



(51) International Patent Classification:

C12N 15/10 (2006.01) C12N 9/22 (2006.01)
C12N 9/12 (2006.01)

(21) International Application Number:

PCT/US2022/053377

(22) International Filing Date:

19 December 2022 (19.12.2022)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

63/291,550 20 December 2021 (20.12.2021) US
63/347,709 01 June 2022 (01.06.2022) US

(71) Applicant: **INSCRIPTA, INC.** [US/US]; 5500 Central Ave., Suite 220, Boulder, Colorado 80301 (US).

(72) Inventors: **ZIMMERMAN, Erik**; 5500 Central Ave., Suite 220, Boulder, Colorado 80301 (US). **FELDMAN, Emily**; 5500 Central Ave., Suite 220, Boulder, Colorado 80301 (US). **SILTANEN, Christian**; 5500 Central Ave., Suite 220, Boulder, Colorado 80301 (US). **BARRETO FELISBINO, Marina**; 5500 Central Ave., Suite 220, Boulder, Colorado 80301 (US).

(74) Agent: **MARSH, David** et al.; Arnold & Porter Kaye Scholer LLP, 601 Massachusetts Ave. NW, Washington, District of Columbia 20001 (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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, 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).

(54) Title: TARGETED GENOMIC BARCODING FOR TRACKING OF EDITING EVENTS

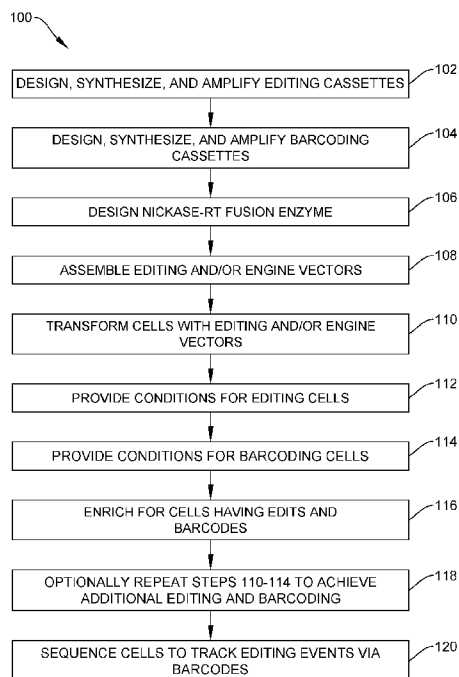


FIG. 1A

(57) Abstract: The present disclosure provides compositions of matter, methods and instruments for nucleic acid-guided nickase/reverse transcriptase fusion enzyme editing of nucleic acids in live mammalian cells, and for tracking of editing events.



Published:

— *with international search report (Art. 21(3))*

TARGETED GENOMIC BARCODING FOR TRACKING OF EDITING EVENTS**CROSS REFERENCE TO RELATED APPLICATIONS**

5 [0001] This application claims the benefit of U.S. Provisional Application No. 63/291,550, filed December 20, 2021, and U.S. Provisional Application No. 63/347,709, filed June 1, 2022, the contents of which are incorporated herein by reference in their entireties.

10

FIELD OF THE INVENTION

[0002] This invention relates to compositions of matter, methods, and instruments for tracking nucleic acid-guided editing events in live cells, particularly mammalian cells.

BACKGROUND OF THE INVENTION

15

[0003] In the following discussion certain articles and methods will be described for background and introductory purposes. Nothing contained herein is to be construed as an “admission” of prior art. Applicant expressly reserves the right to demonstrate, where appropriate, that the methods referenced herein do not constitute prior art under the applicable statutory provisions.

20

[0004] The ability to make precise, targeted changes to the genome of living cells has been a long-standing goal in biomedical research and development. Recently, various nucleases have been identified that allow manipulation of gene sequence, and hence gene function. The nucleases include nucleic acid-guided nucleases, which enable researchers to generate permanent edits in live cells. Of course, it is not only desirable to attain the highest editing rates possible in a cell population, but also to track the genomic edits in the cells, especially when multiple rounds of editing are performed and/or combinatorial libraries of edits are prepared. However, current tracking methods are inefficient and may lead to random genomic integration of tracking sequences, and/or require successive rounds of editing for targeted integration.

25

[0005] There is thus a need in the art of nucleic acid-guided nuclease editing for improved methods, compositions, modules, and instruments for efficient tracking of genomic edits, particularly in mammalian cells. The present disclosure addresses this need.

SUMMARY OF THE INVENTION

5 [0006] This Summary is provided to introduce a selection of concepts in a simplified form that are further described below in the Detailed Description. This Summary is not intended to identify key or essential features of the claimed subject matter, nor is it intended to be used to limit the scope of the claimed subject matter. Other features, details, utilities, and advantages of the claimed subject matter will be apparent from the following written
10 Detailed Description including those aspects illustrated in the accompanying drawings and defined in the appended claims.

[0007] In some aspects, the present disclosure provides a method for performing nucleic acid-guided nuclease/reverse transcriptase fusion editing in a genome of a live cell, comprising: (a) providing the live cell, wherein the live cell comprises a first target locus
15 and a second target locus; (b) providing a nucleic acid-guided nuclease/reverse transcriptase fusion enzyme; (c) providing a CF editing cassette, the CF editing cassette comprising: (i) a nucleic acid sequence encoding a first CFgRNA having a region of complementarity to a sequence of the first target locus; and (ii) a nucleic acid sequence encoding a first repair template; (d) providing a CF barcoding cassette, the CF barcoding
20 cassette comprising: (i) a nucleic acid sequence encoding a second CFgRNA having a region of complementarity to a sequence of the second target locus; and (ii) a nucleic acid sequence encoding a second repair template; (e) providing conditions to allow the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme, the first CFgRNA, and the first repair template to bind to the first target locus; (f) allowing the nucleic acid-guided
25 nuclease/reverse transcriptase fusion enzyme, the first CFgRNA, and the first repair template to edit the first target locus; (g) providing conditions to allow the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme, the second CFgRNA, and the second repair template to bind to the second target locus; and (h) allowing the nucleic acid-guided
30 nuclease/reverse transcriptase fusion enzyme, the second CFgRNA, and the second repair template to integrate the barcode into the second target locus.

[0008] In some aspects, the present disclosure provides an editing system comprising one or more vectors comprising: a nucleic acid sequence encoding a nucleic acid-guided nuclease/reverse transcriptase fusion enzyme; a CF editing cassette comprising: a nucleic acid sequence encoding a first CFgRNA having a region of complementarity to a sequence of a first target locus in a cell; and a nucleic acid sequence encoding a first repair template;
5 a CF barcoding cassette comprising: a nucleic acid sequence encoding a second CFgRNA having a region of complementarity to a sequence of a second target locus in the cell; and a nucleic acid sequence encoding a second repair template.

[0009] In some aspects, the present disclosure provides a vector comprising: a nucleic acid sequence encoding a nucleic acid-guided nuclease/reverse transcriptase fusion enzyme; a CF editing cassette comprising: a nucleic acid sequence encoding a first CFgRNA having a region of complementarity to a sequence of a first target locus in a cell; and a nucleic acid sequence encoding a first repair template; a CF barcoding cassette comprising: a nucleic acid sequence encoding a second CFgRNA having a region of complementarity to a sequence of a second target locus in the cell; and a nucleic acid sequence encoding a second repair template.
10
15

[0010] In some aspects, the present disclosure provides a method for performing nucleic acid-guided nuclease/reverse transcriptase fusion editing in a genome of a live cell, comprising: (a) providing a live cell suitable for the editing; (b) introducing a nucleic acid-guided nuclease/reverse transcriptase fusion enzyme; (c) providing conditions to allow the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme to bind to a first target locus; (d) allowing the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme to edit the first target locus; (e) providing conditions to allow the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme to bind to a second target locus; (f) allowing the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme to integrate a barcode into the second target locus.
20
25

[0011] These aspects and other features and advantages of the invention are described below in more detail.

30

BRIEF DESCRIPTION OF THE DRAWINGS

[0012] The foregoing and other features and advantages of the present invention will be more fully understood from the following detailed description of illustrative embodiments taken in conjunction with the accompanying drawings in which:

5 [0013] FIG. 1A is a simplified block diagram of an example of a method for editing and barcoding live cells to track editing events utilizing a CREATE Fusion (CF) editing cassette, a CF barcoding cassette, and a nucleic acid-guided nickase/reverse transcriptase fusion (“nickase-RT fusion”) enzyme. FIG. 1B is a simplified graphic depiction of the sequencing step of FIG. 1A for tracking editing events. FIG. 1C is a simplified graphic depiction of the mechanism of nucleic acid-guided nickase/reverse transcriptase fusion enzyme editing, generally. FIG. 1D schematically depicts an example of a single-vector system for trackable nickase-reverse transcriptase (RT) fusion editing of live cells, the single vector comprising a nickase-RT fusion enzyme (“Editing Enzyme”), a CF editing cassette (“Library gRNA”), a selectable marker (“PuroR”), and a CF barcoding cassette
10 (“Barcoding gRNA”), wherein a mouse U6 promoter (“mU6”), a Human elongation factor-1 alpha (EF-1 alpha) promoter (“EF1a”), a human U6 promoter (“hU6”), and a phosphoglycerate kinase promoter (“PGK”) are depicted as examples. FIG. 1E schematically depicts an example of a multi-vector system for trackable nickase-RT fusion editing of live cells, wherein a first vector comprises a CF barcoding cassette (“Barcoding gRNA”), and a second vector comprises a nickase-RT fusion enzyme (“Editing Enzyme”), a CF editing cassette (“Library gRNA”), and a selectable marker (“PuroR”). FIG. 1F schematically depicts an example of a CF editing cassette (top) and CF barcoding cassette (bottom) for trackable nickase-RT fusion editing of live cells. FIG. 1G is a simplified graphic depiction of an example of a mechanism for selection of successfully edited and/or
20 barcoded live cells.

[0014] FIGs. 2A – 2C depict three different views of an automated multi-module cell processing instrument for performing trackable nucleic acid-guided nuclease editing employing a split protein reporter system.

[0015] FIG. 3A depicts one aspect of a rotating growth vial for use with the cell growth module described herein and in relation to FIGs. 3B – 3D. FIG. 3B illustrates a perspective view of one aspect of a rotating growth vial in a cell growth module housing. FIG. 3C
30

depicts a cut-away view of the cell growth module from FIG. 3B. FIG. 3D illustrates the cell growth module of FIG. 3B coupled to LED, detector, and temperature regulating components.

[0016] FIG. 4A depicts retentate (top) and permeate (bottom) members for use in a tangential flow filtration module (*e.g.*, cell growth and/or concentration module), as well as the retentate and permeate members assembled into a tangential flow assembly (bottom). FIG. 4B depicts two side perspective views of a reservoir assembly of a tangential flow filtration module. FIGS. 4C – 4E depict an example of a top, with fluidic and pneumatic ports and gasket suitable for the reservoir assemblies shown in FIG. 4B.

[0017] FIG. 5A depicts an example of a combination reagent cartridge and electroporation device (*e.g.*, transformation module) that may be used in a multi-module cell processing instrument. FIG. 5B is a top perspective view of one aspect of an example of a flow-through electroporation device that may be part of a reagent cartridge. FIG. 5C depicts a bottom perspective view of one aspect of an example of a flow-through electroporation device that may be part of a reagent cartridge. FIGS. 5D – 5F depict a top perspective view, a top view of a cross section, and a side perspective view of a cross section of a flow-through electroporation device (FTEP) = useful in a multi-module automated cell processing instrument such as that shown in FIGS. 2A – 2C.

[0018] FIG. 6A depicts a simplified graphic of a workflow for singulating, editing and normalizing cells in a solid wall device. FIGS. 6B – 6D depict an aspect of a solid wall isolation incubation and normalization (SWIIN) module. FIG. 6E depicts the aspect of the SWIIN module in FIGS. 6B – 6D further comprising a heater and a heated cover.

[0019] FIG. 7 is a simplified process diagram of an aspect of an example of an automated multi-module cell processing instrument comprising a solid wall singulation/growth/editing/normalization module for recursive and trackable cell editing—including mammalian cell editing—in a system using a nickase-RT fusion enzyme and a genome-integrating CFgRNA.

[0020] FIGS. 8A – 8C illustrate an example of simultaneous barcoding and editing carried out in mammalian cells using CF barcoding cassettes and CF editing cassettes. FIG. 8A schematically illustrates an example of separate editing and barcoding plasmid designs for simultaneous barcoding and editing. FIG. 8B graphically illustrates the editing rates

observed in mammalian cells when transfected with a single editing plasmid, or dual editing and barcoding plasmids, and FIG. 8C graphically illustrates the barcoding rates observed in mammalian cells when transfected with the dual editing and barcoding plasmids.

5 [0021] FIGs. 9A – 9C illustrate another example of simultaneous barcoding and editing carried out in mammalian cells using CF barcoding cassettes and CF editing cassettes. FIG. 9A schematically illustrates an example of separate editing and barcoding plasmid designs (“Dual Plasmid”), as well as an example of a combined (*i.e.*, in tandem or distal configuration) editing and barcoding plasmid design (“Single Plasmid”), for simultaneous
10 barcoding and editing. FIG. 9B graphically illustrates the editing rates observed in in mammalian cells when transfected with a single tandem plasmid, a single distal plasmid, or dual editing and barcoding plasmids. FIG. 9C graphically illustrates the barcoding rates observed in mammalian cells when transfected with a single tandem plasmid, a single distal plasmid, or dual editing and barcoding plasmids.

15 [0022] FIGs. 10A and 10B illustrate another example of simultaneous barcoding and editing carried out in mammalian cells using CF barcoding cassettes and CF editing cassettes. FIG. 10A graphically illustrates the editing rates observed in mammalian cells when transfected with dual editing and barcoding plasmids versus a negative barcoding control. FIG. 10B depicts barcode insertion rates at a desired target locus as measured by
20 RNAseq.

[0023] FIG. 11 illustrates an example of barcoding in mammalian cells using CF barcoding cassettes comprising nucleic acid sequences encoding CFgRNAs designed for targeting one of a plurality of genomic loci corresponding to 3' untranslated regions (UTRs) of transcribed genes. In particular, FIG. 11 graphically illustrates the barcoding rates
25 observed in mammalian cells for each of the 96 different 3' UTR-corresponding loci targeted in an experiment.

[0024] FIG. 12 illustrates an example of simultaneous barcoding and editing in mammalian cells using CF barcoding cassettes comprising nucleic acid sequences encoding CFgRNAs designed for targeting one of a plurality of genomic loci
30 corresponding to 3' UTRs of transcribed genes. More particularly, FIG. 12 graphically

illustrates both editing and barcoding rates observed in mammalian cells when transfected with a single plasmid or dual plasmid system for four different 3' UTR barcoding loci.

[0025] FIG. 13 illustrates an example of barcoding in mammalian cells using CF barcoding cassettes comprising nucleic acid sequences encoding CFgRNAs designed for targeting one of a plurality of genomic loci corresponding to 3' UTRs of transcribed genes. More particularly, FIG. 13 illustrates relative expression levels of barcoded and non-barcoded transcripts as determined by single-cell RNASeq.

[0026] FIGs. 14A – 14C illustrate an example of simultaneous barcoding and editing in mammalian cells using a CFgRNA design for barcoded hemagglutinin (HA) epitope tag (“HA-tag”) knock-in edits. FIG. 14A depicts an example of a CFgRNA design for barcoded HA epitope tag knock-in edits. FIG. 14B depicts an example of a HA epitope tag barcoding scheme using synonymous codons, where there are 6.7×10^7 unique barcode sequences with a synonymous amino acid translation. FIG. 14C depicts a screening of CFgRNAs in induced pluripotent stem cells (iPSCs) for knock-in and successful expression and surface display of HA epitope tags at five endogenous surface receptor targets (BST2, CD151, CD63, CD81, and CD9), where the expression level of the HA-tag (y-axis) is measured by flow cytometry with PE (Phycoerythrin) labeled anti-HA antibodies, and where the screened CFgRNAs are on the x-axis.

[0027] FIGs. 15A – 15C illustrate an example of simultaneous barcoding and editing in mammalian cells (iPSCs), where the cells are co-transfected with CFgRNAs targeting a GFP-to-BFP edit and a CD81-HA tag knock-in barcoding edit. FIG. 15A depicts that physical magnetic sorting (magnetic-activated cell sorting “MACS”) with anti-HA antibody functionalized beads enriches the population of cells with a successful CD81-HA knock-in edit from 2.2% (left) to 84.1% (right). FIG. 15B depicts that GREEN FLUORESCENT PROTEIN (GFP)-to-BLUE FLUORESCENT PROTEIN (BFP) edit rates (y-axis) improved from 17% BFP+ in unsorted cells to 50% BFP+ in MACS enriched cells. FIG. 15C depicts edit rates at various endogenous targets, where the cells are co-transfected with CFgRNAs targeting various edits at the endogenous loci and a CD81-HA tag knock-in barcoding edit. Sorting for HA knock-in via MACS or fluorescence-activated cell sorting (“FACS”) improves edit rates at most endogenous targets relative to the unsorted samples.

[0028] FIGs. 16A – 16D illustrate examples of hypoxanthine phosphoribosyltransferase (HPRT) loss-of-function edits allowing for negative selection by resistance to 6-thioguanine (6-TG). FIG. 16A depicts an example of a scheme for HPRT disruption by a frame shift barcode insertion. FIG. 16B depicts an example of a scheme for HPRT knock-out that is used in FIGs. 16C and 16D. FIG. 16C depicts iPSCs co-transfected with CFgRNAs targeting a GFP-to-BFP edit and an HPRT knockout edit (“HPRT DF”), wherein 6-TG treatment selects for cells with the HPRT DF edit. FIG. 16D depicts that negative selection for HPRT DF with 6-TG supplemented media improves the GFP-to-BFP edit rate (y-axis) from approximately 60% BFP+ to approximately 80% BFP+.

10 [0029] It should be understood that the drawings are not necessarily to scale, and that like reference numbers refer to like features.

DETAILED DESCRIPTION

[0030] All the functionalities described in connection with one aspect are intended to be applicable to the additional aspects described herein except where expressly stated or where the feature or function is incompatible with the additional aspects. For example, where a given feature or function is expressly described in connection with one aspect but not expressly mentioned in connection with an alternative aspect, it should be understood that the feature or function may be deployed, utilized, or implemented in connection with the alternative aspect unless the feature or function is incompatible with the alternative aspect.

[0031] The practice of the techniques described herein may employ, unless otherwise indicated, conventional techniques and descriptions of organic chemistry, polymer technology, molecular biology (including recombinant techniques), cell biology, biochemistry and sequencing technology, which are within the skill of those who practice in the art. Such conventional techniques include polymer array synthesis, hybridization and ligation of polynucleotides, and detection of hybridization using a label. Specific illustrations of suitable techniques can be had by reference to the examples herein. However, other equivalent conventional procedures can, of course, also be used. Such conventional techniques and descriptions can be found in standard laboratory manuals such as Green, et al., Eds. (1999), *Genome Analysis: A Laboratory Manual Series* (Vols. I-IV); Weiner, Gabriel, Stephens, Eds. (2007), *Genetic Variation: A Laboratory Manual*;

Dieffenbach, Dveksler, Eds. (2003), *PCR Primer: A Laboratory Manual*; Mount (2004), *Bioinformatics: Sequence and Genome Analysis*; Sambrook and Russell (2006), *Condensed Protocols from Molecular Cloning: A Laboratory Manual*; and Sambrook and Russell (2002), *Molecular Cloning: A Laboratory Manual* (all from Cold Spring Harbor Laboratory Press); Stryer, L. (1995) *Biochemistry* (4th Ed.) W.H. Freeman, New York N.Y.; Gait, "Oligonucleotide Synthesis: A Practical Approach" 1984, IRL Press, London; Nelson and Cox (2000), *Lehninger, Principles of Biochemistry* 3rd Ed., W. H. Freeman Pub., New York, N.Y.; Berg et al. (2002) *Biochemistry*, 5th Ed., W.H. Freeman Pub., New York, N.Y.; all of which are herein incorporated in their entirety by reference for all purposes. CRISPR-specific techniques can be found in, e.g., *Genome Editing and Engineering from TALENs and CRISPRs to Molecular Surgery*, Appasani and Church (2018); and *CRISPR: Methods and Protocols*, Lindgren and Charpentier (2015); both of which are herein incorporated in their entirety by reference for all purposes.

[0032] Note that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "an oligonucleotide" refers to one or more oligonucleotides, and reference to "an automated system" includes reference to equivalent steps and methods for use with the system known to those skilled in the art, and so forth. Additionally, it is to be understood that terms such as "left," "right," "top," "bottom," "front," "rear," "side," "height," "length," "width," "upper," "lower," "interior," "exterior," "inner," "outer" that may be used herein merely describe points of reference and do not necessarily limit aspects of the present disclosure to any particular orientation or configuration. Furthermore, terms such as "first," "second," "third," etc., merely identify one of a number of portions, components, steps, operations, functions, and/or points of reference as disclosed herein, and likewise do not necessarily limit aspects of the present disclosure to any particular configuration or orientation.

[0033] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. All publications mentioned herein are incorporated by reference herein in their entireties.

[0034] When a range of numbers is provided herein, the range is understood to be inclusive of the edges of the range as well as any number between the defined edges of the range. For example, “between 1 and 10” includes any number between 1 and 10, as well as the number 1 and the number 10.

5 [0035] The term “about” means plus or minus 10% of the numerical value of the number with which it is being used. For example, “about 100” refers to numbers between (and including) 90 and 110.

[0036] When a grouping of alternatives is presented, any and all combinations of the members that make up that grouping of alternatives is specifically envisioned. For
10 example, if an item is selected from a group consisting of A, B, C, and D, the inventors specifically envisions each alternative individually (*e.g.* A alone, B alone, etc.), as well as combinations such as A, B, and D; A and C; B and C; etc.

[0037] The term “and/or” when used in a list of two or more items means any one of the listed items by itself or in combination with any one or more of the other listed items.
15 For example, the expression “A and/or B” is intended to mean either or both of A and B - *i.e.*, A alone, B alone, or A and B in combination. The expression “A, B and/or C” is intended to mean A alone, B alone, C alone, A and B in combination, A and C in combination, B and C in combination, or A, B, and C in combination.

[0038] In the following description, numerous specific details are set forth to provide
20 a more thorough understanding of the present invention. However, it will be apparent to one of ordinary skill in the art that the present invention may be practiced without one or more of these specific details. In other instances, well-known features and procedures well known to those skilled in the art have not been described in order to avoid obscuring the invention.

25 [0039] The term “complementary” as used herein refers to Watson-Crick base pairing between nucleotides and specifically refers to nucleotides hydrogen bonded to one another with thymine or uracil residues linked to adenine residues by two hydrogen bonds and cytosine and guanine residues linked by three hydrogen bonds. The terms “percent complementarity” or “percent complementary” as used herein in reference to two
30 nucleotide sequences is similar to the concept of percent identity but refers to the percentage of nucleotides of a query sequence that optimally base-pair or hybridize to

nucleotides a subject sequence when the query and subject sequences are linearly arranged and optimally base paired without secondary folding structures, such as loops, stems or hairpins. Such a percent complementarity can be between two DNA strands, two RNA strands, or a DNA strand and a RNA strand. The “percent complementarity” can be

5 calculated by (i) optimally base-pairing or hybridizing the two nucleotide sequences in a linear and fully extended arrangement (*e.g.*, without folding or secondary structures) over a window of comparison, (ii) determining the number of positions that base-pair between the two sequences over the window of comparison to yield the number of complementary positions, (iii) dividing the number of complementary positions by the total number of

10 positions in the window of comparison, and (iv) multiplying this quotient by 100% to yield the percent complementarity of the two sequences. Optimal base pairing of two sequences can be determined based on the known pairings of nucleotide bases, such as G-C, A-T, and A-U, through hydrogen binding. If the “percent complementarity” is being calculated in relation to a reference sequence without specifying a particular comparison window, then

15 the percent identity is determined by dividing the number of complementary positions between the two linear sequences by the total length of the reference sequence. Thus, for purposes of the present application, when two sequences (query and subject) are optimally base-paired (with allowance for mismatches or non-base-paired nucleotides), the “percent complementarity” for the query sequence is equal to the number of base-paired positions

20 between the two sequences divided by the total number of positions in the query sequence over its length, which is then multiplied by 100%. In general, a nucleic acid includes a nucleotide sequence described as having a “percent complementarity” or being a “percent complementary” to a specified second nucleotide sequence. For example, a nucleotide sequence may have 70%, 80%, 90%, 95%, 99%, or 100% complementarity to a specified

25 second nucleotide sequence, indicating that, for example, 7 of 10, 8 of 10, 9 of 10, 19 of 20, 99 of 100, or 10 of 10 nucleotides, respectively, of a sequence are complementary to the specified second nucleotide sequence. For example, the nucleotide sequence 3'-TCGA-5' is 100% complementary to the nucleotide sequence 5'-AGCT-3'; and the nucleotide sequence 3'-TCGA-5' is 100% complementary to a region of the nucleotide sequence 5'-

30 TAGCTG-3'.

[0040] The term DNA "control sequences" refers collectively to promoter sequences, polyadenylation signals, transcription termination sequences, upstream regulatory domains, origins of replication, internal ribosome entry sites, nuclear localization sequences, enhancers, and the like, which collectively provide for the replication, transcription and translation of a coding sequence in a recipient cell. Not all of these types of control sequences need to be present so long as a selected coding sequence is capable of being replicated, transcribed and—for some components—translated in an appropriate host cell.

[0041] A "regulatory sequence" or "regulatory region" refers to the region of a gene where RNA polymerase and other accessory transcription modulator proteins (*e.g.*, transcription factors) bind and interact to control transcription of the gene. Non-limiting examples of regulatory sequences or regions include promoters, enhancers, and terminators. Regulatory sequences or regions are capable of increasing or decreasing gene expression. As a result, these elements can control net protein expression from the gene.

[0042] The terms "CREATE fusion gRNA" or "CFgRNA" refer to a gRNA engineered to function with a nucleic acid-guided nickase/reverse transcriptase fusion enzyme (a "nickase-RT fusion") where the CFgRNA is designed to bind to and facilitate editing or barcoding of one or both DNA strands in a target locus of a cell genome. In certain aspects, "CREATE fusion gRNA" or "CFgRNA" refer to one of two gRNAs engineered to function with a nucleic acid-guided nickase/reverse transcriptase fusion enzyme (a "nickase-RT fusion") where the two CFgRNAs are designed to bind to and edit/barcode opposite DNA strands in a target locus. The two CFgRNAs specific to a target locus have regions of complementarity to one another at least at the site of the edit and preferably at regions 5' and 3' to the site of the edit. The term "complementary CFgRNAs" refers to two CFgRNAs engineered to bind to opposite DNA strands in a target locus which often create the complementary edit at a site in the target locus.

[0043] The terms "CREATE fusion barcoding cassette" or "CF barcoding cassette" in the context of the current methods and compositions refers to a nucleic acid molecule comprising a coding sequence for transcription of a CREATE fusion gRNA or "CFgRNA" to effect barcoding in a nucleic acid-guided nickase/reverse transcriptase fusion system where the CFgRNA is designed to bind to and facilitate incorporation of a barcode

sequence into one or both DNA strands in a target locus. In certain aspects, "CF barcoding cassette" refers to a nucleic acid molecule comprising a coding sequence for transcription of two gRNAs to effect barcoding in a nucleic acid-guided nickase/reverse transcriptase fusion system where the two gRNAs are designed to bind to and integrate barcode sequences into opposite DNA strands in a target locus.

[0044] The terms "CREATE fusion editing cassette" or "CF editing cassette" in the context of the current methods and compositions refers to a nucleic acid molecule comprising a coding sequence for transcription of a CREATE fusion gRNA or "CFgRNA" to effect editing in a nucleic acid-guided nickase/reverse transcriptase fusion system where the CFgRNA is designed to bind to and facilitate editing of one or both DNA strands in a target locus. In certain aspects, "CF editing cassette" refers to a nucleic acid molecule comprising a coding sequence for transcription of two gRNAs to effect editing in a nucleic acid-guided nickase/reverse transcriptase fusion system where the two gRNAs are designed to bind to and edit opposite DNA strands in a target locus.

[0045] The terms "CREATE fusion editing system" or "CF editing system" refer to the combination of a nucleic acid-guided nickase enzyme/reverse transcriptase fusion protein ("nickase-RT fusion") and a CREATE fusion editing cassette ("CF editing cassette") to effect editing in live cells. In certain aspects, a CF editing system further includes a CREATE fusion barcoding cassette ("CF barcoding cassette").

[0046] The terms "guide nucleic acid" or "guide RNA" or "gRNA" refer to a polynucleotide comprising 1) a guide sequence capable of hybridizing to a target genomic locus, and 2) a scaffold sequence capable of interacting or complexing with a nucleic acid-guided nuclease.

[0047] "Homology" or "identity" or "similarity" refers to sequence similarity between two peptides or, more often in the context of the present disclosure, between two nucleic acid molecules. The term "homologous region" or "homology arm" refers to a region on a donor DNA with a certain degree of homology with a target genomic DNA sequence. Homology can be determined by comparing a position in each sequence which may be aligned for purposes of comparison. When a position in the compared sequence is occupied by the same base or amino acid, then the molecules are homologous at that position. A

degree of homology between sequences is a function of the number of matching or homologous positions shared by the sequences.

[0048] The terms “percent identity” or “percent identical” as used herein in reference to two or more nucleotide or amino acid sequences is calculated by (i) comparing two optimally aligned sequences (nucleotide or amino acid) over a window of comparison (the “alignable” region or regions), (ii) determining the number of positions at which the identical nucleic acid base (for nucleotide sequences) or amino acid residue (for proteins and polypeptides) occurs in both sequences to yield the number of matched positions, (iii) dividing the number of matched positions by the total number of positions in the window of comparison, and then (iv) multiplying this quotient by 100% to yield the percent identity. If the “percent identity” is being calculated in relation to a reference sequence without a particular comparison window being specified, then the percent identity is determined by dividing the number of matched positions over the region of alignment by the total length of the reference sequence. Accordingly, for purposes of the present application, when two sequences (query and subject) are optimally aligned (with allowance for gaps in their alignment), the “percent identity” for the query sequence is equal to the number of identical positions between the two sequences divided by the total number of positions in the query sequence over its length (or a comparison window), which is then multiplied by 100%. When percentage of sequence identity is used in reference to amino acids it is recognized that residue positions which are not identical often differ by conservative amino acid substitutions, where amino acid residues are substituted for other amino acid residues with similar chemical properties (*e.g.*, charge or hydrophobicity) and therefore do not change the functional properties of the molecule. When sequences differ in conservative substitutions, the percent sequence identity can be adjusted upwards to correct for the conservative nature of the substitution. Sequences that differ by such conservative substitutions are said to have “sequence similarity” or “similarity.”

[0049] For optimal alignment of sequences to calculate their percent identity, various pair-wise or multiple sequence alignment algorithms and programs are known in the art, such as ClustalW or Basic Local Alignment Search Tool® (BLAST™), etc., that can be used to compare the sequence identity or similarity between two or more nucleotide or amino acid sequences. Although other alignment and comparison methods are known in the art, the

alignment and percent identity between two sequences (including the percent identity ranges described above) can be as determined by the ClustalW algorithm, *see, e.g.*, Chenna *et al.*, "Multiple sequence alignment with the Clustal series of programs," *Nucleic Acids Research* 31: 3497-3500 (2003); Thompson *et al.*, "Clustal W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice," *Nucleic Acids Research* 22: 4673-4680 (1994); Larkin MA *et al.*, "Clustal W and Clustal X version 2.0," *Bioinformatics* 23: 2947-48 (2007); and Altschul *et al.* "Basic local alignment search tool." *J. Mol. Biol.* 215:403-410 (1990), the entire contents and disclosures of which are incorporated herein by reference.

10 [0050] The term "nucleic acid" or "polynucleotide" refers to deoxyribonucleic acids (DNA) or ribonucleic acids (RNA) and polymers thereof in either single- or double-stranded form. Unless otherwise indicated, the terms encompass nucleic acids containing known analogues or natural nucleotides that have similar binding properties as the reference nucleic acid and are metabolized in a manner similar to naturally occurring
15 nucleotides. Unless otherwise indicated, in addition to the sequence specifically stated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (*e.g.*, degenerate codon substitutions), alleles, orthologues, SNPs, and complementary sequences. The term nucleic acid is used interchangeably with DNA, RNA, cDNA, gene, and mRNA encoded by a gene.

20 [0051] As used herein, "nucleic acid-guided nickase/reverse transcriptase fusion" or "nickase-RT fusion" refers to a nucleic acid-guided nickase—or nucleic acid-guided nuclease or CRISPR nuclease that has been engineered to act as a nickase rather than a nuclease that initiates double-stranded DNA breaks—where the nucleic acid-guided nickase is fused to a reverse transcriptase, which is an enzyme used to generate cDNA from
25 an RNA template. In certain aspects, "nucleic acid-guided nickase/reverse transcriptase fusion" or "nickase-RT fusion" refers to two or more nucleic acid-guided nickases—or nucleic acid-guided nucleases or CRISPR nucleases that have been engineered to act as nickases rather than nucleases that initiate double-stranded DNA breaks—where the nucleic acid-guided nickases are fused to a reverse transcriptase. For information regarding
30 nickase-RT fusions see, *e.g.*, USPN 10,689,669 and USSN 16/740,421.

[0052] “Nucleic acid-guided editing components” refers to one or both of a nickase-RT fusion and CREATE fusion guide nucleic acids (CFgRNAs).

[0053] “Operably linked” refers to an arrangement of elements where the components so described are configured so as to perform their usual function. Thus, control sequences operably linked to a coding sequence are capable of effecting the transcription, and in some cases, the translation, of a coding sequence. The control sequences need not be contiguous with the coding sequence so long as they function to direct the expression of the coding sequence. Thus, for example, intervening untranslated yet transcribed sequences can be present between a promoter sequence and the coding sequence and the promoter sequence can still be considered “operably linked” to the coding sequence. In fact, such sequences need not reside on the same contiguous DNA molecule (*e.g.* chromosome) and may still have interactions resulting in altered regulation.

[0054] A “PAM mutation” refers to one or more edits to a target sequence that removes, mutates, or otherwise renders inactive a protospacer adjacent motif (PAM) or spacer region in the target sequence.

[0055] A “promoter” or “promoter sequence” is a DNA regulatory region capable of binding RNA polymerase and initiating transcription of a polynucleotide or polypeptide coding sequence such as messenger RNA, ribosomal RNA, small nuclear or nucleolar RNA, guide RNA, or any kind of RNA. In some aspects, a promoter is an endogenous promoter, synthetically produced, varied, or derived from a known or naturally occurring promoter sequence or other promoter sequence. In some aspects, a promoter is a constitutive promoter. In some aspects, a promoter is an inducible promoter. In some aspects, a promoter is a heterologous promoter.

[0056] A “terminator” or “terminator sequence” or “transcription termination sequence” refers to a DNA regulatory region of a gene that signals termination of transcription of the gene to an RNA polymerase. Without being limiting, terminators cause transcription of an operably linked nucleic acid molecule to stop.

[0057] A “coding sequence” or “coding region” refers to the region of a gene’s DNA or RNA which codes for a protein. In DNA, the coding region of a gene is flanked by the promoter sequence on the 5’ end of the template strand and the termination sequence on the 3’ end. After transcription, the coding region in an mRNA is flanked by the 5’

untranslated region (5'-UTR) and 3' untranslated region (3'-UTR), the 5' cap, and poly-A tail.

[0058] A “non-coding sequence” or “non-coding region” refers to the region of a gene’s DNA which does not code for a protein. However, some non-coding DNA is transcribed into functional non-coding RNA molecules (*e.g.*, transfer RNA, microRNA, siRNA, piRNA, ribosomal RNA, and regulatory RNAs). Other functional non-coding DNA include, for example, regulatory sequences of a gene that control its expression.

[0059] As used herein “gene product” refers to a biochemical material, either RNA or protein, resulting from expression of a gene. In some aspects, a gene product is an RNA molecule, *e.g.*, transfer RNA, microRNA, siRNA, piRNA, ribosomal RNA, or regulatory RNA. In some aspects, the gene product is a protein. In some aspects, the gene product is an enzyme. In some aspects, the gene product is a membrane protein. In some aspects, the gene product is a protein involved in the expression of a gene. In some aspects, the gene product is a transcription factor. In some aspects, the gene product is a coactivator protein. In some aspects, the gene product is a corepressor protein. In some aspects, the gene product is a chromatin-binding protein.

[0060] As used herein, the terms “protein,” “peptide,” and “polypeptide” are used interchangeably herein and refer to a polymer of amino acid residues. In some aspects, proteins are made up entirely of amino acids transcribed by any class of any RNA polymerase I, II or III.

[0061] As used herein, the term “repair template” in the context of a CREATE fusion editing system employing a nickase-RT fusion enzyme refers to a nucleic acid (*e.g.*, a ribonucleic acid) that is designed to serve as a template (including a desired edit or barcode) to be incorporated into target DNA via reverse transcription (*e.g.*, by reverse transcriptase).

[0062] As used herein the term “selectable marker” refers to a gene introduced into a cell, which confers a trait suitable for artificial selection. General use selectable markers are well-known to those of ordinary skill in the art. Drug selectable markers such as ampicillin/carbenicillin, kanamycin, chloramphenicol, nourseothricin N-acetyl transferase, erythromycin, tetracycline, gentamicin, bleomycin, streptomycin, puromycin, hygromycin, blasticidin, and G418 may be employed. In other aspects, selectable markers include, but are not limited to human nerve growth factor receptor (detected with a MAb, such as

described in U.S. Pat. No. 6,365,373); truncated human growth factor receptor (detected with MAb); mutant human dihydrofolate reductase (DHFR; fluorescent MTX substrate available); secreted alkaline phosphatase (SEAP; fluorescent substrate available); human thymidylate synthase (TS; confers resistance to anti-cancer agent fluorodeoxyuridine);
5 human glutathione S-transferase alpha (GSTA1; conjugates glutathione to the stem cell selective alkylator busulfan; chemoprotective selectable marker in CD34+cells); CD24 cell surface antigen in hematopoietic stem cells; human CAD gene to confer resistance to N-phosphonacetyl-L-aspartate (PALA); human multi-drug resistance-1 (MDR-1; P-glycoprotein surface protein selectable by increased drug resistance or enriched by FACS);
10 human CD25 (IL-2 α ; detectable by Mab-FITC); Methylguanine-DNA methyltransferase (MGMT; selectable by carmustine); rhamnose; and Cytidine deaminase (CD; selectable by Ara-C). In some aspects, a selectable marker comprises an antibiotic resistance gene. In some aspects, a selectable marker comprises a puromycin resistance gene. "Selective medium" as used herein refers to cell growth medium to which has been added a chemical
15 compound or biological moiety that selects for or against selectable markers. In some aspects, a selectable marker provides a phenotypic handle for live-cell selection. In some aspects, the selection is a positive selection for a knock-in edit. In some aspects, the selection is a negative selection for a knockout edit. In some aspects, a HA epitope tag can be used as a phenotypic selection marker. In some aspects, 6-GT can be used to select for
20 HPRT knockout edits.

[0063] A "locus" refers to a fixed position in a genome. In some aspects, a locus comprises a coding region. In some aspects, a locus comprises a non-coding region. In some aspects, a locus comprises a gene. In an aspect, a locus comprises at least 1 nucleotide. In an aspect, a locus comprises at least 10 nucleotides. In an aspect, a locus
25 comprises at least 25 nucleotides. In an aspect, a locus comprises at least 50 nucleotides. In an aspect, a locus comprises at least 100 nucleotides. In an aspect, a locus comprises at least 250 nucleotides. In an aspect, a locus comprises at least 500 nucleotides. In an aspect, a locus comprises at least 1000 nucleotides. In an aspect, a locus comprises at least 2500 nucleotides. In an aspect, a locus comprises at least 5000 nucleotides.

30 [0064] The terms "target sequence", "target genomic DNA locus", "target locus", or "target genomic locus" refer to any locus *in vitro* or *in vivo*, or in a nucleic acid (*e.g.*,

genome or episome) of a cell or population of cells, in which a change of at least one nucleotide is desired using a nucleic acid-guided nuclease editing system. The target sequence can be a genomic locus or extrachromosomal locus. In some aspects, a target locus refers to a position in a genome targeted to be edited by the nucleic acid-guided
5 nuclease/reverse transcriptase fusion enzyme and the CF editing cassette. In some aspects, a target locus comprises a gene, including its regulatory regions and coding regions. In some aspects, a target locus comprises a regulatory region of a gene, *e.g.*, a promoter region or a terminator region. In some aspects, the target locus is within a nuclear genome. In some aspects, the target locus is within a mitochondrial genome. In some aspects, the target
10 locus is within a vector.

[0065] In some aspects, an "integration locus" refers to a position in a genome targeted for the integration of a CF editing cassette. In some aspects, an integration locus comprises a coding region. In some aspects, an integration locus comprises a non-coding region. In some aspects, an integration locus comprises a "safe harbor locus." A "safe harbor locus"
15 as used herein refers to an intergenic region that has a reduced potential for the CF editing cassette integration adversely affecting genes neighboring (*e.g.*, within 10 kb) the integrated CF editing cassette.

[0066] The term "gene" refers to a nucleic acid region which includes a coding region operably linked to a suitable regulatory region capable of regulating the expression of a
20 gene product (*e.g.*, a protein or functional non-coding RNA) in some manner. Genes include untranslated regulatory regions (*e.g.*, promoters, enhancers, repressors, etc.) in the DNA before (upstream) and after (downstream) the coding region (open reading frame, ORF), and, where applicable, intervening sequences (*e.g.*, introns) between individual coding regions (*e.g.*, exons).

