derived artificial chromosome (PAC), bacteriophages such as lambda phage or M13 phage, and animal viruses.

Illustrative examples of viruses useful as vectors include, without limitation, retrovirus (including lentivirus and integration deficient lentivirus), adenovirus, adeno-associated virus, herpesvirus (*e.g.*, herpes simplex virus), poxvirus, baculovirus, papillomavirus, and 25 papovavirus (*e.g.*, SV40).

Illustrative examples of expression vectors include, but are not limited to pClneo vectors (Promega) for expression in mammalian cells; pLenti4/V5-DEST<sup>TM</sup>, pLenti6/V5-DEST<sup>TM</sup>, and pLenti6.2/V5-GW/lacZ (Invitrogen) for lentivirus-mediated gene transfer and expression in mammalian cells. In particular embodiments, coding sequences of polypeptides

disclosed herein can be ligated into such expression vectors for the expression of the polypeptides in mammalian cells.

In particular embodiments, the vector is an episomal vector or a vector that is maintained extrachromosomally. As used herein, the term “episomal” refers to a vector that is able to replicate without integration into host’s chromosomal DNA and without gradual loss from a dividing host cell also meaning that said vector replicates extrachromosomally or episomally.

In particular embodiments, to achieve efficient translation of each of the plurality of polypeptides, the polynucleotide sequences can be separated by one or more IRES sequences or polynucleotide sequences encoding self-cleaving polypeptides.

As used herein, an “internal ribosome entry site” or “IRES” refers to an element that promotes direct internal ribosome entry to the initiation codon, such as ATG, of a cistron (a protein encoding region), thereby leading to the cap-independent translation of the gene. *See, e.g.*, Jackson *et al.*, 1990. *Trends Biochem Sci* 15(12):477-83) and Jackson and Kaminski.

15 1995. *RNA* 1(10):985-1000. Examples of IRES generally employed by those of skill in the art include those described in U.S. Pat. No. 6,692,736. Further examples of “IRES” known in the art include, but are not limited to IRES obtainable from picornavirus (Jackson *et al.*, 1990) and IRES obtainable from viral or cellular mRNA sources, such as for example, immunoglobulin heavy-chain binding protein (BiP), the vascular endothelial growth factor (VEGF) (Huez *et al.* 20 1998. *Mol. Cell. Biol.* 18(11):6178-6190), the fibroblast growth factor 2 (FGF-2), and insulin-like growth factor (IGFII), the translational initiation factor eIF4G and yeast transcription factors TFIID and HAP4, the encephelomyocarditis virus (EMCV) which is commercially available from Novagen (Duke *et al.*, 1992. *J. Virol* 66(3):1602-9) and the VEGF IRES (Huez *et al.*, 1998. *Mol Cell Biol* 18(11):6178-90). IRES have also been reported in viral genomes of 25 Picornaviridae, Dicistroviridae and Flaviviridae species and in HCV, Friend murine leukemia virus (FrMLV) and Moloney murine leukemia virus (MoMLV).

In one embodiment, the IRES used in polynucleotides contemplated herein is an EMCV IRES.

30 In particular embodiments, the polynucleotides comprise polynucleotides that have a consensus Kozak sequence and that encode a desired polypeptide. As used herein, the term

“Kozak sequence” refers to a short nucleotide sequence that greatly facilitates the initial binding of mRNA to the small subunit of the ribosome and increases translation. The consensus Kozak sequence is (GCC)RCCATGG (SEQ ID NO: 66), where R is a purine (A or G) (Kozak, 1986. *Cell*. 44(2):283-92, and Kozak, 1987. *Nucleic Acids Res*. 15(20):8125-48).

5 Elements directing the efficient termination and polyadenylation of the heterologous nucleic acid transcripts increases heterologous gene expression.

In particular embodiments, vectors comprise a polyadenylation sequence 3' of a polynucleotide encoding a polypeptide to be expressed. The term “poly(A) site” or “poly(A) sequence” as used herein denotes a DNA sequence which directs both the termination and 10 polyadenylation of the nascent RNA transcript by RNA polymerase II. Polyadenylation sequences can promote mRNA stability by addition of a poly(A) tail to the 3' end of the coding sequence and thus, contribute to increased translational efficiency. Cleavage and polyadenylation is directed by a poly(A) sequence in the RNA. The core poly(A) sequence for mammalian pre-mRNAs has two recognition elements flanking a cleavage-polyadenylation 15 site. Typically, an almost invariant AAUAAA hexamer lies 20-50 nucleotides upstream of a more variable element rich in U or GU residues. Cleavage of the nascent transcript occurs between these two elements and is coupled to the addition of up to 250 adenosines to the 5' cleavage product. In particular embodiments, the core poly(A) sequence is a synthetic poly(A) sequence (e.g., AATAAA, ATTAAA, AGTAAA). Illustrative examples of poly(A) sequences 20 include, but are not limited to an SV40 poly(A) sequence, a bovine growth hormone poly(A) sequence (BGH<sub>p</sub>A), a rabbit β-globin poly(A) sequence (rβgpA), or another suitable heterologous or endogenous poly(A) sequence known in the art.

In particular embodiments, polynucleotides encoding one or more nuclease variants, megaTALs, end-processing enzymes, or fusion polypeptides may be introduced into 25 hematopoietic cells, e.g., T cells, by both non-viral and viral methods. In particular embodiments, delivery of one or more polynucleotides encoding nucleases and/or donor repair templates may be provided by the same method or by different methods, and/or by the same vector or by different vectors.

The term “vector” is used herein to refer to a nucleic acid molecule capable transferring 30 or transporting another nucleic acid molecule. The transferred nucleic acid is generally linked

to, *e.g.*, inserted into, the vector nucleic acid molecule. A vector may include sequences that direct autonomous replication in a cell, or may include sequences sufficient to allow integration into host cell DNA. In particular embodiments, non-viral vectors are used to deliver one or more polynucleotides contemplated herein to a T cell.

5 Illustrative examples of non-viral vectors include, but are not limited to plasmids (*e.g.*, DNA plasmids or RNA plasmids), transposons, cosmids, and bacterial artificial chromosomes.

10 Illustrative methods of non-viral delivery of polynucleotides contemplated in particular embodiments include, but are not limited to: electroporation, sonoporation, lipofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, nanoparticles, polycation or lipid:nucleic acid conjugates, naked DNA, artificial virions, DEAE-dextran-mediated transfer, gene gun, and heat-shock.

15 Illustrative examples of polynucleotide delivery systems suitable for use in particular embodiments contemplated in particular embodiments include, but are not limited to those provided by Amaxa Biosystems, Maxcyte, Inc., BTX Molecular Delivery Systems, and Copernicus Therapeutics Inc. Lipofection reagents are sold commercially (*e.g.*, Transfectam™ and Lipofectin™). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides have been described in the literature. See *e.g.*, Liu *et al.* (2003) *Gene Therapy*. 10:180–187; and Balazs *et al.* (2011) *Journal of Drug*  
20 *Delivery*. 2011:1-12. Antibody-targeted, bacterially derived, non-living nanocell-based delivery is also contemplated in particular embodiments.

## F. CELLS

25 The genome edited cells manufactured by the methods contemplated in particular embodiments comprise one or more gene edits in a TCR $\alpha$  gene and provide improved cell-based therapeutics for the prevention, treatment, or amelioration of at least one symptom, of a cancer, GVHD, infectious disease, autoimmune disease, immunodeficiency or condition associated therewith. Without wishing to be bound to any particular theory, it is believed that the genome edited immune effector cells manufactured by the methods contemplated herein

are imbued with superior properties, including increased improved safety, efficacy, and durability *in vivo*.

Genome edited cells contemplated in particular embodiments may be autologous/autogeneic (“self”) or non-autologous (“non-self,” *e.g.*, allogeneic, syngeneic or 5 xenogeneic). “Autologous,” as used herein, refers to cells from the same subject. “Allogeneic,” as used herein, refers to cells of the same species that differ genetically to the cell in comparison. “Syngeneic,” as used herein, refers to cells of a different subject that are genetically identical to the cell in comparison. “Xenogeneic,” as used herein, refers to cells of a different species to the cell in comparison. In preferred embodiments, the cells are obtained 10 from a mammalian subject. In a more preferred embodiment, the cells are obtained from a primate subject, optionally a non-human primate. In the most preferred embodiment, the cells are obtained from a human subject.

An “isolated cell” refers to a non-naturally occurring cell, *e.g.*, a cell that does not exist in nature, a modified cell, an engineered cell, *etc.*, that has been obtained from an *in vivo* tissue 15 or organ and is substantially free of extracellular matrix.

As used herein, the term “population of cells” refers to a plurality of cells that may be made up of any number and/or combination of homogenous or heterogeneous cell types, as described elsewhere herein. For example, for transduction of T cells, a population of cells may be isolated or obtained from peripheral blood. A population of cells may comprise about 10%, 20 about 20%, about 30%, about 40%, about 50%, about 60%, about 70%, about 80%, about 90%, or about 100% of the target cell type to be edited. In certain embodiments, T cells may be isolated or purified from a population of heterogeneous cells using methods known in the art.

Illustrative examples of cell types whose genome can be edited using the compositions 25 and methods contemplated herein include, but are not limited to, cell lines, primary cells, stem cells, progenitor cells, and differentiated cells, and mixtures thereof.

In a preferred embodiment, the genome editing compositions and methods are used to edit hematopoietic cells, more preferably immune cells, and even more preferably T cells.

The terms “T cell” or “T lymphocyte” are art-recognized and are intended to include 30 thymocytes, immune effector cells, regulatory T cells, naïve T lymphocytes, immature T

lymphocytes, mature T lymphocytes, resting T lymphocytes, or activated T lymphocytes. A T cell can be a T helper (Th) cell, for example a T helper 1 (Th1) or a T helper 2 (Th2) cell. The T cell can be a helper T cell (HTL; CD4<sup>+</sup> T cell) CD4<sup>+</sup> T cell, a cytotoxic T cell (CTL; CD8<sup>+</sup> T cell), a tumor infiltrating cytotoxic T cell (TIL; CD8<sup>+</sup> T cell), CD4<sup>+</sup>CD8<sup>+</sup> T cell, CD4<sup>+</sup>CD8<sup>-</sup> T 5 cell, or any other subset of T cells. In one embodiment, the T cell is an immune effector cell. In one embodiment, the T cell is an NKT cell. Other illustrative populations of T cells suitable for use in particular embodiments include naïve T cells and memory T cells.

In various embodiments, genome edited cells comprise immune effector cells comprising a TCR $\alpha$  gene edited by the compositions and methods contemplated herein. 10 An “immune effector cell,” is any cell of the immune system that has one or more effector functions (e.g., cytotoxic cell killing activity, secretion of cytokines, induction of ADCC and/or CDC). Illustrative immune effector cells contemplated in particular embodiments are T lymphocytes, in particular cytotoxic T cells (CTLs; CD8<sup>+</sup> T cells), TILs, and helper T cells (HTLs; CD4<sup>+</sup> T cells). In one embodiment, immune effector cells include natural 15 killer (NK) cells. In one embodiment, immune effector cells include natural killer T (NKT) cells.

“Potent T cells,” and “young T cells,” are used interchangeably in particular embodiments and refer to T cell phenotypes wherein the T cell is capable of proliferation and a concomitant decrease in differentiation. In particular embodiments, the young T cell has the 20 phenotype of a “naïve T cell.” In particular embodiments, young T cells comprise one or more of, or all of the following biological markers: CD62L, CCR7, CD28, CD27, CD122, CD127, CD197, and CD38. In one embodiment, young T cells comprise one or more of, or all of the following biological markers: CD62L, CD127, CD197, and CD38. In one embodiment, the young T cells lack expression of CD57, CD244, CD160, PD-1, CTLA4, and LAG3.

25 T cells can be obtained from a number of sources including, but not limited to, peripheral blood mononuclear cells, bone marrow, lymph nodes tissue, cord blood, thymus issue, tissue from a site of infection, ascites, pleural effusion, spleen tissue, and tumors.

In particular embodiments, a population of cells comprises immune effector cells or T 30 cells comprising a homing endonuclease variant or megaTAL that targets exon 1 of the constant region of the TCR $\alpha$  gene as contemplated herein.

In particular embodiments, a population of cells comprises immune effector cells or T cells comprising a homing endonuclease variant polypeptide or megaTAL polypeptide that targets exon 1 of the constant region of the TCR $\alpha$  gene as contemplated herein.

5 In particular embodiments, a population of cells comprises immune effector cells or T cells comprising a vector encoding a homing endonuclease variant or megaTAL that targets exon 1 of the constant region of the TCR $\alpha$  gene as contemplated herein.

In particular embodiments, a population of cells comprises immune effector cells or T cells comprising an mRNA encoding a homing endonuclease variant or megaTAL that targets exon 1 of the constant region of the TCR $\alpha$  gene as contemplated herein.

10 **G. COMPOSITIONS AND FORMULATIONS**

The compositions contemplated in particular embodiments may comprise one or more polypeptides, polynucleotides, vectors comprising same, and genome editing compositions and genome edited cell compositions, as contemplated herein. The genome editing compositions and methods contemplated in particular embodiments are useful for editing a target site in the 15 human T cell receptor alpha (TCR $\alpha$ ) gene in a cell or a population of cells. In preferred embodiments, a genome editing composition is used to edit a TCR $\alpha$  gene in a hematopoietic cell, *e.g.*, a T cell or an immune effector cell.

In various embodiments, the compositions contemplated herein comprise a nuclease variant, and optionally an end-processing enzyme, *e.g.*, a 3'-5' exonuclease (Trex2). The 20 nuclease variant may be in the form of an mRNA that is introduced into a cell via polynucleotide delivery methods disclosed *supra*, *e.g.*, electroporation, lipid nanoparticles, *etc.* In one embodiment, a composition comprising an mRNA encoding a homing endonuclease variant or megaTAL, and optionally a 3'-5' exonuclease, is introduced in a cell via polynucleotide delivery methods disclosed *supra*. The composition may be used to generate a 25 genome edited cell or population of genome edited cells by error prone NHEJ.

In various embodiments, the compositions contemplated herein comprise a donor repair template. The composition may be delivered to a cell that expresses or will express nuclease variant, and optionally an end-processing enzyme. In one embodiment, the composition may be delivered to a cell that expresses or will express a homing endonuclease

variant or megaTAL, and optionally a 3'-5' exonuclease. Expression of the gene editing enzymes in the presence of the donor repair template can be used to generate a genome edited cell or population of genome edited cells by HDR.

Compositions include, but are not limited to pharmaceutical compositions. A 5 “pharmaceutical composition” refers to a composition formulated in pharmaceutically-acceptable or physiologically-acceptable solutions for administration to a cell or an animal, either alone, or in combination with one or more other modalities of therapy. It will also be understood that, if desired, the compositions may be administered in combination with other agents as well, such as, *e.g.*, cytokines, growth factors, hormones, small molecules, 10 chemotherapeutics, pro-drugs, drugs, antibodies, or other various pharmaceutically-active agents. There is virtually no limit to other components that may also be included in the compositions, provided that the additional agents do not adversely affect the composition.

