



(51) International Patent Classification:

Not classified

(21) International Application Number:

PCT/US2022/078821

(22) International Filing Date:

27 October 2022 (27.10.2022)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

63/272,649 27 October 2021 (27.10.2021) US
63/407,593 16 September 2022 (16.09.2022) US
63/416,766 17 October 2022 (17.10.2022) US

(71) Applicant: **MISSION BIO, INC.** [US/US]; 400 E Jamie Ct, Suite 100, South San Francisco, California 94080 (US).

(72) Inventors: **DHINGRA, Dalia**; c/o Mission Bio, Inc., 400 E Jamie Ct, Suite 100, South San Francisco, California 94080 (US). **SCIAMBI, Adam**; c/o Mission Bio, Inc., 400 E Jamie

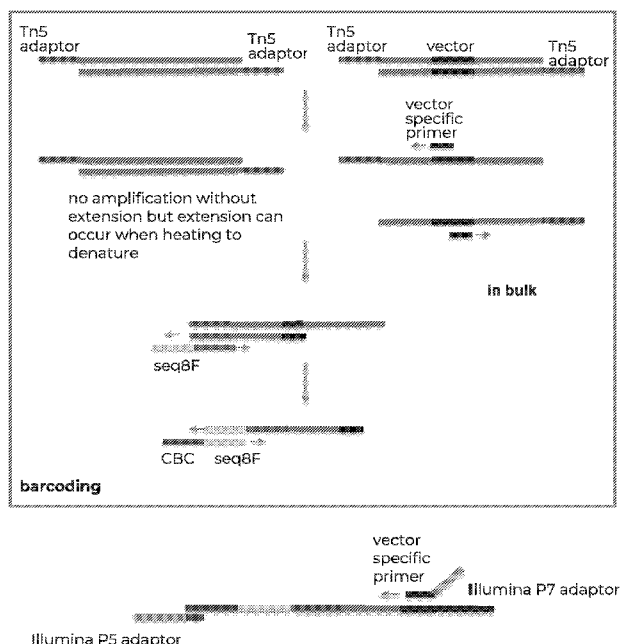
Ct, Suite 100, South San Francisco, California 94080 (US). **LI, Chieh-Yuan**; c/o Mission Bio, Inc., 400 E Jamie Ct, Suite 100, South San Francisco, California 94080 (US).

(74) Agent: **ZHANG, Clark** et al.; Goodwin Procter LLP, 100 Northern Avenue, IP Docketing Dept./7th Fl, Boston, Massachusetts 02210 (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.

(54) Title: SINGLE CELL VIRAL INTEGRATION SITE DETECTION

FIG. 2



(57) Abstract: The invention provides single-cell analysis using combined DNA sequencing of genomic DNA, which is performed to determine whether a segment of foreign DNA has been integrated. Individual cells are encapsulated with reagents, including primers, and lysed. Cell lysates include genomic DNA and the genomic DNA contains an integration site including a foreign DNA segment, if present. Segments of DNA are primed, amplified, and sequenced to generate sequence reads of the genomic DNA. Sequence reads from the DNA-seq reveal the presence or absence of a foreign DNA segment, and the genomic locus of the integration site, if present.

(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).

Declarations under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
- *of inventorship (Rule 4.17(iv))*

Published:

- *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

SINGLE CELL VIRAL INTEGRATION SITE DETECTION**CROSS REFERENCE**

[0001] This application claims the benefit of priority of U.S. Provisional Application Ser. No. 63/272,649, filed October 27, 2021, U.S. Provisional Application Ser. No. 63/407,593, filed September 16, 2022, and U.S. Provisional Application Ser. No. 63/416,766, filed October 17, 2022, the contents of which are hereby incorporated by reference in their entirety for all purposes.

BACKGROUND

[0002] At the forefront of medicine, gene therapy using viral vectors and transposons provide genetic and cell-based technologies to treat diseases. Characterization of transposition or the integration of foreign DNA (*e.g.*, a DNA segment from a viral vector) is a valuable quality control to determine the cells or the proportion of cells that were successfully transduced. Moreover, the characterization of transposition or foreign DNA integration in bulk DNA or even in single cells would provide insight into the effects of these genetic rearrangements on cellular function. For example, single cell analysis provides single cell resolution for better understanding co-occurrence of specific integration sites with somatic genomic variations (*e.g.*, copy number variants (CNVs) and single nucleotide variants (SNVs)), as well as the advantage to select off-target integrations that could lead to clonal expansion. However, methods of determining foreign DNA integration or DNA transposition in single cells and in bulk has remained difficult to execute. An outstanding need exists for improved, scalable methods for determining transpositions and vector integration.

SUMMARY

[0003] Disclosed herein are methods for performing bulk DNA or single-cell analysis to detect viral integration sites. In various embodiments, viral nucleic acids are introduced and integrated into genomic DNA of a cell. Such viral nucleic acids can be a viral plasmid, modified viral plasmid, or nucleic acids from a virus. Examples of viruses include adeno-associated viruses (AAVs), adenoviruses, herpes simplex virus, and lentiviruses (*e.g.*, human immunodeficiency virus (HIV)). Methods disclosed herein involve detecting and/or

confirming the occurrence and optionally, genomic loci of vector integration without prior knowledge of the integration site loci. In cell and gene therapy, vector integration and site analysis pose safety concerns. Thus, methods disclosed herein identify the potential of adverse effects resulting from vector integration. The invention is also based, at least in part, on the unexpected advantage that the same methods can be adapted for bulk DNA as well as for use in detecting translocation or genetic editing of a DNA segment of genomic DNA of a cell. Thus, methods disclosed herein can be used to scale-up single cell or bulk DNA analyses for detecting vector integration, DNA translocations, or genetic editing of a DNA segment of interest.

[0004] In various embodiments, the single-cell analysis involves analyzing an analyte of a single cell to detect vector integration sites, DNA translocations, or genetic editing of a DNA segment of interest. In particular embodiments, the analyte of the single cell is DNA. For example, the DNA can be genomic DNA. As another example, the DNA can be foreign DNA, such as viral DNA. The methods disclosed herein enable detection of rare integration events and is not dependent on proximity to restriction enzyme or Alu priming sites. It can be combined with protein expression and other DNA readouts (e.g., vector copy number or single nucleotide variants) for a more comprehensive view of the vector integration. For example, protein expression analysis can be performed by staining cells with oligonucleotide-tagged antibodies prior to loading them on a single-cell analysis device (e.g., Tapestri®).

[0005] In various embodiments, the single-cell analysis involves performing tagmentation on the single cells. In various embodiments, tagmentation can be performed *in situ*, in a tube, in a first droplet, or in a second droplet. Here, tagmentation may not involve an extension step. In various embodiments, protease and a detergent are provided in a first droplet (or other reaction vessel such as, e.g., a well or a tube (collectively, “tube”)) for lysing a cell and/or digesting chromatin to release genomic DNA. Within a second droplet (or tube), polymerase chain reaction (PCR) is performed with a foreign DNA segment-specific primer with a different adaptor and a bridging primer to attach a cell barcode. In various embodiments, primers can be incorporated into the barcoding droplet (e.g., second droplet or a tube (e.g., a second tube)) that will amplify the vector and a control region enabling the determination of vector copy number. Additionally, because, in various embodiments, extension was not performed in the tagmentation reaction, there should be minimal amplification of the fragments that do not contain the vector sequence. In various embodiments, droplets are broken followed by library PCR and sequencing. The libraries

contain a portion of the host sequence as well as a portion of the vector sequence allowing for integration site confirmation.

[0006] Accordingly, in one aspect, the disclosure relates to a method for detecting integration of a vector including a foreign DNA segment into genomic DNA of a cell, the method including: providing, within a droplet, the genomic DNA of the cell and reagents, the genomic DNA including an integration site where the vector including the foreign DNA segment is integrated into the genomic DNA, wherein the reagents include a foreign DNA segment-specific primer and a second primer; within the droplet, generating one or more amplicons including the integrated foreign DNA segment, if present, using at least the hybridized foreign DNA segment-specific and second primers; and determining the presence or absence of the one or more amplicons, wherein integration of the foreign DNA segment in the genomic DNA of the cell is detected by determining the presence of the one or more amplicons; and wherein no integration of the foreign DNA segment is detected by determining the absence of the one or more amplicons.

[0007] In various embodiments of the foregoing aspect, using at least the hybridized foreign DNA segment-specific and second primers includes: hybridizing the foreign DNA segment-specific primer to the foreign DNA segment, if present in the integration site; extending the hybridized foreign DNA segment-specific primer to generate an extension product; and hybridizing the second primer to a sequence of the extension product.

[0008] In various embodiments of the foregoing aspect, the extension product includes a sequence derived from a transposase adapter sequence.

[0009] In various embodiments of the foregoing aspect, the transposase is a Tn5 transposase.

[0010] In various embodiments of the foregoing aspect, the transposase adapter is a Tn5 transposase adapter.

[0011] In various embodiments of the foregoing aspect, the sequence of the extension product includes a sequence derived from the genomic DNA.

[0012] In various embodiments of the foregoing aspect, using at least the hybridized foreign DNA segment-specific and second primers includes: hybridizing the foreign DNA segment-specific primer to the foreign DNA segment, if present in the integration site, and hybridizing the second primer to a sequence present in the genomic DNA or to a sequence present in the foreign DNA segment.

[0013] In another aspect, the disclosure relates to a method for detecting integration of a vector including a foreign DNA segment into genomic DNA of a cell, the method including: providing, within a droplet, the genomic DNA of the cell and reagents, the genomic DNA including an integration site where the vector including the foreign DNA segment is

5 integrated into the genomic DNA, wherein the reagents include a foreign DNA segment-specific primer; within the droplet, generating one or more amplicons including the integrated foreign DNA segment, if present, using at least the hybridized foreign DNA segment-specific primer; and determining the presence or absence of the one or more amplicons, wherein integration of the foreign DNA segment in the genomic DNA of the cell is detected by
10 determining the presence of the one or more amplicons; and wherein no integration of the foreign DNA segment is detected by determining the absence of the one or more amplicons.

[0014] In various embodiments of either of the foregoing aspects, the method further includes sequencing or determining the length of the one or more amplicons.

[0015] In various embodiments of either of the foregoing aspects, the method further
15 includes analyzing the one or more amplicons sequence and/or the one or more amplicons size to identify the amplicon identity, the genomic locus of the integration site, the number of integration sites, or the orientation of the integration, optionally wherein the number of integration sites includes the vector copy number.

[0016] In another aspect, the disclosure relates to a method for detecting a proportion of
20 cells in a population of cells having integration of a vector including a foreign DNA segment into genomic DNA of the cells, the method including: for each of one or more cells in the population of cells: providing, within a droplet, the genomic DNA of the cell and reagents, the genomic DNA including an integration site where the vector including the foreign DNA segment is integrated into the genomic DNA, wherein the reagents include a foreign DNA
25 segment-specific primer and a second primer; within the droplet, generating one or more amplicons including the integrated foreign DNA segment, if present, using at least the hybridized foreign DNA segment-specific and second primers; and sequencing the generated one or more amplicons; and determining a proportion of the cells in the population of cells having integration of the foreign DNA segment in genomic DNA of the cells based on the
30 sequenced one or more amplicons.

[0017] In various embodiments of any of the foregoing aspects, within the droplet, the method further includes exposing the cell to the reagents, wherein the reagents include a protease and a detergent and lysing the cell using the protease and the detergent.

[0018] In various embodiments of any of the foregoing aspects, the detergent is a pluronic detergent.

[0019] In various embodiments of any of the foregoing aspects, sequencing the generated one or more amplicons further includes characterizing a number of integration sites in the genomic DNA.

[0020] In various embodiments of any of the foregoing aspects, the foreign DNA segment is viral DNA, modified viral DNA, or DNA from a viral vector.

[0021] In various embodiments of any of the foregoing aspects, the DNA from a viral vector includes a transgene encoding a protein of interest or a reporter gene.

[0022] In various embodiments of any of the foregoing aspects, the DNA from a viral vector includes a transgene encoding a protein of interest.

[0023] In various embodiments of any of the foregoing aspects, the method further includes transducing the cell or the population of cells with the viral DNA, the modified viral DNA, or a viral vector.

[0024] In various embodiments of any of the foregoing aspects, the viral DNA, modified viral DNA, or viral vector is derived from an adeno-associated virus (AAV), adenovirus, herpes simplex virus, lentivirus, retrovirus, poxvirus, baculovirus, or vaccinia virus.

[0025] In various embodiments of any of the foregoing aspects, the reagents include a cell buffer and/or a lysis buffer.

[0026] In various embodiments of any of the foregoing aspects, the lysis buffer includes one or more of a reverse primer, a protease, a detergent, an RNA reverse transcriptase, an RNAase inhibitor, a transposase, and a magnesium buffer.

[0027] In various embodiments of any of the foregoing aspects, the lysis buffer includes a protease, a detergent, a transposase, and a magnesium buffer.

[0028] In various embodiments, the transposase is preloaded with an adapter.

[0029] In various embodiments of any of the foregoing aspects, the magnesium buffer includes magnesium, Tris, potassium, [tris(hydroxymethyl)methylamino]propanesulfonic acid (TAPS), dimethylformamide (DMF), and/or poly(ethylene glycol) (PEG).

[0030] In various embodiments of any of the foregoing aspects, the droplet is a water-in-oil emulsion, wherein an oil solution of the water-in-oil emulsion includes one or more of an oil and a non-ionic surfactant.

[0031] In various embodiments of any of the foregoing aspects, the oil includes a fluoruous oil.

[0032] In various embodiments of any of the foregoing aspects, the non-ionic surfactant is a fluoruous non-ionic surfactant.

[0033] In various embodiments of any of the foregoing aspects, the reagents further include a barcode primer including a barcode identification sequence.

5 [0034] In various embodiments of any of the foregoing aspects, the barcode primer is a bead barcode primer.

[0035] In various embodiments, the second primer is a second foreign DNA segment-specific primer, and wherein the method further includes: hybridizing the foreign DNA segment-specific primer to a sequence derived from a transposase adapter sequence.

10 [0036] In various embodiments, the reagents include a transposase.

[0037] In various embodiments, the transposase is a Tn5 transposase.

[0038] In various embodiments, within the droplet, the method further includes tagging the genomic DNA using the reagents to obtain tagmented DNA fragments, wherein at least one of the tagmented DNA fragments includes the foreign DNA segment.

15 [0039] In various embodiments, extending includes extension of the at least one of the tagmented DNA fragments.

[0040] In various embodiments, tagging the genomic DNA using the reagents includes inserting adaptor sequences to obtain tagmented DNA fragments including the adaptor sequences.

20 [0041] In various embodiments, tagging the genomic DNA using the reagents does not include performing an extension to fill one or more gaps.

[0042] In various embodiments, each of the tagmented DNA fragments include at most one adaptor sequence.

25 [0043] In various embodiments, genomic DNA of the cell and reagents are provided in a first droplet that differs from the droplet in which the genomic DNA is tagmented.

[0044] In various embodiments, genomic DNA of the cell and reagents are provided in the same droplet as the droplet in which the genomic DNA is tagmented.

30 [0045] In various embodiments, the second primer is a repeat sequence-specific primer, and wherein the method further includes: hybridizing the repeat sequence-specific primer to a repeat sequence present in the genomic DNA.

[0046] In various embodiments, repeat sequence-specific primer is an Alu1, an Alu2, a LINE1, an 16S, an 18S primer, or any combination thereof.

[0047] In various embodiments of any of the foregoing aspects, extending includes performing nucleic acid extension.

[0048] In various embodiments of any of the foregoing aspects, performing nucleic acid extension includes performing primer extension.

5 [0049] In various embodiments, performing nucleic acid extension includes extending the foreign DNA segment-specific primer to produce the one or more amplicons including a constant region sequence and the foreign DNA segment-specific primer.

[0050] In various embodiments, performing nucleic acid extension further includes producing the one or more amplicons including a complement sequence of the foreign DNA
10 segment.

[0051] In various embodiments, performing nucleic acid extension includes extending the barcode identification sequence to produce the one or more amplicons including a first read sequence, the barcode identification sequence, and a constant region sequence.

[0052] In various embodiments, performing nucleic acid extension includes extending the
15 second foreign DNA segment-specific primer to produce the one or more amplicons including the second foreign DNA segment-specific primer and a second read sequence.

[0053] In various embodiments, performing nucleic acid extension includes extending the repeat sequence-specific primer to produce the one or more amplicons including a constant region sequence and the repeat sequence-specific primer.

20 [0054] In various embodiments of any of the foregoing aspects, the reagents further include a read 1 sequencing primer and/or a read 2 sequencing primer.

[0055] In various embodiments of any of the foregoing aspects, the method further includes breaking an emulsion that includes the droplet and performing nucleic acid extension, wherein performing nucleic acid extension includes performing polymerase chain
25 reaction (PCR).

[0056] In various embodiments of any of the foregoing aspects, PCR includes extending the read 1 sequencing primer to produce the one or more amplicons including a first index sequence and a first read sequence.

[0057] In various embodiments of any of the foregoing aspects, performing PCR includes
30 extending the read 2 sequencing primer to produce the one or more amplicons including the second read sequence and a second index sequence.

[0058] In various embodiments of any of the foregoing aspects, the foreign DNA segment includes an inverted terminal repeat region (ITR), a rep gene, a cap gene, a long terminal

repeat (LTR) region, a gag gene, a pol gene, a tat gene, a rev gene, a IX gene, a IVa2 gene, an L1 gene, an L2 gene, an L3 gene, an L4 gene, an L5 gene, an E2B gene, an E2A gene, an E2A-L gene, an E4 gene, a gene encoding a capsomer protein, a gene encoding a capsid protein, a gene encoding a core protein, a gene encoding a viral non-structural protein, or a gene encoding a viral packing protein.

[0059] In various embodiments of any of the foregoing aspects, the foreign DNA segment includes an LTR.

[0060] In various embodiments of any of the foregoing aspects, the foreign DNA segment-specific primer or the second foreign DNA segment-specific primer includes the nucleic acid sequence of any one of SEQ ID NOs: 1-11.

[0061] In various embodiments of any of the foregoing aspects, the repeat sequence-specific primer includes the nucleic acid sequence of any one of SEQ ID NOs: 12-25.

[0062] In some embodiments, the one or more amplicons include from 5'-to-3': the first index sequence, the first read sequence, the barcode identification sequence, the constant region sequence, the foreign DNA segment-specific primer, the complement sequence of the foreign DNA segment, the second foreign DNA segment-specific primer, the second read sequence, and the second index sequence.

[0063] In some embodiments, the one or more amplicons include from 5'-to-3': the first index sequence, the barcode identification sequence, the constant region sequence, the foreign DNA segment-specific primer, the complement sequence of the foreign DNA segment, the second foreign DNA segment-specific primer, and the second index sequence.

[0064] In some embodiments, the one or more amplicons include from 5'-to-3': the first index sequence, the first read sequence, the barcode identification sequence, the constant region sequence, the repeat sequence-specific primer, the complement sequence of the foreign DNA segment, the second foreign DNA segment-specific primer, the second read sequence, and the second index sequence.

[0065] In various embodiments of any of the foregoing aspects, the genomic DNA further includes one or more additional integration sites where copies of the foreign DNA segment are integrated into the genomic DNA.

[0066] In various embodiments of any of the foregoing aspects, the method further includes determining a vector copy number of the foreign DNA segment across the integration site and the one or more additional integration sites.

[0067] In various embodiments of any of the foregoing aspects, determining the vector copy number includes: identifying a first amplicon including a sequence of the foreign DNA segment and a second amplicon including a sequence of the foreign DNA segment, wherein the first amplicon and the second amplicon include different start sites; and determining
5 whether a portion of the sequence of the foreign DNA segment of the first amplicon overlaps with a portion of the sequence of the foreign DNA segment of the second amplicon.

[0068] In various embodiments of any of the foregoing aspects, the different start sites of the first amplicon and the second amplicon correspond to different Tn5 insertion sites.

[0069] In various embodiments of any of the foregoing aspects, the first amplicon and
10 second amplicon share a common termination site.

[0070] In various embodiments of any of the foregoing aspects, the common termination sites of the first amplicon and second amplicon correspond to the foreign DNA segment-specific primer.

[0071] In various embodiments of any of the foregoing aspects, responsive to the
15 determination that the portion of the sequence of the foreign DNA segment of the first amplicon overlaps with a portion of the sequence of the foreign DNA segment of the second amplicon, determining that the vector copy number is at least 2.

[0072] In various embodiments of any of the foregoing aspects, responsive to the determination that the portion of the sequence of the foreign DNA segment of the first
20 amplicon does not overlap with a portion of the sequence of the foreign DNA segment of the second amplicon, determining that the vector copy number is 1.

[0073] In some embodiments, the method further includes determining one or more mutations of the cell or the population of cells.

[0074] In some embodiments, the one or more mutations include a single nucleotide
25 variant (SNV) or a copy number variation (CNV).

[0075] In some embodiments, the one or more mutations include a SNV and a CNV.

[0076] In some embodiments, the method further includes determining one or more analytes expressed by the cell or the population of cells.

[0077] In some embodiments, the cell or the population of cells are bound to at least one
30 analyte-bound antibody-conjugated oligonucleotide.

[0078] In some embodiments, the antibody-conjugated oligonucleotide includes a PCR handle, a tag sequence, and a capture sequence.

[0079] In some embodiments, determining one or more mutations includes: performing a nucleic acid amplification reaction within the droplet using the antibody-conjugated oligonucleotide to generate an additional one or more amplicons, the additional one or more amplicons including an amplicon derived from the oligonucleotide; determining a presence or absence of an analyte using the second one or more amplicons; and characterizing the presence or absence of the analyte.

[0080] In some embodiments, determining presence or absence of the analyte includes determining an expression level of the analyte, the analyte bound by the antibody conjugated to the oligonucleotide.

[0081] In various embodiments of any of the foregoing aspects, the method further includes generating a targeted DNA library or a targeted protein library.

[0082] In another aspect, the disclosure relates to a method for detecting integration of a vector including a foreign DNA segment into genomic DNA of a cell, the method including: providing, in a bulk setting, the genomic DNA of the cell and reagents, the genomic DNA including an integration site where the vector including the foreign DNA segment is integrated into the genomic DNA, wherein the reagents include a foreign DNA segment-specific primer and a second primer; in a bulk setting, generating one or more amplicons including the integrated foreign DNA segment, if present, using at least the hybridized foreign DNA segment-specific and second primers; and determining the presence or absence of the one or more amplicons, wherein integration of the foreign DNA segment in the genomic DNA of the cell is detected by determining the presence of the one or more amplicons; and wherein no integration of the foreign DNA segment is detected by determining the absence of the one or more amplicons.

[0083] In another aspect, the disclosure relates to a method for detecting translocation of a DNA segment in genomic DNA of a cell, the method including: providing, within a droplet, the genomic DNA of the cell and reagents, the genomic DNA including an integration site where the translocated DNA segment is integrated into the genomic DNA, wherein the reagents include a translocated DNA segment-specific primer and a second primer; within the droplet, generating one or more amplicons including the integrated translocated DNA segment, if present, using at least the hybridized translocated DNA segment-specific and second primers; and determining the presence or absence of the one or more amplicons, wherein integration of the translocated DNA segment in the genomic DNA of the cell is detected by determining the presence of the one or more amplicons; and wherein no

integration of the translocated DNA segment is detected by determining the absence of the one or more amplicons.

[0084] In another aspect, the disclosure relates to a method for detecting genetic editing of a DNA segment of genomic DNA of a cell, the method including: providing, within a droplet, the genomic DNA of the cell and reagents, the genomic DNA including an integration site where the DNA segment is integrated into the genomic DNA by the genetic editing, wherein the reagents include a DNA segment-specific primer and a second primer; within the droplet, generating one or more amplicons including the integrated DNA segment, if present, using at least the hybridized DNA segment-specific and second primers; and determining the presence or absence of the one or more amplicons, wherein integration of the DNA segment in the genomic DNA of the cell is detected by determining the presence of the one or more amplicons; and wherein no integration of the DNA segment is detected by determining the absence of the one or more amplicons.

[0085] In some embodiments, genetic editing includes use of a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas9 system, a meganuclease, a zinc finger nuclease (ZFN), a transposase, an integrase, or a recombinase.

BRIEF DESCRIPTION OF THE DRAWINGS

[0086] These and other features, aspects, and advantages of the present invention will become better understood with regard to the following description and accompanying drawings.

[0087] **FIG. 1** is a set of schematics depicting a two-step workflow including a first step (left inset) of encapsulating a cell within an emulsion (*e.g.*, a droplet or a tube) and exposing the cell to reagents, which may include a protease and a detergent, that cause the cell to lyse. Within a droplet (or a tube), the genomic DNA of the cell, which may include an integration site where foreign DNA has been integrated, is then exposed to the reagents. The reagents may include a transposase (*e.g.*, a bead-linked Tn5) and a transposase adaptor (*e.g.*, Tn5 adaptor). For example, a bead-linked transposase can mediate tagmentation of the genomic DNA, including the fragmentation of the genomic DNA and ligation of transposase adaptors to the genomic DNA. The second step (right inset) includes amplifying the tagmented DNA fragments including the foreign DNA. Such amplification may include primer extension with reagents provided, such as one or more viral DNA-specific primer (“vector specific primer”) and a barcode primer including a barcode identification sequence (“CBC”). In various

embodiments, two viral DNA-specific primers (*e.g.*, a first viral DNA-specific primer and a second viral DNA-specific primer), denoted by left and right arrows may be provided. Primer extension of the first viral DNA-specific primer can mediate extension of the primer to produce a nucleic acid molecule including a constant region sequence (“seq8F”) and the first viral-DNA-specific primer, while primer extension of the second viral DNA-specific primer can produce a nucleic acid molecule including the second viral-DNA-specific primer and a read sequence (*e.g.*, a second read sequence). Primer extension of a barcode primer can mediate extending the barcode identification sequence to produce a nucleic acid molecule including a read sequence (*e.g.*, a first read sequence) and the barcode identification sequence (CBC).

[0088] FIG. 2 is a set of schematics depicting the amplification step of the two-flow workflow generally described in FIG. 1, which includes amplifying the tagmented DNA fragments including the foreign DNA. Such amplification may include primer extension with reagents provided, such as one or more viral DNA-specific primer (“vector specific primer”), a barcode primer including a barcode identification sequence (“CBC”), a read 1 sequence primer, and a read 2 sequence primer. In various embodiments, two viral DNA-specific primers (*e.g.*, a first viral DNA-specific primer and a second viral DNA-specific primer), denoted by left and right arrows may be provided. Primer extension of a barcode primer can mediate extending the barcode identification sequence to produce a nucleic acid molecule including the barcode identification sequence and a constant region sequence (“seq8F”). Primer extension of the second viral DNA-specific primer can produce a nucleic acid molecule including the second viral-DNA-specific primer and an index sequence. Primer extension of the read 1 sequence primer can produce a nucleic acid molecule including a first index sequence to which an adaptor may bind (*e.g.*, an Illumina P5 adaptor). Primer extension of the read 2 sequence primer can produce a nucleic acid molecule including a second index sequence to which an adaptor may bind (*e.g.*, an Illumina P7 adaptor).

[0089] FIG. 3 is a graph of amplicon fragment sizes from gel electrophoresis following the two-step workflow described in FIG. 2.

[0090] FIG. 4 is a schematic depicting a two-step workflow including a first step (not shown) of encapsulating a cell within an emulsion (*e.g.*, a droplet or a tube) and exposing the cell to reagents, which may include a protease and a detergent, that cause the cell to lyse. Within a droplet (or a tube), the genomic DNA of the cell, which may include an integration site where foreign DNA has been integrated, is then exposed to the reagents. The second step

(shown inset) includes amplifying the genomic DNA including the foreign DNA. Such amplification may include primer extension with reagents provided, such as an Alu primer, a barcode primer including a barcode identification sequence (“cell barcode”), and one or more viral DNA-specific primer (“vector specific primer”). Primer extension of an Alu primer can mediate extension of the primer to produce a nucleic acid molecule including a constant region sequence (“const”). Primer extension of the barcode primer can mediate extension of the primer to produce a nucleic acid molecule including the barcode identification sequence and a constant region sequence. Primer extension of a first viral DNA-specific primer can mediate extension of the primer to produce a nucleic acid molecule including the viral-DNA-specific primer and an index sequence.

[0091] FIG. 5 is a set of schematics further depicting the amplification step of the two-flow workflow described in FIG. 4. As shown in the top inset, the amplification step includes amplifying the genomic DNA including the foreign DNA. Such amplification may include primer extension with reagents provided, such as an Alu primer, a barcode primer including a barcode identification sequence (“cell barcode”), and one or more viral DNA-specific primer (“vector specific primer”). Primer extension of an Alu primer can mediate extension of the primer to produce a nucleic acid molecule including a constant region sequence (“const”). Primer extension of the barcode primer can mediate extension of the primer to produce a nucleic acid molecule including the barcode identification sequence and a constant region sequence. Primer extension of a first viral DNA-specific primer can mediate extension of the primer to produce a nucleic acid molecule including the viral-DNA-specific primer and an index sequence. As shown in the right inset, additional reagents include a read 1 sequence primer and one or more adaptors. Primer extension of the read 1 sequence primer can produce a nucleic acid molecule including a first index sequence to which an adaptor may bind (*e.g.*, an Illumina P5 adaptor). Primer extension of the viral DNA-specific primer can produce a nucleic acid molecule including the second viral-DNA-specific primer and an index sequence to which an adaptor may bind (*e.g.*, an Illumina P7 adaptor).

[0092] FIGS. 6A-6D are graphs of amplicon fragment sizes using different primers as determined by gel electrophoresis following the two-step workflow described in FIG. 5.

[0093] FIG. 7 is a graph of the mapped sequence reads from an experiment which combines the detection of viral integration using repeat sequence-specific primers, as described in FIG. 4, with the detection of a target DNA, as described in FIG. 10. NST cells were transduced with a viral vector, which integrates at a known integration site, and the

nucleic acids of the lysate, which entail gDNA of the cell having an integrated foreign DNA segment, were probed with a viral DNA-specific primer to a long terminal repeat (LTR) as well as a repeat sequence-specific primer.

[0094] FIG. 8 is a graph showing the sequence mapping of single-cell lysates probed with primers for the detection of viral integration, as described in FIG. 7. Left-aligned reads on the leftmost side and middle of the graph indicate two 5' LTR priming sites, while the alignment of the reads on the rightmost side of the graph display 3' LTR priming site.

[0095] FIG. 9 is a schematic depicting an exemplary nucleic acid molecule produced by primer extension, as described in FIG. 1. From top to bottom, the amplification entails primer extension of the first viral DNA-specific primer to produce a nucleic acid molecule including a constant region sequence ("Constant Region") and the first viral-DNA-specific primer ("GSP-FWD"), while primer extension of the second viral DNA-specific primer produces a nucleic acid molecule including the second viral-DNA-specific primer ("GSP-REV") and a read sequence ("Read 2"). Primer extension of a barcode primer (*e.g.*, a bead barcode primer) produces a nucleic acid molecule including a read sequence ("Read 1"), the barcode identification sequence ("Bead Barcode"), and the constant region sequence. Primer extension of a read 1 sequence primer produces nucleic acid molecule including a read sequence (*e.g.*, a first read sequence) and an index sequence (*e.g.*, a first index sequence), in which the index sequence is used to amplify the amplicons containing the cell barcodes into libraries. When an adaptor is provided (*e.g.*, a P5 adaptor), the adaptor will bind to the first read sequence ("P5 + Index 1"). Primer extension of a read 2 sequence primer produces nucleic acid molecule including a read sequence (*e.g.*, a second read sequence) and an index sequence (*e.g.*, a second index sequence). When an adaptor is provided (*e.g.*, a P7 adaptor), the adaptor will bind to the first read sequence ("Index 2 + P7"). Following the multi-step amplification process, the nucleic acid molecule includes from 5'-to-3': an adaptor ("P5"), the first index sequence ("Index 1"), the first read sequence ("Read 1"), the barcode identification sequence ("Bead Barcode"), the constant region sequence ("Constant Region"), the first viral DNA-specific primer ("GSP-FWD"), the complement sequence of the foreign DNA ("Region of Interest"), the second viral DNA-specific primer ("GSP-REV"), the second read sequence ("Read 2"), and the second index sequence and an adaptor ("Index 2 + P7").

[0096] FIG. 10 is a set of schematics, depicting a two-step workflow, as described in FIG. 1, including a first step of encapsulating a cell within an emulsion (*e.g.*, a droplet or a tube) and exposing the cell to reagents, which may include a protease and a detergent, that

cause the cell to lyse. In addition to the reagents and steps described in FIG. 1, this figure further depicts (right side) reagents including two additional primers (“GSP rev” and “GSP fwd”) which can bind to a target DNA, such as a putative single nucleotide variant (SNV) or a copy number variation (CNV) present in the genomic DNA (gDNA), thereby enabling the detection of one or more mutations of the cell or the population of cells in a targeted DNA library. Abbreviations: Out VSP, viral DNA-specific primer; CBC, barcode primer including a barcode identification sequence; seq8F, constant region sequence.

[0097] FIG. 11 is a schematic of the mapped sequence reads from an experiment which combines the detection of viral integration with the detection of a target DNA, as described in FIG. 10. NST cells were transduced with a viral vector, which integrates at a known integration site, and the nucleic acids of the lysate, which entail gDNA of the cell having an integrated foreign DNA segment, were probed with a viral DNA-specific primer to a long terminal repeat (LTR).

[0098] FIG. 12 is a schematic of the mapped sequence reads from an experiment which combines the detection of viral integration with the detection of a target DNA, as described in FIG. 11. Nucleic acids of the lysate probed with three viral DNA-specific primers, including primers to a first 3' LTR (“Primers 1+5;” top), a 5' LTR (“Primers 4+6;” middle), and a second 3' LTR (“3LTR_2 + 3LTR_1;” bottom).