[0067] The term "variant" refers to a polypeptide or polynucleotide that differs from a reference polypeptide or polynucleotide. A typical variant of a polypeptide differs in amino acid sequence from another reference polypeptide. Generally, differences may be limited so that the sequences of the reference polypeptide and the variant are closely similar overall
25 (*e.g.*, at least 90% identical) and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more modifications (*e.g.*,
30 substitutions, additions, and/or deletions). A variant of a polypeptide may be a

conservatively modified variant (*e.g.*, at least 95% identical to the reference polypeptide). A substituted or inserted amino acid residue may or may not be one encoded by the genetic code (*e.g.*, a non-natural amino acid). A variant of a polypeptide may be naturally occurring, such as an allelic variant, or it may be a variant that is not known to occur naturally.

5 [0068] A “vector” is any of a variety of nucleic acids that comprise a desired sequence or sequences to be delivered to and/or expressed in a cell. Vectors are typically composed of DNA, although RNA vectors are also available. Vectors include, but are not limited to, plasmids, fosmids, phagemids, virus genomes, BACs, YACs, PACs, synthetic
10 chromosomes, and the like. In the present disclosure, a single vector may include a coding sequence for a nickase-RT fusion enzyme and a CF editing cassette and/or CFgRNA sequence to be transcribed. In other aspects, however, two vectors — *e.g.*, an engine vector comprising the coding sequence for the nickase-RT fusion enzyme, and an editing vector, comprising the CFgRNA sequence to be transcribed — may be used.

15 [0069] As used herein, a “mutation” refers to an inheritable genetic modification introduced into a gene to alter the expression or activity of a product encoded by the gene. In some aspects, “mutation,” “modification,” and “edit” may be used interchangeably in the present disclosure. In some aspects, a modification can be in any sequence region of a gene, for example, in a promoter, 5' UTR, exon, 3' UTR, or terminator region. In some
20 aspects, a modification is in the regulatory region of a gene. In some aspects, a modification is in the coding region of a gene. In some aspects, a modification is in an exon. In some aspects, a modification is in an intron. In some aspects, a modification spans an intron/exon junction. In some aspects, a modification reduces, inhibits, or eliminates the expression or activity of a gene product as compared to an unmodified control. In some aspects, a
25 modification increases, elevates, strengthens, or augments the expression or activity of a gene product as compared to an unmodified control.

[0070] In some aspects, a mutation, or modification is a “non-natural” or “non-naturally occurring” mutation or modification. As used herein, a “non-natural” or “non-naturally occurring” mutation or modification refers to a non-spontaneous mutation or modification
30 generated via human intervention, and does not correspond to a spontaneous mutation or modification generated without human intervention. Non-limiting examples of human

intervention include mutagenesis (*e.g.*, chemical mutagenesis, ionizing radiation mutagenesis) and targeted genetic modifications (*e.g.*, nucleic-acid guided nuclease-based methods, CREATE fusion-based methods, CRISPR-based methods, TALEN-based methods, zinc finger-based methods). Non-natural mutations or modifications and non-naturally occurring mutations or modifications do not include spontaneous mutations that arise naturally (*e.g.*, via aberrant DNA replication).

[0071] Several types of mutations or modifications are known in the art. In some aspects, a mutation or modification comprises an insertion. An “insertion” refers to the addition of one or more nucleotides or amino acids to a given polynucleotide or amino acid sequence, respectively, as compared to an endogenous reference polynucleotide or amino acid sequence. In an aspect, an insertion comprises an insertion of at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 25, at least 50, at least 75, at least 100, at least 200, at least 300, at least 400, at least 500, at least 750, at least 1000, or at least 2500 nucleotides.

[0072] In some aspects, a mutation or modification comprises a deletion. A “deletion” refers to the removal of one or more nucleotides or amino acids to a given polynucleotide or amino acid sequence, respectively, as compared to an endogenous reference polynucleotide or amino acid sequence. In an aspect, a deletion comprises a deletion of at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 25, at least 50, at least 75, at least 100, at least 200, at least 300, at least 400, at least 500, at least 750, at least 1000, or at least 2500 nucleotides.

[0073] In some aspects, a mutation or modification comprises a substitution or a swap. A “substitution” or “swap” refers to the replacement of one or more nucleotides or amino acids to a given polynucleotide or amino acid sequence, respectively, as compared to an endogenous reference polynucleotide or amino acid sequence. In some aspects, a “substitution allele” refers to a nucleic acid sequence at a particular locus comprising a substitution. In an aspect, a substitution comprises the substitution of at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, or at least 50 nucleotides. When more than 1 nucleotide is substituted, the substitutions can be contiguous or non-contiguous.

[0074] In some aspects, a mutation or modification comprises an inversion. An “inversion” refers to when a segment of a polynucleotide or amino acid sequence is reversed end-to-end. In an aspect, an inversion comprises an inversion of at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 25, at least 50, at least 75, at least 100, at least 200, at least 300, at least 400, at least 500, at least 750, at least 1000, or at least 2500 nucleotides.

[0075] In some aspects, a mutation or modification provided herein comprises a mutation selected from the group consisting of an insertion, a deletion, a substitution, and an inversion. In some aspects, a mutation or modification provided herein comprises an insertion. In some aspects, a mutation or modification provided herein comprises a deletion. In some aspects, a mutation or modification provided herein comprises a substitution. In some aspects, a mutation or modification provided herein comprises an inversion.

[0076] In some aspects, a mutation or modification comprises one or more mutation types selected from the group consisting of a nonsense mutation, a missense mutation, a frameshift mutation, a splice-site mutation, and any combinations thereof. As used herein, a “nonsense mutation” refers to a mutation to a nucleic acid sequence that introduces a premature stop codon to an amino acid sequence by the nucleic acid sequence. As used herein, a “missense mutation” refers to a mutation to a nucleic acid sequence that causes a substitution within the amino acid sequence encoded by the nucleic acid sequence. As used herein, a “frameshift mutation” refers to an insertion or deletion to a nucleic acid sequence that shifts the frame for translating the nucleic acid sequence to an amino acid sequence. A “splice-site mutation” refers to a mutation in a nucleic acid sequence that causes an intron to be retained for protein translation, or, alternatively, for an exon to be excluded from protein translation. Splice-site mutations can cause nonsense, missense, or frameshift mutations.

[0077] Mutations or modifications in coding regions of genes (*e.g.*, exonic mutations) can result in a truncated protein or polypeptide when a mutated messenger RNA (mRNA) is translated into a protein or polypeptide. In some aspects, this disclosure provides a mutation that results in the truncation of a protein or polypeptide. As used herein, a “truncated” protein or polypeptide comprises at least one fewer amino acid as compared to an

endogenous control protein or polypeptide. For example, if endogenous Protein A comprises 100 amino acids, a truncated version of Protein A can comprise between 1 and 99 amino acids.

5 [0078] Without being limited by any scientific theory, one way to cause a protein or polypeptide truncation is by the introduction of a premature stop codon in an mRNA transcript of an endogenous gene. In some aspects, this disclosure provides a mutation that results in a premature stop codon in an mRNA transcript of an endogenous gene. As used herein, a “stop codon” refers to a nucleotide triplet within an mRNA transcript that signals a termination of protein translation. A “premature stop codon” refers to a stop codon
10 positioned earlier (*e.g.*, on the 5'-side) than the normal stop codon position in an endogenous mRNA transcript. Without being limiting, several stop codons are known in the art, including “UAG,” “UAA,” “UGA,” “TAG,” “TAA,” and “TGA.” In some aspects, multiple (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more than 10) premature stop codons are introduced.

15 [0079] In some aspects, a mutation or modification provided herein comprises a null mutation. As used herein, a “null mutation” or “knockout edit” refers to a mutation that confers a decreased function or complete loss-of-function for a protein encoded by a gene comprising the mutation, or, alternatively, a mutation that confers a decreased function or complete loss-of-function for a small RNA encoded by a genomic locus. A null mutation
20 can cause lack or decrease of mRNA transcript production, small RNA transcript production, protein function, or a combination thereof. As used herein, a “null allele” refers to a nucleic acid sequence at a particular locus where a null mutation has conferred a decreased function or complete loss-of-function to the allele.

[0080] In some aspects, a “synonymous edit” or “synonymous substitution” is the
25 substitution of one base for another in an exon of a gene coding for a protein, such that the produced amino acid sequence is not modified. This is possible because the genetic code is “degenerate”, meaning that some amino acids are coded for by more than one three-base-pair codon; since some of the codons for a given amino acid differ by just one base pair from others coding for the same amino acid, a mutation that replaces the “normal” base by
30 one of the alternatives will result in incorporation of the same amino acid into the growing polypeptide chain when the gene is translated.

[0081] In an aspect, an edit is a “knock-in” edit. As used herein, a “knock-in edit” or a “knock-in mutation” the substitution or replacement of a non-functioning or low-functioning allele of a gene with a functional or higher-functioning allele of the gene. Knock-in edits are sometimes referred to in the art as gain-of-function edits.

5 [0082] In some aspects, “codon optimization” refers to experimental approaches designed to improve the codon composition of a recombinant gene based on various criteria without altering the amino acid sequence. This is possible because most amino acids are encoded by more than one codon. Codon optimization may be used to improve gene expression and increase the translation efficiency of a gene of interest by accommodating for codon bias
10 of the host organism. In some aspects, a nucleic acid molecule provided herein encodes a polypeptide that is codon optimized for a prokaryote. In some aspects, a nucleic acid molecule provided herein encodes a polypeptide that is codon optimized for an *Escherichia coli* cell. In some aspects, a nucleic acid molecule provided herein encodes a polypeptide that is codon optimized for a eukaryote. In some aspects, a nucleic acid molecule provided
15 herein encodes a polypeptide that is codon optimized for a mammalian cell. In some aspects, a nucleic acid molecule provided herein encodes a polypeptide that is codon optimized for a human cell. In some aspects, a nucleic acid molecule provided herein encodes a polypeptide that is codon optimized for a non-human mammalian cell. In some aspects, a nucleic acid molecule provided herein encodes a polypeptide that is codon
20 optimized for a fungal cell. In some aspects, a nucleic acid molecule provided herein encodes a polypeptide that is codon optimized for a *Saccharomyces cerevisiae* cell. In some aspects, a nucleic acid molecule provided herein encodes a polypeptide that is codon optimized for a plant cell. In some aspects, a nucleic acid molecule provided herein encodes a polypeptide that is codon optimized for an archaeal cell.

25 [0083] The present disclosure includes method of trackable nucleic acid-guided nuclease editing in cell populations, *e.g.*, prokaryotic, archaeal, and eukaryotic cells. In some aspects, the cells include mammalian cells. In some aspects, the cells include human cells. In some aspects, the cells include non-human mammalian cells. In some aspects, the cells include bacterial cells. In some aspects, the cells include *E. coli* cells. In some aspects, the
30 cells include fungal cells. In some aspects, the cells include *S. cerevisiae* cells. In some aspects, the cells include plant cells.

[0084] In some aspects, a mutation or modification provided herein can be positioned in any part of a gene. In some aspects, a mutation or modification provided herein can be positioned in the coding region of a gene. In some aspects, a mutation or modification provided herein can be positioned in the non-coding region of a gene. In some aspects, a mutation or modification provided herein can be positioned in the regulatory region of a gene. In some aspects, a mutation or modification provided herein is positioned within an exon of a gene. In some aspects, a mutation or modification provided herein is positioned within an intron of a gene. In some aspects, a mutation or modification provided herein is positioned within an exon and an intron of a gene. In a further aspect, a mutation or modification provided herein is positioned within a 5'-untranslated region (UTR) of a gene. In still another aspect, a mutation or modification provided herein is positioned within a 3'-UTR of a gene. In yet another aspect, a mutation or modification provided herein is positioned within a promoter of a gene. In yet another aspect, a mutation or modification provided herein is positioned within a terminator of a gene.

[0085] The present disclosure relates to methods and compositions for improved tracking of nucleic acid-guided nuclease editing. With the present compositions and methods, targeted editing and tracking of the intended edit(s) is facilitated using a barcoding gRNA covalently linked to a barcode sequence and designed to precisely insert the barcode sequence into a desired genomic locus. When introduced into cells along with an editing gRNA covalently linked to an intended edit and a corresponding nucleic acid-guided nuclease or nickase, the barcoding gRNA facilitates simultaneous editing and tracking (*e.g.*, barcoding) of the edit(s), wherein the edit is incorporated into a first target locus and the corresponding barcode is integrated into a second, separate target locus. The integrated barcode may then be tracked or analyzed via genomic sequencing (*e.g.*, amplicon-based next-generation sequencing), or RNA sequencing ("RNASeq") if the second target locus is a gene-coding region, to identify the edit, plasmid, or other construct that was co-delivered with the barcoding gRNA. And, because the barcode is integrated into the genome, the barcode may be tracked beyond the timeframe of any transient plasmid reagents utilized to facilitate editing, cell differentiation, and the like.

[0086] In certain aspects, the barcoding gRNA is a component of a barcoding cassette for performing tracking of nucleic acid-guided nuclease editing, the barcoding cassette

comprising the barcoding gRNA having a region of complementarity to a sequence of a target locus in which a barcode sequence is to be integrated, and a barcode sequence for integration into the cell genome having a unique sequence by which a corresponding edit may be identified.

5 [0087] In certain aspects, the barcoding gRNA is a CREATE fusion gRNA (“CFgRNA,” defined *infra*) and the barcoding cassette is a CREATE fusion barcoding cassette (“CF barcoding cassette,” defined *infra*) comprising from 5' to 3': (A) a nucleic acid sequence encoding a barcoding gRNA having a region of complementarity to a sequence of a target locus in which a barcode sequence is to be integrated, the barcoding gRNA comprising: a
10 guide or spacer sequence, and a scaffold region recognized by a corresponding nuclease or nickase; and (B) a nucleic acid sequence encoding a repair template covalently linked to the barcoding gRNA comprising from 5' to 3': an optional post-barcode homology region, a barcode sequence, a nick-to-barcode region, and a primer binding site (PBS). In some aspects, the components of the barcoding cassette are contiguous. In some aspects, the
15 barcoding cassette is agnostic to the order of the barcoding gRNA and repair template. In some aspects, the barcoding gRNA is under the control of a promoter at the 5' end of the barcoding cassette.

[0088] In certain aspects, the nick-to-barcode region of the repair template is between 1 nucleotide and 100 nucleotides in length, between 1 nucleotide and 75 nucleotides in
20 length, between 1 nucleotide and 50 nucleotides in length, between 2 nucleotides and 250 nucleotides in length, between 5 nucleotides and 150 nucleotides in length, or between 1 nucleotide and 150 nucleotides in length. In some aspects of this method, the nick-to-barcode region of the repair template is up to 10,000 nucleotides in length, up to 5000 nucleotides in length, up to 3000 nucleotides in length, up to 1000 nucleotides in length,
25 up to 500 nucleotides in length, up to 250 nucleotides in length, up to 100 nucleotides in length, up to 50 nucleotides in length, or up to 25 nucleotides in length.

[0089] In certain aspects, the post-barcode homology region of the repair template is between 2 nucleotides and 20 nucleotides in length, between 2 nucleotides and 15 nucleotides in length, between 2 nucleotides and 50 nucleotides in length, between 4
30 nucleotides and 40 nucleotides in length, between 3 nucleotides and 30 nucleotides in length or between 5 nucleotides and 25 nucleotides in length.

[0090] In certain aspects, the editing gRNA is a component of an editing cassette for performing nucleic acid-guided nuclease editing, the editing cassette comprising the editing gRNA having a region of complementarity to a sequence of a target locus in which an edit is to be incorporated, and an edit for incorporation into the cell genome.

5 [0091] In certain aspects of the present disclosure, the editing gRNA is a CREATE fusion gRNA (“CFgRNA,” defined *infra*) and the editing cassette is a CREATE fusion editing cassette (“CF editing cassette,” defined *infra*) comprising from 5' to 3': (A) a nucleic acid sequence encoding an editing gRNA having a region of complementarity to a sequence of a target locus in which an edit is to be incorporated, the editing gRNA comprising: a guide
10 or spacer sequence, and a scaffold region recognized by a corresponding nuclease or nickase; and (B) a nucleic acid sequence encoding a repair template covalently linked to the editing gRNA comprising from 5' to 3': an optional post-edit homology region, an edit, an optional nick-to-edit region, and a primer binding site (“PBS”). In some aspects, the components of the editing cassette are contiguous. In some aspects, the editing cassette is
15 agnostic to the order of the editing gRNA and repair template. In some aspects, the editing gRNA is under the control of a promoter at the 5' end of the CF editing cassette.

[0092] In certain aspects, the nick-to-edit region of the repair template is between 1 nucleotide and 100 nucleotides in length, between 1 nucleotide and 75 nucleotides in length, between 1 nucleotide and 50 nucleotides in length, between 2 nucleotides and 250
20 nucleotides in length, between 5 nucleotides and 150 nucleotides in length, or between 1 nucleotide and 150 nucleotides in length. In some aspects of this method, the nick-to-edit region of the repair template is up to 10,000 nucleotides in length, up to 5000 nucleotides in length, up to 3000 nucleotides in length, up to 1000 nucleotides in length, up to 500 nucleotides in length, up to 250 nucleotides in length, up to 100 nucleotides in length, up
25 to 50 nucleotides in length, or up to 25 nucleotides in length.

[0093] In certain aspects, the post-edit homology region of the repair template is between 2 nucleotides and 20 nucleotides in length, between 2 nucleotides and 15 nucleotides in length, between 2 nucleotides and 50 nucleotides in length, between 4 nucleotides and 40
30 nucleotides in length, between 3 nucleotides and 30 nucleotides in length, or between 5 nucleotides and 25 nucleotides in length.

[0094] In certain aspects, the editing cassette is designed to facilitate incorporation of an intended edit at a first target locus (*e.g.*, target site or target region) of the cells, and the barcoding cassette is designed to facilitate integration of a barcode at a second target locus of the cells different from the first target locus.

5 [0095] In certain aspects, the editing cassette and/or the barcoding cassette (*e.g.*, the repair templates) further comprise an edit (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 edits) to immunize a target locus to prevent re-nicking or re-cutting thereof. As discussed herein, in some aspects, an edit to immunize a target locus to prevent re-nicking is one that alters the proto-spacer adjacent motif (PAM) (or other element) such that subsequent binding at the target
10 locus by the nucleic acid-guided polypeptide (*e.g.*, nuclease, nickase, inactive nuclease or inactive nickase) is impaired or prevented.

[0096] In certain aspects, the editing cassette and/or the barcoding cassette further comprise an RNA G-quadruplex region at a 3' end of the repair template to stabilize the cassette and improve target nicking or cleavage efficiency without inducing off-target
15 activity.

[0097] In certain aspects, the editing cassette and/or barcoding cassette further comprise an amplification priming site or subpool primer binding sequence at a 3' end thereof. In specific aspects, the editing cassette and/or barcoding cassette further comprise a melting temperature booster sequence at a 5' end thereof, which is a short protective DNA
20 buffer sequence. In addition, in specific aspects, the editing cassette and/or barcoding cassette comprise regions of homology to a vector for gap-repair insertion of the cassette into the vector, such as an editing vector or engine vector.

[0098] In some aspects, a region of complementarity between the barcoding gRNA and a target locus is between 4 nucleotides and 120 nucleotides in length, between 5 nucleotides
25 and 80 nucleotides in length, or between 6 nucleotides and 60 nucleotides in length. In certain aspects, a region of complementarity between the barcoding gRNA and a target locus is between 1 nucleotide and 10 nucleotides in length, between 10 nucleotides and 20 nucleotides in length, between 20 nucleotides and 50 nucleotides in length, or between 50 nucleotides and 100 nucleotides in length.

30 [0099] In some aspects, the barcode sequence of the repair template of the barcoding cassette is between 1 nucleotide and 750 nucleotides in length, between 1 nucleotide and

500 nucleotides in length, between 1 nucleotide and 150 nucleotides in length, between 1 nucleotide and 100 nucleotides in length, between 4 nucleotides and 50 nucleotides in length, or between 4 nucleotides and 25 nucleotides in length.

[00100] In some aspects, a region of complementarity between the editing gRNA and a target locus is between 4 nucleotides and 120 nucleotides in length, between 5 nucleotides and 80 nucleotides in length, between 6 nucleotides and 60 nucleotides in length, between 1 nucleotide and 10 nucleotides in length, between 10 nucleotides and 20 nucleotides in length, between 20 nucleotides and 50 nucleotides in length, or between 50 nucleotides and 100 nucleotides in length.

10 [00101] In some aspects, the edit region of the repair template of the editing cassette is between 1 nucleotide and 750 nucleotides in length, between 1 nucleotide and 500 nucleotides in length, between 1 nucleotide and 150 nucleotides in length, between 1 nucleotide and 10 nucleotides in length, between 10 nucleotides and 20 nucleotides in length, between 20 nucleotides and 50 nucleotides in length, between 50 nucleotides and 15 100 nucleotides in length, between 100 nucleotides and 250 nucleotides in length, between 250 nucleotides and 500 nucleotides in length, or between 500 nucleotides and 750 nucleotides in length.

[00102] In certain aspects, the edit region of the repair template of the editing cassette comprises two or more edits, or three or more edits, or four or more edits, or five or more 20 edits.

[00103] In some aspects, the edit created by the editing cassette in a target locus includes one or more nucleotide swaps in the target locus.

[00104] In some aspects, the edit created by the editing cassette in a target locus is an insertion in the target locus.

25 [00105] In some aspects, the edit created by the editing cassette is an insertion of recombinase sites, protein degron tags, promoters, terminators, alternative-splice sites, CpG islands, etc.

[00106] In some aspects, the edit created by the editing cassette in a target locus is a deletion in the target locus.

30 [00107] In some aspects, the editing cassette is designed to provide a deletion of between 1 nucleotide and 750 nucleotides at a target locus. In some aspects, the editing

cassette is designed to provide a deletion of between 1 nucleotide and 10 nucleotides, between 10 nucleotides and 20 nucleotides, between 20 nucleotides and 50 nucleotides, between 50 nucleotides and 100 nucleotides, between 100 nucleotides and 200 nucleotides, between 200 nucleotides and 500 nucleotides, or between 250 nucleotides and 750 nucleotides at a target locus.

[00108] In some aspects, the edit created is a deletion of introns, exons, repetitive elements, promoters, terminators, insulators, CpG islands, non-coding elements, retrotransposons, etc.

[00109] In some aspects, the edit comprises several types of edits and/or comprises more than one of one or more types of edits. For example, in some aspects, the edit comprises two or more nucleotide swaps or substitutions (*e.g.*, 2, 3, 4, 5, or between 1 and 20 nucleotide swaps), some or all of which can be adjacent to each other or nonadjacent to each other. In some aspects, the edit comprises one or more nucleotide swaps (*e.g.*, 2, 3, 4, 5, or between 1 and 20 nucleotide swaps) and an insertion of one or more nucleotides (*e.g.*, 2, 3, 4, 5, or between 1 and 20 nucleotides). In some aspects, the edit comprises one or more nucleotide swaps (*e.g.*, 2, 3, 4, 5, or between 1 and 20 nucleotide swaps) and a deletion of one or more nucleotides (*e.g.*, 2, 3, 4, 5, or between 1 and 20 nucleotides).

[00110] In some aspects, the edit created by the editing cassette in a target locus is in a coding region in the target locus.

[00111] In some aspects, the edit created by the editing cassette in a target locus is in a noncoding region in the target locus.

[00112] In some aspects, a barcode sequence integrated at a second target locus facilitates tracking of an incorporated edit at a first target locus. In some aspects, the second target locus comprises a neutral integration site, or “safe spot,” that facilitates stable integration of the barcode sequence without significant impact on cell growth or function. In some aspects, the second target locus is a safe harbor locus disposed centrally in a large intergenic region to reduce the potential of barcode sequence integration adversely affecting genes neighboring the integrated barcode. In some aspects, where a plurality of barcodes are integrated (*e.g.*, during recursive or iterative editing methods), the barcodes are embedded into one or more clustered neutral safe harbor loci. In some aspects, the

integration locus of the barcode sequence is elected based on inspection of a host GFP fusion localization database.

[00113] In some aspects, the second target locus is disposed within a coding region (*e.g.*, exon). In some aspects, the second target locus is disposed within a noncoding region (*e.g.*, intronic or intergenic region). In some aspects, the second target locus comprises the adeno-associated virus site 1 (“AAVS1”), the chemokine (C-C motif) receptor 5 (“CCR5”) gene, the DNA methyltransferase 3B (“DMNT3b”) gene, the eukaryotic translation initiation factor 4E-binding protein 2 (“4EBP2”) gene, the ornithine decarboxylase antizyme 1 (“OAZ1”) gene, or an orthologue of the Rosa26 locus.

10 [00114] In some aspects, the second target locus is adjacent to the first target locus, or within close proximity to the first target locus. As used herein, “close proximity” refers to within 5000 nucleotides.

[00115] In some aspects, successfully barcoded and/or edited cells may be enriched for based on the selection of the second target locus. For example, in specific aspects, the second target locus may be disposed within the coding region of a non-essential cell surface receptor such that integration of a barcode sequence in the second target locus eliminates the receptor from the cell surface. Accordingly, antibody and/or bead-based affinity purification may be utilized to remove cells that were not successfully barcoded, leaving only barcoded cells (*e.g.*, negative selection). In some aspects, the barcode sequence may comprise a frameshifting edit. In some aspects, the barcode sequence may comprise a frameshifting edit and be 8 or more nucleotides in length, wherein the number of nucleotides is not a multiple of 3 (*e.g.*, 10, 11, 13, 14, etc.). In some aspects, the barcode sequence may comprise an in-frame STOP codon (TAGTGA) edit and be 9 or more nucleotides in length.

25 [00116] In some aspects, the second target locus may be disposed within the coding region of the CD9 cell surface glycoprotein, CD81 cell surface receptor, CD63 cell surface receptor, or other non-essential cell surface receptor.

[00117] In specific aspects, the second target locus may be disposed within a locus corresponding to the 5' and/or 3' untranslated region (UTR) of a gene, including CD81, OAZ2, CD9, CD63, SGK1, CARHSP1, MAP4, SLC38A1, WNK1, DIAPH1, LRRC8A, 30 FAF2, NKTR, TBC1D16, GJC1, NUCKS1, CAPZB, TBC1D16, MPP6, WDR83OS,

PMEPA1, SERINC5, HTT, SLC29A1, PPP3CA, EZR, HEBP2, HTT, SLC7A1, LSM14A, ERBB2, CYP51A1, GPATCH8, and/or the like. In some aspects, the second target locus may be disposed within a locus corresponding to a 5' and/or 3' UTR of an mRNA, so that the integrated barcode may be detected by RNA sequencing methods.

5 [00118] In some aspects, the integrated barcode sequence is tracked or analyzed via RNA sequencing (*e.g.*, transcriptome sequencing) or genomic sequencing. In some aspects, the integrated barcode sequence is tracked or analyzed via single cell whole genome sequencing methods, which enable combinatorial and/or linked edit tracking.

[00119] In some aspects, the nuclease includes a MAD-series nuclease, nickase, or a variant (*e.g.*, orthologue) thereof. In some aspects, the nuclease includes a MAD1, 10 MAD2, MAD3, MAD4, MAD5, MAD6, MAD7, MAD8, MAD9, MAD10, MAD11, MAD12, MAD13, MAD14, MAD15, MAD16, MAD17, MAD18, MAD19, MAD20, MAD2001, MAD2007, MAD2008, MAD2009, MAD2011, MAD2017, MAD2019, MAD297, MAD298, MAD299, or other MAD-series nuclease, nickase, variants thereof, 15 and/or combinations thereof.

[00120] In some aspects, the nuclease includes a Cas9 nuclease (also known as Csn1 and Csx12), nickase, or a variant thereof.

[00121] In some aspects, the nuclease includes C2c1, C2c2, C2c3, Cas1, Cas1B, Cas2, 20 Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas10, Cpf1, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx100, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, or similar nuclease, nickase, variants thereof, and/or combinations thereof.

[00122] In some aspects, such as aspects wherein a CF barcoding cassette and/or CF 25 editing cassette is utilized, the nuclease is a fusion protein—*e.g.*, a nucleic acid-guided nickase/reverse transcriptase fusion enzyme (a “nickase-RT fusion”)—that retains certain characteristics of nucleic acid-directed nucleases (*e.g.*, the binding specificity and ability to cleave one or more DNA strands in a targeted manner) combined with another enzymatic activity, namely, reverse transcriptase activity. In some aspects, the reverse transcriptase portion of the nickase-RT fusion may use a CFgRNA (*e.g.*, the barcoding gRNA in a CF 30 barcoding cassette or editing gRNA in a CF editing cassette) to synthesize and edit at a “flap” created by the nickase portion on one or both DNA strands of a target locus, thereby

circumventing the endogenous mismatch repair systems to integrate a barcode or incorporate an edit.

[00123] In some aspects, the barcoding cassette, the editing cassette, and the nuclease are introduced into the cells on a single vector (*e.g.*, a single-part system). In certain aspects, the barcoding cassette, the editing cassette, and/or the nuclease are introduced into the cells as a multi-part system, wherein the barcoding cassette may be introduced separately from the editing cassette and/or the nuclease. In some aspects, the barcoding cassette may be comprised on a first vector, and the editing cassette and/or the nuclease may be comprised on a second vector co-delivered with the first vector. In some aspects, the nuclease is introduced to the cells before the introduction of the barcoding cassette and the editing cassette. In some aspects, one or more of the barcoding cassette, the editing cassette, and the nuclease are introduced into the cells prior to the introduction of the remaining components for effecting editing and/or barcoding. In some aspects, the cell comprises the barcoding cassette, the editing cassette, and the nuclease. In some aspects, the cell comprises the barcoding cassette and the editing cassette. In some aspects, the cell comprises the nuclease.

[00124] In some aspects, the nuclease is introduced into the cells as a DNA molecule coding for the nuclease separately or linked to the barcoding cassette and/or the editing cassette. In some aspects, the nuclease may be introduced separately into the cells in protein form or as part of a complex. In some aspects, the same nuclease may be utilized to incorporate the intended edit into a first genomic locus of the cells and the barcode into a second genomic locus of the cells. In some aspects, different nucleases may be utilized to incorporate the intended edit into a first genomic locus of the cells and the barcode into a second genomic locus of the cells. In some aspects, the nuclease is endogenously expressed in the cells. In some aspects, the nuclease is transiently expressed in the cells. In some aspects, the nuclease is delivered into the cells as a protein molecule. In some aspects, the nuclease is delivered into the cells as a complex with nucleic acid.

[00125] In some aspects, the barcoding cassette, the editing cassette, and/or the nucleic acid-guided nuclease are introduced into the cell on a linear or circular plasmid. In some aspects, the barcoding cassette, the editing cassette, and/or the nucleic acid-guided nuclease are under the control of a constitutive or inducible promoter at a 5' end thereof.

[00126] In some aspects, a vector comprising the barcoding cassette, the editing cassette, and/or the nucleic acid-guided nuclease further comprises an origin of replication and a selectable marker component, *e.g.*, an antibiotic resistance gene or a fluorescent protein gene, for selection or enrichment of cells that have been edited and/or barcoded. In some aspects, the selectable marker may be utilized for selective enrichment of edited and/or barcoded cells. In some aspects, the selectable marker comprises an antibiotic resistance gene or a fluorescent protein. In some aspects, the selectable marker comprises the PuroR gene.

[00127] In some aspects, there is provided a library of vector or plasmid backbones, and/or a library of editing cassettes, and/or a library of barcoding cassettes to be transformed into cells. In some aspects, the utilization of a library of cassettes and/or a library of vector or plasmid backbones enables combinatorial or multiplex editing in the cells. In some aspects, a library of cassettes or vectors may comprise cassettes or vectors that have any combination of common elements and non-common or different elements as compared to other cassettes or vectors within the pool. In some aspects, a library of editing cassettes may comprise common priming sites or common nick-to-edit or post-edit homology regions, while also containing non-common or unique edits. In some aspects, a library of barcoding cassettes may comprise common priming sites or common nick-to-barcode or post-barcode homology regions, while also containing non-common or unique barcode sequences. In some aspects, combinations of common and non-common elements are advantageous for multiplexing or combinatorial techniques disclosed herein.

[00128] In some aspects, a library of cassettes comprises at least 2 cassettes, at least 10 cassettes, at least 100 cassettes, at least 500 cassettes, at least 1,000 cassettes, at least 5,000 cassettes, at least 10,000 cassettes, at least 100,000 cassettes, or at least 1,000,000 cassettes. In some aspects, a library of cassettes comprises between 5 cassettes and 1,000,000 cassettes, between 100 cassettes and 500,000 cassettes, between 1,000 cassettes and 100,000 cassettes, between 1,000 cassettes and 10,000 cassettes, or between 10,000 cassettes and 50,000 cassettes.

[00129] In some aspects, one or more editing cassettes in a library of editing cassettes each comprise a different editing gRNA targeting a different target locus within the cell genome. In some aspects, one or more editing cassettes in a library of editing cassettes each

comprise a different edit to be incorporated within the cell genome. In some aspects, one or more barcoding cassettes in the library of barcoding cassettes each comprise a different barcoding gRNA targeting a different target locus within the cell genome. In some aspects, one or more barcoding cassettes in a library of barcoding cassettes each comprise a
5 different barcode to be incorporated within the cell genome.

[00130] In some aspects, there is provided a trackable library comprising a plurality of cassettes or a plurality of vectors comprising a cassettes as disclosed herein. In some aspects, within the trackable library are distinct editing cassette and barcoding cassette combinations, which when sequenced upon editing, facilitate tracking of editing events in
10 a population of cells. Accordingly, when edits and barcodes are incorporated into a target genome, the incorporation of an edit is determined based on sequenced the barcode.

[00131] In some aspects, there is provided a gene-wide or genome-wide library of cassettes or vectors comprising cassettes as disclosed herein.

[00132] In some aspects, there are provided methods of recursive or iterative rounds of
15 editing operations. In some aspects, during each round of editing, a new or unique barcode is incorporated into the cell genome, such that following multiple editing rounds to construct combinatorial diversity throughout the genome, sequencing of the barcodes can be used to reconstruct each combinatorial genotype or to confirm that the edit from each round or operation has been incorporated into the genome. In some aspects, methods
20 disclosed herein comprise 2 or more rounds of editing, 3 or more rounds of editing, 5 or more rounds of editing, 7 or more rounds of editing, or 10 or more rounds of editing. In some aspects, methods disclosed herein comprises one round of editing.

[00133] In some aspects, one or more unique barcodes can be inserted in each round of multiple iterative or recursive editing operations. In some aspects, the unique barcodes may
25 be inserted adjacent or in proximity to each other (*e.g.*, in a single target region), or at a distance and/or in separate target regions.

[00134] In some aspects, recursive or iterative editing methods may be used. In some aspects, recursive or iterative editing methods may be used for analyzing combinatorial mutational effects on large populations, or for inserting entire pathways within cells.

[00135] In some aspects, the methods described herein facilitate parallel analysis of two
30 or more target proteins.

[00136] In some aspects, the methods described herein enable preparation of a comprehensive library of genetic variations encompassing all residue changes of one or more target proteins, such as one or more target proteins that contributed to a trait.

[00137] The present disclosure provides, in selected aspects, modules, instruments, and systems for automated multi-module cell processing for trackable nucleic acid-guided genome editing in multiple cells. Automated systems for cell processing that may be used for can be found, *e.g.*, in U.S. Pat. Nos. USPNs 10,253,316; 10,329,559; 10,323,242; 10,421,959; 10,465,185; 10,519,437; 10,584,333; 10,584,334; 10,647,982; 10,689,645; 10,738,301; and 10,738,663.

[00138] In some aspects, the automated multi-module cell processing instruments of the present disclosure are designed for recursive genome editing, *e.g.*, sequentially introducing multiple edits into genomes inside one or more cells of a cell population through two or more editing operations within the instruments.

[00139] In some aspects, the methods, compositions, modules, and instruments described herein may be utilized for efficient tracking of barcodes utilized during editing, for efficient tracking of ribonucleoprotein (RNP) based transfections, and for efficient tracking of non-plasmid based barcode delivery via homologous recombination (HR) or non-homologous end joining (NHEJ) based integration.

Nucleic Acid-Guided Nickase/Reverse Transcriptase Fusion Protein Genome Editing, Generally

[00140] Certain aspects described herein provide an alternative to traditional nucleic acid-guided nuclease editing (*e.g.*, RNA-guided nuclease or CRISPR editing) used to introduce desired edits to a population of cells; that is, the compositions and methods described herein may employ a nucleic acid-guided nickase/reverse transcriptase fusion enzyme (“nickase-RT fusion”) as opposed to a nucleic acid-guided nuclease (*e.g.*, a “CRISPR nuclease”). The nickase-RT fusion employed herein differs from traditional CRISPR editing in that instead of initiating double-stranded breaks in the target genome and homologous recombination to effect an intended edit and/or integrate a barcode corresponding with an intended edit, the nickase initiates a nick in a single strand of the target genome, *e.g.*, the non-complementary strand. Further, the fusion of the nickase to a

reverse transcriptase, in combination with an editing or barcoding cassette comprising a gRNA and repair template, eliminates the need for a donor DNA to be incorporated by homologous recombination. Instead, a nucleic acid sequence encoding the repair template of the corresponding cassette—typically a ribonucleic acid—may serve as a template for the reverse transcription (“RT”) portion of the fusion enzyme to add an intended edit or barcode to the nicked strand at the target locus. That is, utilization of a nickase-RT fusion enables incorporation of the edit or barcode in the target genome by copying an RNA sequence (*e.g.*, at the RNA level) rather than replacing a portion of the target locus with a donor DNA (*e.g.*, at the DNA level).

10 [00141] The nickase—functioning as a single-strand cutter and having the specificity of a nucleic acid-guided nuclease—engages the target locus and nicks a strand of the target locus creating one or more free 3' terminal nucleotides. The 3' end of the repair template encoded by the editing cassette or barcoding cassette is then annealed to the nicked strand, and the reverse transcriptase utilizes the 3' terminal nucleotide(s) of the nicked strand to copy the repair template and create a “flap” containing the desired edit or barcode. Thereafter, endogenous repair mechanisms of the cells repair the nick in favor of the desired edit by hybridizing the flap to the wild-type (*e.g.*, unedited) DNA strand. In summary, in certain aspects, the present methods and compositions are drawn to using the nickase-RT fusion to nick a strand of DNA at the target locus and, using an editing cassette or barcoding cassette, to effect the desired edit or barcode on the strand via the reverse transcriptase portion of the nickase-RT fusion.

15 [00142] Generally, nucleic acid-guided nuclease editing typically begins with a nucleic acid-guided nuclease complexing with an appropriate guide nucleic acid in a cell which can cut the genome of the cell at a desired location. The guide nucleic acid helps the nucleic acid-guided nuclease recognize and cut the DNA at a specific target sequence. By manipulating the nucleotide sequence of the guide nucleic acid, the nucleic acid-guided nuclease may be programmed to target any DNA sequence for cleavage as long as an appropriate protospacer adjacent motif (PAM) is nearby. For some nucleic acid-guided nucleases, two separate guide nucleic acid molecules that combine to function as a guide nucleic acid are used, *e.g.*, a CRISPR RNA (crRNA) and trans-activating CRISPR RNA

20

25

30

(tracrRNA). For other nucleic acid-guided nucleases, the guide nucleic acid may be a single guide nucleic acid that includes both the crRNA and tracrRNA sequences.

[00143] In general, a guide nucleic acid (*e.g.*, gRNA or CFgRNA) complexes with a compatible nucleic acid-guided nuclease and can then hybridize with a target sequence, thereby directing the nuclease to the target sequence. A guide nucleic acid can be DNA or RNA; alternatively, a guide nucleic acid may comprise both DNA and RNA. In some aspects, a guide nucleic acid may comprise modified or non-naturally occurring nucleotides. In the present methods and compositions, the guide nucleic acid is RNA.

[00144] A guide nucleic acid comprises a guide sequence, where the guide sequence (as opposed to the scaffold sequence portion of the guide nucleic acid) is a polynucleotide sequence having sufficient complementarity with a target sequence to hybridize with the target sequence and direct sequence-specific binding of a complexed nucleic acid-guided nuclease to the target sequence. The degree of complementarity between a guide sequence and the corresponding target sequence, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 92.5%, 95%, 97.5%, 99%, or more. Optimal alignment may be determined with the use of any suitable algorithm for aligning sequences (*e.g.*, without being limiting, BLAST™). In some aspects, a guide sequence is about or more than about 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75, or more nucleotides in length. In some aspects, a guide sequence is less than about 75, 50, 45, 40, 35, 30, 25, 20 nucleotides in length. Preferably the guide sequence is between 10 nucleotides and 30 nucleotides long, between 15 nucleotides and 20 nucleotides long, or 15, 16, 17, 18, 19, or 20 nucleotides in length.

[00145] In some aspects of the present methods and compositions, the guide nucleic acids are provided as mRNAs or sequences to be expressed from a plasmid or vector, and/or as sequences to be expressed from a cassette optionally inserted into a plasmid or vector, and comprise both the guide sequence and the scaffold sequence as a single transcript. The guide nucleic acids are engineered to target a desired target sequence by altering the guide sequence so that the guide sequence is complementary to a desired target sequence, thereby allowing hybridization between the guide sequence and the target sequence. In general, to generate an edit or integrate a barcode in the target sequence, the gRNA/nuclease complex

binds to a target sequence as determined by the guide RNA, and the nuclease recognizes a protospacer adjacent motif (PAM) sequence adjacent to the target sequence. The target sequence can be any polynucleotide endogenous or exogenous to a prokaryotic or eukaryotic cell, or *in vitro*. For example, the target sequence can be a polynucleotide residing in the nucleus of a eukaryotic cell. A target sequence can be a sequence encoding a gene product (*e.g.*, a protein) or a non-coding sequence (*e.g.*, a regulatory polynucleotide, an intron, a PAM, or “junk” DNA).