The phrase “pharmaceutically acceptable” is employed herein to refer to those compounds, materials, compositions, and/or dosage forms which are, within the scope of sound 15 medical judgment, suitable for use in contact with the tissues of human beings and animals without excessive toxicity, irritation, allergic response, or other problem or complication, commensurate with a reasonable benefit/risk ratio.

The term “pharmaceutically acceptable carrier” refers to a diluent, adjuvant, excipient, or vehicle with which the therapeutic cells are administered. Illustrative examples of 20 pharmaceutical carriers can be sterile liquids, such as cell culture media, water and oils, including those of petroleum, animal, vegetable or synthetic origin, such as peanut oil, soybean oil, mineral oil, sesame oil and the like. Saline solutions and aqueous dextrose and glycerol solutions can also be employed as liquid carriers, particularly for injectable solutions. Suitable pharmaceutical excipients in particular embodiments, include starch, glucose, lactose, sucrose, 25 gelatin, malt, rice, flour, chalk, silica gel, sodium stearate, glycerol monostearate, talc, sodium chloride, dried skim milk, glycerol, propylene, glycol, water, ethanol and the like. Except insofar as any conventional media or agent is incompatible with the active ingredient, its use in the therapeutic compositions is contemplated. Supplementary active ingredients can also be incorporated into the compositions.

In one embodiment, a composition comprising a pharmaceutically acceptable carrier is suitable for administration to a subject. In particular embodiments, a composition comprising a carrier is suitable for parenteral administration, *e.g.*, intravascular (intravenous or intraarterial), intraperitoneal or intramuscular administration.

5 In particular embodiments, a composition comprising a pharmaceutically acceptable carrier is suitable for intraventricular, intraspinal, or intrathecal administration. Pharmaceutically acceptable carriers include sterile aqueous solutions, cell culture media, or dispersions. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is

10 incompatible with the transduced cells, use thereof in the pharmaceutical compositions is contemplated.

In particular embodiments, compositions contemplated herein comprise genetically modified T cells and a pharmaceutically acceptable carrier. A composition comprising a cell-based composition contemplated herein can be administered separately by enteral or

15 parenteral administration methods or in combination with other suitable compounds to effect the desired treatment goals.

The pharmaceutically acceptable carrier must be of sufficiently high purity and of sufficiently low toxicity to render it suitable for administration to the human subject being treated. It further should maintain or increase the stability of the composition. The

20 pharmaceutically acceptable carrier can be liquid or solid and is selected, with the planned manner of administration in mind, to provide for the desired bulk, consistency, *etc.*, when combined with other components of the composition. For example, the pharmaceutically acceptable carrier can be, without limitation, a binding agent (*e.g.*, pregelatinized maize starch, polyvinylpyrrolidone or hydroxypropyl methylcellulose, *etc.*), a filler (*e.g.*, lactose

25 and other sugars, microcrystalline cellulose, pectin, gelatin, calcium sulfate, ethyl cellulose, polyacrylates, calcium hydrogen phosphate, *etc.*), a lubricant (*e.g.*, magnesium stearate, talc, silica, colloidal silicon dioxide, stearic acid, metallic stearates, hydrogenated vegetable oils, corn starch, polyethylene glycols, sodium benzoate, sodium acetate, *etc.*), a disintegrant (*e.g.*, starch, sodium starch glycolate, *etc.*), or a wetting agent (*e.g.*, sodium

30 lauryl sulfate, *etc.*). Other suitable pharmaceutically acceptable carriers for the

compositions contemplated herein include, but are not limited to, water, salt solutions, alcohols, polyethylene glycols, gelatins, amyloses, magnesium stearates, talcs, silicic acids, viscous paraffins, hydroxymethylcelluloses, polyvinylpyrrolidones and the like.

Such carrier solutions also can contain buffers, diluents and other suitable additives. The term “buffer” as used herein refers to a solution or liquid whose chemical makeup neutralizes acids or bases without a significant change in pH. Examples of buffers contemplated herein include, but are not limited to, Dulbecco’s phosphate buffered saline (PBS), Ringer’s solution, 5% dextrose in water (D5W), normal/physiologic saline (0.9% NaCl).

The pharmaceutically acceptable carriers may be present in amounts sufficient to maintain a pH of the composition of about 7. Alternatively, the composition has a pH in a range from about 6.8 to about 7.4, *e.g.*, 6.8, 6.9, 7.0, 7.1, 7.2, 7.3, and 7.4. In still another embodiment, the composition has a pH of about 7.4.

Compositions contemplated herein may comprise a nontoxic pharmaceutically acceptable medium. The compositions may be a suspension. The term “suspension” as used herein refers to non-adherent conditions in which cells are not attached to a solid support. For example, cells maintained as a suspension may be stirred or agitated and are not adhered to a support, such as a culture dish.

In particular embodiments, compositions contemplated herein are formulated in a suspension, where the genome edited T cells are dispersed within an acceptable liquid medium or solution, *e.g.*, saline or serum-free medium, in an intravenous (IV) bag or the like. Acceptable diluents include, but are not limited to water, PlasmaLyte, Ringer’s solution, isotonic sodium chloride (saline) solution, serum-free cell culture medium, and medium suitable for cryogenic storage, *e.g.*, Cryostor® medium.

In certain embodiments, a pharmaceutically acceptable carrier is substantially free of natural proteins of human or animal origin, and suitable for storing a composition comprising a population of genome edited T cells. The therapeutic composition is intended to be administered into a human patient, and thus is substantially free of cell culture components such as bovine serum albumin, horse serum, and fetal bovine serum.

In some embodiments, compositions are formulated in a pharmaceutically acceptable cell culture medium. Such compositions are suitable for administration to human subjects. In particular embodiments, the pharmaceutically acceptable cell culture medium is a serum free medium.

5        Serum-free medium has several advantages over serum containing medium, including a simplified and better defined composition, a reduced degree of contaminants, elimination of a potential source of infectious agents, and lower cost. In various embodiments, the serum-free medium is animal-free, and may optionally be protein-free. Optionally, the medium may contain biopharmaceutically acceptable recombinant 10 proteins. “Animal-free” medium refers to medium wherein the components are derived from non-animal sources. Recombinant proteins replace native animal proteins in animal-free medium and the nutrients are obtained from synthetic, plant or microbial sources. “Protein-free” medium, in contrast, is defined as substantially free of protein.

15      Illustrative examples of serum-free media used in particular compositions includes, but is not limited to QBSF-60 (Quality Biological, Inc.), StemPro-34 (Life Technologies), and X-VIVO 10.

In a preferred embodiment, the compositions comprising genome edited T cells are formulated in PlasmaLyte.

20      In various embodiments, compositions comprising genome edited T cells are formulated in a cryopreservation medium. For example, cryopreservation media with cryopreservation agents may be used to maintain a high cell viability outcome post-thaw. Illustrative examples of cryopreservation media used in particular compositions includes, but is not limited to, CryoStor CS10, CryoStor CS5, and CryoStor CS2.

25      In one embodiment, the compositions are formulated in a solution comprising 50:50 PlasmaLyte A to CryoStor CS10.

30      In particular embodiments, the composition is substantially free of mycoplasma, endotoxin, and microbial contamination. By “substantially free” with respect to endotoxin is meant that there is less endotoxin per dose of cells than is allowed by the FDA for a biologic, which is a total endotoxin of 5 EU/kg body weight per day, which for an average 70 kg person is 350 EU per total dose of cells. In particular embodiments, compositions

comprising hematopoietic stem or progenitor cells transduced with a retroviral vector contemplated herein contain about 0.5 EU/mL to about 5.0 EU/mL, or about 0.5 EU/mL, 1.0 EU/mL, 1.5 EU/mL, 2.0 EU/mL, 2.5 EU/mL, 3.0 EU/mL, 3.5 EU/mL, 4.0 EU/mL, 4.5 EU/mL, or 5.0 EU/mL.

5 In certain embodiments, compositions and formulations suitable for the delivery of polynucleotides are contemplated including, but not limited to, one or more mRNAs encoding one or more reprogrammed nucleases, and optionally end-processing enzymes.

Exemplary formulations for *ex vivo* delivery may also include the use of various transfection agents known in the art, such as calcium phosphate, electroporation, heat 10 shock and various liposome formulations (*i.e.*, lipid-mediated transfection). Liposomes, as described in greater detail below, are lipid bilayers entrapping a fraction of aqueous fluid. DNA spontaneously associates to the external surface of cationic liposomes (by virtue of its charge) and these liposomes will interact with the cell membrane.

In particular embodiments, formulation of pharmaceutically-acceptable carrier 15 solutions is well-known to those of skill in the art, as is the development of suitable dosing and treatment regimens for using the particular compositions described herein in a variety of treatment regimens, including *e.g.*, enteral and parenteral, *e.g.*, intravascular, intravenous, intrarterial, intraosseously, intraventricular, intracerebral, intracranial, 20 intraspinal, intrathecal, and intramedullary administration and formulation. It would be understood by the skilled artisan that particular embodiments contemplated herein may comprise other formulations, such as those that are well known in the pharmaceutical art, and are described, for example, in *Remington: The Science and Practice of Pharmacy*, volume I and volume II. 22<sup>nd</sup> Edition. Edited by Loyd V. Allen Jr. Philadelphia, PA: 25 Pharmaceutical Press; 2012, which is incorporated by reference herein, in its entirety.

25

All publications, patent applications, and issued patents cited in this specification are herein incorporated by reference as if each individual publication, patent application, or issued patent were specifically and individually indicated to be incorporated by reference.

Although the foregoing embodiments have been described in some detail by way of 30 illustration and example for purposes of clarity of understanding, it will be readily apparent to

one of ordinary skill in the art in light of the teachings contemplated herein that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims. The following examples are provided by way of illustration only and not by way of limitation. Those of skill in the art will readily recognize a variety of noncritical  
5 parameters that could be changed or modified to yield essentially similar results.

## EXAMPLES

EXAMPLE 1IMPROVING SELECTIVITY OF AN I-ONU1 VARIANT THAT TARGETS THE HUMAN TCR $\alpha$  GENE

Regions of sub-optimal selectivity were identified in the DNA recognition 5 interfacing regions of an I-OnuI variant (SEQ ID NO: 6) that targets exon 1 in the constant region of the human TCR $\alpha$  gene (“the original TCR $\alpha$  I-OnuI variant”). Positions p4, p5, and p6 in the target site were found to be relatively tolerant to base pair substitutions. The nucleotides differ at these positions in the TCR $\alpha$  on-target site (CCG at p456) (SEQ ID NO: 17) and a frequent off-target site in intron 6 of the KAT2B gene (ACA at p456, *see* 10 Osborne et. al., 2015) (Figure 1, SEQ ID NO: 21). An I-OnuI variant comprising a domain with enhanced selectivity at these positions was developed to improve overall selectivity properties as well as a specific reduction of activity at the KAT2B off-target site (SEQ ID NO: 21).

Inspection of the crystal structure of the wild-type I-OnuI protein (*e.g.*, SEQ ID 15 NOs: 1-5) complexed with its natural target site identified a subset of 8 amino acid residues that are likely contacting or proximal to positions p456. Local analysis of another I-OnuI variant that targets the human TCR $\alpha$  gene identified amino acids that have significantly greater selectivity for CCG in comparison with the original TCR $\alpha$  targeting I-OnuI variant. The amino acids responsible for high CCG triplet selectivity in the TCR $\alpha$  targeting I-OnuI 20 variant were inserted into the original TCR $\alpha$  I-OnuI variant to generated the TCR $\alpha$ 2.0 I-OnuI variant (SEQ ID NO: 7) (Figure 2). TCR $\alpha$ 2.0 I-OnuI variant enzymatic activity was assessed against all 64-nucleotide combinations at the p456 substrate positions using yeast surface display (Figure 3). The TCR $\alpha$ 2.0 I-OnuI variant exhibited significantly better 25 substrate discrimination, marked by cleaving far fewer substrates than the original TCR $\alpha$  I-OnuI variant.

EXAMPLE 2IMPROVING CATALYTIC ACTIVITY OF AN TCR $\alpha$  2.0 MEGATAL

Disruption of the TCR $\alpha$  locus prevents trafficking of CD3 to the cell surface. CD3 expression can be measured using fluorescently labeled anti-CD3 antibody and flow 5 cytometry. Gene editing efficiency is inversely proportional to CD3 fluorescence.

Primary human T cells were activated with CD3 and CD28, and electroporated with *in vitro* transcribed mRNA encoding either the original TCR $\alpha$  HE variant formatted as a megaTAL (e.g., SEQ ID NO: 13) or the TCR $\alpha$ 2.0 HE variant formatted as a megaTAL 10 (e.g., SEQ ID NO: 14). Despite displaying significantly enhanced selectivity at the p456 substrate positions, the TCR $\alpha$ 2.0 megaTAL displayed reduced catalytic activity, indicated by increased CD3 expression compared to the original TCR $\alpha$  megaTAL (Figure 4). Genomic DNA was subsequently isolated from megaTAL treated cells and the reduced catalytic activity was confirmed using PCR coupled with Tracking of Indels by DEcomposition (TIDE, *see* Brinkman *et al.*, 2014) at the target site.

15 A single amino acid substitution, E178D, that increased catalytic activity in another I-OnuI variant was inserted at the active site of the TCR $\alpha$ 2.0 megaTAL to generate the TCR $\alpha$ 2.1 megaTAL (SEQ ID NO: 8). *In vitro* transcribed mRNA encoding the TCR $\alpha$ 2.1 megaTAL (e.g., SEQ ID NO: 15) was electroporated into primary human T cells. The TCR $\alpha$ 2.1 megaTAL increased CD3 knockdown 10-fold compared to the TCR $\alpha$ 2.0 20 megaTAL. Figure 5.

EXAMPLE 3THE TCR $\alpha$ 2.1 I-ONUI VARIANT AND TCR $\alpha$ 2.1 MEGATAL SHOW  
ENHANCED ON- TO OFF- TARGET DISCRIMINATION

25 Therapeutic genome editing carries a degree of risk attributed to very low frequency off-target editing events. Genome editing enzymes can be characterized in depth for off-target editing events using existing methods (see Tsai *et al.* 2016).

Potential off-target sites were identified using bioinformatic algorithms that scan the human genome for 22-bp sequences resembling the TCR $\alpha$  target sequence (SEQ ID NO: 17). The top 764 22-bp sequences returned from the bioinformatics search were assayed individually for cleavage by the original TCR $\alpha$  HE variant and the TCR $\alpha$ 2.1 HE variant using yeast surface display. The TCR $\alpha$ 2.1 HE variant displayed increased selectivity for the TCR $\alpha$  target sequence compared to the original TCR $\alpha$  HE variant (Figure 6), consistent with results obtained in Example 1.

A second bioinformatics search was conducted using the sequence information of the top 50 cleaved substrates from the 764 substrate panel to further inform the search. 10 These were assayed individually for cleavage by the original TCR $\alpha$  and TCR $\alpha$ 2.1 HE variants. The refined TCR $\alpha$ 2.1 HE variant displayed an enhanced selectivity profile compared to the original TCR $\alpha$  HE variant.

