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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 August 2021 (19.08.2021)

WIPO | PCT

(10) International Publication Number
WO 2021/163611 A1

(51) International Patent Classification:
G16B 20/30 (2019.01) *C12Q 1/68* (2018.01)

(21) International Application Number:
PCT/US2021/018020

(22) International Filing Date:
12 February 2021 (12.02.2021)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

62/976,270	13 February 2020 (13.02.2020)	US
62/979,986	21 February 2020 (21.02.2020)	US
63/114,378	16 November 2020 (16.11.2020)	US
63/125,331	14 December 2020 (14.12.2020)	US

(71) Applicant: **10X GENOMICS, INC.** [US/US]; 6230 Stoneridge Mall Road, Pleasanton, California 94588-3260 (US).

(72) Inventors: **MAHESHWARI, Arundhati Shamoni**; c/o 6230 Stoneridge Mall Road, Pleasanton, California 94588 (US). **WANG, Li**; c/o 6230 Stoneridge Mall Road, Pleasanton, California 94588 (US).

(74) Agent: **HO, Cheuk Hei**; WILSON SONSINI GOODRICH & ROSATI, PC, 650 Page Mill Road, Palo Alto, California 94304 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

WO 2021/163611 A1

(54) **Title:** METHODS FOR CHARACTERIZING CELLS USING GENE EXPRESSION AND CHROMATIN ACCESSIBILITY

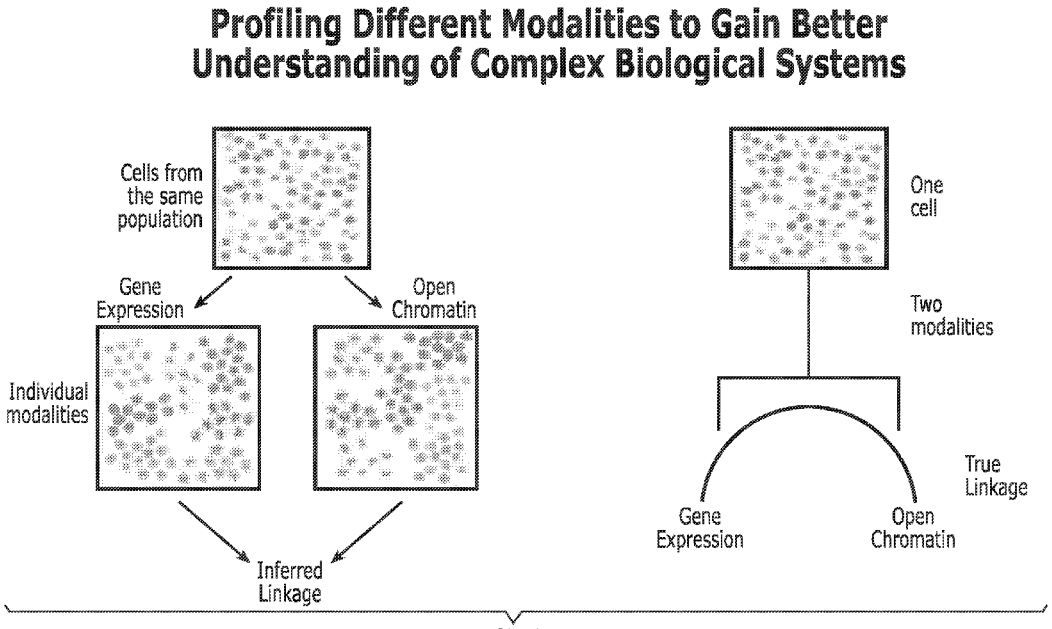


FIG. 24

(57) **Abstract:** While robust high-throughput systems for assaying either transcription or chromatin accessibility (e.g., using ATAC-seq) at single cell resolution are now widespread, researchers and clinicians typically must split cell samples and analyze each modality separately and computationally infer linkages between the gene expression and chromatin accessibility data. The present disclosure provides a high-throughput solution that simultaneously measures gene expression and chromatin accessibility from single cells or nuclei and methods of using this data to directly infer linkages between gene expression and chromatin accessibility data to chart regulatory pathways and functional characterization of complex disease pathology.

[Continued on next page]

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

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

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

METHODS FOR CHARACTERIZING CELLS USING GENE EXPRESSION AND CHROMATIN ACCESSIBILITY

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Patent Application No. 62/976,270, filed February 13, 2020, U.S. Provisional Patent Application No. 62/979,986, filed February 21, 2020, U.S. Provisional Patent Application No. 63/114,378, filed November 16, 2020, and U.S. Provisional Patent Application No. 63/125,331, filed December 14, 2020, each of which applications is herein incorporated by reference in its entirety for all purposes.

BACKGROUND

[0002] While robust high-throughput systems for assaying either transcription or chromatin accessibility (e.g., using ATAC-seq) at single cell resolution are now widespread, researchers and clinicians typically must split cell samples and analyze each modality separately and computationally infer linkages between the gene expression and chromatin accessibility data.

SUMMARY

[0003] Disclosed herein, in some embodiments, are methods for characterizing cells or cell nuclei. In an aspect, the present disclosure provides a method for characterizing cells or cell nuclei, comprising: providing a plurality of partitions comprising a plurality of cells or cell nuclei and a plurality of particles, wherein a partition of the plurality of partitions comprises a cell or cell nucleus of the plurality of cells or cell nuclei and a particle of the plurality of particles, wherein (i) the plurality of cells or cell nuclei comprises a plurality of nucleic acid molecules, wherein the plurality of nucleic acid molecules comprises a plurality of ribonucleic acid (RNA) molecules and a plurality of deoxyribonucleic acid (DNA) molecules; and (ii) the plurality of particles comprises a plurality of nucleic acid barcode molecules coupled thereto, wherein a nucleic acid barcode molecule of the plurality of nucleic acid barcode molecules comprises a nucleic acid barcode sequence of a plurality of nucleic acid barcode sequences, and wherein the particle comprises a unique nucleic acid barcode sequence of the plurality of nucleic acid barcode sequences; within the plurality of partitions, using nucleic acid barcode molecules of the plurality of nucleic acid barcode molecules and nucleic acid molecules of the plurality of nucleic acid molecules to generate a plurality of barcoded nucleic acid molecules, wherein the plurality of barcoded nucleic acid molecules comprises (i) a first subset comprising sequences corresponding to RNA molecules of the plurality of RNA molecules and (ii) a second subset

comprising sequences corresponding to DNA molecules of the plurality of DNA molecules, wherein a barcoded nucleic acid molecule of the plurality of barcoded nucleic acid molecules comprises a sequence corresponding to a nucleic acid barcode sequence of the plurality of nucleic acid barcode sequences; processing the plurality of barcoded nucleic acid molecules, or derivatives thereof, to generate sequence information corresponding to the RNA molecules and the DNA molecules; and using the sequence information to identify characteristics of the plurality of cells or cell nuclei. In some embodiments, the plurality of cells or cell nuclei comprise cell types. In some embodiments, the cell types are selected from the group consisting of monocytes, natural killer cells, B cells, T cells, granulocytes, dendritic cells, and stromal cells. In some embodiments, the B cells are selected from the group consisting of replicating B cells, normal B cells, and tumor B cells. In some embodiments, the B cells are selected from the group consisting of naïve B cells, memory B cells, plasmablast B cells, lymphoplasmacytoid cells, B-1 cells, regulatory B cells, and plasma B cells. In some embodiments, the T cells are selected from the group consisting of replicating T cells and normal T cells. In some embodiments, the T cells are selected from the group consisting of helper T cells, cytotoxic T cells, memory T cells, regulatory T cells, natural killer T cells, mucosal associated invariant T (MAIT) cells, gamma delta T cells, effector T cells, and naïve T cells. In some embodiments, the monocytes are selected from the group consisting of monocytes characterized by high level expression of CD14 cell surface receptors and monocytes characterized by high level expression of CD16 cell surface receptors. In some embodiments, the dendritic cells are selected from conventional dendritic cells and plasmacytoid dendritic cells.