[00146] As described above, in certain aspects, the guide nucleic acids may be part of editing cassettes or barcoding cassettes that also encode for repair templates, which are used as templates for reverse transcription by the reverse transcriptase portion of the nickase-RT fusion. Each repair template generally comprises a desired edit or barcode corresponding with a desired edit to be incorporated into the target DNA sequence. Accordingly, the edit or barcode is integrated into the target DNA sequence via copying of the repair template by the nickase-RT fusion, therefore not depending on HDR mechanisms between the target genome and a donor nucleic acid to effect the edit or barcode.

[00147] The target sequence is associated with a proto-spacer adjacent motif (PAM), which is a short nucleotide sequence recognized by the gRNA/nuclease complex. The precise preferred PAM sequence and length requirements for different nucleic acid-guided nucleases vary; however, PAMs typically are 2-8 base-pair sequences adjacent or in proximity to the target sequence and, depending on the nuclease, can be 5' or 3' to the target sequence. Engineering of the PAM-interacting domain of a nucleic acid-guided nuclease may allow for alteration of PAM specificity, improve target site recognition fidelity, decrease target site recognition fidelity, or increase the versatility of a nucleic acid-guided nuclease.

[00148] In certain aspects, the editing or barcoding of a cellular target sequence both introduces a desired DNA change to the cellular target sequence, *e.g.*, the genomic DNA of a cell, and removes, mutates, or renders inactive a PAM region or spacer region in the cellular target sequence. Rendering the PAM at the cellular target sequence inactive precludes additional editing of the cell genome at that cellular target sequence, *e.g.*, upon subsequent exposure to a nucleic acid-guided nuclease complexed with a synthetic guide nucleic acid in later rounds of editing.

[00149] The range of target sequences that nucleic acid-guided nucleases can recognize is constrained by the need for a specific PAM to be located near the desired target sequence. As a result, it often can be difficult to target edits with the precision that is necessary for genome editing. It has been found that nucleases can recognize some PAMs very well (*e.g.*, canonical PAMs), and other PAMs less well or poorly (*e.g.*, non-canonical PAMs).

[00150] As for the nuclease or nickase-RT fusion component of the nucleic acid-guided nuclease editing system, a polynucleotide sequence encoding the nucleic acid-guided nuclease or nickase-RT fusion can be codon optimized for expression in particular cell types, such as archaeal, prokaryotic or eukaryotic cells. Eukaryotic cells can be yeast, fungi, algae, plant, animal, or human cells. Eukaryotic cells may be those of or derived from a particular organism, such as a mammal, including but not limited to human, mouse, rat, rabbit, dog, or non-human mammals including non-human primates. The choice of nucleic acid-guided nuclease or nickase-RT fusion to be employed depends on many factors, such as what type of edit is to be made in the target sequence and whether an appropriate PAM is located close to the desired target sequence.

[00151] Nucleases of use in the methods described herein include but are not limited to nickases engineered from nucleic acid-guided nucleases such as Cas9, Cas 12/Cpf1, MAD2, MAD2007, MAD2017, MAD2019, MAD297, MAD298, MAD299, MAD7®, or other MADZYME®, variants thereof, and nuclease or nickase fusions thereof. Nickase-RT fusion enzymes typically comprise one or more CRISPR nucleic acid-guided nucleases, each engineered to nick one DNA strand in the target DNA rather than making a double-stranded cut, and the nickase portion(s) are fused to a reverse transcriptase. In certain aspects of the present methods, the nickase-RT fusion nicks both strands of the target locus, albeit where the two nicks are staggered rather than at the same position which would result in a double-stranded cut. As with the guide nucleic acid, the nucleases or nickases may be encoded by one or more DNA sequences on a vector (*e.g.*, an engine vector or an editing or barcoding vector also comprising the editing and/or barcoding cassette) and be under the control of a promoter—including inducible or constitutive promoters—or the nickase-RT fusion may be delivered as a protein or RNA-protein complex.

[00152] In addition to a nucleic acid sequence encoding the gRNA and a nucleic acid sequence encoding the repair template, an editing cassette, barcoding cassette, or editing

vector backbone may comprise one or more primer sites. The primer sites can be used to amplify the cassette or editing vector backbone by using oligonucleotide primers; for example, if the primer sites flank one or more of the other components of the cassette or editing vector backbone, *e.g.*, the nucleic acid sequence encoding the gRNA and/or the nucleic acid sequence encoding the repair template.

[00153] Additionally, in some aspects, a vector encoding the nickase-RT fusion enzyme and/or the CF editing cassette further encodes one or more nuclear localization sequences (NLSs), such as about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs. In some aspects, the engineered nuclease comprises NLSs at or near the amino-terminus, NLSs at or near the carboxy-terminus, or a combination.

Improved Nucleic Acid-Guided Nickase/Reverse Transcriptase Fusion Editing and Tracking of Edits

[00154] Creating a library of genomic edits requires tracking (*e.g.*, identification) of editing events. Traditionally, in order to track editing events during one or more rounds of nucleic acid-guided nuclease editing, lentivector-based barcodes or episomal components are introduced into the host cells along with the editing guide nucleic acids, donor DNA, and/or nucleases for integration into the cell genomes. However, random integration of lentivector-based systems may adversely affect phenotype-genotype reagents, and episomes are inefficient and have low establishment rates, leading to a loss in library diversity. The present disclosure addresses the deficiencies of these and other trackable integration technologies.

[00155] In particular, aspects of the present disclosure provide compositions of matter, methods and instruments for nucleic acid-guided nickase/reverse transcriptase fusion (“nickase-RT fusion”) editing of live cells using editing and barcoding cassettes, *e.g.*, CREATE fusion editing and barcoding cassettes, each comprising a gRNA covalently linked to a repair template comprising an intended edit or barcode. The editing cassettes and barcoding cassettes are engineered to edit and barcode genomic DNA, respectively, at separate target loci, wherein each barcode may correspond with one or more editing events. Thus, nickase-RT fusion editing events may be tracked on a one-to-one basis utilizing single cell genomic DNA or RNA sequencing methods to identify integrated barcodes. In

such examples, each integrated barcode may serve as a proxy for one or more corresponding edits.

[00156] Utilizing the compositions and methods described herein, a single nickase-RT fusion enzyme may be used to facilitate the incorporation of a desired edit into a cell genome at a first target locus, as well as the integration of a barcode sequence into the genome at a second target locus. Further, a single barcode integration locus, *e.g.*, a safe harbor locus, intronic region, or non-essential gene exon, once optimized, may enable consistent integration of multiple trackable barcodes, thereby facilitating tracking of multiple editing events that may target many different genomic loci via sequencing of the single barcode integration locus. Accordingly, sequencing of a single locus enables relative quantitation of design diversity and abundance without having to directly sequence the exact targeted locus or construct used in an edit-driving library.

[00157] FIG. 1A is a simplified block diagram of an example of a method 100 for editing live cells via nucleic acid-guided nickase/reverse transcriptase fusion (“nickase-RT fusion”) editing and for tracking the editing events. In particular, the exemplary method 100 utilizes a nickase-RT fusion enzyme in combination with a CREATE fusion editing cassette (“CF editing cassette”) and a CREATE fusion barcoding cassette (“CF barcoding cassette”), as described above, to effect an intended edit in the cell genome and further integrate a barcode corresponding with the intended edit.

[00158] Looking at FIG. 1A, method 100 begins at 102 by designing and synthesizing CF editing cassettes, such as a library of CF editing cassettes, each comprising a covalently-linked editing CFgRNA and repair template designed to incorporate an edit into one or both DNA strands at a first target locus. That is, each CF editing cassette encodes an editing CFgRNA sequence and a repair template sequence to be reverse transcribed comprising a desired genome edit, as well as a PAM and/or spacer mutation(s). Once the CF editing cassettes have been synthesized, the individual cassettes may be amplified.

[00159] Similarly, at 104, CF barcoding cassettes (*e.g.*, a library of CF barcoding cassettes) are designed and synthesized, each cassette comprising a covalently-linked barcoding CFgRNA and repair template designed to incorporate a unique barcode into one or both DNA strands at a second target locus, which may be different from the first target locus. In other words, each CF barcoding cassette encodes a barcoding CFgRNA sequence

and a repair template sequence to be reverse transcribed comprising a barcode sequence corresponding with an edit carried by the CF editing cassettes, as well as a PAM and/or spacer mutation(s). Once the CF barcoding cassettes have been synthesized, the individual cassettes may be amplified.

5 [00160] In certain aspects, the second target locus comprises a neutral integration site, or “safe spot,” that facilitates stable integration of the barcode sequence without significant impact on cell growth or function. In certain aspects, the second target locus is a safe harbor locus disposed centrally in a large intergenic or intronic region to reduce the potential of barcode sequence integration adversely affecting genes neighboring the integrated barcode.

10 In certain aspects, the integration locus of the barcode sequence is elected based on inspection of a host GFP fusion localization database. In further aspects, the second target locus is adjacent to the first target locus, or within close proximity to the first target locus.

[00161] At 106, a nickase-RT fusion enzyme is designed. As described above, the nickase-RT fusion enzyme comprises, in order from amino terminus to carboxy terminus, 15 or from carboxy terminus to amino terminus, a nucleic acid-guided nickase and a reverse transcriptase. The nickase-RT fusion enzyme may be delivered to the cells as a coding sequence in a vector (in some aspects under the control of an inducible promoter), such as the same or different vector as the CF editing cassette and/or CF barcoding cassette, or the nickase-RT fusion enzyme may be delivered to the cells as a protein or protein complex.

20 In method 100, the nickase-RT fusion enzyme is delivered to the cells via a coding sequence in an editing vector further comprising at least the CF editing cassette.

[00162] At 108, the CF editing cassettes, the CF barcoding cassettes, and/or the nickase-RT fusion enzymes are assembled with vector backbones, such as plasmid backbones, to create editing vectors, *e.g.*, a library of editing vectors. In certain aspects, a CF editing 25 cassette, CF barcoding cassette, and nickase-RT fusion enzyme are assembled together on a single editing vector. An example of an editing vector comprising all the aforementioned components is illustrated in FIG 1D. In certain other aspects, however, the CF editing cassette and nickase-RT fusion enzyme are assembled into an editing vector together, and the CF barcoding cassette is assembled into a separate barcoding vector. In still other 30 aspects, the CF barcoding cassette and nickase-RT fusion enzyme are assembled into a vector together, and the CF editing cassette is assembled into a separate vector.

[00163] At 110, the engine and editing vectors are introduced into the live cells. A variety of delivery systems may be used to introduce (*e.g.*, transform, transfect, or transduce) nucleic acid-guided nickase fusion editing system components into a host cell 110. These delivery systems include the use of yeast systems, lipofection systems, 5 microinjection systems, biolistic systems, virosomes, liposomes, immunoliposomes, polycations, lipid:nucleic acid conjugates, virions, artificial virions, viral vectors, electroporation, cell permeable peptides, nanoparticles, nanowires, exosomes. Alternatively, molecular trojan horse liposomes may be used to deliver nucleic acid-guided nuclease components across the blood brain barrier. Of particular interest is the use of 10 electroporation, particularly flow-through electroporation (either as a stand-alone instrument or as a module in an automated multi-module system) as described in, *e.g.*, USPNs 10,253,316, issued 09 April 2019; 10,329,559, issued 25 June 2019; 10,323,242, issued 18 June 2019; 10,421,959, issued 24 September 2019; 10,465,185, issued 05 November 2019; 10,519,437, issued 31 December 2019; 10,584,333, issued 10 March 15 2020; 10,584,334, issued 10 March 2020; 10,647,982, issued 12 May 2020; 10,689,645, issued 23 June 2020; 10,738,301, issued 11 August 2020; 10,738,663, issued 29 September 2020; and 10,894,958, issued 19 January 2021.

[00164] Once transformed 110, the next steps in method 100 include providing conditions for nucleic acid-guided nuclease editing 112 and for barcoding 114. "Providing 20 conditions" includes incubation of the cells in appropriate medium and may also include providing conditions to induce transcription of an inducible promoter (*e.g.*, adding antibiotics, adding inducers, increasing temperature) for transcription of a CF editing cassette, CF barcoding cassette, and/or nickase-RT fusion enzyme. In certain aspects, the conditions for editing 112 and for genomic integration of the barcode 114 are the same and 25 thus, these steps are performed simultaneously. In certain aspects, the conditions for editing 112 and for genomic integration of the barcode 114 are different (*e.g.*, the barcoding CFgRNA of the CF barcoding cassette may be under the control of a different inducible promoter than other components of the editing system), and these steps may be performed either simultaneously or in sequence.

30 [00165] Once editing and barcoding is complete, the cells are allowed to recover and are preferably enriched for cells that have been edited and/or cells in which the barcode

has integrated into the genome 116. Enrichment can be performed directly, such as via cells from the population that express a selectable marker, or by using surrogates, *e.g.*, cell surface handles co-introduced with one or more components of the editing components. At this point in method 100, the cells can be characterized phenotypically or genotypically or, optionally, steps 102 to 114 or steps 110 to 114 may be repeated to make additional trackable edits 118 in recursive or iterative editing rounds. In certain aspects, steps 102 to 114 are repeated to create or construct a defined combination of edits or a combinatorial library.

[00166] After recovery and enrichment of edited cells, the genomic DNA or RNA transcripts of the cells may be sequenced to track or analyze the editing events 120, wherein the integrated barcode(s) serve as accurate proxies for corresponding edits. For example, the cells may be lysed and DNA or RNA extracted, purified, amplified, prepared into libraries, and sequenced to track for integrated barcodes. A simplified graphic depiction of step 120 is depicted in FIG. 1B, wherein a library of edits is sequenced to track editing events. In certain aspects, amplicons of genomic DNA are sequenced via any suitable high-throughput method, such as single molecule real time (SMRT) sequencing, nanopore sequencing, sequencing by synthesis (SBS) or Illumina sequencing, Ion Torrent sequencing, sequencing by ligation (SBL), combinatorial probe anchor synthesis (cPAS) sequencing, parallel pyrosequencing, microfluidic methods, etc. In certain aspects, the transcriptome of the cells is sequencing via any suitable high-throughput RNA sequencing (RNA-Seq) method, such as bulk or scRNA-Seq.

[00167] FIG. 1C is a simplified graphic depiction of the mechanism of a nucleic acid-guided nickase enzyme/reverse transcriptase fusion enzyme edit. At left in FIG. 1C, a nickase-RT fusion enzyme and CFgRNA of a CF editing cassette (or CF barcoding cassette) are shown bound to a target locus of the cell genome, where the target locus in the context of the methods and compositions herein is a locus of approximately 8 to 500 nucleotides in length, or 10 to 400 nucleotides in length, or 10 to 300 nucleotides in length. In one step, the nickase-RT fusion enzyme and the CFgRNA bind to the target locus and the nickase nicks a single DNA strand at the target locus, thus creating a 3' "flap." In order for the nickase-RT fusion enzyme and the CFgRNA to bind to the target locus and nick the genomic DNA, there must be a protospacer adjacent motif (PAM) appropriately located in

or adjacent (*e.g.*, downstream) to the target locus and on the strand to be nicked and edited. The CFgRNA must also be complementary to a region of the strand to be edited and must include the desired edit to be incorporated.

[00168] At right in FIG. 1C shows the previously formed flap, where the reverse transcriptase (RT) portion of the nickase-RT fusion enzyme adds nucleotides to extend the 3' free end of the nicked strand using the repair template of the CF editing cassette, which includes the desired edit, as a template. The regions of the DNA strands that are synthesized by the RT may include a nick-to-edit region, an edit region, and a post-edit homology (PEH) region. The nick-to-edit region and the post-edit homology (PEH) region are complementary to the unedited (*e.g.*, wild-type (wt)) strand, thus facilitating resolution of the edited flap with the unedited strand via endogenous repair mechanisms, *e.g.*, homology-directed repair (HDR), recombination pathways, or other DNA repair pathways. The target locus may resolve into either wild-type, where the desired edit is not incorporated, or into an edited target locus. Once the DNA flap containing the edit is synthesized, an equilibrium is established between the newly synthesized 3' flap and the wt 5' flap. The equilibrium can be affected by the length of the edit, nick-to-edit distance, and/or post edit homology region. In order for the newly synthesized flap to be incorporated into the genome, the 5' flap is likely degraded by an exonuclease. This allows the 3' flap to anneal to the DNA, and a polymerase then likely fills in any missing nucleotides and a DNA ligase seals the nick.

[00169] At this stage, one DNA strand contains the edit while the second DNA strand does not. A mismatch repair or DNA replication process is likely responsible for copying the edit into both strands. Note that DNA replication and mismatch repair can also favor the wt strand as opposed to the edited strand. If the flap equilibration favors the wt (wildtype) 5' flap, the newly synthesized flap is likely degraded and sealed in the same manner described above.

[00170] Although described with reference to a CF editing cassette, the mechanism depicted in FIG. 1C may also be utilized for integration of a barcode with a CF barcoding cassette. For example, the same nickase-RT fusion enzyme may be utilized to create and add to a 3' flap, wherein the RT portion of the fusion enzyme utilizes the repair template of the CF barcoding cassette as a template to add the barcode to the nicked strand. Accordingly, the regions of the DNA strands that are synthesized by the RT may include a

nick-to-barcode region, a barcode region, and a post-barcode homology (PBH) region. The nick-to-barcode region and the post-barcode homology (PBH) region, similarly to the nick-to-edit region and PEH region, are complementary to the wild-type strand, thus facilitating resolution of the barcoded flap with the wild-type strand via endogenous repair mechanisms, *e.g.*, homology-directed repair (HDR) or other recombination pathways, before being copied into both strands.

[00171] FIG. 1D schematically depicts an example of a single-vector system for trackable nickase-RT fusion editing of live cells according to aspects described herein. Note that the layout of the vector in FIG. 1D is only an example of, and does not limit aspects of, the present disclosure to any particular arrangement or orientation of components. As shown, the “combined editing and barcoding” vector comprises a nickase-RT fusion enzyme, a CF editing cassette comprising a nucleic acid sequence encoding an editing CFgRNA and repair template, and a CF barcoding cassette comprising a nucleic acid sequence encoding a barcoding CFgRNA and repair template. Also shown in FIG. 1D are one or more mammalian promoters (*e.g.*, inducible or constitutive), which may be integrated into the vector at 5' ends of the nickase-RT fusion enzyme, the CF editing cassette, the CF barcoding cassette, and/or other components to drive transcription thereof. For example, the nickase-RT fusion enzyme in FIG. 1D is depicted as being under the transcriptional control of a eukaryotic translation elongation factor 1 α (“EF1A”) promoter, the CF editing cassette is under the transcriptional control of a human U6 (“hU6”) promoter, and the CF barcoding cassette is under the transcriptional control of a mouse U6 (“mU6”) promoter. However, other promoters are also contemplated.

[00172] In certain aspects, the single-vector system further includes a selectable marker to facilitate enrichment or selection of cells successfully transformed with the vector, *e.g.*, while performing the method 100. Accordingly, the selectable marker may be used to “tag” and enrich for transformation events, and may also be under the transcriptional control of a promoter. Examples of suitable markers include antibiotic resistance genes and fluorescent proteins. In FIG. 1D, the selectable marker comprises the PuroR gene, which is under the control of a phosphoglycerate kinase (“PGK”) promoter at a 5' end thereof.

[00173] FIG. 1E schematically depicts an exemplary multi-vector system for trackable nickase-RT fusion editing of live cells according to aspects described herein. In particular,

FIG. 1E depicts a two-vector system, wherein a nickase-RT fusion enzyme, a CF editing cassette comprising a nucleic acid sequence encoding an editing CFgRNA and repair template, and a selectable marker are disposed on a first “editing” vector, and the CF barcoding cassette comprising a nucleic acid sequence encoding a barcoding CFgRNA and repair template is disposed on a second “barcoding” vector. Note that the layout of the two vectors in FIG. 1E is only an example, and does not limit aspects of the present disclosure to any particular arrangement or orientation of components. For example, in certain aspects, the nickase-RT fusion enzyme and/or selectable marker may be integrated into the barcoding vector with the CF barcoding cassette instead of the editing vector.

10 [00174] Similar to FIG. 1D, the nickase-RT fusion enzyme, the CF editing cassette, the CF barcoding cassette, and/or other components of the multiple vectors may be under the control of one or more inducible or constitutive promoters. For purposes of illustration, the same promoters in FIG. 1D are depicted in FIG. 1E. However, other promoters are also contemplated.

15 [00175] FIG. 1F is a simplified graphic depiction of an example of a double-stranded CF editing cassette (top) and CF barcoding cassette (bottom) for trackable nickase-RT fusion editing of live cells. Note that the layouts of the cassettes in FIG. 1F are only examples and does not limit aspects of the present disclosure to any particular arrangement or orientation of components thereof. In FIG. 1F, each of the CF editing cassette and CF barcoding cassette comprise a nucleic acid sequence encoding a CFgRNA covalently linked to a repair template. More particularly, the CF editing cassette comprises from 5' to 3' an optional GG transcription initiation sequence to facilitate high efficiency transcription (e.g., by T7 RNA polymerase) (denoted “GG”); an editing CFgRNA configured to target a first target locus and comprising from 5' to 3' a spacer region (denoted “SR”) having complementarity to the first target locus and a scaffold or repeat region (denoted “CR”); the repair template comprising from 5' to 3' an optional post-edit homology region (denoted “PEH”), an edit (denoted “E”), a nick-to-edit region (denoted “NE”), and a primer binding site (denoted “PBS”); an RNA G-quadruplex region (denoted “QG”); and a PolyT transcription terminator (denoted “TT”).

30 [00176] Similarly, the CF barcoding cassette comprises from 5' to 3' an optional GG transcription initiation sequence (denoted “GG”); a barcoding CFgRNA configured to

target a second target locus and comprising from 5' to 3' a spacer region (denoted "SR") having complementarity to the second target locus and a scaffold (denoted "CR"); the repair template comprising from 5' to 3' an optional post-barcode homology region (denoted "PBH"), a barcode sequence (denoted "BC"), a nick-to-barcode region (denoted "NB"), and a primer binding site (denoted "PBS"); an RNA G-quadruplex region (denoted "QG"); and a PolyT transcription terminator (denoted "TT").

[00177] FIG. 1G is a simplified graphic depiction of an example of a mechanism for selection of successfully edited and/or barcoded live cells. In the example of FIG. 1G, the barcoding CFgRNA is designed to target a second target locus within the coding region of a non-essential cell surface receptor such that a barcode is integrated into said coding region, thus preventing expression of a functional cell surface receptor gene. Accordingly, successful barcode integration results in elimination of the receptor from cell surfaces. Antibody and/or bead-based affinity purification (*e.g.*, affinity depletion) may then be utilized to remove cells that were not successfully barcoded (*e.g.*, still expressing the cell surface receptor), leaving only barcoded cells via negative selection (right of FIG. 1G).

[00178] In such examples, the barcode sequence may comprise a frameshifting edit (bottom of FIG. 1G) and be 8 or more nucleotides in length, wherein the number of nucleotides is not a multiple of 3 (*e.g.*, 10, 11, 13, 14, etc.). Alternatively, the barcode sequence may comprise an in-frame STOP codon (TAGTGA) (center of FIG. 1G) edit and be 9 or more nucleotides in length.

Automated Cell Editing Instruments and Modules to Perform Nucleic Acid-Guided Nuclease Editing in Cells

Automated Cell Editing Instruments

[00179] FIG. 2A depicts an example of an automated multi-module cell processing instrument 200 to, *e.g.*, perform one of the exemplified novel methods using the novel nickase-RT fusion and CFgRNA compositions described herein. The instrument 200, for example, may be and preferably is designed as a stand-alone desktop instrument for use within a laboratory environment. The instrument 200 may incorporate a mixture of reusable and disposable components for performing the various integrated processes in conducting automated genome cleavage and/or editing in cells without human intervention. Illustrated

is a gantry 202, providing an automated mechanical motion system (actuator) (not shown) that supplies XYZ axis motion control to, *e.g.*, an automated (*e.g.*, robotic) liquid handling system 258 including, *e.g.*, an air displacement pipettor 232 which allows for cell processing among multiple modules without human intervention. In some automated multi-module cell processing instruments, the air displacement pipettor 232 is moved by gantry 202 and the various modules and reagent cartridges remain stationary; however, in other aspects, the liquid handling system 258 may stay stationary while the various modules and reagent cartridges are moved. Also included in the automated multi-module cell processing instrument 200 are reagent cartridges 210 comprising reservoirs 212 and transformation module 230 (*e.g.*, a flow-through electroporation device as described in detail in relation to FIGs. 5B – 5F), as well as wash reservoirs 206, cell input reservoir 251 and cell output reservoir 253. The wash reservoirs 206 may be configured to accommodate large tubes, for example, wash solutions, or solutions that are used often throughout an iterative process. Although two of the reagent cartridges 210 comprise a wash reservoir 206 in FIG. 2A, the wash reservoirs instead could be included in a wash cartridge where the reagent and wash cartridges are separate cartridges. In such a case, the reagent cartridge 210 and wash cartridge 204 may be identical except for the consumables (reagents or other components contained within the various inserts) inserted therein.

[00180] In some implementations, the reagent cartridges 210 are disposable kits comprising reagents and cells for use in the automated multi-module cell processing/editing instrument 200. For example, a user may open and position each of the reagent cartridges 210 comprising various desired inserts and reagents within the chassis of the automated multi-module cell editing instrument 200 prior to activating cell processing. Further, each of the reagent cartridges 210 may be inserted into receptacles in the chassis having different temperature zones appropriate for the reagents contained therein.

[00181] Also illustrated in FIG. 2A is the robotic liquid handling system 258 including the gantry 202 and air displacement pipettor 232. In some examples, the robotic handling system 258 may include an automated liquid handling system such as those manufactured by Tecan Group Ltd. of Mannedorf, Switzerland, Hamilton Company of Reno, NV (see, *e.g.*, WO2018015544A1), or Beckman Coulter, Inc. of Fort Collins, CO. (see, *e.g.*,

US20160018427A1). Pipette tips may be provided in a pipette transfer tip supply (not shown) for use with the air displacement pipettor 232.

[00182] Inserts or components of the reagent cartridges 210, in some implementations, are marked with machine-readable indicia (not shown), such as bar codes, for recognition by the robotic handling system 258. For example, the robotic liquid handling system 258 may scan one or more inserts within each of the reagent cartridges 210 to confirm contents. In other implementations, machine-readable indicia may be marked upon each reagent cartridge 210, and a processing system (not shown, but see element 237 of FIG. 2B) of the automated multi-module cell editing instrument 200 may identify a stored materials map based upon the machine-readable indicia. In the aspect illustrated in FIG. 2A, a cell growth module comprises a cell growth vial 218 (described in greater detail below in relation to FIGs. 3A – 3D). Additionally seen is the TFF module 222 (described above in detail in relation to FIGs. 4A – 4E). Also illustrated as part of the automated multi-module cell processing instrument 200 of FIG. 2A is a singulation module 240 (*e.g.*, a solid wall isolation, incubation and normalization device (SWIIN device) is shown here) described herein in relation to FIGs. 6C – 6F, served by, *e.g.*, robotic liquid handling system 258 and air displacement pipettor 232. Additionally seen is a selection module 220. Also note the placement of three heatsinks 255.

[00183] FIG 2B is a simplified representation of the contents of the example of a multi-module cell processing instrument 200 depicted in FIG. 2A. Cartridge-based source materials (such as in reagent cartridges 210), for example, may be positioned in designated areas on a deck of the instrument 200 for access by an air displacement pipettor 232. The deck of the multi-module cell processing instrument 200 may include a protection sink such that contaminants spilling, dripping, or overflowing from any of the modules of the instrument 200 are contained within a lip of the protection sink. Also seen are reagent cartridges 210, which are shown disposed with thermal assemblies 211 which can create temperature zones appropriate for different regions. Note that one of the reagent cartridges also comprises a flow-through electroporation device 230 (FTEP), served by FTEP interface (*e.g.*, manifold arm) and actuator 231. Also seen is TFF module 222 with adjacent thermal assembly 225, where the TFF module is served by TFF interface (*e.g.*, manifold arm) and actuator 233. Thermal assemblies 225, 235, and 245 encompass thermal electric

devices such as Peltier devices, as well as heatsinks, fans and coolers. The rotating growth vial 218 is within a growth module 234, where the growth module is served by two thermal assemblies 235. Selection module is seen at 220. Also seen is the SWIIN module 240, comprising a SWIIN cartridge 241, where the SWIIN module also comprises a thermal assembly 245, illumination 243 (in this aspect, backlighting), evaporation and condensation control 249, and where the SWIIN module is served by SWIIN interface (*e.g.*, manifold arm) and actuator 247. Also seen in this view is touch screen display 201, display actuator 203, illumination 205 (one on either side of multi-module cell processing instrument 200), and cameras 239 (one illumination device on either side of multi-module cell processing instrument 200). Finally, element 237 comprises electronics, such as circuit control boards, high-voltage amplifiers, power supplies, and power entry; as well as pneumatics, such as pumps, valves and sensors.

[00184] FIG. 2C illustrates a front perspective view of multi-module cell processing instrument 200 for use in as a desktop version of the automated multi-module cell editing instrument 200. For example, a chassis 290 may have a width of about 24–48 inches, a height of about 24–48 inches and a depth of about 24–48 inches. Chassis 290 may be and preferably is designed to hold all modules and disposable supplies used in automated cell processing and to perform all processes required without human intervention; that is, chassis 290 is configured to provide an integrated, stand-alone automated multi-module cell processing instrument. As illustrated in FIG. 2C, chassis 290 includes touch screen display 201, cooling grate 264, which allows for air flow via an internal fan (not shown). The touch screen display provides information to a user regarding the processing status of the automated multi-module cell editing instrument 200 and accepts inputs from the user for conducting the cell processing. In this aspect, the chassis 290 is lifted by adjustable feet 270a, 270b, 270c and 270d (feet 270a – 270c are shown in this FIG. 2C). Adjustable feet 270a – 270d, for example, allow for additional air flow beneath the chassis 290.

[00185] Inside the chassis 290, in some implementations, will be most or all of the components described in relation to FIGs. 2A and 2B, including the robotic liquid handling system disposed along a gantry, reagent cartridges 210 including a flow-through electroporation device, a rotating growth vial 218 in a cell growth module 234, a tangential flow filtration module 222, a SWIIN module 240 as well as interfaces and actuators for the

various modules. In addition, chassis 290 houses control circuitry, liquid handling tubes, air pump controls, valves, sensors, thermal assemblies (*e.g.*, heating and cooling units) and other control mechanisms. For examples of multi-module cell editing instruments, see USPNs 10,253,316; 10,329,559; 10,323,242; 10,421,959; 10,465,185; 10,519,437; 5 10,584,333; 10,584,334; 10,647,982; 10,689,645; 10,738,301; 10,738,663 and USSNs 16/412,175 and 16/988,694, all of which are herein incorporated by reference in their entirety.

The Rotating Cell Growth Module

10 [00186] FIG. 3A shows one aspect of a rotating growth vial 300 for use with the cell growth device and in the automated multi-module cell processing instruments described herein. The rotating growth vial 300 is an optically-transparent container having an open end 304 for receiving liquid media and cells, a central vial region 306 that defines the primary container for growing cells, a tapered-to-constricted region 318 defining at least one light path 310, a closed end 316, and a drive engagement mechanism 312. The rotating 15 growth vial 300 has a central longitudinal axis 320 around which the vial rotates, and the light path 310 is generally perpendicular to the longitudinal axis of the vial. The first light path 310 is positioned in the lower constricted portion of the tapered-to-constricted region 318. Optionally, some aspects of the rotating growth vial 300 have a second light path 308 20 in the tapered region of the tapered-to-constricted region 318. Both light paths in this aspect are positioned in a region of the rotating growth vial that is constantly filled with the cell culture (cells + growth media) and are not affected by the rotational speed of the growth vial. The first light path 310 is shorter than the second light path 308 allowing for sensitive measurement of OD values when the OD values of the cell culture in the vial are at a high 25 level (*e.g.*, later in the cell growth process), whereas the second light path 308 allows for sensitive measurement of OD values when the OD values of the cell culture in the vial are at a lower level (*e.g.*, earlier in the cell growth process).

[00187] The drive engagement mechanism 312 engages with a motor (not shown) to rotate the vial. In some aspects, the motor drives the drive engagement mechanism 312 30 such that the rotating growth vial 300 is rotated in one direction only, and in other aspects, the rotating growth vial 300 is rotated in a first direction for a first amount of time or

periodicity, rotated in a second direction (*e.g.*, the opposite direction) for a second amount of time or periodicity, and this process may be repeated so that the rotating growth vial 300 (and the cell culture contents) are subjected to an oscillating motion. Further, the choice of whether the culture is subjected to oscillation and the periodicity therefor may be selected by the user. The first amount of time and the second amount of time may be the same or may be different. The amount of time may be 1, 2, 3, 4, 5, or more seconds, or may be 1, 2, 3, 4 or more minutes. In another aspect, in an early stage of cell growth the rotating growth vial 400 may be oscillated at a first periodicity (*e.g.*, every 60 seconds), and then a later stage of cell growth the rotating growth vial 300 may be oscillated at a second periodicity (*e.g.*, every one second) different from the first periodicity.

[00188] The rotating growth vial 300 may be reusable or, preferably, the rotating growth vial is consumable. In some aspects, the rotating growth vial is consumable and is presented to the user pre-filled with growth medium, where the vial is hermetically sealed at the open end 304 with a foil seal. A medium-filled rotating growth vial packaged in such a manner may be part of a kit for use with a stand-alone cell growth device or with a cell growth module that is part of an automated multi-module cell processing system. To introduce cells into the vial, a user need only pipette up a desired volume of cells and use the pipette tip to punch through the foil seal of the vial. Open end 304 may optionally include an extended lip 302 to overlap and engage with the cell growth device. In automated systems, the rotating growth vial 300 may be tagged with a barcode or other identifying means that can be read by a scanner or camera (not shown) that is part of the automated system.

[00189] The volume of the rotating growth vial 300 and the volume of the cell culture (including growth medium) may vary greatly, but the volume of the rotating growth vial 300 must be large enough to generate a specified total number of cells. In practice, the volume of the rotating growth vial 300 may range from between 1 and 250 mL, between 2 and 100 mL, between 5 and 80 mL, between 10 and 50 mL, or between 12 and 35 mL. Likewise, the volume of the cell culture (cells + growth media) should be appropriate to allow proper aeration and mixing in the rotating growth vial 400. Proper aeration promotes uniform cellular respiration within the growth media. Thus, the volume of the cell culture should be approximately 5-85% of the volume of the growth vial or between 20% and 60%

of the volume of the growth vial. For example, for a 30 mL growth vial, the volume of the cell culture would be from about 1.5 mL to about 26 mL, or from 6 mL to about 18 mL.

[00190] The rotating growth vial 300 preferably is fabricated from a bio-compatible optically transparent material—or at least the portion of the vial comprising the light path(s) is transparent. Additionally, material from which the rotating growth vial is fabricated should be able to be cooled to about 4°C or lower and heated to about 55°C or higher to accommodate both temperature-based cell assays and long-term storage at low temperatures. Further, the material that is used to fabricate the vial must be able to withstand temperatures up to 55°C without deformation while spinning. Suitable materials include cyclic olefin copolymer (COC), glass, polyvinyl chloride, polyethylene, polyamide, polypropylene, polycarbonate, poly(methyl methacrylate (PMMA), polysulfone, polyurethane, and co-polymers of these and other polymers. Preferred materials include polypropylene, polycarbonate, or polystyrene. In some aspects, the rotating growth vial is inexpensively fabricated by, *e.g.*, injection molding or extrusion.

[00191] FIG. 3B is a perspective view of one aspect of a cell growth device 330. FIG. 3C depicts a cut-away view of the cell growth device 330 from FIG. 3B. In both figures, the rotating growth vial 300 is seen positioned inside a main housing 336 with the extended lip 302 of the rotating growth vial 300 extending above the main housing 336. Additionally, end housings 352, a lower housing 332 and flanges 334 are indicated in both figures. Flanges 334 are used to attach the cell growth device 330 to heating/cooling means or other structure (not shown). FIG. 3C depicts additional detail. In FIG. 3C, upper bearing 342 and lower bearing 340 are shown positioned within main housing 336. Upper bearing 342 and lower bearing 340 support the vertical load of rotating growth vial 300. Lower housing 332 contains the drive motor 338. The cell growth device 330 of FIG. 3C comprises two light paths: a primary light path 344, and a secondary light path 350. Light path 344 corresponds to light path 310 positioned in the constricted portion of the tapered-to-constricted portion of the rotating growth vial 300, and light path 350 corresponds to light path 308 in the tapered portion of the tapered-to-constricted portion of the rotating growth via 316. Light paths 310 and 308 are not shown in FIG. 3C but may be seen in FIG. 3A. In addition to light paths 344 and 340, there is an emission board 348 to illuminate the light path(s), and

detector board 346 to detect the light after the light travels through the cell culture liquid in the rotating growth vial 300.

[00192] The motor 338 engages with drive mechanism 312 and is used to rotate the rotating growth vial 300. In some aspects, motor 338 is a brushless DC type drive motor with built-in drive controls that can be set to hold a constant revolution per minute (RPM) between 0 and about 3000 RPM. Alternatively, other motor types such as a stepper, servo, brushed DC, and the like can be used. Optionally, the motor 338 may also have direction control to allow reversing of the rotational direction, and a tachometer to sense and report actual RPM. The motor is controlled by a processor (not shown) according to, *e.g.*, standard protocols programmed into the processor and/or user input, and the motor may be configured to vary RPM to cause axial precession of the cell culture thereby enhancing mixing, *e.g.*, to prevent cell aggregation, increase aeration, and optimize cellular respiration.

[00193] Main housing 336, end housings 352 and lower housing 332 of the cell growth device 330 may be fabricated from any suitable, robust material including aluminum, stainless steel, and other thermally conductive materials, including plastics. These structures or portions thereof can be created through various techniques, *e.g.*, metal fabrication, injection molding, creation of structural layers that are fused, etc. Whereas the rotating growth vial 300 is envisioned in some aspects to be reusable, but preferably is consumable, the other components of the cell growth device 330 are preferably reusable and function as a stand-alone benchtop device or as a module in a multi-module cell processing system.

[00194] The processor (not shown) of the cell growth device 330 may be programmed with information to be used as a “blank” or control for the growing cell culture. A “blank” or control is a vessel containing cell growth medium only, which yields 100% transmittance and 0 OD (optical density), while the cell sample will deflect light rays and will have a lower percent transmittance and higher OD. As the cells grow in the media and become denser, transmittance will decrease and OD will increase. The processor (not shown) of the cell growth device 330 may be programmed to use wavelength values for blanks commensurate with the growth media typically used in cell culture (whether, *e.g.*, mammalian cells, bacterial cells, animal cells, yeast cells, etc.). Alternatively, a second

spectrophotometer and vessel may be included in the cell growth device 330, where the second spectrophotometer is used to read a blank at designated intervals.

[00195] FIG. 3D illustrates a cell growth device 330 as part of an assembly comprising the cell growth device 330 of FIG. 3B coupled to light source 390, detector 392, and thermal components 394. The rotating growth vial 300 is inserted into the cell growth device. Components of the light source 390 and detector 392 (*e.g.*, such as a photodiode with gain control to cover 5-log) are coupled to the main housing of the cell growth device. The lower housing 332 that houses the motor that rotates the rotating growth vial 300 is illustrated, as is one of the flanges 334 that secures the cell growth device 330 to the assembly. Also, the thermal components 394 illustrated are a Peltier device or thermoelectric cooler. In this aspect, thermal control is accomplished by attachment and electrical integration of the cell growth device 330 to the thermal components 394 via the flange 334 on the base of the lower housing 332. Thermoelectric coolers are capable of “pumping” heat to either side of a junction, either cooling a surface or heating a surface depending on the direction of current flow. In one aspect, a thermistor is used to measure the temperature of the main housing and then, through a standard electronic proportional-integral-derivative (PID) controller loop, the rotating growth vial 300 is controlled to approximately +/- 0.5°C.

[00196] In use, cells are inoculated (cells can be pipetted, *e.g.*, from an automated liquid handling system or by a user) into pre-filled growth media of a rotating growth vial 300 by piercing through the foil seal or film. The programmed software of the cell growth device 330 sets the control temperature for growth, typically 30°C, then slowly starts the rotation of the rotating growth vial 300. The cell/growth media mixture slowly moves vertically up the wall due to centrifugal force allowing the rotating growth vial 300 to expose a large surface area of the mixture to a normal oxygen environment. The growth monitoring system takes either continuous readings of the OD or OD measurements at pre-set or pre-programmed time intervals. These measurements are stored in internal memory and if requested the software plots the measurements versus time to display a growth curve. If enhanced mixing is required, *e.g.*, to optimize growth conditions, the speed of the vial rotation can be varied to cause an axial precession of the liquid, and/or a complete directional change can be performed at programmed intervals. The growth monitoring can

be programmed to automatically terminate the growth stage at a pre-determined OD, and then quickly cool the mixture to a lower temperature to inhibit further growth.

[00197] One application for the cell growth device 330 is to constantly measure the optical density of a growing cell culture. One advantage of the described cell growth device is that optical density can be measured continuously (kinetic monitoring) or at specific time intervals; e.g., every 5, 10, 15, 20, 30, 45, or 60 seconds, or every 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 minutes. While the cell growth device 330 has been described in the context of measuring the OD of a growing cell culture, it should, however, be understood by a skilled artisan given the teachings of the present specification that other cell growth parameters can be measured in addition to or instead of cell culture OD. As with optional measure of cell growth in relation to the solid wall device or module described *supra*, spectroscopy using visible, ultraviolet (UV), or near infrared (NIR) light allows monitoring the concentration of nutrients and/or wastes in the cell culture and other spectroscopic measurements may be made; that is, other spectral properties can be measured via, e.g., dielectric impedance spectroscopy, visible fluorescence, fluorescence polarization, or luminescence. Additionally, the cell growth device 330 may include additional sensors for measuring, e.g., dissolved oxygen, carbon dioxide, pH, conductivity, and the like. For additional details regarding rotating growth vials and cell growth devices see USPNs 10,435,662; 10,443,031; 10,590,375, and 10,717,959.