[0099] FIGS. 13A-13C are a set of graphs showing the relative panel uniformity and percentage (%) of DNA completeness (FIG. 13A), genotypic mapping (FIG. 13B), and reads of Tn5 integration (FIG. 13C), respectively, of the same experiment described in FIGs. 9-11 which combines the detection of viral integration with the detection of a target DNA.

[00100] FIG. 14 is a graph showing detection of a viral integration site in transduced Jurkat cells, as compared to control, non-transduced Raji cells in an experiment which combines the detection of viral integration with the detection of a target DNA as described in FIGs. 9-11. The x-axis shows the number of reads for a target DNA, while the y-axis shows the number of reads of a particular integration site.

[00101] FIG. 15 is a graph showing the sequence mapping of single-cell lysates probed with primers for the detection of viral integration. Non-aligned reads on the leftmost side of the graph indicate unique Tn5 insertion sites, while the alignment of the reads on the rightmost side of the graph display a viral DNA-specific primer and read sequence, which was consistent across cells due to the identical site of integration of the vector in the cells.

[00102] FIG. 16 is a schematic depicting how a method described herein may be used to estimate the vector copy number of viral DNA in a single cell using counts of the unique Tn5 insertion sites, which are random. As described in FIG. 1, a sequence may be tagged randomly and a transposase adapter may be inserted at the respective site. As depicted in the schematic herein, in a single cell, two unique Tn5 insertion sites were inserted into the gDNA (depicted by the two circular sector symbols at “Position 1” and “Position 3,” respectively). Following the amplification and sequencing steps described in FIG. 1, the sequence map would show two unique sequence reads, both having genome:vector junctions. This count of two genome:vector junctions, within a single cell, indicates that two vector copies exist in the single cell.

[00103] FIGs. 17A-17B are a set of schematics depicting another method, alternative to the method described in FIG. 16, which may be used to estimate the vector copy number of viral DNA in a single cell. As described in FIG. 16, in a single cell, Tn5 may integrate randomly into two unique locations, such as two positions in the foreign DNA sequence (depicted by the two circular sector symbols at “Position 2” and “Position 4,” respectively). Following the amplification and sequencing steps described in FIG. 1, the sequence map would contain two amplicons with an overlapping sequence of a portion of the vector sequence (depicted by vertical dashed lines). This overlapping read of the vector sequence, within a single cell, indicates that two vector copies exist in the single cell (FIG. 17A). When a non-overlapping read is detected, it does not provide information of another vector copy number (FIG. 17B), and it is discarded from vector copy number analyses.

[00104] FIG. 18 provides a schematic depicting how the methods of the disclosure may be used to estimate the vector copy number of viral DNA in a single cell. Exemplary amplicons from the schematics described in FIGs. 16, 17A, and 17B are outlined in bold rectangles (top) and are overlaid upon an exemplary sequence map (bottom). Overlapping amplicons indicate the vector copy numbers in a single cell, as described in FIG. 17A.

DETAILED DESCRIPTION

Definitions

[00105] Definitions of common terms in cell biology and molecular biology can be found in “The Merck Manual of Diagnosis and Therapy,” 19th Edition, published by Merck Research Laboratories, 2006 (ISBN 0-911910-19-0); Robert S. Porter et al. (eds.), The Encyclopedia of Molecular Biology, published by Blackwell Science Ltd., 1994 (ISBN 0-

632-02182-9). Definitions of common terms in molecular biology can also be found in Benjamin Lewin, *Genes X*, published by Jones & Bartlett Publishing, 2009 (ISBN-10: 0763766321); Kendrew et al. (eds.), *Molecular Biology and Biotechnology: a Comprehensive Desk Reference*, published by VCH Publishers, inc., 1995 (ISBN 1-56081-569-8) and *Current Protocols in Protein Sciences* 2009, Wiley Intersciences, Coligan et al., eds.

[00106] The features and other details of the disclosure will now be more particularly described. Certain terms employed in the specification, examples and appended claims are collected here. These definitions should be read in light of the remainder of the disclosure and understood as by a person of skill in the art. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by a person of ordinary skill in the art. Any terms not directly defined herein shall be understood to have the meanings commonly associated with them as understood within the art of the invention. Certain terms are discussed herein to provide additional guidance to the practitioner in describing the compositions, devices, methods and the like of aspects of the invention, and how to make or use them. It will be appreciated that the same thing may be said in more than one way. Consequently, alternative language and synonyms may be used for any one or more of the terms discussed herein. No significance is to be placed upon whether or not a term is elaborated or discussed herein. Some synonyms or substitutable methods, materials and the like are provided. Recital of one or a few synonyms or equivalents does not exclude use of other synonyms or equivalents, unless it is explicitly stated. Use of examples, including examples of terms, is for illustrative purposes only and does not limit the scope and meaning of the aspects of the invention herein.

[00107] The singular terms “a,” “an,” and “the” include plural referents unless context clearly indicates otherwise. Similarly, the word “or” is intended to include “and” unless the context clearly indicates otherwise. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of this disclosure, suitable methods and materials are described below. The abbreviation, “*e.g.*,” is derived from the Latin *exempli gratia*, and is used herein to indicate a non-limiting example. Thus, the abbreviation “*e.g.*,” is synonymous with the term “for example.”

[00108] As used herein, the term “adapter” or “adaptor” is a single-stranded or a double-stranded nucleic acid molecule that can be linked to the end of other nucleic acids. In various embodiments, an adapter is a short, chemically synthesized, double-stranded nucleic acid

molecule which can be used to link the ends of two other nucleic acid molecules. In various embodiments, an adaptor is a double-stranded nucleic acid (*e.g.*, oligonucleotide) that includes single-stranded nucleotide overhangs at the 5' and/or 3' ends. In various embodiments, the single-stranded overhangs are 1, 2, or more nucleotides. For example, 5 adaptors used in tagmentation may be referred to herein as Tn5 adapters. In various embodiments, adaptors include additional nucleic acid sequence for cloning or analysis of the integration of foreign DNA.

[00109] The terms “amplify,” “amplifying,” “amplification reaction,” and variants thereof, refer generally to any action or process whereby at least a portion of a nucleic acid molecule (referred to as a template nucleic acid molecule) is replicated or copied 10 into at least one additional nucleic acid molecule. The additional nucleic acid molecule optionally includes a sequence that is substantially identical or substantially complementary to at least some portion of the template nucleic acid molecule. The template nucleic acid molecule can be single-stranded or double-stranded and the 15 additional nucleic acid molecule can independently be single-stranded or double-stranded. In various embodiments, amplification includes a template-dependent *in vitro* enzyme-catalyzed reaction for the production of at least one copy of at least some portion of the nucleic acid molecule or the production of at least one copy of a nucleic acid sequence that is complementary to at least some portion of the nucleic acid molecule. 20 Amplification optionally includes linear or exponential replication of a nucleic acid molecule. In various embodiments, such amplification is performed using isothermal conditions; in other embodiments, such amplification can include thermocycling. In various embodiments, the amplification is a multiplex amplification that includes the simultaneous amplification of a plurality of target sequences in a single amplification 25 reaction. At least some of the target sequences can be situated, on the same nucleic acid molecule or on different target nucleic acid molecules included in the single amplification reaction. In various embodiments, “amplification” includes amplification of at least some portion of DNA-based nucleic acids. The amplification reaction can include single- or double-stranded nucleic acid substrates and can further include any of the amplification 30 processes known to one of ordinary skill in the art. In various embodiments, the amplification reaction includes polymerase chain reaction (PCR). In various embodiments, the amplification reaction includes an isothermal amplification reaction such as Loop-mediated isothermal amplification (LAMP). In the present invention, the

terms “synthesis” and “amplification” of nucleic acid are used. The synthesis of nucleic acid in the present invention means the elongation or extension of nucleic acid from an oligonucleotide serving as the origin of synthesis. If not only this synthesis but also the formation of other nucleic acid and the elongation or extension reaction of this formed nucleic acid occur continuously, a series of these reactions is comprehensively called amplification. The polynucleic acid produced by the amplification technology employed is generically referred to as an “amplicon” or “amplification product.”

[00110] Any nucleic acid amplification method may be utilized, such as a PCR-based assay, *e.g.*, quantitative PCR (qPCR), or an isothermal amplification may be used to detect the presence of certain nucleic acids, *e.g.*, genes of interest, present in discrete entities or one or more components thereof, *e.g.*, cells encapsulated therein. Such assays can be applied to discrete entities within a microfluidic device or a portion thereof or any other suitable location. The conditions of such amplification or PCR-based assays may include detecting nucleic acid amplification over time and may vary in one or more ways.

[00111] PCR relies upon polymerase extension of annealed primers at each cycle. As used herein, the term “polymerase extension” means the template-dependent incorporation of at least one complementary nucleotide, by a nucleic acid polymerase, onto the 3' end of an annealed primer. Polymerase extension preferably adds more than one nucleotide, preferably up to and including nucleotides corresponding to the full length of the template. Conditions for polymerase extension vary with the identity of the polymerase. The temperature used for polymerase extension is generally based upon the known activity properties of the enzyme. Although, where annealing temperatures are to be, for example, below the optimal temperatures for the enzyme, it will often be acceptable to use a lower extension temperature. In general, although the enzymes retain at least partial activity below their optimal extension temperatures, polymerase extension by the most commonly used thermostable polymerases (*e.g.*, Taq polymerase and variants thereof) is performed at 65 °C to 75 °C, preferably about 68 °C to 72 °C.

[00112] A number of nucleic acid polymerases can be used in the amplification reactions utilized in certain embodiments provided herein, including any enzyme that can catalyze the polymerization of nucleotides (including analogs thereof) into a nucleic acid strand. Such nucleotide polymerization can occur in a template-dependent fashion. Such polymerases can include, without limitation, naturally-occurring polymerases and any subunits and truncations thereof, mutant polymerases, variant polymerases, recombinant,

fusion or otherwise engineered polymerases, chemically modified polymerases, synthetic molecules or assemblies, and any analogs, derivatives or fragments thereof that retain the ability to catalyze such polymerization. Optionally, the polymerase can be a mutant polymerase including one or more mutations involving the replacement of one or more amino acids with other amino acids, the insertion or deletion of one or more amino acids from the polymerase, or the linkage of parts of two or more polymerase. The polymerase can include one or more active sites at which nucleotide binding and/or catalysis of nucleotide polymerization can occur. Some exemplary polymerases include, without limitation, DNA polymerases and RNA polymerases. The term “polymerase” and variants thereof, as used herein, also includes fusion proteins including at least two portions linked to each other, where the first portion includes a peptide that can catalyze the polymerization of nucleotides into a nucleic acid strand and is linked to a second portion that includes a second polypeptide. In various embodiments, the second polypeptide can include a reporter enzyme or a processivity-enhancing domain. Optionally, the polymerase can possess 5’ exonuclease activity or terminal transferase activity. In various embodiments, the polymerase can be optionally reactivated, for example through the use of heat, chemicals or re-addition of new amounts of polymerase into a reaction mixture. In various embodiments, the polymerase can include a hot-start polymerase or an aptamer-based polymerase that optionally can be reactivated.

[00113] The term “analyte” refers to a component of a cell. Cell analytes can be informative for understanding a state, behavior, or trajectory of a cell. Therefore, performing single-cell analysis of one or more analytes of a cell using the systems and methods described herein are informative for determining a state or behavior of a cell. Examples of an analyte include a nucleic acid (*e.g.*, RNA, DNA, and cDNA), a protein, a peptide, an antibody, an antibody fragment, a polysaccharide, a sugar, a lipid, a small molecule, or combinations thereof. In particular embodiments, a bulk DNA or single-cell analysis involves analyzing two different analytes such as protein and DNA. In particular embodiments, a bulk DNA or single-cell analysis involves analyzing three or more different analytes of a cell, such as RNA, DNA, and protein. In particular embodiments, an analyte refers to genomic DNA of a cell. Here, the genomic DNA of the cell may or may not include an integration site at which foreign DNA is integrated.

[00114] The term “antibody” encompasses monoclonal antibodies (including full length monoclonal antibodies), polyclonal antibodies, multispecific antibodies (*e.g.*, bispecific

antibodies), and antibody fragments that are antigen-binding, *e.g.*, an antibody or an antigen-binding fragment thereof. “Antibody fragment,” and all grammatical variants thereof, as used herein are defined as a portion of an intact antibody including the antigen binding site or variable region of the intact antibody, wherein the portion is free of the constant heavy chain domains (i.e., CH2, CH3, and CH4, depending on antibody isotype) of the Fc region of the intact antibody. Examples of antibody fragments include Fab, Fab’, Fab’-SH, F(ab’)₂, and Fv fragments; diabodies; any antibody fragment that is a polypeptide having a primary structure consisting of one uninterrupted sequence of contiguous amino acid residues (referred to herein as a “single-chain antibody fragment” or “single chain polypeptide”).

[00115] A “barcode” nucleic acid identification sequence can be incorporated into a nucleic acid primer or linked to a primer to enable independent sequencing and identification to be associated with one another via a barcode which relates information and identification that originated from molecules that existed within the same sample. There are numerous techniques that can be used to attach barcodes to the nucleic acids within a discrete entity. For example, the target nucleic acids may or may not be first amplified and fragmented into shorter pieces. The molecules can be combined with discrete entities, *e.g.*, droplets, containing the barcodes. The barcodes can then be attached to the molecules using, for example, splicing by overlap extension. In this approach, the initial target molecules can have “adaptor” sequences added, which are molecules of a known sequence to which primers can be synthesized. When combined with the barcodes, primers can be used that are complementary to the adaptor sequences and the barcode sequences, such that the product amplicons of both target nucleic acids and barcodes can anneal to one another and, via an extension reaction such as DNA polymerization, be extended onto one another, generating a double-stranded product including the target nucleic acids attached to the barcode sequence. Alternatively, the primers that amplify that target can themselves be barcoded so that, upon annealing and extending onto the target, the amplicon produced has the barcode sequence incorporated into it. This can be applied with a number of amplification strategies, including specific amplification with PCR or non-specific amplification with, for example, multiple displacement amplification (MDA). An alternative enzymatic reaction that can be used to attach barcodes to nucleic acids is ligation, including blunt or sticky end ligation. In this approach, the DNA barcodes are incubated with the nucleic acid targets and ligase enzyme, resulting in the ligation of the barcode to the targets. The ends of the nucleic

acids can be modified, as needed, for ligation by a number of techniques, including by using adaptors introduced with ligase or fragments to enable greater control over the number of barcodes added to the end of the molecule. In various embodiments, the barcode primer is a bead barcode primer.

5 **[00116]** The terms “cell” and “host cell” are used interchangeably and refer to one or more cells into which foreign DNA has been introduced, including the progeny of such cells.

10 **[00117]** The phrase “cell genotype” refers to the genetic makeup of the cell and can refer to one or more genes and/or the combination of alleles (*e.g.*, homozygous or heterozygous) of a cell. The phrase cell genotype further encompasses one or more mutations of the cell including polymorphisms, single nucleotide polymorphisms (SNPs), single nucleotide variants (SNVs), insertions, deletions, knock-ins, knock-outs, copy number variations (CNVs), duplications, translocations, and loss of heterozygosity (LOH). In various embodiments, a cell phenotype is determined using bulk DNA or single-cell analysis. In various embodiments, the cell phenotype can refer to the expression of a panel of genes.

15 **[00118]** The phrase “cell phenotype” refers to the cell expression of one or more proteins (*e.g.*, cellular proteomics). In various embodiments, a cell phenotype is determined using bulk DNA or single-cell analysis. In various embodiments, the cell phenotype can refer to the expression of a panel of proteins.

20 **[00119]** “Complementarity” refers to the ability of a nucleic acid to form hydrogen bond(s) or hybridize with another nucleic acid sequence by either traditional Watson-Crick or other non-traditional types. As used herein “hybridization,” refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under low, medium, or highly stringent conditions, including when that sequence is present in a complex mixture (*e.g.*, total cellular) DNA or RNA. See *e.g.* Ausubel, et al., Current
25 Protocols In Molecular Biology, John Wiley & Sons, New York, N.Y., 1993. If a nucleotide at a certain position of a polynucleotide is capable of forming a Watson-Crick pairing with a nucleotide at the same position in an anti-parallel DNA or RNA strand, then the polynucleotide and the DNA or RNA molecule are complementary to each other at that position. The polynucleotide and the DNA or RNA molecule are “substantially
30 complementary” to each other when a sufficient number of corresponding positions in each molecule are occupied by nucleotides that can hybridize or anneal with each other in order to affect the desired process. A complementary sequence is a sequence capable of

annealing under stringent conditions to provide a 3'-terminal serving as the origin of synthesis of complementary chain.

[00120] Throughout the specification and claims, the word "comprise," or variations such as "comprises" or "comprising," will be understood to imply the inclusion of a stated word or group of words but not the exclusion of any other word or group of words.

[00121] As used herein, the term "determining," as used in the context of "determining the presence or absence of" an amplicon refers to determining the presence or lack thereof of the amplicon. In various embodiments, determining the presence or absence of an amplicon occurs when the amplicon or fragment thereof has been fully or partially separated from other components of a sample or composition, and also can include determining the charge-to-mass ratio, the mass, the amount, the absorbance, the fluorescence, or other property of the nucleic acid molecule or fragment thereof. In particular embodiments, determining the presence or absence of an amplicon occurs through sequencing methods (e.g., by sequencing a sequence of the amplicon).

[00122] In various embodiments, the discrete entities as described herein are droplets. The terms "emulsion," "drop," "droplet," and "microdroplet" are used interchangeably herein, to refer to small, generally spherically structures, containing at least a first fluid phase, e.g., an aqueous phase (e.g., water), bounded by a second fluid phase (e.g., oil) which is immiscible with the first fluid phase. In various embodiments, droplets according to the present disclosure may contain a first fluid phase, e.g., oil, bounded by a second immiscible fluid phase, e.g. an aqueous phase fluid (e.g., water). In various embodiments, the second fluid phase will be an immiscible phase carrier fluid. Thus droplets according to the present disclosure may be provided as aqueous-in-oil emulsions or oil-in-aqueous emulsions.

Droplets may be sized and/or shaped as described herein for discrete entities. For example, droplets according to the present disclosure generally range from 1 μm to 1000 μm , inclusive, in diameter. Droplets according to the present disclosure may be used to encapsulate cells, nucleic acids (e.g., DNA), enzymes, reagents, reaction mixture, and a variety of other components. The term emulsion may be used to refer to an emulsion produced in, on, or by a microfluidic device and/or flowed from or applied by a microfluidic device.

[00123] The term "foreign DNA segment-specific primer," also referred to herein as a "vector-specific primer" refers to a primer that is complementary to a sequence of foreign DNA. In various embodiments, foreign DNA segment-specific primers are single-stranded or double-stranded polynucleotides, such as an oligonucleotide, that include at least one

sequence that is at least partially complementary to a target nucleic acid sequence (e.g., a segment of foreign DNA). An exemplary foreign DNA segment-specific primer includes a primer targeted to viral DNA (e.g., a viral DNA-specific primer). “A sequence of the foreign DNA” refers to one or more regions of the foreign DNA e.g., to which the foreign DNA segment-specific primer (e.g., a foreign DNA segment-specific primer and/or a second foreign DNA segment-specific primer) bind. The primers act to delimit the region of the original foreign polynucleotide which is exponentially amplified during amplification.

[00124] “Identity,” as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, “identity” also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as determined by the match between strings of such sequences. “Identity” and “similarity” can be readily calculated by known methods, including, but not limited to, those described in Computational Molecular Biology, Lesk, A. M., ed., Oxford University Press, New York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D. W., ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part I, Griffin, A. M., and Griffin, H. G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., Siam J. Applied Math., 48:1073 (1988). In addition, values for percentage identity can be obtained from amino acid and nucleotide sequence alignments generated using the default settings for the AlignX component of Vector NTI Suite 8.0 (Informax, Frederick, Md.). Preferred methods to determine identity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Example computer program methods to determine identity and similarity between two sequences include, but are not limited to, the GCG program package (Devereux, J., et al., *Nucleic Acids Research* 12(1): 387 (1984)), BLASTP, BLASTN, and FASTA (Atschul, S. F. et al., *J. Molec. Biol.* 215:403-410 (1990)). The BLAST X program is publicly available from NCBI and other sources (BLAST Manual, Altschul, S., et al., NCBINLM NIH Bethesda, Md. 20894; Altschul, S., et al., *J. Mol. Biol.* 215:403-410 (1990)). The well-known Smith Waterman algorithm may also be used to determine identity.

[00125] As used herein, the terms “integrates,” “integration,” and “integration sites” refer generally to instances in which foreign DNA *e.g.*, of a vector such as of a viral vector has translocated into the nucleus of a host genome and integrated into the genomic DNA of the host. This stands in contrast to non-integrating vectors, in which foreign DNA may remain in the cytoplasm of the host in, for example, an episomal form.

[00126] An “ITR” is a palindromic nucleic acid, *e.g.*, an inverted terminal repeat, that is about 120 nucleotides to about 250 nucleotides in length and capable of forming a hairpin. The term “ITR” includes the site of the viral genome replication that can be recognized and bound by a parvoviral protein (*e.g.*, Rep78/68). An ITR may be from any adeno-associated virus (AAV), with serotype 2 being preferred. An ITR includes a replication protein binding element (RBE) and a terminal resolution sequences (TRS). The term “ITR” does not require a wild-type parvoviral ITR (*e.g.*, a wild-type nucleic acid sequence may be altered by insertion, deletion, truncation, or missense mutations), as long as the ITR functions to mediate virus packaging, replication, integration, and/or provirus rescue, and the like.

[00127] An “LTR” is a “long terminal repeat” that is generated as a DNA duplex at both ends of the retrovirus when a retrovirus integrates into a host genome. The 5' LTR includes a U3, R, and U5 nucleic acid element. The 3' LTR also includes U3, R, and U5 nucleic acid element. In a replication competent retrovirus, LTRs also contain an active RNA polymerase II promoter which allows transcription of the integrated provirus by host cell RNA polymerase II to generate new copies of the retroviral RNA genome.

[00128] As used herein, the term “modified” in reference to a nucleic acid or oligonucleotide refers to any variation made to a given nucleic acid or oligonucleotide, such as a oligonucleotide's length, nucleic acid sequence, chemical structure, or post-translational modifications.

[00129] The terms “nucleic acid,” “polynucleotides,” and “oligonucleotides” refers to biopolymers of nucleotides and, unless the context indicates otherwise, includes modified and unmodified nucleotides, and both DNA and RNA, and modified nucleic acid backbones. For example, in certain embodiments, the nucleic acid is a peptide nucleic acid (PNA) or a locked nucleic acid (LNA). In various embodiments, the methods as described herein are performed using DNA as the nucleic acid template for amplification. However, nucleic acid whose nucleotide is replaced by an artificial derivative or modified nucleic acid from natural DNA or RNA is also included in the nucleic acid of the present invention insofar as it functions as a template for synthesis of complementary chain. The

nucleic acid of the present invention is generally contained in a biological sample. The biological sample includes animal, plant or microbial tissues, cells, cultures and excretions, or extracts therefrom. In certain aspects, the biological sample includes intracellular parasitic genomic DNA or RNA such as virus or mycoplasma. The nucleic acid may be derived from nucleic acid contained in said biological sample. For example, genomic DNA, or cDNA synthesized from mRNA, or nucleic acid amplified on the basis of nucleic acid derived from the biological sample, are preferably used in the described methods. Unless denoted otherwise, whenever an oligonucleotide sequence is represented, it will be understood that the nucleotides are in 5' to 3' order from left to right and that "A" denotes deoxyadenosine, "C" denotes deoxycytidine, "G" denotes deoxyguanosine, "T" denotes deoxythymidine, and "U" denotes uridine. Oligonucleotides are said to have "5' ends" and "3' ends" because mononucleotides are, in some cases, reacted to form oligonucleotides via attachment of the 5' phosphate or equivalent group of one nucleotide to the 3' hydroxyl or equivalent group of its neighboring nucleotide, optionally via a phosphodiester or other suitable linkage.

[00130] Primers and oligonucleotides used in embodiments herein include nucleotides. A nucleotide includes any compound, including without limitation any naturally occurring nucleotide or analog thereof, which can bind selectively to, or can be polymerized by a polymerase. Typically, but not necessarily, selective binding of the nucleotide to the polymerase is followed by polymerization of the nucleotide into a nucleic acid strand by the polymerase; occasionally however the nucleotide may dissociate from the polymerase without becoming incorporated into the nucleic acid strand, an event referred to herein as a "non-productive" event. Such nucleotides include not only naturally occurring nucleotides but also any analogs, regardless of their structure, that can bind selectively to, or can be polymerized by, a polymerase. While naturally occurring nucleotides typically include base, sugar and phosphate moieties, the nucleotides of the present disclosure can include compounds lacking anyone, some or all of such moieties. For example, the nucleotide can optionally include a chain of phosphorus atoms including three, four, five, six, seven, eight, nine, ten, or more phosphorus atoms. In various embodiments, the phosphorus chain can be attached to any carbon of a sugar ring, such as the 5' carbon. The phosphorus chain can be linked to the sugar with an intervening O or S. In various embodiments, one or more phosphorus atoms in the chain can be part of a phosphate group having P and O. In another embodiment, the phosphorus atoms in

the chain can be linked together with intervening O, NH, S, methylene, substituted methylene, ethylene, substituted ethylene, CNH₂, C(O), C(CH₂), CH₂CH₂, or C(OH)CH₂R (where R can be a 4-pyridine or 1-imidazole). In various embodiments, the phosphorus atoms in the chain can have side groups having O, BH₃, or S. In the
5 phosphorus chain, a phosphorus atom with a side group other than O can be a substituted phosphate group. In the phosphorus chain, phosphorus atoms with an intervening atom other than O can be a substituted phosphate group. Some examples of nucleotide analogs are described in Xu, U.S. Pat. No. 7,405,281.

[00131] As used herein, “primer” refers to a DNA or RNA polynucleotide molecule or an
10 analog thereof capable of specifically annealing to a polynucleotide template and providing a 3' end that serves as a substrate for a template-dependent polymerase to produce an extension product which is complementary to the polynucleotide template. A primer useful in the methods described herein is generally single-stranded, and a primer and its complement can anneal to form a double-stranded polynucleotide. Primers according to the methods and
15 compositions described herein can be less than or equal to 300 nucleotides in length, *e.g.*, less than or equal to 300, or 250, or 200, or 150, or 100, or 90, or 80, or 70, or 60, or 50, or 40, or 30 or fewer, or 20 or fewer, or 15 or fewer, but at least 10 nucleotides in length. Methods of making primers are well known in the art, and numerous commercial sources offer oligonucleotide synthesis services suitable for providing primers according to the methods
20 and compositions described herein, *e.g.* INVITROGEN™ Custom DNA Oligos; Life Technologies; Grand Island, N.Y. or custom DNA Oligos from IDT; Coralville, Iowa.

[00132] “Percent (%) nucleic acid sequence identity” with respect to a reference polynucleotide sequence is defined as the percentage of nucleic acids in a candidate sequence that are identical with the nucleic acids in the reference polynucleotide sequence, after
25 aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any conservative substitutions as part of the sequence identity. Alignment for purposes of determining percent nucleic acid sequence identity can be achieved in various ways that are within the skill in the art, for instance, using publicly available computer software such as BLAST, BLAST-2, or Megalign (DNASTAR) software.

[00133] Those skilled in the art can determine appropriate parameters for aligning
30 sequences, including any algorithms needed to achieve maximal alignment over the full length of the sequences being compared. For purposes herein, however, % nucleic acid sequence identity values are generated using the sequence comparison computer program

BLAST. The % nucleic acid sequence identity of a given nucleic acid sequence A to, with, or against a given nucleic acid sequence B (which can alternatively be phrased as a given nucleic acid sequence A that has or includes a certain % nucleic acid sequence identity to, with, or against a given nucleic acid sequence B) is calculated as follows:

$$100 \text{ multiplied by (the fraction } X/Y \text{)}$$

where X is the number of nucleotides scored as identical matches by the sequence alignment program BLAST in that program's alignment of A and B, and where Y is the total number of nucleic acids in B. It will be appreciated that where the length of nucleic acid sequence A is not equal to the length of nucleic acid sequence B, the % nucleic acid sequence identity of A to B will not equal the % nucleic acid sequence identity of B to A.

[00134] As used herein, a "population" refers to a group of at least two (*e.g.*, at least 2, 3, 4, 5, 10, or 15 or more) cells.

[00135] As used herein, the term "reagents" refers to a mixture of components for carrying out a given process, such as the amplification of genomic DNA that includes the integration of foreign DNA. Such reagents may include components including, but not limited to, proteases, cell buffer (*e.g.*, including a detergent, a density-match agent, and a phosphate buffer), and a lysis buffer (*e.g.*, including a reverse primer, a protease, a detergent, an RNA reverse transcriptase, an RNAase inhibitor, a transposase, and a magnesium buffer).

[00136] The term "repeat sequence-specific primer" refers to a primer that is complementary to a repeat sequence (*e.g.*, an Alu repeat element) of DNA. For example, in various embodiments, the repeat sequence-specific primer is an Alu primer. Repeat sequence-specific primers are generally single-stranded or double-stranded polynucleotides, such as an oligonucleotide, that include at least one sequence that is at least partially complementary to a target nucleic acid sequence. The primer acts to delimit the region of the original polynucleotide which is exponentially amplified during amplification. In various embodiments, the repeat sequence-specific primer is an Alu1, an Alu2, a LINE1 a 16S, or an 18S primer.

[00137] As used herein, "sequencing" refers to the determination of the order of nucleotides in a nucleic acid molecule (*e.g.*, an amplicon). Traditional sequencing methods generate sequence information randomly (*e.g.* "shotgun" sequencing) or between two known sequences which are used to design primers. In contrast, the methods described herein, in various embodiments, allow for determining the nucleotide sequence (*e.g.* sequencing) upstream or downstream of a single region of known sequence with a high level of specificity

and sensitivity. Examples of sequencing include, but are not limited to, “next generation sequencing,” which refers to high-throughput sequencing methods that allow millions to billions of molecules to be sequenced in parallel. Examples of next-generation sequencing methods include, but are not limited to, sequencing by synthesis, sequencing by ligation, sequencing by hybridization, polony sequencing, ion semiconductor sequencing, and pyrosequencing. By attaching the primer to the solid substrate and the complementary sequence to the nucleic acid molecule, the nucleic acid molecule can be hybridized to the solid substrate via the primer and then multiple copies are generated in distinct regions on the solid substrate by using a polymerase. Consequently, during the sequencing process, nucleotides at a particular location may be sequenced multiple times (*e.g.*, hundreds or thousands of times) - this depth of coverage is referred to as “deep sequencing”. Examples of high-throughput nucleic acid sequencing techniques include parallel bead arrays, sequencing by synthesis, sequencing by ligation, capillary electrophoresis, “biochips,” microarrays, parallel microchips, single-molecule sequencing, as well as sequencing by platforms provided by Illumina, BGI, Qiagen, Thermo-Fisher, and Roche, including modalities such as molecular arrays (see *e.g.*, *Science* 311:1544-1546, 2006).

[00138] In various embodiments, sequencing refers to a next-generation sequencing method, wherein reads from a single molecule sequencing device are used for sequencing a single molecule of DNA. Unlike next-generation sequencing methods that rely on amplification to clone many DNA molecules in parallel for sequencing in a phased approach, single-molecule sequencing interrogates single molecules of DNA and thus amplifies them. Single molecule sequencing provides methods that include stopping the sequencing reaction after each base incorporation (‘wash-and-scan’ cycles) and methods that do not require interruptions between read steps. Examples of single molecule sequencing methods include single molecule real-time sequencing (Pacific Biosciences), nanopore-based sequencing (Oxford Nanopore), duplex blocked nanopore sequencing, and direct imaging of DNA using a developed microscope.

[00139] As used herein, the terms “tagmentation,” “tagment,” or “tagmenting” refer to transforming a nucleic acid, *e.g.*, a DNA, into adaptor-modified templates in solution ready for cluster formation and sequencing by the use of transposase-mediated fragmentation and tagging. This process often involves the modification of the nucleic acid by a transposome complex including transposase enzyme complexed with adaptors including transposon end sequence. Tagmentation results in the simultaneous fragmentation of the nucleic acid and

ligation of the adaptors to the 5' ends of both strands of duplex fragments. Following a purification step to remove the transposase enzyme, additional sequences are added to the ends of the adapted fragments by PCR. As used herein, the term "transposome complex" refers to a transposase enzyme non-covalently bound to a double-stranded nucleic acid. For example, the complex can be a transposase enzyme preincubated with double-stranded transposon DNA under conditions that support non-covalent complex formation. Double-stranded transposon DNA can include, without limitation, Tn5 DNA, a portion of Tn5 DNA, a transposon end composition, a mixture of transposon end compositions or other double-stranded DNAs capable of interacting with a transposase such as the hyperactive Tn5 transposase.

[00140] As used herein, the terms "transduction" and "transduce" refer to a method of introducing a vector construct or a part thereof into a cell and subsequent expression, such as expression of a transgene encoded by the vector construct in the cell.

[00141] As used herein, the term "transgene" refers to a recombinant nucleic acid (*e.g.*, DNA or cDNA) encoding a gene product. The gene product may be an RNA. In addition to the coding region for the gene product, the transgene may include or be operably linked to one or more elements to facilitate or enhance expression, such as a promoter, enhancer(s), destabilizing domains(s), response element(s), reporter element(s), insulator element(s), polyadenylation signal(s), and other functional elements.