A second assay employing integrase deficient lentiviral vectors (IDLV, *see* Gabriel 2011) was used to identify additional off-target sites generated *in vivo*. Briefly, IDLV is 15 captured at the site of double strand breaks within enzyme treated and mock treated cells. The frequency and location of IDLV integration can then be mapped and compared between the two populations. Increased IDLV uptake frequency at a particular genomic site in enzyme treated samples compared to mock treated samples indicates enzymatic cleavage. A primary off-target site of the original TCR $\alpha$  megaTAL was identified in intron 20 six of the KAT2B gene (SEQ ID NO: 21, Osborne et al. 2015) using this method.

The putative off-target sites identified were evaluated by deep sequencing of amplicons generated from megaTAL treated donor derived primary T cells (Figure 7). The TCR $\alpha$ 2.1 megaTAL eliminated off-target editing at the KAT2B gene locus and displayed enhanced selectivity for the on-target site, with reduced activity at the potential off-target 25 sites compared to the original TCR $\alpha$  megaTAL.

EXAMPLE 4EXTENSION OF THE TAL DNA BINDING DOMAIN INCREASES TCR $\alpha$  GENOME EDITING

MegaTALs include TALE DNA binding domain fusions to the meganuclease architecture to increase binding affinity, or addressing, toward the desired target sequence 5 (see Boissel *et al.* 2013). To further increase on-target editing efficiency at the TCR $\alpha$  locus an additional TALE RVD was appended to the C-terminal end the TALE DNA binding domain to increase the total length of the TALE DNA binding domain array from 10.5 to 11.5 RVD repeats and generate the 11.5 RVD TCR $\alpha$  2.1 megaTAL (e.g., SEQ ID NO: 12). The 11.5 RVD TALE DNA binding domain recognizes the original TALE binding site 10 (SEQ ID NO: 18) with an additional nucleotide on the 3' end (SEQ ID NO: 19).

Human primary T cells were electroporated with *in vitro* transcribed mRNA encoding the 10.5 RVD and 11.5 RVD TCR $\alpha$ 2.1 megaTALs. The 11.5 RVD TCR $\alpha$ 2.1 megaTAL increased TCR $\alpha$  editing 30-50%, as measured by CD3 knockdown using flow cytometry, compared to TCR $\alpha$  editing with the 10.5 RVD TCR $\alpha$ 2.1 megaTAL (Figure 8).

15

EXAMPLE 5

## VARIANT N-TERMINAL TALE ARCHITECTURES

MegaTALs are hybrid nucleases that include homing endonuclease fused to a TAL effector (TALE) architecture comprising an N-terminal domain, a DNA binding domain, and a C-terminal domain. Some TALE architectures comprise a minimal N-terminal 20 region ('delta-154') of approximately 136 amino acids, a variable number of 34 amino acid RVD-containing repeats (terminating with a truncated 20 amino acid repeat), and a minimal C-terminal region of approximately 63 amino acids.

The TALE architecture fused to the homing endonuclease described in the previous examples includes sequences derived from the *Xanthomonas oryzae* TALE proteins. In 25 addition to the *X. oryzae* TALE domain sequences, minimal N-terminal TALE domain (NTD) regions can be derived from other members of *Xanthomonas* family such as *Xanthomonas translucens* and *Xanthomonas citri*. Multiple sequence alignments of N-

terminal TALE domains showed that the *X. Oryzae* shares 85% and 87% sequence similarity with *X. translucens* and *X. citri*, respectively. To assess if TAL activity could be further enhanced using alternative species NTDs, hybrid NTDs were generated by rationally substituting and combining the following amino acid positions in endogenous *X. translucens* and *X. citri* sequences: V152G, D153K, D165E, E166K, I167L, P169L, 5 A173P, T174I, Q180E, M183I, H191Q, V205I, V207A, I224V, V227I, A236V, G249S, G257S, V271I (e.g., SEQ ID NO:21); V152G, D153K, T170K, V171A, V183I, A200P, A236T, A239V, A244V, E245Q, G257S, L260V, V271I, (e.g., SEQ ID NO:24); T170K, A200P, Q230H, W231C, S232G, A236T, A239V, A244V, E245Q, G257S (e.g., SEQ ID 10 NO: 23). The hybrid sequences (e.g., SEQ ID NOs: 22, 23, 24) showed 80% similarity to *X. Oryzae* N-terminal TAL domain. Each of these minimal TALE domains was individually grafted onto the N-terminus of the TALE DNA binding domain of the original TCR $\alpha$  I-OnuI variant megaTAL. *In vitro* transcribed mRNAs encoding the variant megaTALs comprising TALE N-terminal domain (NTD) variants or *X. oryzae* NTDs were 15 electroporated into human primary T cells to assess editing rates at the TCR $\alpha$  locus. Flow cytometric analysis of CD3 expression demonstrated that the TCR $\alpha$  megaTALs comprising NTD variant domains showed similar activity, and in some cases slightly higher activity, relative to the original TCR $\alpha$  megaTAL comprising *X. oryzae* NTD (Figure 9).

20 In general, in the following claims, the terms used should not be construed to limit the claims to the specific embodiments disclosed in the specification and the claims, but should be construed to include all possible embodiments along with the full scope of equivalents to which such claims are entitled. Accordingly, the claims are not limited by the disclosure.

## CLAIMS

What is claimed is:

1. A polypeptide comprising an I-OnuI homing endonuclease (HE) variant that cleaves a target site in the human T cell receptor alpha (TCR $\alpha$ ) gene, wherein the variant comprises the amino acid substitutions: L26I, R28D, N32R, K34N, S35E, V37N, G38R, S40R, E42S, G44R, V68K, A70T, G73S, N75R, S78M, K80R, L138M, T143N, S159P, S176A, C180H, F182G, I186K, S188V, S190G, K191T, L192A, G193K, Q195Y, Q197G, V199R, S201A, T203S, K207R, Y223S, K225R, S233R, D236E, and V238E of any one of SEQ ID NOs: 1-5, or a biologically active fragment thereof.
2. A polypeptide comprising an I-OnuI homing endonuclease (HE) variant that cleaves a target site in the human T cell receptor alpha (TCR $\alpha$ ) gene, wherein the variant comprises the amino acid substitutions: L26I, R28D, N32R, K34N, S35E, V37N, G38R, S40R, E42S, G44R, V68K, A70T, G73S, N75R, S78M, K80R, L138M, T143N, S159P, S176A, E178D, C180H, F182G, I186K, S188V, S190G, K191T, L192A, G193K, Q195Y, Q197G, V199R, S201A, T203S, K207R, Y223S, K225R, S233R, D236E, and V238E of any one of SEQ ID NOs: 1-5, or a biologically active fragment thereof.
3. The polypeptide of claim 1 or claim 2, wherein the biologically active fragment lacks the 1, 2, 3, 4, 5, 6, 7, or 8 N-terminal amino acids compared to a corresponding wild type I-OnuI HE.
4. The polypeptide of claim 3, wherein the biologically active fragment lacks the 4 N-terminal amino acids compared to a corresponding wild type I-OnuI HE.
5. The polypeptide of claim 3, wherein the biologically active fragment lacks the 8 N-terminal amino acids compared to a corresponding wild type I-OnuI HE.

6. The polypeptide of claim 1 or claim 2, wherein the biologically active fragment lacks the 1, 2, 3, 4, or 5 C-terminal amino acids compared to a corresponding wild type I-OnuI HE.

7. The polypeptide of claim 6, wherein the biologically active fragment lacks the C-terminal amino acid compared to a corresponding wild type I-OnuI HE.

8. The polypeptide of claim 6, wherein the biologically active fragment lacks the 2 C-terminal amino acids compared to a corresponding wild type I-OnuI HE.

9. The polypeptide of any one of claims 1 to 8, wherein the I-OnuI HE variant comprises an amino acid sequence that is at least 99% identical to the amino acid sequence set forth in any one of SEQ ID NOs: 7-8, or a biologically active fragment thereof.

10. The polypeptide of any one of claims 1 to 9, wherein the HE variant comprises the amino acid sequence set forth in SEQ ID NO: 7, or a biologically active fragment thereof.

11. The polypeptide of any one of claims 1 to 9, wherein the HE variant comprises the amino acid sequence set forth in SEQ ID NO: 8, or a biologically active fragment thereof.

12. The polypeptide of any one of claims 1 to 11, wherein the polypeptide binds the polynucleotide sequence set forth in SEQ ID NO: 17.

13. The polypeptide of any one of claims 1 to 12, further comprising a DNA binding domain.

14. The polypeptide of claim 13, wherein the DNA binding domain is selected from the group consisting of: a TALE DNA binding domain and a zinc finger DNA binding domain.

15. The polypeptide of claim 14, wherein the TALE DNA binding domain comprises about 9.5 TALE repeat units, about 10.5 TALE repeat units, about 11.5 TALE repeat units, about 12.5 TALE repeat units, about 13.5 TALE repeat units, about 14.5 TALE repeat units, or about 15.5 TALE repeat units.

16. The polypeptide of claim 14 or claim 15, wherein the TALE DNA binding domain comprises 11.5 TALE repeat units and binds the polynucleotide sequence set forth in SEQ ID NO: 19.

17. The polypeptide of claim 14 or claim 15, wherein the TALE DNA binding domain comprises 10.5 TALE repeat units and binds the polynucleotide sequence set forth in SEQ ID NO: 18.

18. The polypeptide of any one of claims 14 to 17, wherein the polypeptide binds and cleaves the polynucleotide sequence set forth in SEQ ID NO: 20.

19. The polypeptide of claim 14, wherein the zinc finger DNA binding domain comprises 2, 3, 4, 5, 6, 7, or 8 zinc finger motifs.

20. The polypeptide of any one of claims 1 to 19, further comprising a peptide linker and an end-processing enzyme or biologically active fragment thereof.

21. The polypeptide of any one of claims 1 to 20, further comprising a viral self-cleaving 2A peptide and an end-processing enzyme or biologically active fragment thereof.

22. The polypeptide of claim 20 or claim 21, wherein the end-processing enzyme or biologically active fragment thereof has 5'-3' exonuclease, 5'-3' alkaline exonuclease, 3'-5' exonuclease, 5' flap endonuclease, helicase or template-independent DNA polymerase activity.

23. The polypeptide of any one of claims 20 to 22, wherein the end-processing enzyme comprises Trex2 or a biologically active fragment thereof.

24. The polypeptide of any one of claims 1 to 23, wherein the polypeptide comprises the amino acid sequence set forth in any one of SEQ ID NOs: 10 to 12, or a biologically active fragment thereof.

25. The polypeptide of claim 24, wherein the polypeptide comprises the amino acid sequence set forth in SEQ ID NO: 10, or a biologically active fragment thereof.

26. The polypeptide of claim 24, wherein the polypeptide comprises the amino acid sequence set forth in SEQ ID NO: 11, or a biologically active fragment thereof.

27. The polypeptide of claim 24, wherein the polypeptide comprises the amino acid sequence set forth in SEQ ID NO: 12, or a biologically active fragment thereof.

28. A polynucleotide encoding the polypeptide of any one of claims 1 to 27.

29. An mRNA encoding the polypeptide of any one of claims 1 to 27.

30. A cDNA encoding the polypeptide of any one of claims 1 to 27.

31. A vector comprising a polynucleotide encoding the polypeptide of any one of claims 1 to 27.

32. A cell comprising the polypeptide of any one of claims 1 to 27.

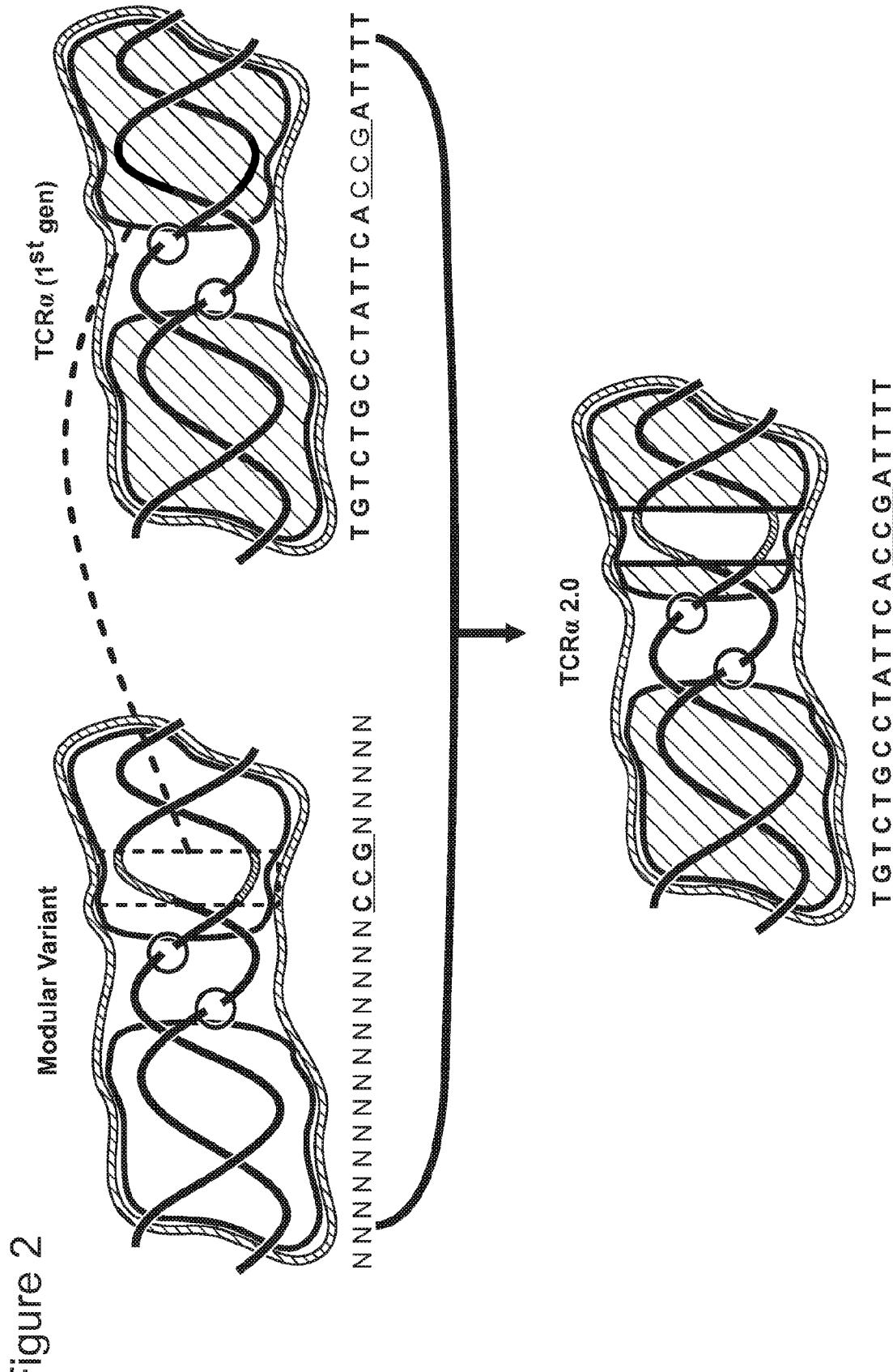
33. A cell comprising a polynucleotide encoding the polypeptide of any one of claims 1 to 27.