The Cell Concentration Module

[00198] As described above in relation to the rotating growth vial and cell growth module, in order to obtain an adequate number of cells for transformation or transfection, cells typically are grown to a specific optical density in medium appropriate for the growth of the cells of interest; however, for effective transformation or transfection, it is desirable to decrease the volume of the cells as well as render the cells competent via buffer or medium exchange. Thus, one sub-component or module that is desired in cell processing systems to perform the methods described herein is a module or component that can grow, perform buffer exchange, and/or concentrate cells and render them competent so that they may be transformed or transfected with the nucleic acids needed for engineering or editing the cell's genome.

[00199] FIG. 4A shows a retentate member 422 (top), permeate member 420 (middle) and a tangential flow assembly 410 (bottom) comprising the retentate member 422, membrane 424 (not seen in FIG. 4A), and permeate member 420 (also not seen). In FIG. 4A, retentate member 422 comprises a tangential flow channel 402, which has a serpentine configuration that initiates at one lower corner of retentate member 422—specifically at retentate port 428—traverses across and up then down and across retentate member 422, ending in the other lower corner of retentate member 422 at a second retentate port 428. Also seen on retentate member 422 are energy directors 491, which circumscribe the region where a membrane or filter (not seen in this FIG. 4A) is seated, as well as interdigitate between areas of channel 402. Energy directors 491 in this aspect mate with and serve to facilitate ultrasonic welding or bonding of retentate member 422 with permeate/filtrate member 420 via the energy director component 491 on permeate/filtrate member 420 (at right). Additionally, countersinks 423 can be seen, two on the bottom one at the top middle of retentate member 422. Countersinks 423 are used to couple and tangential flow assembly 410 to a reservoir assembly (not seen in this FIG. 4A but see FIG. 4B).

[00200] Permeate/filtrate member 420 is seen in the middle of FIG. 4A and comprises, in addition to energy director 491, through-holes for retentate ports 428 at each bottom corner (which mate with the through-holes for retentate ports 428 at the bottom corners of retentate member 422), as well as a tangential flow channel 402 and two permeate/filtrate ports 426 positioned at the top and center of permeate member 420. The tangential flow channel 402 structure in this aspect has a serpentine configuration and an undulating geometry, although other geometries may be used. Permeate member 420 also comprises countersinks 423, coincident with the countersinks 423 on retentate member 420.

[00201] On the left of FIG. 4A is a tangential flow assembly 410 comprising the retentate member 422 and permeate member 420 seen in this FIG. 4A. In this view, retentate member 422 is “on top” of the view, a membrane (not seen in this view of the assembly) would be adjacent and under retentate member 422 and permeate member 420 (also not seen in this view of the assembly) is adjacent to and beneath the membrane. Again countersinks 423 are seen, where the countersinks in the retentate member 422 and the permeate member 420 are coincident and configured to mate with threads or mating

elements for the countersinks disposed on a reservoir assembly (not seen in FIG. 4A but see FIG. 4B).

[00202] A membrane or filter is disposed between the retentate and permeate members, where fluids can flow through the membrane but cells cannot and are thus retained in the flow channel disposed in the retentate member. Filters or membranes appropriate for use in the TFF device/module are those that are solvent resistant, are contamination free during filtration, and are able to retain the types and sizes of cells of interest. For example, in order to retain small cell types such as bacterial cells, pore sizes can be as low as 0.2 μm , however for other cell types, the pore sizes can be as high as 20 μm . Indeed, the pore sizes useful in the TFF device/module include filters with sizes from 0.20 μm , 0.21 μm , 0.22 μm , 0.23 μm , 0.24 μm , 0.25 μm , 0.26 μm , 0.27 μm , 0.28 μm , 0.29 μm , 0.30 μm , 0.31 μm , 0.32 μm , 0.33 μm , 0.34 μm , 0.35 μm , 0.36 μm , 0.37 μm , 0.38 μm , 0.39 μm , 0.40 μm , 0.41 μm , 0.42 μm , 0.43 μm , 0.44 μm , 0.45 μm , 0.46 μm , 0.47 μm , 0.48 μm , 0.49 μm , 0.50 μm and larger. The filters may be fabricated from any suitable non-reactive material including cellulose mixed ester (cellulose nitrate and acetate) (CME), polycarbonate (PC), polyvinylidene fluoride (PVDF), polyethersulfone (PES), polytetrafluoroethylene (PTFE), nylon, glass fiber, or metal substrates as in the case of laser or electrochemical etching.

[00203] The length of the channel structure 402 may vary depending on the volume of the cell culture to be grown and the optical density of the cell culture to be concentrated. The length of the channel structure typically is between 60 mm and 300 mm, or between 70 mm and 200 mm, or between 80 mm and 100 mm. The cross-section configuration of the flow channel 402 may be round, elliptical, oval, square, rectangular, trapezoidal, or irregular. If square, rectangular, or another shape with generally straight sides, the cross section may be between about 10 μm and 1000 μm wide, or between 200 μm and 800 μm wide, or between 300 μm and 700 μm wide, or between 400 μm and 600 μm wide; and between about 10 μm and 1000 μm high, or between 200 μm and 800 μm high, or between 300 μm and 700 μm high, or between 400 μm and 600 μm high. If the cross section of the flow channel 402 is generally round, oval or elliptical, the radius of the channel may be between about 50 μm and 1000 μm in hydraulic radius, or between 5 μm and 800 μm in hydraulic radius, or between 200 μm and 700 μm in hydraulic radius, or between 300 μm and 600 μm wide in hydraulic radius, or between about 200 and 500 μm in hydraulic radius.

Moreover, the volume of the channel in the retentate 422 and permeate 420 members may be different depending on the depth of the channel in each member.

[00204] FIG. 4B shows front perspective (right) and rear perspective (left) views of a reservoir assembly 450 configured to be used with the tangential flow assembly 410 seen in FIG. 4A. Seen in the front perspective view (*e.g.*, “front” being the side of reservoir assembly 450 that is coupled to the tangential flow assembly 410 seen in FIG. 4A) are retentate reservoirs 452 on either side of permeate reservoir 454. Also seen are permeate ports 426, retentate ports 428, and three threads or mating elements 425 for countersinks 423 (countersinks 423 not seen in this FIG. 4B). Threads or mating elements 425 for countersinks 423 are configured to mate or couple the tangential flow assembly 410 (seen in FIG. 4A) to reservoir assembly 450. Alternatively or in addition, fasteners, sonic welding or heat stakes may be used to mate or couple the tangential flow assembly 410 to reservoir assembly 450. In addition gasket 445 is seen covering the top of reservoir assembly 450. Gasket 445 is described in detail in relation to FIG. 4E. At left in FIG. 4B is a rear perspective view of reservoir assembly 1250, where “rear” is the side of reservoir assembly 450 that is not coupled to the tangential flow assembly. Seen are retentate reservoirs 452, permeate reservoir 454, and gasket 445.

[00205] The TFF device may be fabricated from any robust material in which channels (and channel branches) may be milled including stainless steel, silicon, glass, aluminum, or plastics including cyclic-olefin copolymer (COC), cyclo-olefin polymer (COP), polystyrene, polyvinyl chloride, polyethylene, polyamide, polyethylene, polypropylene, acrylonitrile butadiene, polycarbonate, polyetheretherketone (PEEK), poly(methyl methacrylate) (PMMA), polysulfone, and polyurethane, and co-polymers of these and other polymers. If the TFF device/module is disposable, preferably it is made of plastic. In some aspects, the material used to fabricate the TFF device/module is thermally-conductive so that the cell culture may be heated or cooled to a desired temperature. In certain aspects, the TFF device is formed by precision mechanical machining, laser machining, electro discharge machining (for metal devices); wet or dry etching (for silicon devices); dry or wet etching, powder or sandblasting, photostructuring (for glass devices); or thermoforming, injection molding, hot embossing, or laser machining (for plastic devices) using the materials mentioned above that are amenable to this mass production techniques.

[00206] FIG. 4C depicts a top-down view of the reservoir assemblies 450 shown in FIG. 4B. FIG. 4D depicts a cover 444 for reservoir assembly 450 shown in FIG. 4B and 4E depicts a gasket 445 that in operation is disposed on cover 444 of reservoir assemblies 450 shown in FIG. 4B. FIG. 4C is a top-down view of reservoir assembly 450, showing the
5 tops of the two retentate reservoirs 452, one on either side of permeate reservoir 454. Also seen are grooves 432 that will mate with a pneumatic port (not shown), and fluid channels 434 that reside at the bottom of retentate reservoirs 452, which fluidically couple the retentate reservoirs 452 with the retentate ports 428 (not shown), via the through-holes for the retentate ports in permeate member 420 and membrane 424 (also not shown). FIG. 4D
10 depicts a cover 444 that is configured to be disposed upon the top of reservoir assembly 450. Cover 444 has round cut-outs at the top of retentate reservoirs 452 and permeate/filtrate reservoir 454. Again, at the bottom of retentate reservoirs 452 fluid channels 434 can be seen, where fluid channels 434 fluidically couple retentate reservoirs 452 with the retentate ports 428 (not shown). Also shown are three pneumatic ports 430
15 for each retentate reservoir 452 and permeate/filtrate reservoir 454. FIG. 4E depicts a gasket 445 that is configured to be disposed upon the cover 444 of reservoir assembly 450. Seen are three fluid transfer ports 442 for each retentate reservoir 452 and for permeate/filtrate reservoir 454. Again, three pneumatic ports 430, for each retentate reservoir 452 and for permeate/filtrate reservoir 454, are shown.

[00207] The overall work flow for cell growth comprises loading a cell culture to be
20 grown into a first retentate reservoir, optionally bubbling air or an appropriate gas through the cell culture, passing or flowing the cell culture through the first retentate port then tangentially through the TFF channel structure while collecting medium or buffer through one or both of the permeate ports 406, collecting the cell culture through a second retentate
25 port 404 into a second retentate reservoir, optionally adding additional or different medium to the cell culture and optionally bubbling air or gas through the cell culture, then repeating the process, all while measuring, *e.g.*, the optical density of the cell culture in the retentate reservoirs continuously or at desired intervals. Measurements of optical densities at programmed time intervals are accomplished using a 600 nm Light Emitting Diode (LED)
30 that has been columnated through an optic into the retentate reservoir(s) containing the growing cells. The light continues through a collection optic to the detection system which

consists of a (digital) gain-controlled silicone photodiode. Generally, optical density is shown as the absolute value of the logarithm with base 10 of the power transmission factors of an optical attenuator: $OD = -\log_{10}(\text{Power out/Power in})$. Since OD is the measure of optical attenuation—that is, the sum of absorption, scattering, and reflection—the TFF device OD measurement records the overall power transmission, so as the cells grow and become denser in population, the OD (the loss of signal) increases. The OD system is pre-calibrated against OD standards with these values stored in an on-board memory accessible by the measurement program.

[00208] In the channel structure, the membrane bifurcating the flow channels retains the cells on one side of the membrane (the retentate side 422) and allows unwanted medium or buffer to flow across the membrane into a filtrate or permeate side (*e.g.*, permeate member 420) of the device. Bubbling air or other appropriate gas through the cell culture both aerates and mixes the culture to enhance cell growth. During the process, medium that is removed during the flow through the channel structure is removed through the permeate/filtrate ports 406. Alternatively, cells can be grown in one reservoir with bubbling or agitation without passing the cells through the TFF channel from one reservoir to the other.

[00209] The overall work flow for cell concentration using the TFF device/module involves flowing a cell culture or cell sample tangentially through the channel structure. As with the cell growth process, the membrane bifurcating the flow channels retains the cells on one side of the membrane and allows unwanted medium or buffer to flow across the membrane into a permeate/filtrate side (*e.g.*, permeate member 420) of the device. In this process, a fixed volume of cells in medium or buffer is driven through the device until the cell sample is collected into one of the retentate ports 404, and the medium/buffer that has passed through the membrane is collected through one or both of the permeate/filtrate ports 406. All types of prokaryotic and eukaryotic cells—both adherent and non-adherent cells—can be grown in the TFF device. Adherent cells may be grown on beads or other cell scaffolds suspended in medium that flow through the TFF device.

[00210] The medium or buffer used to suspend the cells in the cell concentration device/module may be any suitable medium or buffer for the type of cells being transformed or transfected, such as LB, SOC, TPD, YPG, YPAD, MEM, DMEM, IMDM,

RPMI, Hanks', PBS and Ringer's solution, where the media may be provided in a reagent cartridge as part of a kit. For culture of adherent cells, cells may be disposed on beads, microcarriers, or other type of scaffold suspended in medium. Most normal mammalian tissue-derived cells— except those derived from the hematopoietic system—are anchorage dependent and need a surface or cell culture support for normal proliferation. In the rotating growth vial described herein, microcarrier technology is leveraged. Microcarriers of particular use typically have a diameter of 100-300 μm and have a density slightly greater than that of the culture medium (thus facilitating an easy separation of cells and medium for, *e.g.*, medium exchange) yet the density must also be sufficiently low to allow complete suspension of the carriers at a minimum stirring rate in order to avoid hydrodynamic damage to the cells. Many different types of microcarriers are available, and different microcarriers are optimized for different types of cells. There are positively charged carriers, such as Cytodex 1 (dextran-based, GE Healthcare), DE-52 (cellulose-based, Sigma-Aldrich Labware), DE-53 (cellulose-based, Sigma-Aldrich Labware), and HLX 11-170 (polystyrene-based); collagen- or ECM- (extracellular matrix) coated carriers, such as Cytodex 3 (dextran-based, GE Healthcare) or HyQ-sphere Pro-F 102-4 (polystyrene-based, Thermo Scientific); non-charged carriers, like HyQ-sphere P 102-4 (Thermo Scientific); or macroporous carriers based on gelatin (Cultisphere, Percell Biolytica) or cellulose (Cytopore, GE Healthcare).

[00211] In both the cell growth and concentration processes, passing the cell sample through the TFF device and collecting the cells in one of the retentate ports 404 while collecting the medium in one of the permeate/filtrate ports 406 is considered “one pass” of the cell sample. The transfer between retentate reservoirs “flips” the culture. The retentate and permeate ports collecting the cells and medium, respectively, for a given pass reside on the same end of TFF device/module with fluidic connections arranged so that there are two distinct flow layers for the retentate and permeate/filtrate sides, but if the retentate port 404 resides on the retentate member of device/module (that is, the cells are driven through the channel above the membrane and the filtrate (medium) passes to the portion of the channel below the membrane), the permeate/filtrate port 406 will reside on the permeate member of device/module and vice versa (that is, if the cell sample is driven through the channel below the membrane, the filtrate (medium) passes to the portion of the channel

above the membrane). Due to the high pressures used to transfer the cell culture and fluids through the flow channel of the TFF device, the effect of gravity is negligible.

[00212] At the conclusion of a “pass” in either of the growth and concentration processes, the cell sample is collected by passing through the retentate port 404 and into the retentate reservoir (not shown). To initiate another “pass”, the cell sample is passed again through the TFF device, this time in a flow direction that is reversed from the first pass. The cell sample is collected by passing through the retentate port 404 and into retentate reservoir (not shown) on the opposite end of the device/module from the retentate port 404 that was used to collect cells during the first pass. Likewise, the medium/buffer that passes through the membrane on the second pass is collected through the permeate port 406 on the opposite end of the device/module from the permeate port 406 that was used to collect the filtrate during the first pass, or through both ports. This alternating process of passing the retentate (the concentrated cell sample) through the device/module is repeated until the cells have been grown to a desired optical density, and/or concentrated to a desired volume, and both permeate ports (*e.g.*, if there are more than one) can be open during the passes to reduce operating time. In addition, buffer exchange may be effected by adding a desired buffer (or fresh medium) to the cell sample in the retentate reservoir, before initiating another “pass”, and repeating this process until the old medium or buffer is diluted and filtered out and the cells reside in fresh medium or buffer. Note that buffer exchange and cell growth may (and typically do) take place simultaneously, and buffer exchange and cell concentration may (and typically do) take place simultaneously. For further information and alternative aspects on TFFs see, *e.g.*, USSNs 62/728,365, filed 07 September 2018; 62/857,599, filed 05 June 2019; and 62/867,415, filed 27 June 2019.

25 The Cell Transformation Module

[00213] FIG. 5A depicts an example of a combination reagent cartridge and electroporation device 500 (“cartridge”) that may be used in an automated multi-module cell processing instrument along with the TFF module. In addition, in certain aspects the material used to fabricate the cartridge is thermally-conductive, as in certain aspects the cartridge 500 contacts a thermal device (not shown), such as a Peltier device or thermoelectric cooler, that heats or cools reagents in the reagent reservoirs or reservoirs

504. Reagent reservoirs or reservoirs 504 may be reservoirs into which individual tubes of reagents are inserted as shown in FIG. 5A, or the reagent reservoirs may hold the reagents without inserted tubes. Additionally, the reservoirs in a reagent cartridge may be configured for any combination of tubes, co-joined tubes, and direct-fill of reagents.

5 [00214] In one aspect, the reagent reservoirs or reservoirs 504 of reagent cartridge 500 are configured to hold various size tubes, including, *e.g.*, 250 mL tubes, 25 mL tubes, 10 mL tubes, 5 mL tubes, and Eppendorf or microcentrifuge tubes. In yet another aspect, all reservoirs may be configured to hold the same size tube, *e.g.*, 5 mL tubes, and reservoir inserts may be used to accommodate smaller tubes in the reagent reservoir. In yet another
10 aspect—particularly in an aspect where the reagent cartridge is disposable—the reagent reservoirs hold reagents without inserted tubes. In this disposable aspect, the reagent cartridge may be part of a kit, where the reagent cartridge is pre-filled with reagents and the receptacles or reservoirs sealed with, *e.g.*, foil, heat seal acrylic or the like and presented to a consumer where the reagent cartridge can then be used in an automated multi-module
15 cell processing instrument. As one of ordinary skill in the art will appreciate given the present disclosure, the reagents contained in the reagent cartridge will vary depending on work flow; that is, the reagents will vary depending on the processes to which the cells are subjected in the automated multi-module cell processing instrument, *e.g.*, protein production, cell transformation and culture, cell editing, etc.

20 [00215] Reagents such as cell samples, enzymes, buffers, nucleic acid vectors, expression cassettes, proteins or peptides, reaction components (such as, *e.g.*, MgCl₂, dNTPs, nucleic acid assembly reagents, gap repair reagents, and the like), wash solutions, ethanol, and magnetic beads for nucleic acid purification and isolation, etc. may be positioned in the reagent cartridge at a known position. In some aspects of cartridge 500,
25 the cartridge comprises a script (not shown) readable by a processor (not shown) for dispensing the reagents. Also, the cartridge 500 as one component in an automated multi-module cell processing instrument may comprise a script specifying two, three, four, five, ten or more processes to be performed by the automated multi-module cell processing instrument. In certain aspects, the reagent cartridge is disposable and is pre-packaged with
30 reagents tailored to performing specific cell processing protocols, *e.g.*, genome editing or protein production. Because the reagent cartridge contents vary while components/modules

of the automated multi-module cell processing instrument or system may not, the script associated with a particular reagent cartridge matches the reagents used and cell processes performed. Thus, *e.g.*, reagent cartridges may be pre-packaged with reagents for genome editing and a script that specifies the process steps for performing genome editing in an automated multi-module cell processing instrument, or, *e.g.*, reagents for protein expression and a script that specifies the process steps for performing protein expression in an automated multi-module cell processing instrument.

[00216] For example, the reagent cartridge may comprise a script to pipette competent cells from a reservoir, transfer the cells to a transformation module, pipette a nucleic acid solution comprising a vector with expression cassette from another reservoir in the reagent cartridge, transfer the nucleic acid solution to the transformation module, initiate the transformation process for a specified time, then move the transformed cells to yet another reservoir in the reagent cassette or to another module such as a cell growth module in the automated multi-module cell processing instrument. In another example, the reagent cartridge may comprise a script to transfer a nucleic acid solution comprising a vector from a reservoir in the reagent cassette, nucleic acid solution comprising editing oligonucleotide cassettes in a reservoir in the reagent cassette, and a nucleic acid assembly mix from another reservoir to the nucleic acid assembly/desalting module, if present. The script may also specify process steps performed by other modules in the automated multi-module cell processing instrument. For example, the script may specify that the nucleic acid assembly/desalting reservoir be heated to 50°C for 30 minutes to generate an assembled product; and desalting and resuspension of the assembled product via magnetic bead-based nucleic acid purification involving a series of pipette transfers and mixing of magnetic beads, ethanol wash, and buffer.

[00217] As described in relation to FIGs. 5B and 5C below, the examples of reagent cartridges for use in the automated multi-module cell processing instruments may include one or more electroporation devices, preferably flow-through electroporation (FTEP) devices. In yet other aspects, the reagent cartridge is separate from the transformation module. Electroporation is a widely-used method for permeabilization of cell membranes that works by temporarily generating pores in the cell membranes with electrical stimulation. Applications of electroporation include the delivery of DNA, RNA, siRNA,

peptides, proteins, antibodies, drugs or other substances to a variety of cells such as mammalian cells (including human cells), plant cells, archaea, yeasts, other eukaryotic cells, bacteria, and other cell types. In some aspects, a cell is a prokaryotic cell. In some aspects, a cell is an archaea cell. In some aspects, a cell is a bacterial cell. In some aspects, a cell is an *Escherichia coli* cell. In some aspects, a cell is a eukaryotic cell. In some aspects, a cell is an animal cell. In some aspects, a cell is a mammalian cell. In some aspects, a cell is a human cell. In some aspects, the cell is an induced pluripotent stem cell (iPSC). In some aspects, a cell is a non-human animal cell. In some aspects, a cell is a non-human mammalian cell. In some aspects, a cell is a primate cell. In some aspects, a cell is a rodent cell. In some aspects, a cell is a plant cell. In some aspects, a cell is a fungal cell. In some aspects, a cell is a yeast cell. In some aspects, a cell is a *Saccharomyces cerevisiae* cell. In some aspects, a cell is a *Schizosaccharomyces pombe* cell.

[00218] Electrical stimulation may also be used for cell fusion in the production of hybridomas or other fused cells. During a typical electroporation procedure, cells are suspended in a buffer or medium that is favorable for cell survival. For bacterial cell electroporation, low conductance mediums, such as water, glycerol solutions and the like, are often used to reduce the heat production by transient high current. In traditional electroporation devices, the cells and material to be electroporated into the cells (collectively “the cell sample”) are placed in a cuvette embedded with two flat electrodes for electrical discharge. For example, Bio-Rad (Hercules, Calif.) makes the GENE PULSER XCELL™ line of products to electroporate cells in cuvettes. Traditionally, electroporation requires high field strength; however, the flow-through electroporation devices included in the reagent cartridges achieve high efficiency cell electroporation with low toxicity. The reagent cartridges of the disclosure allow for particularly easy integration with robotic liquid handling instrumentation that is typically used in automated instruments and systems such as air displacement pipettors. Such automated instrumentation includes, but is not limited to, off-the-shelf automated liquid handling systems from Tecan (Mannedorf, Switzerland), Hamilton (Reno, NV), Beckman Coulter (Fort Collins, CO), etc.

[00219] FIGs. 5B and 5C are top perspective and bottom perspective views, respectively, of an example of an FTEP device 550 that may be part of (e.g., a component

in) reagent cartridge 500 in FIG. 5A or may be a stand-alone module; that is, not a part of a reagent cartridge or other module. FIG. 5B depicts an FTEP device 550. The FTEP device 550 has wells that define cell sample inlets 552 and cell sample outlets 554. FIG. 5C is a bottom perspective view of the FTEP device 550 of FIG. 5B. An inlet well 552 and an outlet well 554 can be seen in this view. Also seen in FIG. 5C are the bottom of an inlet 562 corresponding to well 552, the bottom of an outlet 564 corresponding to the outlet well 554, the bottom of a defined flow channel 566 and the bottom of two electrodes 568 on either side of flow channel 566. The FTEP devices may comprise push-pull pneumatic means to allow multi-pass electroporation procedures; that is, cells to electroporated may be “pulled” from the inlet toward the outlet for one pass of electroporation, then be “pushed” from the outlet end of the FTEP device toward the inlet end to pass between the electrodes again for another pass of electroporation. Further, this process may be repeated one to many times. For additional information regarding FTEP devices, see, *e.g.*, USPNs 10,435,713; 10,443,074; 10,323,258; 10,508,288; 10,415,058; 10,851,389; and 10,557,150. Further, other aspects of the reagent cartridge may provide or accommodate electroporation devices that are not configured as FTEP devices, such as those described in USPN 10,738,327. For reagent cartridges useful in the present automated multi-module cell processing instruments, see, *e.g.*, USPN 10,376,889; 10,406,525; 10,478,822; 10,576,474; and 10,639,637.

[00220] Additional details of the FTEP devices are illustrated in FIGs. 5D - 5F. Note that in the FTEP devices in FIGs. 5D -- 5F the electrodes are placed such that a first electrode is placed between an inlet and a narrowed region of the flow channel, and the second electrode is placed between the narrowed region of the flow channel and an outlet. FIG. 5D shows a top planar view of an FTEP device 550 having an inlet 552 for introducing a fluid containing cells and exogenous material into FTEP device 550 and an outlet 554 for removing the transformed cells from the FTEP following electroporation. The electrodes 568 are introduced through channels (not shown) in the device. FIG. 5E shows a cutaway view from the top of the FTEP device 550, with the inlet 552, outlet 554, and electrodes 568 positioned with respect to a flow channel 566. FIG. 5F shows a side cutaway view of FTEP device 550 with the inlet 552 and inlet channel 572, and outlet 554 and outlet channel 574. The electrodes 568 are positioned in electrode channels 576 so that they are in fluid

communication with the flow channel 566, but not directly in the path of the cells traveling through the flow channel 566. Note that the first electrode is placed between the inlet and the narrowed region of the flow channel, and the second electrode is placed between the narrowed region of the flow channel and the outlet. The electrodes 568 in this aspect of the device are positioned in the electrode channels 576 which are generally perpendicular to the flow channel 566 such that the fluid containing the cells and exogenous material flows from the inlet channel 572 through the flow channel 566 to the outlet channel 574, and in the process fluid flows into the electrode channels 576 to be in contact with the electrodes 568. In this aspect, the inlet channel, outlet channel and electrode channels all originate from the same planar side of the device. In certain aspects, however, the electrodes may be introduced from a different planar side of the FTEP device than the inlet and outlet channels.

[00221] In the FTEP devices of the disclosure, the toxicity level of the transformation results in greater than 30% viable cells after electroporation, preferably greater than 35%, 40%, 45%, 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95% or even 99% viable cells following transformation, depending on the cell type and the nucleic acids being introduced into the cells.

[00222] The housing of the FTEP device can be made from many materials depending on whether the FTEP device is to be reused, autoclaved, or is disposable, including stainless steel, silicon, glass, resin, polyvinyl chloride, polyethylene, polyamide, polystyrene, polyethylene, polypropylene, acrylonitrile butadiene, polycarbonate, polyetherethketone (PEEK), polysulfone and polyurethane, co-polymers of these and other polymers. Similarly, the walls of the channels in the device can be made of any suitable material including silicone, resin, glass, glass fiber, polyvinyl chloride, polyethylene, polyamide, polyethylene, polypropylene, acrylonitrile butadiene, polycarbonate, polyetherethketone (PEEK), polysulfone and polyurethane, co-polymers of these and other polymers. Preferred materials include crystal styrene, cyclo-olefin polymer (COP) and cyclic olephin co-polymers (COC), which allow the device to be formed entirely by injection molding in one piece with the exception of the electrodes and, *e.g.*, a bottom sealing film if present.

[00223] The FTEP devices described herein (or portions of the FTEP devices) can be created or fabricated via various techniques, *e.g.*, as entire devices or by creation of

structural layers that are fused or otherwise coupled. For example, for metal FTEP devices, fabrication may include precision mechanical machining or laser machining; for silicon FTEP devices, fabrication may include dry or wet etching; for glass FTEP devices, fabrication may include dry or wet etching, powderblasting, sandblasting, or photostructuring; and for plastic FTEP devices fabrication may include thermoforming, injection molding, hot embossing, or laser machining. The components of the FTEP devices may be manufactured separately and then assembled, or certain components of the FTEP devices (or even the entire FTEP device except for the electrodes) may be manufactured (*e.g.*, using 3D printing) or molded (*e.g.*, using injection molding) as a single entity, with other components added after molding. For example, housing and channels may be manufactured or molded as a single entity, with the electrodes later added to form the FTEP unit. Alternatively, the FTEP device may also be formed in two or more parallel layers, *e.g.*, a layer with the horizontal channel and filter, a layer with the vertical channels, and a layer with the inlet and outlet ports, which are manufactured and/or molded individually and assembled following manufacture.

[00224] In specific aspects, the FTEP device can be manufactured using a circuit board as a base, with the electrodes, filter and/or the flow channel formed in the desired configuration on the circuit board, and the remaining housing of the device containing, *e.g.*, the one or more inlet and outlet channels and/or the flow channel formed as a separate layer that is then sealed onto the circuit board. The sealing of the top of the housing onto the circuit board provides the desired configuration of the different elements of the FTEP devices of the disclosure. Also, two to many FTEP devices may be manufactured on a single substrate, then separated from one another thereafter or used in parallel. In certain aspects, the FTEP devices are reusable and, in some aspects, the FTEP devices are disposable. In additional aspects, the FTEP devices may be autoclavable.

[00225] The electrodes 508 can be formed from any suitable metal, such as copper, stainless steel, titanium, aluminum, brass, silver, rhodium, gold or platinum, or graphite. One preferred electrode material is alloy 303 (UNS330300) austenitic stainless steel. An applied electric field can destroy electrodes made from of metals like aluminum. If a multiple-use (*e.g.*, non-disposable) flow-through FTEP device is desired-as opposed to a disposable, one-use flow-through FTEP device-the electrode plates can be coated with

metals resistant to electrochemical corrosion. Conductive coatings like noble metals, *e.g.*, gold, can be used to protect the electrode plates.

[00226] As mentioned, the FTEP devices may comprise push-pull pneumatic means to allow multi-pass electroporation procedures; that is, cells to electroporated may be "pulled" from the inlet toward the outlet for one pass of electroporation, then be "pushed" from the outlet end of the flow-through FTEP device toward the inlet end to pass between the electrodes again for another pass of electroporation. This process may be repeated one to many times.

[00227] Depending on the type of cells to be electroporated (*e.g.*, bacterial, yeast, mammalian) and the configuration of the electrodes, the distance between the electrodes in the flow channel can vary widely. For example, where the flow channel decreases in width, the flow channel may narrow to between 10 μm and 5 mm, or between 25 μm and 3 mm, or between 50 μm and 2 mm, or between 75 μm and 1 mm. The distance between the electrodes in the flow channel may be between 1 mm and 10 mm, or between 2 mm and 8 mm, or between 3 mm and 7 mm, or between 4 mm and 6 mm. The overall size of the FTEP device may be between 3 cm and 15 cm in length, or between 4 cm and 12 cm in length, or between 4.5 cm and 10 cm in length. The overall width of the FTEP device may be between 0.5 cm and 5 cm, or between 0.75 cm and 3 cm, or between 1 cm and 2.5 cm, or between 1 cm and 1.5 cm.

[00228] The region of the flow channel that is narrowed is wide enough so that at least two cells can fit in the narrowed portion side-by-side. For example, a typical bacterial cell is 1 μm in diameter; thus, the narrowed portion of the flow channel of the FTEP device used to transform such bacterial cells will be at least 2 μm wide. In another example, if a mammalian cell is approximately 50 μm in diameter, the narrowed portion of the flow channel of the FTEP device used to transform such mammalian cells will be at least 100 μm wide. That is, the narrowed portion of the FTEP device will not physically contort or "squeeze" the cells being transformed.

[00229] In aspects of the FTEP device where reservoirs are used to introduce cells and exogenous material into the FTEP device, the reservoirs range in volume from between 100 μL and 10 mL, or between 500 μL and 75 mL, or between 1 mL and 5 mL. The flow rate in the FTEP ranges from between 0.1 mL and 5 mL per minute, or between 0.5 mL

and 3 mL per minute, or between 1.0 mL and 2.5 mL per minute. The pressure in the FTEP device ranges from between 1 and 30 PSI, or between 2 and 10 PSI, or between 3 and 5 PSI.

[00230] To avoid different field intensities between the electrodes, the electrodes should be arranged in parallel. Furthermore, the surface of the electrodes should be as smooth as possible without pin holes or peaks. Electrodes having a roughness Rz of between 1 μm and 10 μm are preferred. In another aspect of the invention, the flow-through electroporation device comprises at least one additional electrode which applies a ground potential to the FTEP device.

10

Cell Singulation and Enrichment Device

[00231] FIG. 6A depicts a solid wall device 6050 and a workflow for singulating cells in microwells in the solid wall device. At the top left of the figure (i), there is depicted solid wall device 6050 with microwells 6052. A section 6054 of substrate 6050 is shown at (ii), also depicting microwells 6052. At (iii), a side cross-section of solid wall device 6050 is shown, and microwells 6052 have been loaded, where, in this aspect, Poisson or substantial Poisson loading has taken place; that is, each microwell has one or no cells, and the likelihood that any one microwell has more than one cell is low. At (iv), workflow 6040 is illustrated where substrate 6050 having microwells 6052 shows microwells 6056 with one cell per microwell, microwells 6057 with no cells in the microwells, and one microwell 6060 with two cells in the microwell. In step 6051, the cells in the microwells are allowed to double approximately 2-150 times to form clonal colonies (v), then editing is allowed to occur 6053.

15

20

[00232] After editing 6053, many cells in the colonies of cells that have been edited die as a result of the nicks caused by active editing or by fitness effects from the edits themselves and there is a lag in growth for the edited cells that do survive but must repair and recover following editing (microwells 6058), where cells that do not undergo editing thrive (microwells 6059) (vi). All cells are allowed to continue grow to establish colonies and normalize, where the colonies of edited cells in microwells 6058 catch up in size and/or cell number with the cells in microwells 6059 that do not undergo editing (vii). Once the cell colonies are normalized, either pooling 6060 of all cells in the microwells can take

30

place, in which case the cells are enriched for edited cells by eliminating the bias from non-editing cells and fitness effects from editing; alternatively, colony growth in the microwells is monitored after editing, and slow growing colonies (*e.g.*, the cells in microwells 6058) are identified and selected 6061 (*e.g.*, “cherry picked”) resulting in even greater enrichment
5 of edited cells.

[00233] In growing the cells, the medium used will depend, on the type of cells being edited—*e.g.*, bacterial, yeast or mammalian. For example, medium for yeast cell growth includes LB, SOC, TPD, YPG, YPAD, MEM and DMEM.

[00234] A module useful for performing the method depicted in FIG. 6A is a solid
10 wall isolation, incubation, and normalization (SWIIN) module. FIG. 6B depicts an aspect of a SWIIN module 650 from an exploded top perspective view. In SWIIN module 650 the retentate member is formed on the bottom of a top of a SWIIN module component and the permeate member is formed on the top of the bottom of a SWIIN module component.

[00235] The SWIIN module 650 in FIG. 6B comprises from the top down, a reservoir
15 gasket or cover 658, a retentate member 604 (where a retentate flow channel cannot be seen in this FIG. 6B), a perforated member 601 swaged with a filter (filter not seen in FIG. 6B), a permeate member 608 comprising integrated reservoirs (permeate reservoirs 652 and retentate reservoirs 654), and two reservoir seals 662, which seal the bottom of permeate reservoirs 652 and retentate reservoirs 654. A permeate channel 660a can be seen
20 disposed on the top of permeate member 608, defined by a raised portion 676 of serpentine channel 660a, and ultrasonic tabs 664 can be seen disposed on the top of permeate member 608 as well. The perforations that form the wells on perforated member 601 are not seen in this FIG. 6B; however, through-holes 666 to accommodate the ultrasonic tabs 664 are seen. In addition, supports 670 are disposed at either end of SWIIN module 650 to support
25 SWIIN module 650 and to elevate permeate member 608 and retentate member 604 above reservoirs 652 and 654 to minimize bubbles or air entering the fluid path from the permeate reservoir to serpentine channel 660a or the fluid path from the retentate reservoir to serpentine channel 660b (neither fluid path is seen in this FIG. 6B).

[00236] In this FIG. 6B, it can be seen that the serpentine channel 660a that is disposed
30 on the top of permeate member 608 traverses permeate member 608 for most of the length of permeate member 608 except for the portion of permeate member 608 that comprises

permeate reservoirs 652 and retentate reservoirs 654 and for most of the width of permeate member 608. As used herein with respect to the distribution channels in the retentate member or permeate member, “most of the length” means about 95% of the length of the retentate member or permeate member, or about 90%, 85%, 80%, 75%, or 70% of the length of the retentate member or permeate member. As used herein with respect to the distribution channels in the retentate member or permeate member, “most of the width” means about 95% of the width of the retentate member or permeate member, or about 90%, 85%, 80%, 75%, or 70% of the width of the retentate member or permeate member.

[00237] In this aspect of a SWIIN module, the perforated member includes through-holes to accommodate ultrasonic tabs disposed on the permeate member. Thus, in this aspect the perforated member is fabricated from 316 stainless steel, and the perforations form the walls of microwells while a filter or membrane is used to form the bottom of the microwells. Typically, the perforations (microwells) are approximately 150 μm to 200 μm in diameter, and the perforated member is approximately 125 μm deep, resulting in microwells having a volume of approximately 2.5 nL, with a total of approximately 200,000 microwells. The distance between the microwells is approximately 279 μm center-to-center. Though here the microwells have a volume of approximately 2.5 nL, the volume of the microwells may be between 1 nL and 25 nL, or preferably between 2 nL and 10 nL, and even more preferably between 2 nL and 4 nL. As for the filter or membrane, like the filter described previously, filters appropriate for use are solvent resistant, contamination free during filtration, and are able to retain the types and sizes of cells of interest. For example, in order to retain small cell types such as bacterial cells, pore sizes can be as low as 0.10 μm , however for other cell types (*e.g.*, such as for mammalian cells), the pore sizes can be as high as from 10.0 μm to 20.0 μm or more. Indeed, the pore sizes useful in the cell concentration device/module include filters with sizes from 0.10 μm , 0.11 μm , 0.12 μm , 0.13 μm , 0.14 μm , 0.15 μm , 0.16 μm , 0.17 μm , 0.18 μm , 0.19 μm , 0.20 μm , 0.21 μm , 0.22 μm , 0.23 μm , 0.24 μm , 0.25 μm , 0.26 μm , 0.27 μm , 0.28 μm , 0.29 μm , 0.30 μm , 0.31 μm , 0.32 μm , 0.33 μm , 0.34 μm , 0.35 μm , 0.36 μm , 0.37 μm , 0.38 μm , 0.39 μm , 0.40 μm , 0.41 μm , 0.42 μm , 0.43 μm , 0.44 μm , 0.45 μm , 0.46 μm , 0.47 μm , 0.48 μm , 0.49 μm , 0.50 μm and larger. The filters may be fabricated from any suitable material including cellulose mixed ester (cellulose nitrate and acetate) (CME), polycarbonate (PC), polyvinylidene

fluoride (PVDF), polyethersulfone (PES), polytetrafluoroethylene (PTFE), nylon, or glass fiber.

[00238] The cross-section configuration of the mated serpentine channel may be round, elliptical, oval, square, rectangular, trapezoidal, or irregular. If square, rectangular, or another shape with generally straight sides, the cross section may be between about 2 mm and 15 mm wide, or between 3 mm and 12 mm wide, or between 5 mm and 10 mm wide. If the cross section of the mated serpentine channel is generally round, oval or elliptical, the radius of the channel may be between about 3 mm and 20 mm in hydraulic radius, or between 5 mm and 15 mm in hydraulic radius, or between 8 mm and 12 mm in hydraulic radius.

[00239] Serpentine channels 660a and 660b can have approximately the same volume or a different volume. For example, each “side” or portion 660a, 660b of the serpentine channel may have a volume of, *e.g.*, 2 mL, or serpentine channel 660a of permeate member 608 may have a volume of 2 mL, and the serpentine channel 660b of retentate member 604 may have a volume of, *e.g.*, 3 mL. The volume of fluid in the serpentine channel may range from about 2 mL to about 80 mL, or from about 4 mL to about 60 mL, or from about 5 mL to about 40 mL, or from about 6 mL to about 20 mL (note these volumes apply to a SWIIN module comprising a, *e.g.*, 50-500K perforation member). The volume of the reservoirs may range between 5 mL and 50 mL, or between 7 mL and 40 mL, or between 8 mL and 30 mL or between 10 mL and 20 mL, and the volumes of all reservoirs may be the same or the volumes of the reservoirs may differ (*e.g.*, the volume of the permeate reservoirs is greater than that of the retentate reservoirs).