[00142] A "transposase" means an enzyme that is capable of forming a functional complex with a transposon end-containing composition (*e.g.*, transposons, transposon ends, transposon end compositions) and catalyzing insertion or transposition of the transposon end-containing composition into the double-stranded target nucleic acid with which it is incubated, for example, in an *in vitro* transposition reaction. A transposase as presented herein can also include integrases from retrotransposons and retroviruses. Transposases, transposomes, and transposome complexes are generally known to those of skill in the art, as exemplified by the disclosure of US 2010/0120098, the content of which is incorporated herein by reference in its entirety. Although many embodiments described herein refer to Tn5 transposase and/or hyperactive Tn5 transposase, it will be appreciated that any transposition system that is capable of inserting a transposon end with sufficient efficiency to 5'-tag and fragment a target nucleic acid for its intended purpose can be used in the present invention. In particular embodiments, a preferred transposition system is capable of inserting the transposon end in a random or in an almost random manner to 5'-tag and fragment the target nucleic acid.

[00143] As used herein, the term “vector” includes a nucleic acid vector, *e.g.*, a DNA vector, such as a plasmid, an RNA vector, or another suitable replicon (*e.g.*, viral vector). A variety of vectors have been developed for the delivery of polynucleotides encoding exogenous (*e.g.*, foreign) polynucleotides or proteins into a prokaryotic or eukaryotic cell.

5 Examples of such expression vectors are disclosed in, *e.g.*, WO 1994/011026; incorporated herein by reference as it pertains to vectors suitable for the expression of a nucleic acid molecule of interest. Expression vectors suitable for use with the compositions and methods described herein contain a polynucleotide sequence as well as, *e.g.*, additional sequence elements used for the expression of heterologous nucleic acid materials (*e.g.*, a nucleic acid
10 molecule) in a mammalian cell. Certain vectors that can be used for the expression of the nucleic acid molecules described herein include plasmids that contain regulatory sequences, such as promoter and enhancer regions, which direct gene transcription. Other useful vectors for expression of nucleic acid molecule agents disclosed herein contain polynucleotide sequences that enhance the rate of translation of these polynucleotides or improve the
15 stability or nuclear export of the RNA that results from gene transcription. These sequence elements include, *e.g.*, 5' and 3' untranslated regions, an IRES, and polyA in order to direct efficient transcription of the gene carried on the expression vector. The expression vectors suitable for use with the compositions and methods described herein may also contain a polynucleotide encoding a marker for selection of cells that contain such a vector. Examples
20 of a suitable marker are genes that encode resistance to antibiotics, such as ampicillin, chloramphenicol, kanamycin, nourseothricin, or zeocin.

[00144] As use herein, the phrase “method of the disclosure” refers to those methods which are disclosed herein, generically.

25 **Overview of Methods of the Disclosure**

[00145] Provided herein are embodiments for detecting integration of a vector including a foreign DNA segment into genomic DNA of a cell using bulk DNA or single-cell analysis and DNA-sequencing (DNA-seq). Examples of foreign DNA segments include viral DNA, modified viral DNA, or DNA from a viral vector.

30 [00146] Generally, the single-cell analysis involves a workflow for processing single cells and performing sequencing to obtain sequencing reads of analytes of the single cells. Single-cell analysis may also be performed upon a population of cells (*e.g.*, a population of cells having integration of a vector including a foreign DNA segment into genomic DNA of the

cells) or for a plurality of cells to determine cellular genotypes and phenotypes of individual cells. In various embodiments, the single-cell analysis involves performing targeted DNA-seq to generate sequence reads derived from genomic DNA that are used to determine the cell genotype (*e.g.*, cell mutations such as CNVs and/or SNVs). In various embodiments, the single-cell analysis involves performing sequencing of oligonucleotides that are linked to antibodies, where an antibody exhibits binding affinity for a specific analyte expressed by a cell. Thus, sequence reads derived from the antibody-conjugated oligonucleotides are used to determine the cell phenotype (*e.g.*, expression or presence of one or more analytes of the cell).

[00147] In various embodiments, the single-cell analysis involves performing both targeted DNA-seq analysis and protein expression analysis. The combination of cellular genotypes and phenotypes across cells in a population (*e.g.*, a population of heterogeneous cancer cells) is useful for discerning subpopulations of cells, a subpopulation being characterized by a combination of a genotype and a phenotype. Subpopulations of cells may represent a subpopulation that was previously unknown, or a subpopulation that is unlikely to be detected using either cell genotype or phenotype alone.

[00148] In various embodiments, the workflow for processing a single cell enables the determination of the presence or absence of integration of a foreign DNA segment in the genomic DNA of the cell. For example, in various embodiments, integration of a foreign DNA segment in the genomic DNA of a cell is detected by determining the presence of one or more amplicons. In various embodiments, a cell is exposed to reagents that include a foreign DNA segment-specific primer and a second primer (*e.g.*, a second foreign DNA segment-specific primer or a repeat sequence-specific primer), as well as proteases and, in some instances, transposases. DNA-seq can be performed to obtain sequencing reads of nucleic acid molecules (*e.g.*, amplicons) derived from genomic DNA. The sequencing reads obtained from DNA-seq are analyzed to determine the presence or absence of integration of a vector including a foreign DNA segment.

[00149] For example, the present disclosure provides methods for detecting integration of a vector including a foreign DNA segment into genomic DNA of a cell, the method including: (a) providing, within a droplet (or a tube), the genomic DNA of the cell and reagents, the genomic DNA including an integration site where the vector including the foreign DNA segment is integrated into the genomic DNA, wherein the reagents include a foreign DNA segment-specific primer and a second primer; (b) within the droplet, generating

one or more amplicons including the integrated foreign DNA segment, if present, using at least the hybridized foreign DNA segment-specific and second primers; and (c) determining the presence or absence of the one or more amplicons, wherein integration of the foreign DNA segment in the genomic DNA of the cell is detected by determining the presence of the one or more amplicons; and wherein no integration of the foreign DNA segment is detected by determining the absence of the one or more amplicons.

[00150] Such methods can also be performed in a tube to detect integration of a vector including a foreign DNA segment into genomic DNA of a cell. For example, the disclosure also provides a method for detecting integration of a vector including a foreign DNA segment into genomic DNA of a cell, the method including: (a) providing, in a bulk setting, the genomic DNA of the cell and reagents, the genomic DNA including an integration site where the vector including the foreign DNA segment is integrated into the genomic DNA, wherein the reagents include a foreign DNA segment-specific primer and a second primer; (b) within the tube, generating one or more amplicons including the integrated foreign DNA segment, if present, using at least the hybridized foreign DNA segment-specific and second primers; and (c) determining the presence or absence of the one or more amplicons, wherein integration of the foreign DNA segment in the genomic DNA of the cell is detected by determining the presence of the one or more amplicons; and wherein no integration of the foreign DNA segment is detected by determining the absence of the one or more amplicons.

[00151] In various embodiments, a second primer is not used. Therefore, the disclosure also provides methods for detecting integration of a vector including a foreign DNA segment into genomic DNA of a cell, the method including: (a) providing, within a droplet (or a tube), the genomic DNA of the cell and reagents, the genomic DNA including an integration site where the vector including the foreign DNA segment is integrated into the genomic DNA, wherein the reagents include a foreign DNA segment-specific primer; (b) within the droplet, generating one or more amplicons including the integrated foreign DNA segment, if present, using at least the hybridized foreign DNA segment-specific primer; and (c) determining the presence or absence of the one or more amplicons, wherein integration of the foreign DNA segment in the genomic DNA of the cell is detected by determining the presence of the one or more amplicons; and wherein no integration of the foreign DNA segment is detected by determining the absence of the one or more amplicons.

[00152] The disclosure also provides methods for detecting a proportion of cells in a population of cells having integration of a vector including a foreign DNA segment into

genomic DNA of the cells, the method including: (i) for each of one or more cells in the population of cells: (a) providing, within a droplet (or a tube), the genomic DNA of the cell and reagents, the genomic DNA including an integration site where the vector including the foreign DNA segment is integrated into the genomic DNA, wherein the reagents include a foreign DNA segment-specific primer and a second primer; (b) within the droplet, generating one or more amplicons including the integrated foreign DNA segment, if present, using at least the hybridized foreign DNA segment-specific and second primers; and (c) sequencing the generated one or more amplicons; and (ii) determining a proportion of the cells in the population of cells having integration of the foreign DNA segment in genomic DNA of the cells based on the sequenced one or more amplicons.

[00153] Alternatively, for example, bulk DNA or single-cell methods provided herein can be adapted for detecting translocation of a DNA segment in genomic DNA of a cell or for detecting genetic editing of a DNA segment of genomic DNA of a cell.

[00154] For example, the disclosure additionally provides a method for detecting translocation of a DNA segment in genomic DNA of a cell, the method including: (a) providing, within a droplet (or a tube), the genomic DNA of the cell and reagents, the genomic DNA including an integration site where the translocated DNA segment is integrated into the genomic DNA, wherein the reagents include a translocated DNA segment-specific primer and a second primer; (b) within the droplet, generating one or more amplicons including the integrated translocated DNA segment, if present, using at least the hybridized translocated DNA segment-specific and second primers; and (c) determining the presence or absence of the one or more amplicons, wherein integration of the translocated DNA segment in the genomic DNA of the cell is detected by determining the presence of the one or more amplicons; and wherein no integration of the translocated DNA segment is detected by determining the absence of the one or more amplicons.

[00155] The disclosure also provides a method for detecting genetic editing of a DNA segment of genomic DNA of a cell, the method including: (a) providing, within a droplet (or a tube), the genomic DNA of the cell and reagents, the genomic DNA including an integration site where the DNA segment is integrated into the genomic DNA, wherein the reagents include a DNA segment-specific primer and a second primer; (b) within the droplet, generating one or more amplicons including the integrated DNA segment, if present, using at least the hybridized DNA segment-specific and second primers; and (c) determining the presence or absence of the one or more amplicons, wherein integration of the DNA segment

in the genomic DNA of the cell is detected by determining the presence of the one or more amplicons; and wherein no integration of the DNA segment is detected by determining the absence of the one or more amplicons

5 **Methods for Performing Single-Cell Analysis**

Encapsulation, Analyte Release, Barcoding, and Amplification

[00156] Embodiments provided herein involve encapsulating one or more cells to perform single-cell analysis on the one or more cells. In various embodiments, the one or more cells can be isolated from a test sample obtained from a subject or a patient. In various
10 embodiments, the one or more host cells. In various embodiments, the test sample is obtained from host cells following treatment of the cells (*e.g.*, following transduction with viral DNA, modified viral DNA, or a viral vector). Thus, single-cell analysis of the cells enables cellular and cellular quantification of the transduction of a foreign DNA segment.

[00157] In various embodiments, the disclosure provides providing, within a droplet (or a
15 tube), the genomic DNA of a cell and reagents, the genomic DNA potentially including an integration site where the vector including a foreign DNA segment is integrated into the genomic DNA, wherein the reagents include a foreign DNA segment-specific primer and a second primer (*e.g.*, a second foreign DNA segment-specific primer or a repeat sequence-specific primer).

[00158] In various embodiments, the second primer is a second foreign DNA segment-specific primer, and the method includes: incubating the second foreign DNA segment-specific primer under conditions to promote hybridization of the second foreign DNA segment-specific primer to a second vector sequence, if present in the integration site. In various embodiments, the method further includes incubating the reaction mixture under
20 conditions to promote amplification of the genomic DNA integration site including the foreign DNA segment, if present, using the hybridized second foreign DNA segment-specific primer.

[00159] For example, in various embodiments, the foreign DNA segment-specific primer and/or the second foreign DNA segment-specific primer has the nucleic acid sequence of any
25 one of SEQ ID NOs: 1-11, as found in **Table 1**, below.

Table 1. Exemplary Foreign DNA segment-specific primers

SEQ ID NO.	Target	Sequence
1	AAV2	AACTCCATCACTAGGGGTTCC
2	AAV2	CGGCCTCAGTGAGCGA
3	LTR 4	GATCTGGTCTAACCAGAGAG
4	LTR 6	CCCAGTACAAGCAAAAAGCAG
5	LTR 1	AGTAGTGTGTGCCCGTCTGT
6	LTR 5	GATCCCTCAGACCCTTTTAGTC
7	LTR 7	GACTAAAAGGGTCTGAGGGATC
8	LTR 8	GATCCCTCAGACCCTTTTAGTCAGAGT
9	LTR 9	TAAAGCTTGCCTTGAGTGCTTCA
10	LTR 10	CTCTGGTAACTAGAGATCCCTCA
11	LTR 11	CTGAGGGATCTCTAGTTACC

[00160] In various embodiments, a foreign DNA segment-specific primer is the nucleic acid of SEQ ID NO: 1. In various embodiments, a foreign DNA segment-specific primer is the nucleic acid of SEQ ID NO: 2. In various embodiments, a foreign DNA segment-specific primer is the nucleic acid of SEQ ID NO: 3. In various embodiments, a foreign DNA segment-specific primer is the nucleic acid of SEQ ID NO: 4. In various embodiments, a foreign DNA segment-specific primer is the nucleic acid of SEQ ID NO: 5. In various embodiments, a foreign DNA segment-specific primer is the nucleic acid of SEQ ID NO: 6. In various embodiments, a foreign DNA segment-specific primer is the nucleic acid of SEQ ID NO: 7. In various embodiments, a foreign DNA segment-specific primer is the nucleic acid of SEQ ID NO: 8. In various embodiments, a foreign DNA segment-specific primer is the nucleic acid of SEQ ID NO: 9. In various embodiments, a foreign DNA segment-specific primer is the nucleic acid of SEQ ID NO: 10. In various embodiments, a foreign DNA segment-specific primer is the nucleic acid of SEQ ID NO: 11.

[00161] Alternatively, for example, in various embodiments, the second primer is a repeat sequence-specific primer, and the method includes: generating a reaction mixture by incubating the foreign DNA segment-specific primer and the repeat sequence-specific primer under conditions to promote hybridization of the repeat sequence-specific primer to a repeat sequence present in the genomic DNA.

- [00162]** In various embodiments, the repeat sequence-specific primer is an Alu1, an Alu2, a LINE1, a 16S, or an 18S primer. For example, in various embodiments, the repeat sequence-specific primer is an Alu1 primer (*e.g.*, a primer having the nucleic acid sequence of any one of SEQ ID NOs: 12, 14, or 16). In various embodiments, the repeat sequence-specific primer is an Alu2 primer (*e.g.*, a primer having the nucleic acid sequence of any one of SEQ ID NOs: 13, 15, or 17). In various embodiments, the repeat sequence-specific primer is a LINE1 primer (*e.g.*, a primer having the nucleic acid sequence of any one of SEQ ID NOs: 22-25). In various embodiments, the repeat sequence-specific primer is an 18S primer (*e.g.*, a primer having the nucleic acid sequence of any one of SEQ ID NOs: 18-21).
- [00163]** For example, in various embodiments, the repeat sequence-specific primer has the nucleic acid sequence of any one of SEQ ID NOs 12-25, as found in **Table 2**, below.

Table 2. Exemplary Repeat Sequence-Specific Primers

SEQ ID NO.	Name	Sequence
12	Alu 1A	TCCCAGCTACTGGGGAGGCTGAGG
13	Alu 2A	GCCTCCCAAAGTGCTGGGATTACAG
14	Alu 1B	TGTAATCCCAGCTACTCGGG
15	Alu 2B	GAACTCCTGACCTCAGG
16	Alu 1D	CCTCAGCCTCCCCAGTAGCTGGGA
17	Alu 2D	CTGTAATCCCAGCACTTTGGGAGGC
18	18S 1A	CGCTCCACCAACTAAGAACG
19	18S 2A	CGTTCTTAGTTGGTGGAGCG
20	18S 1D	CGCTCCACCAACTAAGAACG
21	18S 2D	CTCAACACGGGAAACCTCAC
22	LINE1 1A	GCAGTGTGTAGAGGGAA
23	LINE1 2A	GTGGTCAATTTTGAATAGGTGT
24	LINE1 1D	ACACCTATTCCAAAATTGACCAC
25	LINE1 2D	TTCCCTCTACACACTGC

[00164] In various embodiments, a repeat sequence-specific primer is the nucleic acid of SEQ ID NO: 12. In various embodiments, a repeat sequence-specific primer is the nucleic acid of SEQ ID NO: 13. In various embodiments, a repeat sequence-specific primer is the nucleic acid of SEQ ID NO: 14. In various embodiments, a repeat sequence-specific primer is the nucleic acid of SEQ ID NO: 15. In various embodiments, a repeat sequence-specific primer is the nucleic acid of SEQ ID NO: 16. In various embodiments, a repeat sequence-specific primer is the nucleic acid of SEQ ID NO: 17. In various embodiments, a repeat sequence-specific primer is the nucleic acid of SEQ ID NO: 18. In various embodiments, a repeat sequence-specific primer is the nucleic acid of SEQ ID NO: 19. In various
embodiments, a repeat sequence-specific primer is the nucleic acid of SEQ ID NO: 20. In various embodiments, a repeat sequence-specific primer is the nucleic acid of SEQ ID NO: 21. In various embodiments, a repeat sequence-specific primer is the nucleic acid of SEQ ID NO: 22. In various embodiments, a repeat sequence-specific primer is the nucleic acid of SEQ ID NO: 23. In various embodiments, a repeat sequence-specific primer is the nucleic acid of SEQ ID NO: 24. In various embodiments, a repeat sequence-specific primer is the nucleic acid of SEQ ID NO: 25.

[00165] In various embodiments, a repeat sequence-specific primer is a combination of one or one or more repeat sequence-specific primers, such as SEQ ID NOs: SEQ ID NOs: 14, 15, 18, 19, 22, and/or 23).

[00166] In various embodiments, encapsulating a cell (*e.g.*, within a droplet or a tube) with reagents is accomplished by combining an aqueous phase including the cell and reagents with an immiscible oil phase. In various embodiments, an aqueous phase including the cell and reagents are flowed together with a flowing immiscible oil phase such that water in oil emulsions are formed, where at least one emulsion includes a single cell and the reagents. In various embodiments the immiscible oil phase includes a fluoruous oil, a fluoruous non-ionic surfactant, or both. In various embodiments, emulsions can have an internal volume of about 0.001 picoliters to 1000 picoliters or more and can range from 0.1 μm to 1000 μm in diameter.

[00167] In various embodiments, the aqueous phase including the cell and reagents need not be simultaneously flowing with the immiscible oil phase. For example, the aqueous phase can be flowed to contact a stationary reservoir of the immiscible oil phase, thereby enabling the budding of water in oil emulsions within the stationary oil reservoir.

[00168] In various embodiments, combining the aqueous phase and the immiscible oil phase can be performed in a microfluidic device. For example, the aqueous phase can flow through a microchannel of the microfluidic device to contact the immiscible oil phase, which is simultaneously flowing through a separate microchannel or is held in a stationary reservoir of the microfluidic device. The encapsulated cell and reagents within an emulsion can then be flowed through the microfluidic device to undergo cell lysis.

[00169] Further example embodiments of adding reagents and cells to emulsions can include merging emulsions that separately contain the cells and reagents or picoinjecting reagents into an emulsion. Further description of example embodiments is described in US Application No. 14/420,646, which is hereby incorporated by reference in its entirety.

[00170] The encapsulated cell in an emulsion is lysed to generate cell lysate. In various embodiments, a cell is lysed by lysing agents that are present in the reagents. For example, the reagents can include a lysis buffer (*e.g.*, protease and a detergent) or a cell buffer, such as a cell buffer including a detergent such as NP40 (*e.g.*, Tergitol-type NP-40 or nonyl phenoxypolyethoxylethanol) which lyses the cell membrane. In various embodiments, cell lysis may also, or instead, rely on techniques that do not involve a lysing agent in the reagent. For example, lysis may be achieved by mechanical techniques that may employ various geometric features to effect piercing, shearing, abrading, etc. of cells. Other types of mechanical breakage such as acoustic techniques may also be used. Further, thermal energy can also be used to lyse cells. Any convenient means of effecting cell lysis may be employed in the methods Provided herein. The lysed cell may include analytes within the cytoplasm of the cell such as genomic DNA (*e.g.*, genomic DNA having a foreign DNA segment integrated).

[00171] In various embodiments, the cell buffer includes one or more of a detergent, a density-match agent, and a phosphate buffer. In various embodiments, the detergent is a pluronic detergent. In various embodiments, the density-match agent is optiprep.

[00172] In various embodiments, the lysis buffer includes one or more of a reverse primer, a protease, a detergent, an RNA reverse transcriptase, an RNAase inhibitor, a transposase, and a magnesium buffer. In various embodiments, the lysis buffer includes a protease, a detergent, a transposase, and a magnesium buffer.

[00173] In various embodiments, the magnesium buffer includes magnesium, Tris, potassium, [tris(hydroxymethyl)methylamino]propanesulfonic acid (TAPS), dimethylformamide (DMF), and/or poly(ethylene glycol) (PEG). For example, in various

embodiments, the magnesium buffer includes magnesium and Tris. In various embodiments, the magnesium buffer includes magnesium, Tris, and potassium. In various embodiments, the magnesium buffer includes magnesium and TAPS. In various embodiments, any of the above described magnesium buffers further includes DMF and/or PEG.

5 [00174] In various embodiments, the reaction mixture includes components, such as primers, for performing a nucleic acid reaction on target nucleic acids. Primers may include a foreign DNA segment-specific primer and a second primer (*e.g.*, a second foreign DNA segment-specific primer or a repeat sequence-specific primer). Additional primers may include a barcode primer including a barcode identification sequence (*e.g.*, a bead barcode primer), a read 1 sequencing primer, and/or a read 2 sequencing primer. For example, in 10 various embodiments, the method includes performing nucleic acid extension including extending a barcode primer including a barcode identification sequence. In various embodiments, the method includes performing nucleic acid extension including extending a read 1 sequencing primer. In various embodiments, the method includes performing nucleic acid extension including extending a read 2 sequencing primer. 15

[00175] In various embodiments, an additional primers may hybridize to a sequence present in the genomic DNA or a segment of the foreign DNA, if it exists. For example, in various embodiments, an additional primer may hybridize to a sequence present in the genomic DNA. In various embodiments, an additional primer may hybridize to a segment of 20 the foreign DNA, if it exists.

[00176] In various embodiments, a cell lysate is encapsulated with a reaction mixture and a barcode primer including a barcode identification sequence (*e.g.*, a bead barcode primer) by combining an aqueous phase including the reaction mixture and the barcode with the cell lysate and an immiscible oil phase. In various embodiments, an aqueous phase including the reaction mixture and the barcode are flowed together with a flowing cell lysate and a flowing immiscible oil phase such that water in oil emulsions are formed, where at least one emulsion includes a cell lysate, the reaction mixture, and the barcode. In various embodiments the immiscible oil phase includes a fluorous oil, a fluorous non-ionic surfactant, or both. In various embodiments, emulsions can have an internal volume of about 0.001 picoliters to 25 1000 picoliters or more and can range from 0.1 μm to 1000 μm in diameter. 30

[00177] In various embodiments, combining the aqueous phase and the immiscible oil phase can be performed in a microfluidic device. For example, the aqueous phase can flow through a microchannel of the microfluidic device to contact the immiscible oil phase, which

is simultaneously flowing through a separate microchannel or is held in a stationary reservoir of the microfluidic device. The encapsulated cell lysate, reaction mixture, and barcode within an emulsion can then be flowed through the microfluidic device to perform amplification of target nucleic acids.

- 5 **[00178]** Further example embodiments of adding reaction mixture and barcodes to emulsions can include merging emulsions that separately contain the cell lysate and reaction mixture and barcodes or picoinjecting the reaction mixture and/or barcode into an emulsion.
- 10 **[00179]** Once the reaction mixture and barcode are added to an emulsion, the emulsion may be incubated under conditions that facilitates the nucleic acid amplification reaction (*e.g.*, nucleic acid extension *e.g.*, primer extension). In various embodiments, the emulsion may be incubated on the same microfluidic device as was used to add the reaction mixture and/or barcode, or may be incubated on a separate device. In certain embodiments, incubating the emulsion under conditions that facilitates nucleic acid amplification is performed on the same microfluidic device used to encapsulate the cells and lyse the cells. Incubating the
- 15 emulsions may take a variety of forms. In certain aspects, the emulsions containing the reaction mix, barcode, and cell lysate may be flowed through a channel that incubates the emulsions under conditions effective for nucleic acid amplification. Flowing the microdroplets through a channel may involve a channel that snakes over various temperature zones maintained at temperatures effective for PCR. Such channels may, for example, cycle
- 20 over two or more temperature zones, wherein at least one zone is maintained at about 65 °C and at least one zone is maintained at about 95 °C. As the drops move through such zones, their temperature cycles, as needed for nucleic acid amplification. The number of zones, and the respective temperature of each zone, may be readily determined by those of skill in the art to achieve the desired nucleic acid amplification.
- 25 **[00180]** In various embodiments, nucleic acid extension includes extending a foreign DNA segment-specific primer to produce one or more amplicons including a constant region sequence and a foreign DNA segment-specific primer. In various embodiments, performing nucleic acid extension includes extending a second foreign DNA segment-specific primer to produce a one or more amplicons including a constant region sequence and a second foreign
- 30 DNA segment-specific primer.
- [00181]** In various embodiments, nucleic acid extension includes producing one or more amplicons including a complement sequence of a foreign DNA segment.

[00182] In various embodiments, nucleic acid extension includes extending a barcode identification sequence to produce one or more amplicons including a first read sequence, a barcode identification sequence, and a constant region sequence.

[00183] In various embodiments, nucleic acid extension includes extending a second foreign DNA segment-specific primer to produce one or more amplicons including a second foreign DNA segment-specific primer and a second read sequence.

[00184] In various embodiments, nucleic acid extension includes extending a repeat sequence-specific primer (*e.g.*, an Alu primer) to produce one or more amplicons including a constant region sequence and a repeat sequence-specific primer.

[00185] In various embodiments, nucleic acid extension includes extending the read 1 sequencing primer to produce the one or more amplicons including a first index sequence and a first read sequence.

[00186] In various embodiments, nucleic acid extension includes extending the read 2 sequencing primer to produce the one or more amplicons including the second read sequence and a second index sequence.

[00187] In various embodiments, following nucleic acid amplification, emulsions containing the amplified nucleic acids are collected. In various embodiments, the emulsions are collected in a well, such as a well of a microfluidic device. In various embodiments, the emulsions are collected in a reservoir or a tube, such as an Eppendorf tube. In various embodiments, the method further includes breaking an emulsion that includes the droplet and performing nucleic acid extension, such as PCR. Once collected, the amplified nucleic acids across the different emulsions are pooled. In various embodiments, the emulsions are broken by providing an external stimuli to pool the amplified nucleic acids. In various embodiments, the emulsions naturally aggregate over time given the density differences between the aqueous phase and immiscible oil phase. Thus, the amplified nucleic acids pool in the aqueous phase.

[00188] Following pooling, the amplified nucleic acids can undergo further preparation for sequencing. For example, sequencing adapters can be added to the pooled nucleic acids. Example sequencing adapters are P5 and P7 sequencing adapters. The sequencing adapters enable the subsequent sequencing of the nucleic acids.

Tagmentation

[00189] The present disclosure provides, among other things, a method including tagging genomic DNA using reagents (*e.g.*, a transposase and a transposase adapter *e.g.*, a transposase preloaded with the transposase adapter) to obtain tagmented DNA. Tagmentation refers to the modification of DNA by a transposome complex including transposase enzyme and transposon end sequence in which the transposon end sequence further includes adaptor sequence. Tagmentation results in the simultaneous fragmentation of the DNA and ligation of the adaptors to the 5' ends of both strands of duplex fragments.

[00190] In various embodiments, transposon-based technology can be utilized for fragmenting DNA, for example as exemplified in the workflow for Nextera™ DNA sample preparation kits (Illumina, inc.), in which genomic DNA can be fragmented by an engineered transposome that simultaneously fragments and tags input DNA (“tagmentation”) thereby creating a population of fragmented nucleic acid molecules which include unique adapter sequences at the ends of the fragments.

[00191] For example, in various embodiments, the disclosure provides a method including tagging genomic DNA using reagents to obtain tagmented DNA fragments, in which at least one of the tagmented DNA fragments includes a foreign DNA segment. In various embodiments, the method further includes amplification of the at least one of the tagmented DNA fragments.

[00192] Tagmentation may be performed, for example, in a droplet or a tube. In various embodiments, the droplet in which the genomic DNA is tagmented (*e.g.*, a second droplet) differs from the droplet in which genomic DNA of a cell and reagents are provided. In various embodiments, genomic DNA of the cell and reagents are provided in the same droplet as the droplet in which the genomic DNA is tagmented. Alternatively, for example, tagmentation may be performed, for example, in a tube.

[00193] Following a purification step to remove the transposase enzyme, additional sequences can be added to the ends of the adapted fragments, for example by PCR, ligation, or any other suitable methodology known to those of skill in the art.

[00194] The method of the invention can use any transposase that can accept a transposase end sequence and fragment a target nucleic acid, attaching a transferred end, but not a non-transferred end. A transposome includes at least a transposase enzyme and a transposase recognition site. In some such systems, termed ‘transposomes,’ the transposase can form a functional complex with a transposon recognition site that is capable of catalyzing a

transposition reaction. The transposase or integrase may bind to the transposase recognition site and insert the transposase recognition site into a target nucleic acid in a process sometimes termed “tagmentation.” In some such insertion events, one strand of the transposase recognition site may be transferred into the target nucleic acid.

- 5 **[00195]** In standard sample preparation methods, each template contains an adaptor at either end of the insert and often a number of steps are included to both modify the DNA or RNA and to purify the desired products of the modification reactions. These steps are performed in solution prior to the addition of the adapted fragments to a droplet (or tube) where they are coupled to the surface by a primer extension reaction that copies the
- 10 hybridized fragment onto the end of a primer covalently attached to the surface. These ‘seeding’ templates then give rise to monoclonal clusters of copied templates through several cycles of amplification.
- [00196]** In various embodiments, an additional primers may hybridize to a transposase adapter, which may have integrated in the genomic DNA or a segment of the foreign DNA, if
- 15 it exists. For example, in various embodiments, an additional primer may hybridize to a transposase adapter sequence present in the genomic DNA. In various embodiments, an additional primer may hybridize to a transposase adapter sequence present in a segment of the foreign DNA, if it exists.
- [00197]** In various embodiments, an adapter is a Tn5 adapter. In various embodiments, a
- 20 Tn5 adapter has the nucleic acid sequence of any one of SEQ ID NOs: 26-29 (**Table 3**).

SEQ ID NO.	Name	Sequence
26	Tn5 adaptor 1	GCGAGCAGTGTCAGCAGATGTGTATAAGAGACAG
27	Tn5 adaptor 2	GCGAGCAGTGTCAGCAGATGTGATCAAGAGACAG
28	Tn5 adaptor 3	GTACTCGCAGTAGTCAGATGTGTATAAGAGACAG
29	Tn5 adaptor 4	ACTCGCAGTAGTCAGATGTGTATAAGAGACAG

[00198] In various embodiments, a Tn5 adapter has the nucleic acid sequence of SEQ ID NO: 26. In various embodiments, a Tn5 adapter has the nucleic acid sequence of any one of SEQ ID NO: 27. In various embodiments, a Tn5 adapter has the nucleic acid sequence of any

one of SEQ ID NO: 28. In various embodiments, a Tn5 adapter has the nucleic acid sequence of any one of SEQ ID NO: 29.

[00199] The number of steps included to transform DNA into adaptor-modified templates in solution ready for cluster formation and sequencing can be minimized by the use of transposase-mediated fragmentation and tagging.

[00200] For example, in various embodiments, the disclosure provides a method including fragmenting the genomic DNA using the reagents includes inserting adaptor sequences to obtain tagmented DNA fragments including the adaptor sequences.

[00201] In various embodiments, each of the tagmented DNA fragments include at most one adaptor sequence.

[00202] Various embodiments can include the use of a hyperactive Tn5 transposase and a Tn5-type transposase recognition site (Goryshin and Reznikoff, *J. Biol. Chem.*, 273:7367 (1998)), or MuA transposase and a Mu transposase recognition site including R1 and R2 end sequences (Mizuuchi, K., *Cell*, 35: 785, 1983; Savilahti, H, et al., *EMBO J.*, 14: 4893, 1995).

An exemplary transposase recognition site that forms a complex with a hyperactive Tn5 transposase is the EZ-Tn5™ Transposase, Epicentre Biotechnologies, Madison, Wis..

[00203] More examples of transposition systems that can be used with certain embodiments provided herein include *Staphylococcus aureus* Tn552 (Colegio et al., *J. Bacteriol.*, 183: 2384-8, 2001; Kirby C et al., *Mol. Microbiol.*, 43: 173-86, 2002), Ty1 (Devine & Boeke, *Nucleic Acids Res.*, 22: 3765-72, 1994 and International Publication WO 95/23875), Transposon Tn7 (Craig, N L, *Science*. 271: 1512, 1996; Craig, N L, Review in: *Curr Top Microbiol Immunol.*, 204:27-48, 1996), Tn/O and IS10 (Kleckner N, et al., *Curr Top Microbiol Immunol.*, 204:49-82, 1996), Mariner transposase (Lampe D J, et al., *EMBO J.*, 15: 5470-9, 1996), Tc1 (Plasterk R H, *Curr. Topics Microbiol. Immunol.*, 204: 125-43, 1996), P Element (Gloor, G B, *Methods Mol. Biol.*, 260: 97-114, 2004), Tn3 (Ichikawa & Ohtsubo, *J Biol. Chem.* 265:18829-32, 1990), bacterial insertion sequences (Ohtsubo & Sekine, *Curr. Top. Microbiol. Immunol.* 204: 1-26, 1996), retroviruses (Brown, et al., *Proc Natl Acad Sci USA*, 86:2525-9, 1989), and retrotransposon of yeast (Boeke & Corces, *Annu Rev Microbiol.* 43:403-34, 1989). More examples include IS5, Tn10, Tn903, IS911, and engineered versions of transposase family enzymes (Zhang et al., (2009) *PLoS Genet.* 5:e1000689. Epub 2009 Oct. 16; Wilson C. et al (2007) *J. Microbiol. Methods* 71:332-5).