[0004] In some embodiments, the sequences corresponding to the DNA molecules of the plurality of DNA molecules correspond to regions of accessible chromatin. In some embodiments, the RNA molecules of the plurality of RNA molecules comprise messenger RNA (mRNA) molecules. In some embodiments, the sequence information comprises a first plurality of sequencing reads corresponding to the DNA molecules and a second plurality of sequencing reads corresponding to the RNA molecules. In some embodiments, the sequence information comprises a plurality of sequencing reads associated with individual cells or cell nuclei of the plurality of cells or cell nuclei. In some embodiments, (d) comprises determining a linked signature of the cell or cell nucleus of the plurality of cells or cell nuclei using the sequence information, which linked signature of the cell or cell nucleus links a first data set comprising sequence information corresponding to DNA molecules of the cell or cell nucleus and a second data set comprising sequence information corresponding to RNA molecules of the cell or cell nucleus. In some embodiments, (d) comprises using the sequence information to cluster cells or

cell nuclei of the plurality of cells or cell nuclei by gene expression signatures and/or by regions of accessible chromatin signatures. In some embodiments, (d) comprises (i) using the sequence information to cluster cells or cell nuclei of the plurality of cells or cell nuclei by regions of accessible chromatin signatures, (ii) using the sequence information to cluster cells or cell nuclei of the plurality of cells or cell nuclei by gene expression signatures, and (iii) using the sequence information and the cells or cell nuclei clustered by the gene expression signatures to further characterize the cells or cell nuclei clustered by the regions of accessible chromatin. In some embodiments, (d) comprises (i) using the sequence information to cluster cells or cell nuclei of the plurality of cells or cell nuclei by regions of accessible chromatin signatures, (ii) using the sequence information to cluster cells or cell nuclei of the plurality of cells or cell nuclei by gene expression signatures, and (iii) using the sequence information and the cells or cell nuclei clustered by the regions of accessible chromatin signatures to further characterize the cells or cell nuclei clustered by the gene expression signatures.

[0005] In some embodiments, the plurality of cells or cell nuclei is derived from a sample comprising a tumor or suspected of comprising a tumor. In some embodiments, the method further comprises processing the sequence information corresponding to the RNA molecules and the DNA molecules with sequence information generated from a control sample. In some embodiments, the sample is derived from a bodily fluid. In some embodiments, the sample is derived from a biopsy. In some embodiments, the tumor is a B cell lymphoma tumor. In some embodiments, the method further comprises using the sequence information to identify a presence of a tumor cell or cell nucleus in the sample. In some embodiments, the method further comprises (e) using the sequence information to identify a cell type, a cell state, a tumor-specific gene expression pattern, or a tumor-specific differentially accessible region of chromatin in the sample. In some embodiments, the method further comprises, based at least in part on (e), identifying a therapeutic regimen for treatment of a subject from which the sample derives. In some embodiments, the therapeutic regimen comprises administration of a therapeutically effective amount of an agent targeting one or more targets identified in the tumor-specific gene expression pattern or the tumor-specific differentially accessible region of chromatin.

[0006] In some embodiments, the plurality of partitions comprises a plurality of droplets. In some embodiments, the plurality of cells or cell nuclei comprises a plurality of transposed nuclei. In some embodiments, the plurality of particles comprises a plurality of gel beads. In some embodiments, the plurality of nucleic acid barcode molecules are releasably coupled to the plurality of particles. In some embodiments, the nucleic acid barcode molecules of the plurality

of nucleic acid barcode molecules are releasable from the particles of the plurality of particles upon application of a stimulus.

[0007] In some embodiments, the stimulus is a chemical stimulus. In some embodiments, the stimulus comprises a reducing agent. In some embodiments, the plurality of nucleic acid barcode molecules are coupled to the plurality of particles via a plurality of labile moieties. In some embodiments, the method further comprises generating the plurality of partitions using a microfluidic device. In some embodiments, the method further comprises recovering the plurality of barcoded nucleic acid molecules from the at least the subset of the plurality of partitions. In some embodiments, the method further comprises, prior to (b), lysing or permeabilizing the plurality of cells or cell nuclei to provide access to the plurality of nucleic acid molecules therein. In some embodiments, the method further comprises, prior to (a), processing open chromatin structures of the plurality of cells or cell nuclei with a transposase to provide the plurality of DNA molecules. In some embodiments, within the at least the subset of the plurality of partitions, reverse transcribing the plurality of RNA molecules to provide a plurality of complementary DNA (cDNA) molecules.

[0008] Disclosed herein, in some embodiments, are methods for identifying a genetic feature. In an aspect, the present disclosure provides a method for identifying a genetic feature, comprising

[0009] A method for identifying a genetic feature, comprising: (a) providing a first data set corresponding to regions of accessible chromatin of a plurality of deoxyribonucleic acid (DNA) molecules of a plurality of cells or cell nuclei and a second data set corresponding to a plurality of ribonucleic acid (RNA) molecules of the plurality of cells or cell nuclei, wherein the first data set comprises a first plurality of sequencing reads corresponding to sequences of the regions of accessible chromatin and a plurality of nucleic acid barcode sequences, and wherein the second data set comprises a second plurality of sequencing reads corresponding to sequences of the plurality of RNA molecules and the plurality of nucleic acid barcode sequences, wherein a cell or cell nucleus of the plurality of cells or cell nuclei corresponds to a nucleic acid barcode sequence of the plurality of nucleic acid barcode sequences; (b) using the plurality of nucleic acid barcode sequences of the first data set and the second data set to identify first sequencing reads of the first plurality of sequencing reads and second sequencing reads of the second plurality of sequencing reads as corresponding to cells or cell nuclei of the plurality of cells or cell nuclei, thereby generating a third data set comprising sequence information corresponding to regions of accessible chromatin and RNA molecules associated with cells or cell nuclei of the plurality of cells or cell nuclei; (c) using the sequence information to identify cell types of the

cells or cell nuclei; (d) using the sequence information corresponding to the RNA molecules to identify an expressed protein of a cell type of the cell types; and (e) using the sequence information corresponding to the regions of accessible chromatin to identify a genetic feature corresponding to the expressed protein.

[0010] In some embodiments, the cell types are selected from the group consisting of monocytes, natural killer cells, B cells, T cells, granulocytes, dendritic cells, and stromal cells. In some embodiments, the B cells are selected from the group consisting of replicating B cells, normal B cells, and tumor B cells. In some embodiments, the B cells are selected from the group consisting of naïve B cells, memory B cells, plasmablast B cells, lymphoplasmacytoid cells, B-1 cells, regulatory B cells, and plasma B cells. In some embodiments, the T cells are selected from the group consisting of replicating T cells and normal T cells. In some embodiments, the T cells are selected from the group consisting of helper T cells, cytotoxic T cells, memory T cells, regulatory T cells, natural killer T cells, mucosal associated invariant T (MAIT) cells, gamma delta T cells, effector T cells, and naïve T cells. In some embodiments, the monocytes are selected from the group consisting of monocytes characterized by high level expression of CD14 cell surface receptors and monocytes characterized by high level expression of CD16 cell surface receptors. In some embodiments, the dendritic cells are selected from conventional dendritic cells and plasmacytoid dendritic cells. In some embodiments, the plurality of cells or cell nuclei comprises at least 500 cells or cell nuclei. In some embodiments, the plurality of cells or cell nuclei comprises at least 1,000 cells or cell nuclei. In some embodiments, the plurality of cells or cell nuclei comprises at least 10,000 cells or cell nuclei.

[0011] In some embodiments, the genetic feature is a cis-regulatory element. In some embodiments, the cis-regulatory element is a promoter. In some embodiments, the cis-regulatory element is an enhancer. In some embodiments, the expressed protein is a cytokine. In some embodiments, the plurality of RNA molecules comprises a plurality of messenger RNA (mRNA) molecules. In some embodiments, the method further comprises determining a linked signature of the cell or cell nucleus of the plurality of cells or cell nuclei using the first data set and the second data set, which linked signature of the cell or cell nucleus links a fourth data set comprising sequence information corresponding to DNA molecules of the cell or cell nucleus and a fifth data set comprising sequence information corresponding to RNA molecules of the cell or cell nucleus. In some embodiments, (c) comprises using the sequence information to cluster cells or cell nuclei of the plurality of cells or cell nuclei by regions of accessible chromatin signatures. In some embodiments, (c) comprises using the sequence information to cluster cells or cell nuclei of the plurality of cells or cell nuclei by gene expression signatures. In

some embodiments, (c) comprises (i) using the sequence information to cluster cells or cell nuclei of the plurality of cells or cell nuclei by regions of accessible chromatin signatures, (ii) using the sequence information to cluster cells or cell nuclei of the plurality of cells or cell nuclei by gene expression signatures, and (iii) using the sequence information and the cells or cell nuclei clustered by the gene expression signatures to further characterize the cells or cell nuclei clustered by the regions of accessible chromatin. In some embodiments, (c) comprises (i) using the sequence information to cluster cells or cell nuclei of the plurality of cells or cell nuclei by regions of accessible chromatin signatures, (ii) using the sequence information to cluster cells or cell nuclei of the plurality of cells or cell nuclei by gene expression signatures, and (iii) using the sequence information and the cells or cell nuclei clustered by the regions of accessible chromatin signatures to further characterize the cells or cell nuclei clustered by the gene expression signatures.