[00240] The serpentine channel portions 660a and 660b of the permeate member 608 and retentate member 604, respectively, are approximately 200 mm long, 130 mm wide, and 4 mm thick, though in other aspects, the retentate and permeate members can be between 75 mm and 400 mm in length, or between 100 mm and 300 mm in length, or between 150 mm and 250 mm in length; between 50 mm and 250 mm in width, or between 75 mm and 200 mm in width, or between 100 mm and 150 mm in width; and between 2 mm and 15 mm in thickness, or between 4 mm and 10 mm in thickness, or between 5 mm and 8 mm in thickness. In some aspects, the retentate (and permeate) members may be fabricated from PMMA (poly(methyl methacrylate) or other materials may be used,

including polycarbonate, cyclic olefin co-polymer (COC), glass, polyvinyl chloride, polyethylene, polyamide, polypropylene, polysulfone, polyurethane, and co-polymers of these and other polymers. Preferably at least the retentate member is fabricated from a transparent material so that the cells can be visualized (see, *e.g.*, FIG. 6E and the description thereof). For example, a video camera may be used to monitor cell growth by, *e.g.*, density change measurements based on an image of an empty well, with phase contrast, or if, *e.g.*, a chromogenic marker, such as a chromogenic protein, is used to add a distinguishable color to the cells. Chromogenic markers such as blitzen blue, dreidel teal, virginia violet, vixen purple, prancer purple, tinsel purple, maccabee purple, donner magenta, cupid pink, seraphina pink, scrooge orange, and leor orange (the Chromogenic Protein Paintbox, all available from ATUM (Newark, CA)) obviate the need to use fluorescence, although fluorescent cell markers, fluorescent proteins, and chemiluminescent cell markers may also be used.

[00241] Because the retentate member preferably is transparent, colony growth in the SWIIN module can be monitored by automated devices such as those sold by JoVE (ScanLag™ system, Cambridge, MA) (also see Levin-Reisman, et al., *Nature Methods*, 7:737-39 (2010)). Cell growth for, *e.g.*, mammalian cells may be monitored by, *e.g.*, the growth monitor sold by IncuCyte (Ann Arbor, MI) (see also, Choudhry, *PLoS One*, 11(2):e0148469 (2016)). Further, automated colony pickers may be employed, such as those sold by, *e.g.*, TECAN (Pickolo™ system, Mannedorf, Switzerland); Hudson Inc. (RapidPick™, Springfield, NJ); Molecular Devices (QPix 400™ system, San Jose, CA); and Singer Instruments (PIXL™ system, Somerset, UK).

[00242] Due to the heating and cooling of the SWIIN module, condensation may accumulate on the retentate member which may interfere with accurate visualization of the growing cell colonies. Condensation of the SWIIN module 650 may be controlled by, *e.g.*, moving heated air over the top of (*e.g.*, retentate member) of the SWIIN module 650, or by applying a transparent heated lid over at least the serpentine channel portion 660b of the retentate member 604. See, *e.g.*, FIG. 6E and the description thereof *infra*.

[00243] In SWIIN module 650 cells and medium—at a dilution appropriate for Poisson or substantial Poisson distribution of the cells in the microwells of the perforated member—are flowed into serpentine channel 660b from ports in retentate member 604,

and the cells settle in the microwells while the medium passes through the filter into serpentine channel 660a in permeate member 608. The cells are retained in the microwells of perforated member 601 as the cells cannot travel through filter 603. Appropriate medium may be introduced into permeate member 608 through permeate ports 611. The medium
5 flows upward through filter 603 to nourish the cells in the microwells (perforations) of perforated member 601. Additionally, buffer exchange can be effected by cycling medium through the retentate and permeate members. In operation, the cells are deposited into the microwells, are grown for an initial, *e.g.*, between 2 and 100 doublings, editing is induced by, *e.g.*, raising the temperature of the SWIIN to 42°C to induce a temperature inducible
10 promoter or by removing growth medium from the permeate member and replacing the growth medium with a medium comprising a chemical component that induces an inducible promoter.

[00244] Once editing has taken place, the temperature of the SWIIN may be decreased, or the inducing medium may be removed and replaced with fresh medium
15 lacking the chemical component thereby de-activating the inducible promoter. The cells then continue to grow in the SWIIN module 650 until the growth of the cell colonies in the microwells is normalized. For the normalization protocol, once the colonies are normalized, the colonies are flushed from the microwells by applying fluid or air pressure (or both) to the permeate member serpentine channel 660a and thus to filter 603 and pooled.
20 Alternatively, if cherry picking is desired, the growth of the cell colonies in the microwells is monitored, and slow-growing colonies are directly selected; or, fast-growing colonies are eliminated.

[00245] FIG. 6C is a top perspective view of a SWIIN module with the retentate and perforated members in partial cross section. In this FIG. 6C, it can be seen that serpentine
25 channel 660a is disposed on the top of permeate member 608 is defined by raised portions 676 and traverses permeate member 608 for most of the length and width of permeate member 608 except for the portion of permeate member 608 that comprises the permeate and retentate reservoirs (note only one retentate reservoir 652 can be seen). Moving from left to right, reservoir gasket 658 is disposed upon the integrated reservoir cover 678 (cover
30 not seen in this FIG. 6C) of retentate member 604. Gasket 658 comprises reservoir access apertures 632a, 632b, 632c, and 632d, as well as pneumatic ports 633a, 633b, 633c and

633d. Also at the far left end is support 670. Disposed under permeate reservoir 652 can be seen one of two reservoir seals 662. In addition to the retentate member being in cross section, the perforated member 601 and filter 603 (filter 603 is not seen in this FIG. 6C) are in cross section. Note that there are a number of ultrasonic tabs 664 disposed at the right end of SWIIN module 650 and on raised portion 676 which defines the channel turns of serpentine channel 660a, including ultrasonic tabs 664 extending through through-holes 666 of perforated member 601. There is also a support 670 at the end distal reservoirs 652, 654 of permeate member 608.

[00246] FIG. 6D is a side perspective view of an assembled SWIIN module 650, including, from right to left, reservoir gasket 658 disposed upon integrated reservoir cover 678 (not seen) of retentate member 604. Gasket 658 may be fabricated from rubber, silicone, nitrile rubber, polytetrafluoroethylene, a plastic polymer such as polychlorotrifluoroethylene, or other flexible, compressible material. Gasket 658 comprises reservoir access apertures 632a, 632b, 632c, and 632d, as well as pneumatic ports 633a, 633b, 633c and 633d. Also at the far-left end is support 670 of permeate member 608. In addition, permeate reservoir 652 can be seen, as well as one reservoir seal 662. At the far-right end is a second support 670.

[00247] Imaging of cell colonies growing in the wells of the SWIIN is desired in most implementations for, *e.g.*, monitoring both cell growth and device performance and imaging is necessary for cherry-picking implementations. Real-time monitoring of cell growth in the SWIIN requires backlighting, retentate plate (top plate) condensation management and a system-level approach to temperature control, air flow, and thermal management. In some implementations, imaging employs a camera or CCD device with sufficient resolution to be able to image individual wells. For example, in some configurations a camera with a 9-pixel pitch is used (that is, there are 9 pixels center-to-center for each well). Processing the images may, in some implementations, utilize reading the images in grayscale, rating each pixel from low to high, where wells with no cells will be brightest (due to full or nearly-full light transmission from the backlight) and wells with cells will be dim (due to cells blocking light transmission from the backlight). After processing the images, thresholding is performed to determine which pixels will be called “bright” or “dim”, spot finding is performed to find bright pixels and arrange them into

blocks, and then the spots are arranged on a hexagonal grid of pixels that correspond to the spots. Once arranged, the measure of intensity of each well is extracted, by, *e.g.*, looking at one or more pixels in the middle of the spot, looking at several to many pixels at random or pre-set positions, or averaging X number of pixels in the spot. In addition, background
5 intensity may be subtracted. Thresholding is again used to call each well positive (*e.g.*, containing cells) or negative (*e.g.*, no cells in the well). The imaging information may be used in several ways, including taking images at time points for monitoring cell growth. Monitoring cell growth can be used to, *e.g.*, remove the “muffin tops” of fast-growing cells followed by removal of all cells or removal of cells in “rounds” as described above, or
10 recover cells from specific wells (*e.g.*, slow-growing cell colonies); alternatively, wells containing fast-growing cells can be identified and areas of UV light covering the fast-growing cell colonies can be projected (or rastered with shutters) onto the SWIIN to irradiate or inhibit growth of those cells. Imaging may also be used to assure proper fluid flow in the serpentine channel 660.

15 [00248] FIG. 6E depicts the aspect of the SWIIN module in FIGs. 6B – 6D further comprising a heat management system including a heater and a heated cover. The heater cover facilitates the condensation management that is required for imaging. Assembly 698 comprises a SWIIN module 650 seen lengthwise in cross section, where one permeate reservoir 652 is seen. Disposed immediately upon SWIIN module 650 is cover 694 and
20 disposed immediately below SWIIN module 650 is backlight 680, which allows for imaging. Beneath and adjacent to the backlight and SWIIN module is insulation 682, which is disposed over a heatsink 684. In this FIG. 6E, the fins of the heatsink would be in-out of the page. In addition there is also axial fan 686 and heat sink 688, as well as two thermoelectric coolers 692, and a controller 690 to control the pneumatics, thermoelectric
25 coolers, fan, solenoid valves, etc. The arrows denote cool air coming into the unit and hot air being removed from the unit. It should be noted that control of heating allows for growth of many different types of cells (prokaryotic and eukaryotic) as well as strains of cells that are, *e.g.*, temperature sensitive, etc., and allows use of temperature-sensitive promoters. Temperature control allows for protocols to be adjusted to account for differences in
30 transformation efficiency, cell growth and viability. For more details regarding solid wall isolation incubation and normalization devices see USSNs 16/399,988, filed 30 April 2019;

16/454,865, filed 26 June 2019; and 16/540,606, filed 14 August 2019. For alternative isolation, incubation and normalization modules, see USSN 16/536,049, filed 08 August 2019.

5 Use of the Automated Multi-Module Cell Processing Instrument

[00249] FIG. 7 illustrates an aspect of a multi-module cell processing instrument. This aspect depicts an example of a system that performs recursive and trackable CFgRNA/nickase-RT fusion editing on a cell population. The cell processing instrument 700 may include a housing 726, a reservoir for storing cells to be transformed or transfected 10 702, and a cell growth module (comprising, *e.g.*, a rotating growth vial) 704. The cells to be transformed are transferred from a reservoir 702 to the cell growth module 704 to be cultured until the cells hit a target OD. Once the cells hit the target OD, the growth module may cool or freeze the cells for later processing or transfer the cells to a cell concentration (*e.g.*, filtration) module 706 where the cells are subjected to buffer exchange and rendered 15 electrocompetent and the volume of the cells may be reduced substantially. Once the cells have been concentrated to an appropriate volume, the cells are transferred to electroporation device 708 or other transformation module. In addition to the reservoir for storing cells 702, the multi-module cell processing instrument includes a reservoir for storing the engine and editing vectors or engine + editing vectors or vectors and proteins 20 to be introduced into the electrocompetent cell population 722. The vector backbones and editing cassettes are transferred to the electroporation device 708, which already contains the cell culture grown to a target OD. In the electroporation device 708, the nucleic acids (or nucleic acids and proteins) are electroporated into the cells. Following electroporation, the cells are transferred into an optional recovery and dilution module 710, where the cells 25 recover briefly post-transformation.

[00250] After recovery, the cells may be transferred to a storage module 712, where the cells can be stored at, *e.g.*, 4°C or -20°C for later processing, or the cells may be diluted and transferred to a selection/singulation/growth/induction/editing/normalization (SWIIN) module 720. In the SWIIN 720, the cells are arrayed such that there is an average of one to 30 twenty or fifty or so cells per microwell. The arrayed cells may be in selection medium to select for cells that have been transformed or transfected with the editing vector(s). Once

singulated, the cells grow through 2 to 50 doublings and establish colonies. Once colonies are established, editing is induced by providing conditions (*e.g.*, temperature, addition of an inducing or repressing chemical) to induce editing. Editing is then initiated and allowed to proceed, the cells are allowed to grow to terminal size (*e.g.*, normalization of the colonies) in the microwells and then are treated to conditions that cure the editing vector from this round. Once cured, the cells can be flushed out of the microwells and pooled, then transferred to the storage (or recovery) unit 712 or can be transferred back to the growth module 704 for another round of editing. In between pooling and transfer to a growth module, there typically is one or more additional steps, such as cell recovery, medium exchange (rendering the cells electrocompetent), cell concentration (typically concurrently with medium exchange by, *e.g.*, filtration).

[00251] Note that the selection/singulation/growth/induction/editing/normalization and curing modules may be the same module, where all processes are performed in, *e.g.*, a solid wall device, or selection and/or dilution may take place in a separate vessel before the cells are transferred to the solid wall singulation/growth/induction/editing/normalization/editing module (SWIIN). Similarly, the cells may be pooled after normalization, transferred to a separate vessel, and cured in the separate vessel. Once the putatively-edited cells are pooled, they may be subjected to another round of editing, beginning with growth, cell concentration and treatment to render electrocompetent, and transformation by yet another donor nucleic acid in another editing cassette via the electroporation module 708.

[00252] In electroporation device 708, the cells selected from the first round of editing are transformed by a second set of editing vectors and the cycle is repeated until the cells have been transformed and edited by a desired number of, *e.g.*, CF editing cassettes. The multi-module cell processing instrument exemplified in FIG. 7 is controlled by a processor 724 configured to operate the instrument based on user input or is controlled by one or more scripts including at least one script associated with the reagent cartridge. The processor 724 may control the timing, duration, and temperature of various processes, the dispensing of reagents, and other operations of the various modules of the instrument 700. For example, a script or the processor may control the dispensing of cells, reagents, vectors, and editing oligonucleotides; which editing oligonucleotides are used for cell editing and

in what order; the time, temperature and other conditions used in the recovery and expression module, the wavelength at which OD is read in the cell growth module, the target OD to which the cells are grown, and the target time at which the cells will reach the target OD. In addition, the processor may be programmed to notify a user (*e.g.*, via an application) as to the progress of the cells in the automated multi-module cell processing instrument.

5 [00253] It should be apparent to one of ordinary skill in the art given the present disclosure that the process described may be recursive and multiplexed; that is, cells may go through the workflow described in relation to FIG. 7, then the resulting edited culture may go through another (or several or many) rounds of additional editing (*e.g.*, recursive editing) with different editing cassettes (or ribozyme-containing editing cassettes). For example, the cells from round 1 of editing may be diluted and an aliquot of the edited cells edited by editing cassette A may be combined with editing cassette B, an aliquot of the edited cells edited by editing cassette A may be combined with editing cassette C, an aliquot of the edited cells edited by editing cassette A may be combined with editing cassette D, and so on for a second round of editing. After round two, an aliquot of each of the double-edited cells may be subjected to a third round of editing, where, *e.g.*, aliquots of each of the AB-, AC-, AD-edited cells are combined with additional editing cassettes, such as editing cassettes X, Y, and Z. That is, double-edited cells AB may be combined with and edited by editing cassettes X, Y, and Z to produce triple-edited edited cells ABX, ABY, and ABZ; double-edited cells AC may be combined with and edited by editing cassettes X, Y, and Z to produce triple-edited cells ACX, ACY, and ACZ; and double-edited cells AD may be combined with and edited by editing cassettes X, Y, and Z to produce triple-edited cells ADX, ADY, and ADZ, and so on. In this process, many permutations and combinations of edits can be executed, leading to very diverse cell populations and cell libraries.

15 20 25 30 [00254] In any recursive process, it is advantageous to “cure” the editing vectors comprising the CF editing cassette. “Curing” is a process in which one or more editing vectors used in the prior round of editing is eliminated from the transformed cells. (See, *e.g.*, curing can be accomplished by, *e.g.*, cleaving the editing vector(s) using a curing plasmid thereby rendering the editing vectors nonfunctional; diluting the editing vector(s)

in the cell population via cell growth (that is, the more growth cycles the cells go through, the fewer daughter cells will retain the editing vector(s)), or by, *e.g.*, utilizing a heat-sensitive origin of replication on the editing vector. The conditions for curing will depend on the mechanism used for curing; that is, in this example, how the curing plasmid cleaves the editing vector.

[00255] A variety of further modifications and improvements in and to the composition, methods, and modified cells of the present disclosure will be apparent to those skilled in the art. The following non-limiting, embodiments are specifically envisioned:

1. A method for performing nucleic acid-guided nuclease/reverse transcriptase fusion editing in a genome of a live cell, comprising:

(a) providing the live cell, wherein the live cell comprises a first target locus and a second target locus;

(b) providing a nucleic acid-guided nuclease/reverse transcriptase fusion enzyme;

(c) providing a CF editing cassette, the CF editing cassette comprising:

(i) a nucleic acid sequence encoding a first CFgRNA having a region of complementarity to a sequence of the first target locus; and

(ii) a nucleic acid sequence encoding a first repair template;

(d) providing a CF barcoding cassette, the CF barcoding cassette comprising:

(i) a nucleic acid sequence encoding a second CFgRNA having a region of complementarity to a sequence of the second target locus; and

(ii) a nucleic acid sequence encoding a second repair template;

(e) providing conditions to allow the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme, the first CFgRNA, and the first repair template to bind to the first target locus;

(f) allowing the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme, the first CFgRNA, and the first repair template to edit the first target locus;

(g) providing conditions to allow the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme, the second CFgRNA, and the second repair template to bind to the second target locus; and

(h) allowing the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme, the second CFgRNA, and the second repair template to integrate the barcode into the second target locus.

2. The method of embodiment 1, wherein the first CFgRNA comprises a spacer region and a structural region recognized by the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme.
3. The method of embodiment 1 or 2, wherein the first repair template comprises an edit and a primer binding site (PBS).
4. The method of embodiment 3, wherein the first repair template further comprises a post-edit homology region.
5. The method of embodiment 3 or 4, wherein the first repair template comprises a nick-to-edit region.
6. The method of any one of embodiments 1 to 5, wherein the second CFgRNA comprises a spacer region and a structural region recognized by the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme.
7. The method of any one of embodiments 1 to 6, wherein the second repair template comprises a barcode and a primer binding site (PBS).
8. The method of embodiment 7, wherein the second repair template further comprises a post-barcode homology region.
9. The method of embodiment 7 or 8, wherein the second repair template comprises a nick-to-barcode region.
10. The method of any one of embodiments 1 to 9, further comprising:
sequencing the genome or a transcriptome of the cell to track for integration of the barcode, the integration of the barcode representing a nucleic acid-guided nuclease/reverse transcriptase fusion editing event effected by the CF editing cassette.
11. The method of embodiment 10, wherein the genome is sequenced by single-cell amplicon-based next-generation sequencing.
12. The method of embodiment 10, wherein the transcriptome is sequenced by single-cell RNA sequencing.

13. The method of any one of embodiments 1 to 12, wherein the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme comprises a nucleic acid-guided nickase and a reverse transcriptase.
14. The method of embodiment 13, wherein the nucleic acid-guided nickase
5 comprises a MAD nickase or a variant thereof.
15. The method of embodiment 13, wherein the nucleic acid-guided nickase comprises a Cas nickase or a variant thereof.
16. The method of any one of embodiments 1 to 15, wherein a nucleic acid sequence encoding the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme and the
10 CF editing cassette are assembled into a vector for introduction of the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme and CF editing cassette into the cell.
17. The method of embodiment 16, wherein the CF barcoding cassette is further assembled into the vector for introduction of the CF barcoding cassette into the cell.
18. The method of any one of embodiments 1 to 15, wherein the CF barcoding
15 cassette is assembled into a barcoding vector for introduction of the CF barcoding cassette into the cell, and wherein the CF editing cassette is assembled into an editing vector for introduction of the CF editing cassette into the cell, wherein the editing vector is different from the barcoding vector.
19. The method of any one of embodiments 1 to 18, further comprising:
20 providing a selectable marker.
20. The method of embodiment 19, wherein the selectable marker is for selection and enrichment of cells having an integrated barcode or an effected edit.
21. The method of embodiment 19 or 20, further comprising selecting and enriching for cells having an integrated barcode or an effected edit.
- 25 22. The method of any one of embodiments 1 to 21, wherein the second target locus is a safe harbor locus disposed centrally in an intergenic or intronic region of the cell.
23. The method of any one of embodiments 1 to 21, wherein the second target locus is disposed within a coding region of the cell.
24. The method of any one of embodiments 1 to 21, wherein the second target locus
30 is disposed within a noncoding region of the cell.

25. The method of any one of embodiments 1 to 24, wherein the CF editing cassette further comprises an edit to immunize the first target locus and prevent re-nicking.
26. The method of any one of embodiments 1 to 25, wherein the CF barcoding cassette further comprises an edit to immunize the second target locus and prevent re-nicking.
27. An editing system comprising one or more vectors comprising:
a nucleic acid sequence encoding a nucleic acid-guided nuclease/reverse transcriptase fusion enzyme;
a CF editing cassette comprising:
10 a nucleic acid sequence encoding a first CFgRNA having a region of complementarity to a sequence of a first target locus in a cell; and
a nucleic acid sequence encoding a first repair template;
a CF barcoding cassette comprising:
a nucleic acid sequence encoding a second CFgRNA having a region of
15 complementarity to a sequence of a second target locus in the cell; and
a nucleic acid sequence encoding a second repair template.
28. The editing system of embodiment 27, wherein the first CFgRNA comprises a spacer region and a structural region recognized by the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme.
29. The editing system of embodiment 27 or 28, wherein the first repair template
20 comprises an edit and a primer binding site (PBS).
30. The editing system of embodiment 29, wherein the first repair template further comprises a post-edit homology region.
31. The editing system of embodiment 29 or 30, wherein the first repair template
25 further comprises a nick-to-edit region.
32. The editing system of any one of embodiments 27 to 31, wherein the second CFgRNA comprises a spacer region and a structural region recognized by the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme.
33. The editing system of any one of embodiments 27 to 32, wherein the second
30 repair template comprises a barcode and a primer binding site (PBS).

34. The editing system of embodiment 33, wherein the second repair template further comprises a post-barcode homology region.
35. The editing system of embodiment 33 or 34, wherein the second repair template comprises a nick-to-barcode region.
- 5 36. The editing system of any one of embodiments 27 to 35, wherein the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme comprises a nucleic acid-guided nickase and a reverse transcriptase.
37. The editing system of embodiment 36, wherein the nucleic acid-guided nickase comprises a MAD nickase or a variant thereof.
- 10 38. The editing system of embodiment 36, wherein the nucleic acid-guided nickase comprises a Cas nickase or a variant thereof.
39. The editing system of any one of embodiments 27 to 38, wherein the one or more vectors comprises an editing vector, and wherein the editing vector comprises the CF editing cassette.
- 15 40. The editing system of embodiment 39, wherein the editing vector further comprises a nucleic acid sequence encoding the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme.
41. The editing system of embodiment 39 or 40, wherein the editing vector further comprises the CF barcoding cassette.
- 20 42. The editing system of any one of embodiments 27 to 38, wherein an editing vector comprises the CF editing cassette, and a barcoding vector comprises the CF barcoding cassette, and wherein the editing vector is different than the barcoding vector.
43. The editing system of any one of embodiments 27 to 38, wherein one or more of the one or more vectors comprises a selectable marker.
- 25 44. The editing system of embodiment 43, wherein the selectable marker is for selection and enrichment of cells having an integrated barcode or an effected edit.
45. The editing system of any one of embodiments 27 to 44, wherein the second target locus is a safe harbor locus disposed centrally in an intergenic or intronic region of the cell.
- 30 46. The editing system of any one of embodiments 27 to 44, wherein the second target locus is disposed within a coding region of the cell.

47. The editing system of any one of embodiments 27 to 44, wherein the second target locus is disposed within a noncoding region of the cell.
48. The editing system of any one of embodiments 27 to 47, wherein the CF editing cassette further comprises an edit to immunize the first target locus and prevent re-nicking.
- 5 49. The editing system of any one of embodiments 27 to 48, wherein the CF barcoding cassette further comprises an edit to immunize the second target locus and prevent re-nicking.
50. A vector comprising:
- 10 a nucleic acid sequence encoding a nucleic acid-guided nuclease/reverse transcriptase fusion enzyme;
- a CF editing cassette comprising:
- a nucleic acid sequence encoding a first CFgRNA having a region of complementarity to a sequence of a first target locus in a cell; and
- 15 a nucleic acid sequence encoding a first repair template;
- a CF barcoding cassette comprising:
- a nucleic acid sequence encoding a second CFgRNA having a region of complementarity to a sequence of a second target locus in the cell; and
- a nucleic acid sequence encoding a second repair template.
- 20 51. The vector of embodiment 50, wherein the first CFgRNA comprises a spacer region and a structural region recognized by the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme.
52. The vector of embodiment 50 or 51, wherein the first repair template comprises an edit and a primer binding site (PBS).
- 25 53. The vector of embodiment 52, wherein the first repair template further comprises a post-edit homology region.
54. The vector of embodiment 52 or 53, wherein the first repair template further comprises a nick-to-edit region.
55. The vector of any one of embodiments 50 to 54, wherein the second CFgRNA
- 30 comprises a spacer region and a structural region recognized by the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme.

56. The vector of any one of embodiments 50 to 55, wherein the second repair template comprises a barcode and a primer binding site (PBS).
57. The vector of any one of embodiments 56, wherein the second repair template further comprises a post-barcode homology region.
- 5 58. The vector of embodiment 56 or 57, wherein the second repair template comprises a nick-to-barcode region.
59. The vector of any one of embodiments 50 to 58, wherein the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme comprises a nucleic acid-guided nickase and a reverse transcriptase.
- 10 60. The vector of embodiment 59, wherein the nucleic acid-guided nickase comprises a MAD nickase or a variant thereof.
61. The vector of embodiment 59, wherein the nucleic acid-guided nickase comprises a Cas nickase or a variant thereof.
62. The vector of any one of embodiments 50 to 61, wherein the vector further
15 comprises a selectable marker.
63. The vector of embodiment 62, wherein the selectable marker is for selection and enrichment of cells having an integrated barcode or an effected edit.
64. The vector of embodiment 62 or 63, wherein the selectable marker is a puromycin resistance gene.
- 20 65. The vector of any one of embodiments 50 to 64, wherein the second target locus is a safe harbor locus disposed centrally in an intergenic or intronic region of the cell.
66. The vector of any one of embodiments 50 to 64, wherein the second target locus is disposed within a coding region of the cell.
67. The vector of any one of embodiments 50 to 64, wherein the second target locus
25 is disposed within a noncoding region of the cell.
68. The vector of any one of embodiments 50 to 67, wherein the CF editing cassette further comprises an edit to immunize the first target locus and prevent re-nicking.
69. The vector of any one of embodiments 50 to 68, wherein the CF barcoding cassette further comprises an edit to immunize the second target locus and prevent re-
30 nicking.

70. A method for performing nucleic acid-guided nuclease/reverse transcriptase fusion editing in a genome of a live cell, comprising:

(a) providing a live cell suitable for the editing;

(b) introducing a nucleic acid-guided nuclease/reverse transcriptase fusion
5 enzyme;

(c) providing conditions to allow the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme to bind to a first target locus;

(d) allowing the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme to edit the first target locus;

10 (e) providing conditions to allow the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme to bind to a second target locus;

(f) allowing the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme to integrate a barcode into the second target locus.

71. The method of embodiment 70, wherein the live cell comprises a CF editing
15 cassette, the CF editing cassette comprising a nucleic acid sequence encoding a first CFgRNA having a region of complementarity to a sequence of the first target locus, and a nucleic acid sequence encoding a first repair template.

72. The method of embodiment 71, wherein the first CFgRNA comprises a spacer
20 region and a structural region recognized by the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme.

73. The method of any one of embodiments 70 to 72, wherein the first repair template comprises an edit and a primer binding site (PBS).

74. The method of any one of embodiments 70 to 73, wherein the live cell comprises a
25 CF barcoding cassette, the CF barcoding cassette comprising a nucleic acid sequence encoding a second CFgRNA having a region of complementarity to a sequence of the second target locus, and a nucleic acid sequence encoding a second repair template.

75. The method of embodiment 74, wherein the second CFgRNA comprises a spacer
region and a structural region recognized by the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme.

30 76. The method of any one of embodiments 70 to 75, wherein the second repair template comprises a barcode and a primer binding site (PBS).

EXAMPLES

[00256] The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the present invention, and are not intended to limit the scope of what the inventors regard as their invention, nor are they intended to represent or imply that the experiments below are all of or the only experiments performed. It will be appreciated by persons skilled in the art that numerous variations and/or modifications may be made to the invention as shown in the specific aspects without departing from the spirit or scope of the invention as broadly described. The present aspects are, therefore, to be considered in all respects as illustrative and not restrictive.

Example I: GFP to BFP Conversion Assay

[00257] A GFP to BFP reporter cell line is created using mammalian cells with a stably integrated genomic copy of the GFP gene (HEK293T-GFP). These cell lines enable phenotypic detection of genomic edits of different classes by various different mechanisms, including flow cytometry, fluorescent cell imaging, and genotypic detection by sequencing of the genome-integrated GFP gene. Lack of editing, or perfect repair of cut events in the GFP gene, result in cells that remain GFP-positive. Cut events that are repaired by the Non-Homologous End-Joining (NHEJ) pathway often result in nucleotide insertion or deletion events (indels), resulting in frame-shift mutations in the coding sequence that cause loss of GFP gene expression and fluorescence. Cut events that are repaired by the Homology-Directed Repair (HDR) pathway using the GFP to BFP HDR donor as a repair template or by the use of CFgRNAs, *e.g.*, complementary CFgRNAs, result in conversion of the cell fluorescence profile from that of GFP to that of BFP.

Example II: CREATE Fusion Editing—MAD2007 Nickase

[00258] CREATE Fusion Editing (CFE) is a technique that uses a nucleic acid nickase fusion protein (*e.g.*, MAD2007 nickase) fused to a peptide with reverse transcriptase activity along with a nucleic acid encoding a gRNA comprising a region complementary to a target region of a nucleic acid in one or more cells, which comprises a mutation of at

least one nucleotide relative to the target region in the one or more cells and a protospacer adjacent motif (PAM) mutation.

[00259] In a first design, a nickase enzyme derived from the MAD2007 nuclease (*see, e.g.*, USPNs 9,982,279 and 10,337,028), *e.g.*, Cas9 H840A nickase or MAD7® nickase (5 *see, e.g.*, USSNs 16/837,212 and 17/084,522), is fused to an engineered reverse transcriptase (RT) on the C-terminus and cloned downstream of a CMV promoter. In this instance, the RT used is derived from Moloney Murine Leukemia Virus (M-MLV).

[00260] RNA guides (gRNAs) are designed that are complementary to a single region proximal to the EGFP-to-BFP editing site. The gRNA is extended on the 3' end to include 10 a region of 13 bp that include the TY-to-SH edit and a second region of 13 bp that is complementary to the nicked EGFP DNA sequence. This allows the nicked genomic DNA to anneal to the 3' end of the gRNA which can then be extended by the reverse transcriptase to incorporate the edit in the genome. A second gRNA targets a region in the EGFP DNA sequence that is 86 bp upstream of the edit site. This gRNA is designed such that it enables 15 the nickase to cut the opposite strand relative to gRNA. Both of these gRNAs are cloned downstream of a U6 promoter. A poly-T sequence is also included that terminates the transcription of the gRNA.

[00261] The plasmids are transformed into NEB Stable *E. coli* (Ipswich, NY) and grown overnight in 25 mL LB cultures. The following day the plasmids are purified from *E. coli* 20 using the Qiagen Midi Prep kit (Venlo, Netherlands). The purified plasmid is then RNase A (ThermoFisher, Waltham, Mass) treated and re-purified using the DNA Clean and Concentrator kit (Zymo, Irvine, CA).

[00262] HEK293T cells are cultured in DMEM medium which is supplemented with 10% FBS and 1× Penicillin and Streptomycin. 100 ng of total DNA (50 ng of gRNA 25 plasmid and 50 ng of CFE plasmids) is mixed with 1 µL of PolyFect (Qiagen, Venlo, Netherlands) in 25 µL of OptiMEM in a 96 well plate. The complex is incubated for 10 minutes and then 20,000 HEK293T cells resuspended in 100 µL of DMEM are added to the mixture. The resulting mixture is then incubated for 80 hours at 37 °C and 5% CO₂.

[00263] The cells are harvested from flat bottom 96 well plates using TrypLE Express 30 reagent (ThermoFisher, Waltham, MA) and transferred to v-bottom 96 well plate. The plate is then spun down at 500 × g for 5 minutes. The TrypLE solution is then aspirated and the

cell pellet is resuspended in FACS buffer (1X PBS, 1 % FBS, 1 mM EDTA and 0.5% BSA). The GFP+, BFP+ and RFP+ cells are then analyzed on the Attune NxT flow cytometer and the data is analyzed on FlowJo software.

5 [00264] The RFP+BFP+ cells that are identified are indicative of the proportion of enriched cells that have undergone precise or imprecise editing process. BFP+ cells indicate cells that have undergone successful editing process and express BFP. The GFP-

10 [00265] In this experiment, the edit is immediately 3' of the gRNA, and 3' of the edit is a further region complementary to the nicked genome, although the intended edit could also be present further 5' within the region homologous to the nicked genome. A nickase RT fusion enzyme (Cas9 H840A nickase) creates a nick in the target site and the nicked DNA anneals to its complementary sequence on the 3' end of the gRNA. The RT then extends the DNA, thereby incorporating the intended edit directly in the genome.

15 [00266] The effectiveness of CREATE Fusion Editing in GFP+ HEK293T cells is tested. In the assay system devised, a successful precise edit results in a BFP+ cell whereas an imprecise edit turns the cell both BFP and GFP negative. CREATE Fusion gRNA in combination with CFE2.1 or CFE2.2 gives approximately 40-45% BFP+ cells indicating that almost half the cell population undergoes successful editing (data not shown). The

20 GFP- cells are ~10% of the population. The use of a second nicking gRNA, as described in Anzalone et al. (Nature, 576(7785):149-157 (2019)) does not increase the precision edit rate any further; in fact, it significantly increases the imprecisely edited, GFP-negative cell population and the editing rate is lower.

25 [00267] Previous literature has shown that double nicks on opposite strands (<90 bp away) do result in a double strand break which tend to be repaired via NHEJ resulting in imprecise insertions or deletions. Overall, the results indicate that CREATE Fusion Editing predominantly yields precisely edited cells and that the imprecisely edited cells proportion is much lower (data not shown).

30 [00268] An enrichment handle, specifically a fluorescent reporter (RFP) linked to nuclease expression is included in this experimentation as a proxy for cells receiving the editing machinery. When only the RFP-positive cells are analyzed (computational

enrichment) after 3 to 4 cell divisions, up to 75% of the cells are BFP+ when tested with gRNA (data not shown), indicating uptake or expression-linked reporters can be used to enrich for a population of cells with higher rates of CREATE Fusion-mediated gene editing. In fact, the combined use of CREATE Fusion Editing and the described enrichment methods result in a significantly improved rate of intended edits (data not shown).

Example III: CREATE Fusion Editing

[00269] CREATE Fusion Editing is carried out in mammalian cells using a single guide RNA covalently linked to a homology arm having an intended edit to the native sequence and an edit that disrupts nuclease cleavage at this site. Briefly, lentiviral vectors are produced using the following protocol: 1000 ng of lentiviral transfer plasmid containing the CREATE Fusion cassettes along with 1500 ng of lentiviral packaging plasmids (ViraSafe Lentivirus Packaging System Cell BioLabs) are transfected into HEK293T cells using Lipofectamine LTX in 6-well plates. Media containing the lentivirus is collected 72 hours post transfection. Two clones of a lentiviral CREATE Fusion gRNA-HA design are chosen, and an empty lentiviral backbone is included as negative control.

[00270] The day before the transduction, 200,000 HEK293T cells are seeded in six well plates. Different volumes of CREATE lentivirus (10 μ L to 1000 μ L) are added to HEK293T cells in six well plates along with 10 μ g/mL of Polybrene. 48 hours after transduction, media with 15 μ g/mL of Blasticidin is added to the wells. Cells are maintained in selection for one week. Following selection, the well with lowest number of surviving cells is selected for future experiments (<5 % cells).

[00271] The experimental constructs or wild-type SpCas9 are electroporated into HEK293T cells using the Neon Transfection System (Thermo Fisher Scientific). Briefly, 400 ng of total plasmid DNA is mixed with 100,000 cells in Buffer R in a total of 15 μ L volume. The 10 μ L Neon tip is used to electroporate cells using 2 pulses of 20 ms and 1150 v. Cells are analyzed on the flow cytometer 80 hours post electroporation. Unenriched editing rates of up to 15% are achieved from single copy delivery of gRNA (data not shown).

[00272] When the editing is combined with computational selection of RFP+ cells, however, enriched editing rates of up to 30% are achieved from a single copy delivery

gRNA. This enrichment via selection of cells receiving the editing machinery is shown to result in a 2-fold increase in precise, complete intended edits (data not shown). Two or more enrichment/delivery steps can also be used to achieve higher editing rates of CREATE Fusion Editing in an automated instrument, e.g., use of a module for cell handle
5 enrichment and identification of cells having BFP expression. When the method enriches for cells that have higher gRNA expression levels, the editing rate is even further increased, and thus a growth and/or enrichment module of the instrument may include gRNA enrichment.

10 *Example IV: Dual Plasmid CREATE Fusion Editing and Barcoding*

[00273] CREATE Fusion Barcoding (“CFB”) is simultaneously carried out with CREATE Fusion Editing in mammalian iPS-GFP cells using CF barcoding cassettes and CF editing cassettes having different CFgRNAs for targeting separate genomic loci. CF editing cassettes include a single CFgRNA covalently linked to a repair template for
15 targeting a model GFP locus and effecting a GFP-to-BFP swap mutation at this site. CF barcoding cassettes include a single CFgRNA covalently linked to a repair template for targeting either a model eukaryotic translation initiation factor 4E-binding protein 2 (“4EBP2”) locus or model DNA methyltransferase 3 beta (“DNMT3b”) locus and integrating a 9 bp barcode at these sites. The CF editing cassettes are assembled into editing
20 plasmids encoding the CF editing cassettes and a CREATE Fusion Enzyme (“CFE”), and the CF barcoding cassettes are assembled into separate barcoding plasmids. FIG. 8A schematically illustrates the assembled editing and barcoding plasmids.

[00274] The editing and barcoding plasmids are co-transfected into the iPS-GFP cells, and conditions for editing/barcoding are provided. FIG. 8B graphically illustrates the
25 editing rates that are observed in the iPS-GFP cells when the cells are transfected with a single editing (targeting GFP) plasmid, or dual editing (targeting GFP) and barcoding (either targeting 4EBP2 or DNMT3b) plasmids. The cells in FIG. 8B are then sorted on BFP +/- gates by fluorescence-activated cell sorting (“FACS”), and barcode insertion rates are measured by next-generation sequencing (“NGS”), with the results shown in FIG. 8C.
30 45% or 89% of the editing cells (BFP+) are successfully barcoded at either the 4EBP2 locus or DNMT3b locus, respectively.

Example V: Single Plasmid CREATE Fusion Editing and Barcoding

[00275] Similar to Example IV, CREATE Fusion Barcoding (“CFB”) is simultaneously carried out with CREATE Fusion Editing in mammalian iPS-GFP cells using CF barcoding cassettes and CF editing cassettes having different CFgRNAs for targeting separate loci. CF editing cassettes include a single CFgRNA covalently linked to a repair template for targeting a model GFP locus and effecting a GFP-to-BFP swap mutation at this site. CF barcoding cassettes include a single CFgRNA covalently linked to a repair template for targeting a model DNA methyltransferase 3 beta (“DNMT3b”) locus and integrating either a 9 bp barcode (MAM10 or PacI insertion) or an 18 bp barcode (ISceI insertion) at these sites. The CF editing cassettes and CF barcoding cassettes are either assembled into a single plasmid (e.g., in “tandem”) further encoding a CREATE fusion enzyme (“CFE”), or the CF editing cassettes and CF barcoding cassettes are assembled into separate editing and barcoding plasmids, respectively. FIG. 9A schematically illustrates the assembled plasmids.

[00276] The single or dual plasmids are transfected into the iPS-GFP cells, and conditions for editing/barcoding are provided. FIG. 9B graphically illustrates the editing rates that are observed in the iPS-GFP cells when the cells are transfected with a single tandem (targeting GFP and DNMT3b) plasmid, or dual editing (targeting GFP) and barcoding (targeting DNMT3b) plasmids. Similarly, FIG. 9C graphically illustrates the barcoding rates that are observed in the iPS-GFP cells when the cells are transfected with a single tandem (targeting GFP and DNMT3b) plasmid, or dual editing (targeting GFP) and barcoding (targeting DNMT3b) plasmids. As shown, editing and barcoding rates are both improved utilizing a single tandem plasmid as compared to dual plasmid transfection. Barcoding rates are sensitive to barcode insertion sequence and length, with no detectable ISceI (18bp) insertions detected by qPCR.

Example VI: Detection of CREATE Fusion Barcodes by RNAseq

[00277] CREATE Fusion Barcoding (“CFB”) is simultaneously carried out with CREATE Fusion Editing in mammalian iPS-GFP cells using CF barcoding cassettes and CF editing cassettes having different CFgRNAs for targeting separate loci. CF editing

cassettes include a single CFgRNA covalently linked to a repair template for targeting a model GFP locus and effecting a GFP-to-BFP swap mutation at this site. CF barcoding cassettes include a single CFgRNA covalently linked to a repair template for targeting a model ornithine decarboxylase antizyme 1 (“OAZ1”) locus and effecting a 2 bp swap at this site. The CF editing cassettes and CF barcoding cassettes are assembled into separate editing and barcoding plasmids, and are co-transfected into the iPS-GFP cells. FIG. 10A graphically illustrates the editing rates observed in the iPS-GFP cells when the cells are transfected with dual editing (targeting GFP) and barcoding (targeting OAZ1) plasmids versus a negative barcoding control. The cells in FIG. 10A are sorted on BFP +/- gates by fluorescence-activated cell sorting (“FACS”), and barcode insertion rates at OAZ1 transcripts are measured by RNAseq, with the results shown in FIG. 10B.