Additionally, the methods and compositions provided herein are useful with transposase of *Vibrio* species, including *Vibrio harveyi*, as set forth in greater detail in the disclosures of

US 2014/0093916 and 2012/0301925, each of which is incorporated by reference in its entirety. In various embodiments, the transposase is a Tn5 transposase.

[00204] The adapters that are added to the 5'- and/or 3'-end of a nucleic acid can include a universal sequence. A universal sequence is a region of nucleotide sequence that is common to, i.e., shared by, two or more nucleic acid molecules. Optionally, the two or more nucleic acid molecules also have regions of sequence differences. Thus, for example, the 5' adapters can include identical or universal nucleic acid sequences and the 3' adapters can include identical or universal sequences. A universal sequence that may be present in different members of a plurality of nucleic acid molecules can allow the replication or amplification of multiple different sequences using a single universal primer that is complementary to the universal sequence.

[00205] In various embodiments, the transposase adapter is a Tn5 transposase adapter. An extension product of such an adapter may be used to hybridize a second primer (*e.g.*, a second foreign DNA segment-specific primer). In various embodiments, the transposase adapter may be preloaded to the transposase.

[00206] In various embodiments, tagmenting the genomic DNA using the reagents does not include performing an extension to fill one or more gaps.

Exemplary Barcoding of Antibody-Conjugated Oligonucleotide and Genomic DNA

[00207] The methods provided herein can be used to determine one or more analytes expressed in bulk DNA or by a cell or a population of cells. In various embodiments, the one or more analytes include genomic DNA (*e.g.*, for single nucleotide variants and/or copy number variations). In various embodiments, the one or more analytes include proteins. In various embodiments, the one or more embodiments include both genomic DNA and protein expression. Further details for performing single-cell analysis of genomic DNA and protein expression is described in WO 2021/030447, which is incorporated by reference in its entirety.

[00208] The determination of one or more analytes expressed in bulk DNA or by a cell or a population of cells can be identified by a finding that a cell or a population of cells (*e.g.*, at least one cell within the population or one analyte within the bulk DNA) is bound to at least one analyte-bound antibody-conjugated oligonucleotide. In various embodiments, an antibody oligonucleotide includes a PCR handle, a tag sequence (*e.g.*, an antibody tag), and a capture sequence that links the oligonucleotide to the antibody. In various embodiments, the

antibody oligonucleotide is conjugated to a region of the antibody, such that the antibody's ability to bind a target epitope is unaffected. For example, the antibody oligonucleotide can be linked to a Fc region of the antibody, thereby leaving the variable regions of the antibody unaffected and available for epitope binding. In various the antibody oligonucleotide can include a unique molecular identifier (UMI). In various embodiments, the UMI can be inserted before or after the antibody tag. In various embodiments, the UMI can flank either end of the antibody tag. In various embodiments, the UMI enables the identification of the particular antibody oligonucleotide and antibody combination.

[00209] In various embodiments, the antibody oligonucleotide includes more than one PCR handle. For example, the antibody oligonucleotide can include two PCR handles, one on each end of the antibody oligonucleotide. In various embodiments, one of the PCR handles of the antibody oligonucleotide is conjugated to the antibody. Here, a foreign DNA segment-specific primer and an optional second primer can be provided that hybridize with the two PCR handles, thereby enabling amplification of the antibody oligonucleotide.

[00210] In various embodiments, the second primer comprises a cell barcode

[00211] Generally, the antibody tag of the antibody oligonucleotide enables the subsequent identification of the antibody (and corresponding protein). For example, the antibody tag can serve as an identifier *e.g.*, a barcode for identifying the type of protein for which the antibody binds to. In various embodiments, antibodies that bind to the same target are each linked to the same antibody tag. For example antibodies that bind to the same epitope of a target protein are each linked to the same antibody tag, thereby enabling the subsequent determination of the presence of the target protein. In various embodiments, antibodies that bind different epitopes of the same target protein can be linked to the same antibody tag, thereby enabling the subsequent determination of the presence of the target protein.

[00212] In various embodiments, an oligonucleotide sequence is encoded by its nucleobase sequence and thus confers a combinatorial tag space far exceeding what is possible with conventional approaches using fluorescence. For example, a modest tag length of ten bases provides over a million unique sequences, sufficient to label an antibody against every epitope in the human proteome. Indeed, with this approach, the limit to multiplexing is not the availability of unique tag sequences but, rather, that of specific antibodies that can detect the epitopes of interest in a multiplexed reaction.

[00213] A primer may include a PCR handle and a common sequence. The PCR handle of the primer may be complementary to the PCR handle of the antibody oligonucleotide. Thus,

the primer primes the antibody oligonucleotide given the hybridization of the PCR handles. In various embodiments, extension occurs from the PCR handle of the antibody oligonucleotide. In various embodiments, extension occurs from the PCR handle of the primer, thereby generating a nucleic acid with the antibody tag and capture sequence.

5 **[00214]** A barcode (*e.g.*, cell barcode) can be releasably attached to a bead and further linked to a common sequence. The common sequence linked to the cell barcode can be complementary to the common sequence linked to the PCR handle, antibody tag, and capture sequence. The antibody oligonucleotide can be extended to include the common sequence and cell barcode.

10 **[00215]** In various embodiments, the antibody oligonucleotide can be amplified, thereby generating amplicons with the cell barcode, common sequence, PCR handle, antibody tag, and capture sequence. In various embodiments, the capture sequence contains a biotin oligonucleotide capture site, which enables streptavidin bead enrichment prior to library preparation. In various embodiments, the barcoded antibody-oligonucleotides can be enriched
15 by size separation from the amplified genomic DNA targets.

[00216] In various embodiments, determining the presence or absence of the analyte includes determining an expression level of the analyte, in which the analyte is bound by the antibody conjugated to the oligonucleotide. Using such methods, one may generate a targeted DNA library or a targeted protein library, Provided below in the section titled ‘Targeted
20 Panels.’

[00217] Such antibody-conjugated oligonucleotides may be used to determine one or more mutations in a cell, a population of cells, or in bulk DNA (*e.g.*, cell lysate in bulk DNA). For example, in various embodiments, the disclosure provides determining one or more mutations by performing a nucleic acid amplification reaction within a droplet or tube using an
25 antibody-conjugated oligonucleotide to (a) generate one or more amplicons, the one or more amplicons including an amplicon derived from the oligonucleotide; (b) determining a presence or absence of an analyte using the one or more amplicons; and (c) characterizing the presence or absence of the analyte.

30 **Exemplary Processing of DNA for Detecting Viral Integration Site**

[00218] This section provides several exemplary methods of processing of DNA, such as genomic DNA, that may be performed in the methods provided herein. In various embodiments, the DNA can include a viral integration site in which foreign DNA has been

integrated into the DNA. In various embodiments, genomic DNA further includes one or more additional integration sites where copies of the foreign DNA segment are integrated into the genomic DNA. Thus, the exemplary methods of processing DNA disclosed herein can be used to detect presence of the viral integration site and/or determine vector copy number of the foreign DNA.

Tagmentation-Based Methodology for Detecting Viral Integration Site

[00219] In various embodiments, an exemplary method for processing DNA can involve a tagmentation-based methodology. In various embodiments, the tagmentation-based methodology involves a two-step process in which a first step involves encapsulating and lysing a cell, followed by a second step involving amplification and barcoding of amplicons including foreign DNA sequences. Generally, the tagmentation-based methodology includes a step of tagmenting genomic DNA of the cell. In various embodiments, the tagmentation occurs during a first step of the two-step process (e.g., in a droplet involving lysis of the cell).

In various embodiments, the tagmentation occurs during a second step of the two-step process (e.g., in a droplet involving amplification and barcoding of amplicons).

[00220] In various embodiments, the methodology begins with encapsulating a single cell in a droplet, followed by lysing the cell within the droplet to generate a cell lysate. In various embodiments, a cell is lysed by lysing agents. For example, the lysing reagents can include a detergent such as NP-40 and/or a protease. The detergent and/or the protease can lyse the cell membrane.

[00221] In various embodiments, tagmentation can be performed in this droplet within which the cell was lysed. For example, the cell can be encapsulated with lysing agents as well as tagmentation reagents (e.g., a transposase and a transposase adapter e.g., a transposase preloaded with the transposase adapter) in the droplet. Thus, the cell is lysed within the droplet and genomic DNA, including foreign DNA integrated into the genomic DNA, if present, undergoes tagmentation. The left panel of **FIG. 1** shows an example process in which tagmentation occurs in this droplet. Specifically, a transposase with transposase adapters cleaves the genomic DNA at tagmentation sites and inserts adapters (e.g., Tn5 adapters) at the ends of the cleaved fragments. Further details of the tagmentation process are described herein.

[00222] In various embodiments, the genomic DNA of the cell is contacted with reagents. In various embodiments, the genomic DNA of the cell is encapsulated in a droplet, such as a second droplet that differs from the droplet in which the cell was lysed. The right panel of **FIG. 1** (labeled as “barcoding”) shows an exemplary process that may occur in the droplet (e.g., a second droplet). The reagents may include primers, such as at least a foreign DNA segment-specific primer (referred to in the right panel of **FIG. 1** as a “vector specific primer”) and, optionally, a second primer, where at least the foreign DNA segment-specific primer hybridizes with a segment of the foreign DNA segment. In various embodiments, the reagents are provided in a reaction mixture, which includes the primer(s) that are capable of acting as a point of initiation of synthesis along a complementary strand when placed under conditions in which synthesis of a primer extension product which is complementary to a nucleic acid strand is catalyzed. In various embodiments, the reaction mixture includes the four different deoxyribonucleoside triphosphates (adenine, guanine, cytosine, and thymine). In various embodiments, the reaction mixture includes enzymes for nucleic acid amplification.

[00223] The exemplary method may then entail hybridizing a foreign DNA segment-specific primer to the foreign DNA segment, if present in the integration site and extending the hybridized foreign DNA segment-specific primer to generate an extension product including a sequence derived from a transposase adapter sequence. Specifically, as shown in the right column of the right panel of **FIG. 1**, the foreign DNA segment-specific primer (referred to as “vector specific primer”) can contact a sequence of the vector in the tagmented DNA. In various embodiments, such as the embodiment shown in the right panel of **FIG. 1**, the vector specific primer may have a constant sequence that does not hybridize with a sequence of the foreign DNA. Here, the constant sequence may be useful for subsequently incorporating adapters, such as library sequencing adapters. In various embodiments, such as the embodiment shown in **FIG. 2**, the vector specific primer may only include a sequence that hybridizes with a sequence of the foreign DNA. Specifically, the vector specific primer shown in **FIG. 2** does not include the constant sequence shown in the right panel of **FIG. 1**. Thus, library sequencing adapters can be later incorporated in bulk (e.g., as shown below in **FIG. 2** as the “Illumina P7 adaptor”).

[00224] In various embodiments, a reaction mixture can be generated by incubating the foreign DNA segment-specific primer under conditions to promote hybridization of the foreign DNA segment-specific primer to a foreign DNA segment, if present in the integration

site. Extension is initiated beginning at the vector specific primer (as shown by the directional arrow) to generate an extension product that includes the sequence derived from a transposase adapter sequence (annotated as “Tn5 adapter” in **FIG. 1**). Then, within the emulsion (*e.g.*, droplet), the reaction mixture can be incubated under conditions to promote amplification of the genomic DNA integration site including the foreign DNA segment, if present, using the hybridized vector-specific primer to generate one or more amplicons including the integrated foreign DNA segment, if present, and a sequence derived from a transposase adapter sequence. In contrast, in the left column of the right panel of **FIG. 1**, the tagged DNA does not include a foreign DNA sequence. Thus, this tagged DNA does not undergo extension or amplification because the foreign DNA segment-specific primer does not hybridize with the tagged DNA.

[00225] As shown in the right panel of **FIG. 1**, the method may then include hybridizing a second foreign DNA segment-specific primer (annotated as “seq8F” in **FIG. 1** or “Z” in **FIG. 7**) to the sequence derived from a transposase adapter sequence. The second foreign DNA segment-specific primer can be linked to a constant region, such as a PCR handle. The PCR handle of the foreign DNA segment-specific primer is complementary to a PCR handle linked to a barcode primer including a barcode identification sequence (*e.g.*, a bead barcode primer) sequence (annotated as “CBC” in **FIG. 1** or “Bead Barcode” in **FIG. 9**). Thus, synthesis can occur off of the second foreign DNA segment-specific primer.

[00226] Here, the amplified nucleic acid includes sequences of a first index sequence (P5 sequence adapter; annotated as “P5+Index 1” in **FIG. 9**), a first read sequence (annotated as “Read 1” in **FIG. 9**), the barcode (CBC” in **FIG. 1** or “Bead Barcode” in **FIG. 9**), a constant region sequence (the first PCR handle; annotated as “Constant Region” in **FIG. 9**), the foreign DNA segment-specific primer (the forward primer; annotated as “GSP-FWD in **FIG. 9**), the complement sequence of the foreign DNA segment (cDNA; annotated as “Region of Interest” in **FIG. 9**), the second foreign DNA segment-specific primer (the reverse primer; annotated as “GSP-REV” in **FIG. 9**), an optional second read sequence (the second PCR handle; annotated as “Read 2” in **FIG. 9**), and the second index sequence (a P7 sequence adapter; annotated as “Index 2+P7” in **FIG. 9**). In one scenario, the read 2 sequence can be included in the second PCR handle linked to the reverse primer sequence. In another scenario, the read 2 sequence can be included in the P7 sequence adapter.

[00227] To generate the amplicons, the droplet can be exposed to an increased temperature range (*e.g.*, increased relative to physiological temperatures), such as a temperature between

40 °C - 60 °C. In various embodiments, the emulsion can be exposed to an increased temperature of 40 °C, 41 °C, 42 °C, 43 °C, 44 °C, 45 °C, 46 °C, 47 °C, 48 °C, 49 °C, 50 °C.

[00228] Although both panels of **FIG. 1** show an embodiment in which tagmentation occurs in a first droplet followed by barcoding in a second droplet, in various embodiments, the tagmentation may occur in the second droplet. For example, the method may include lysing the cell in a first droplet. Next, the method further includes providing, in a second droplet, genomic DNA of the cell and reagents (*e.g.*, at least a transposase and a transposase adapter), the genomic DNA including an integration site where a foreign DNA segment is integrated into the genomic DNA and within a droplet, tagmenting the genomic DNA using the reagents to obtain tagmented DNA fragments, in which at least one of the tagmented DNA fragments includes the foreign DNA segment.

Repeat Sequence Methodology for Detecting Viral Integration Site

[00229] In various embodiments, an exemplary method for processing DNA can involve a repeat sequence methodology. In various embodiments, the repeat sequence methodology involves a two-step process in which a first step involves encapsulating and lysing a cell, followed by a second step involving amplification and barcoding of amplicons including foreign DNA sequences using a primer that targets a repeat sequence of the genomic DNA. In particular embodiments, the method includes using a foreign DNA segment-specific primer and a repeat sequence-specific primer. In such an embodiment, the foreign DNA segment-specific primer is hybridized to the foreign DNA segment, if present in the integration site, and the second primer is hybridized to a sequence present in the genomic DNA. The repeat sequence-specific primer may hybridize to a repeat sequence present in the genomic DNA.

[00230] In various embodiments, the methodology begins with encapsulating a single cell in a droplet, followed by lysing the cell within the droplet to generate a cell lysate. In various embodiments, a cell is lysed by lysing agents. For example, the lysing reagents can include a detergent such as NP-40 and/or a protease.

[00231] In various embodiments, the genomic DNA of the cell is contacted with reagents. In various embodiments, the genomic DNA of the cell is encapsulated in a droplet, such as a second droplet that differs from the droplet in which the cell was lysed. A reaction mixture can be generated by incubating the foreign DNA segment-specific primer (referred to as “vector specific primer”) and repeat sequence-specific primer under conditions to promote

hybridization of the foreign DNA segment-specific primer to a foreign DNA segment sequence, if present in the integration site and hybridization of the repeat sequence-specific primer to a repeat sequence present in the genomic DNA.

[00232] In various embodiments, such as the embodiment shown in **FIG. 4**, the vector specific primer may have a constant sequence that does not hybridize with a sequence of the foreign DNA. Here, the constant sequence may be useful for subsequently incorporating adapters, such as library sequencing adapters. In various embodiments, such as the embodiment shown in **FIG. 5**, the vector specific primer may only include a sequence that hybridizes with a sequence of the foreign DNA. Specifically, the vector specific primer shown in **FIG. 5** does not include the constant sequence shown in the right panel of **FIG. 4**. Thus, library sequencing adapters can be later incorporated in bulk (e.g., as shown below in **FIG. 5** as the “Illumina P7 adaptor”).

[00233] The foreign DNA segment-specific primer hybridizes to a sequence of the foreign DNA integrated into genomic DNA. Then, within the emulsion (e.g., droplet), the reaction mixture can be incubated under conditions to promote amplification of the genomic DNA integration site including the foreign DNA segment, if present, using the hybridized vector-specific primer to generate one or more amplicons including the integrated foreign DNA segment, if present. For example, DNA extension begins at the vector specific primer as indicated by the arrow to generate an extension product. Here, the extension product includes a sequence of the vector specific primer. Next, the extension product can be primed by a repeat sequence-specific primer (shown in **FIG. 4** as an “Alu primer,” though other repeat sequence-specific primers are described herein). The repeat sequence-specific primer can be linked to a constant region, such as a PCR handle (annotated as “const” in **FIG. 4** and **FIG. 5**). The PCR handle of the foreign DNA segment-specific primer is complementary to a PCR handle linked to a barcode primer including a barcode identification sequence (e.g., a bead barcode primer) sequence (annotated as “cell barcode” in **FIG. 4** and **FIG. 5**). Thus, synthesis can occur off of the repeat sequence-specific primer.

[00234] In various embodiments, the cell barcode can be directly linked to a sequence of a repeat sequence specific primer. For example, instead of hybridizing a constant region of the repeat sequence specific primer to a constant region of the cell barcode, as shown in **FIG. 4** and **FIG. 5**, the sequence of the repeat sequence specific primer can be directly linked to the cell barcode sequence.

[00235] In various embodiments, the amplified nucleic acid includes sequences of the first index sequence, the first read sequence, the barcode identification sequence, the constant region sequence, the repeat sequence-specific primer, the complement sequence of the foreign DNA segment, the second foreign DNA segment-specific primer, the second read sequence, and the second index sequence. In one scenario, the read 2 sequence can be included in the second PCR handle linked to the second (*e.g.*, reverse) primer sequence. In another scenario, the read 2 sequence can be included in the P7 sequence adapter.

[00236] Alternatively, in a third exemplary embodiment, a second primer is not utilized, such that the methods described above are adapted, though includes using the hybridized foreign DNA segment-specific primer to the foreign DNA segment, if present in the integration site.

Sequencing and Read Alignment

[00237] Amplified nucleic acids are sequenced to obtain sequence reads for generating a sequencing library. Sequence reads can be achieved with commercially available next generation sequencing (NGS) platforms, including platforms that perform any of sequencing by synthesis, sequencing by ligation, pyrosequencing, using reversible terminator chemistry, using phospholinked fluorescent nucleotides, or real-time sequencing. As an example, amplified nucleic acids may be sequenced on an Illumina MiSeq platform.

[00238] When pyrosequencing, libraries of NGS fragments are cloned *in situ* and amplified by capture of one matrix molecule using granules coated with oligonucleotides complementary to adapters. Each granule containing a matrix of the same type is placed in a microbubble of the “water-in-oil” type and the matrix is cloned amplified using a method called emulsion PCR. After amplification, the emulsion is destroyed and the granules are stacked in separate wells of a titration picoplate acting as a flow cell during sequencing reactions. The ordered multiple administration of each of the four dNTP reagents into the flow cell occurs in the presence of sequencing enzymes and a luminescent reporter, such as luciferase. In the case where a suitable dNTP is added to the 3' end of the sequencing primer, the resulting ATP produces a flash of luminescence within the well, which is recorded using a CCD camera. It is possible to achieve a read length of more than or equal to 400 bases, and it is possible to obtain 10^6 readings of the sequence, resulting in up to 500 million base pairs (megabytes) of the sequence. Additional details for pyrosequencing are described in Voelkerding et al., *Clinical Chem.*, 55: 641-658, 2009; MacLean et al., *Nature Rev.*

Microbiol., 7: 287-296; US Patent No. 6,210,891; US Patent No. 6,258,568; each of which is hereby incorporated by reference in its entirety.

[00239] On the Solexa/Illumina platform, sequencing data is produced in the form of short readings. In this method, fragments of a library of NGS fragments are captured on the surface of a flow cell that is coated with oligonucleotide anchor molecules. An anchor molecule is used as a PCR primer, but due to the length of the matrix and its proximity to other nearby anchor oligonucleotides, elongation by PCR leads to the formation of a “vault” of the molecule with its hybridization with the neighboring anchor oligonucleotide and the formation of a bridging structure on the surface of the flow cell. These DNA loops are denatured and cleaved. Straight chains are then sequenced using reversibly stained terminators. The nucleotides included in the sequence are determined by detecting fluorescence after inclusion, where each fluorescent and blocking agent is removed prior to the next dNTP addition cycle. Additional details for sequencing using the Illumina platform are found in Voelkerding et al., *Clinical Chem.*, 55: 641-658, 2009; MacLean et al., *Nature Rev. Microbiol.*, 7: 287-296; US Patent No. 6,833,246; US Patent No. 7,115,400; US Patent No. 6,969,488; each of which is hereby incorporated by reference in its entirety.

[00240] Sequencing of nucleic acid molecules using SOLiD technology includes clonal amplification of the library of NGS fragments using emulsion PCR. After that, the granules containing the matrix are immobilized on the derivatized surface of the glass flow cell and annealed with a primer complementary to the adapter oligonucleotide. However, instead of using the indicated primer for 3' extension, it is used to obtain a 5' phosphate group for ligation for test probes containing two probe-specific bases followed by 6 degenerate bases and one of four fluorescent labels. In the SOLiD system, test probes have 16 possible combinations of two bases at the 3' end of each probe and one of four fluorescent dyes at the 5' end. The color of the fluorescent dye and, thus, the identity of each probe, corresponds to a certain color space coding scheme. After many cycles of alignment of the probe, ligation of the probe and detection of a fluorescent signal, denaturation followed by a second sequencing cycle using a primer that is shifted by one base compared to the original primer. In this way, the sequence of the matrix can be reconstructed by calculation; matrix bases are checked twice, which leads to increased accuracy. Additional details for sequencing using SOLiD technology are found in Voelkerding et al., *Clinical Chem.*, 55: 641-658, 2009; MacLean et al., *Nature Rev. Microbiol.*, 7: 287-296; US Patent No. 5,912,148; US Patent No. 6,130,073; each of which is incorporated by reference in its entirety.

[00241] In particular embodiments, HeliScope from Helicos BioSciences is used.

Sequencing is achieved by the addition of polymerase and serial additions of fluorescently-labeled dNTP reagents. Switching on leads to the appearance of a fluorescent signal

corresponding to dNTP, and the specified signal is captured by a CCD camera before each

dNTP addition cycle. The reading length of the sequence varies from 25-50 nucleotides with a total yield exceeding 1 billion nucleotide pairs per analytical work cycle. Additional details for performing sequencing using HeliScope are found in Voelkerding et al., *Clinical Chem.*, 55: 641-658, 2009; MacLean et al., *Nature Rev. Microbiol.*, 7: 287-296; US Patent No.

7,169,560; US Patent No. 7,282,337; US Patent No. 7,482,120; US Patent No. 7,501,245; US Patent No. 6,818,395; US Patent No. 6,911,345; and US Patent No. 7,501,245; each of which is incorporated by reference in its entirety.

[00242] In various embodiments, a Roche sequencing system is used. Sequencing involves two-steps. In the first step, DNA is cut into fragments of approximately 300-800 base pairs, and these fragments have blunt ends. Oligonucleotide adapters are then ligated to the ends of

the fragments. The adapter serve as primers for amplification and sequencing of fragments.

Fragments can be attached to DNA-capture beads, for example, streptavidin-coated beads, using, for example, an adapter that contains a 5'-biotin tag. Fragments attached to the granules are amplified by PCR within the droplets (or a tube) of an oil-water emulsion. The result is multiple copies of cloned amplified DNA fragments on each bead. At the second stage, the granules are captured in wells (several picoliters in volume). Pyrosequencing is carried out on each DNA fragment in parallel. Adding one or more nucleotides leads to the generation of a light signal, which is recorded on a CCD camera of the sequencing instrument. The signal intensity is proportional to the number of nucleotides included.

Pyrosequencing uses pyrophosphate (PPi), which is released upon the addition of a

nucleotide. PPi is converted to ATP using ATP sulfurylase in the presence of adenosine 5' phosphosulfate. Luciferase uses ATP to convert luciferin to oxyluciferin, and as a result of this reaction, light is generated that is detected and analyzed. Additional details for performing sequencing is found in Margulies et al. (2005) *Nature* 437: 376-380, which is hereby incorporated by reference in its entirety.

[00243] In various embodiments, PCR methods used may include sequence-specific PCR, foreign DNA-specific PCR, or linear amplification PCR.

[00244] Ion Torrent technology is a DNA sequencing method based on the detection of hydrogen ions that are released during DNA polymerization. The microwell contains a

fragment of a library of NGS fragments to be sequenced. Under the microwell layer is the hypersensitive ion sensor ISFET. All layers are contained within a semiconductor CMOS chip, similar to the chip used in the electronics industry. When a dNTP is incorporated into a growing complementary chain, a hydrogen ion is released that excites a hypersensitive ion sensor. If homopolymer repeats are present in the sequence of the template, multiple dNTP molecules will be included in one cycle. This results in a corresponding amount of hydrogen atoms being released and in proportion to a higher electrical signal. This technology is different from other sequencing technologies that do not use modified nucleotides or optical devices. Additional details for Ion Torrent Technology are found in *Science* 327 (5970): 1190 (2010); U.S Patent Application Publication Nos. 2009/0026082, 2009/0127589, 2010/0301398, 2010/0197507, 2010/0188073, and 2010/0137143, each of which is incorporated by reference in its entirety.

[00245] In various embodiments, sequencing reads obtained from the NGS methods can be filtered by quality and grouped by barcode sequence using any algorithms known in the art, *e.g.*, Python script barcodeCleanup.py. In various embodiments, a given sequencing read may be discarded if more than about 20% of its bases have a quality score (Q-score) less than Q20, indicating a base call accuracy of about 99%. In various embodiments, a given sequencing read may be discarded if more than about 5%, about 10%, about 15%, about 20%, about 25%, about 30% have a Q-score less than Q10, Q20, Q30, Q40, Q50, Q60, or more, indicating a base call accuracy of about 90%, about 99%, about 99.9%, about 99.99%, about 99.999%, about 99.9999%, or more, respectively.

[00246] In various embodiments, all sequencing reads associated with a barcode containing less than 50 reads may be discarded to ensure that all barcode groups, representing single cells, contain a sufficient number of high-quality reads. In various embodiments, all sequencing reads associated with a barcode containing less than 30, less than 40, less than 50, less than 60, less than 70, less than 80, less than 90, less than 100 or more may be discarded to ensure the quality of the barcode groups representing single cells.

[00247] Sequence reads with common barcode sequences (*e.g.*, meaning that sequence reads originated from the same cell) may be aligned to a reference genome using known methods in the art to determine alignment position information. The alignment position information may indicate a beginning position and an end position of a region in the reference genome that corresponds to a beginning nucleotide base and end nucleotide base of a given sequence read. A region in the reference genome may be associated with a target

gene or a segment of a gene. Exemplary aligner algorithms include BWA, Bowtie, Spliced Transcripts Alignment to a Reference (STAR), Tophat, and HISAT2. Further details for aligning sequence reads to reference sequences are described in US Application No.

16/279,315, which is hereby incorporated by reference in its entirety. In various

5 embodiments, an output file having a sequence alignment map (SAM) format or binary alignment map (BAM) format may be generated and output for subsequent analysis, such as for determining cell trajectory.

[00248] Sequencing may be performed to determine the length of a nucleic acid (*e.g.*, an amplicon). Analysis of size of a nucleic acid may also be performed to identify the genomic
10 locus of one or more integration sites (*e.g.*, an integration site of foreign DNA into genomic DNA). For example, in various embodiments, the disclosure provides a method in which sequencing generated amplicons further includes characterizing a number of integration sites in the genomic DNA or a number of copies of the foreign DNA segment (*e.g.*, vector copy number).

15 **[00249]** Sequencing may also be analyzed to identify the amplicon identity (*e.g.*, unique reads rather than PCR duplicates), the genomic locus of the integration site, the number of integration sites, or the orientation of the integration, optionally in which the number of integration sites includes the vector copy number. For example, in various embodiments, the one or more amplicons sequence and/or the one or more amplicons size is analyzed to
20 identify the amplicon identity, such as unique reads, rather than PCR duplicates. In various embodiments, the one or more amplicons sequence and/or the one or more amplicons size is analyzed to identify the genomic locus of the integration site. In various embodiments, the one or more amplicons sequence and/or the one or more amplicons size is analyzed to identify the number of integration sites (*e.g.*, vector copy number). In various embodiments,
25 the one or more amplicons sequence and/or the one or more amplicons size is analyzed to identify the orientation of the integration.

[00250] In various embodiments, using the sequenced amplicons, the unique number of genome and vector integration sites can be counted to determine the vector copy number. In such instances, the number of unique genomic coordinates identified determines the vector
30 copy number per cell. Alternatively, or additionally, in various embodiments, using the sequenced amplicons, the unique Tn5 insertion sites on the foreign DNA segment, if it exists, can be counted. In such instances, when overlapping sequences of the foreign DNA segment exist, that count can be used to determine the vector copy number. For example, by assessing

the range of unique Tn5 insertion sites on a foreign DNA segment, the vector copy number per cell can be estimated based upon overlaying regions.

[00251] In various embodiments, the method further includes determining a vector copy number of the foreign DNA segment across the integration site and the one or more additional integration sites.

[00252] In various embodiments, determining the vector copy number includes: identifying a first amplicon including a sequence of the foreign DNA segment and a second amplicon including a sequence of the foreign DNA segment, wherein the first amplicon and the second amplicon include different start sites; and determining whether a portion of the sequence of the foreign DNA segment of the first amplicon overlaps with a portion of the sequence of the foreign DNA segment of the second amplicon. In various embodiments, the different start sites of the first amplicon and the second amplicon correspond to different Tn5 insertion sites. In various embodiments, the first amplicon and second amplicon share a common termination site.

[00253] In various embodiments, the common termination sites of the first amplicon and second amplicon correspond to the foreign DNA segment-specific primer.

[00254] In various embodiments, responsive to the determination that the portion of the sequence of the foreign DNA segment of the first amplicon overlaps with a portion of the sequence of the foreign DNA segment of the second amplicon, determining that the vector copy number is at least 2. In various embodiments, responsive to the determination that the portion of the sequence of the foreign DNA segment of the first amplicon does not overlap with a portion of the sequence of the foreign DNA segment of the second amplicon, determining that the vector copy number is 1.

[00255] In various embodiments, in a single cell, there may be 1, 2, 3, 4, 5, or more integration sites, which can be determined by i) counting the unique number of genome and vector integration sites and/or ii) counting the number of overlapping sequences of the foreign DNA segment that exist in the one or more amplicons).

[00256] In various embodiments, using sequencing, the cellular genotype and cellular phenotype of the cell is used to identify cellular subpopulations. For example, the cell can be derived from a population of cells. In such embodiments, the cellular genotype and cellular phenotype of the cell is analyzed in conjunction with cellular genotypes and cellular phenotypes of other cells derived from the population of cells. In various embodiments, analyzing the cellular genotypes and cellular phenotypes of the population of cells involves

performing one or both of a dimensional reduction analysis and a clustering analysis, such that cells with similar genotypes or phenotypes are localized within clusters. In various embodiments, heterogeneous subpopulations of cells can be identified from individual clusters. In various embodiments, heterogeneous subpopulations of cells can be identified from even within the clusters themselves.

[00257] Identifying subpopulations of cells with differing combinations of genotypes and phenotypes can be useful for discovering subpopulations of cells in cell populations. As one example, a subpopulation of cells can refer to a diseased (e.g., cancer) cell subpopulation. Thus, detection and/or identification of the presence of a diseased cell subpopulation is useful for diagnosing a subject with said disease. As another example, the population of cells may be a population of diseased cells previously thought to be homogeneous. Thus, analyzing the cellular genotypes and phenotypes of cells in the diseased cells is helpful in understanding the heterogeneity of the diseased cells, which can be used to guide the development or selection of treatments for targeting the various subpopulations of cells.

[00258] Following sequencing, it may be determined that a sequenced nucleic acid (e.g., an amplicon) includes from 5'-to-3': a first index sequence, a first read sequence, a barcode identification sequence, a constant region sequence, a foreign DNA segment-specific primer, a complement sequence of a foreign DNA segment, a second foreign DNA segment-specific primer, a second read sequence, and a second index sequence.

[00259] Alternatively, for example, following sequencing, it may be determined that a sequenced nucleic acid (e.g., an amplicon) includes from 5'-to-3': a first index sequence, a barcode identification sequence, a constant region sequence, a foreign DNA segment-specific primer, a complement sequence of a foreign DNA segment, a second foreign DNA segment-specific primer, and a second index sequence.

[00260] Alternatively, for example, following sequencing, it may be determined that a sequenced nucleic acid (e.g., an amplicon) includes 5'-to-3': a first index sequence, a first read sequence, a barcode identification sequence, a constant region sequence, a repeat sequence-specific primer, a complement sequence of a foreign DNA segment, a second foreign DNA segment-specific primer, a second read sequence, and a second index sequence.

Cellular Genotype and Phenotype

[00261] Sequencing reads of nucleic acids derived from genomic DNA can be analyzed to determine cellular phenotypes and cellular genotypes.