[0012] In some embodiments, the plurality of cells or cell nuclei is derived from a sample comprising a tumor or suspected of comprising a tumor. In some embodiments, the sample is derived from a bodily fluid. In some embodiments, the sample is derived from a biopsy. In some embodiments, the tumor is a B cell lymphoma tumor. In some embodiments, the method further comprises using the sequence information to identify a presence of a tumor cell or cell nucleus in the sample. In some embodiments, the method further comprises (f) using the sequence information to identify a cell type, a cell state, a tumor-specific gene expression pattern, or a tumor-specific differentially accessible region of chromatin in the sample. In some embodiments, the method further comprises, based at least in part on (f), identifying a therapeutic regimen for treatment of a subject from which the sample derives. In some embodiments, the therapeutic regimen comprises administration of a therapeutically effective amount of an agent targeting one or more targets identified in the tumor-specific gene expression pattern or the tumor-specific differentially accessible region of chromatin.

[0013] Disclosed herein, in some embodiments, are systems for identifying a genetic feature. In an aspect, the present disclosure provides a system for identifying a genetic feature, comprising: one or more databases comprising a first data set corresponding to regions of accessible chromatin of a plurality of deoxyribonucleic acid (DNA) molecules of a plurality of cells or cell nuclei and a second data set corresponding to a plurality of ribonucleic acid (RNA) molecules of the plurality of cells or cell nuclei, wherein the first data set comprises a first plurality of sequencing reads corresponding to sequences of the regions of accessible chromatin and a plurality of nucleic acid barcode sequences, and wherein the second data set comprises a second plurality of sequencing reads corresponding to sequences of the plurality of RNA

molecules and the plurality of nucleic acid barcode sequences, wherein a cell or cell nucleus of the plurality of cells or cell nuclei corresponds to a nucleic acid barcode sequence of the plurality of nucleic acid barcode sequences; and one or more computer processors operably coupled to the one or more databases, wherein the one or more computer processors are individually or collectively programmed to: (i) use the plurality of nucleic acid barcode sequences of the first data set and the second data set to identify first sequencing reads of the first plurality of sequencing reads and second sequencing reads of the second plurality of sequencing reads as corresponding to cells or cell nuclei of the plurality of cells or cell nuclei, thereby generating a third data set comprising sequence information corresponding to regions of accessible chromatin and RNA molecules associated with cells or cell nuclei of the plurality of cells or cell nuclei; (ii) use the sequence information to identify cell types of the cells or cell nuclei; (iii) use the sequence information corresponding to the RNA molecules to identify an expressed protein of a cell type of the cell types; and (iv) use the sequence information corresponding to the regions of accessible chromatin to identify a genetic feature corresponding to the expressed protein.

[0014] In some embodiments, the cell types are selected from the group consisting of monocytes, natural killer cells, B cells, T cells, granulocytes, dendritic cells, and stromal cells. In some embodiments, the B cells are selected from the group consisting of replicating B cells, normal B cells, and tumor B cells. In some embodiments, the B cells are selected from the group consisting of naïve B cells, memory B cells, plasmablast B cells, lymphoplasmacytoid cells, B-1 cells, regulatory B cells, and plasma B cells. In some embodiments, the T cells are selected from the group consisting of replicating T cells and normal T cells. In some embodiments, the T cells are selected from the group consisting of helper T cells, cytotoxic T cells, memory T cells, regulatory T cells, natural killer T cells, mucosal associated invariant T (MAIT) cells, gamma delta T cells, effector T cells, and naïve T cells. In some embodiments, the monocytes are selected from the group consisting of monocytes characterized by high level expression of CD14 cell surface receptors and monocytes characterized by high level expression of CD16 cell surface receptors. In some embodiments, the dendritic cells are selected from conventional dendritic cells and plasmacytoid dendritic cells. In some embodiments, the plurality of cells or cell nuclei comprises at least 500 cells or cell nuclei. In some embodiments, the plurality of cells or cell nuclei comprises at least 1,000 cells or cell nuclei. In some embodiments, the plurality of cells or cell nuclei comprises at least 10,000 cells or cell nuclei.

[0015] In some embodiments, the genetic feature is a cis-regulatory element. In some embodiments, the cis-regulatory element is a promoter. In some embodiments, the cis-regulatory element is an enhancer. In some embodiments, the expressed protein is a cytokine. In some

embodiments, the plurality of RNA molecules comprises a plurality of messenger RNA (mRNA) molecules. In some embodiments, the one or more computer processors are individually or collectively programmed to determine a linked signature of the cell or cell nucleus of the plurality of cells or cell nuclei using the first data set and the second data set, which linked signature of the cell or cell nucleus links a fourth data set comprising sequence information corresponding to DNA molecules of the cell or cell nucleus and a fifth data set comprising sequence information corresponding to RNA molecules of the cell or cell nucleus.

[0016] In some embodiments, the one or more computer processors are individually or collectively programmed to use the sequence information in (ii) to cluster cells or cell nuclei of the plurality of cells or cell nuclei by regions of accessible chromatin signatures. In some embodiments, in (ii), the one or more computer processors are individually or collectively programmed to use the sequence information in (ii) to cluster cells or cell nuclei of the plurality of cells or cell nuclei by gene expression signatures. In some embodiments, , in (ii), the one or more computer processors are individually or collectively programmed to use the sequence information (1) to cluster cells or cell nuclei of the plurality of cells or cell nuclei by regions of accessible chromatin signatures, (2) to cluster cells or cell nuclei of the plurality of cells or cell nuclei by gene expression signatures, and (3) with the cells or cell nuclei clustered by the gene expression signatures to further characterize the cells or cell nuclei clustered by the regions of accessible chromatin. In some embodiments, in (ii) the one or more computer processors are individually or collectively programmed to use the sequence information (1) to cluster cells or cell nuclei of the plurality of cells or cell nuclei by regions of accessible chromatin signatures, (2) to cluster cells or cell nuclei of the plurality of cells or cell nuclei by gene expression signatures, and (3) with the cells or cell nuclei clustered by the regions of accessible chromatin signatures to further characterize the cells or cell nuclei clustered by the gene expression signatures. In some embodiments, the one or more computer processors are individually or collectively further programmed to generate an output that relates to detecting a disease or condition in the sample, the output comprising the regions of accessible chromatin signatures and gene expression signatures, presence or absence of the disease or condition; or a level of progression of the disease or condition.

[0017] In some embodiments, the plurality of cells or cell nuclei is derived from a sample comprising a tumor or suspected of comprising a tumor. In some embodiments, the sample is derived from a bodily fluid. In some embodiments, the sample is derived from a biopsy. In some embodiments, the tumor is a B cell lymphoma tumor. In some embodiments, the one or more computer processors are individually or collectively further programmed to use the sequence

information to identify a presence of a tumor cell or cell nucleus in the sample. In some embodiments, the one or more computer processors are individually or collectively further programmed to compare the sequence information to sequence information from a control sample.

[0018] In some embodiments, the one or more computer processors are individually or collectively further programmed to use the sequence information to identify a cell type, a cell state, a tumor-specific gene expression pattern, or a tumor-specific differentially accessible region of chromatin in the sample. In some embodiments, the one or more computer processors are individually or collectively further programmed to, based at least in part on the use of the sequence information to identify the cell type, the cell state, the tumor-specific gene expression pattern, or the tumor-specific differentially accessible region of chromatin in the sample, identify a therapeutic regimen for treatment of a subject from which the sample derives. In some embodiments, the therapeutic regimen comprises administration of a therapeutically effective amount of an agent targeting one or more targets identified in the tumor-specific gene expression pattern or the tumor-specific differentially accessible region of chromatin. In some embodiments, the system is used to monitor efficacy of the therapeutic regimen for treatment.