34. A cell comprising the vector of claim 27.
35. A cell edited by the polypeptide of any one of claims 1-27.
36. The cell of any one of claims 32 to 35, wherein the cell is a hematopoietic cell.
37. The cell of any one of claims 32 to 36, wherein the cell is a T cell.
38. The cell of any one of claims 32 to 37, wherein the cell is a CD3<sup>+</sup>, CD4<sup>+</sup>, and/or CD8<sup>+</sup> cell.
39. The cell of any one of claims 32 to 38, wherein the cell is an immune effector cell.
40. The cell of any one of claims 32 to 39, wherein the cell is a cytotoxic T lymphocytes (CTLs), a tumor infiltrating lymphocytes (TILs), or a helper T cells.
41. The cell of any one of claims 32 to 39, wherein the cell is a natural killer (NK) cell or natural killer T (NKT) cell.
42. The cell of any one of claims 32 to 41, wherein the source of the cell is peripheral blood mononuclear cells, bone marrow, lymph nodes tissue, cord blood, thymus issue, tissue from a site of infection, ascites, pleural effusion, spleen tissue, or tumors.

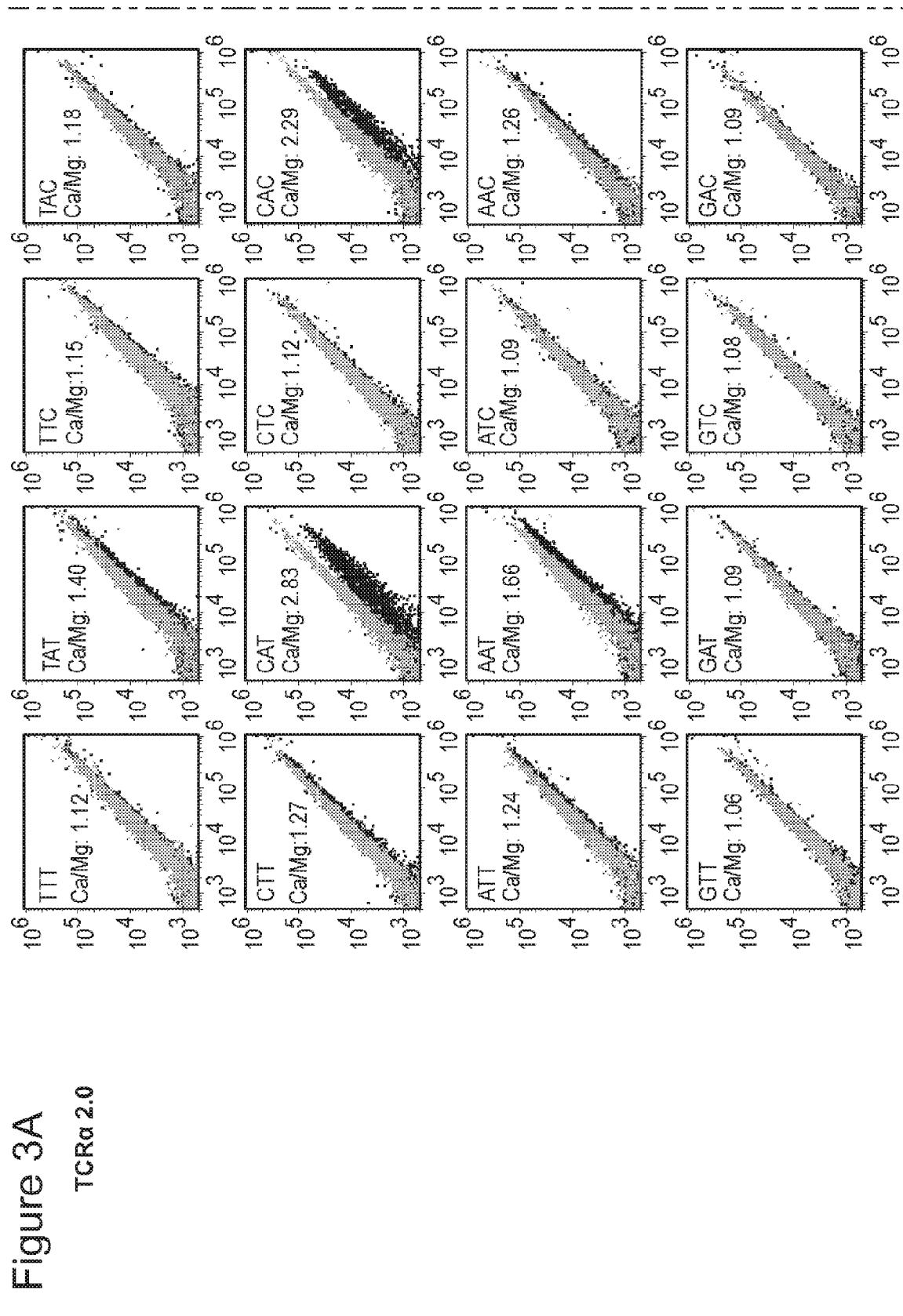
Figure 1

	n <sub>11</sub>	n <sub>10</sub>	n <sub>9</sub>	n <sub>8</sub>	n <sub>7</sub>	n <sub>6</sub>	n <sub>5</sub>	n <sub>4</sub>	n <sub>3</sub>	n <sub>2</sub>	n <sub>1</sub>	p <sub>1</sub>	p <sub>2</sub>	p <sub>3</sub>	p <sub>4</sub>	p <sub>5</sub>	p <sub>6</sub>	p <sub>7</sub>	p <sub>8</sub>	p <sub>9</sub>	p <sub>10</sub>	p <sub>11</sub>
TCRa	T	G	T	C	T	G	C	C	T	A	T	T	C	A	C	C	G	A	T	T	T	T
KAT2B	A	G	T	A	T	G	C	C	T	A	T	T	C	T	A	C	A	A	G	A	A	A