[00262] In various embodiments, determining a cell genotype refers to determining one or more nucleotides or sequences that are present in the genome of the cell. For example, determining a cell genotype can refer to determining presence or absence of a sequence of foreign DNA. As another example, determining a cell genotype can refer to determining one or more mutations in the genome of the cell. In particular embodiments, the Tapestri® Insights software is implemented to identify the one or more mutations in the genome of the cell. In various embodiments, the one or more mutations include single nucleotide changes (*e.g.*, SNVs) or short sequences of nucleotide changes (*e.g.*, short indels). Here, aligned sequence reads derived from genomic DNA of the cell are analyzed against the reference genome to determine differences between likely nucleotide bases present in the cell mutations corresponding nucleotide bases present in the reference genome. In various embodiments, identifying SNVs and/or short indels can be accomplished by implementing any publicly available SNV caller algorithms including, but not limited to:

BWA <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5852328/-/bb0015>, NovoAlign, Torrent Mapping Alignment Program (TMAP), VarScan2, qSNP, Shimmer, RADIA, SOAPSnv, VarDict, SNVMix2, SPLINTER, SNVer, OutLyzer, Pisces, ISOWN, SomVarIUS, and SiNVICT.

[00263] In various embodiments, the one or more mutations include structural variants such as CNVs and/or mutations that encompass long sequences (*e.g.*, long indels). Here, split-reads and *de novo* assembly methods can be used to identify CNVs and/or longer indels. In various embodiments, the CNV caller workflow involves one or more of the following steps: binning, GC content correction, mappability correction, removal of outlier bins, removal of outlier cells, segmentation, and calling of absolute numbers. Further details of CNV caller workflows are described in Fan, X. et al, Methods for Copy Number Aberration Detection from Single-cell DNA Sequencing Data, *bioRxiv* 696179, which is hereby incorporated by reference in its entirety. In various embodiments, identifying CNVs and/or long indels can be accomplished by implementing any publicly available CNV caller including, but not limited to: HMMcopy, SeqSeg, CNV-seq, rSW-seq, FREEC, CNASeg, ReadDepth, CNVator, seqCBS, seqCNA, m-HMM, Ginkgo, nbCNV, AneuFinder, SCNV, and CNV IFTV.

[00264] In various embodiments, sequence reads are pre-processed prior to their use in identifying one or more mutations of the cell genome. For example, reads from a cell are normalized by the cell's total read count and grouped by hierarchical clustering based on amplicon read distribution. Amplicon counts from the cell is divided by the median of the

corresponding amplicons from a control group (*e.g.*, a control cell cluster with known CNVs). Thus, normalized percentage of sequencing reads were used to calculate CNVs for each gene.

[00265] In various embodiments, sequence reads used to determine the cellular genotype can be derived from various regions of a cell genome. These regions of the cell genome include both coding regions and non-coding regions (*e.g.*, introns, regulatory elements, transcription factor binding sites, and chromosomal translocation junctions). Therefore, one or more mutations (*e.g.*, SNVs, CNVs, and indels) can be identified in both coding and non-coding regions. The single-cell workflow analysis detailed above that directly determines cellular genotypes from genomic DNA enables the identification of mutations from both coding and non-coding regions, whereas less direct methods (*e.g.*, those that reverse transcribe RNA) only identify mutations from coding regions.

[00266] To determine a cell phenotype, sequence reads derived from antibody-conjugated oligonucleotides are analyzed. Specifically, the sequence of the antibody tag of the antibody oligonucleotide is sequenced. The presence of the sequence read indicates that the corresponding antibody (on which the oligonucleotide was conjugated) had previously been bound to an analyte of the cell. In other words, the presence of the sequence read indicates that the cell expressed the target analyte.

[00267] In various embodiments, determining a cell phenotype involves quantifying a level of expression of a target analyte. In various embodiments, quantifying a level of expression of a target analyte involves normalizing the sequence reads derived from antibody-conjugated oligonucleotides. In various embodiments, normalizing the sequence reads involves performing a centered log ratio (CLR) transformation. In various embodiments, normalizing the sequence reads involves performing Denoised and Scaled by Background (DSB). Additional description of DSB normalization is found in Mulè, M. et al. "Normalizing and denoising protein expression data from droplet-based single cell profiling." *bioRxiv* 2020.02.24.963603, which is hereby incorporated by reference in its entirety.

[00268] In various embodiments, a cell phenotype can refer to the cell expression of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 100, 500, 1000, 5000, or 10,000 target analytes. Therefore, the single-cell workflow analysis can yield an expression profile for a plurality of target analytes of a cell.

[00269] In various embodiments, the genotype and the phenotype of the cell can be used to classify the cell. For example, the cell can be classified within a population of cells that share at least the genotype, share at least the phenotype, or share at least both the genotype and the phenotype of the cell. In various embodiments, the single-cell workflow analysis is conducted on each cell in a population of cells. Therefore, the cell genotype and cell phenotype of each cell in the population can be used to classify each cell to gain an understanding as to the distribution of cells in the population. In various embodiments, the classified cells provide insight as to the subpopulations that are present. In various embodiments, classifying a cell involves comparing the genotype and phenotype of the cell against a library of known cell populations that are characterized by known genotypes and phenotypes. Therefore, if the cell shares a genotype, shares a phenotype, or shares both a genotype and phenotype with a known cell population, the cell can be classified in a category of the known cell population.

[00270] To provide an example, the population of cells can be obtained from a subject suspected of having cancer, each cell in the population can be analyzed using the single-cell workflow to determine each cell's genotype and phenotype. Cells are classified according to their genotypes and phenotypes by comparing to genotypes and phenotypes of known reference cells. Thus, classifying cells in the population using their genotypes and phenotypes reveals a distribution of cells which can guide the selection of a cancer treatment for the subject. For example, if a large proportion of cells in the population are classified with a known cell population that are known to be resistant to particular therapies, then alternative therapies that are more likely to be efficacious can be selected for treating the cancer.

[00271] In various embodiments, the genotype and the phenotype of the cell are used to identify subpopulations within a population of cells. This is useful for discovering new subpopulations that were not previously known. For example, a cell population previously thought be homogeneous can be analyzed to reveal multiple subpopulations of cells with different genotype and phenotype combinations. In various embodiments, a cell population may reveal two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, thirteen, fourteen, fifteen, sixteen, seventeen, eighteen, nineteen, or twenty different subpopulations.

[00272] In various embodiments, the single-cell workflow analysis is conducted on each cell in a population of cells and the cell genotypes and cell phenotypes of cells in the population are used to identify subpopulations of cells that are characterized by genotypes and phenotypes. In various embodiments, using the genotypes and phenotypes of the cells to identify subpopulations involves performing a dimensionality reduction analysis. In various

embodiments, using the genotypes and phenotypes of the cells to identify subpopulations involves performing an unsupervised clustering analysis. In various embodiments, using the genotypes and phenotypes of the cells to identify subpopulations involves performing a dimensionality reduction analysis and an unsupervised clustering analysis.

5 **[00273]** Examples of unsupervised cluster analysis include hierarchical clustering, k-means clustering, clustering using mixture models, density based spatial clustering of applications with noise (DBSCAN), ordering points to identify the clustering structure (OPTICS), or combinations thereof. Examples of dimensionality reduction analysis include principal component analysis (PCA), kernel PCA, graph-based kernel PCA, linear
10 discriminant analysis, generalized discriminant analysis, autoencoder, non-negative matrix factorization, T-distributed stochastic neighbor embedding (t-SNE), or uniform manifold approximation and projection (UMAP) and dens-UMAP.

[00274] In particular embodiments, a dimensionality reduction analysis and unsupervised clustering is performed on at least one of either cellular genotypes or cellular phenotypes of
15 cells in the population. Thus, clusters of cells are generated according to at least one of either cellular genotypes or cellular phenotypes of the cells. In particular embodiments, clusters of cells are generated according to detected SNVs for one or more genes. In particular embodiments, clusters of cells are generated according to detected SNVs for two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, thirteen, fourteen, fifteen, sixteen, seventeen,
20 eighteen, nineteen, twenty, twenty five, thirty, forty, fifty, sixty, seventy, eighty, ninety, or one hundred genes. In particular embodiments, clusters of cells are generated according to detected CNVs for one or more genes. In particular embodiments, clusters of cells are generated according to detected CNVs for two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, thirteen, fourteen, fifteen, sixteen, seventeen, eighteen, nineteen, twenty,
25 twenty five, thirty, forty, fifty, sixty, seventy, eighty, ninety, or one hundred genes. In particular embodiments, clusters of cells are generated according to levels of analyte expression for one or more analytes. In particular embodiments, clusters of cells are generated according to levels of analyte expression for two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, thirteen, fourteen, fifteen, sixteen, seventeen, eighteen, nineteen,
30 twenty, twenty five, thirty, forty, fifty, sixty, seventy, eighty, ninety, or one hundred analytes.

[00275] In various embodiments individual cells in clusters are labeled using the other of the cellular genotypes or cellular phenotypes to reveal any subpopulations of cells either within clusters or across the clusters. As one example, cellular phenotypes (*e.g.*, analyte

expression) can be used to generate clusters of cells and cellular genotypes (*e.g.*, mutations) are used to label cells in the clusters. As another example, cellular genotypes are used to generate clusters of cells and cellular phenotypes are used to label cells in the clusters.

[00276] To provide a specific example, dimensionality reduction analysis and

5 unsupervised clustering is performed on cellular phenotypes of cells. Specifically, dimensionality reduction analysis can be performed on normalized sequence read values (*e.g.*, CLR values) derived from antibody oligonucleotides. Then, unsupervised clustering is performed on the CLR normalized sequence read values in the dimensionally reduced space to generate clusters of cells. Here, cells that have similar analyte expression profiles may be
10 clustered in a common cluster whereas cells that have dissimilar analyte expression profiles may be clustered in different clusters. Cellular genotypes of the cells can be used to label individual cells within clusters. For example, individual cells within clusters can be labeled as having a particular mutation (*e.g.*, a particular SNV on a gene or an increase/decrease in copy number for a particular gene). In some scenarios, individual cells within clusters can be
15 labeled as having more than one mutation (*e.g.*, SNVs on one or more genes or increase/decrease in copy number of one or more genes).

[00277] As another example, a dimensionality reduction analysis and unsupervised clustering is performed on cellular genotypes of cells. Specifically, dimensionality reduction analysis can be performed according to mutations (*e.g.*, SNVs and/or CNVs) of one or more
20 genes identified within the cells. Then, unsupervised clustering is performed in the dimensionally reduced space to generate clusters of cells. Here, cells that have similar genotypes (*e.g.*, mutations of one or more genes) may be clustered in a common cluster whereas cells that have dissimilar genotypes may be clustered in different clusters. Cellular phenotypes of the cells can be used to label individual cells within clusters. For example,
25 individual cells within clusters can be labeled as expressing or not expressing a particular analyte. In some scenarios, individual cells within clusters can be labeled as expressing more than one analyte or not expressing more than one analyte.

[00278] In various embodiments, a dimensionality reduction analysis and unsupervised clustering is performed on both cellular genotypes and cellular phenotypes of cells. Here,
30 cells that have similar genotypes (*e.g.*, mutations of one or more genes) and phenotypes may be clustered in a common cluster whereas cells that have dissimilar genotypes and phenotypes may be clustered in different clusters.

[00279] Analyzing the labeled clusters of cells can, in some scenarios, reveal subpopulations of cells that have particular combinations of genotypes (*e.g.*, mutations) and phenotypes (*e.g.*, analyte expression). In various embodiments, a subpopulation of cells can refer to a cluster of cells that have a common phenotype and common genotype. For example, a subpopulation of cells can refer to a cluster of cells that express an analyte and have a SNV at a particular position of a gene. As another example, a subpopulation of cells can refer to a cluster of cells that do not an analyte and have an increased copy number of a gene. Any combination of cellular phenotype (*e.g.*, expression or lack of expression of an analyte) and cellular genotype (*e.g.*, presence or absence of one or more SNVs or increase/decrease in copy number of a gene) of a cluster of cells can be identified as a subpopulation.

Targeted Panels

[00280] Embodiments disclosed herein include targeted DNA libraries for interrogating one or more genes as well as targeted protein libraries for interrogating expression and/or expression levels of one or more proteins.

[00281] In various embodiments, the targeted gene panel includes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, or 1000 genes. In various embodiments, the targeted protein panel includes at least 1, at least 2, at least 5, at least 10, at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 200, at least 300, at least 400, at least 500, or at least 1000 genes.

[00282] In various embodiments, the targeted protein panel includes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, or 1000 proteins. In various embodiments, the targeted protein panel includes at least 1, at least 2, at least 5, at least 10, at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 200, at least 300, at least 400, at least 500, or at least 1000 proteins.

[00283] In various embodiments, the targeted protein panel includes one or more proteins of HLA-DR, CD10, CD117, CD11b, CD123, CD13, CD138, CD14, CD141, CD15, CD16, CD163, CD19, CD193 (CCR3), CD1c, CD2, CD203c, CD209, CD22, CD25, CD3, CD30, CD303, CD304, CD33, CD34, CD4, CD42b, CD45RA, CD5, CD56, CD62P (P-Selectin), CD64, CD68, CD69, CD38, CD7, CD71, CD83, CD90 (Thy1), Fc epsilon RI alpha, Siglec-8, CD235a, CD49d, CD45, CD8, CD45RO, mouse IgG1, kappa, mouse IgG2a, kappa, mouse

IgG2b, kappa, CD103, CD62L, CD11c, CD44, CD27, CD81, CD319 (SLAMF7), CD269 (BCMA), CD99, CD164, KCNJ3, CXCR4 (CD184), CD109, CD53, CD74, HLA-DR, DP, DQ, HLA-A, B, C, ROR1, Annexin A1, and CD20.

5 Methods for the Delivery of Foreign Nucleic Acids to Target Cells

[00284] A target cell can be transduced with foreign DNA provided herein. For example, the disclosure herein provides transduction of a cell or a population of cells with foreign DNA, including viral DNA, modified viral DNA, or a viral vector, such as with the methods Provided herein.

10 [00285] Techniques that can be used to introduce a nucleic acid molecule into a mammalian cell are well known in the art. For example, electroporation can be used to permeabilize mammalian cells (*e.g.*, human target cells) by the application of an electrostatic potential to the cell of interest. Mammalian cells, such as human cells, subjected to an external electric field in this manner are subsequently predisposed to the uptake of foreign
15 nucleic acids. Electroporation of mammalian cells is described in detail, *e.g.*, in Chu et al., *Nucleic Acids Research* 15:1311 (1987), the disclosure of which is incorporated herein by reference. A similar technique, Nucleofection™, utilizes an applied electric field in order to stimulate the uptake of foreign polynucleotides into the nucleus of a eukaryotic cell.

[00286] Nucleofection™ and protocols useful for performing this technique are described
20 in detail, *e.g.*, in Distler et al., *Experimental Dermatology* 14:315 (2005), as well as in US 2010/0317114, the disclosures of each of which are incorporated herein by reference.

[00287] Additional techniques useful for the transfection of target cells are the squeeze-poration methodology. This technique induces the rapid mechanical deformation of cells in order to stimulate the uptake of foreign DNA through membranous pores that form in
25 response to the applied stress. This technology is advantageous in that a vector is not included for delivery of nucleic acids into a cell, such as a human target cell. Squeeze-poration is described in detail, *e.g.*, in Sharei et al., *JoVE* 81:e50980 (2013), the disclosure of which is incorporated herein by reference.

[00288] Lipofection represents another technique useful for transfection of target cells.
30 This method involves the loading of nucleic acids into a liposome, which often presents cationic functional groups, such as quaternary or protonated amines, towards the liposome exterior. This promotes electrostatic interactions between the liposome and a cell due to the anionic nature of the cell membrane, which ultimately leads to uptake of the foreign nucleic

acids, for example, by direct fusion of the liposome with the cell membrane or by endocytosis of the complex. Lipofection is described in detail, for example, in US 7,442,386, the disclosure of which is incorporated herein by reference.

[00289] Similar techniques that exploit ionic interactions with the cell membrane to provoke the uptake of foreign nucleic acids are contacting a cell with a cationic polymer-nucleic acid complex. Exemplary cationic molecules that associate with polynucleotides so as to impart a positive charge favorable for interaction with the cell membrane are activated dendrimers (described, *e.g.*, in Dennig, *Topics in Current Chemistry* 228:227 (2003), the disclosure of which is incorporated herein by reference) polyethylenimine, and diethylaminoethyl (DEAE)-dextran, the use of which as a transfection agent is described in detail, for example, in Gulick et al., *Current Protocols in Molecular Biology* 40:1:9.2:9.2.1 (1997), the disclosure of which is incorporated herein by reference. Magnetic beads are another tool that can be used to transfect target cells in a mild and efficient manner, as this methodology utilizes an applied magnetic field in order to direct the uptake of nucleic acids. This technology is described in detail, for example, in US 2010/0227406, the disclosure of which is incorporated herein by reference.

[00290] Another useful tool for inducing the uptake of foreign nucleic acids by target cells is laserfection, also called optical transfection, a technique that involves exposing a cell to electromagnetic radiation of a particular wavelength in order to gently permeabilize the cells and allow polynucleotides to penetrate the cell membrane. The bioactivity of this technique is similar to, and in some cases found superior to, electroporation.

[00291] Impalefection is another technique that can be used to deliver genetic material to target cells. It relies on the use of nanomaterials, such as carbon nanofibers, carbon nanotubes, and nanowires.

[00292] Needle-like nanostructures are synthesized perpendicular to the surface of a substrate. DNA containing the gene, intended for intracellular delivery, is attached to the nanostructure surface. A chip with arrays of these needles is then pressed against cells or tissue. Cells that are impaled by nanostructures can express the delivered gene(s). An example of this technique is described in Shalek et al., *PNAS* 107:1870 (2010), the disclosure of which is incorporated herein by reference.

[00293] Magnetofection can also be used to deliver nucleic acids to target cells. The magnetofection principle is to associate nucleic acids with cationic magnetic nanoparticles. The magnetic nanoparticles are made of iron oxide, which is fully biodegradable, and coated

with specific cationic proprietary molecules varying upon the applications. Their association with the gene vectors (DNA, viral vector) is achieved by salt-induced colloidal aggregation and electrostatic interaction. The magnetic particles are then concentrated on the target cells by the influence of an external magnetic field generated by magnets. This technique is described in detail in Scherer et al., *Gene Therapy* 9:102 (2002), the disclosure of which is incorporated herein by reference.

[00294] Another useful tool for inducing the uptake of foreign nucleic acids by target cells is sonoporation, a technique that involves the use of sound (such as ultrasonic frequencies) for modifying the permeability of the cell plasma membrane. This technique is described in detail, e.g., in Rhodes et al., *Methods in Cell Biology* 82:309 (2007), the disclosure of which is incorporated herein by reference.

[00295] Microvesicles represent another potential vehicle that can be used to modify the genome of a target cell according to the methods described herein. For example, microvesicles that have been induced by the co-overexpression of the glycoprotein VSV-G with, e.g., a genome-modifying protein, such as a nuclease, can be used to efficiently deliver proteins into a cell that subsequently catalyze the site-specific cleavage of an endogenous polynucleotide sequence so as to prepare the genome of the cell for the covalent incorporation of a polynucleotide of interest, such as a gene or regulatory sequence. The use of such vesicles, also referred to as Gesicles, for the genetic modification of eukaryotic cells is described in detail, e.g., in Quinn et al., Genetic Modification of Target Cells by Direct Delivery of Active Protein [abstract]. In: Methylation changes in early embryonic genes in cancer [abstract], in: Proceedings of the 18th Annual Meeting of the American Society of Gene and Cell Therapy; 2015 May 13, Abstract No. 122.

Nucleic Acid Vectors

[00296] Effective intracellular concentrations of foreign DNA encoding a gene (e.g., a transgene encoding a protein of interest or a reporter gene) disclosed herein can be achieved via the stable expression of a vector encoding a coding sequence (e.g., by integration into the nuclear or mitochondrial genome of a mammalian cell). In order to introduce such a gene into a mammalian cell, the gene can be incorporated into a vector.

[00297] Vectors can be introduced into a cell by a variety of methods, including transformation, transfection, direct uptake, projectile bombardment, and by encapsulation of

the vector in a liposome. Examples of suitable methods of transfecting or transforming cells are calcium phosphate precipitation, electroporation, microinjection, infection, lipofection, and direct uptake. Such methods are described in more detail, for example, in Green et al., *Molecular Cloning: A Laboratory Manual*, Fourth Edition (Cold Spring Harbor University Press, New York (2014)); and Ausubel et al., *Current Protocols in Molecular Biology* (John Wiley & Sons, New York (2015)), the disclosures of each of which are incorporated herein by reference.

[00298] The genes disclosed herein can also be introduced into a mammalian cell by targeting a vector containing a polynucleotide encoding such a gene to cell membrane phospholipids. For example, vectors can be targeted to the phospholipids on the extracellular surface of the cell membrane by linking the vector molecule to a VSV-G protein, a viral protein with affinity for all cell membrane phospholipids. Such, a construct can be produced using conventional and routine methods of the art. In addition to achieving high rates of transcription and translation, stable expression of an foreign polynucleotide in a mammalian cell can be achieved by integration of the polynucleotide containing the gene into the nuclear genome of the mammalian cell. A variety of vectors for the delivery and integration of polynucleotides encoding foreign proteins into the nuclear DNA of a mammalian cell have been developed. Examples of expression vectors are disclosed in, *e.g.*, WO 1994/011026 and are incorporated herein by reference. Expression vectors for use in the compositions and methods described herein contain a polynucleotide sequence that encodes a gene as well as, *e.g.*, additional sequence elements used for the expression of these genes and/or the integration of these polynucleotide sequences into the genome of a mammalian cell. Certain vectors that can be used include plasmids that contain regulatory sequences, such as promoter and enhancer regions, which direct gene transcription. Other useful vectors contain polynucleotide sequences that enhance the rate of translation of these genes or improve the stability or nuclear export of the mRNA that results from gene transcription. These sequence elements include, *e.g.*, 5' and 3' UTR regions, an internal ribosomal entry site (IRES), and polyA in order to direct efficient transcription of the gene carried on the expression vector. The expression vectors suitable for use with the compositions and methods described herein may also contain a polynucleotide encoding a marker for selection of cells that contain such a vector. Examples of a suitable marker are genes that encode resistance to antibiotics, such as ampicillin, chloramphenicol, kanamycin, nourseothricin.

Viral Vectors

[00299] Viral genomes provide a rich source of vectors that can be used for the efficient delivery of foreign DNA (*e.g.*, a foreign DNA segment) into a mammalian cell. Viral genomes are particularly useful vectors for gene delivery as the polynucleotides contained within such genomes are, in various embodiments, incorporated into the nuclear genome of a mammalian cell by generalized or specialized transduction. These processes occur as part of the natural viral replication cycle, and do not require added proteins or reagents in order to induce gene integration. Examples of viral vectors are a parvovirus (*e.g.*, adeno-associated viruses (AAV)), retrovirus (*e.g.*, Retroviridae family viral vector), adenovirus (*e.g.*, Ad5, Ad26, Ad34, Ad35, and Ad48), coronavirus, negative strand RNA viruses such as orthomyxovirus (*e.g.*, influenza virus), rhabdovirus (*e.g.*, rabies and vesicular stomatitis virus), paramyxovirus (*e.g.* measles and Sendai), positive strand RNA viruses, such as picornavirus and alphavirus, and double-stranded DNA viruses including adenovirus, herpesvirus (*e.g.*, Herpes Simplex virus types 1 and 2, Epstein-Barr virus, cytomegalovirus), and poxvirus (*e.g.*, vaccinia, modified vaccinia Ankara (MVA), fowlpox and canarypox). Other viruses include Norwalk virus, togavirus, flavivirus, reoviruses, papovavirus, hepadnavirus, human papilloma virus, human foamy virus, and hepatitis virus, for example. Examples of retroviruses are avian leukosis-sarcoma, avian C-type viruses, mammalian C-type, B-type viruses, D-type viruses, oncoretroviruses, HTLV-BLV group, lentivirus (*e.g.*, HIV), alpharetrovirus, gammaretrovirus, spumavirus (Coffin, J. M., *Retroviridae: The viruses and their replication*, Virology, Third Edition (Lippincott-Raven, Philadelphia, (1996))). Other examples are murine leukemia viruses, murine sarcoma viruses, murine mammary tumor virus, bovine leukemia virus, feline leukemia virus, feline sarcoma virus, avian leukemia virus, human T-cell leukemia virus, baboon endogenous virus, Gibbon ape leukemia virus, Pfizer monkey virus, simian immunodeficiency virus, simian sarcoma virus, Rous sarcoma virus and lentiviruses (*e.g.*, HIV). Other examples of vectors are described, for example, in McVey et al., (U.S. Patent No. 5,801,030), the teachings of which are incorporated herein by reference.

[00300] In various embodiments, the viral DNA, modified viral DNA, or viral vector of the disclosure is derived from an AAV, adenovirus, herpes simplex virus, lentivirus (*e.g.*, HIV), retrovirus, poxvirus, baculovirus, or vaccinia virus.

[00301] For example, in various embodiments, a foreign DNA segment disclosed herein may include an inverted terminal repeat region (ITR), a rep gene, a cap gene, a long terminal

repeat (LTR) region, a gag gene, a pol gene, a tat gene, a rev gene, a IX gene, a IVa2 gene, an L1 gene, an L2 gene, an L3 gene, an L4 gene, an L5 gene, an E2B gene, an E2A gene, an E2A-L gene, an E4 gene, a gene encoding a capsomer protein, a gene encoding a capsid protein, a gene encoding a core protein, a gene encoding a viral non-structural protein, or a gene encoding a viral packing protein.

[00302] In various embodiments, the foreign DNA segment includes an LTR.

[00303] Alternatively, or in addition the above, in various embodiments, a foreign DNA segment disclosed herein may include a transgene encoding a protein of interest or a reporter gene. For example, in various embodiments, DNA from a viral vector includes a transgene encoding a protein of interest. In some DNA from a viral vector includes a reporter gene.

Incorporation of target genes by gene editing techniques

[00304] In addition to the above, a variety of tools have been developed that can be used for the incorporation of foreign DNA into a target cell. One such method that can be used for incorporating polynucleotides encoding target genes into target cells involves the use of transposons. Transposons are polynucleotides that encode transposase enzymes and contain a polynucleotide sequence or gene of interest flanked by 5' and 3' excision sites. Once a transposon has been delivered into a cell, expression of the transposase gene commences and results in active enzymes that cleave the gene of interest from the transposon. This activity is mediated by the site-specific recognition of transposon excision sites by the transposase. In some instances, these excision sites may be terminal repeats or inverted terminal repeats. Once excised from the transposon, the gene of interest can be integrated into the genome of a mammalian cell by transposase-catalyzed cleavage of similar excision sites that exist within the nuclear genome of the cell. This allows the gene of interest to be inserted into the cleaved nuclear DNA at the complementary excision sites, and subsequent covalent ligation of the phosphodiester bonds that join the gene of interest to the DNA of the mammalian cell genome completes the incorporation process. In certain cases, the transposon may be a retrotransposon, such that the gene encoding the target gene is first transcribed to an RNA product and then reverse-transcribed to DNA before incorporation in the mammalian cell genome. Exemplary transposon systems are the piggybac transposon (described in detail in, e.g., WO 2010/085699) and the sleeping beauty transposon (described in detail in, e.g., US 2005/0112764), the disclosures of each of which are incorporated herein by reference as they pertain to transposons for use in gene delivery to a cell of interest

[00305] Another tool for the integration of target genes into the genome of a target cell is the clustered regularly interspaced short palindromic repeats (CRISPR)/Cas system, a system that originally evolved as an adaptive defense mechanism in bacteria and archaea against viral infection. The CRISPR/Cas system includes palindromic repeat sequences within plasmid DNA and an associated Cas9 nuclease. This ensemble of DNA and protein directs site specific DNA cleavage of a target sequence by first incorporating foreign DNA into CRISPR loci. Polynucleotides containing these foreign sequences and the repeat-spacer elements of the CRISPR locus are in turn transcribed in a host cell to create a guide RNA, which can subsequently anneal to a target sequence and localize the Cas9 nuclease to this site. In this manner, highly site-specific cas9-mediated DNA cleavage can be engendered in a foreign polynucleotide because the interaction that brings cas9 within close proximity of the target DNA molecule is governed by RNA:DNA hybridization. As a result, one can design a CRISPR/Cas system to cleave any target DNA molecule of interest. This technique has been exploited in order to edit eukaryotic genomes (Hwang et al., *Nature Biotechnology* 31 :227 (2013)) and can be used as an efficient means of site-specifically editing target cell genomes in order to cleave DNA prior to the incorporation of a gene encoding a target gene. The use of CRISPR/Cas to modulate gene expression has been described in, for example, US Patent No. 8,697,359, the disclosure of which is incorporated herein by reference as it pertains to the use of the CRISPR/Cas system for genome editing. Alternative methods for site-specifically cleaving genomic DNA prior to the incorporation of a gene of interest in a target cell include the use of zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs). Unlike the CRISPR/Cas system, these enzymes do not contain a guiding polynucleotide to localize to a specific target sequence. Target specificity is instead controlled by DNA binding domains within these enzymes. The use of ZFNs and TALENs in genome editing applications is described, e.g., in Urnov et al. , *Nature Reviews Genetics* 11 :636 (2010); and in Joung et al., *Nature Reviews Molecular Cell Biology* 14:49 (2013), the disclosure of each of which are incorporated herein by reference as they pertain to compositions and methods for genome editing

[00306] Additional genome editing techniques that can be used to incorporate polynucleotides encoding target genes into the genome of a target cell include the use of ARCUSTM meganucleases that can be rationally designed so as to site-specifically cleave genomic DNA. The use of these enzymes for the incorporation of genes encoding target genes into the genome of a mammalian cell is advantageous in view of the defined structure-

activity relationships that have been established for such enzymes. Single chain meganucleases can be modified at certain amino acid positions in order to create nucleases that selectively cleave DNA at desired locations, enabling the site-specific incorporation of a target gene into the nuclear DNA of a target cell. These single-chain nucleases have been described extensively in, for example, US Patent Nos. 8,021,867 and US 8,445,251, the disclosures of each of which are incorporated herein by reference as they pertain to compositions and methods for genome editing.

ADDITIONAL EMBODIMENTS

[00307] Provided herein, in various embodiments, is a method for detecting integration of foreign DNA in genomic DNA of a cell, the method including: providing genomic DNA of the cell and reagents, the genomic DNA including an integration site where foreign DNA is integrated into the genomic DNA; within a droplet, tagmenting the genomic DNA using the reagents to obtain tagmented DNA fragments, wherein at least one of the tagmented DNA fragments includes the foreign DNA; amplifying the at least one of the tagmented DNA fragments including the foreign DNA to generate amplicons including a sequence derived from the foreign DNA; and determining the presence or absence of the amplicons wherein the presence of the amplicons detects integration of foreign DNA in genomic DNA of the cell.

[00308] In various embodiments, the presence or absence of the amplicons includes sequencing the amplicons.

[00309] In various embodiments, determining the presence or absence of the amplicons further includes analyzing sequenced amplicons to determine one or more integration sites.

[00310] Also provided herein, in various embodiments, is a method for detecting a proportion of cells in a population of cells having integration of foreign DNA in genomic DNA of the cells, the method including: for each cell in the population of cells: providing genomic DNA of the cell and reagents, the genomic DNA including an integration site where foreign DNA is integrated into the genomic DNA; within a droplet, tagmenting the genomic DNA using the reagents to obtain tagmented DNA fragments, wherein at least one of the tagmented DNA fragments includes the foreign DNA; amplifying the at least one of the tagmented DNA fragments including the foreign DNA to generate amplicons including a sequence derived from the foreign DNA; and sequencing the generated amplicons; and determining a proportion of the cells in the population of cells having integration of foreign DNA in genomic DNA of the cells based on the sequenced amplicons.

[00311] In various embodiments, sequencing the generated amplicons further includes characterizing a number of integration sites in the genomic DNA.

[00312] In various embodiments, characterizing the number of integration sites in the genomic DNA includes identifying one or more distinct integration sites in the genomic DNA from the sequenced amplicons.

[00313] In various embodiments, the reagents include a transposase.

[00314] In various embodiments, the transposase is a Tn5 transposase.

[00315] In various embodiments, the foreign DNA is viral DNA or modified viral DNA.

[00316] In various embodiments, the viral DNA is derived from an AAV, adenovirus, herpes simplex virus, or lentivirus.

[00317] In various embodiments, amplifying the at least one of the tagmented DNA fragments including the foreign DNA includes: providing a vector specific primer that hybridizes with a sequence of the foreign DNA; and performing nucleic acid extension using the hybridized vector specific primer.

[00318] In various embodiments, performing nucleic acid extension includes performing primer extension.

[00319] In various embodiments, performing nucleic acid extension includes performing polymerase chain reaction.

[00320] In various embodiments, tagmenting the genomic DNA using the reagents includes inserting adaptor sequences to obtain tagmented DNA fragments including the adaptor sequences.

[00321] In various embodiments, tagmenting the genomic DNA using the reagents does not include performing an extension to fill one or more gaps.

[00322] In various embodiments, each of the tagmented DNA fragments include at most one adaptor sequence.

[00323] In various embodiments, the vector specific primer hybridizes to an ITR.

[00324] In various embodiments, the vector specific primer hybridizes to a LTR region.

[00325] In various embodiments, genomic DNA of the cell and reagents are provided *in situ*.

[00326] In various embodiments, genomic DNA of the cell and reagents are provided in the droplet.

[00327] In various embodiments, genomic DNA of the cell and reagents are provided in a first droplet that differs from the droplet in which the genomic DNA is tagmented.