[0019] Disclosed herein, in some embodiments, are methods for determining a condition of a sample. In an aspect, the present disclosure provides a method for determining a condition of a sample, comprising: generating (i) a first data set comprising sequencing information corresponding to regions of accessible chromatin of a plurality of deoxyribonucleic acid (DNA) molecules of a cell or cell nucleus of the sample, (ii) a second data set comprising sequencing information corresponding to a plurality of ribonucleic acid (RNA) molecules of the cell or cell nucleus, and (iii) a linked signature of the cell or cell nucleus using the first data set and the second data set; using the linked signature of the cell or cell nucleus and a control linked signature of a control cell or cell nucleus of a control sample to determine one or more regions of accessible chromatin of the plurality of DNA molecules or one or more genes expressed from the plurality of RNA molecules indicative of the condition.

[0020] In some embodiments, the method further comprises c) determining a level of the one or more regions of accessible chromatin and/or the one or more genes expressed determined in b) indicative of the condition in one or more samples of an individual suspected of having the condition. In some embodiments, the method further comprises providing a diagnostic evaluation of the condition, a prognostic evaluation of the condition, monitoring of the condition, and/or management of the condition. In some embodiments, a gene associated with the one or more regions of accessible chromatin and/or one or more genes expressed determined

in b) is identified as a target of a therapeutic regimen for treatment of the condition. In some embodiments, the method further comprises administering a therapeutically effective amount of an agent to a subject targeting the target, wherein the sample is derived from the subject. In some embodiments, the method further comprises determining an efficacy of the agent in the subject. In some embodiments, determining the efficacy comprises detecting a presence or absence of a response to the agent by the subject, wherein the response comprises a quantity, degree, or extent of response following administration of a first dose or a subsequent dose of the agent. In some embodiments, the response comprises a differential in gene expression and/or chromatin accessibility of the target between prior to and after administration of the agent.

[0021] In some embodiments, the sample is from a subject having a tumor or suspected of having a tumor. In some embodiments, the condition is a tumor, cancer, malignancy, neoplasm, or other proliferative disease or disorder. In some embodiments, the condition is a B cell malignancy. In some embodiments, the B cell malignancy is B cell lymphoma. In some embodiments, the sample is derived from a bodily fluid. In some embodiments, the sample is derived from a biopsy. In some embodiments, in the method for determining a condition of a sample, a) comprises providing a plurality of sequencing reads corresponding to sequences of the plurality of DNA molecules and the plurality of RNA molecules, wherein the sequencing reads each correspond to the cell or cell nucleus via a nucleic acid barcode sequence. In some embodiments, the method further comprises, prior to b), clustering a plurality of cells or cell nuclei of the sample by respective regions of accessible chromatin signatures, by respective genes expressed, and/or by respective linked signatures of the plurality of cells or cell nuclei.

[0022] In some embodiments, the plurality of cells or cell nuclei are clustered by cell types selected from the group consisting of monocytes, natural killer cells, B cells, T cells, granulocytes, dendritic cells, and stromal cells. In some embodiments, the B cells are selected from the group consisting of replicating B cells, normal B cells, and tumor B cells. In some embodiments, the B cells are selected from the group consisting of naïve B cells, memory B cells, plasmablast B cells, lymphoplasmacytoid cells, B-1 cells, regulatory B cells, and plasma B cells. In some embodiments, the T cells are selected from the group consisting of replicating T cells and normal T cells. In some embodiments, the T cells are selected from the group consisting of helper T cells, cytotoxic T cells, memory T cells, regulatory T cells, natural killer T cells, mucosal associated invariant T (MAIT) cells, gamma delta T cells, effector T cells, and naïve T cells.

[0023] In some embodiments, the method further comprises c) monitoring a level of the one or more regions of accessible chromatin and/or the one or more genes expressed determined in

b) indicative of the condition in the individual. In some embodiments, the method further comprises, prior to step a), generating a plurality of tagged DNA fragments. In some embodiments, the method further comprises, prior to a), generating a plurality of barcoded nucleic acid molecules, wherein the plurality of barcoded nucleic acid molecules comprises (i) a first subset comprising sequences corresponding to regions of accessible chromatin of the plurality of deoxyribonucleic acid (DNA) molecules of the cell or cell nucleus of the sample and (ii) a second subset comprising sequences corresponding to the ribonucleic acid (RNA) molecules of the cell or cell nucleus.

[0024] In some embodiments, the generating is performed within a plurality of partitions. In some embodiments, the method further comprises sequencing the plurality of barcoded nucleic acid molecules. In some embodiments, the first data set is generated by sequencing a first plurality of barcoded nucleic acid molecules comprising sequences corresponding to regions of accessible chromatin of the plurality of deoxyribonucleic acid (DNA) molecules of the cell or cell nucleus of the sample; and the second data set is generated by sequencing a second plurality of barcoded nucleic acid molecules comprising sequences of the ribonucleic acid (RNA) molecules of the cell or cell nucleus.

[0025] Disclosed herein, in some embodiments, are *in vitro* methods of preparing a biological sample. In an aspect, the present disclosure provides an in vitro method of preparing a biological sample, comprising: (a) processing open chromatin structures of T-cells and/or B-cells from the biological sample with a transposase to provide a plurality of DNA molecules; (b) generating a first plurality of barcoded nucleic acid molecules comprising the plurality of DNA molecules processed in (a); (c) generating a second plurality of barcoded nucleic acid molecules comprising a plurality of nucleic acids comprising mRNA sequences or derivatives thereof from the T-cells and/or B-cells from the biological sample; and (d) generating a first and second sequencing library from the first and second plurality of barcoded nucleic acid molecules, respectively, to determine a linked signature of a cell of the T-cells and/or B-cells.

[0026] In some embodiments, wherein step (b) and/or step (c) is performed within a plurality of partitions. In some embodiments, the method further comprises determining a significance level for the linked signature determined in step (d). In some embodiments, wherein step (c) comprises reverse transcribing the plurality of mRNA sequences from the T-cells and/or B-cells from the biological sample to provide a plurality of complementary DNA (cDNA) molecules, and the second plurality of barcoded nucleic acid molecules comprises the cDNA molecules. In some embodiments, wherein step (c) comprises barcoding 3' ends of the mRNA were barcoded. In some embodiments, the method further comprises encapsulating single nuclei of the T-cells

and/or B-cells in droplets prior to step (b). In some embodiments, the method further comprises (e) determining from the first and second sequencing libraries a presence, absence, and/or level of the one or more linked signatures correlated with a condition. In some embodiments, the condition is a tumor, cancer, malignancy, neoplasm, or other proliferative disease or disorder. In some embodiments, the condition is a B cell malignancy. In some embodiments, the B cell malignancy is B cell lymphoma. In some embodiments, the linked signature of a cell or cell nucleus of the T-cells and/or B-cells is compared to a control linked signature of a control cell or cell nucleus of a control sample. In some embodiments, the method comprises providing the biological sample isolated and obtained from an individual. In some aspects, the method further comprises obtaining the biological sample from the individual. In some aspects, the method may, but need not, comprise the further step of obtaining the biological sample from the individual. In some embodiments, the method is performed *ex vivo*. In some embodiments, the method further comprises providing the biological sample isolated and obtained from an individual prior to step (a).

[0027] Disclosed herein, in some embodiments, are systems for characterizing cells. In an aspect, the present disclosure provides a system for characterizing cells, comprising: a plurality of partitions comprising a plurality of cells or cell nuclei and a plurality of particles, wherein a partition of the plurality of partitions comprises a cell or cell nucleus of the plurality of cells or cell nuclei and a particle of the plurality of particles, wherein (i) the plurality of cells or cell nuclei comprises a plurality of nucleic acid molecules, wherein the plurality of nucleic acid molecules comprises a plurality of RNA molecules and a plurality of DNA molecules; and (ii) the plurality of particles comprises a plurality of nucleic acid barcode molecules coupled thereto, wherein a nucleic acid barcode molecule of the plurality of nucleic acid barcode molecules comprises a nucleic acid barcode sequence of a plurality of nucleic acid barcode sequences, and wherein the particle comprises a unique nucleic acid barcode sequence of the plurality of nucleic acid barcode sequences; and one or more computer processors, individually or collectively, programmed to; (a) process a plurality of barcoded nucleic acid molecules, generated in the plurality of partitions using the plurality of nucleic acid barcode molecules and the plurality of nucleic acid molecules, or derivatives thereof, to generate sequence information corresponding to the RNA molecules and the DNA molecules; and (b) use the sequence information to identify characteristics of the plurality of cells or cell nuclei.

[0028] In some embodiments, the characteristics of the plurality of cells or cell nuclei comprise cell types. In some embodiments, the cell types are selected from the group consisting of monocytes, natural killer cells, B cells, T cells, granulocytes, dendritic cells, and stromal cells.