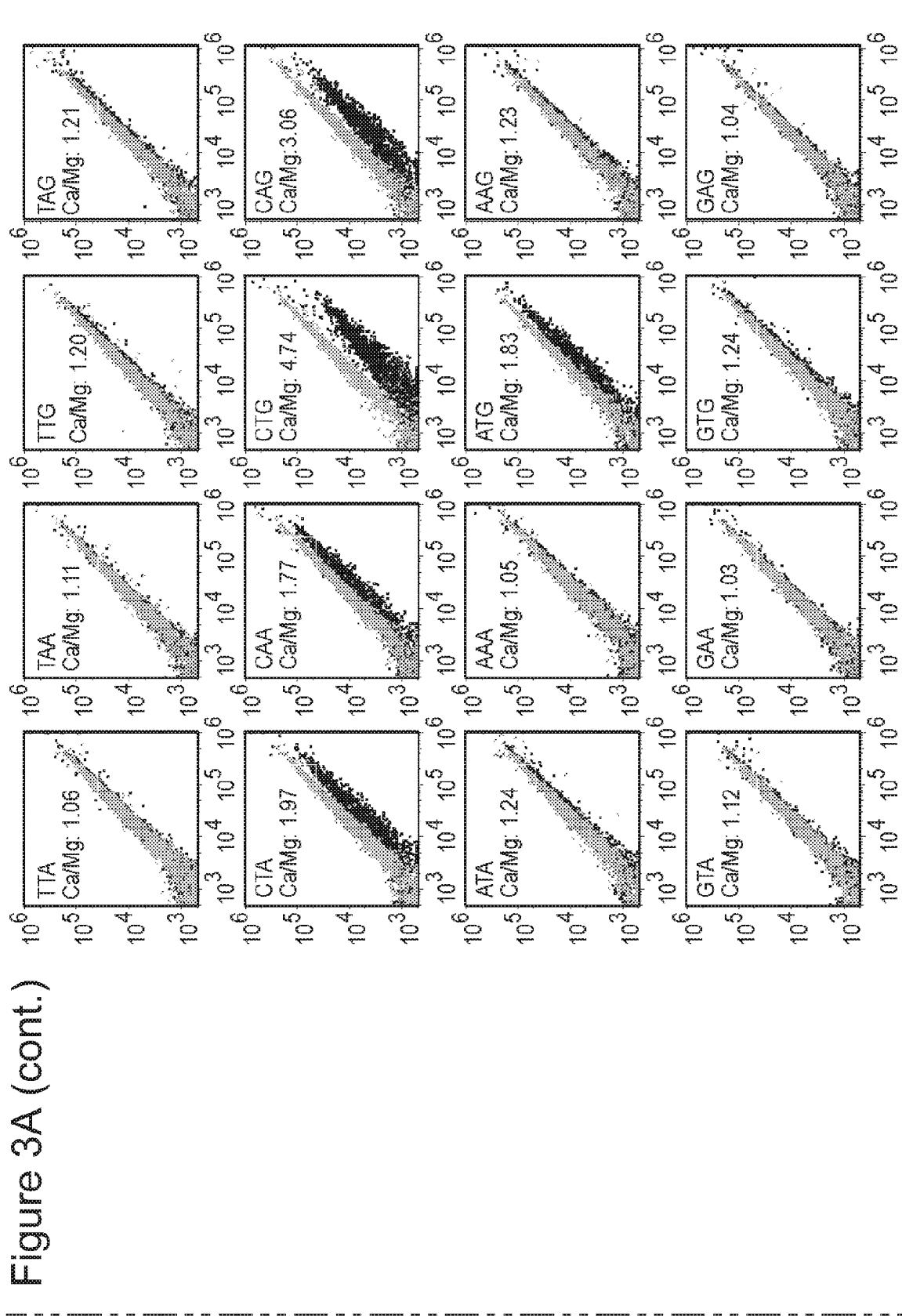
2/23



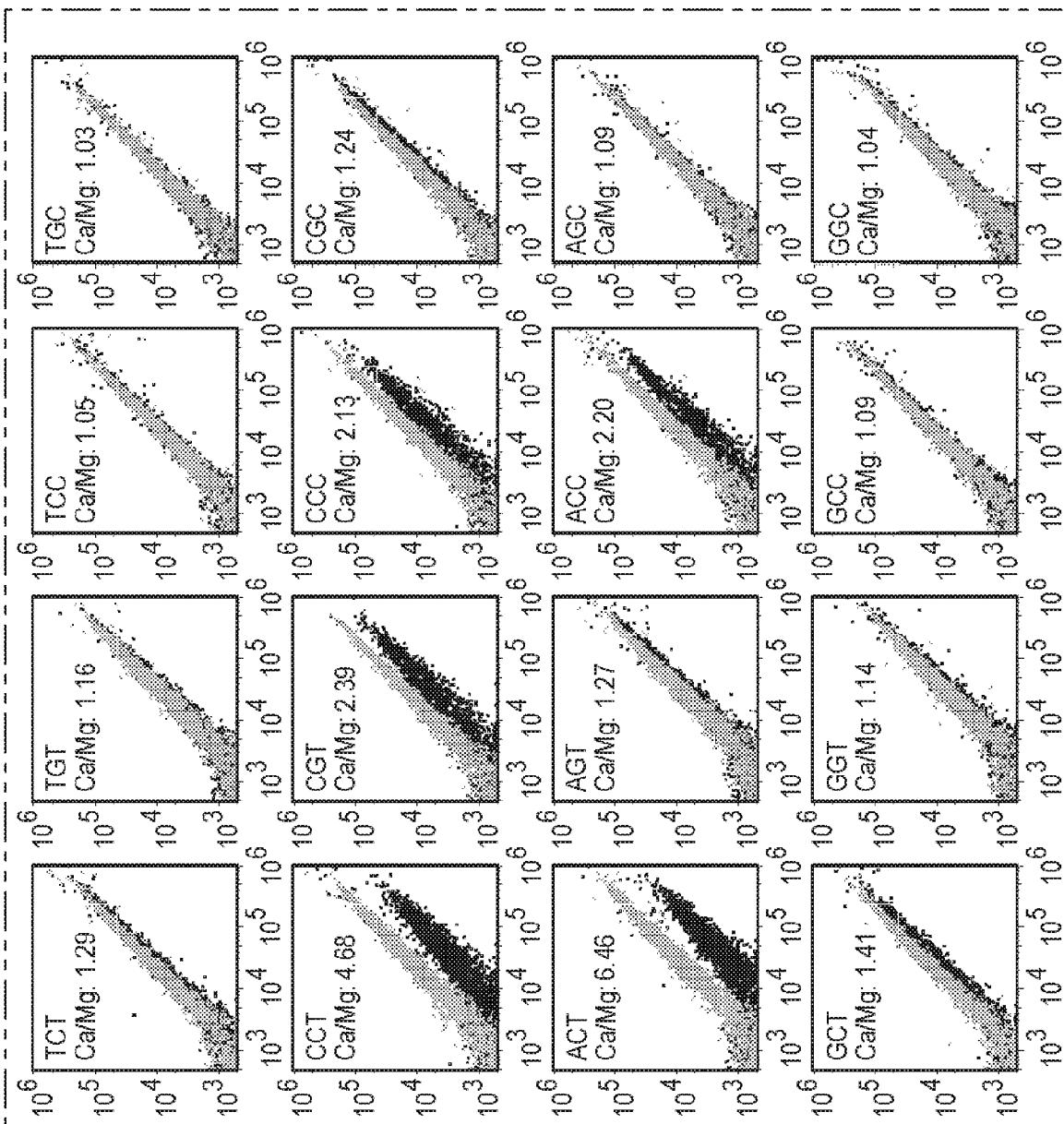
3/23



4/23

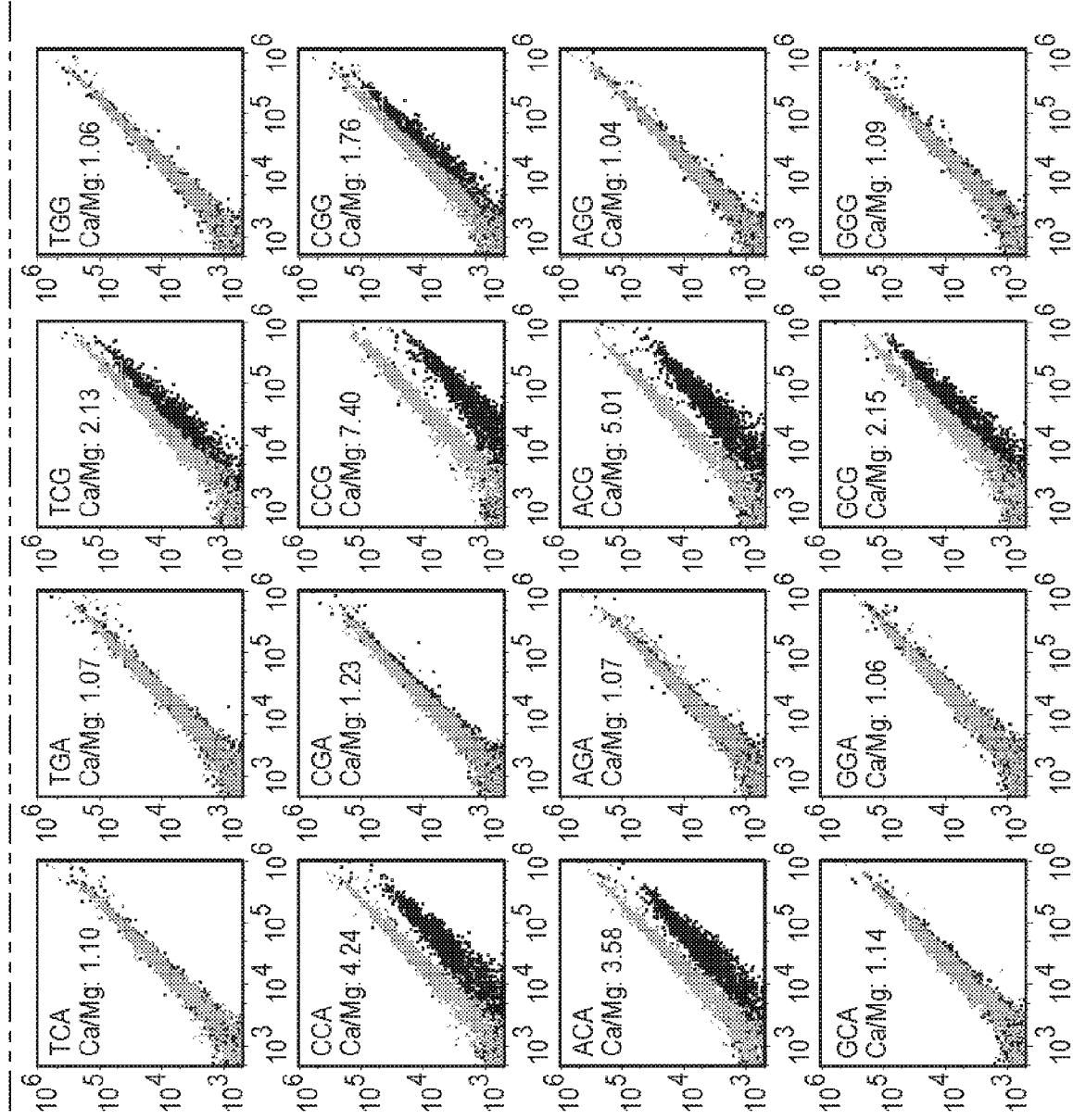


5/23

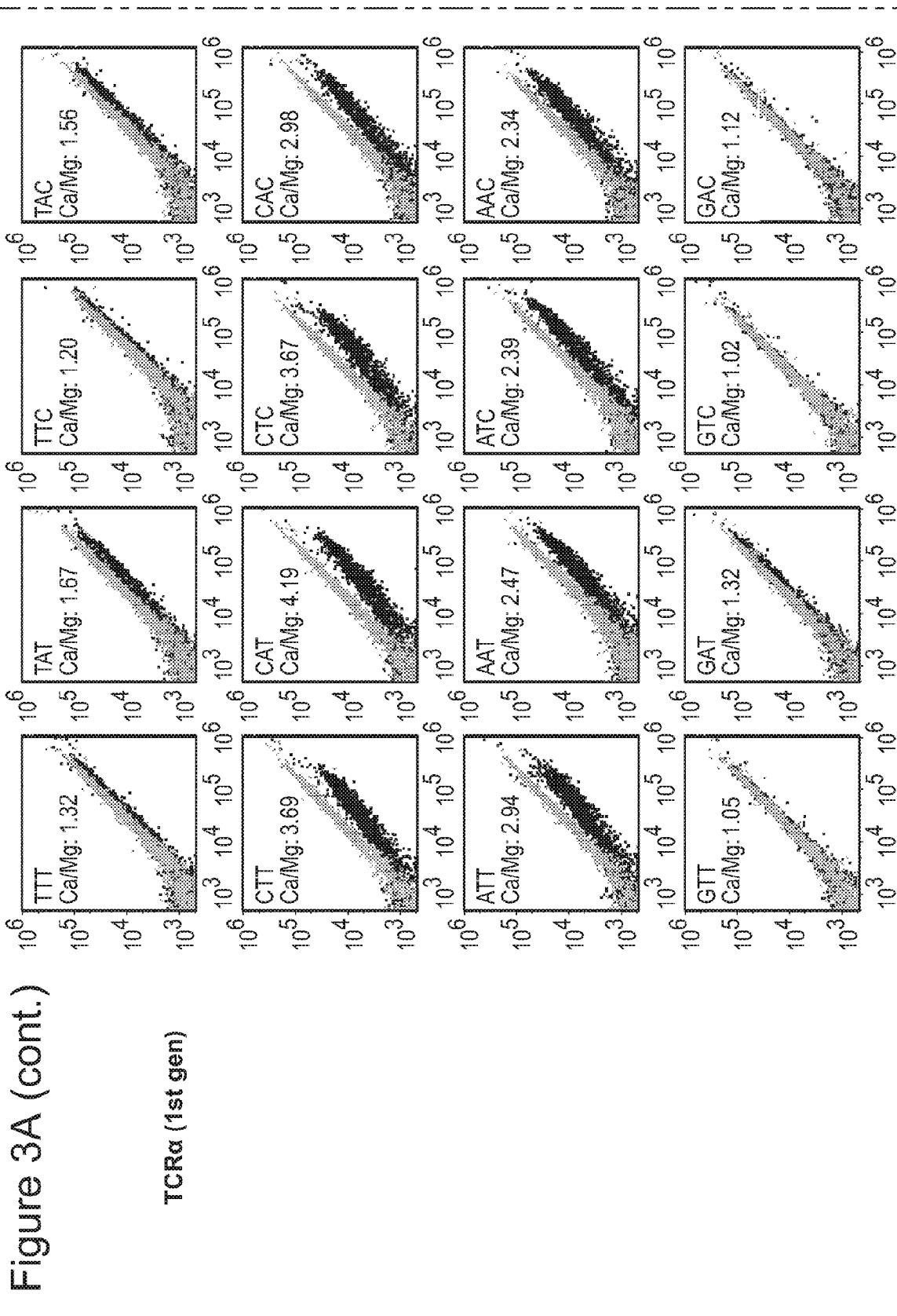


6/23

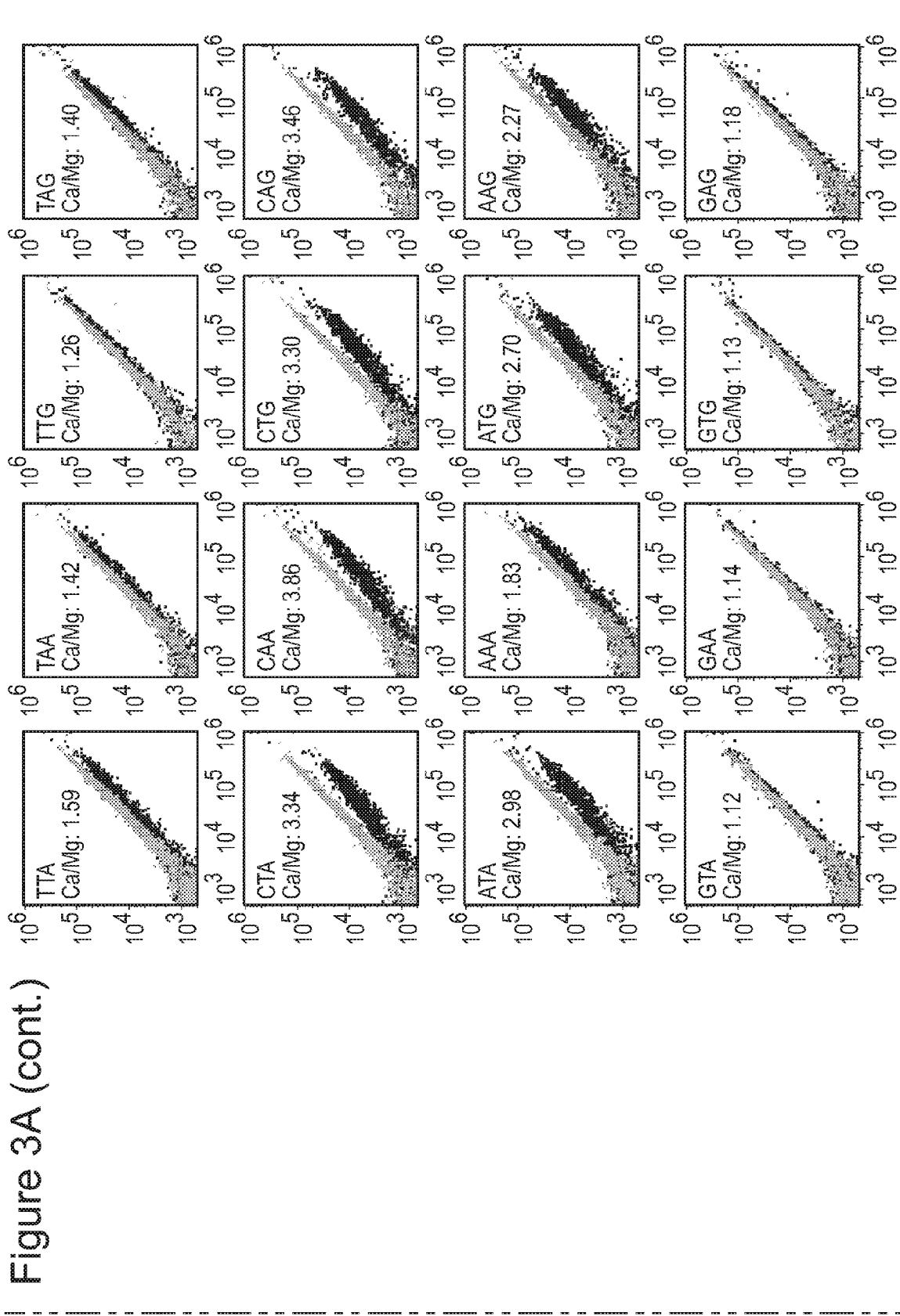
Figure 3A (cont.)