[00328] Also provided herein, in various embodiments, is a method for detecting integration of foreign DNA in genomic DNA of a cell, the method including: providing genomic DNA of the cell and reagents, the genomic DNA including an integration site where foreign DNA is integrated into the genomic DNA; within a droplet, tagmenting the genomic DNA using the reagents to obtain tagmented DNA fragments, wherein at least one of the tagmented DNA fragments includes the foreign DNA; amplifying the at least one of the tagmented DNA fragments including the foreign DNA to generate amplicons including a sequence derived from the foreign DNA; and determining the presence or absence of the amplicons wherein the presence of the amplicons detects integration of foreign DNA in genomic DNA of the cell.

[00329] In various embodiments, determining the presence or absence of the amplicons includes sequencing the amplicons.

[00330] In various embodiments, determining the presence or absence of the amplicons further includes analyzing sequenced amplicons to determine one or more integration sites.

[00331] Also provided herein, in various embodiments, is a method for detecting a proportion of cells in a population of cells having integration of foreign DNA in genomic DNA of the cells, the method including: for each cell in the population of cells: providing genomic DNA of the cell and reagents, the genomic DNA including an integration site where foreign DNA is integrated into the genomic DNA; within a droplet, tagmenting the genomic DNA using the reagents to obtain tagmented DNA fragments, wherein at least one of the tagmented DNA fragments includes the foreign DNA; amplifying the at least one of the tagmented DNA fragments including the foreign DNA to generate amplicons including a sequence derived from the foreign DNA; and sequencing the generated amplicons; and determining a proportion of the cells in the population of cells having integration of foreign DNA in genomic DNA of the cells based on the sequenced amplicons.

[00332] In various embodiments, sequencing the generated amplicons further includes characterizing a number of integration sites in the genomic DNA.

[00333] In various embodiments, characterizing the number of integration sites in the genomic DNA includes identifying one or more distinct integration sites in the genomic DNA from the sequenced amplicons.

[00334] In various embodiments, the reagents include a transposase.

[00335] In various embodiments, the transposase is a Tn5 transposase.

[00336] In various embodiments, the foreign DNA is viral DNA or modified viral DNA.

[00337] In various embodiments, the viral DNA is derived from an AAV, adenovirus, herpes simplex virus, or lentivirus.

[00338] In various embodiments, amplifying the at least one of the tagmented DNA fragments including the foreign DNA includes: providing a vector specific primer that
5 hybridizes with a sequence of the foreign DNA; and performing nucleic acid extension using the hybridized vector specific primer.

[00339] In various embodiments, performing nucleic acid extension includes performing primer extension.

[00340] In various embodiments, performing nucleic acid extension includes performing
10 polymerase chain reaction.

[00341] In various embodiments, tagmenting the genomic DNA using the reagents includes inserting adaptor sequences to obtain tagmented DNA fragments including the adaptor sequences.

[00342] In various embodiments, tagmenting the genomic DNA using the reagents does not
15 include performing an extension to fill one or more gaps.

[00343] In various embodiments, each of the tagmented DNA fragments include at most one adaptor sequence.

[00344] In various embodiments, the vector specific primer hybridizes to an ITR.

[00345] In various embodiments, the vector specific primer hybridizes to a LTR region.

[00346] In various embodiments, genomic DNA of the cell and reagents are provided *in situ*.
20

[00347] In various embodiments, genomic DNA of the cell and reagents are provided in the droplet.

[00348] In various embodiments, genomic DNA of the cell and reagents are provided in a first droplet that differs from the droplet in which the genomic DNA is tagmented.

[00349] In various embodiments, the method further includes determining one or more
25 mutations of the cell or the population of cells.

[00350] In various embodiments, the one or more mutations include a SNV or a CNV.

[00351] In various embodiments, the one or more mutations include a SNV and a CNV.

[00352] In various embodiments, the method further includes determining one or more
30 analytes expressed by the cell or the population of cells.

[00353] In various embodiments, the method further includes the cell or the population of cells are bound to at least one analyte-bound antibody-conjugated oligonucleotide.

[00354] In various embodiments, the method further includes the antibody-conjugated oligonucleotide includes a PCR handle, a tag sequence, and a capture sequence.

[00355] In various embodiments, the method further includes determining one or more mutations includes: performing a nucleic acid amplification reaction within the droplet using the antibody-conjugated oligonucleotide to generate an additional one or more amplicons, the additional one or more amplicons including an amplicon derived from the oligonucleotide; determining a presence or absence of an analyte using the second one or more amplicons; and characterizing the presence or absence of the analyte.

[00356] In various embodiments, the method further includes determining presence or absence of the analyte includes determining an expression level of the analyte bound by the antibody conjugated to the oligonucleotide.

[00357] In various embodiments, the method further includes generating a targeted DNA library or a targeted protein library.

EXAMPLES

[00358] The disclosure is further illustrated by the following examples. The examples are provided for illustrative purposes only, and are not to be construed as limiting the scope or content of the disclosure in any way.

Example 1: A Single-Cell Workflow for Viral Integration Sites and Somatic Genomic Variations

[00359] Characterization of viral vector integration sites in single cells can provide unique insights into the effects of these integrations on cellular function. In bulk assays, the influence of viral integration sites on malignancies or translocations would be difficult to identify. Rare integration sites could also be masked which could result in treatment-emergent serious adverse events (TESAEs). Single cell analysis provides single-cell resolution to better understand co-occurrence of specific integration sites with somatic genomic variations as well as the advantage of select off target integrations that could lead to clonal expansion.

[00360] The disclosed workflow determines viral integration sites for thousands of single cells and identifies gDNA sequence variations, including copy number variants, for these same cells, requiring no prior knowledge of integration sites. Leveraging cell barcoding, novel primer design strategies, including vector-specific and a second primer (*e.g.*, a repeat sequence-specific primer *e.g.*, an Alu primer described in Example 3), and enzymatic

manipulation of cellular contents, next generation sequencing libraries were created containing the viral vector integration sites and regions of interest in the gDNA (**FIG. 3**).

These studies on stable cell lines with integrated lentiviral vectors demonstrate a solution for simultaneous detection of single-cell viral integration sites and somatic variations. Methods involve pairing the viral integration chemistry with a multiplexed PCR panel containing over 300 amplicons covering myeloid targets. Based on cell genotypes, the sensitivity and specificity of integration site detection were determined. This novel workflow, which identifies viral integration sites and co-current somatic genomic variations, provides a better understanding of the relationship between viral integration sites and resulting malignancies, improving the efficacy and safety of therapies.

Example 2. Identification of Viral Integration and Targeted DNA

[00361] NIST VCN2 or Jurkat cells were transduced with a lentivirus. The NIST VCN2 cells or Jurkat were washed in BSA and DPBS, while control Raji cells were not transduced and were washed in DPBS. Cells were combined in a 1:2 ratio (NIST:Raji) for a final concentration of ~3000 cells/uL. The cells were then processed using the workflow process shown in **FIGs. 1 and 2** using the Tapestri®. In particular, single cells were partitioned into emulsions along with reagents. The reagents included a Tn5 mastermix prepared by mixing a Tn5 buffer containing Tris acetate and magnesium chloride, NP-40, proteinase K, and a loaded Tn5 with a custom adaptor. This mastermix was loaded onto the Tapestri® along with the encapsulation oil. The cells were then encapsulated, followed by incubation for cell lysis, tagmentation, and protease treatment. These droplets were then loaded back onto the Tapestri® cartridge for droplet merging with barcoding primer beads and PCR reagents containing polymerase, buffer, and primers for targeted DNA, control regions, and foreign DNA segment-specific primers (e.g., for detecting integration sites). Such foreign DNA segment-specific primers were directed to a long-terminal region (LTR) of the lentivirus (e.g., AGTAGTGTGTGCCCCGTCTGT SEQ ID NO: 5).

[00362] After thermocycling to attach cell barcodes to targeted DNA amplicons, the control region amplicons and the integration site amplicons, the emulsions were broken. A digestion step was performed followed by an inner nested PCR amplification with targeted DNA primers, control region primers, and integration site primers (**FIG. 10**). Sequencing adaptors were then attached through library PCR.

[00363] Following sequencing, integration sites of foreign DNA into the genomic DNA was detected from a number of transduced cells (pseudobulk) processed on Tapestry (FIG. 11). Provided with the random, unique Tn5 insertion sites, the left side of the sequence reads do not align, and are of vary lengths. The rightmost side of the mapped sequence reads, which terminate at the LTR priming sites, do align due to the identical site of lentivirus integration in the NST cells. When different primers were used (e.g., any one of SEQ ID NOs: 1-11), integration sites were also observed (FIG. 12). As depicted by FIG. 14, only transduced cells displayed an integration site of foreign DNA, as would be expected. When separated by barcode, one Tn5 insertion site as well as identical sites of integration per cell were observed (FIG. 15). As above, the rightmost side of the mapped sequence reads, which terminate at the LTR priming sites, do align due to the identical site of integration in the NST cells whilst the left side of the sequence reads do not align due to the random insertion of Tn5 adapters.

[00364] Targeted DNA primers were also included in the reagents, which enabled targeted DNA metrics including panel uniformity and DNA completeness (FIG. 13A), genotypic mapping of the two cellular populations (FIG. 13B), as well as sequence reads of the Tn5 control regions (FIG. 13C).

[00365] Following sequencing of single cells from the above described experiment, the sequence maps from each cell, respectively, were used to estimate the vector copy number of viral DNA in each single cell. Overlapping sequences of amplicons from each sequence map were determined and the number of overlapping amplicons determined the vector copy number, such that two overlapping amplicons determined two vector copies in a single exemplary cell (FIG. 18).

[00366] These results provide evidence that the novel method herein can be used to detect viral integration sites, including counts of the vector copy number in single cells, and co-current somatic genomic variations in a targeted DNA panel.

Example 3: Amplicons of viral DNA – vector-specific priming

[00367] Cells are transduced with viral DNA, modified viral DNA, or a viral vector (e.g., a viral vector including a transgene encoding a protein of interest or a reporter gene). The cells are then processed using the workflow process shown in FIGs. 1 and 2, for example, using the Tapestry®. In particular, single cells are partitioned into emulsions along with reagents. The reagents may include a foreign DNA segment-specific primer, such as a primer directed, for example, to a long-terminal region (LTR) of a lentivirus. Exemplary foreign DNA

segment-specific primers are shown in **Table 4**, below. The reagents may also include a protease, a cell buffer (*e.g.*, including a detergent, a density-match agent, and a phosphate buffer), a lysis buffer (*e.g.*, including a reverse primer, a protease, a detergent, an RNA reverse transcriptase, an RNAase inhibitor, a transposase, and a magnesium buffer), a transposase (*e.g.*, a Tn5 transposase), and a transposase adapter (*e.g.*, a Tn5 transposase adapter). The protease and detergent cause single cells to lyse in the emulsions. A tube containing the encapsulation droplets is incubated, for example, at 55 °C for 10 min then 80 °C for 10 min. Within the droplet (or a tube), the genomic DNA is tagmented using the reagents to obtain tagmented DNA fragments, in which at least one of the tagmented DNA fragments include the foreign DNA. Genomic DNA including the foreign DNA (*e.g.*, viral DNA, modified viral DNA, or a viral vector) is processed. Specifically, within a first emulsion, genomic DNA from single cells are primed with one or more foreign DNA segment-specific primers (*e.g.*, a foreign DNA segment-specific primer and a second foreign DNA segment-specific primer) and an intermediary amplicon including a sequence derived from the foreign DNA will be generated.

Table 4: Foreign DNA segment-specific primers

SEQ ID NO.	Target	Sequence
1	AAV2	AACTCCATCACTAGGGGTTCC
2	AAV2	CGGCCTCAGTGAGCGA
3	LTR 4	GATCTGGTCTAACCAGAGAG
4	LTR 6	CCCAGTACAAGCAAAAAGCAG
5	LTR 1	AGTAGTGTGTGCCCCGTCTGT
6	LTR 5	GATCCCTCAGACCCTTTTAGTC
7	LTR 7	GACTAAAAGGGTCTGAGGGATC
8	LTR 8	GATCCCTCAGACCCTTTTAGTCAGAGT
9	LTR 9	TAAAGCTTGCCTTGAGTGCTTCA
10	LTR 10	CTCTGGTAACTAGAGATCCCTCA
11	LTR 11	CTGAGGGATCTCTAGTTACC

[00368] The cell lysate, including the amplicon including a sequence derived from the foreign DNA, was generated and was then emulsified in a second emulsion with reagents,

such as a barcode primer including a barcode identification sequence, a read 1 sequencing primer, and a read 2 sequencing primer. Nucleic acid amplification is then conducted to generate amplified nucleic acids derived from the amplicon including a sequence derived from the foreign DNA. Such a second intermediary amplicon includes a first read sequence, the barcode identification sequence, the constant region sequence, the foreign DNA segment-specific primer, the complement sequence of the foreign DNA, the second foreign DNA segment-specific primer, and a second read sequence.

[00369] Amplified nucleic acids are pooled in a tube (*e.g.*, PCR tube or Eppendorf tube) and emulsions are broken. The amplified nucleic acids undergo library preparation by adding P5 (*e.g.*, the first index sequence) and P7 sequence (*e.g.*, the second index sequence) adaptors. Nucleic acid sequences are then sequenced to obtain sequence reads. For example, the amplicon includes from 5'-to-3': the first index sequence, the first read sequence, the barcode identification sequence, the constant region sequence, the foreign DNA segment-specific primer, the complement sequence of the foreign DNA, the second foreign DNA segment-specific primer, the second read sequence, and the second index sequence.

Alternatively, for example, the amplicon includes from 5'-to-3': the first index sequence, the barcode identification sequence, the constant region sequence, the foreign DNA segment-specific primer, the complement sequence of the foreign DNA, the second foreign DNA segment-specific primer, and the second index sequence (**FIG. 9**). Sequence reads are clustered according to common barcodes. **FIG. 3** depicts DNA amplicon sizes observed with reads of genomic DNA including the foreign DNA obtained through tagmentation and vector-specific priming. Reads were present at various lengths. This indicates that foreign DNA was integrated into the genomic DNA of cells.

[00370] Using such a workflow, the vector copy number of the foreign DNA in the genomic DNA (gDNA) in each single cell was determined. For example, in a single cell, two unique Tn5 insertion sites were located within the gDNA (depicted by the two circular sector symbols at "Position 1" and "Position 3," respectively; **FIG. 16**). The sequence map showed two unique sequence reads, both having genome:vector junctions. This count of two genome:vector junctions, within a single cell, is used to determine that, for example, two vector copies exist in the single cell. Alternatively, or in addition to, this method, for example, assuming again that a Tn5 integrates randomly into two unique locations in a single cell, such as two positions in the foreign DNA sequence (depicted by the two circular sector symbols at "Position 2" and "Position 4," respectively; **FIG. 17A**), the sequence map of such

a cell contains two amplicons with an overlapping sequence of a portion of the vector sequence (depicted by vertical dashed lines). This overlapping read of the vector sequence, within a single cell, , for example, determines that two vector copies exist in the single cell. If a non-overlapping read are detected (**FIG. 17B**), they are discarded from vector copy number analyses.

Example 4: Amplicons of viral DNA – vector-specific priming and Alu priming

[00371] Cells are transduced with viral DNA, modified viral DNA, or a viral vector (*e.g.*, a viral vector including a transgene encoding a protein of interest or a reporter gene). The cells may be then processed using the workflow process shown in **FIGs. 4 and 5**, for example, using the Tapestri®. In particular, single cells are partitioned into emulsions along with reagents. The reagents include a foreign DNA segment-specific primer, such as a primer directed, for example, to a LTR of a lentivirus. Exemplary foreign DNA segment-specific primers are shown in **Table 2**, above. The reagents also include a protease, a cell buffer (*e.g.*, including a detergent, a density-match agent, and a phosphate buffer), a lysis buffer (*e.g.*, including a reverse primer, a protease, a detergent, an RNA reverse transcriptase, an RNAase inhibitor, a transposase, and a magnesium buffer), and an Alu primer. Exemplary Alu primers are shown in **Table 5**, below. The protease and detergent cause single cells to lyse in the emulsions. A tube containing the encapsulation droplets is incubated, for example, at 55 °C for 10 min then 80 °C for 10 min. Genomic DNA from single cells are primed with a foreign DNA segment-specific primer and an Alu primer and an intermediate amplicon including a sequence derived from the foreign DNA is generated.

Table 5: Alu primers

SEQ ID NO.	Name	Sequence
12	Alu 1A	TCCCAGCTACTGGGGAGGCTGAGG
13	Alu 2A	GCCTCCCAAAGTGCTGGGATTACAG
14	Alu 1B	TGTAATCCCAGCTACTCGGG
15	Alu 2B	GAACTCCTGACCTCAGG
16	Alu 1D	CCTCAGCCTCCCCAGTAGCTGGGA
17	Alu 2D	CTGTAATCCCAGCACTTTGGGAGGC

18	18S 1A	CGCTCCACCAACTAAGAACG
19	18S 2A	CGTTCTTAGTTGGTGGAGCG
20	18S 1D	CGCTCCACCAACTAAGAACG
21	18S 2D	CTCAACACGGGAAACCTCAC
22	LINE1 1A	GCAGTGTGTAGAGGGAA
23	LINE1 2A	GTGGTCAATTTTGGAATAGGTGT
24	LINE1 1D	ACACCTATTCCAAAATTGACCAC
25	LINE1 2D	TTCCCTCTACACACTGC

[00372] The cell lysate, including the amplicon including a sequence derived from the foreign DNA, is generated and is then emulsified in a second emulsion with reagents, such as a barcode primer including a barcode identification sequence, a read 1 sequencing primer, and a read 2 sequencing primer. Nucleic acid amplification is then conducted to generate amplified nucleic acids derived from the amplicon including a sequence derived from the foreign DNA. Such a second intermediary amplicon includes the first read sequence, the barcode identification sequence, a constant region sequence, the Alu primer, the complement sequence of the foreign DNA, the foreign DNA segment-specific primer, and the second read sequence.

[00373] Amplified nucleic acids are pooled in a tube (*e.g.*, PCR tube or Eppendorf tube) and emulsions are broken. The amplified nucleic acids undergo library preparation by adding P5 (*e.g.*, the first index sequence) and P7 sequence (*e.g.*, the second index sequence) adaptors. Nucleic acid sequences are then sequenced to obtain sequence reads. For example, the amplicon includes from 5'-to-3': the first index sequence, the first read sequence, the barcode identification sequence, the constant region sequence, the Alu primer, the complement sequence of the foreign DNA, the second foreign DNA segment-specific primer, the second read sequence, and the second index sequence. Sequence reads are clustered according to common barcodes. **FIGs. 6A-6D** depict DNA amplicon sizes observed with reads of genomic DNA including the foreign DNA obtained through vector-specific and Alu primer priming. Reads are present at various lengths. This indicates that foreign DNA was integrated into the genomic DNA of cells.

Example 5. Viral Integration and Somatic Genomic Variations Detected with Alu Priming

[00374] NIST VCN2 or Jurkat cells were transduced with a lentivirus. The NIST VCN2 cells or Jurkat were washed in BSA and DPBS, while control Raji cells were not transduced and were washed in DPBS. Cells were combined in a 1:2 ratio (NIST:Raji) for a final concentration of ~3000 cells/uL. The cells were then processed using the workflow process shown in **FIG. 4** using the Tapestri®. In particular, single cells were partitioned into emulsions along with reagents. The reagents included an Alu repeat sequence-specific primer. A mastermix including the repeat sequence-specific primer and a foreign DNA segment-specific primer was loaded onto the Tapestri® along with the encapsulation oil. The cells were then encapsulated, followed by incubation for cell lysis and protease treatment. These droplets were then loaded back onto the Tapestri® cartridge for droplet merging with barcoding primer beads and PCR reagents containing polymerase, buffer, and primers for targeted DNA, Alu repeat-sequence-specific PCR, control regions, and foreign DNA segment-specific primers (e.g., for detecting integration sites). Such foreign DNA segment-specific primers were directed to a long-terminal region (LTR) of the lentivirus.

[00375] After barcoding the PCR, the amplicons were separated based upon whether they were for Alu-PCR (e.g., LTR targets, control region targets, and foreign DNA segment targets) or the targeted DNA panel. Following sequencing, multiple LTR priming sites and an integration site of foreign DNA into the genomic DNA were detected from a number of transduced cells processed on Tapestri (**FIG. 7**). When separated by barcode, two 5' LTR priming sites were readily observed (**FIG. 8**).

INCORPORATION BY REFERENCE

[00376] The entire disclosure of each of the Patent documents and scientific articles cited herein is incorporated by reference for all purposes.

EQUIVALENTS

[00377] The disclosure can be embodied in other specific forms without departing from the essential characteristics thereof. The foregoing embodiments therefore are to be considered illustrative rather than limiting on the disclosure described herein. The scope of the disclosure is indicated by the appended claims rather than by the foregoing description,

and all changes that come within the meaning and range of equivalency of the claims are intended to be embraced therein.

WHAT IS CLAIMED IS:

1. A method for detecting integration of a vector comprising a foreign DNA segment into genomic DNA of a cell, the method comprising:
 - providing, within a droplet, the genomic DNA of the cell and reagents, the genomic DNA comprising an integration site where the vector comprising the foreign DNA segment is integrated into the genomic DNA, wherein the reagents comprise a foreign DNA segment-specific primer and a second primer;
 - within the droplet, generating one or more amplicons comprising the integrated foreign DNA segment, if present, using at least the hybridized foreign DNA segment-specific and second primers; and
 - determining the presence or absence of the one or more amplicons, wherein integration of the foreign DNA segment in the genomic DNA of the cell is detected by determining the presence of the one or more amplicons; and wherein no integration of the foreign DNA segment is detected by determining the absence of the one or more amplicons.
2. The method of claim 1, wherein using at least the hybridized foreign DNA segment-specific and second primers comprises:
 - hybridizing the foreign DNA segment-specific primer to the foreign DNA segment, if present in the integration site;
 - extending the hybridized foreign DNA segment-specific primer to generate an extension product; and
 - hybridizing the second primer to a sequence of the extension product.
3. The method of claim 2, wherein the extension product comprises a sequence derived from a transposase adapter sequence.
4. The method of claim 3, wherein the transposase is a Tn5 transposase.
5. The method of claim 4, wherein the transposase adapter is a Tn5 transposase adapter.

6. The method of any one of claims 1-5, wherein the sequence of the extension product comprises a sequence derived from the genomic DNA.
7. The method of claim 1, wherein using at least the hybridized foreign DNA segment-specific and second primers comprises:
hybridizing the foreign DNA segment-specific primer to the foreign DNA segment, if present in the integration site, and hybridizing the second primer to a sequence present in the genomic DNA or to a sequence present in the foreign DNA segment.
8. A method for detecting integration of a vector comprising a foreign DNA segment into genomic DNA of a cell, the method comprising:
providing, within a droplet, the genomic DNA of the cell and reagents, the genomic DNA comprising an integration site where the vector comprising the foreign DNA segment is integrated into the genomic DNA, wherein the reagents comprise a foreign DNA segment-specific primer;
within the droplet, generating one or more amplicons comprising the integrated foreign DNA segment, if present, using at least the hybridized foreign DNA segment-specific primer; and
determining the presence or absence of the one or more amplicons, wherein integration of the foreign DNA segment in the genomic DNA of the cell is detected by determining the presence of the one or more amplicons; and wherein no integration of the foreign DNA segment is detected by determining the absence of the one or more amplicons.
9. The method of any one of claims 1-8, further comprising sequencing or determining the length of the one or more amplicons.
10. The method of claim 9, further comprising analyzing the one or more amplicons sequence and/or the one or more amplicons size to identify the amplicon identity, the genomic locus of the integration site, the number of integration sites, or the orientation of the integration, optionally wherein the number of integration sites comprises the vector copy number.

11. A method for detecting a proportion of cells in a population of cells having integration of a vector comprising a foreign DNA segment into genomic DNA of the cells, the method comprising:

for each of one or more cells in the population of cells:

providing, within a droplet, the genomic DNA of the cell and reagents, the genomic DNA comprising an integration site where the vector comprising the foreign DNA segment is integrated into the genomic DNA, wherein the reagents comprise a foreign DNA segment-specific primer and a second primer;

within the droplet, generating one or more amplicons comprising the integrated foreign DNA segment, if present, using at least the hybridized foreign DNA segment-specific and second primers; and sequencing the generated one or more amplicons; and

determining a proportion of the cells in the population of cells having integration of the foreign DNA segment in genomic DNA of the cells based on the sequenced one or more amplicons.

12. The method of any one of claims 1-11, wherein within the droplet, the method further comprises exposing the cell to the reagents, wherein the reagents comprise a protease and a detergent and lysing the cell using the protease and the detergent.

13. The method of claim 12, wherein the detergent is a pluronic detergent.

14. The method of any one of claims 11-13, wherein sequencing the generated one or more amplicons further comprises characterizing a number of integration sites in the genomic DNA.

15. The method of any one of claims 1-14, wherein the foreign DNA segment is viral DNA, modified viral DNA, or DNA from a viral vector.

16. The method of claim 15, wherein the DNA from a viral vector comprises a transgene encoding a protein of interest or a reporter gene.

17. The method of claim 16, wherein the DNA from a viral vector comprises a transgene encoding a protein of interest.
18. The method of any one of claims 15-17, wherein the method further comprises transducing the cell or the population of cells with the viral DNA, the modified viral DNA, or a viral vector.
19. The method of any one of claims 15-18, wherein the viral DNA, modified viral DNA, or viral vector is derived from an adeno-associated virus (AAV), adenovirus, herpes simplex virus, lentivirus, retrovirus, poxvirus, baculovirus, or vaccinia virus.
20. The method of any one of claims 1-19, wherein the reagents comprise a cell buffer and/or a lysis buffer.
21. The method of claim 20, wherein the lysis buffer comprises one or more of a reverse primer, a protease, a detergent, an RNA reverse transcriptase, an RNAase inhibitor, a transposase, and a magnesium buffer.
22. The method of claim 21, wherein the lysis buffer comprises a protease, a detergent, a transposase, and a magnesium buffer.
23. The method of claim 21 or 22, wherein the transposase is preloaded with an adapter.
24. The method of any one of claims 21-23, wherein the magnesium buffer comprises magnesium, Tris, potassium, [tris(hydroxymethyl)methylamino]propanesulfonic acid (TAPS), dimethylformamide (DMF), and/or poly(ethylene glycol) (PEG).
25. The method of any one of claims 1-24, wherein the droplet is a water-in-oil emulsion, wherein an oil solution of the water-in-oil emulsion comprises one or more of an oil and a non-ionic surfactant.
26. The method of claim 25, wherein the oil comprises a fluoruous oil.

27. The method of claim 25, wherein the non-ionic surfactant is a fluoruous non-ionic surfactant.
28. The method of any one of claims 1-27, wherein the reagents further comprise a barcode primer comprising a barcode identification sequence.
29. The method of claim 28, wherein the barcode primer is a bead barcode primer.
30. The method of any one of claims 3-6 and 9-29, wherein the second primer is a second foreign DNA segment-specific primer, and wherein the method further comprises:
hybridizing the foreign DNA segment-specific primer to a sequence derived from a transposase adapter sequence.
31. The method of any one of claims 1-30, wherein the reagents comprise a transposase.
32. The method of claim 31, wherein the transposase is a Tn5 transposase.
33. The method of claim 31 or 32, wherein within the droplet, the method further comprises tagmenting the genomic DNA using the reagents to obtain tagmented DNA fragments, wherein at least one of the tagmented DNA fragments comprises the foreign DNA segment.
34. The method of claim 33, wherein extending comprises extension of the at least one of the tagmented DNA fragments.
35. The method of claim 33 or 34, wherein tagmenting the genomic DNA using the reagents comprises inserting adaptor sequences to obtain tagmented DNA fragments comprising the adaptor sequences.
36. The method of any one of claims 33-35, wherein tagmenting the genomic DNA using the reagents does not include performing an extension to fill one or more gaps.

37. The method of any one of claims 33-36, wherein each of the tagmented DNA fragments comprise at most one adaptor sequence.
38. The method of any one of claims 33-37, wherein genomic DNA of the cell and reagents are provided in a first droplet that differs from the droplet in which the genomic DNA is tagmented.
39. The method of any one of claims 33-37, wherein genomic DNA of the cell and reagents are provided in the same droplet as the droplet in which the genomic DNA is tagmented.
40. The method of any one of claims 1 and 7-29, wherein the second primer is a repeat sequence-specific primer, and wherein the method further comprises:
hybridizing the repeat sequence-specific primer to a repeat sequence present in the genomic DNA.
41. The method of claim 40, wherein the repeat sequence-specific primer is an Alu1, an Alu2, a LINE1, an 16S, an 18S primer, or any combination thereof.
42. The method of any one of claims 1-41, wherein extending comprises performing nucleic acid extension.
43. The method of claim 42, wherein performing nucleic acid extension comprises performing primer extension.
44. The method of claim 42 or 43, wherein performing nucleic acid extension comprises extending the foreign DNA segment-specific primer to produce the one or more amplicons comprising a constant region sequence and the foreign DNA segment-specific primer.
45. The method of any one of claims 42-44, wherein performing nucleic acid extension further comprises producing the one or more amplicons comprising a complement sequence of the foreign DNA segment.

46. The method of any one of claims 42-45, wherein performing nucleic acid extension comprises extending the barcode identification sequence to produce the one or more amplicons comprising a first read sequence, the barcode identification sequence, and a constant region sequence.
47. The method of any one of claims 42-46, wherein performing nucleic acid extension comprises extending the second foreign DNA segment-specific primer to produce the one or more amplicons comprising the second foreign DNA segment-specific primer and a second read sequence.
48. The method of any one of claims 42-47, wherein performing nucleic acid extension comprises extending the repeat sequence-specific primer to produce the one or more amplicons comprising a constant region sequence and the repeat sequence-specific primer.
49. The method of any one of claims 1-48, wherein the reagents further comprise a read 1 sequencing primer and/or a read 2 sequencing primer.
50. The method of claim 49, wherein the method further comprises breaking an emulsion that comprises the droplet and performing nucleic acid extension, wherein performing nucleic acid extension comprises performing polymerase chain reaction (PCR).
51. The method of claim 50, wherein PCR comprises extending the read 1 sequencing primer to produce the one or more amplicons comprising a first index sequence and a first read sequence.
52. The method of claim 50 or 51, wherein performing PCR comprises extending the read 2 sequencing primer to produce the one or more amplicons comprising the second read sequence and a second index sequence.
53. The method of any one of claims 1-52, wherein the foreign DNA segment comprises an inverted terminal repeat region (ITR), a rep gene, a cap gene, a long terminal repeat (LTR) region, a gag gene, a pol gene, a tat gene, a rev gene, a IX gene, a IVa2 gene, an L1 gene, an L2 gene, an L3 gene, an L4 gene, an L5 gene, an E2B gene, an E2A gene, an E2A-L gene, an

E4 gene, a gene encoding a capsomer protein, a gene encoding a capsid protein, a gene encoding a core protein, a gene encoding a viral non-structural protein, or a gene encoding a viral packing protein.

54. The method of claim 53, wherein the foreign DNA segment comprises an LTR.

55. The method of any one of claims 1-54, wherein the foreign DNA segment-specific primer or the second foreign DNA segment-specific primer comprises the nucleic acid sequence of any one of SEQ ID NOs: 1-11.

56. The method of any one of claims 40-55, wherein the repeat sequence-specific primer comprises the nucleic acid sequence of any one of SEQ ID NOs: 12-25.

57. The method of any one of claims 44-56, wherein the one or more amplicons comprise from 5'-to-3': the first index sequence, the first read sequence, the barcode identification sequence, the constant region sequence, the foreign DNA segment-specific primer, the complement sequence of the foreign DNA segment, the second foreign DNA segment-specific primer, the second read sequence, and the second index sequence.

58. The method of any one of claims 44-57, wherein the one or more amplicons comprise from 5'-to-3': the first index sequence, the barcode identification sequence, the constant region sequence, the foreign DNA segment-specific primer, the complement sequence of the foreign DNA segment, the second foreign DNA segment-specific primer, and the second index sequence.

59. The method of any one of claims 44-58, wherein the one or more amplicons comprise from 5'-to-3': the first index sequence, the first read sequence, the barcode identification sequence, the constant region sequence, the repeat sequence-specific primer, the complement sequence of the foreign DNA segment, the second foreign DNA segment-specific primer, the second read sequence, and the second index sequence.

60. The method of any one of claims 1-59, wherein the genomic DNA further comprises one or more additional integration sites where copies of the foreign DNA segment are integrated into the genomic DNA.

61. The method of claim 60, further comprising determining a vector copy number of the foreign DNA segment across the integration site and the one or more additional integration sites.

62. The method of claim 61, wherein determining the vector copy number comprises: identifying a first amplicon comprising a sequence of the foreign DNA segment and a second amplicon comprising a sequence of the foreign DNA segment, wherein the first amplicon and the second amplicon include different start sites; and determining whether a portion of the sequence of the foreign DNA segment of the first amplicon overlaps with a portion of the sequence of the foreign DNA segment of the second amplicon.

63. The method of claim 62, wherein the different start sites of the first amplicon and the second amplicon correspond to different Tn5 insertion sites.

64. The method of claim 62 or 63, wherein the first amplicon and second amplicon share a common termination site.

65. The method of claim 64, wherein the common termination sites of the first amplicon and second amplicon correspond to the foreign DNA segment-specific primer.

66. The method of any one of claims 62-65, wherein responsive to the determination that the portion of the sequence of the foreign DNA segment of the first amplicon overlaps with a portion of the sequence of the foreign DNA segment of the second amplicon, determining that the vector copy number is at least 2.

67. The method of any one of claims 62-65, wherein responsive to the determination that the portion of the sequence of the foreign DNA segment of the first amplicon does not

overlap with a portion of the sequence of the foreign DNA segment of the second amplicon, determining that the vector copy number is 1.

68. The method of any one of claims 9-67, further comprising determining one or more mutations of the cell or the population of cells.