In some embodiments, the B cells are selected from the group consisting of replicating B cells, normal B cells, and tumor B cells. In some embodiments, the B cells are selected from the group consisting of naïve B cells, memory B cells, plasmablast B cells, lymphoplasmacytoid cells, B-1 cells, regulatory B cells, and plasma B cells. In some embodiments, the T cells are selected from the group consisting of replicating T cells and normal T cells. In some embodiments, the T cells are selected from the group consisting of helper T cells, cytotoxic T cells, memory T cells, regulatory T cells, natural killer T cells, mucosal associated invariant T (MAIT) cells, gamma delta T cells, effector T cells, and naïve T cells. In some embodiments, the monocytes are selected from the group consisting of monocytes characterized by high level expression of CD14 cell surface receptors and monocytes characterized by high level expression of CD16 cell surface receptors. In some embodiments, the dendritic cells are selected from conventional dendritic cells and plasmacytoid dendritic cells. In some embodiments, the sequences corresponding to the DNA molecules of the plurality of DNA molecules correspond to regions of accessible chromatin. In some embodiments, the RNA molecules of the plurality of RNA molecules comprise messenger RNA (mRNA) molecules. In some embodiments, the sequence information comprises a first plurality of sequencing reads corresponding to the DNA molecules and a second plurality of sequencing reads corresponding to the RNA molecules. In some embodiments, the sequence information comprises a plurality of sequencing reads associated with individual cells or cell nuclei of the plurality of cells or cell nuclei.

[0029] In some embodiments, wherein in (b) the one or more computer processors are individually or collectively programmed to: determine a linked signature of the cell or cell nucleus of the plurality of cells or cell nuclei using the sequence information, which linked signature of the cell or cell nucleus links a first data set comprising sequence information corresponding to DNA molecules of the cell or cell nucleus and a second data set comprising sequence information corresponding to RNA molecules of the cell or cell nucleus. In some embodiments, wherein in (b) the one or more computer processors are individually or collectively programmed to: use the sequence information to cluster cells or cell nuclei of the plurality of cells or cell nuclei by gene expression signatures and/or by regions of accessible chromatin signatures. In some embodiments, wherein in (b) the one or more computer processors are individually or collectively programmed to: (i) use the sequence information to cluster cells or cell nuclei of the plurality of cells or cell nuclei by regions of accessible chromatin signatures, (ii) use the sequence information to cluster cells or cell nuclei of the plurality of cells or cell nuclei by gene expression signatures, and (iii) use the sequence information and the cells or cell nuclei clustered by the gene expression signatures to further

characterize the cells or cell nuclei clustered by the regions of accessible chromatin. In some embodiments, wherein in (b) the one or more computer processors are individually or collectively programmed to: (i) use the sequence information to cluster cells or cell nuclei of the plurality of cells or cell nuclei by regions of accessible chromatin signatures, (ii) use the sequence information to cluster cells or cell nuclei of the plurality of cells or cell nuclei by gene expression signatures, and (iii) use the sequence information and the cells or cell nuclei clustered by the regions of accessible chromatin signatures to further characterize the cells or cell nuclei clustered by the gene expression signatures.

[0030] In some embodiments, the plurality of cells or cell nuclei is derived from a sample comprising a tumor or suspected of comprising a tumor. In some embodiments, the one or more computer processors are further, individually or collectively, programmed to process the sequence information corresponding to the RNA molecules and the DNA molecules with sequence information generated from a control sample. In some embodiments, the sample is derived from a bodily fluid. In some embodiments, the sample is derived from a biopsy. In some embodiments, the tumor is a B cell lymphoma tumor. In some embodiments, the one or more computer processors are further, individually or collectively, programmed to use the sequence information to identify a presence of a tumor cell or cell nucleus in the sample. In some embodiments, the one or more computer processors are further, individually or collectively, programmed to: (c) use the sequence information to identify a cell type, a cell state, a tumor-specific gene expression pattern, or a tumor-specific differentially accessible region of chromatin in the sample. In some embodiments, the one or more computer processors are further, individually or collectively, programmed to, based at least in part on (c), identify a therapeutic regimen for treatment of a subject from which the sample derives. In some embodiments, the therapeutic regimen comprises administration of a therapeutically effective amount of an agent targeting one or more targets identified in the tumor-specific gene expression pattern or the tumor-specific differentially accessible region of chromatin.

[0031] In some embodiments, the plurality of partitions comprises a plurality of droplets. In some embodiments, the plurality of cells or cell nuclei comprises a plurality of transposed nuclei. In some embodiments, the plurality of particles comprises a plurality of gel beads. In some embodiments, the plurality of nucleic acid barcode molecules is releasably coupled to the plurality of particles. In some embodiments, the nucleic acid barcode molecules of the plurality of nucleic acid barcode molecules are releasable from the particles of the plurality of particles upon application of a stimulus. In some embodiments, the stimulus is a chemical stimulus. In some embodiments, the stimulus comprises a reducing agent. In some embodiments, the

plurality of nucleic acid barcode molecules is coupled to the plurality of particles via a plurality of labile moieties. In some embodiments, the system further comprises a microfluidic device that generates the plurality of partitions.

[0032] Disclosed herein, in some embodiments, are systems for determining a condition of a sample. In an aspect, a system for determining a condition of a sample, comprises: one or more databases comprising (i) a first data set comprising sequencing information corresponding to regions of accessible chromatin of a plurality of deoxyribonucleic acid (DNA) molecules of a cell or cell nucleus of the sample, (ii) a second data set comprising sequencing information corresponding to a plurality of ribonucleic acid (RNA) molecules of the cell or cell nucleus, and (iii) a linked signature of the cell or cell nucleus using the first data set and the second data set; one or more computer processors operably coupled to the one or more databases, wherein the one or more computer processors are individually or collectively programmed to use the linked signature of the cell or cell nucleus and a control linked signature of a control cell or cell nucleus of a control sample to determine one or more regions of accessible chromatin of the plurality of DNA molecules or one or more genes expressed from the plurality of RNA molecules indicative of the condition.

[0033] In some embodiments, the one or more computer processors are individually or collectively programmed to determine a level of the one or more regions of accessible chromatin and/or the one or more genes expressed indicative of the condition in one or more samples of an individual suspected of having the condition. In some embodiments, the one or more computer processors are individually or collectively programmed to generate an output that relates to providing a diagnostic evaluation of the condition, a prognostic evaluation of the condition, monitoring of the condition, and/or management of the condition. In some embodiments, the one or more computer processors are individually or collectively configured to identify a gene associated with the one or more regions of accessible chromatin and/or the one or more genes expressed as a target of a therapeutic regimen for treatment of the condition.

[0034] In some embodiments, the one or more computer processors are individually or collectively programmed to generate an output that relates to determining an administration regimen of a therapeutically effective amount of an agent to a subject targeting the target, wherein the sample is derived from the subject. In some embodiments, the one or more computer processors are individually or collectively programmed to generate an output that relates to determining an efficacy of an agent targeting the target when administered to a subject, wherein the sample is derived from the subject. In some embodiments, the one or more computer processors are individually or collectively programmed to generate an output that

relates to detecting a presence or absence of a response to the agent by the subject, wherein the response comprises a quantity, degree, or extent of response following administration of a first dose or a subsequent dose of the agent.

[0035] In some embodiments, the response comprises a differential in gene expression and/or chromatin accessibility of the target between prior to and after administration of the agent. In some embodiments, the sample is from a subject having a tumor or suspected of having a tumor. In some embodiments, the condition is a tumor, cancer, malignancy, neoplasm, or other proliferative disease or disorder. In some embodiments, the condition is a B cell malignancy. In some embodiments, the B cell malignancy is B cell lymphoma. In some embodiments, the sample is derived from a bodily fluid. In some embodiments, the sample is derived from a biopsy.