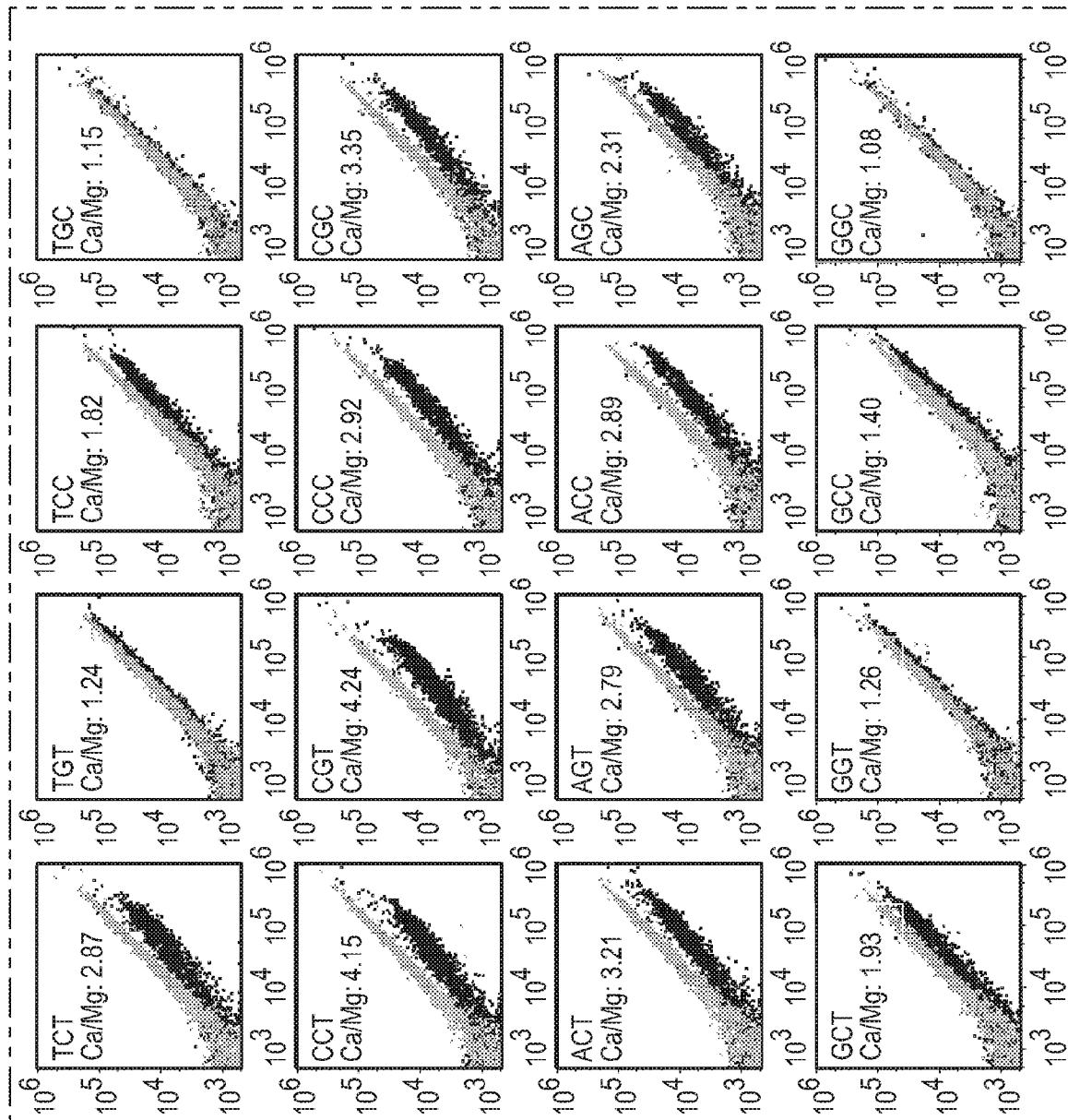
7/23



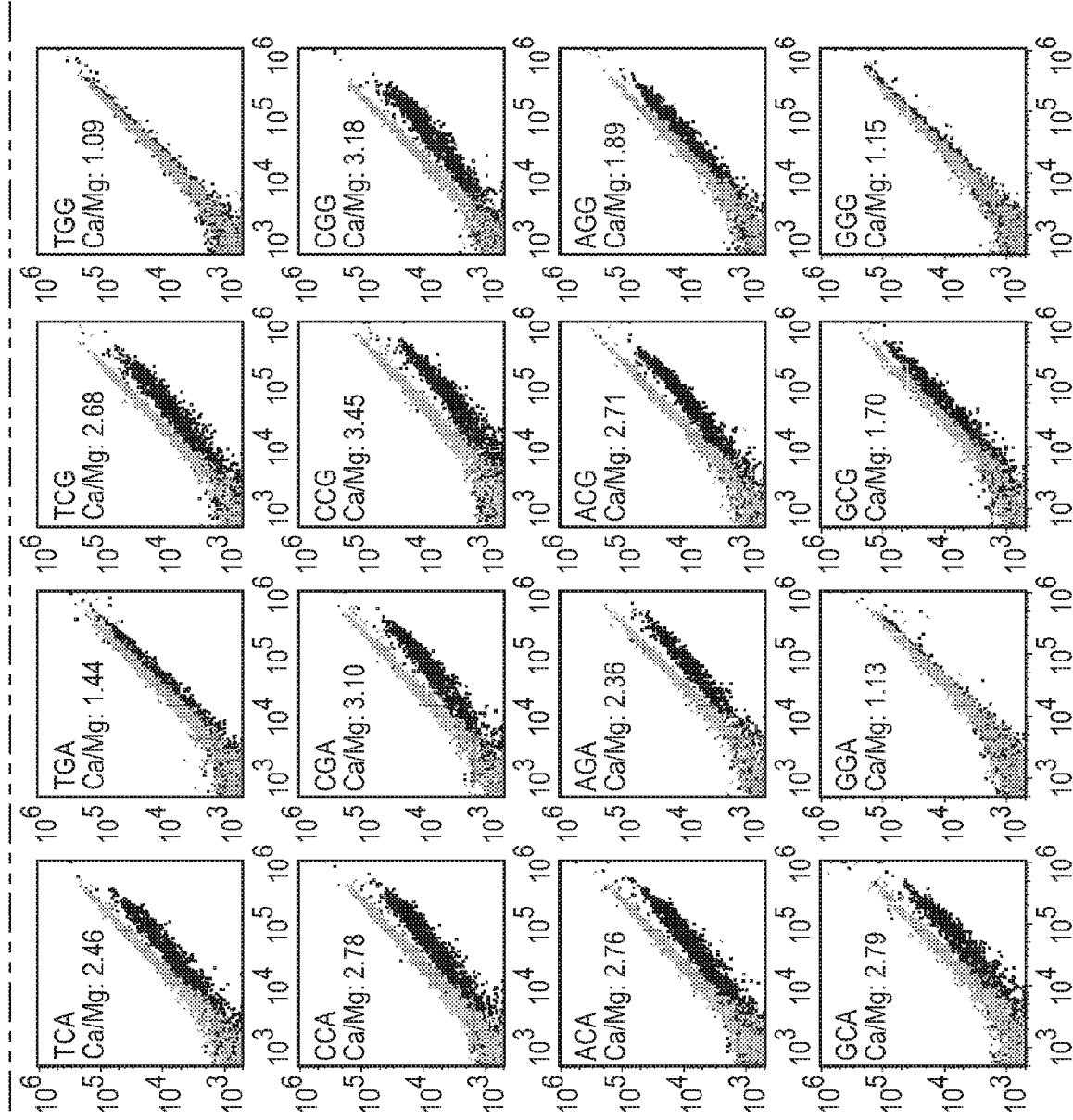
8/23



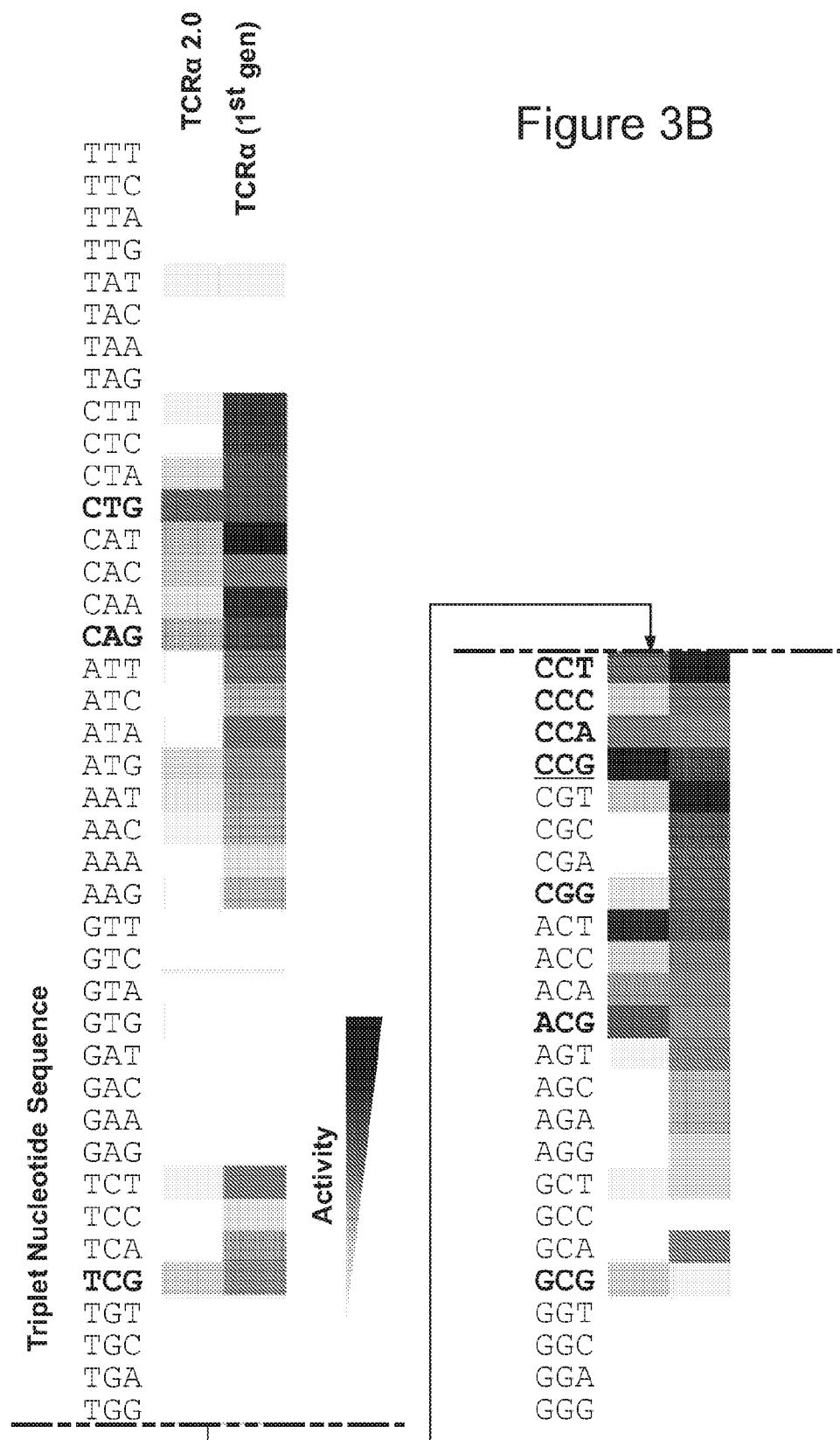
9/23

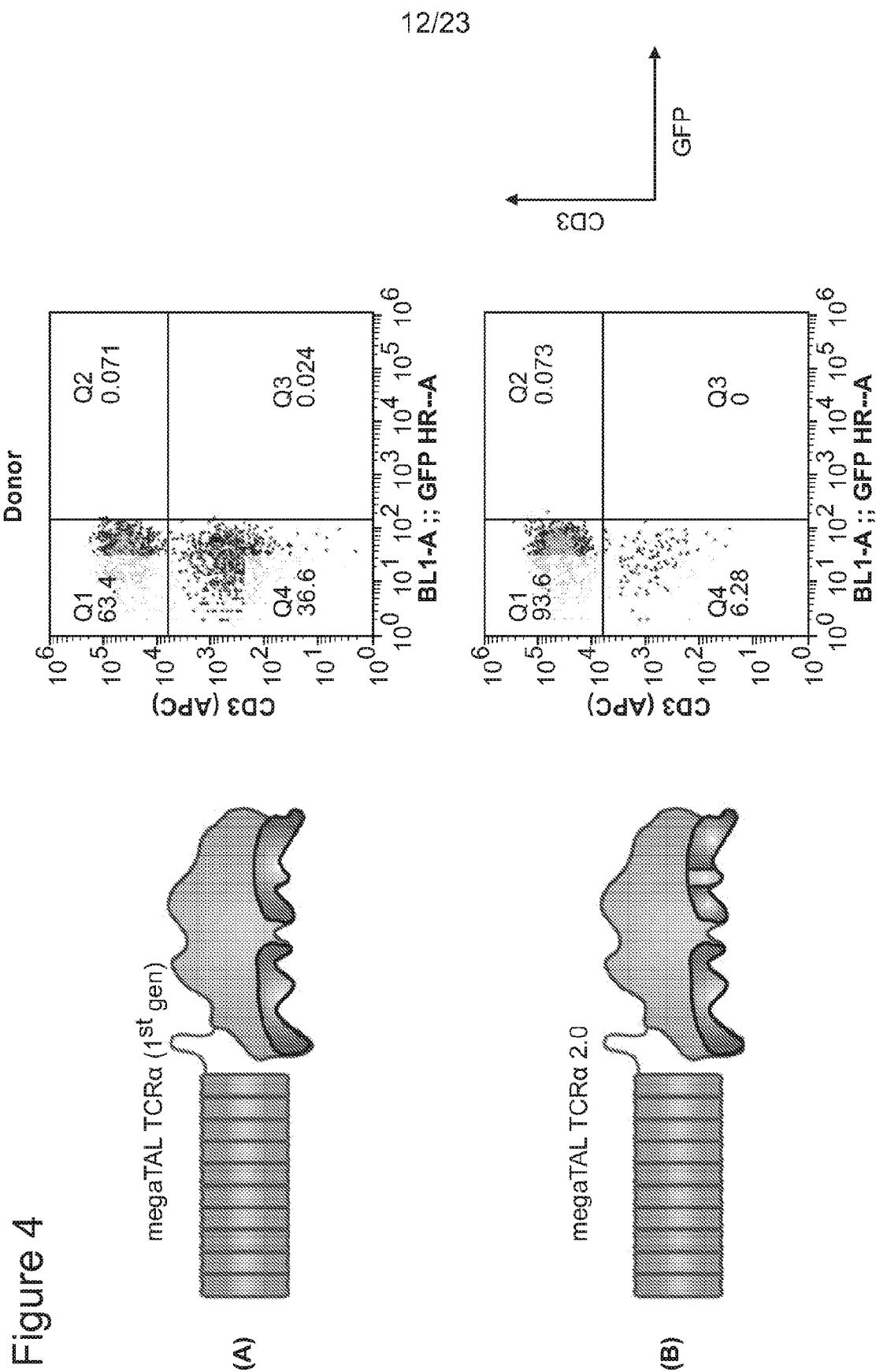
**Figure 3A (cont.)**

10/23

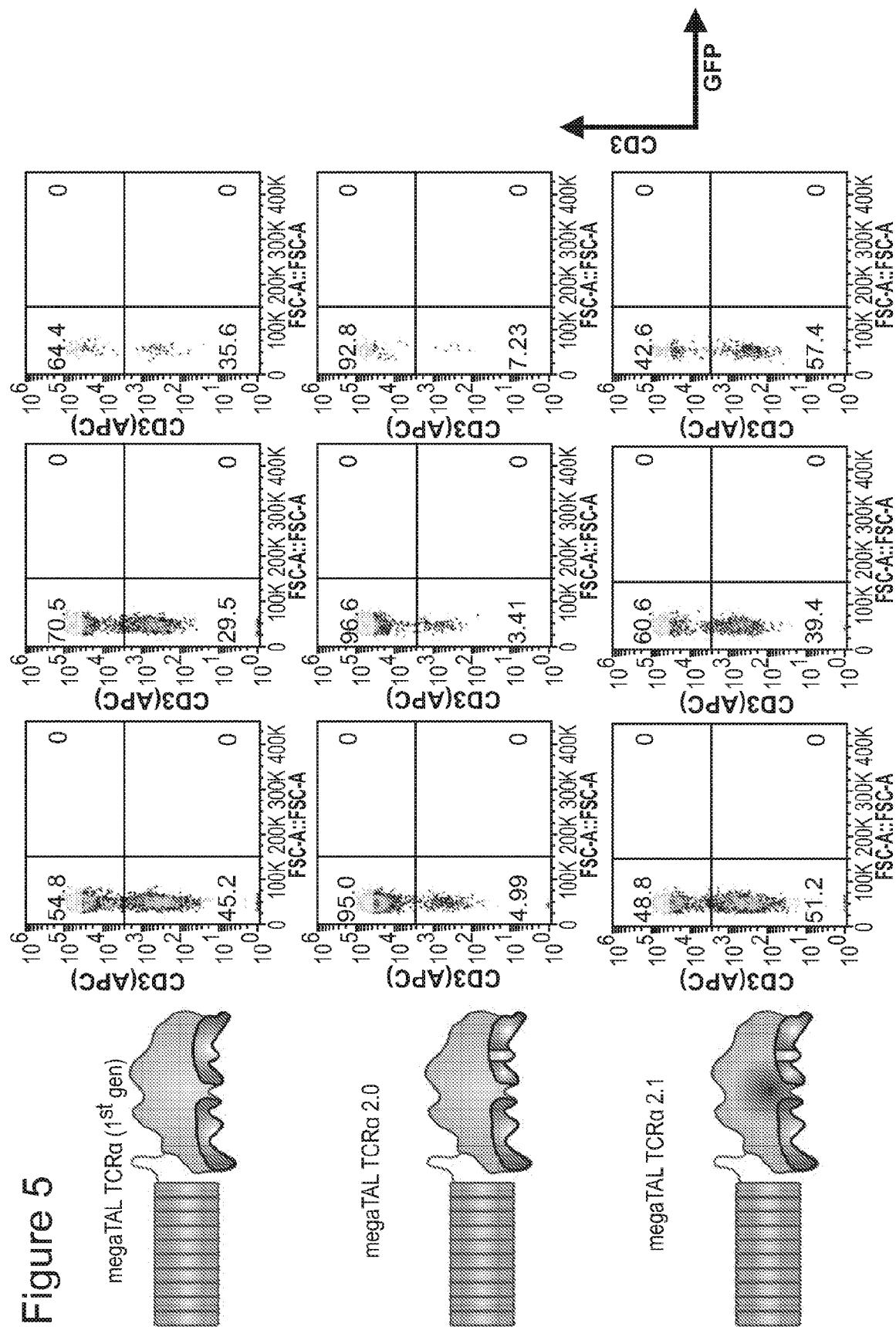
**Figure 3A (cont.)**

11/23





13/23



## Figure 5

14/23