Example VII: CREATE Fusion Barcoding in 3' UTRs of Transcribed Genes

[00278] As previously discussed, barcoding may be carried out in genomic safe harbor loci to reduce the potential of barcode sequence integration adversely affecting genes neighboring the integrated barcode(s). In other words, a target genomic locus for barcoding may include a safe harbor region of the genome that minimally affects the biology of the cell. In this experiment, genomic loci corresponding to 3' untranslated regions (UTRs) regions of transcribed genes are considered and tested for barcoding due to the 3' UTRs being benign to cell function while also being readily detectable by RNAseq. In this experiment, ninety-six (96) different 3'-UTR loci are tested and screened for barcoding efficiency utilizing a 9 bp barcode encoded by barcoding cassettes.

[00279] Briefly, CREATE Fusion Barcoding (“CFB”) is carried out in mammalian iPS-GFP cells using CF barcoding cassettes having CFgRNAs designed for targeting one of a plurality of genomic loci corresponding to 3' UTRs of transcribed genes. The CF barcoding cassettes include a single CFgRNA covalently linked to a repair template for targeting and integrating the 9 bp barcode at a respective 3'-UTR locus.

[00280] With the aforementioned cassettes, each of the ninety-six (96) different 3' UTR loci are individually tested for barcoding efficiency using the following protocol: PGP168-GFP cells are cultured in mTeSR Plus medium (STEMCELL Technologies, Vancouver, Canada) at 37°C and 5% CO₂. 24 hours before transfection, the cells are seeded at 15k cells

per well in Matrigel-coated (Corning Life Sciences, Corning, NY) flat bottom, 96-well culture plates and supplemented with 10 μ M Y-27632 ROCK inhibitor (STEMCELL Technologies). 100 μ L of the medium is replaced (without Y-27632) immediately before transfection. 100 ng of total DNA (50 ng of barcoding gRNAs plus 50 ng of CREATE Fusion Enzyme (CFE) and plasmids) is mixed with 1 μ L Lipofectamine Stem Transfection Reagent (Thermo Fisher Scientific) in 10 μ L of OptiMEM (Thermo Fisher Scientific). The resulting mixture is incubated for 10-30 minutes at room temperature, then added to a single well of PGP168-GFP cells in 96-well culture plates and transferred to a 37°C incubator with 5% CO₂. After 24 hours of incubation, the transfection medium is removed, and replaced with mTeSR Plus Medium and 2 μ g/mL puromycin (InvivoGen, San Diego, CA). 48 hours after transfection, the medium is replaced with mTeSR Plus and 0.25 μ g/mL puromycin.

[00281] Ninety-six hours after transfection, genomic DNA is purified from the cells using a DNAdvanced Kit (Beckman Coulter Life Sciences, Brea, CA) for genomic DNA extraction. Regions of genes containing the target loci of the CFE-gRNA complexes are then amplified via PCR. These PCR amplicons are prepared for next generation sequencing (NGS) using a TruSeq DNA Sample Prep Kit (Illumina, Inc., San Diego, CA) according to the manufacturer's instructions. The samples are then sequenced using an Illumina MiSeq benchtop sequencer and 2x150 reagent kit (Illumina). NGS analysis is performed using a custom NGS analysis and sequencing read alignment pipeline to bin read counts according to sequence identity to target genomic loci with a complete targeted 9 base insertion or wild-type sequence.

[00282] FIG. 11 graphically illustrates the barcoding rates observed in the cells for each of the 96 different 3'-UTR-corresponding loci ("guide1"- "guide96") targeted. The barcoding rates depicted in FIG. 11 are calculated as the fraction of total reads aligned to the respective target genomic locus that contained the 9 bp insertion. As shown, the results indicate that barcoding of 3' UTR loci according to the methods described herein is generally effective with relatively high barcoding rates at some 3' UTR loci (right side of FIG. 11). The varying barcoding efficiency rates also indicate that while 3' UTR loci appear to be effective for barcoding, barcoding efficiency appears to be sensitive to and dependent on the specific 3'-UTR locus targeted. In the example of FIG. 11, the top 28 performing

loci (in terms of barcoding efficiency; “guide96” - “guide69”) include 3’ UTR of the following genes, respectively: CARHSP1, MAP4, SLC38A1, WNK1, DIAPH1, LRRC8A, FAF2, NKTR, TBC1D16, GJC1, NUCKS1, CAPZB, TBC1D16, MPP6, WDR83OS, PMEPA1, SERINC5, HTT, SLC29A1, PPP3CA, EZR, HEBP2, HTT, SLC7A1, LSM14A, ERBB2, CYP51A1, and GPATCH8.

Example VIII: CREATE Fusion Editing and Barcoding in 3’ UTRs of Transcribed Genes

[00283] CREATE Fusion Barcoding (“CFB”) is simultaneously carried out with CREATE Fusion Editing in mammalian iPS-GFP cells using CF barcoding cassettes and CF editing cassettes having different CFgRNAs for targeting separate genomic loci. CF editing cassettes include a single CFgRNA covalently linked to a repair template for targeting a model GFP locus and effecting a GFP-to-BFP swap mutation at this site. CF barcoding cassettes include a single CFgRNA covalently linked to a repair template for targeting one of four different genomic loci corresponding to 3’ UTRs of transcribed genes (“guide1”, “guide2”, “guide3”, and “guide4” – shown in FIG. 12), including a 3’ UTR of CD63, and integrating a 9 bp barcode at each respective site.

[00284] The CF editing cassettes and CF barcoding cassettes are either assembled into a single plasmid (*e.g.*, in “tandem”) further encoding a CREATE fusion enzyme (“CFE”), or the CF editing cassettes and CF barcoding cassettes are assembled into separate editing and barcoding plasmids, respectively.

[00285] Briefly, CF barcoding and editing with the single plasmid system is carried out with the following protocol: PGP168-GFP cells are transfected in 96-well culture plates at 15,000 cells per well with 100 ng total plasmid DNA, as described in Example VII. Here, a CFE, barcoding cassette (*e.g.*, comprising a barcoding gRNA), library editing cassette (*e.g.*, comprising an editing gRNA targeting the GFP-to-BFP conversion target described in Example I), and puromycin deacetylase gene are expressed from the single plasmid. 24 hours after transfection, the transfection medium is replaced with mTeSR Plus and 4, 8, or 10 µg/mL puromycin, and is then reduced to 0.25 µg/mL puromycin at 48 hours after transfection.

[00286] Meanwhile, CF barcoding and editing with the dual-plasmid system is carried out with the following protocol: PGP168-GFP cells are transfected in 96-well culture plates at 15,000 cells per well with 100 ng total plasmid DNA, as described in Example VII. The cells are transfected with 25, 50, or 75 ng of a first plasmid containing a CFE, editing cassette, and puromycin deacetylase gene, as well as 75, 50, or 25 ng of a second plasmid containing a barcoding cassette, respectively. Twenty-four hours after transfection, the transfection medium is replaced with mTeSR Plus and 10 µg/mL puromycin, and is then reduced to 0.25 µg/mL puromycin at 48 hours after transfection.

[00287] Ninety-six hours after transfection for each of the two experiments, genomic DNA are extracted from a first aliquot of each sample and barcoding rates are determined via genomic amplicon sequencing, as described in Example VII. Another aliquot of each sample is then collected and analyzed for GFP-to-BFP conversion by flow cytometry, as described in Example II.

[00288] FIG. 12 graphically illustrates both editing and barcoding rates observed in the cells when the cells are transfected with a single plasmid system (“1-Plasmid,” shown at left) or dual plasmid system (“2-Plasmid,” shown at right) for four different 3’ UTR barcoding loci (“guide1” to “guide4”). The 1-Plasmid graphs at left further depict editing and barcoding rates in samples prepared with different concentrations of puromycin in the selection and maintenance media (e.g., either 4, 8, or 10 µg/mL, as described above). Similarly, the 2-Plasmid graphs at right depict editing and barcoding rates for samples prepared with different ratios of the two plasmids (in FIG. 12, the mass of the barcoding plasmid, “pUCIDT,” is listed, as well as the corresponding barcode (“BC”) gRNA to edit gRNA stoichiometry). As shown, the results generally indicate that simultaneous, high efficiency editing is possible with barcoding of 3’ UTR loci for both single plasmid and dual plasmid systems according to the methods described herein.

Example IX: Detection of CREATE Fusion Barcodes in 3’ UTRs of Transcribed Genes by Single-cell RNAseq

[00289] CREATE fusion barcode insertion in genomic loci corresponding with the 3’UTRs of transcribed genes enables efficient barcode detection and sequencing using common poly-A mRNA sequencing techniques. Here, CREATE Fusion barcodes inserted

into such 3' UTR loci are detected by single-cell RNAseq using the following protocol: PGP168-GFP cells are transfected in triplicate with plasmids comprising a CFE and barcode editing cassette, and are thereafter enriched by puromycin selection as described in Example VII. Ninety-six hours after transfection, samples having about 500 cells per sample are collected and processed with a 10X Genomics Chromium Next GEM Single Cell 3' Reagent Kit v3.1 with feature barcoding for cell multiplexing, per the manufacturer's instructions (10X Genomics, Pleasanton, CA). Sequencing libraries are then prepared using a NextSeq 550 v2.5 150 cycle High Output Kit (Illumina) and sequenced with paired end, dual indexing on an Illumina NextSeq System.

5
10 [00290] NGS analysis is performed using 10X Genomics Cell Ranger software and a custom NGS analysis and sequencing read alignment pipeline to de-multiplex single cell transcriptomes and bin read counts according to sequence identity (e.g., similarity) to barcoded target genomic loci with a complete 9 base insertion or wild-type sequence. Raw counts of unique molecular identifiers associated with barcode or wild-type reads are then used to determine the relative expression level of barcoded and non-barcoded transcripts, as shown in FIG. 13.

[00291] In FIG. 13, complete targeted insertions are labeled as "alt," while a wild-type sequences are labeled as "ref." Barcoded cells are shown and classified as cells with a 9 base barcode insertion in all target genomic loci reads ("alt/alt"), or a mixture of 9 base barcode insertions and wild-type sequences ("alt/ref"), corresponding with cells that are barcoded on both or one allele, respectively. Non-barcoded cells, where no barcodes were detected, are classified and shown as "ref/ref." Lastly, instances where neither the reference sequence nor the barcode were detected are classified are shown in FIG. 13 as "No Call." FIG. 13 thus indicates that barcodes integrated into genomic loci corresponding with the 3' UTR of transcribed genes according to the methods described herein are relatively easy to detect via mRNA sequencing techniques.

Example X: CREATE Fusion Editing and Barcoding - Integration of HA Epitope Tag Knock-in on Cell Surface Receptors

[00292] CREATE fusion barcoding and editing is carried out in mammalian cells (iPSCs), where the CFgRNA design allows for barcoded HA epitope tag knock-in edits at various genomic loci that encodes for endogenous surface receptors (BST2, CD151, CD63,

CD81, and CD9) (FIG. 14A). The HA epitope tag has a barcoded linker using synonymous codons, where there are 6.7×10^7 unique barcode sequences with a synonymous amino acid translation (FIG. 14B). Various CFgRNAs (x-axis) used in this experiment are screened in iPSCs for successful knock-in, expression, and surface display of the HA tags, where expression (y-axis) is measured by flow cytometry with PE labeled anti-HA antibodies (FIG. 14C).

Example XI: CREATE Fusion Editing and Barcoding - Integration of Various Edits at Endogenous Loci and CD81-HA Tag Knock-in Barcoding Edit

[00293] CREATE fusion barcoding and editing is carried out in mammalian cells (iPSCs), where the cells are co-transfected with CFgRNAs targeting a GFP-to-BFP edit and a CD81-HA tag knock-in barcoding edit. Physical magnetic sorting (MACS) is performed with anti-HA antibody functionalized beads, and enriches the population of cells with a successful CD81-HA tag knock-in edit from 2.2% (left) to 84.1% (right) (FIG. 15A). GFP-to-BFP edit rates (y-axis) is improved from 17% BFP+ in unsorted cells to 50% BFP+ in MACS sorted/enriched cells (FIG. 15B). Cells are co-transfected with CFgRNAs targeting various edits at endogenous loci and a CD81-HA tag knock-in barcoding edit. Sorting cells for HA knock-in via MACS or FACS improves edit rates at most endogenous targets relative to the edit rates for the unsorted samples (FIG. 15C).

Example XII: CREATE Fusion Editing and Barcoding - Integration of HPRT Loss-of-function Edit Confers Negative Selection by Resistance to 6-TG

[00294] CREATE fusion barcoding and editing is carried out in mammalian cells using CFgRNA designs for HPRT loss-of-function edit, for example wherein the HPRT loss of function is introduced by a frame shift barcode insertion (FIG. 16A), or wherein the HPRT loss of function is from a HPRT knockout (FIG. 16B). Cells are co-transfected with CFgRNAs targeting a GFP-to-BFP edit and an HRPT knockout edit ("HPRT DF"), and further treated with 6-TG, wherein the 6-TG treatment selects for cells with the HPRT DF edit (FIG 16C). Negative selection for HPRT DF with 6-TG supplemented media improves the GFP-to-BFP edit rate (y-axis) from approximately 60% BFP+ to approximately 80% BFP+ (FIG. 16D).

Example XIII: CREATE Fusion Editing and Barcoding - Integration of Genomic Edits and a CD81-HA Tag Knock-in Barcoding Edit

[00295] CREATE fusion barcoding and editing is carried out in mammalian cells, where the cells are co-transfected with CFgRNAs targeting various genomic edits and a CD81-HA knock-in barcoding edit, wherein each genomic edit CFgRNA design is paired with a unique barcode CFgRNA during co-transfection. HA-tag knock-in provides a phenotypic handle for live-cell selection which enriches the population of cells with the desired genomic edits. NGS data shows genomic DNA-level insertion rates across diverse barcode sequences. NGS data also shows correlation of the barcode - genomic edit pairs, and shows a proof-of-concept for using barcodes to track intended genomic edits.

[00296] Further to the aforementioned examples, in some aspects, the compositions, methods, and modified cells of the current disclosure applies to the use of gRNA. In some aspects, the compositions, methods, and modified cells of the current disclosure applies to the use of any type of gRNA. In some aspects, the compositions, methods, and modified cells of the current disclosure applies to the use of one or more types of gRNAs.

[00297] In some aspects, the compositions, methods, and modified cells of the current disclosure applies to gene editing via endogenous repair mechanisms, *e.g.*, Homology-Directed Repair (HDR), recombination pathways, or other DNA repair pathways. In some aspects, the compositions, methods, and modified cells of the current disclosure applies to HDR-based gene editing. In some aspects, the compositions, methods, and modified cells of the current disclosure applies to any method to introduce a genetic mutation into a genome (*e.g.*, knock-in). In some aspects, the compositions, methods, and modified cells of the current disclosure applies to the use of gRNAs and HDR-based gene editing.

[00298] While this invention is satisfied by aspects in many different forms, as described in detail in connection with preferred aspects of the invention, it is understood that the present disclosure is to be considered as an example of the principles of the invention and is not intended to limit the invention to the specific aspects illustrated and described herein. Numerous variations may be made by persons skilled in the art without departure from the spirit of the invention. The scope of the invention will be measured by the appended claims and their equivalents. The abstract and the title are not to be construed as limiting the scope of the present invention, as their purpose is to enable the appropriate authorities, as well as the general public, to quickly determine the general nature of the invention. In the claims that follow, unless the term “means” is used, none of the features or elements recited

therein should be construed as means-plus-function limitations pursuant to 35 U.S.C. §112,
¶6.

CLAIMS

1. A method for performing nucleic acid-guided nuclease/reverse transcriptase fusion editing in a genome of a live cell, comprising:
- 5 (a) providing the live cell, wherein the live cell comprises a first target locus and a second target locus;
- (b) providing a nucleic acid-guided nuclease/reverse transcriptase fusion enzyme;
- (c) providing a CF editing cassette, the CF editing cassette comprising:
- 10 (i) a nucleic acid sequence encoding a first CFgRNA having a region of complementarity to a sequence of the first target locus; and
- (ii) a nucleic acid sequence encoding a first repair template;
- (d) providing a CF barcoding cassette, the CF barcoding cassette comprising:
- (i) a nucleic acid sequence encoding a second CFgRNA having a region of complementarity to a sequence of the second target locus; and
- 15 (ii) a nucleic acid sequence encoding a second repair template;
- (e) providing conditions to allow the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme, the first CFgRNA, and the first repair template to bind to the first target locus;
- (f) allowing the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme, the first CFgRNA, and the first repair template to edit the first target locus;
- 20 (g) providing conditions to allow the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme, the second CFgRNA, and the second repair template to bind to the second target locus; and
- (h) allowing the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme, the second CFgRNA, and the second repair template to integrate the barcode into the
- 25 second target locus.
2. The method of claim 1, further comprising:
- sequencing the genome or a transcriptome of the cell to track for integration of the barcode, the integration of the barcode representing a nucleic acid-guided nuclease/reverse transcriptase fusion editing event effected by the CF editing cassette.
- 30 3. The method of claim 2, wherein the genome is sequenced by single-cell amplicon-based next-generation sequencing.

4. The method of claim 2, wherein the transcriptome is sequenced by single-cell RNA sequencing.
5. The method of any one of claims 1 to 4, wherein a nucleic acid sequence encoding the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme and the CF editing cassette are assembled into a vector for introduction of the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme and CF editing cassette into the cell.
6. The method of claim 5, wherein the CF barcoding cassette is further assembled into the vector for introduction of the CF barcoding cassette into the cell.
7. The method of any one of claims 1 to 4, wherein the CF barcoding cassette is assembled into a barcoding vector for introduction of the CF barcoding cassette into the cell, and wherein the CF editing cassette is assembled into an editing vector for introduction of the CF editing cassette into the cell, wherein the editing vector is different from the barcoding vector.
8. The method of any one of claims 1 to 7, further comprising:
15 providing a selectable marker.
9. The method of claim 8, wherein the selectable marker is for selection and enrichment of cells having an integrated barcode or an effected edit.
10. The method of claim 8 or 9, further comprising selecting and enriching for cells having an integrated barcode or an effected edit.
- 20 11. An editing system comprising one or more vectors comprising:
a nucleic acid sequence encoding a nucleic acid-guided nuclease/reverse transcriptase fusion enzyme;
a CF editing cassette comprising:
a nucleic acid sequence encoding a first CFgRNA having a region of
25 complementarity to a sequence of a first target locus in a cell; and
a nucleic acid sequence encoding a first repair template;
a CF barcoding cassette comprising:
a nucleic acid sequence encoding a second CFgRNA having a region of
complementarity to a sequence of a second target locus in the cell; and
30 a nucleic acid sequence encoding a second repair template.

12. The editing system of claim 11, wherein the one or more vectors comprises an editing vector, and wherein the editing vector comprises the CF editing cassette.
13. The editing system of claim 12, wherein the editing vector further comprises a nucleic acid sequence encoding the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme.
14. The editing system of claim 12 or 13, wherein the editing vector further comprises the CF barcoding cassette.
15. The editing system of claim 11, wherein an editing vector comprises the CF editing cassette, and a barcoding vector comprises the CF barcoding cassette, and wherein the editing vector is different than the barcoding vector.
16. The editing system of claim 11, wherein one or more of the one or more vectors comprises a selectable marker.
17. A vector comprising:
a nucleic acid sequence encoding a nucleic acid-guided nuclease/reverse transcriptase fusion enzyme;
a CF editing cassette comprising:
a nucleic acid sequence encoding a first CFgRNA having a region of complementarity to a sequence of a first target locus in a cell; and
a nucleic acid sequence encoding a first repair template;
a CF barcoding cassette comprising:
a nucleic acid sequence encoding a second CFgRNA having a region of complementarity to a sequence of a second target locus in the cell; and
a nucleic acid sequence encoding a second repair template.
18. The vector of claim 17, wherein the CF editing cassette further comprises an edit to immunize the first target locus and prevent re-nicking.
19. The vector of claim 17 or 18, wherein the CF barcoding cassette further comprises an edit to immunize the second target locus and prevent re-nicking.
20. A method for performing nucleic acid-guided nuclease/reverse transcriptase fusion editing in a genome of a live cell, comprising:
(a) providing a live cell suitable for the editing;

(b) introducing a nucleic acid-guided nuclease/reverse transcriptase fusion enzyme;

(c) providing conditions to allow the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme to bind to a first target locus;

5 (d) allowing the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme to edit the first target locus;

(e) providing conditions to allow the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme to bind to a second target locus;

10 (f) allowing the nucleic acid-guided nuclease/reverse transcriptase fusion enzyme to integrate a barcode into the second target locus.

1/31

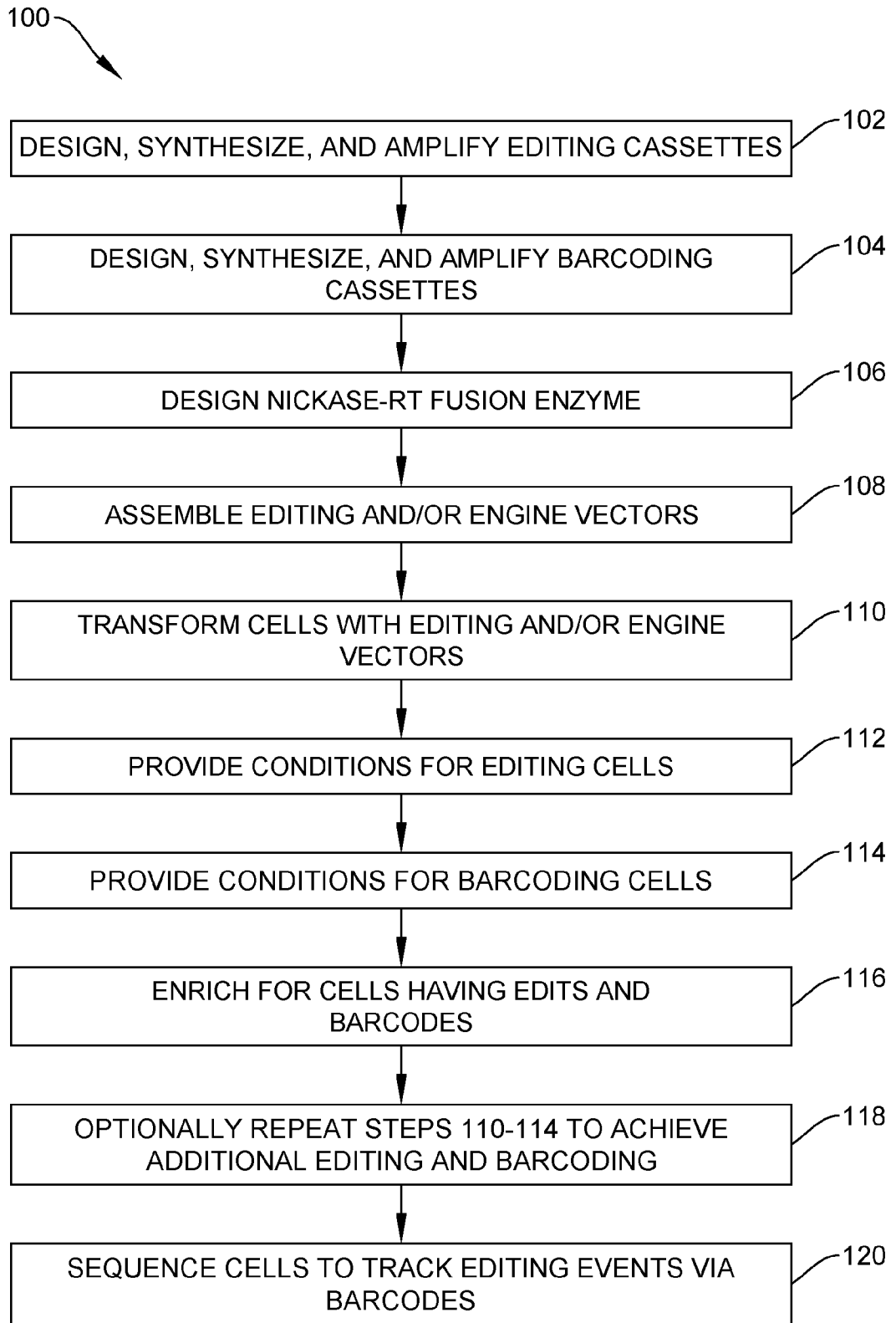


FIG. 1A

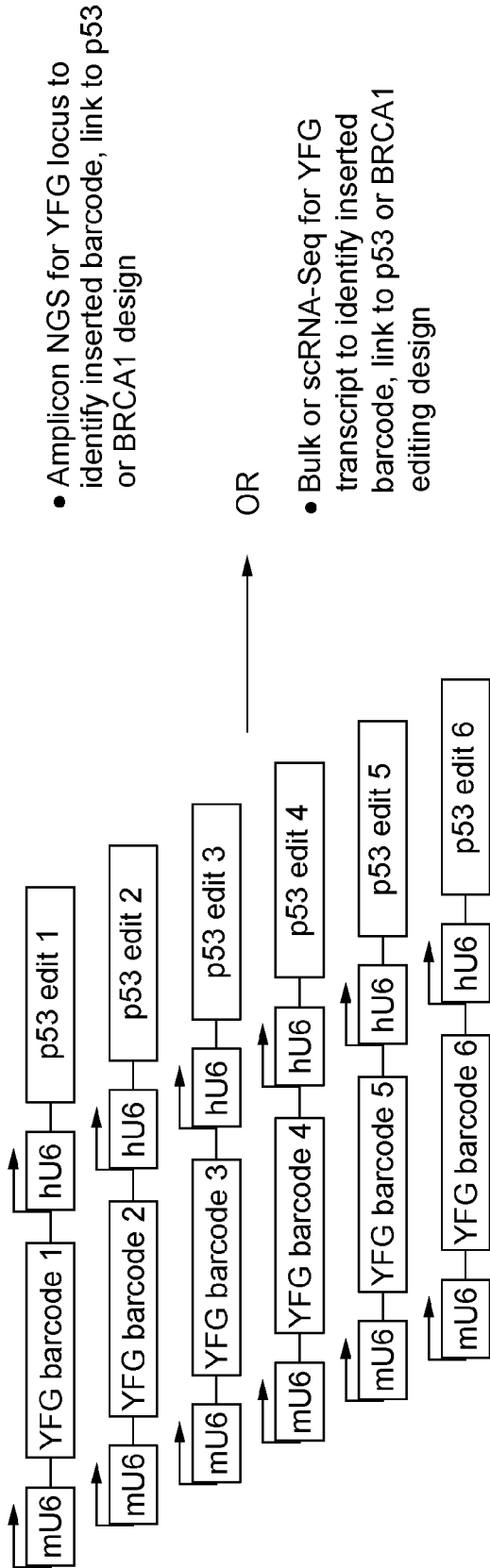


FIG. 1B

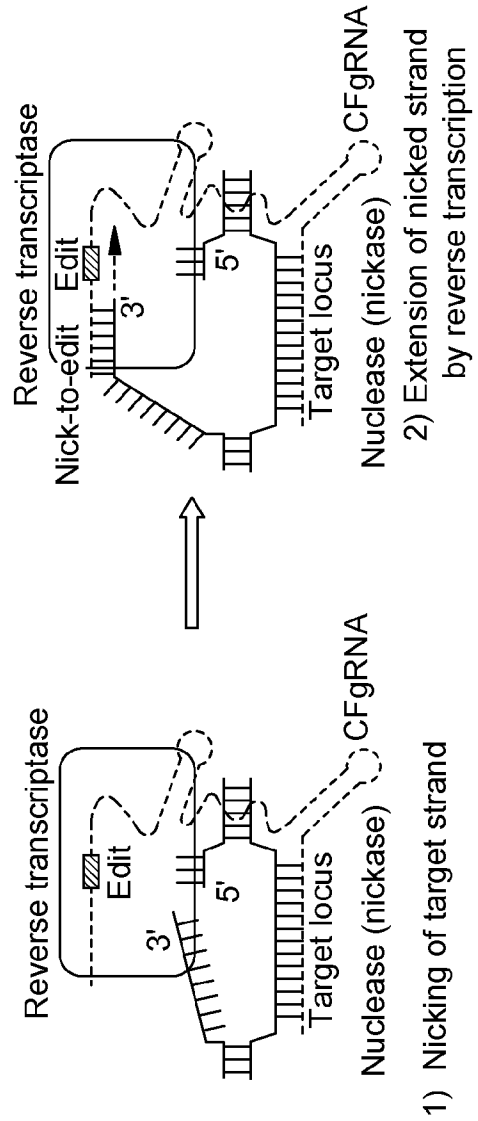
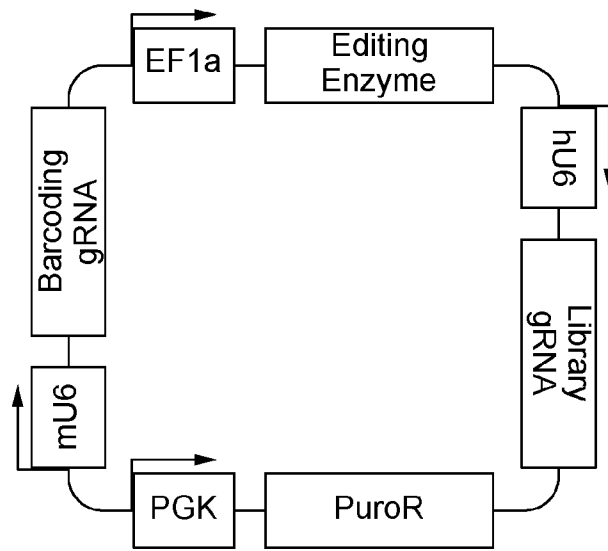