[0036] In some embodiments, the first data set and the second data set comprise a plurality of sequencing reads corresponding to sequences of the plurality of DNA molecules and the plurality of RNA molecules, wherein the sequencing reads each correspond to the cell or cell nucleus via a nucleic acid barcode sequence. In some embodiments, the one or more computer processors are individually or collectively programmed to cluster a plurality of cells or cell nuclei of the sample by respective regions of accessible chromatin signatures, by respective genes expressed, and/or by respective linked signatures of the plurality of cells or cell nuclei. In some embodiments, the plurality of cells or cell nuclei are clustered by cell types selected from the group consisting of monocytes, natural killer cells, B cells, T cells, granulocytes, dendritic cells, and stromal cells. In some embodiments, the B cells are selected from the group consisting of replicating B cells, normal B cells, and tumor B cells. In some embodiments, the B cells are selected from the group consisting of naïve B cells, memory B cells, plasmablast B cells, lymphoplasmacytoid cells, B-1 cells, regulatory B cells, and plasma B cells. In some embodiments, the T cells are selected from the group consisting of replicating T cells and normal T cells. In some embodiments, the T cells are selected from the group consisting of helper T cells, cytotoxic T cells, memory T cells, regulatory T cells, natural killer T cells, mucosal associated invariant T (MAIT) cells, gamma delta T cells, effector T cells, and naïve T cells.

[0037] In some embodiments, the one or more computer processors are individually or collectively programmed to monitor a level of the one or more regions of accessible chromatin and/or the one or more genes expressed indicative of the condition in the individual.

[0038] In some embodiments, the plurality of DNA fragments is tagged. In some embodiments, the first data set comprising sequencing information corresponding to regions of accessible chromatin of a plurality of DNA molecules of a cell or cell nucleus of the sample and

the second data set comprising sequencing information corresponding to a plurality of RNA molecules of the cell or cell nucleus are barcoded with barcoded nucleic acid sequences. In some embodiments, the first data set comprising sequencing information corresponding to regions of accessible chromatin of a plurality of DNA molecules of a cell or cell nucleus of the sample and the second data set comprising sequencing information corresponding to a plurality of RNA molecules of the cell or cell nucleus are barcoded with barcoded nucleic acid sequences within a plurality of partitions.

[0039] In some embodiments, the system further comprises a device or a sequencer configured to sequence the plurality of barcoded nucleic acid sequences. In some embodiments, the first data set is generated by sequencing a first plurality of barcoded nucleic acid sequences comprising sequences corresponding to regions of accessible chromatin of the plurality of deoxyribonucleic acid (DNA) molecules of the cell or cell nucleus of the sample; and the second data set is generated by sequencing a second plurality of barcoded nucleic acid sequences comprising sequences of the ribonucleic acid (RNA) molecules of the cell or cell nucleus.

[0040] In some embodiments, the one or more computer processors are individually or collectively programmed to process the first data set and/or the second data set to generate a filtered first data set and/or a filtered second data set. In some embodiments, the filtered first data set is filtered using motif enrichment. In some embodiments, the filtered second data set is filtered using differential expression analysis. In some embodiments, the one or more computer processors are individually or collectively programmed to process the first data set and/or the second data set to generate a linkage significance. In some embodiments, the one or more computer processors are individually or collectively programmed to process the first filtered data set and/or the second filtered data set to generate an enrichment score. In some embodiments, the filtered first data set and the filtered second data set is used to generate a transcription factor-target gene network. In some embodiments, wherein a gene from the transcription factor-target gene network is identified as a target of a therapeutic regimen for treatment of the condition. In some embodiments, the target is a transcription factor.

[0041] Another aspect of the present disclosure provides a non-transitory computer readable medium comprising machine executable code that, upon execution by one or more computer processors, implements any of the methods above or elsewhere herein.

[0042] Another aspect of the present disclosure provides a system comprising one or more computer processors and computer memory coupled thereto. The computer memory comprises machine executable code that, upon execution by the one or more computer processors, implements any of the methods above or elsewhere herein.

[0043] Additional aspects and advantages of the present disclosure will become readily apparent to those skilled in this art from the following detailed description, wherein only illustrative embodiments of the present disclosure are shown and described. As will be realized, the present disclosure is capable of other and different embodiments, and its several details are capable of modifications in various obvious respects, all without departing from the disclosure. Accordingly, the drawings and description are to be regarded as illustrative in nature, and not as restrictive.

[0044] A sample may be processed for various purposes, such as identification of a type of moiety within the sample. The sample may be a biological sample. Biological samples may be processed, such as for detection of a disease (e.g., cancer) or identification of a particular species. There are various approaches for processing samples, such as polymerase chain reaction (PCR) and sequencing.

[0045] Biological samples may be processed within various reaction environments, such as partitions. Partitions may be wells or droplets. Droplets or wells may be employed to process biological samples in a manner that enables the biological samples to be partitioned and processed separately. For example, such droplets may be fluidically isolated from other droplets, enabling accurate control of respective environments in the droplets.

[0046] Biological samples in partitions may be subjected to various processes, such as chemical processes or physical processes. Samples in partitions may be subjected to heating or cooling, or chemical reactions, such as to yield species that may be qualitatively or quantitatively processed.

INCORPORATION BY REFERENCE

[0047] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference. To the extent publications and patents or patent applications incorporated by reference contradict the disclosure contained in the specification, the specification is intended to supersede and/or take precedence over any such contradictory material.

BRIEF DESCRIPTION OF THE DRAWINGS

[0048] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative

embodiments, in which the principles of the invention are utilized, and the accompanying drawings (also "Figure" and "FIG." herein), of which:

[0049] **FIG. 1** shows an example of a microfluidic channel structure for partitioning individual analyte carriers.

[0050] **FIG. 2** shows an example of a microfluidic channel structure for the controlled partitioning of beads into discrete droplets.

[0051] **FIG. 3** illustrates an example of a barcode carrying bead.

[0052] **FIG. 4** illustrates another example of a barcode carrying bead.

[0053] **FIG. 5** schematically illustrates an example microwell array.

[0054] **FIG. 6** schematically illustrates an example workflow for processing nucleic acid molecules.

[0055] **FIG. 7** shows a computer system that is programmed or otherwise configured to implement methods provided herein.

[0056] **FIGs. 8A** and **8B** show beads for use according to the methods of the present disclosure.

[0057] **FIG. 9** illustrates a transposase-nucleic acid complex comprising a transposase, a first double-stranded oligonucleotide comprising a transposon end sequence and a first primer sequence and a second double-stranded oligonucleotide comprising a transposon end sequence and a second primer sequence.

[0058] **FIG. 10** illustrates a transposase-nucleic acid complex comprising a transposase, a first double-stranded oligonucleotide comprising a transposon end sequence and first and second primer sequences and a second double-stranded oligonucleotide comprising a transposon end sequence and third and fourth primer sequences.

[0059] **FIG. 11** illustrates a transposase-nucleic acid complex comprising a transposase, a first hairpin molecule, and a second hairpin molecule.

[0060] **FIG. 12** illustrates a scheme for tandem ATAC ligation and RNA template switching.

[0061] **FIG. 13** illustrates an additional scheme for tandem ATAC ligation and RNA template switching.

[0062] **FIG. 14** illustrates an exemplary scheme for tandem ATAC ligation and RNA template switching.

[0063] **FIG. 15** illustrates an additional scheme for tandem ATAC ligation and RNA template switching.

[0064] **FIG. 16** illustrates an additional scheme for tandem ATAC ligation and RNA template switching.

[0065] **FIG. 17** illustrates an additional scheme for tandem ATAC ligation and RNA template switching.

[0066] **FIG. 18** illustrates an additional scheme for tandem ATAC ligation and RNA template switching.

[0067] **FIG. 19** illustrates an additional scheme for tandem ATAC ligation and RNA template switching.

[0068] **FIG. 20** illustrates a scheme for T7 mediated linear amplification.

[0069] **FIG. 21** shows a modified workflow T7 mediated linear amplification.

[0070] **FIG. 22** illustrates a scheme for tandem ATAC and RNA processing.

[0071] **FIG. 23** illustrates a scheme for tandem ATAC and RNA processing.

[0072] **FIG. 24** illustrates the differences between computationally inferred linkages and true single cell linkage of gene expression and open chromatin analyses.

[0073] **FIG. 25** illustrates an example of a method to generate single cell accessible chromatin (ATAC) and gene expression libraries.

[0074] **FIG. 26** illustrates cell clustering and annotation of cells using expressed markers.

[0075] **FIG. 27** illustrates cell clustering and annotation of cells using chromatin accessibility (open chromatin).

[0076] **FIG. 28** illustrates concordance between the two read-outs of **FIGs. 26** and **27**.

[0077] **FIGs. 29A-29B** illustrate another representative concordance between the two read-outs of **FIGs. 26** and **27**.

[0078] **FIG. 30** illustrates the transfer gene expression marker-derived annotation into accessible chromatin clustered populations. Clustering cells from ATAC data alone using transcription factor accessibility (i.e., open chromatin) provides annotated cell clusters (left panel) that may lack the specificity to differentiate certain cell types that clustering by gene expression markers can provide. Thus, by using gene expression markers to annotate cells in open chromatin (ATAC) clusters, one may provide additional context around particular cell types.