69. The method of claim 68, wherein the one or more mutations comprise a single nucleotide variant (SNV) or a copy number variation (CNV).

70. The method of claim 69, wherein the one or more mutations comprise a SNV and a CNV.

71. The method of any one of claims 9-70, further comprising determining one or more analytes expressed by the cell or the population of cells.

72. The method of claim 71, wherein the cell or the population of cells are bound to at least one analyte-bound antibody-conjugated oligonucleotide.

73. The method of claim 72, wherein the antibody-conjugated oligonucleotide comprises a PCR handle, a tag sequence, and a capture sequence.

74. The method of claim 72 or 73, wherein determining one or more mutations comprises:
performing a nucleic acid amplification reaction within the droplet using the
antibody-conjugated oligonucleotide to generate an additional one or
more amplicons, the additional one or more amplicons comprising
an amplicon derived from the oligonucleotide;
determining a presence or absence of an analyte using the second one or more
amplicons; and
characterizing the presence or absence of the analyte.

75. The method of claim 74, wherein determining presence or absence of the analyte comprises determining an expression level of the analyte, the analyte bound by the antibody conjugated to the oligonucleotide.

76. The method of any one of claims 1-75, further comprising generating a targeted DNA library or a targeted protein library.
77. A method for detecting integration of a vector comprising a foreign DNA segment into genomic DNA of a cell, the method comprising:
- providing, in a bulk setting, the genomic DNA of the cell and reagents, the genomic DNA comprising an integration site where the vector comprising the foreign DNA segment is integrated into the genomic DNA, wherein the reagents comprise a foreign DNA segment-specific primer and a second primer;
 - in a bulk setting, generating one or more amplicons comprising the integrated foreign DNA segment, if present, using at least the hybridized foreign DNA segment-specific and second primers; and
 - determining the presence or absence of the one or more amplicons, wherein integration of the foreign DNA segment in the genomic DNA of the cell is detected by determining the presence of the one or more amplicons; and wherein no integration of the foreign DNA segment is detected by determining the absence of the one or more amplicons.
78. A method for detecting translocation of a DNA segment in genomic DNA of a cell, the method comprising:
- providing, within a droplet, the genomic DNA of the cell and reagents, the genomic DNA comprising an integration site where the translocated DNA segment is integrated into the genomic DNA, wherein the reagents comprise a translocated DNA segment-specific primer and a second primer;
 - within the droplet, generating one or more amplicons comprising the integrated translocated DNA segment, if present, using at least the hybridized translocated DNA segment-specific and second primers; and
 - determining the presence or absence of the one or more amplicons, wherein integration of the translocated DNA segment in the genomic DNA of the cell is detected by determining the presence of the one or more amplicons; and wherein no integration of the translocated DNA segment is detected by determining the absence of the one or more amplicons.

79. A method for detecting genetic editing of a DNA segment of genomic DNA of a cell, the method comprising:

providing, within a droplet, the genomic DNA of the cell and reagents, the genomic DNA comprising an integration site where the DNA segment is integrated into the genomic DNA by the genetic editing, wherein the reagents comprise a DNA segment-specific primer and a second primer;

within the droplet, generating one or more amplicons comprising the integrated DNA segment, if present, using at least the hybridized DNA segment-specific and second primers; and

determining the presence or absence of the one or more amplicons, wherein integration of the DNA segment in the genomic DNA of the cell is detected by determining the presence of the one or more amplicons; and wherein no integration of the DNA segment is detected by determining the absence of the one or more amplicons.

80. The method of claim 79, wherein genetic editing comprises use of a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas9 system, a meganuclease, a zinc finger nuclease (ZFN), a transposase, an integrase, or a recombinase.

FIG. 1

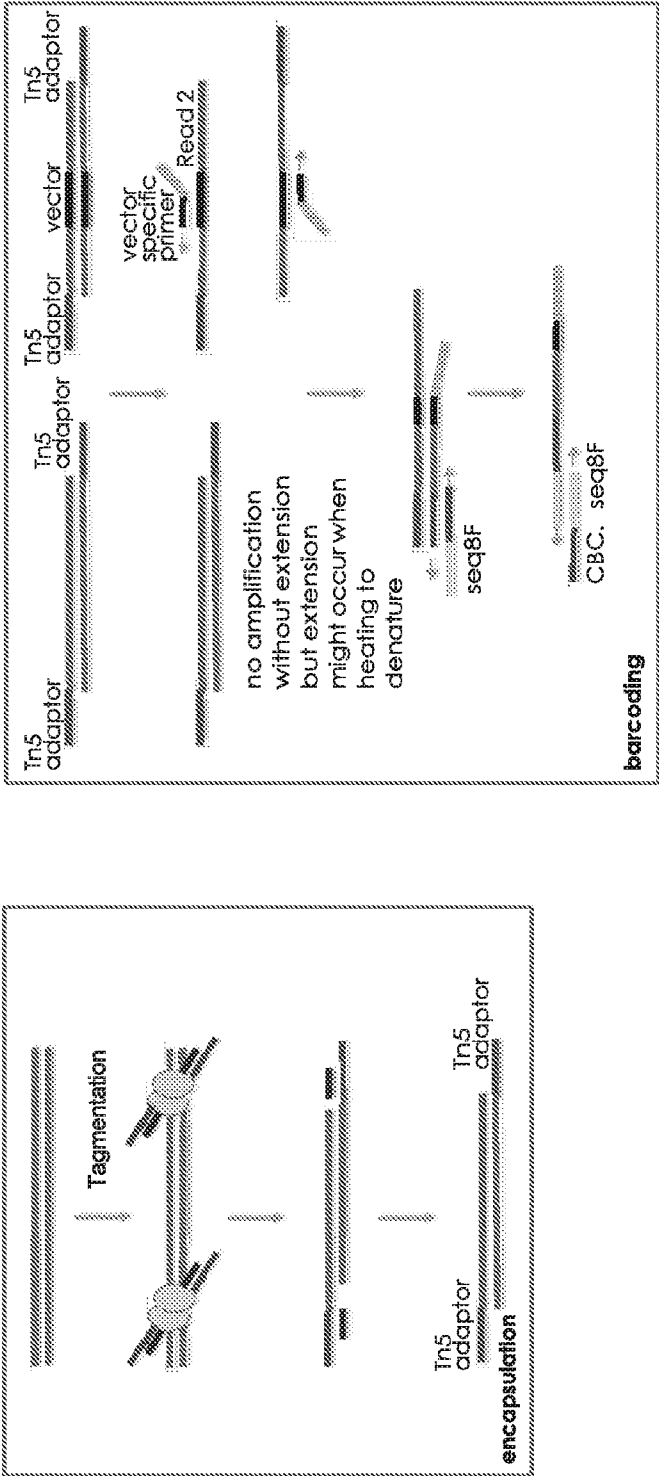


FIG. 2

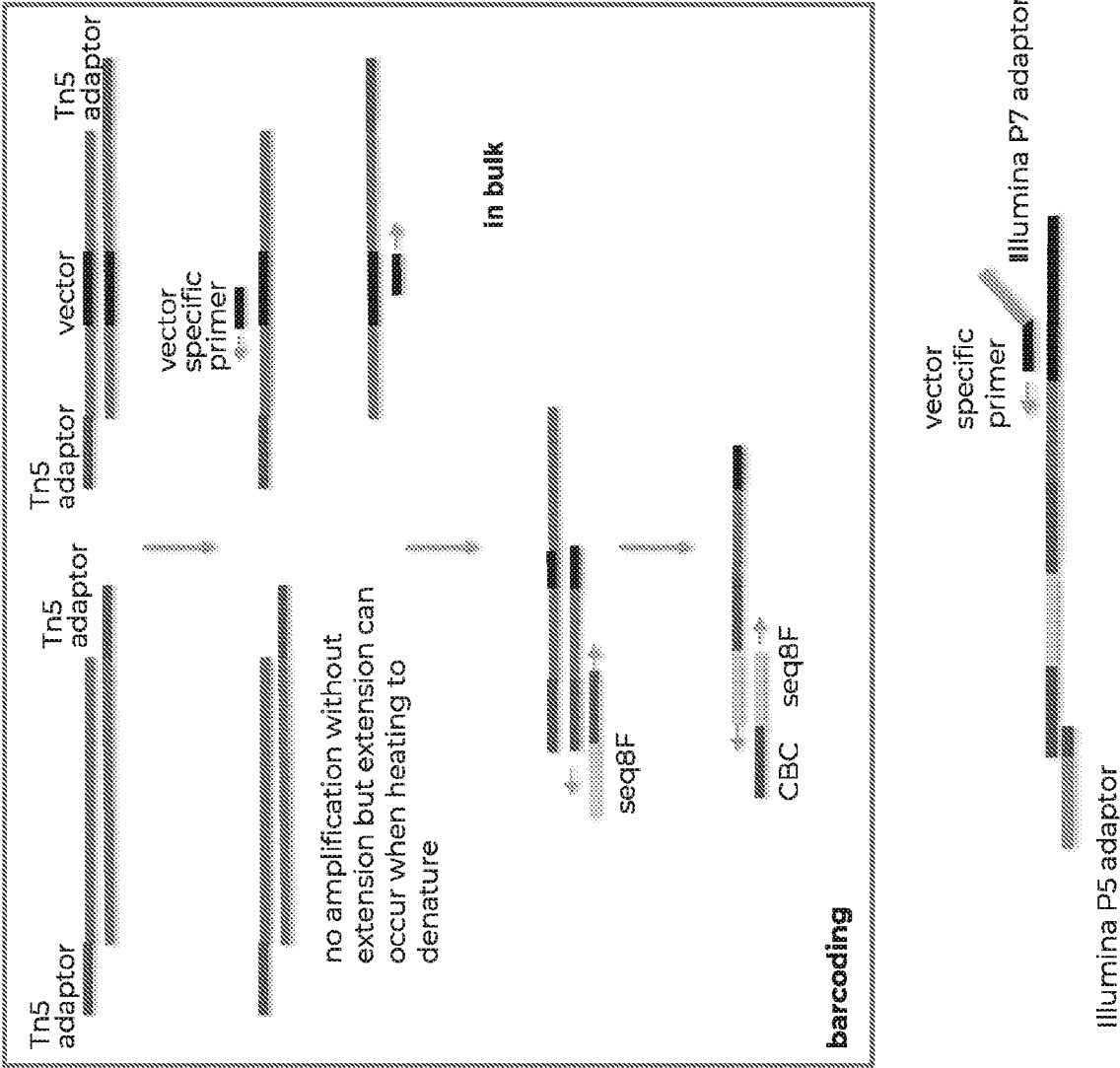
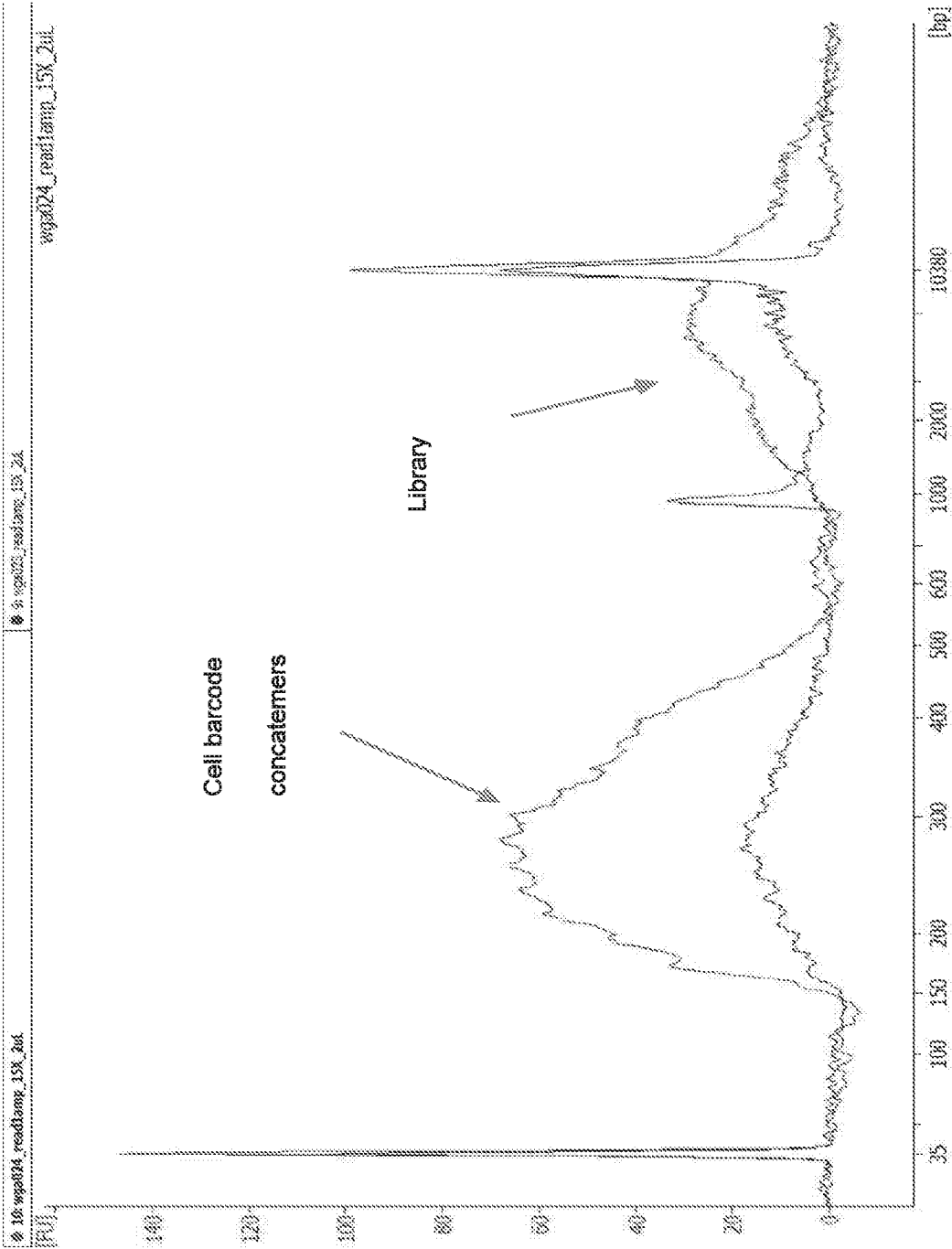


FIG. 3



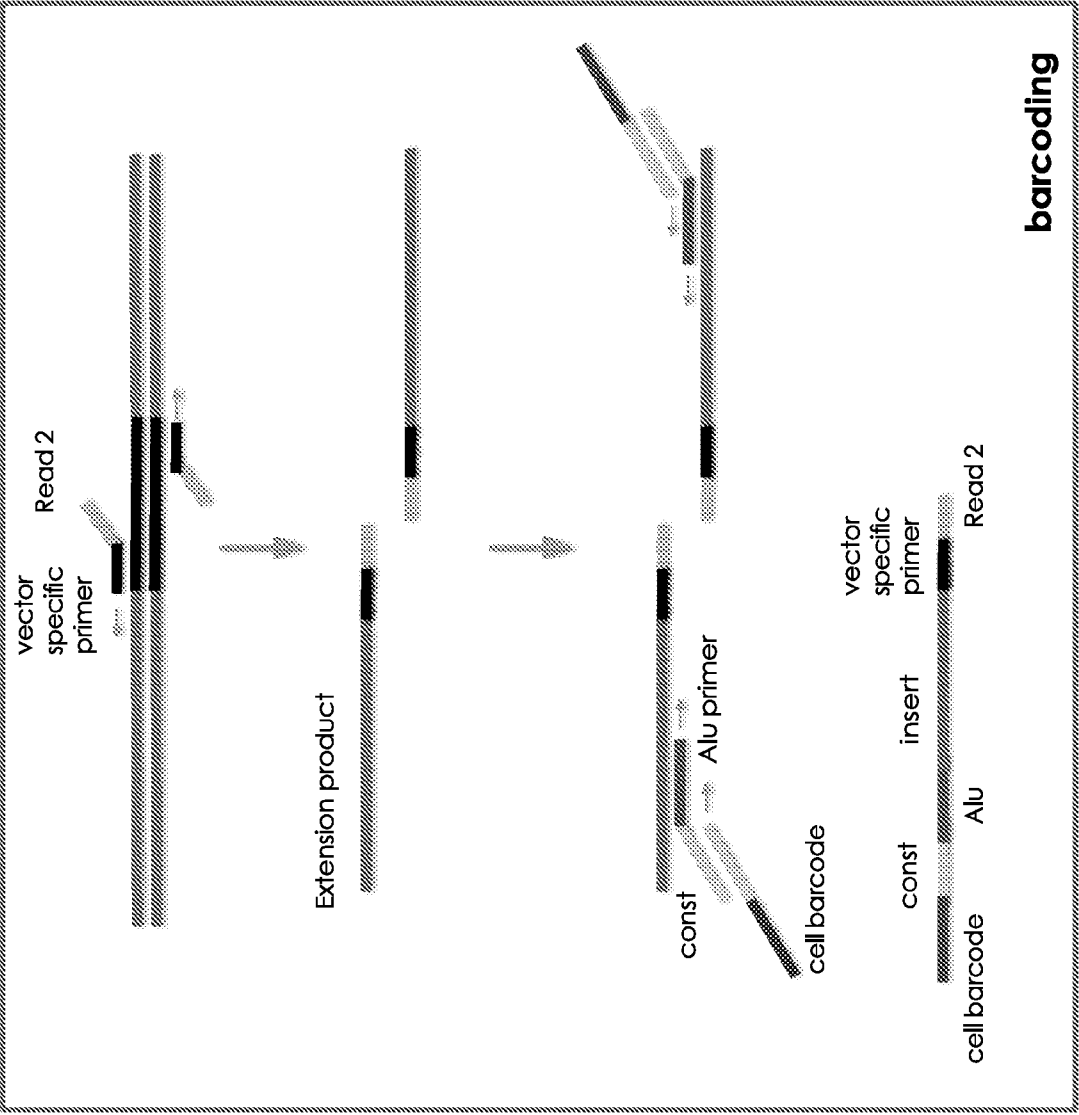
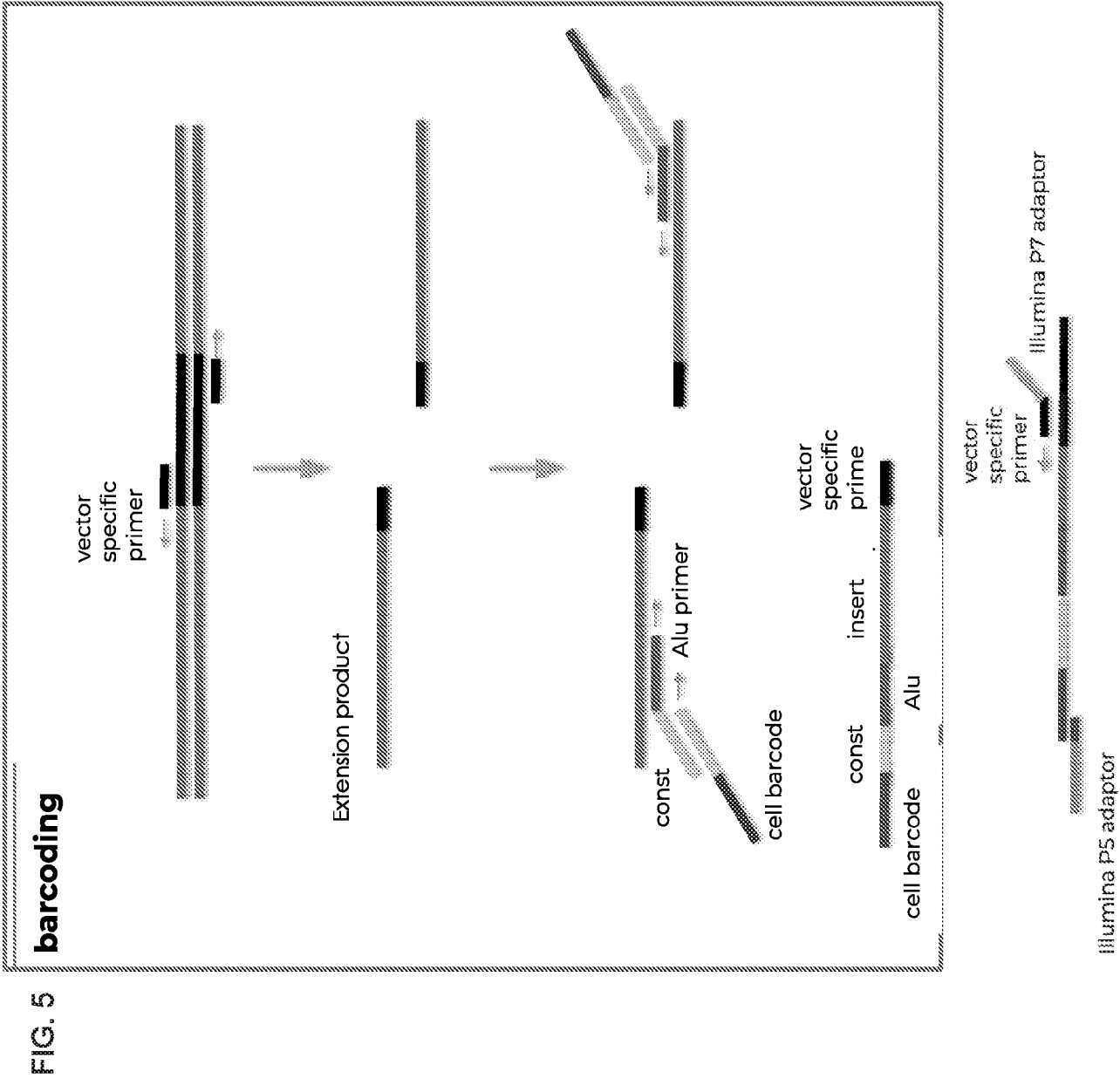