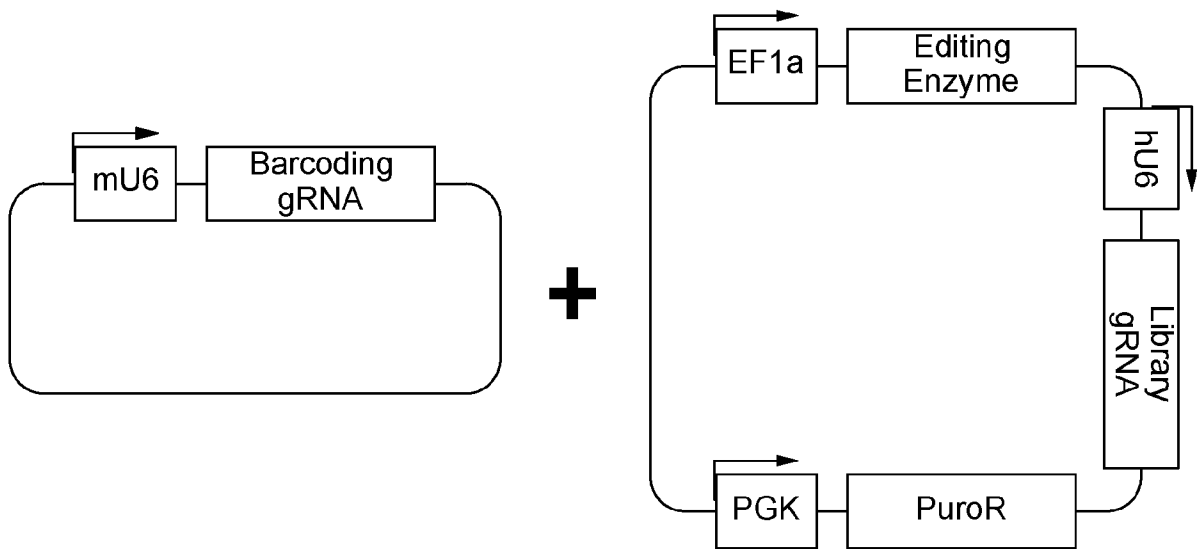
FIG. 1C

3/31



1-vector system

FIG. 1D



2-vector system

FIG. 1E

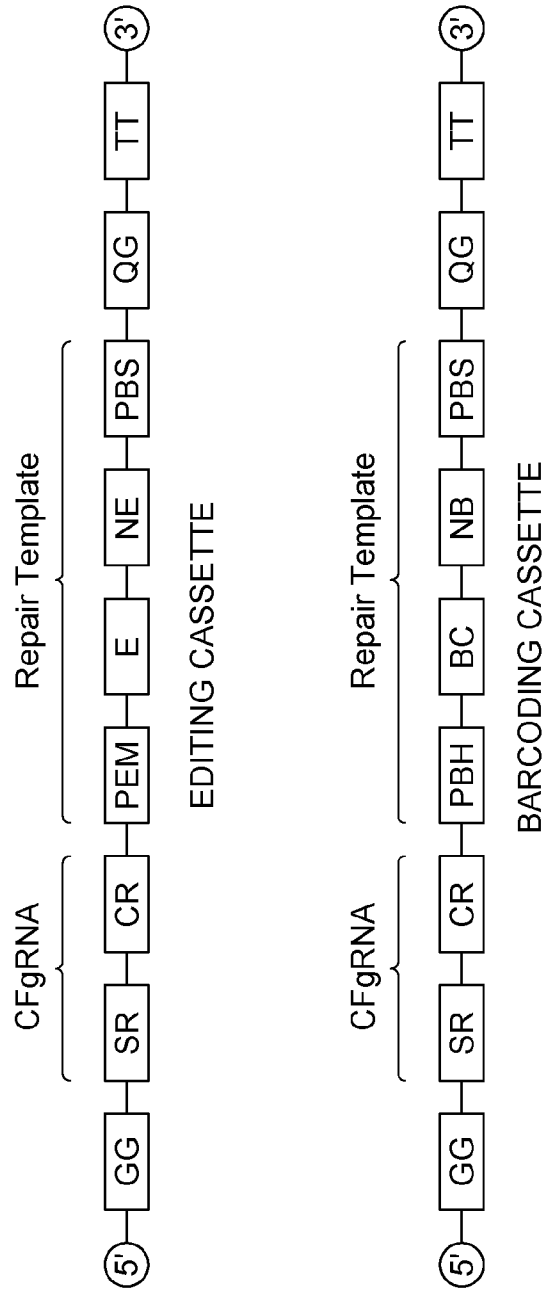


FIG. 1F

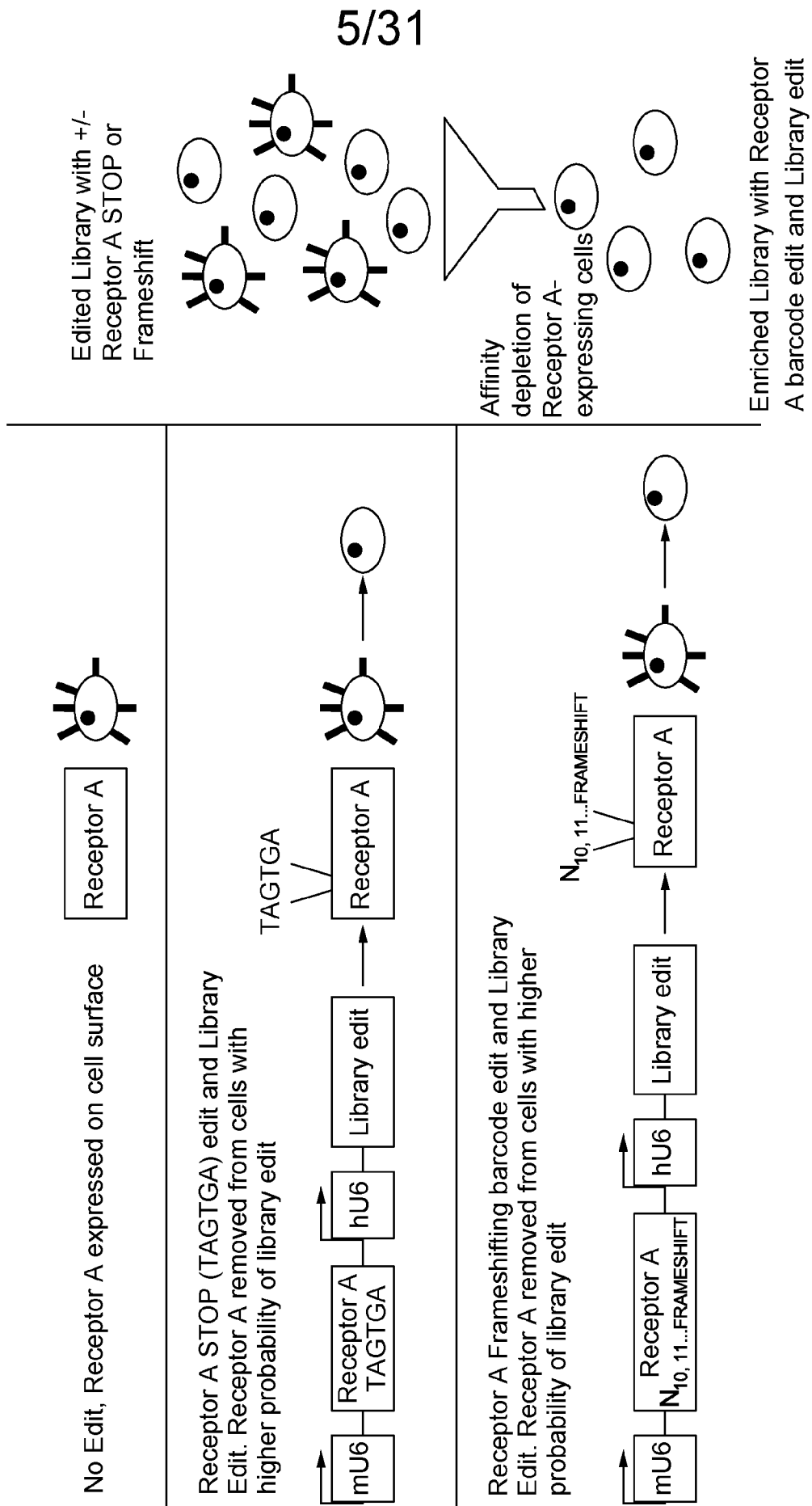


FIG. 1G

8/31

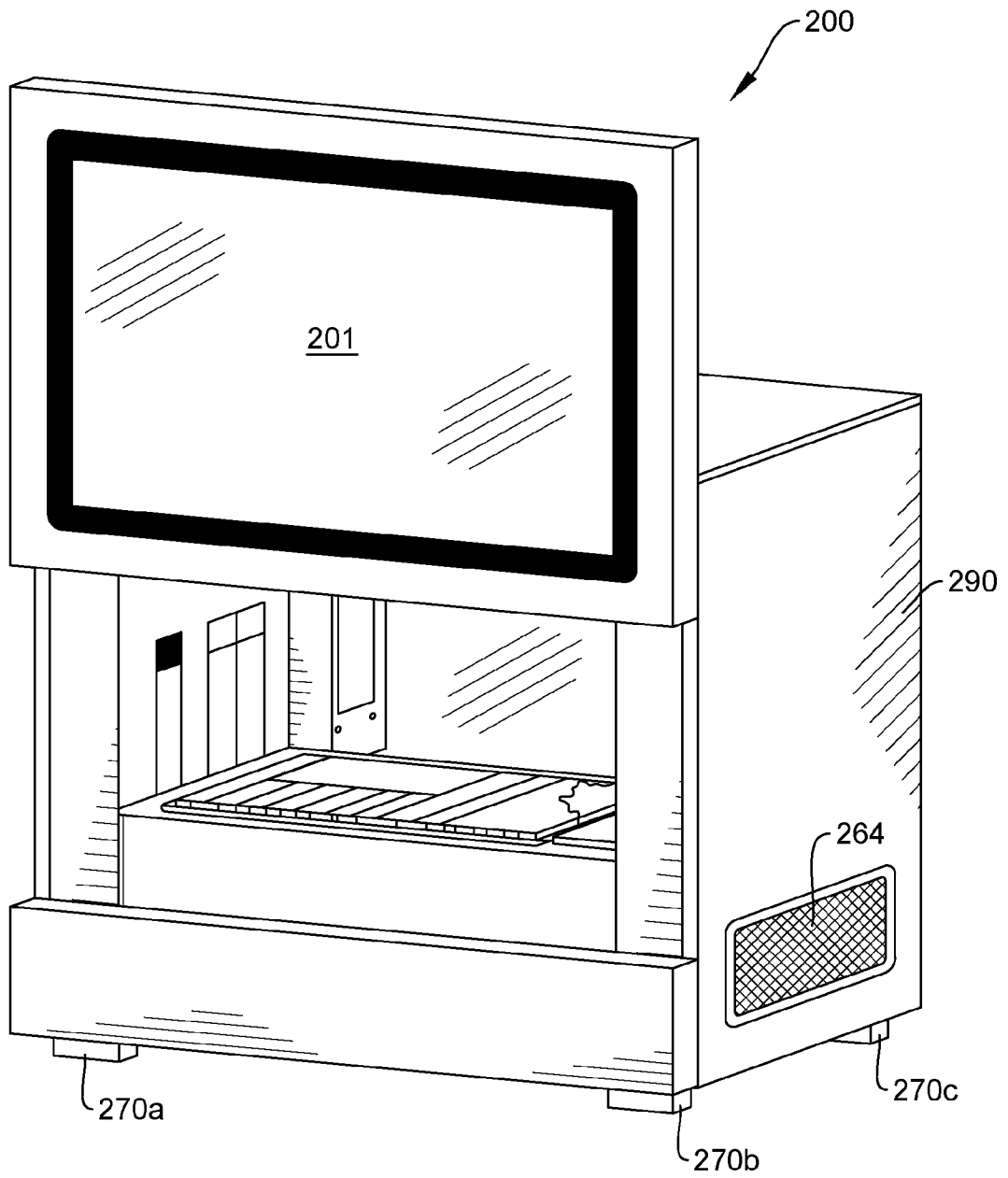


FIG. 2C

9/31

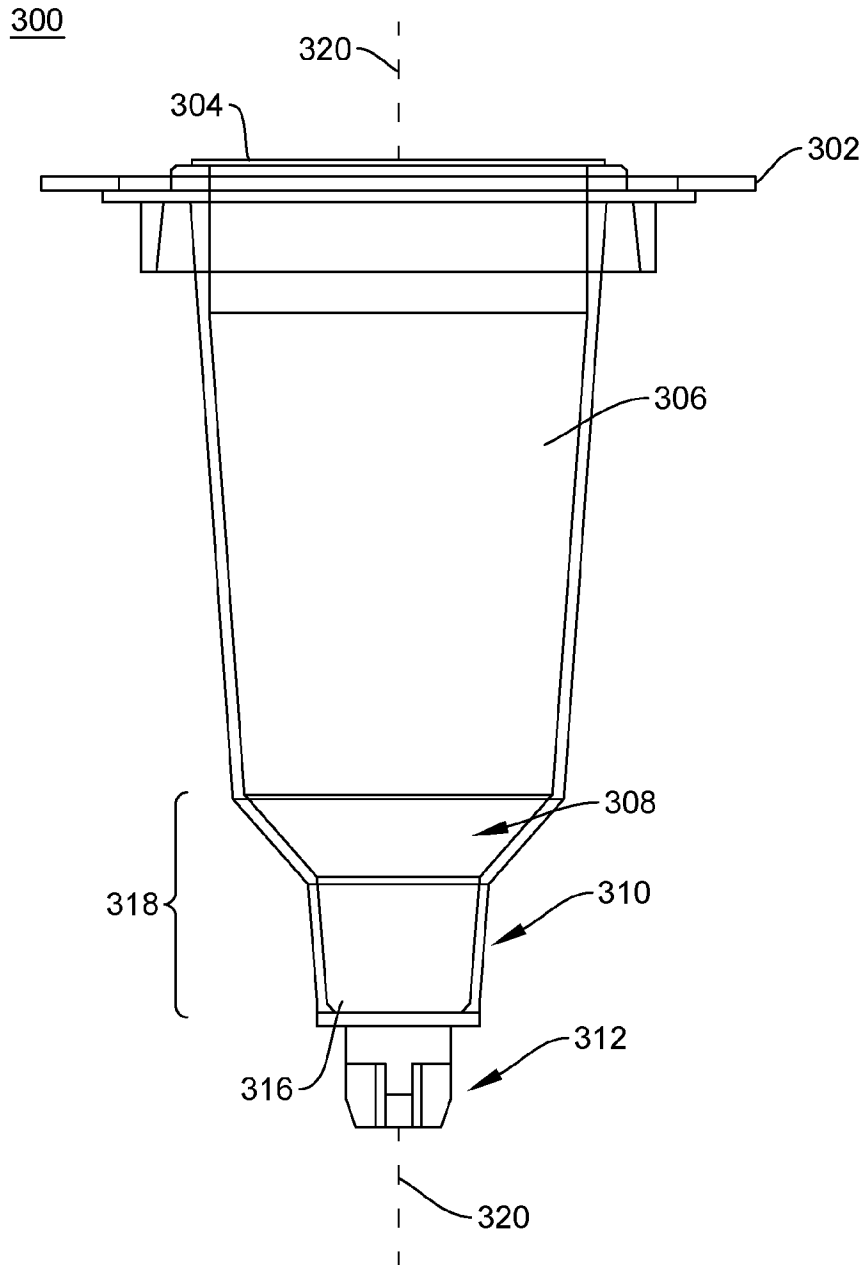


FIG. 3A

10/31

FIG. 3B

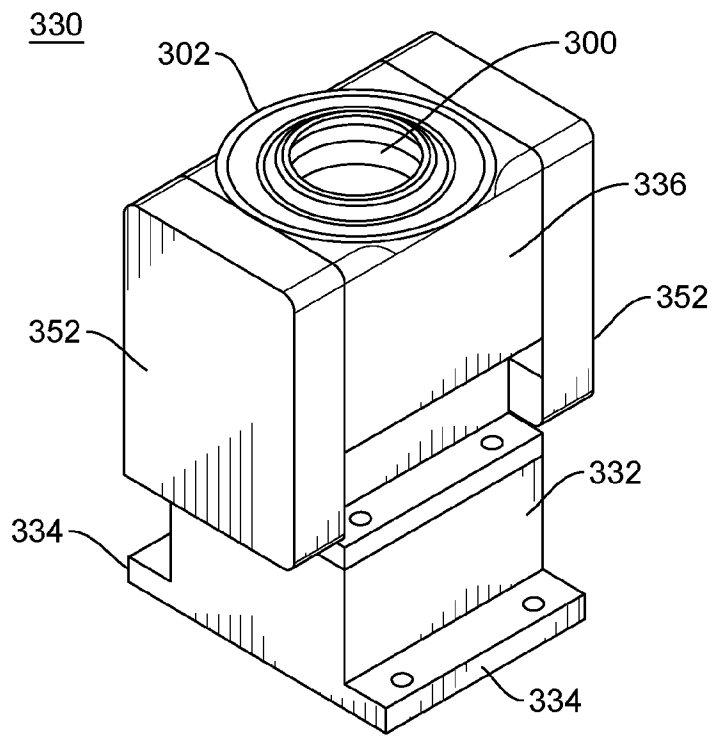
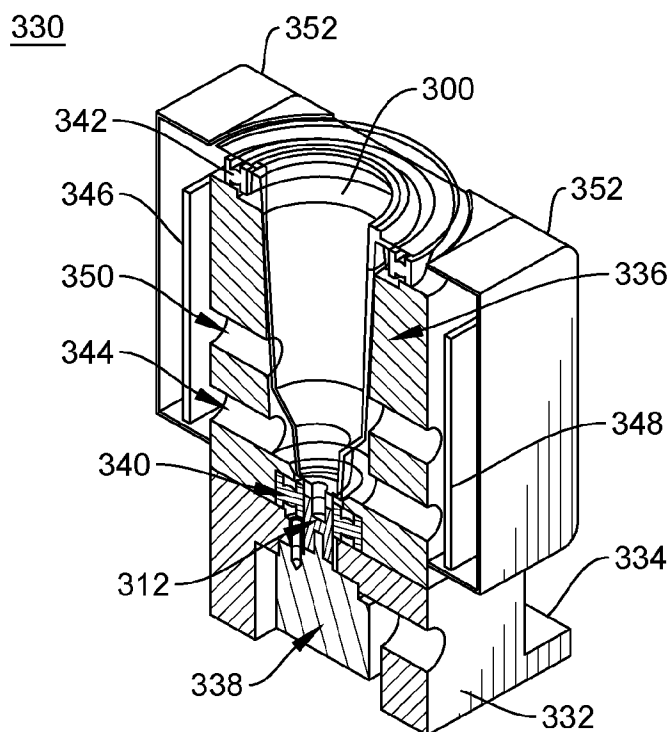


FIG. 3C



11/31

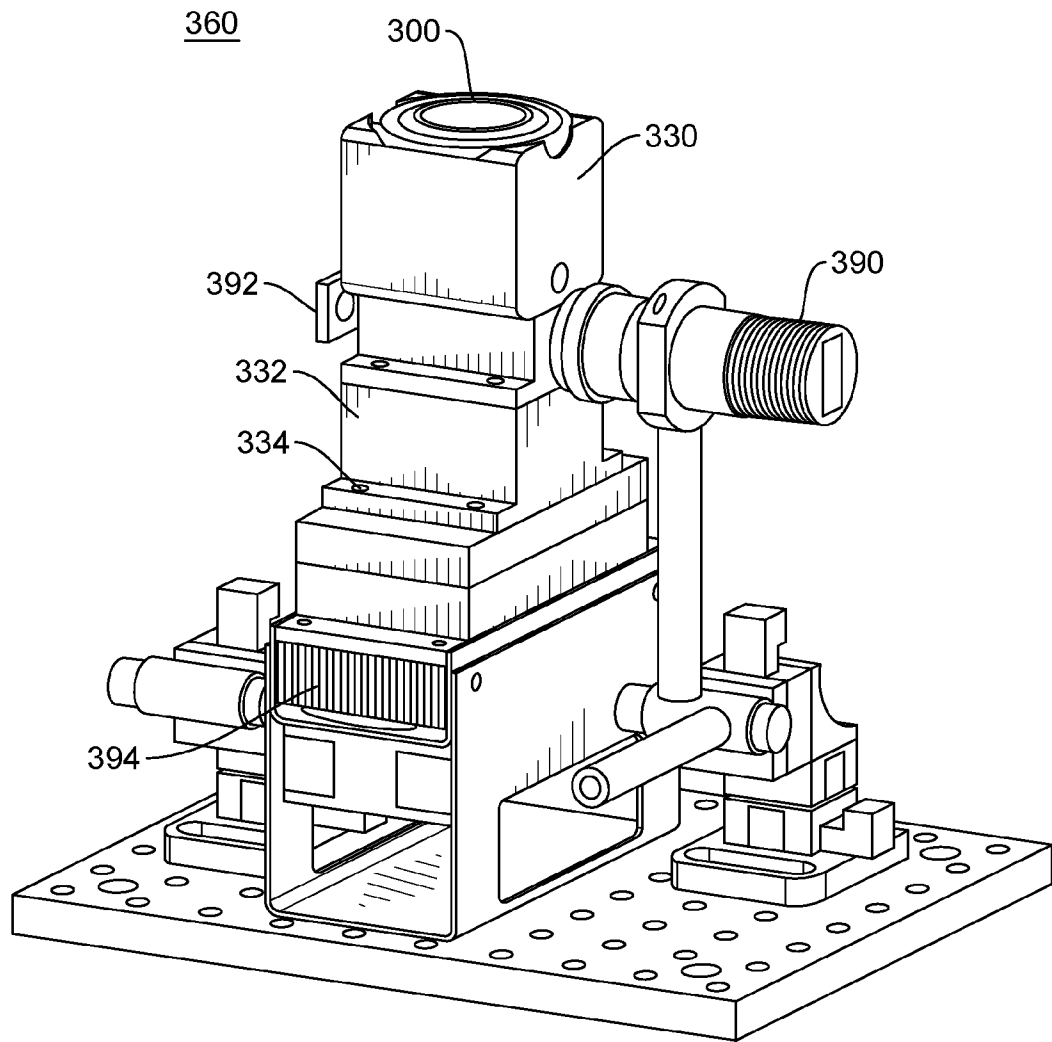
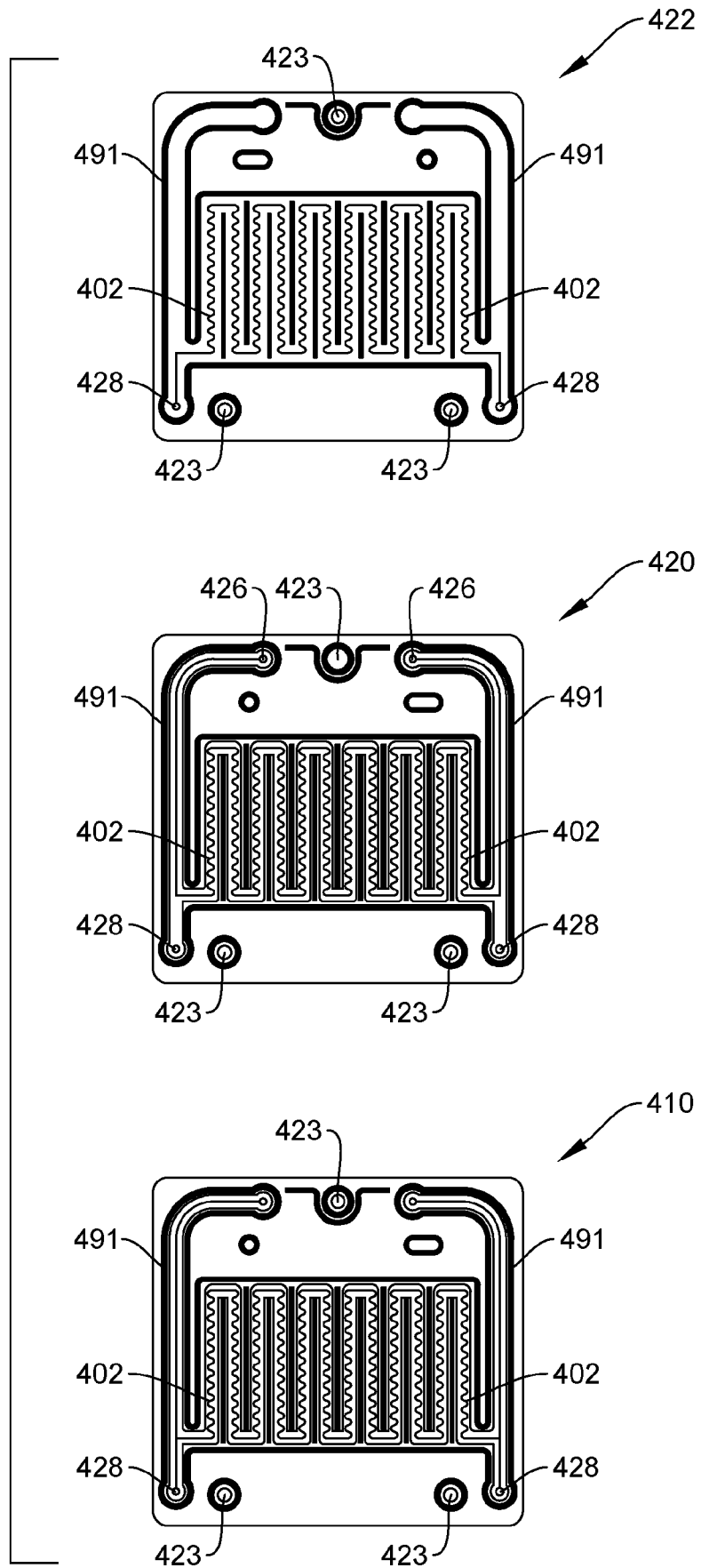


FIG. 3D

12/31

FIG. 4A



13/31

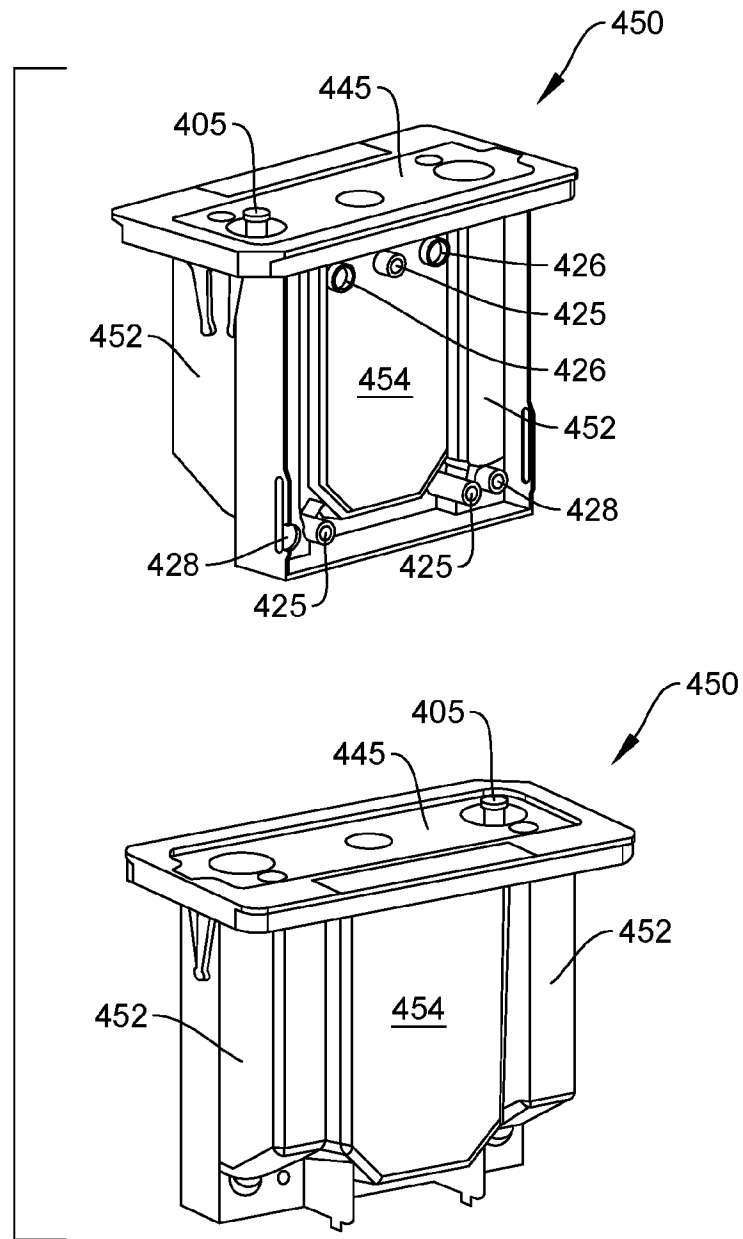


FIG. 4B

14/31

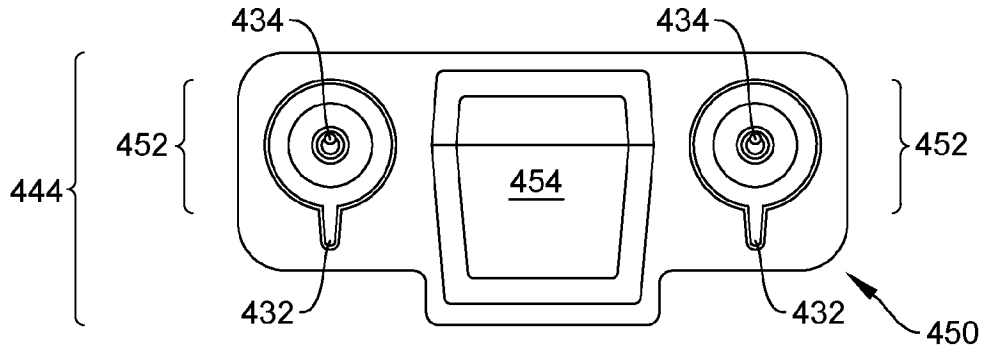


FIG. 4C

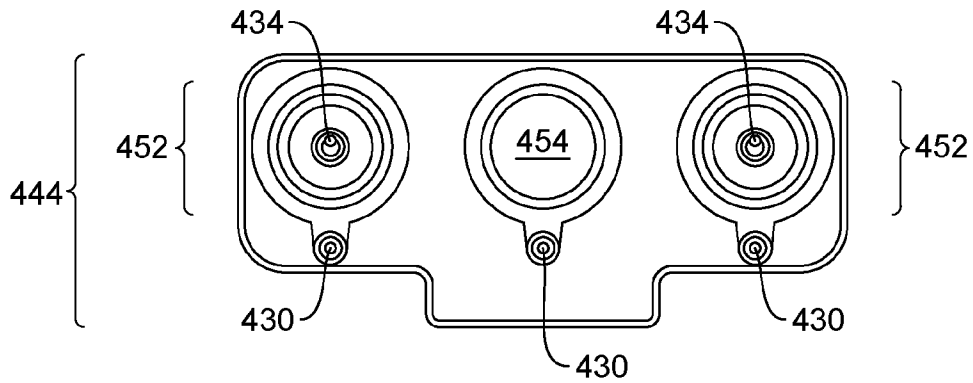


FIG. 4D

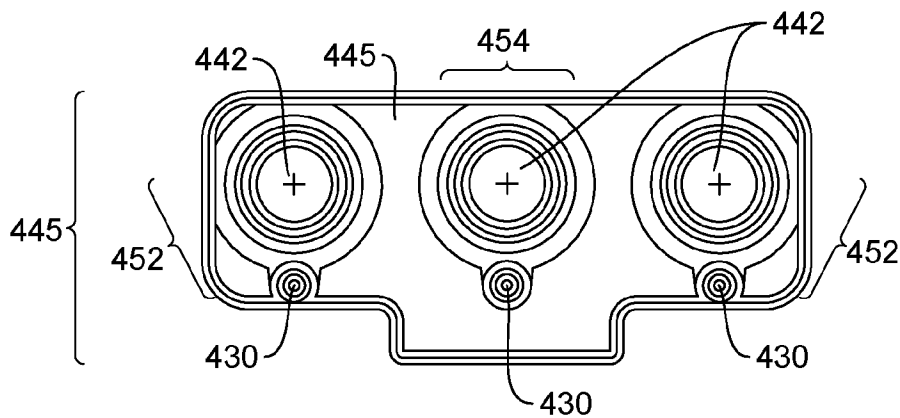


FIG. 4E

15/31

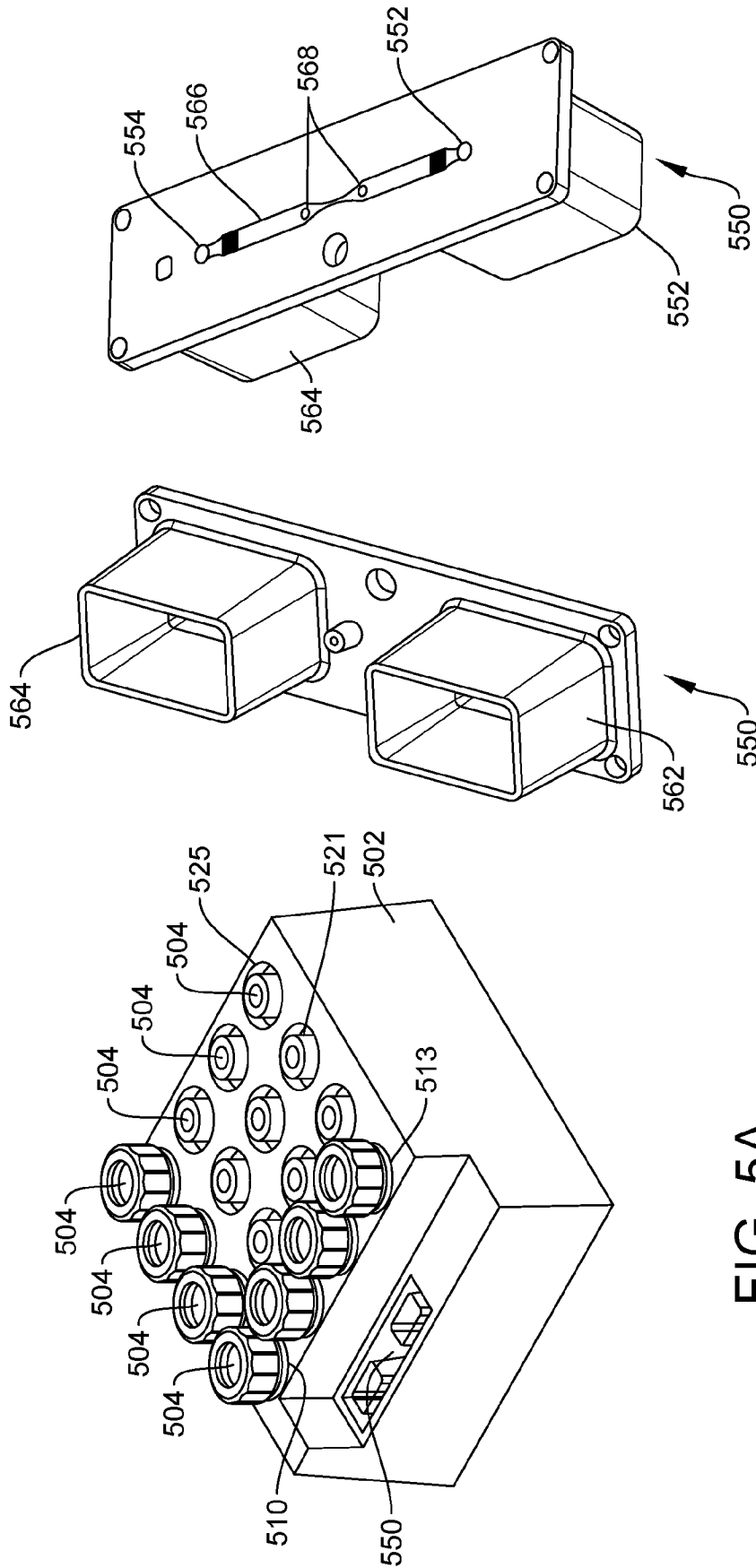


FIG. 5A

FIG. 5B

FIG. 5C

16/31

FIG. 5D

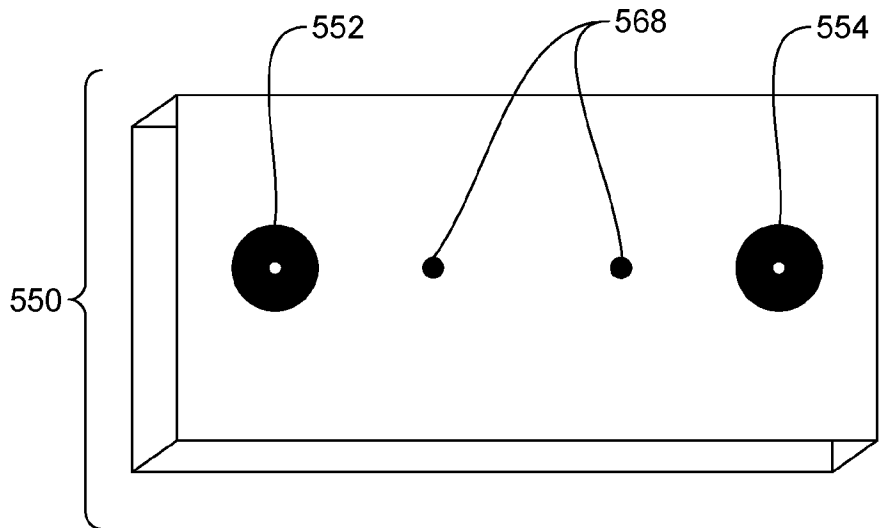


FIG. 5E

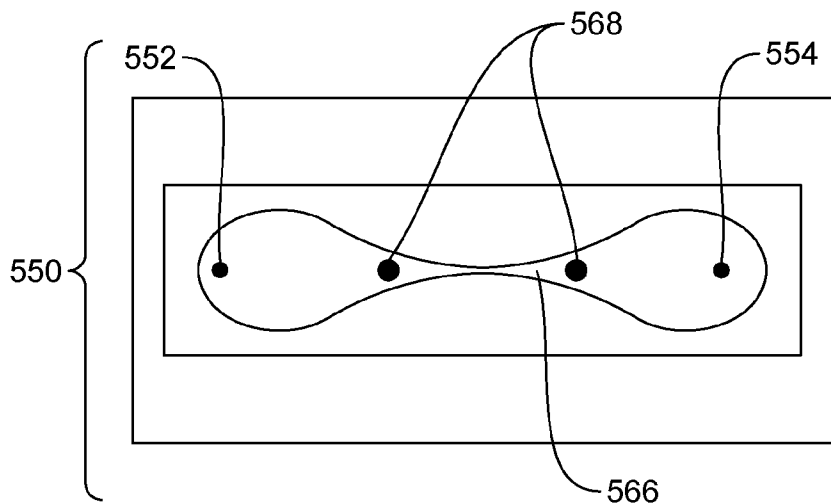
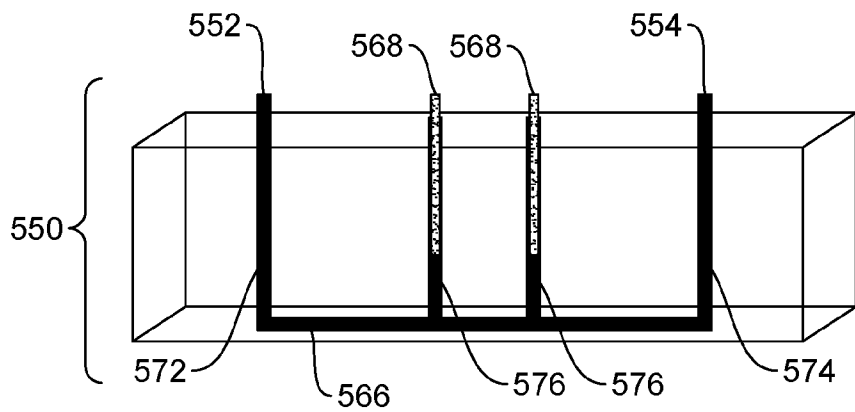