[0079] **FIG. 31** illustrates the identification of novel populations of cells that would otherwise be unidentified and/or unannotated when analyzing either gene expression or regions of open chromatin alone. Here, populations of cells open chromatin analysis alone would show as large clusters of cells (e.g., B cells (blue), top left panel) may be annotated using gene expression markers (top right panel) to further stratify cells clustered by open chromatin (e.g., naïve/memory B cells, bottom panel).

[0080] **FIG. 32** illustrates differential gene expression in the annotated cells in **FIG. 31** and the identification and differentiation of prospective naïve vs. memory B cells. Here, a population of cells identified by gene expression analysis to be a single cluster (naïve/memory B cells, top right panel) is identified as two distinct clusters in the gene expression annotated open chromatin (naïve/memory B cells, top left panel). Looking at differential gene expression in the gene expression annotated open chromatin reveals two distinct populations of cells (subcluster1 and subcluster 2, bottom left panel) that are obscured when viewing gene expression alone (bottom right panel). Gene expression analysis of subcluster1 and subcluster 2 identifies subcluster 1 as prospective memory B cells (relatively higher Ig, relatively lower naïve B cell associated transcripts) and subcluster 2 as prospective naïve B cells (relatively lower Ig, relatively higher naïve B cell associated transcripts).

[0081] **FIG. 33** illustrates the pathological description of a tumor sample analyzed by single cell open chromatin (ATAC-seq) and gene expression analysis.

[0082] **FIG. 34** illustrates cell type annotation of a tumor sample using gene expression ("GEX", left panel) markers and transcription factor ("ATAC", right panel) accessibility.

[0083] **FIG. 35** illustrates the identification of tumor B cells from normal B cells using mutational load (SNVs) and the *BANK1* pathway (markers of B cell hyper-activation).

[0084] **FIG. 36** illustrates gene expression annotation of tumor cells to annotate and identify open chromatin cell populations.

[0085] **FIG. 37** illustrates differential gene expression between normal and tumor B cells. *FCRL5* / *FCRL3* encode members of the immunoglobulin receptor superfamily and the Fc-receptor like family. These genes are implicated in B cell development and lymphomagenesis. *MIR155HG* represents a microRNA host gene. The long RNA transcribed from this gene is expressed at high levels in lymphoma and may function as an oncogene. *RASGRF1* is a guanine nucleotide exchange factor (GEF) and is involved in MAP-Erk pathway. *IL4R* is a receptor to a key inflammatory signaling factor, pro-growth and pro-metastatic. *XAF1* encodes a protein which binds to and counteracts the inhibitory effect of a member of the IAP (inhibitor of apoptosis) protein family. *BANK1* is a tumor suppressor in B cell lymphoma.

[0086] **FIGs. 38A-38C** illustrate, on the basis of covariance of open chromatin and gene expression, the identification of a candidate enhancer region that regulates the expression of the IL4R specifically in the tumor B cells. Signal transducer and activator of transcription (STAT) proteins are critical mediators of cytokine signaling. Among the seven STAT proteins, STAT6 is activated by IL-4 and IL-13 and plays a predominant role in the immune system. Here, gene

expression and accessible chromatin characterization of *Stat3* and *Stat6* and indicate that IL4R-mediated STAT6 signaling pathway is activated in this tumor.

[0087] **FIG. 39** schematically illustrates correspondence between DNA sequencing information and RNA sequencing information obtained using the methods provided herein.

[0088] **FIG. 40** illustrates an example workflow for identifying a cis-regulatory element associated with an expressed protein.

[0089] **FIG. 41** illustrates the identification of tumor B cells from normal B cells using orthogonal lines of evidence.

[0090] **FIG. 42** illustrates the application of gene expression and chromatin data to identify signaling pathways in a tumor system.

[0091] **FIG. 43** illustrates an example flowchart showing a process flow for feature linkage analysis, in accordance with various embodiments.

[0092] **FIG. 44** illustrates another example flowchart showing a process flow for feature linkage analysis, in accordance with various embodiments.

[0093] **FIGs. 45A-45B** illustrate an analysis of open chromatin and gene expression of the IL4R gene and observed feature linkages in tumor B cells.

[0094] **FIG. 46A** illustrates mean expression of genes in normal B cells, tumor B cells and cycling tumor B cells for selected top differentially expressed immune genes, transcription factors and cell cycle genes. **FIG. 46B** illustrates an enriched functional gene set for up-regulated genes in tumor B cells.

[0095] **FIG. 47** illustrates an exemplary workflow of transcription factor-gene network construction.

[0096] **FIG. 48A** illustrates linkage significance distribution for tumor-enriched feature linkages, separated by the overlap of CLL-annotated super-enhancers. **FIG. 48B** illustrates ATAC cut site coverage and inferred feature linkages at the PAX5 locus (left) and per-cell type expression of PAX5 and linked peaks (right).

[0097] **FIG. 49A** illustrates an exemplary analysis workflow of motif enrichment and **FIG. 49B** illustrates the motif enrichment scores of all motifs (left) and the top enriched hits (right).

[0098] **FIG. 50** illustrates a transcription factor regulatory network in tumor B cells.

[0099] **FIG. 51** illustrates feature linkages which can be positively or negatively correlated signals across cells for peaks identified in an ATAC library and transcript levels identified in a gene expression library.

[00100] The instant application may contain at least one drawing executed in color. Copies of this patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

DETAILED DESCRIPTION

[00101] While various embodiments of the invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions may occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed.

[00102] Where values are described as ranges, it will be understood that such disclosure includes the disclosure of all possible sub-ranges within such ranges, as well as specific numerical values that fall within such ranges irrespective of whether a specific numerical value or specific sub-range is expressly stated.

[00103] The terms "a," "an," and "the," as used herein, generally refers to singular and plural references unless the context clearly dictates otherwise.

[00104] Whenever the term "at least," "greater than," or "greater than or equal to" precedes the first numerical value in a series of two or more numerical values, the term "at least," "greater than" or "greater than or equal to" applies to each of the numerical values in that series of numerical values. For example, greater than or equal to 1, 2, or 3 is equivalent to greater than or equal to 1, greater than or equal to 2, or greater than or equal to 3.

[00105] Whenever the term "no more than," "less than," or "less than or equal to" precedes the first numerical value in a series of two or more numerical values, the term "no more than," "less than," or "less than or equal to" applies to each of the numerical values in that series of numerical values. For example, less than or equal to 3, 2, or 1 is equivalent to less than or equal to 3, less than or equal to 2, or less than or equal to 1.

[00106] The term "barcode," as used herein, generally refers to a label, or identifier, that conveys or is capable of conveying information about an analyte. A barcode can be part of an analyte. A barcode can be independent of an analyte. A barcode can be a tag attached to an analyte (e.g., nucleic acid molecule) or a combination of the tag in addition to an endogenous characteristic of the analyte (e.g., size of the analyte or end sequence(s)). A barcode may be unique. Barcodes can have a variety of different formats. For example, barcodes can include: polynucleotide barcodes; random nucleic acid and/or amino acid sequences; and synthetic nucleic acid and/or amino acid sequences. A barcode can be attached to an analyte in a reversible or irreversible manner. A barcode can be added to, for example, a fragment of a

deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) sample before, during, and/or after sequencing of the sample. Barcodes can allow for identification and/or quantification of individual sequencing-reads.

[00107] The term "real time," as used herein, can refer to a response time of less than about 1 second, a tenth of a second, a hundredth of a second, a millisecond, or less. The response time may be greater than 1 second. In some instances, real time can refer to simultaneous or substantially simultaneous processing, detection or identification.

[00108] The term "subject," as used herein, generally refers to an animal, such as a mammal (e.g., human) or avian (e.g., bird), or other organism, such as a plant. For example, the subject can be a vertebrate, a mammal, a rodent (e.g., a mouse), a primate, a simian or a human. Animals may include, but are not limited to, farm animals, sport animals, and pets. A subject can be a healthy or asymptomatic individual, an individual that has or is suspected of having a disease (e.g., cancer) or a pre-disposition to the disease, and/or an individual that is in need of therapy or suspected of needing therapy. A subject can be a patient. A subject can be a microorganism or microbe (e.g., bacteria, fungi, archaea, viruses).