FIG. 4



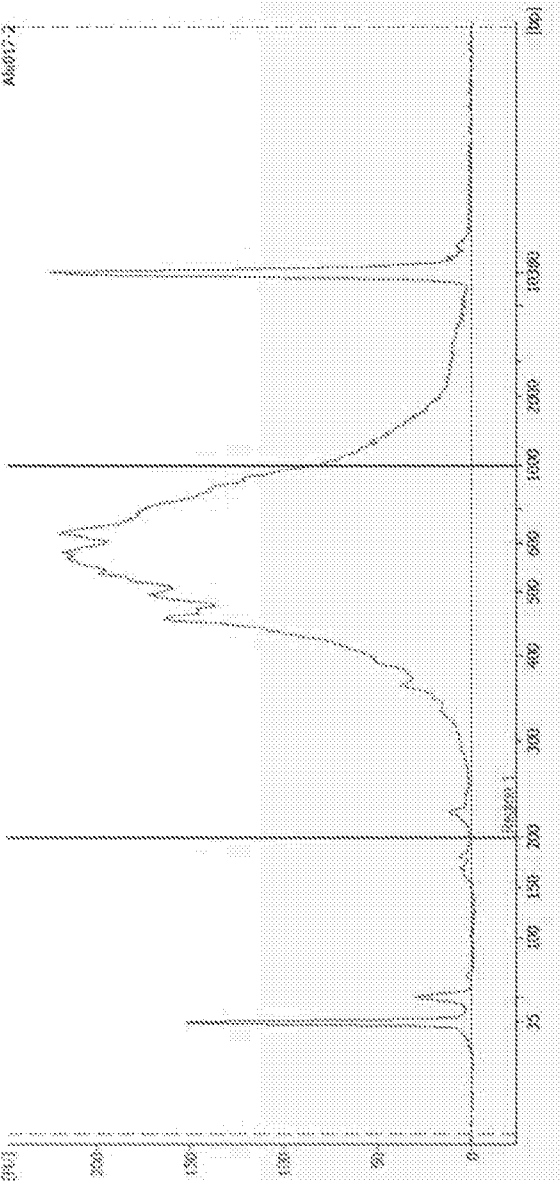


FIG. 6A
Alu2

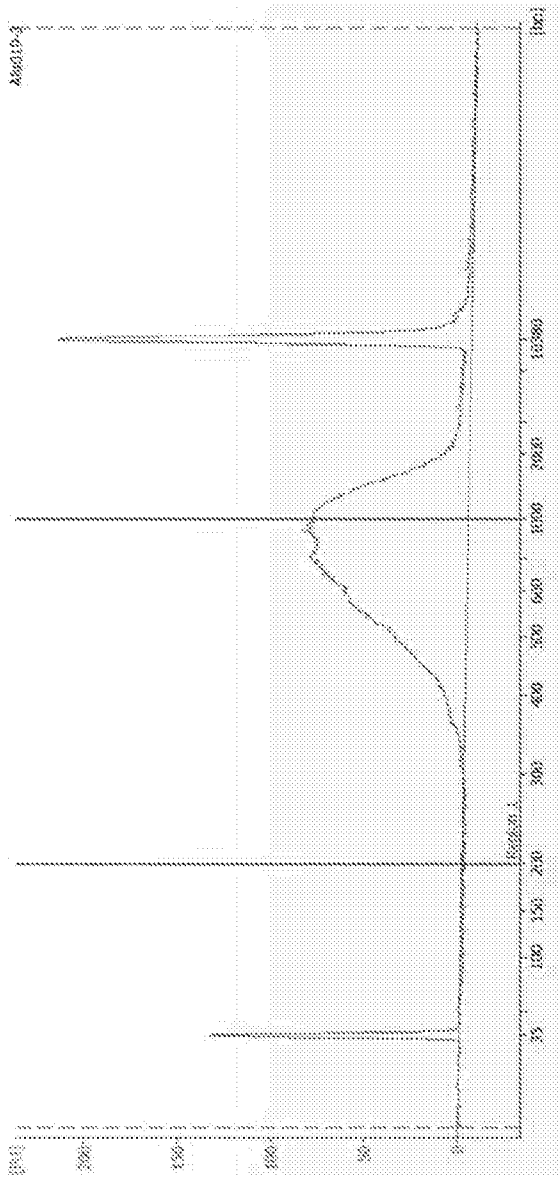


FIG. 6B
Alu1 + Alu2

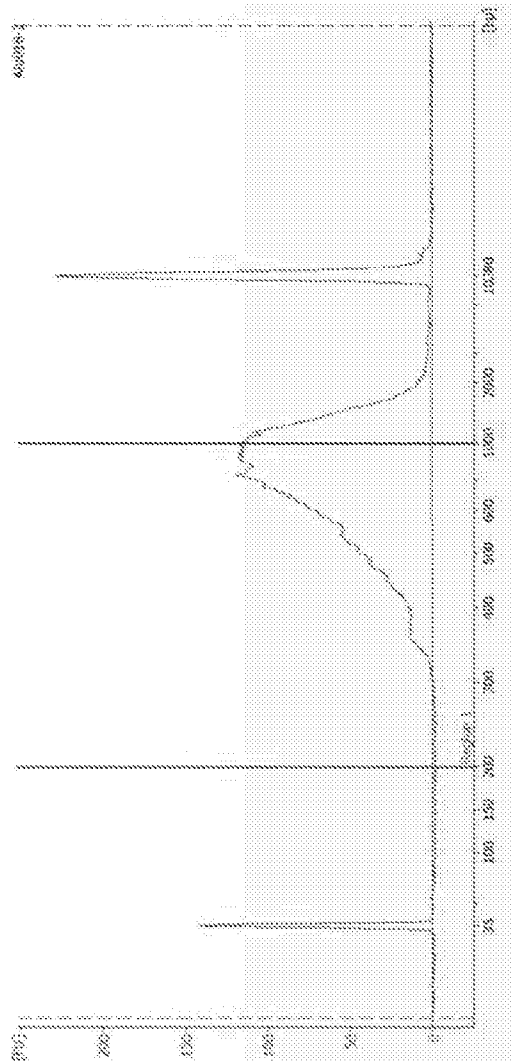


FIG. 6C

Alu fwd2 +
Alu fwdR

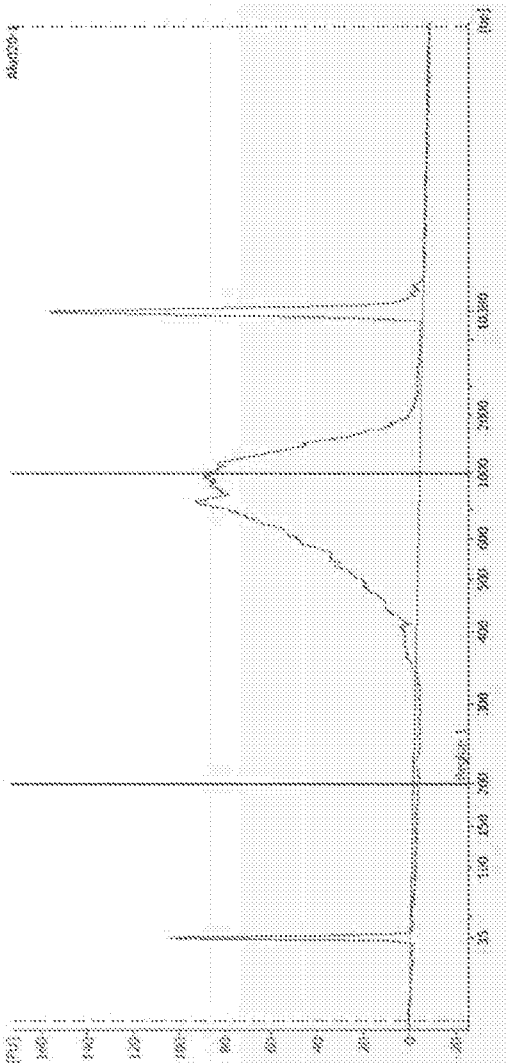


FIG. 6D

Alu fwd2+Alu fwdR + 18S fwd2 + LINE1 fwd2 + LINE1 fwd2 + LINE1 fwdR

FIG. 7

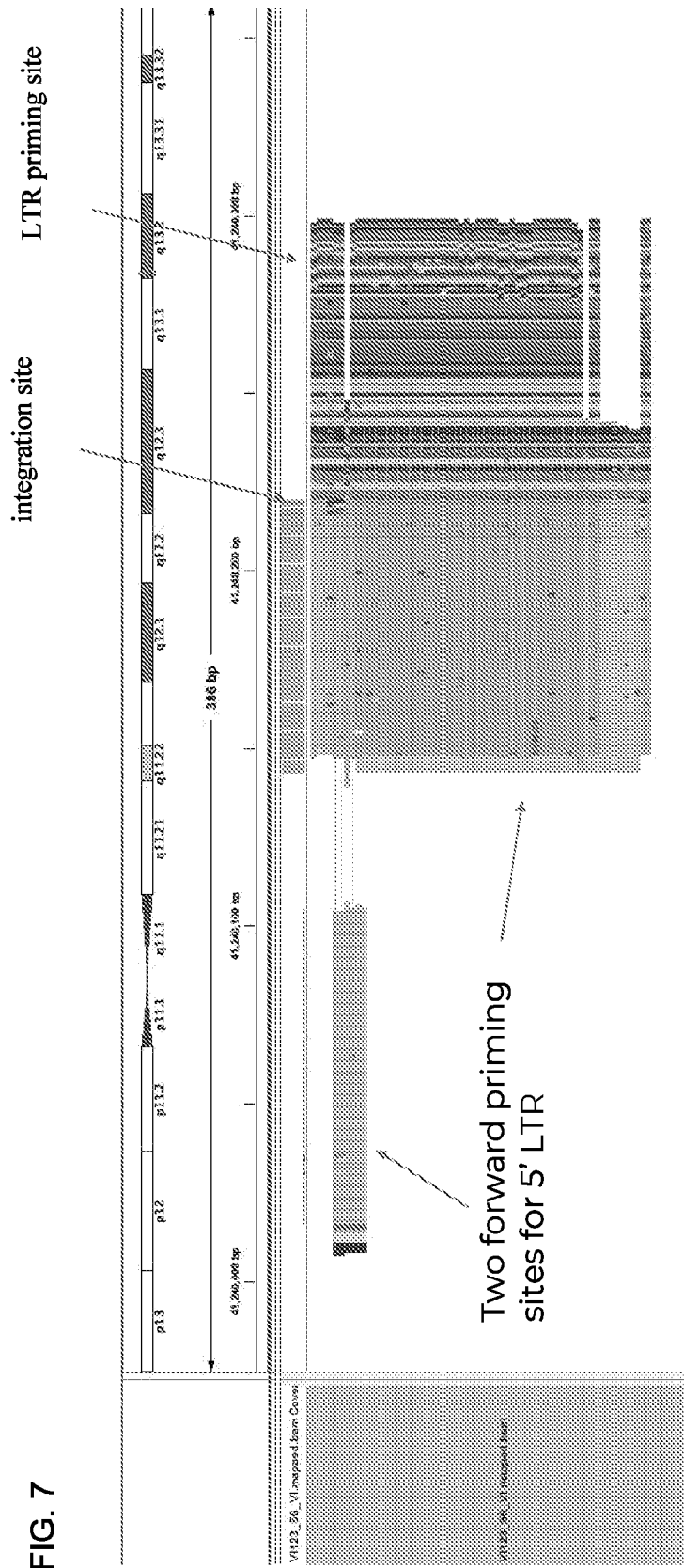


FIG. 9

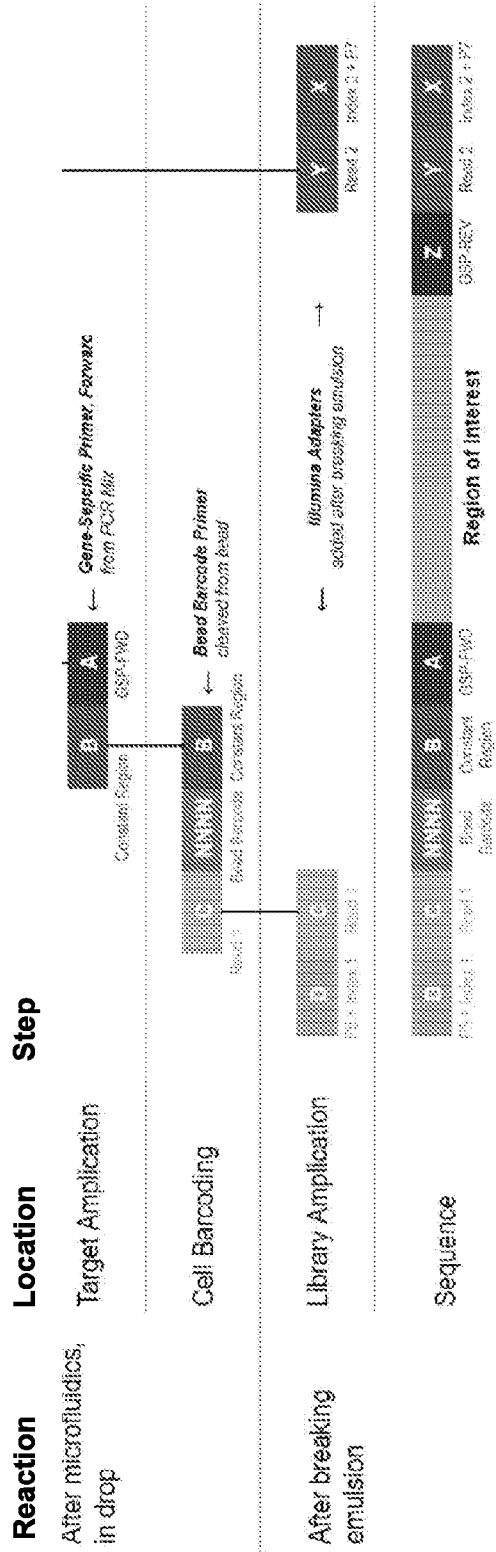


FIG. 10

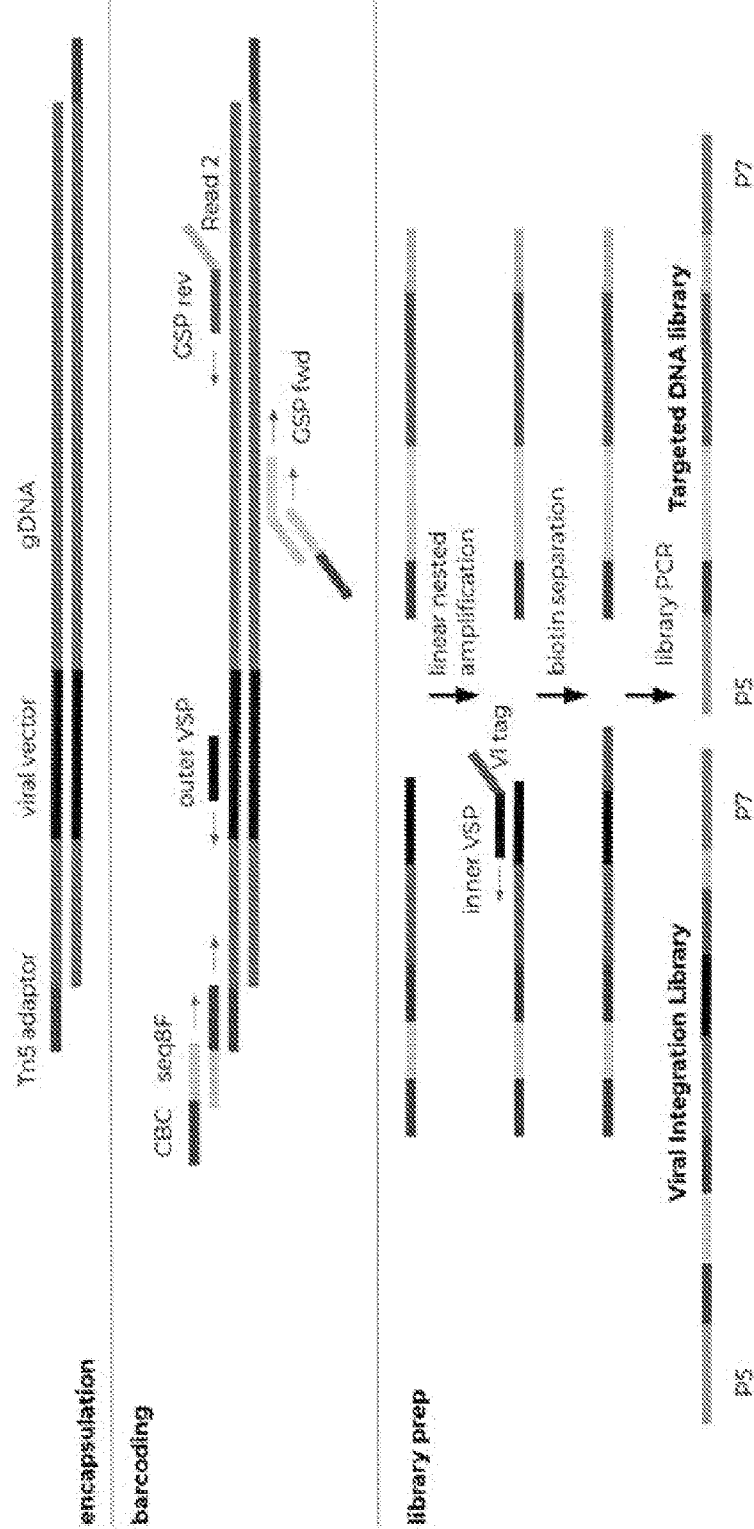


FIG. 11

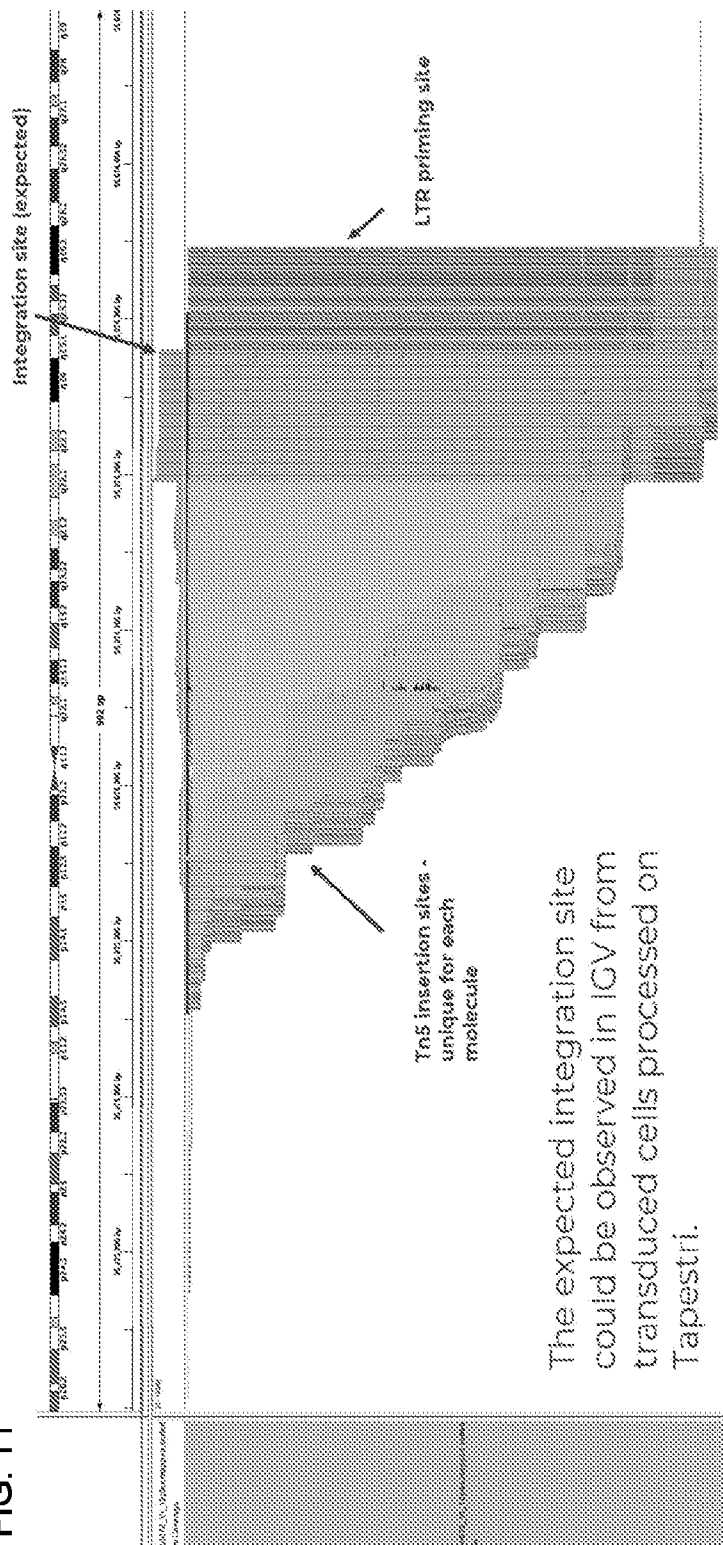


FIG. 12

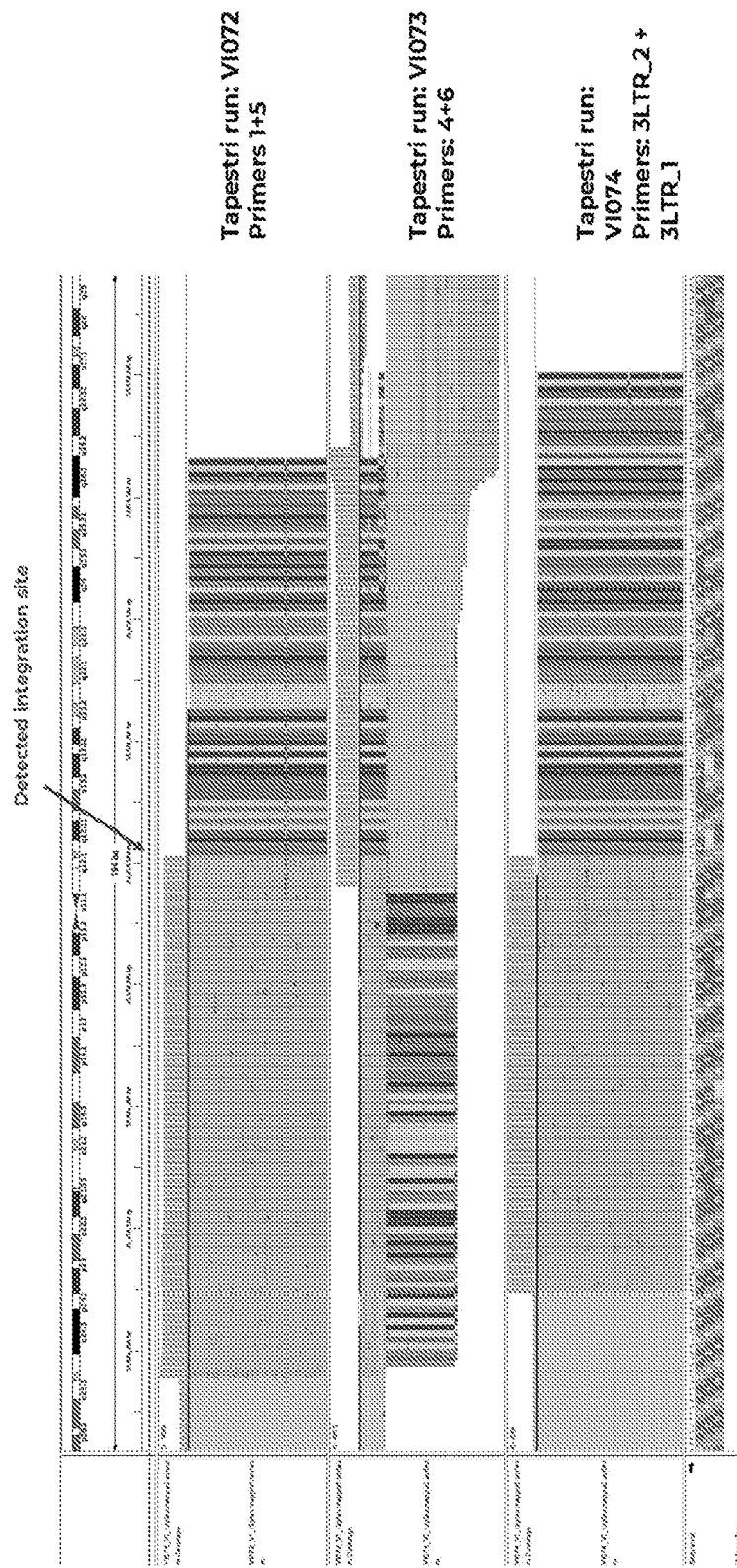


FIG. 13A

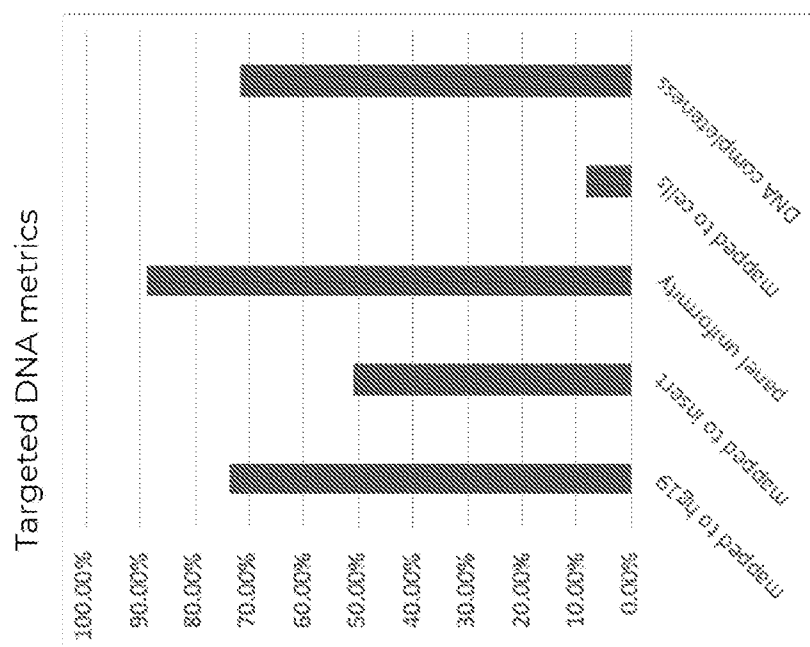


FIG. 13B

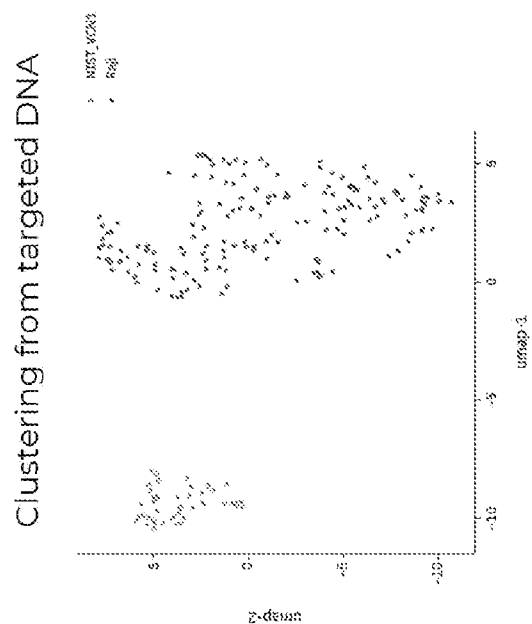


FIG. 13C

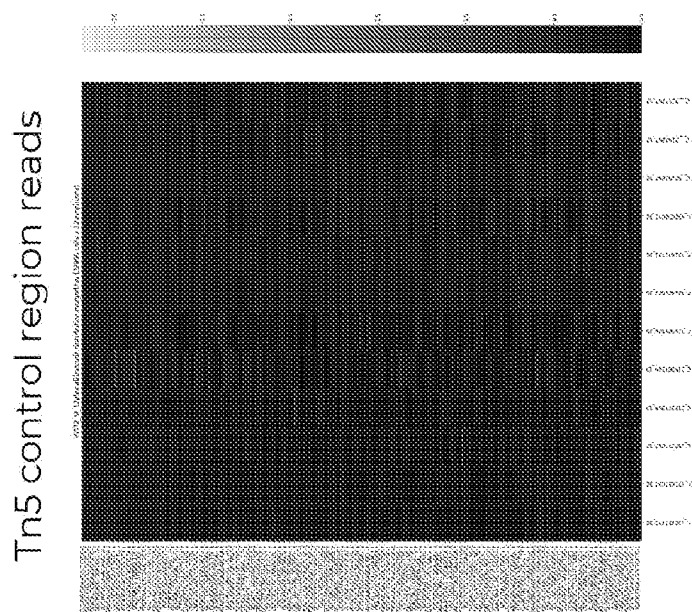


FIG. 14

V1080__25_MYE_V1_fixed (VCN2_1)

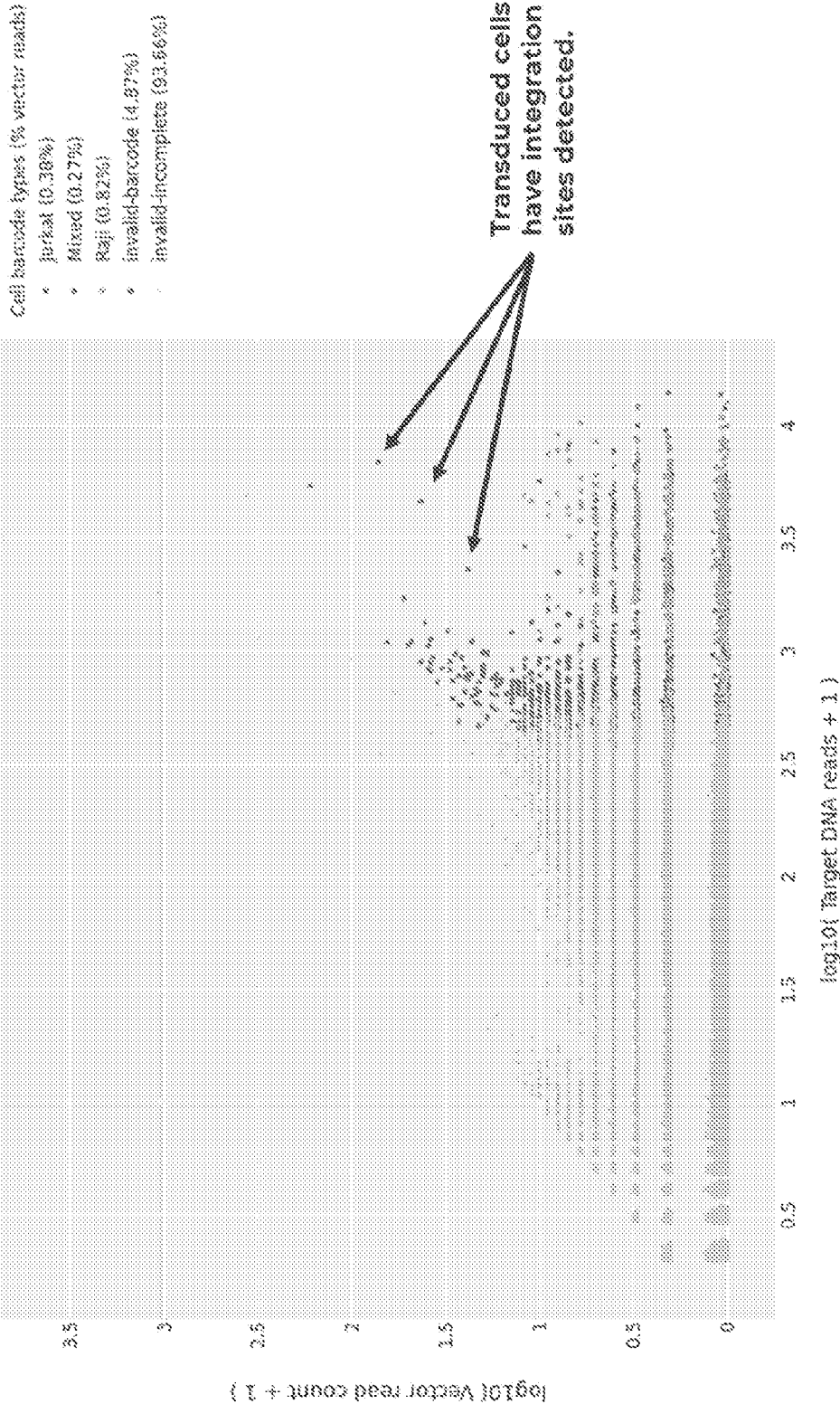


FIG. 15

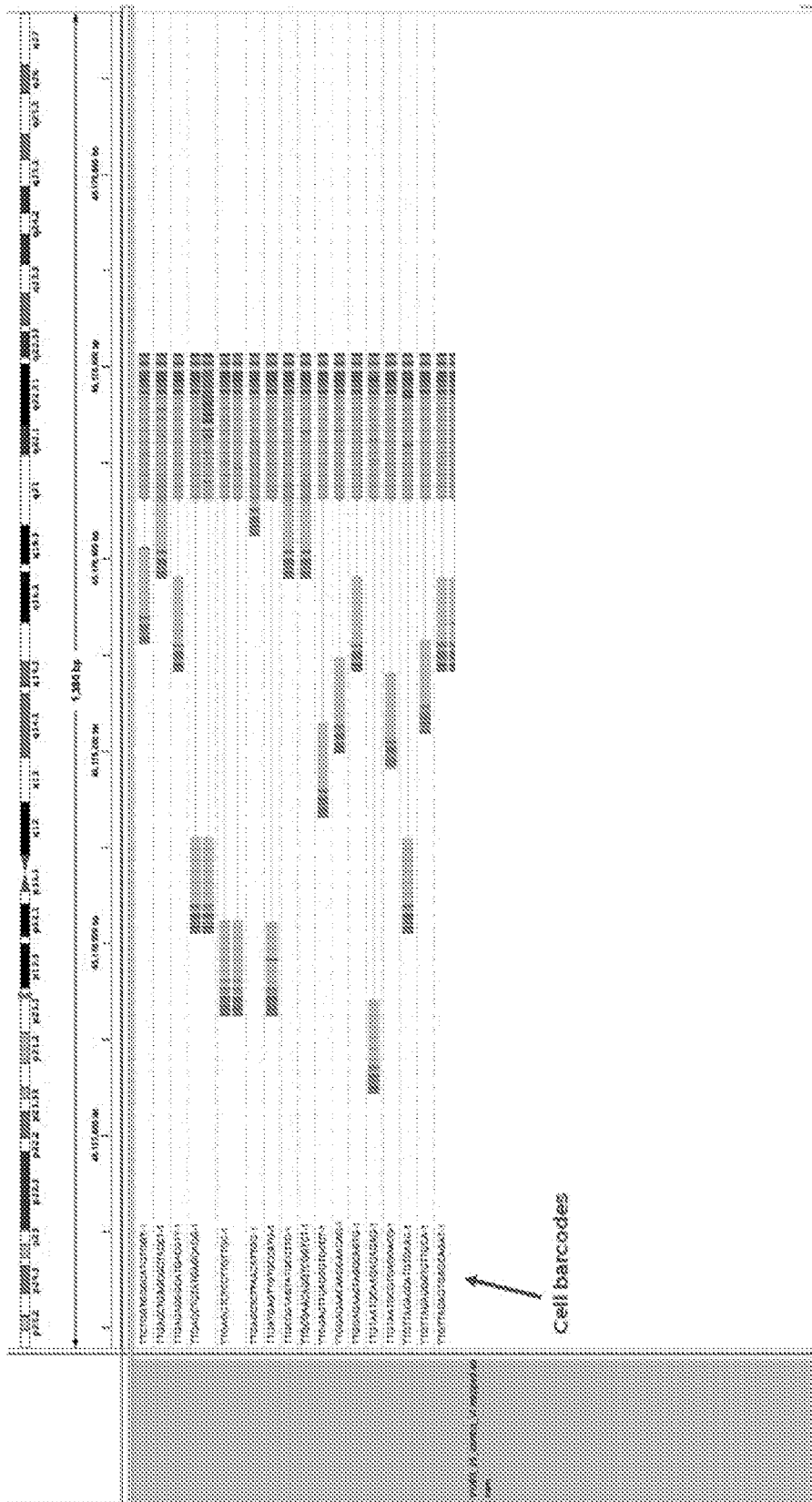


FIG. 16

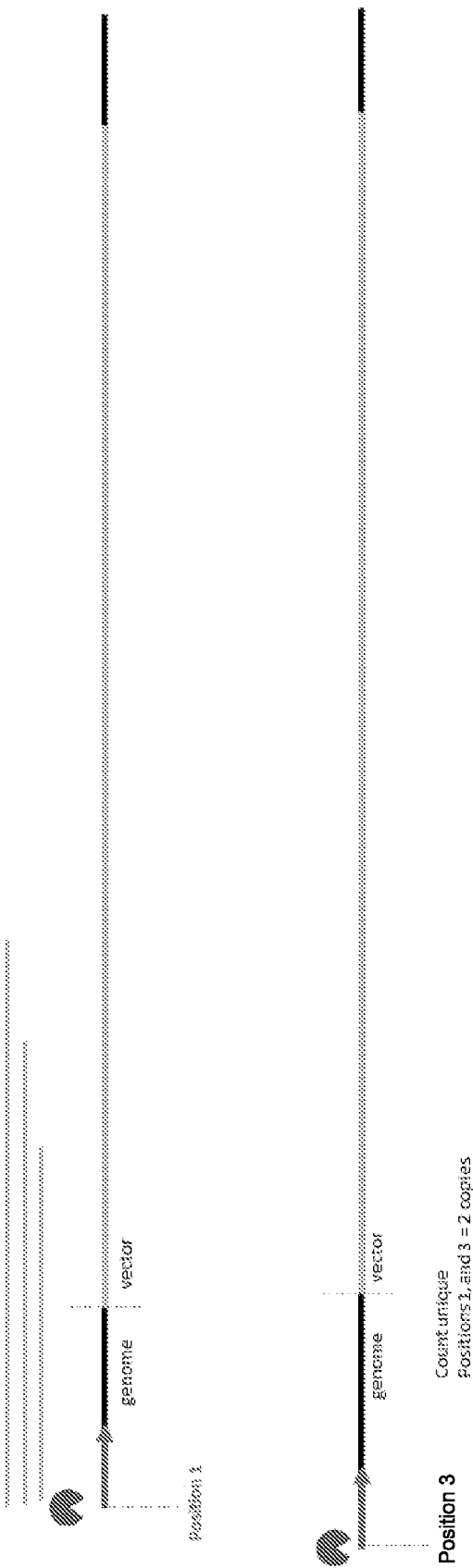


FIG. 17A

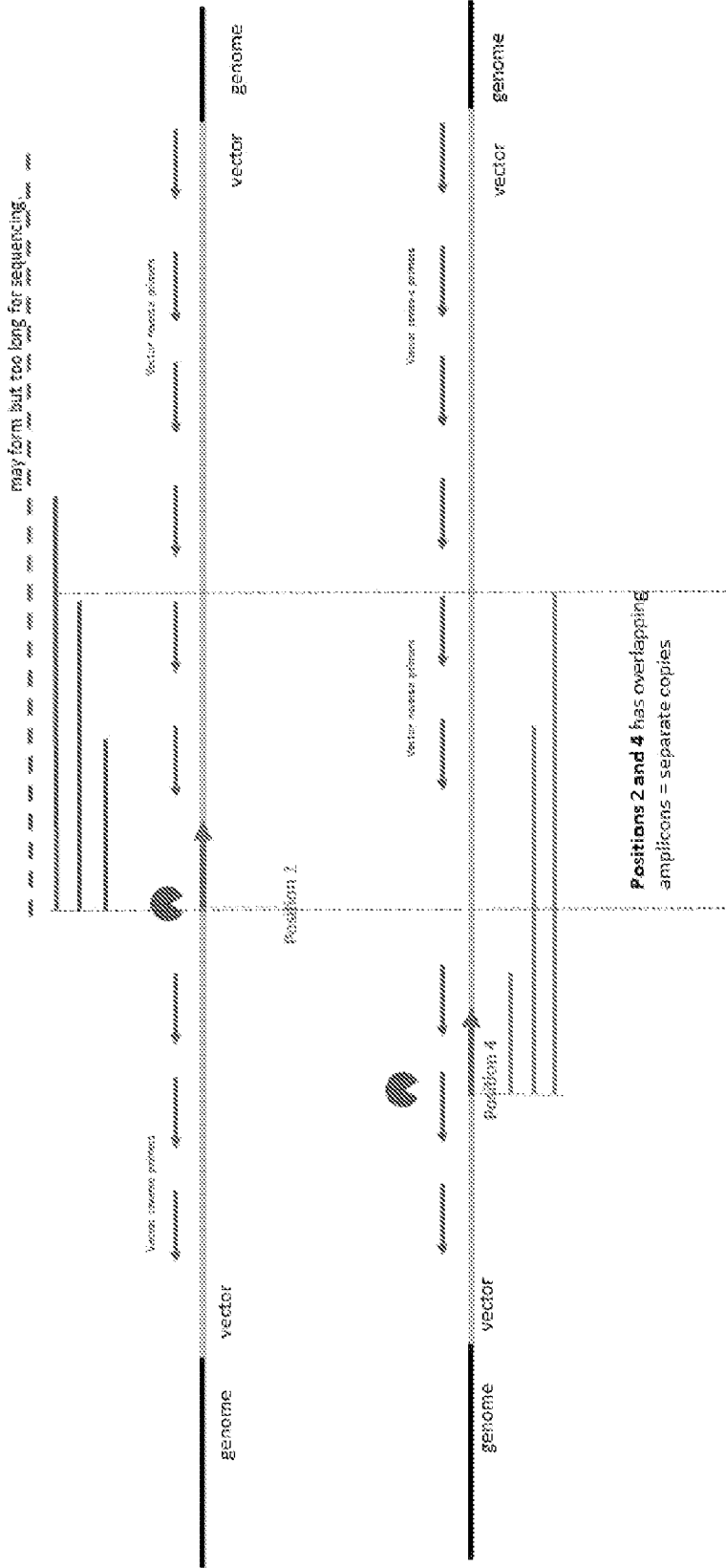


FIG. 17B

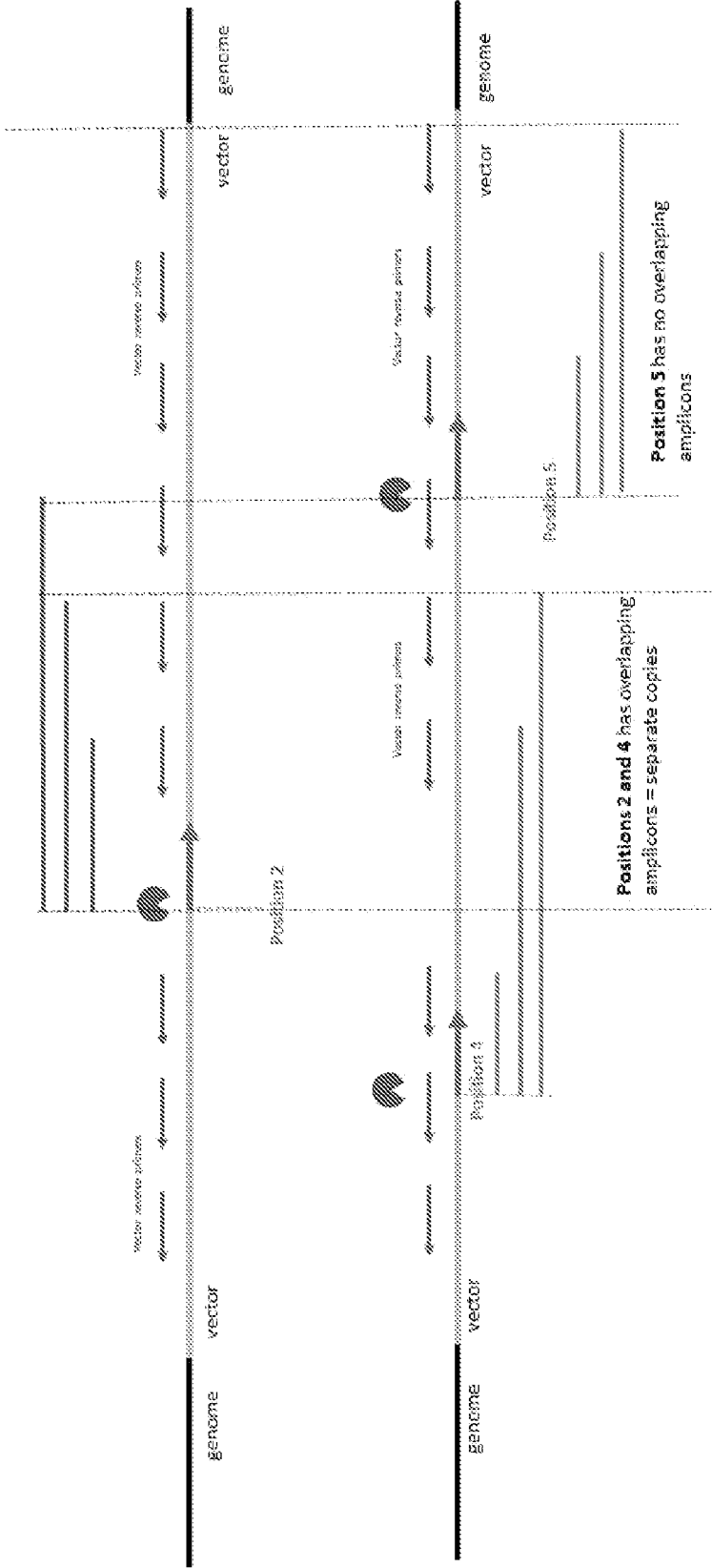


FIG. 18