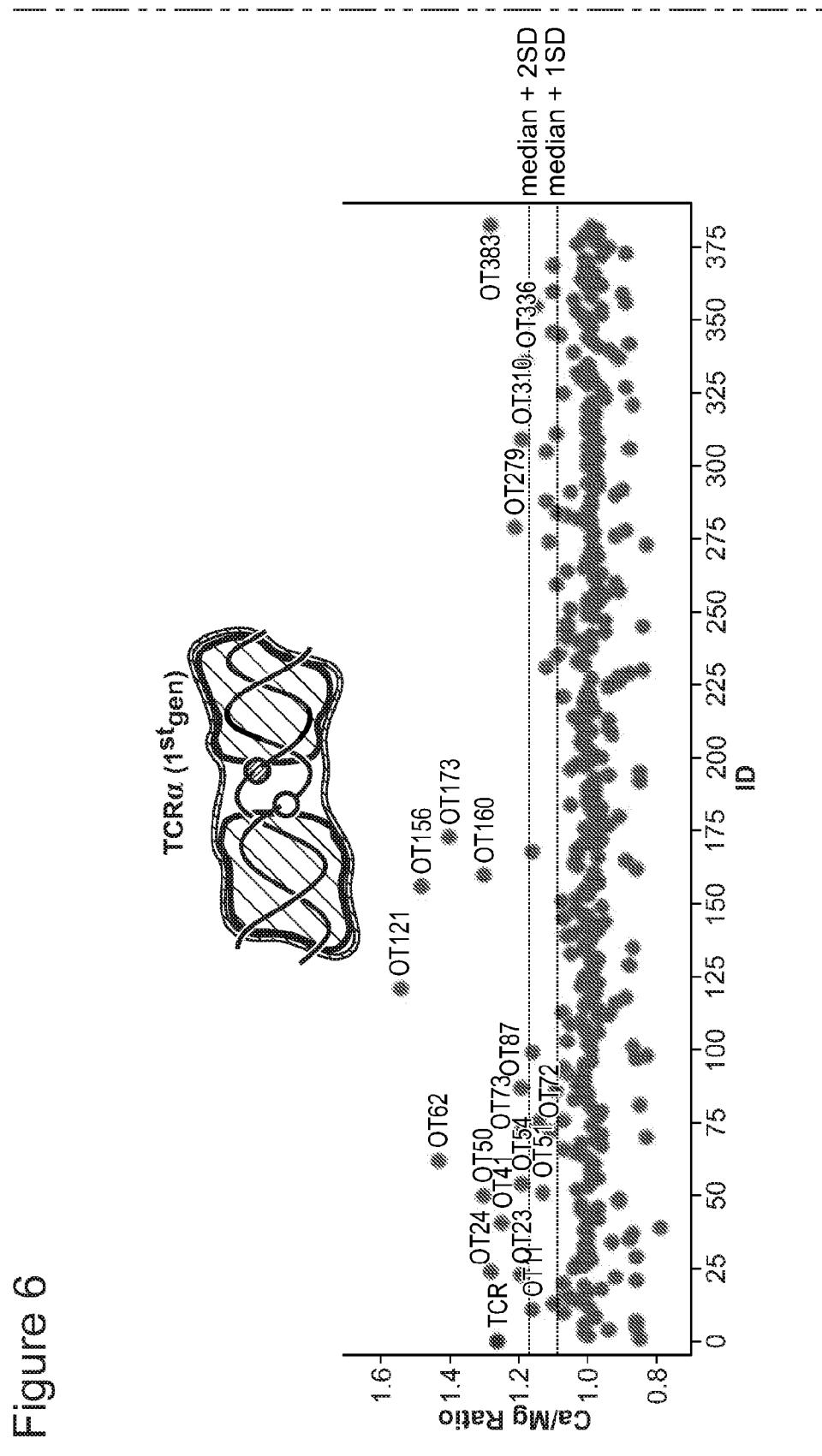
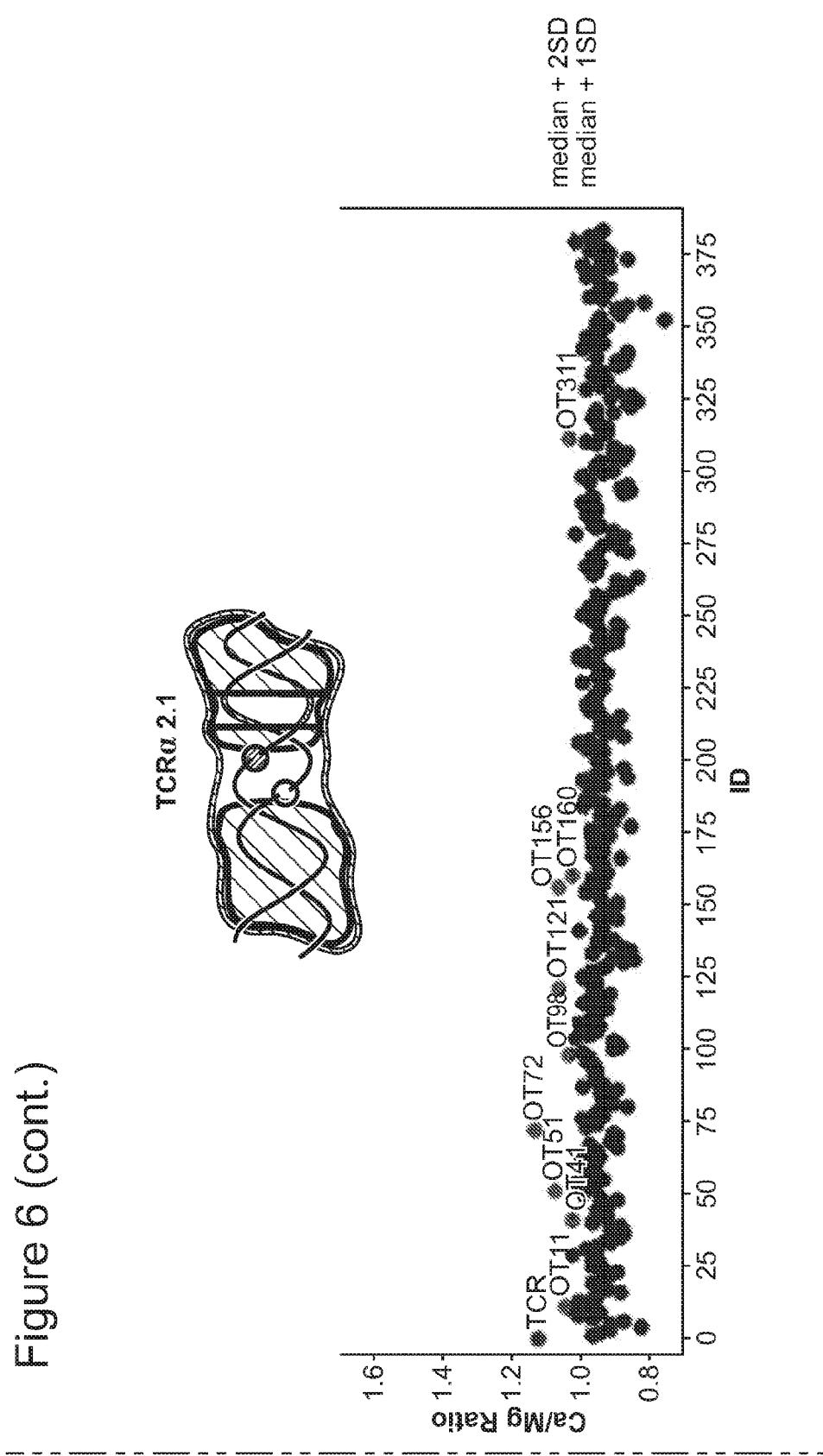


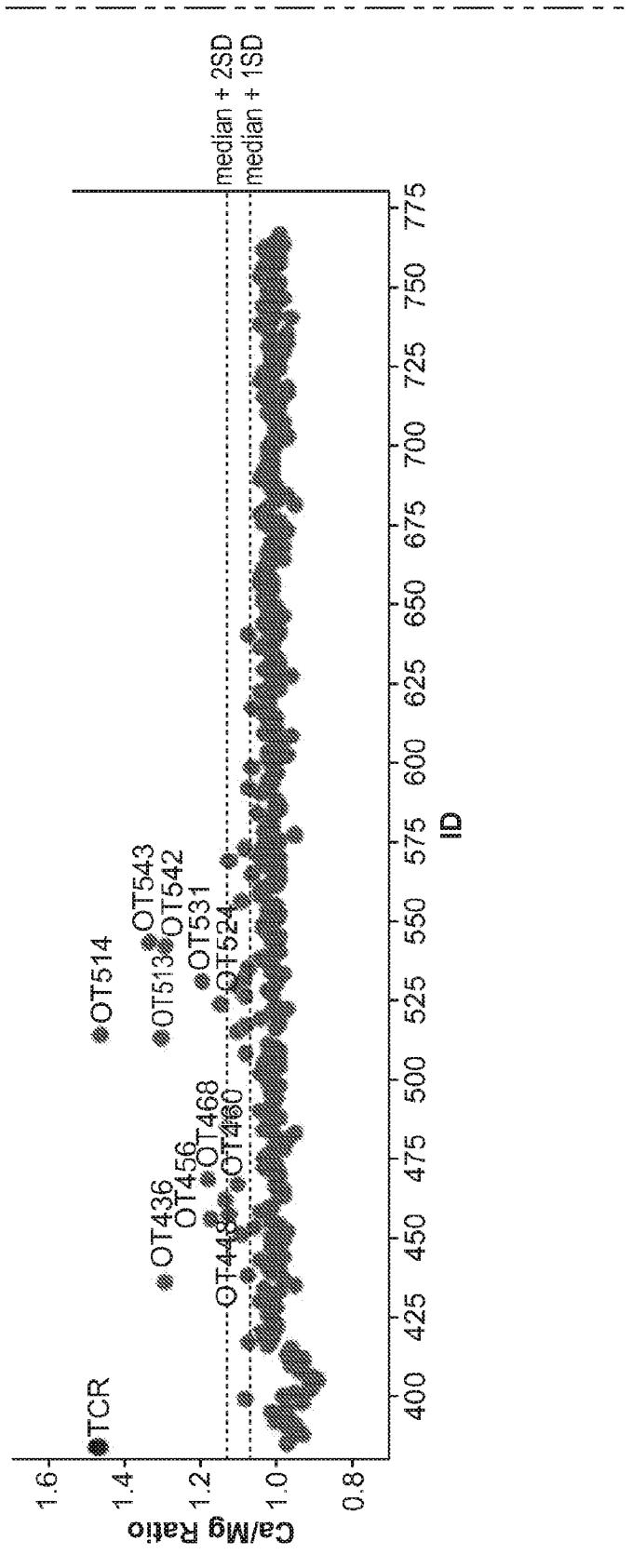
Figure 6

15/23



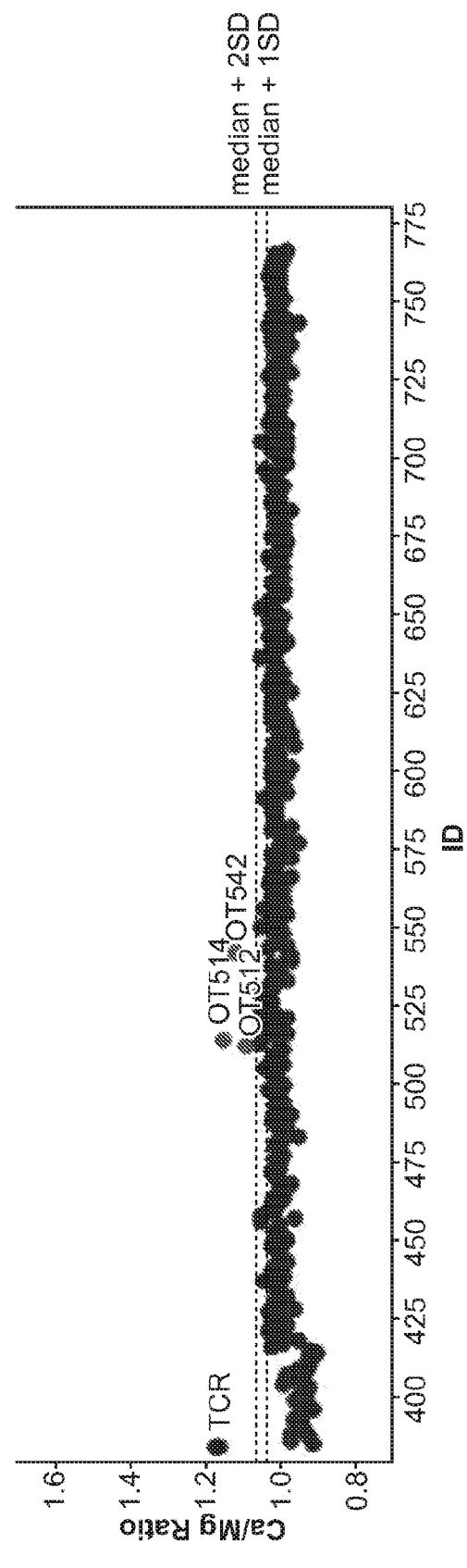
16/23

Figure 6 (cont.)



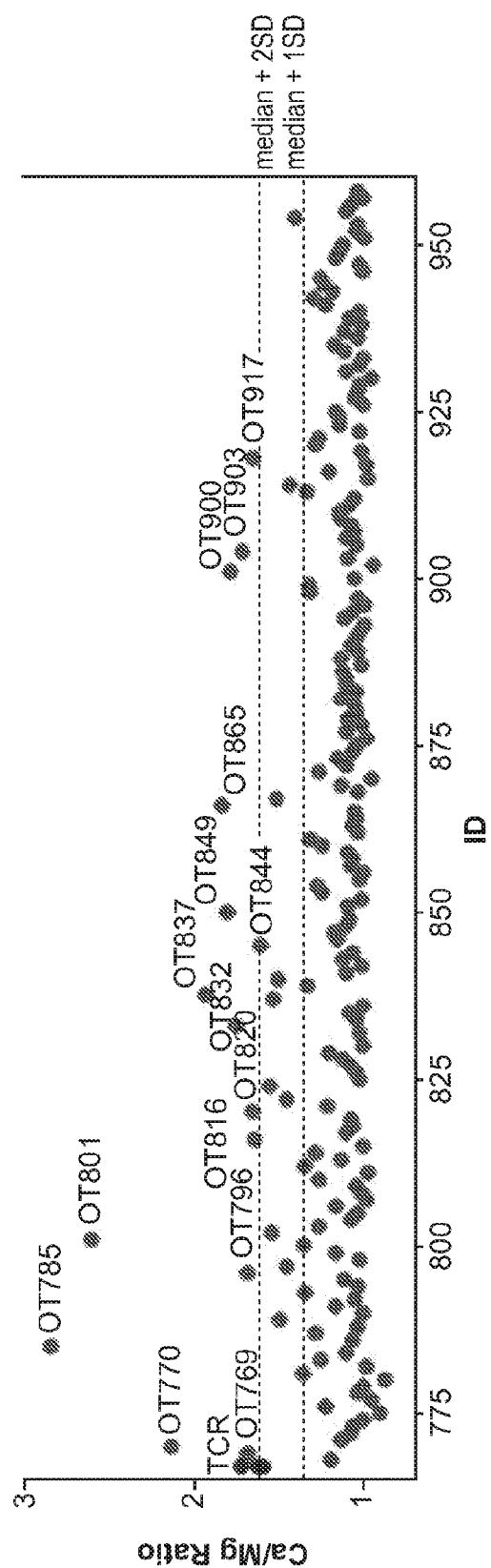
17/23

Figure 6 (cont.)



18/23

Figure 6 (cont.)