FIG. 5F



18/31

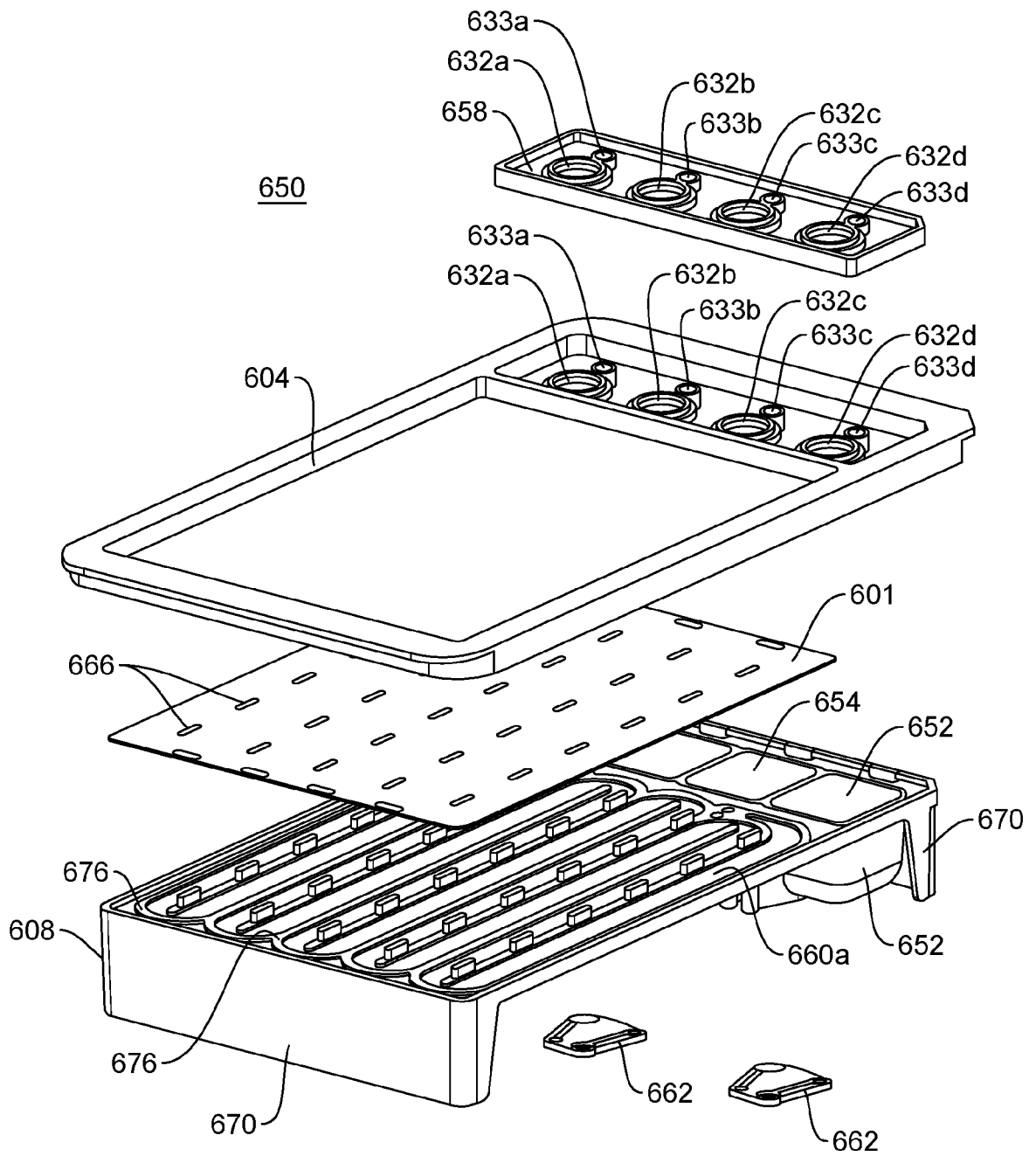


FIG. 6B

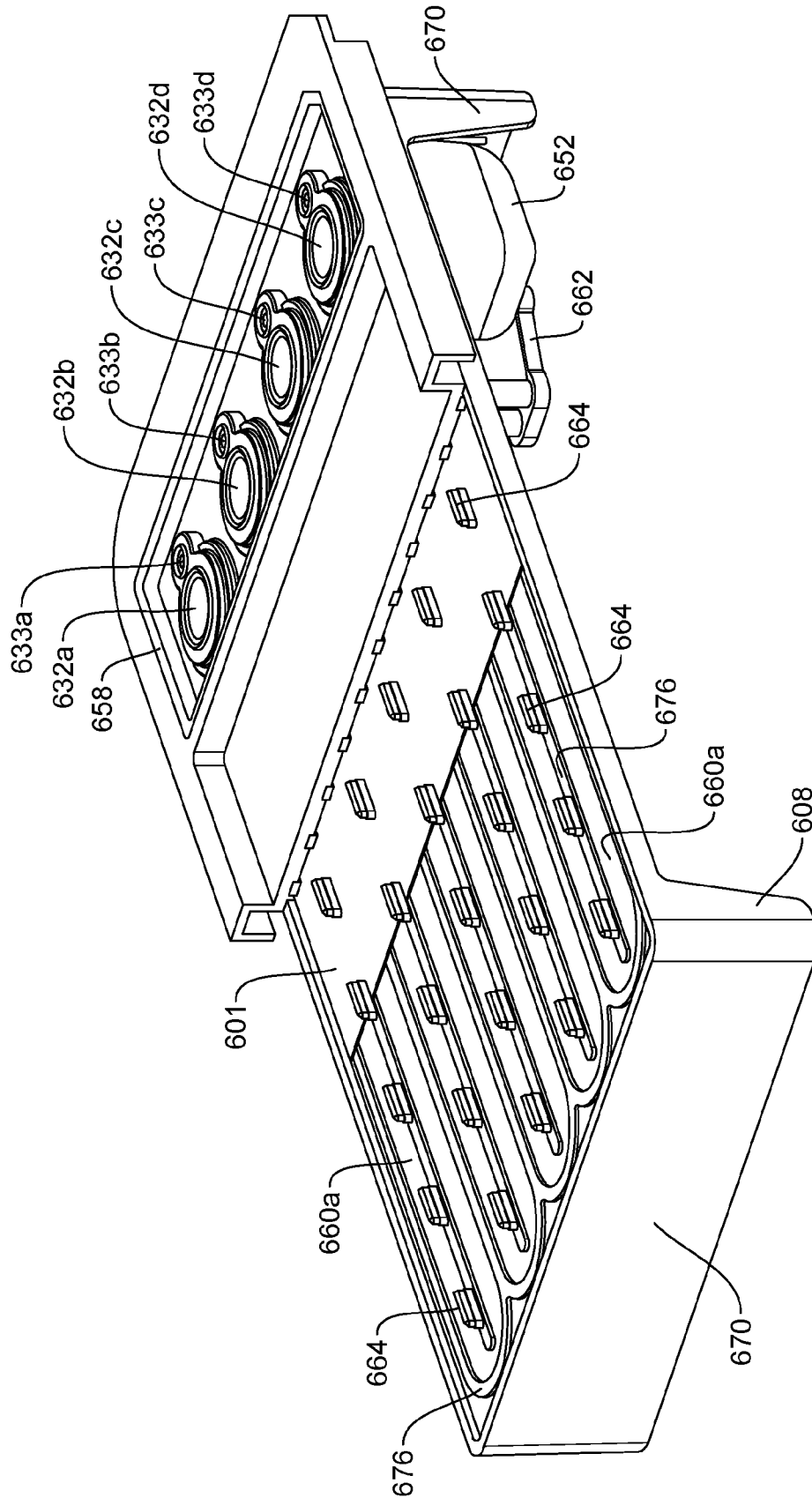


FIG. 6C

20/31

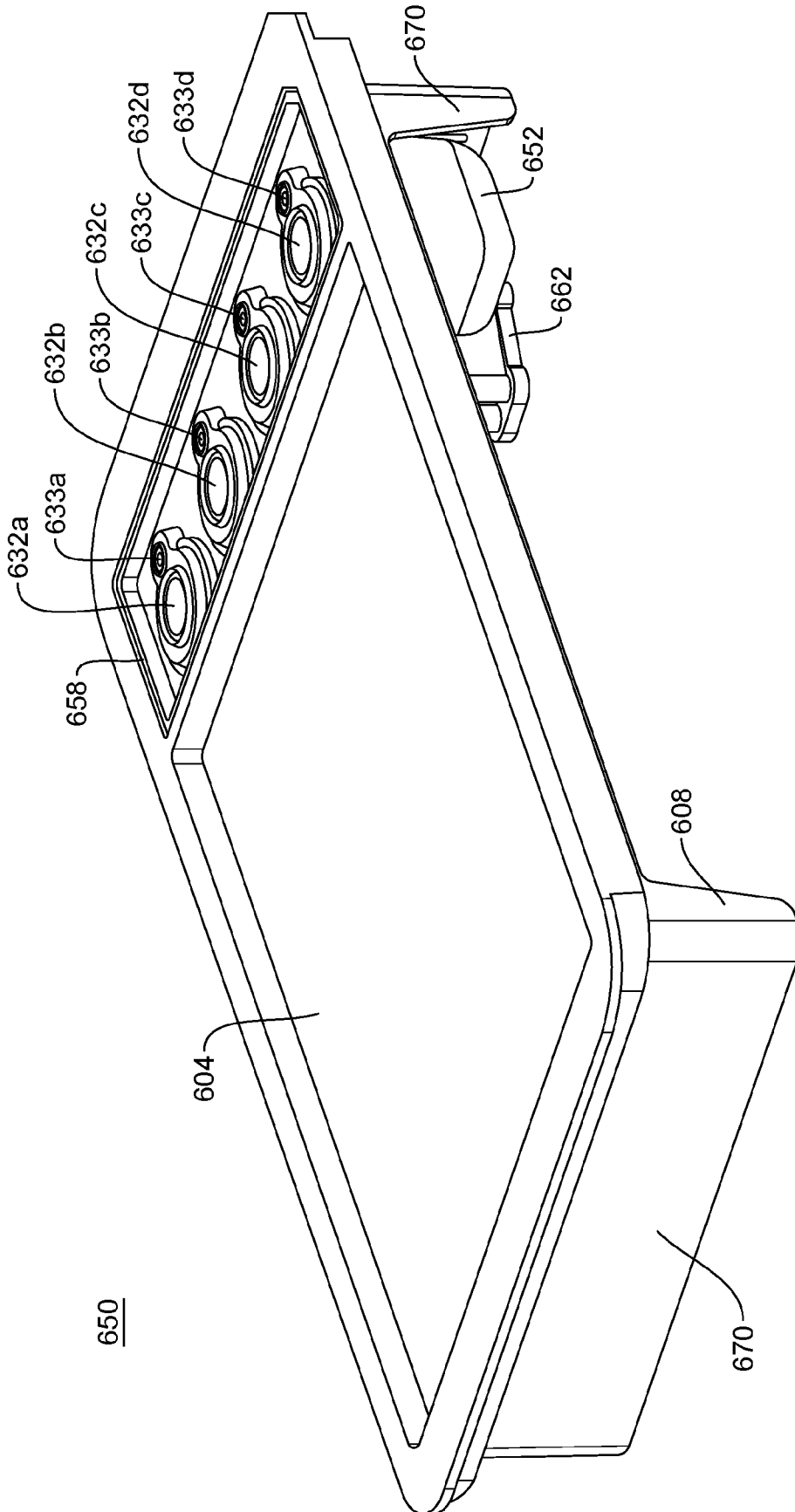


FIG. 6D

21/31

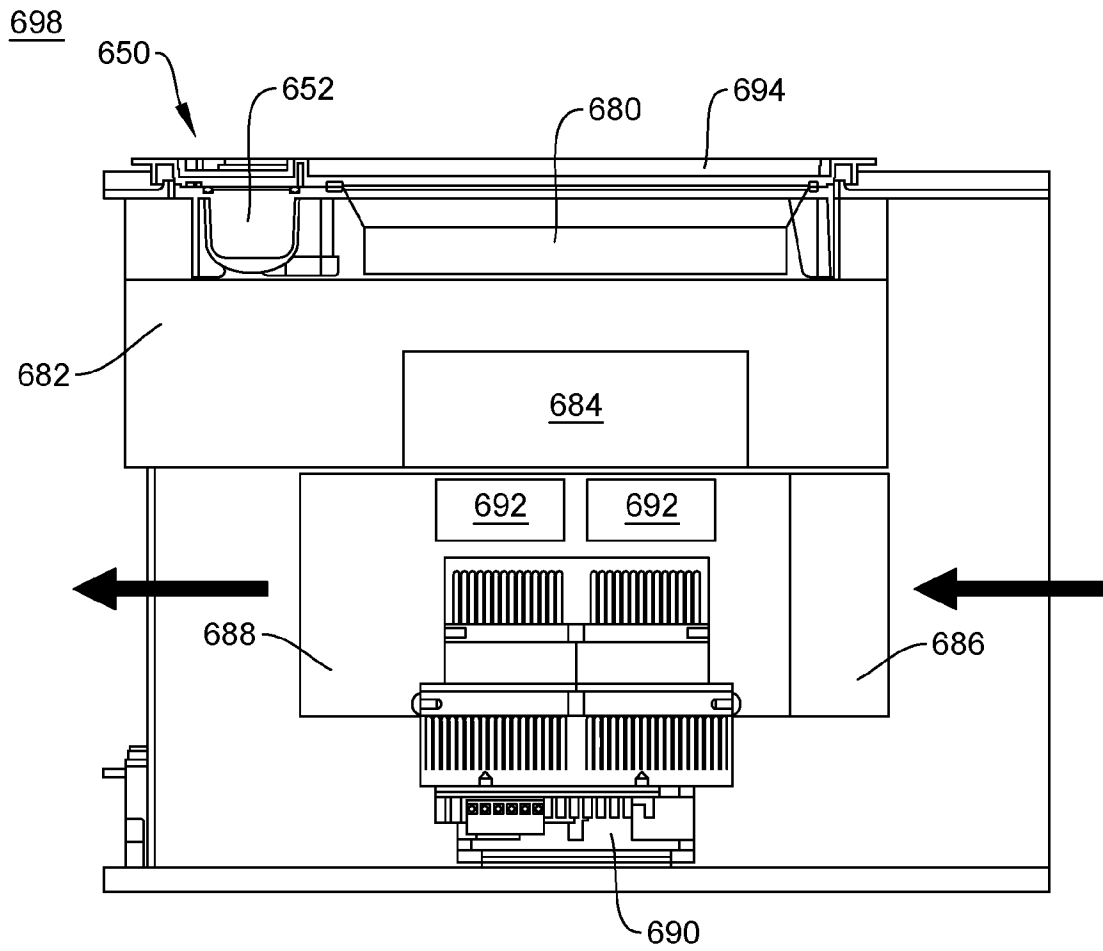


FIG. 6E

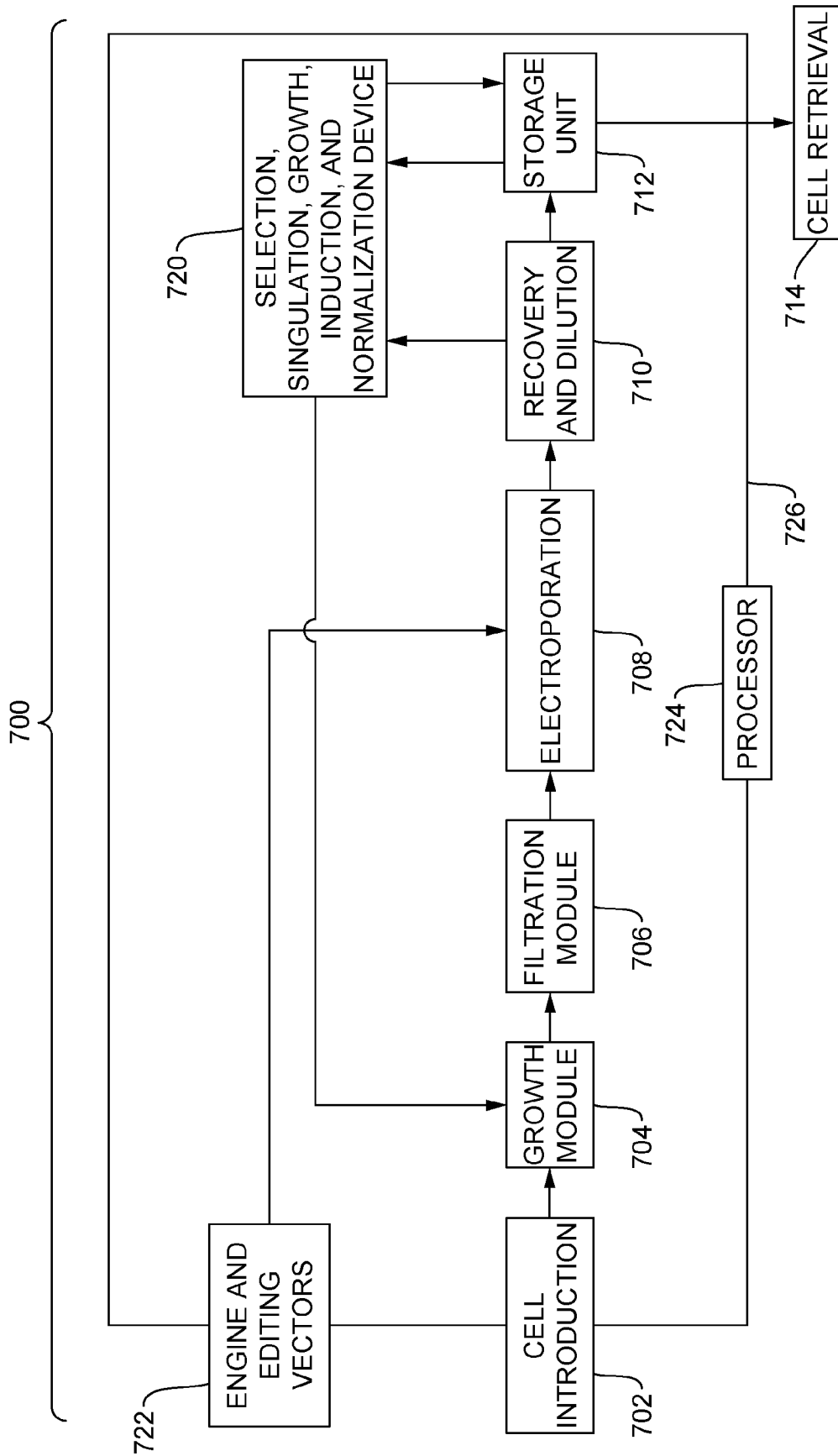


FIG. 7

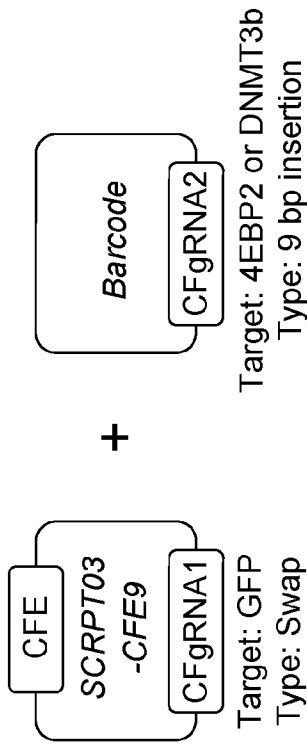


FIG. 8A

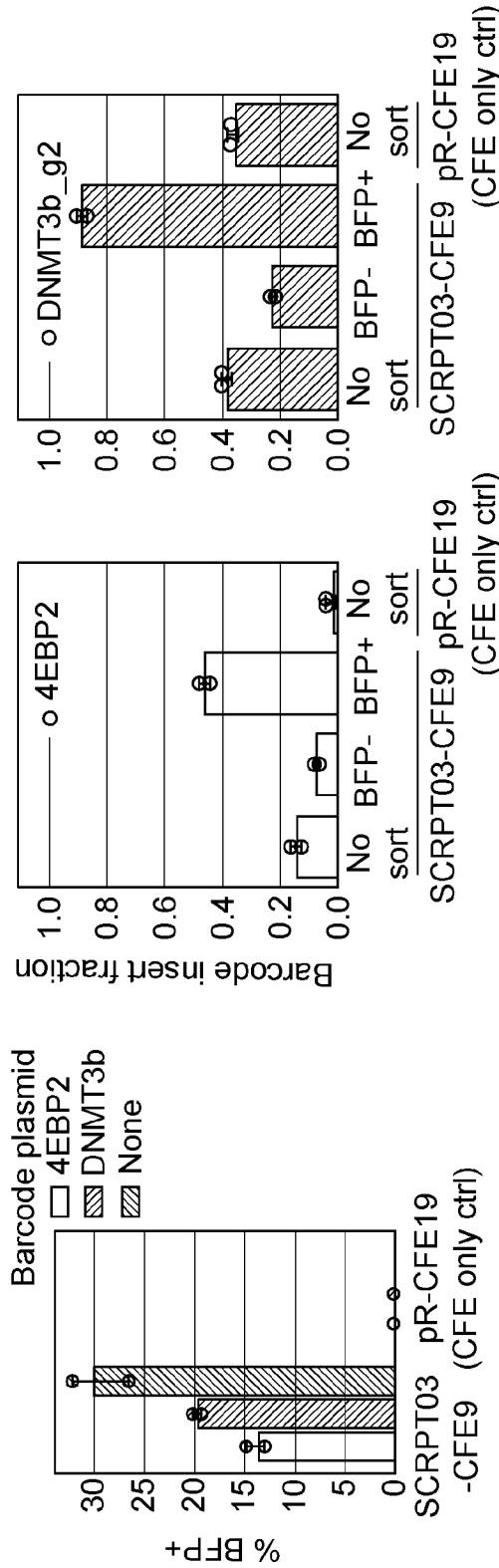


FIG. 8B

FIG. 8C

CREATE-fusion barcode insertion proof-of-concept. A) Editing and barcoding plasmids were co-transfected into iPS-GFP cells. Editing plasmids encode nuclease and CFgRNA targeting a model GFP locus (GFP-to-BFP swap). Barcode plasmids encode CFgRNA containing a 9bp insertion targeting either 4EBP2 or DNMT3b model barcode sites. B) Edit rates in iPS-GFP cells transfected with single (edit only) or dual (edit + barcode) plasmids. C) Cells in (B) were sorted on BFP+/- gates by FACS and 9bp barcode insertion rates were measured by NGS. 45% or 89% of edited cells (BFP+) were successfully barcoded at the 4EBP2 or DNMT3b locus, respectively (arrows).

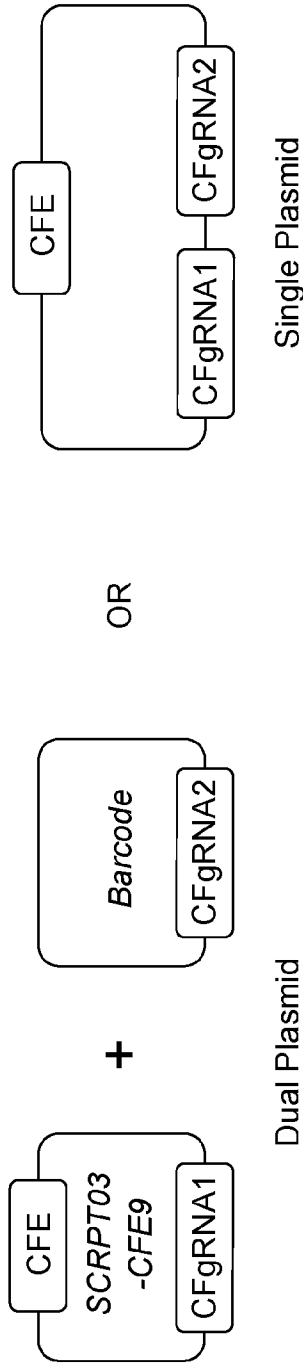


FIG. 9A

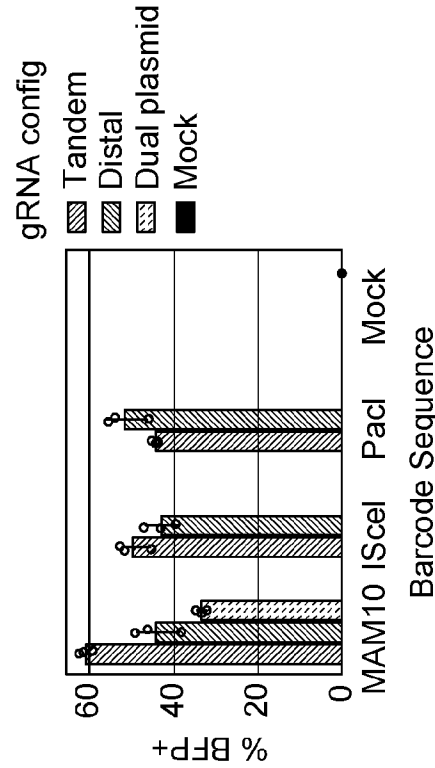


FIG. 9B

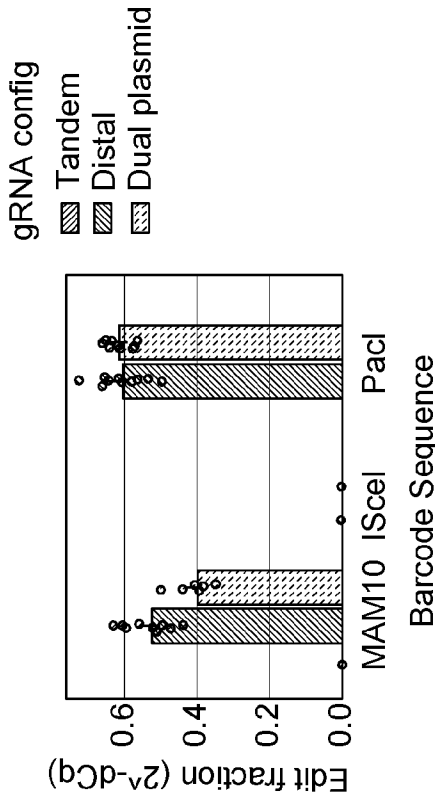


FIG. 9C

CREATE-fusion barcode insertion proof-of-concept. A) Editing and barcoding plasmids were transfected into iPS-GFP cells targeting the GFP locus (GFP-to-BFP swap). Barcoding guides contained either a 9bp insertion (MAM10 and Paci) or 18 bp insertions (IScel) targeting DNMT3b as a model barcode site. Editing and barcoding guides were either delivered on a single plasmid (in "tandem" or "dual" configuration) or co-delivered on dual plasmids B-C) Edit and barcode rate are improved with single plasmid vs. dual plasmid transfection. Barcoding rates are sensitive to barcode insertion sequence and length, with no detectable IScel (18bp) inserts detected by qPCR.

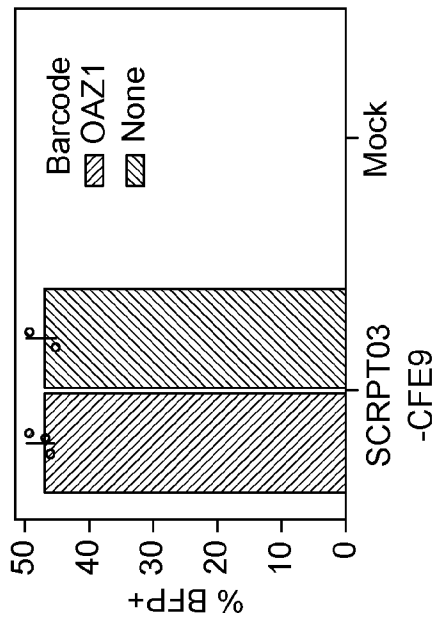


FIG. 10A

Sample	BC	Sort gate	%BFP+	RNAseq OAZ1 Edit%
1	OAZ1	BFP+	89.9%	76%
2	OAZ1	BFP+		80%
3	OAZ1	BFP-	0.27%	18%
4	OAZ1	BFP-		19%
5	None (mock)	NA	NA	0%
6	None (mock)	NA	NA	0%

FIG. 10B

Detection of CREATE-fusion barcodes by RNAseq. A) Editing and barcoding plasmids were co-transfected into iPS-GFP cells. Editing plasmids encode nuclease and CFgRNA targeting a model GFP locus (GFP-to-BFP swap). Barcode plasmids encode CFgRNA containing a 2bp swap targeting OAZ1. GFP-to-BFP edit rates are similar between samples transfected with barcoding plasmids vs. negative controls. B) Cells in (A) were sorted on BFP+/- gates by FACS and barcode insertion rates at OAZ1 transcripts were measured by bulk RNAseq.

26/31

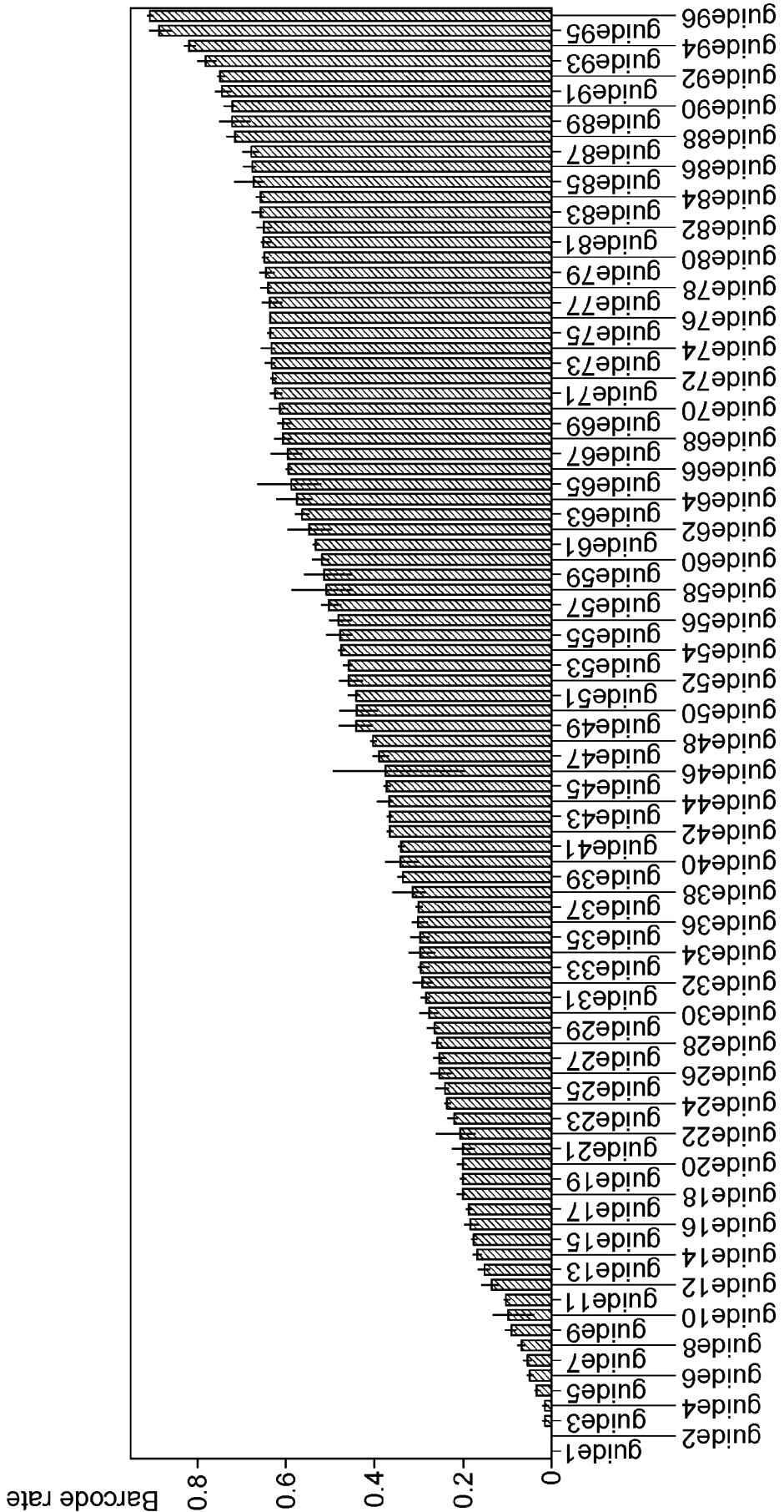


FIG. 11

27/31

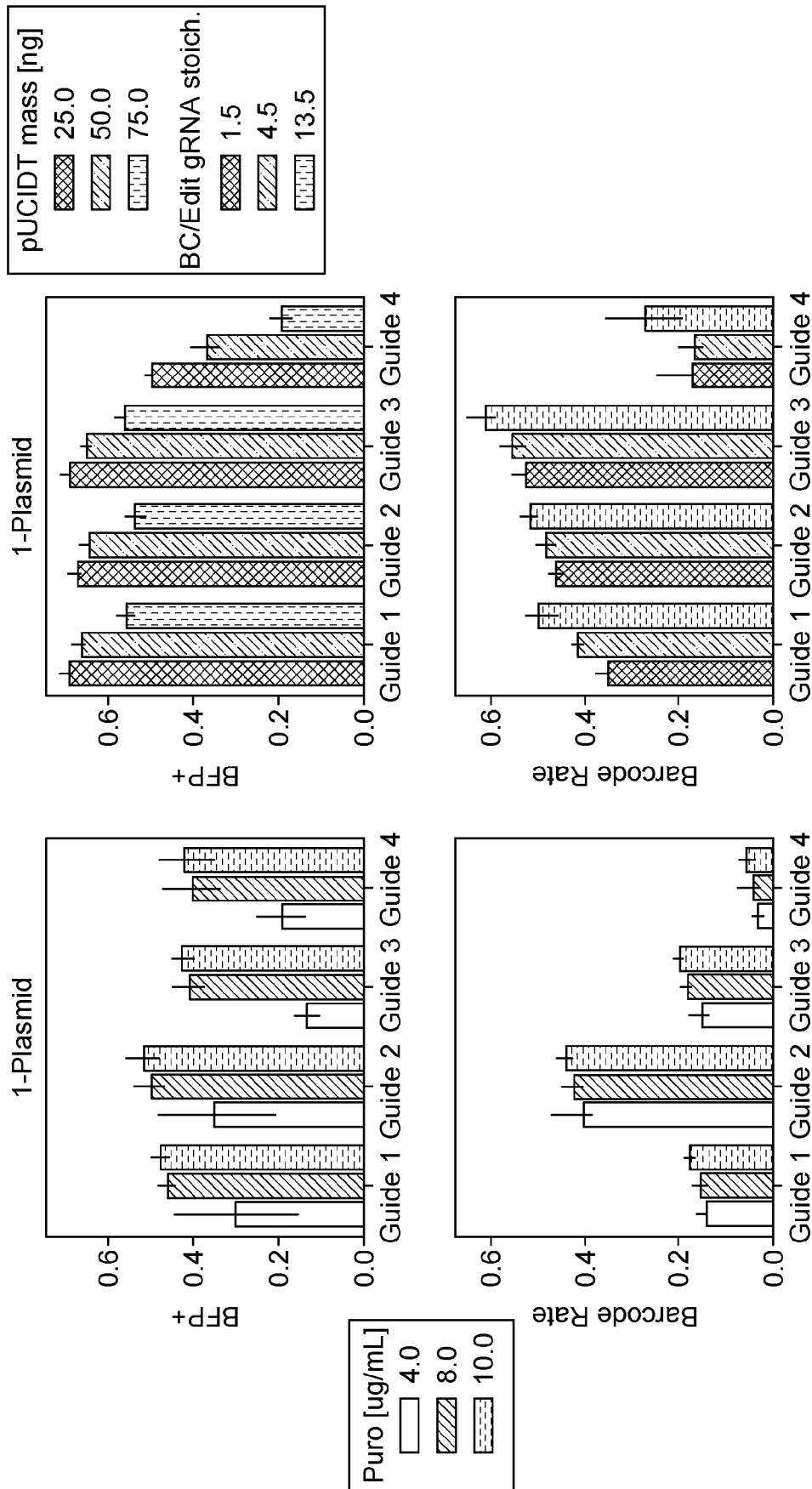


FIG. 12

28/31

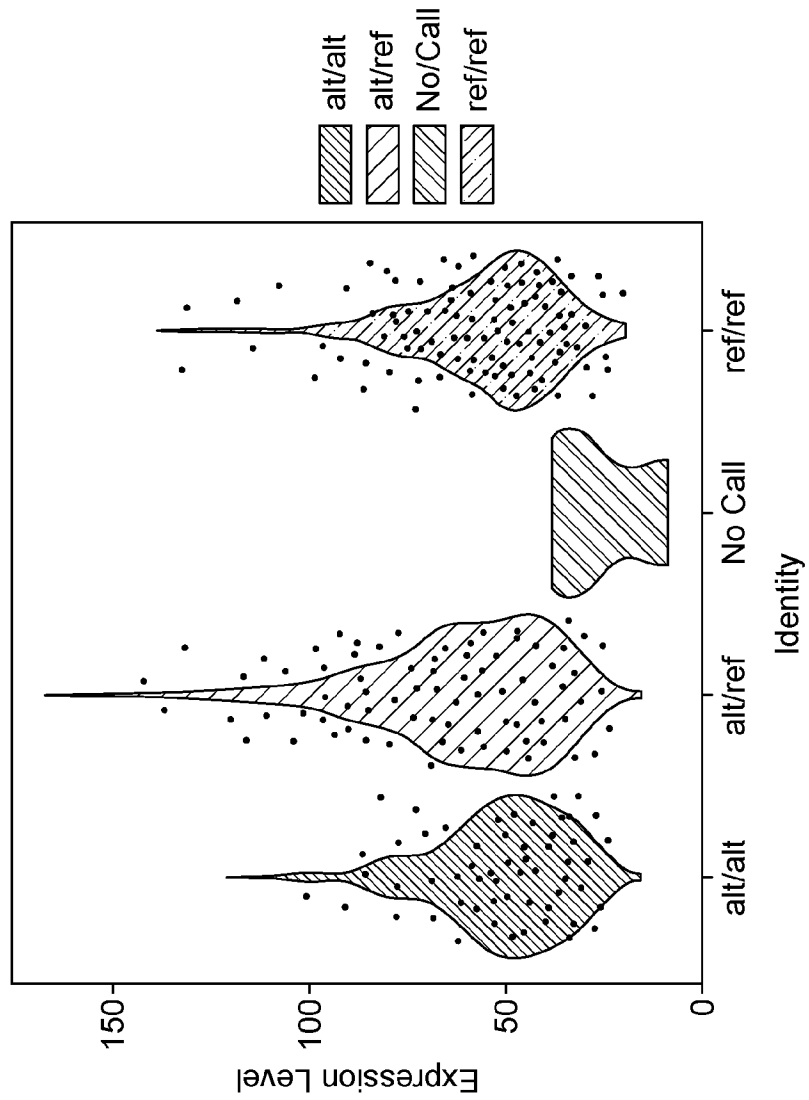


FIG. 13

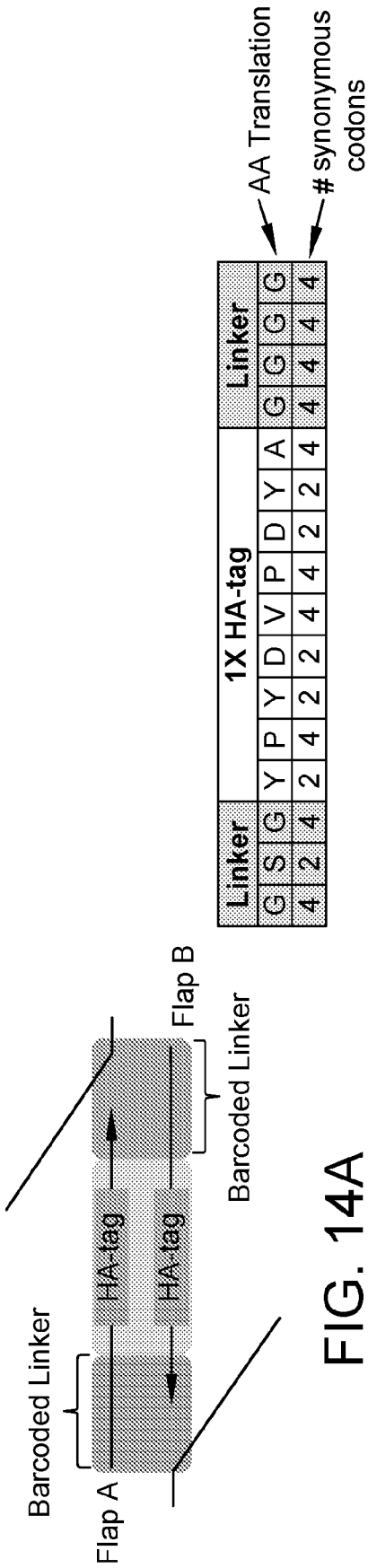


FIG. 14B

FIG. 14A

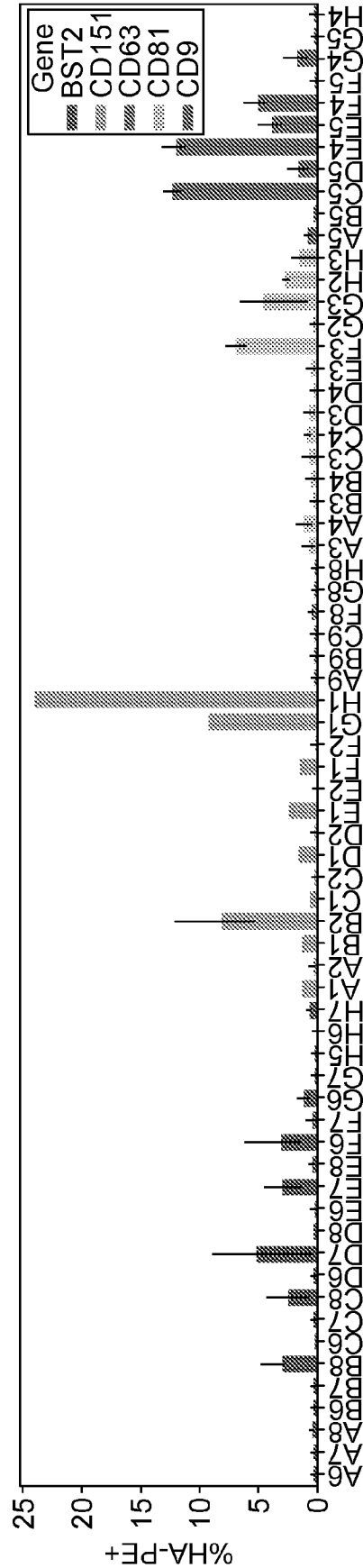


FIG. 14C

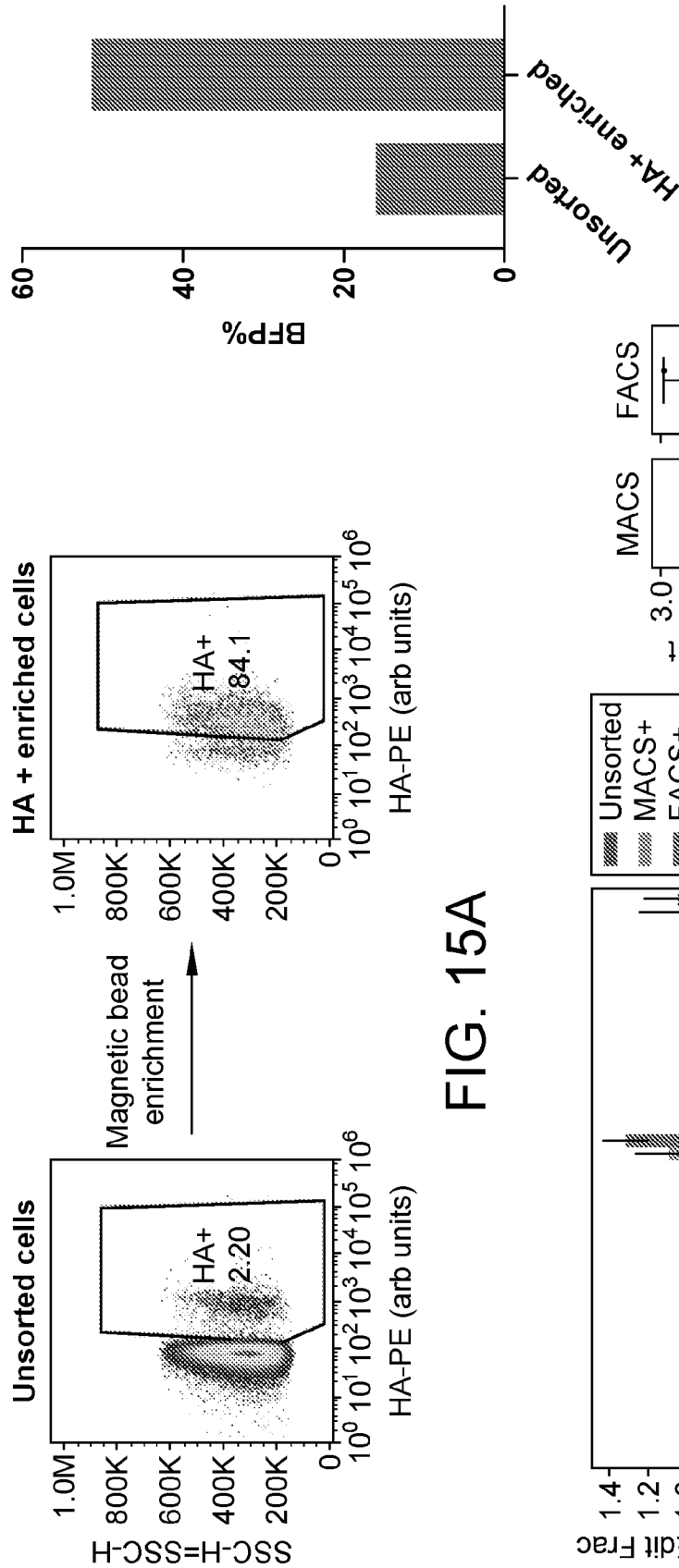


FIG. 15A

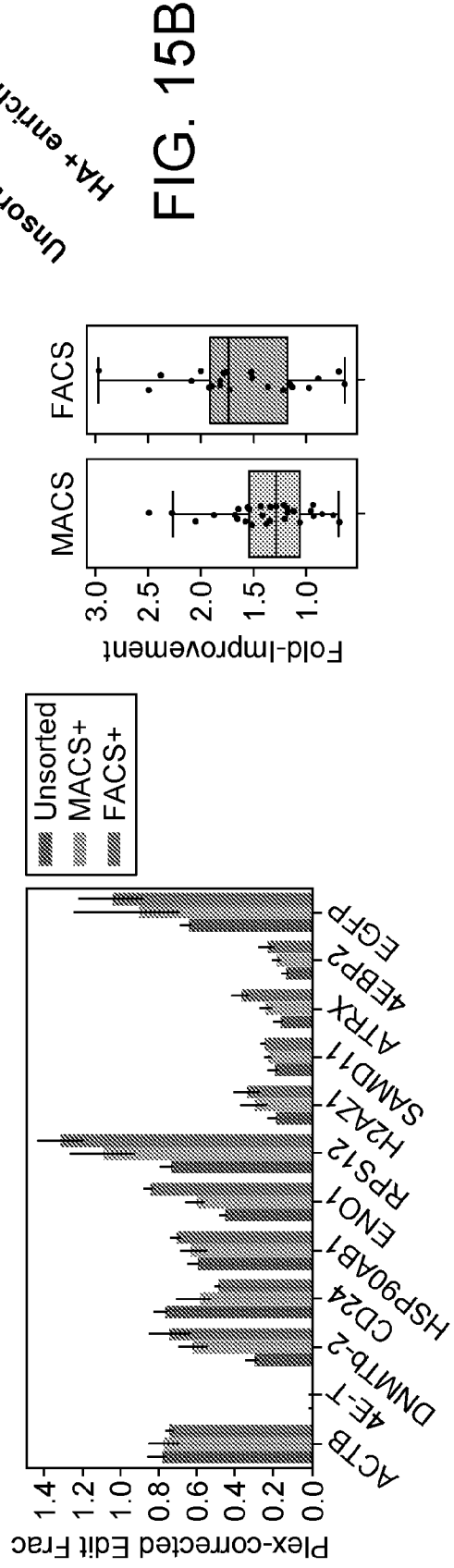


FIG. 15C

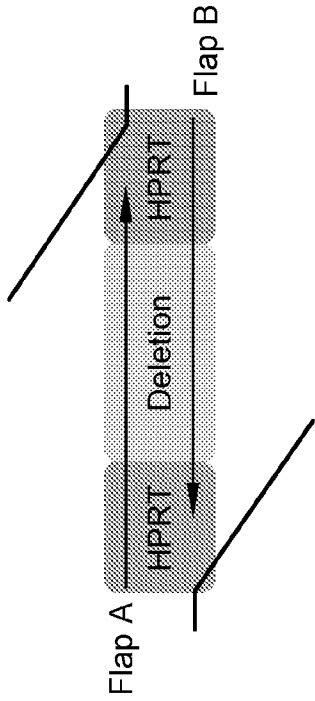


FIG. 16B

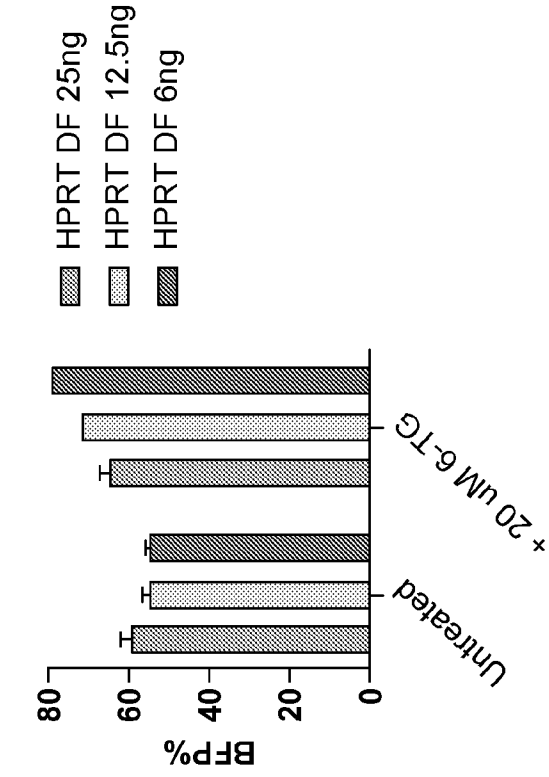


FIG. 16D

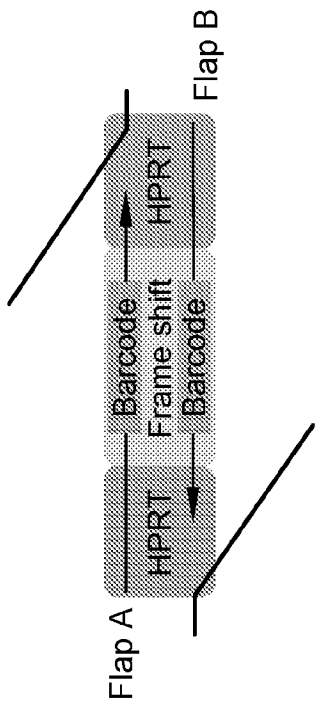


FIG. 16A

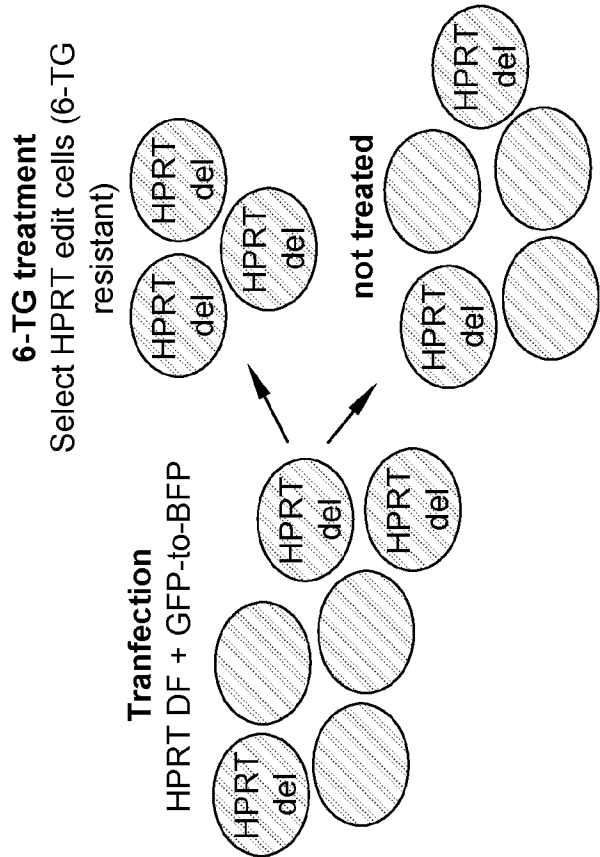


FIG. 16C

INTERNATIONAL SEARCH REPORT

International application No PCT/US2022/053377
--

A. CLASSIFICATION OF SUBJECT MATTER		
INV. C12N15/10	C12N9/12	C12N9/22
ADD.		
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) C12N		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) EPO-Internal, BIOSIS, EMBASE, FSTA		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 10 689 669 B1 (FELDMAN EMILY [US] ET AL) 23 June 2020 (2020-06-23) the whole document, in particular Example X and Example XIII -----	1-20
X	WO 2021/142394 A1 (INSCRIPTA INC [US]) 15 July 2021 (2021-07-15) the whole document, in particular Example X and Example XIII -----	1-20
X	US 2021/214671 A1 (MIR AAMIR [US] ET AL) 15 July 2021 (2021-07-15) the whole document, in particular Example X and Example XIII -----	1-20
A	WO 2021/092204 A1 (SPOTLIGHT THERAPEUTICS [US]) 14 May 2021 (2021-05-14) ----- -/--	1-20
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> 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	"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	
"E" earlier application or patent but published on or after the international filing date	"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	
"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)	"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	
"O" document referring to an oral disclosure, use, exhibition or other means	"&" document member of the same patent family	
"P" document published prior to the international filing date but later than the priority date claimed		
Date of the actual completion of the international search	Date of mailing of the international search report	
13 March 2023	21/03/2023	
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Authorized officer Bassias, Ioannis	

INTERNATIONAL SEARCH REPORT

International application No PCT/US2022/053377
--

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>WO 2021/226558 A1 (BROAD INST INC [US]; HARVARD COLLEGE [US]) 11 November 2021 (2021-11-11) -----</p>	<p align="center">1-20</p>
A	<p>ANZALONE ANDREW V ET AL: "Search-and-replace genome editing without double-strand breaks or donor DNA", NATURE, NATURE PUBLISHING GROUP UK, LONDON, vol. 576, no. 7785, 21 October 2019 (2019-10-21), pages 149-157, XP036953141, ISSN: 0028-0836, DOI: 10.1038/S41586-019-1711-4 [retrieved on 2019-10-21] -----</p>	<p align="center">1-20</p>

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No PCT/US2022/053377
--

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 10689669	B1	23-06-2020	NONE
WO 2021142394	A1	15-07-2021	AU 2021205422 A1 08-09-2022
			CA 3167640 A1 15-07-2021
			CN 115279899 A 01-11-2022
			EP 4087923 A1 16-11-2022
			WO 2021142394 A1 15-07-2021
US 2021214671	A1	15-07-2021	US 2021214671 A1 15-07-2021
			US 2021214672 A1 15-07-2021
WO 2021092204	A1	14-05-2021	NONE
WO 2021226558	A1	11-11-2021	AU 2021267940 A1 08-12-2022
			BR 112022022603 A2 17-01-2023
			CA 3177481 A1 11-11-2021
			EP 4146804 A1 15-03-2023
			IL 297761 A 01-12-2022
			KR 20230019843 A 09-02-2023
			WO 2021226558 A1 11-11-2021