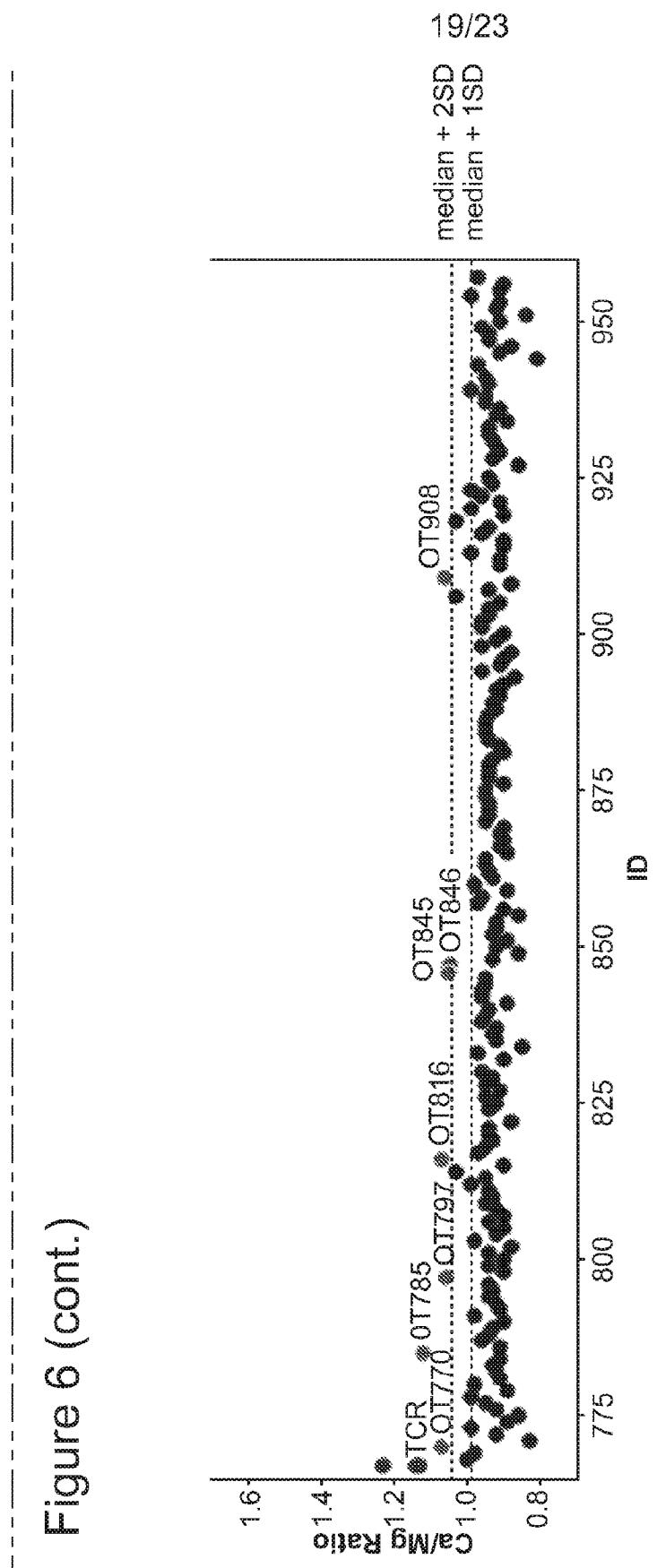


Figure 6 (cont.)

Figure 7

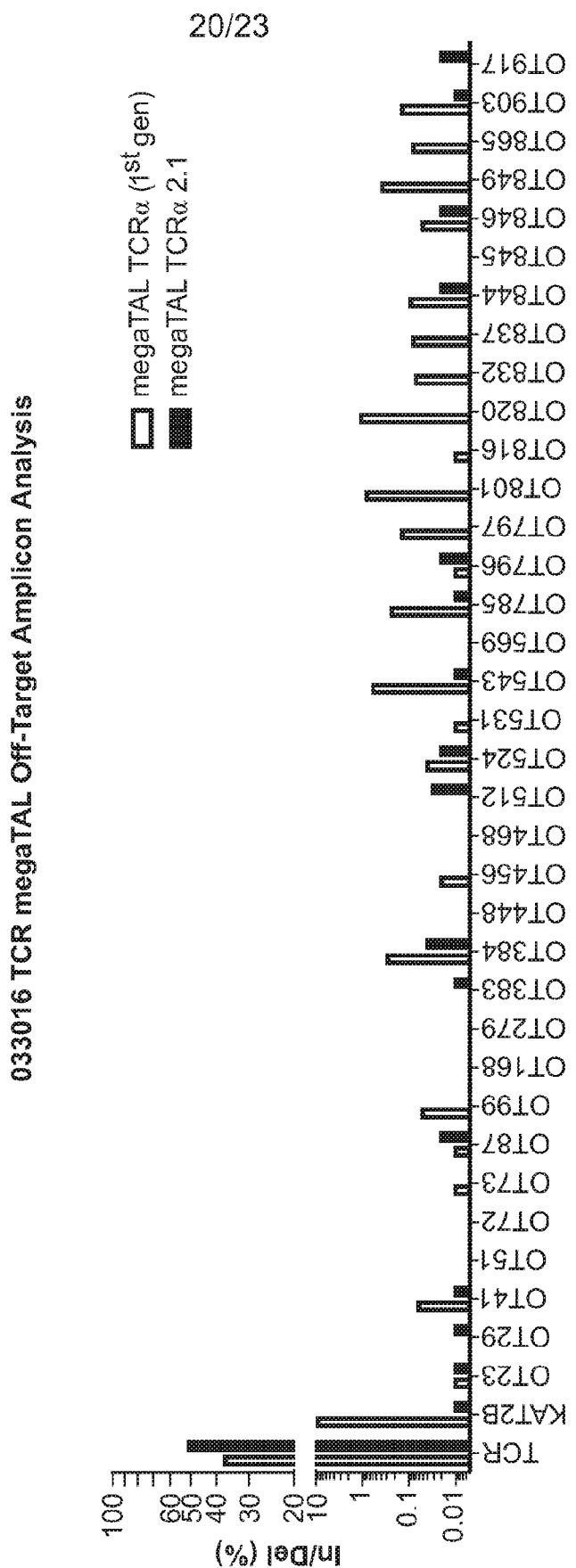
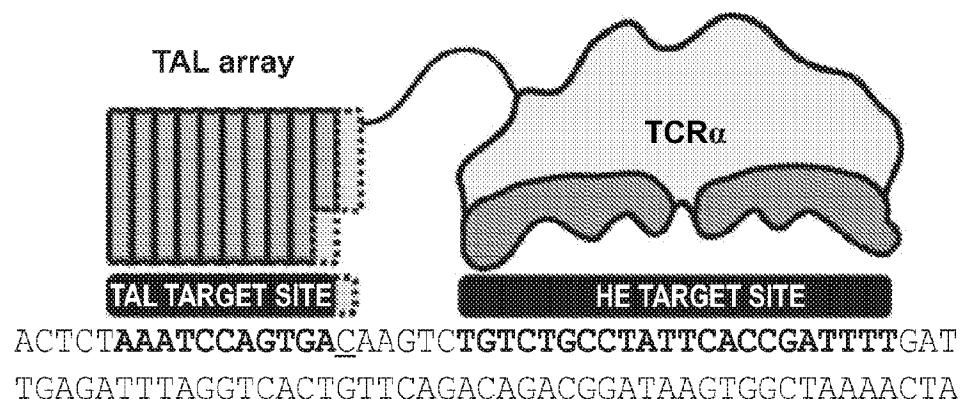
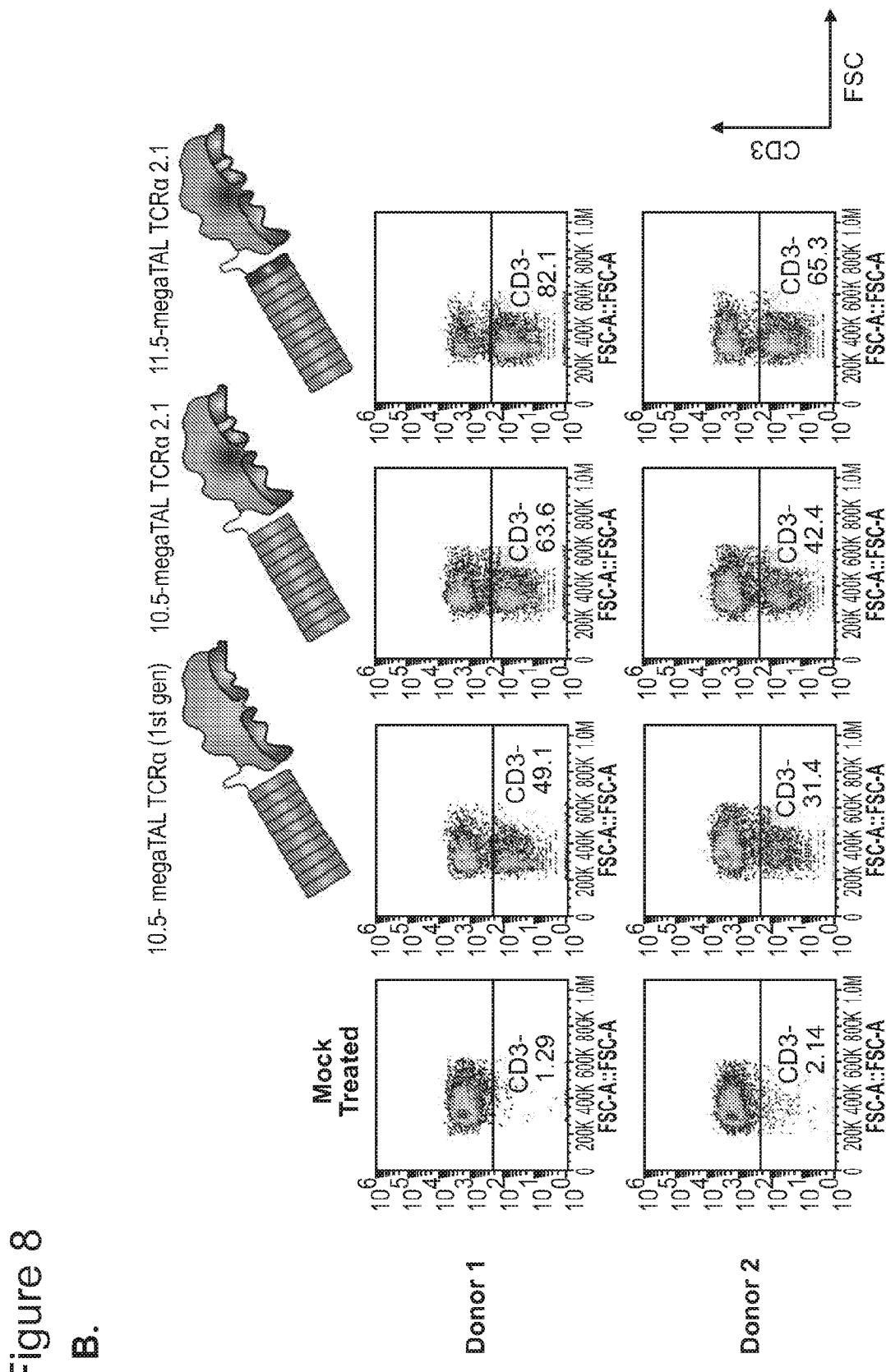


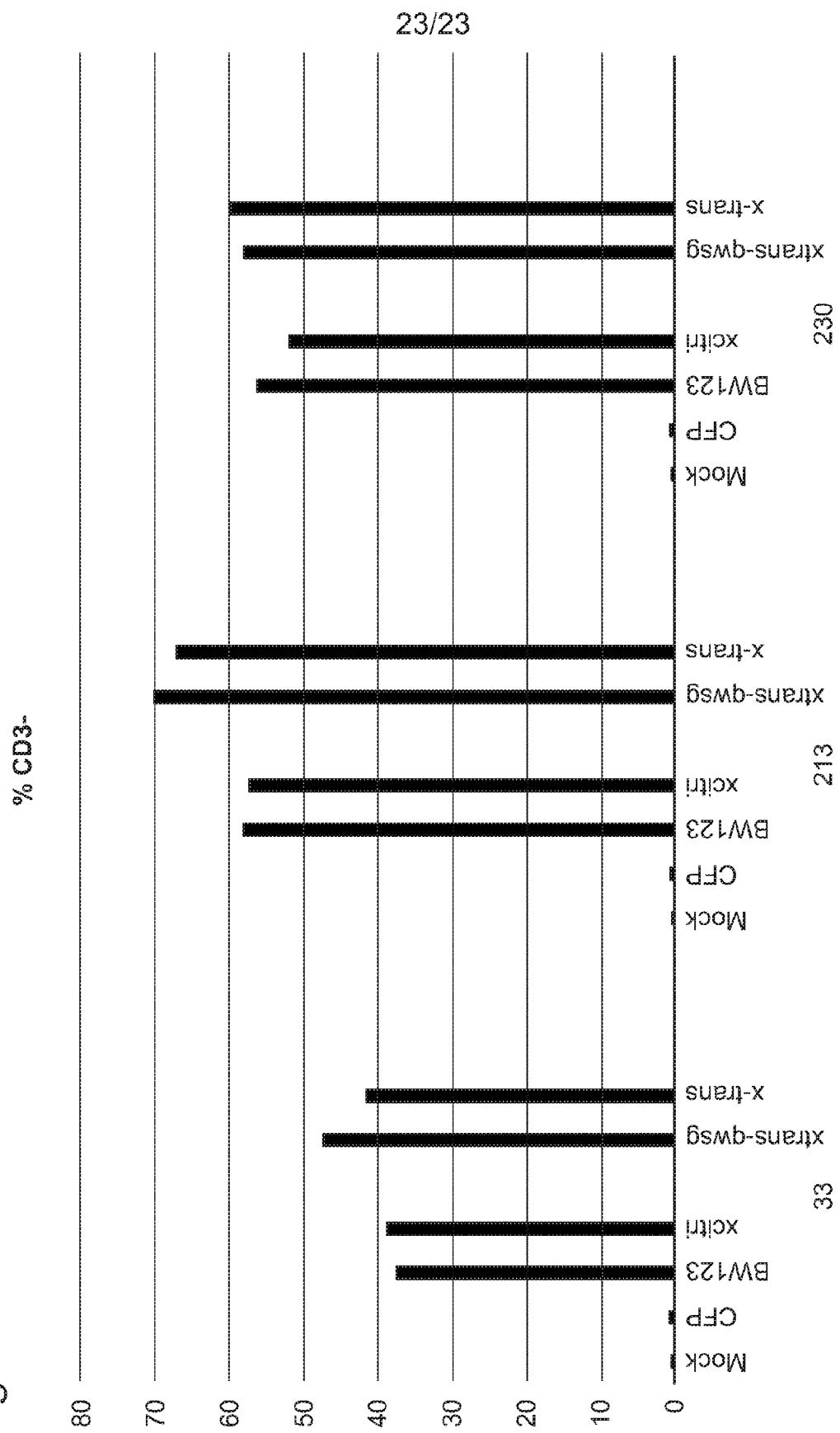
Figure 8

A.



22/23



**Figure 9**

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 17/56178

**A. CLASSIFICATION OF SUBJECT MATTER**  
 IPC(8) - A61K 35/12 (2017.01)  
 CPC - A61K 38/465, A61K 48/00, C12N 9/22

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

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

See Search History Document

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

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

See Search History Document

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X ----- Y	US 2016/0130569 A1 (CELLLECTIS et al.) 12 May 2016 (12.05.2016) para [0016]; [0040]-[0051]; SEQ ID NO: 7.	1, 2 ----- 3-8
Y	US 2014/0208457 A1 (FONFARA et al.) 24 July 2014 (24.07.2014) para [0009]; [0015]; [0020].	3-8

Further documents are listed in the continuation of Box C.

See patent family annex.

* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family
--	--

Date of the actual completion of the international search  29 December 2017	Date of mailing of the international search report  29 JAN 2018
Name and mailing address of the ISA/US  Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer:  Lee W. Young  PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

**INTERNATIONAL SEARCH REPORT**

International application No.

PCT/US 17/56178

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

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

1.  Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:
  
2.  Claims Nos.: because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
  
3.  Claims Nos.: 9-42 because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

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

This International Searching Authority found multiple inventions in this international application, as follows:

1.  As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2.  As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3.  As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
  
4.  No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

**Remark on Protest**

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