[00109] The term "genome," as used herein, generally refers to genomic information from a subject, which may be, for example, at least a portion or an entirety of a subject's hereditary information. A genome can be encoded either in DNA or in RNA. A genome can comprise coding regions (e.g., that code for proteins) as well as non-coding regions. A genome can include the sequence of all chromosomes together in an organism. For example, the human genome ordinarily has a total of 46 chromosomes. The sequence of all of these together may constitute a human genome.

[00110] The terms "adaptor(s)", "adapter(s)" and "tag(s)" may be used synonymously. An adaptor or tag can be coupled to a polynucleotide sequence to be "tagged" by any approach, including ligation, hybridization, or other approaches.

[00111] The term "sequencing," as used herein, generally refers to methods and technologies for determining the sequence of nucleotide bases in one or more polynucleotides. The polynucleotides can be, for example, nucleic acid molecules such as deoxyribonucleic acid (DNA) or ribonucleic acid (RNA), including variants or derivatives thereof (e.g., single stranded DNA). Sequencing can be performed by various systems currently available, such as, without limitation, a sequencing system by Illumina®, Pacific Biosciences (PacBio®), Oxford Nanopore®, or Life Technologies (Ion Torrent®). Alternatively or in addition, sequencing may be performed using nucleic acid amplification, polymerase chain reaction (PCR) (e.g., digital PCR, quantitative PCR, or real time PCR), or isothermal amplification. Such systems may

provide a plurality of raw genetic data corresponding to the genetic information of a subject (e.g., human), as generated by the systems from a sample provided by the subject. In some examples, such systems provide sequencing reads (also "reads" herein). A read may include a string of nucleic acid bases corresponding to a sequence of a nucleic acid molecule that has been sequenced. In some situations, systems and methods provided herein may be used with proteomic information.

[00112] The term "bead," as used herein, generally refers to a particle. The bead may be a solid or semi-solid particle. The bead may be a gel bead. The gel bead may include a polymer matrix (e.g., matrix formed by polymerization or cross-linking). The polymer matrix may include one or more polymers (e.g., polymers having different functional groups or repeat units). Polymers in the polymer matrix may be randomly arranged, such as in random copolymers, and/or have ordered structures, such as in block copolymers. Cross-linking can be via covalent, ionic, or inductive, interactions, or physical entanglement. The bead may be a macromolecule. The bead may be formed of nucleic acid molecules bound together. The bead may be formed via covalent or non-covalent assembly of molecules (e.g., macromolecules), such as monomers or polymers. Such polymers or monomers may be natural or synthetic. Such polymers or monomers may be or include, for example, nucleic acid molecules (e.g., DNA or RNA). The bead may be formed of a polymeric material. The bead may be magnetic or non-magnetic. The bead may be rigid. The bead may be flexible and/or compressible. The bead may be disruptable or dissolvable. The bead may be a solid particle (e.g., a metal-based particle including but not limited to iron oxide, gold or silver) covered with a coating comprising one or more polymers. Such coating may be disruptable or dissolvable.

[00113] As used herein, the term "barcoded nucleic acid molecule" generally refers to a nucleic acid molecule that results from, for example, the processing of a nucleic acid barcode molecule with a nucleic acid sequence (e.g., nucleic acid sequence complementary to a nucleic acid primer sequence encompassed by the nucleic acid barcode molecule). The nucleic acid sequence may be a targeted sequence or a non-targeted sequence. For example, in the methods and systems described herein, hybridization and reverse transcription of a nucleic acid molecule (e.g., a messenger RNA (mRNA) molecule) of a cell with a nucleic acid barcode molecule (e.g., a nucleic acid barcode molecule containing a barcode sequence and a nucleic acid primer sequence complementary to a nucleic acid sequence of the mRNA molecule) results in a barcoded nucleic acid molecule that has a sequence corresponding to the nucleic acid sequence of the mRNA and the barcode sequence (or a reverse complement thereof). A barcoded nucleic acid molecule may serve as a template, such as a template polynucleotide, that can be further

processed (e.g., amplified) and sequenced to obtain the target nucleic acid sequence. For example, in the methods and systems described herein, a barcoded nucleic acid molecule may be further processed (e.g., amplified) and sequenced to obtain the nucleic acid sequence of the mRNA.

[00114] The term "sample," as used herein, generally refers to a biological sample of a subject. The biological sample may comprise any number of macromolecules, for example, cellular macromolecules. The sample may be a cell sample. The sample may be a cell line or cell culture sample. The sample can include one or more cells. The sample can include one or more microbes. The biological sample may be a nucleic acid sample or protein sample. The biological sample may also be a carbohydrate sample or a lipid sample. The biological sample may be derived from another sample. The sample may be a tissue sample, such as a biopsy, core biopsy, needle aspirate, or fine needle aspirate. The sample may be a fluid sample, such as a blood sample, urine sample, or saliva sample. The sample may be a skin sample. The sample may be a cheek swab. The sample may be a plasma or serum sample. The sample may be a cell-free or cell free sample. A cell-free sample may include extracellular polynucleotides. Extracellular polynucleotides may be isolated from a bodily sample that may be selected from the group consisting of blood, plasma, serum, urine, saliva, mucosal excretions, sputum, stool and tears.

[00115] The term "biological particle," as used herein, generally refers to a discrete biological system derived from a biological sample. The biological particle may be a macromolecule. The biological particle may be a small molecule. The biological particle may be a virus. The biological particle may be a cell or derivative of a cell. The biological particle may be an organelle. The biological particle may be a rare cell from a population of cells. The biological particle may be any type of cell, including without limitation prokaryotic cells, eukaryotic cells, bacterial, fungal, plant, mammalian, or other animal cell type, mycoplasmas, normal tissue cells, tumor cells, or any other cell type, whether derived from single cell or multicellular organisms. The biological particle may be a constituent of a cell. The biological particle may be or may include DNA, RNA, organelles, proteins, or any combination thereof. The biological particle may be or may include a matrix (e.g., a gel or polymer matrix) comprising a cell or one or more constituents from a cell (e.g., cell bead), such as DNA, RNA, organelles, proteins, or any combination thereof, from the cell. The biological particle may be obtained from a tissue of a subject. The biological particle may be a hardened cell. Such hardened cell may or may not include a cell wall or cell membrane. The biological particle may include one or more constituents of a cell, but may not include other constituents of the cell. An example of such

constituents is a nucleus or an organelle. A cell may be a live cell. The live cell may be capable of being cultured, for example, being cultured when enclosed in a gel or polymer matrix, or cultured when comprising a gel or polymer matrix.

[00116] The term "macromolecular constituent," as used herein, generally refers to a macromolecule contained within or from a biological particle. The macromolecular constituent may comprise a nucleic acid. In some cases, the biological particle may be a macromolecule. The macromolecular constituent may comprise DNA. The macromolecular constituent may comprise RNA. The RNA may be coding or non-coding. The RNA may be messenger RNA (mRNA), ribosomal RNA (rRNA) or transfer RNA (tRNA), for example. The RNA may be a transcript. The RNA may be small RNA that are less than 200 nucleic acid bases in length, or large RNA that are greater than 200 nucleic acid bases in length. Small RNAs may include 5.8S ribosomal RNA (rRNA), 5S rRNA, transfer RNA (tRNA), microRNA (miRNA), small interfering RNA (siRNA), small nucleolar RNA (snoRNAs), Piwi-interacting RNA (piRNA), tRNA-derived small RNA (tsRNA) and small rDNA-derived RNA (srRNA). The RNA may be double-stranded RNA or single-stranded RNA. The RNA may be circular RNA. The macromolecular constituent may comprise a protein. The macromolecular constituent may comprise a peptide. The macromolecular constituent may comprise a polypeptide.

[00117] The term "molecular tag," as used herein, generally refers to a molecule capable of binding to a macromolecular constituent. The molecular tag may bind to the macromolecular constituent with high affinity. The molecular tag may bind to the macromolecular constituent with high specificity. The molecular tag may comprise a nucleotide sequence. The molecular tag may comprise a nucleic acid sequence. The nucleic acid sequence may be at least a portion or an entirety of the molecular tag. The molecular tag may be a nucleic acid molecule or may be part of a nucleic acid molecule. The molecular tag may be an oligonucleotide or a polypeptide. The molecular tag may comprise a DNA aptamer. The molecular tag may be or comprise a primer. The molecular tag may be, or comprise, a protein. The molecular tag may comprise a polypeptide. The molecular tag may be a barcode.

[00118] The term "partition," as used herein, generally, refers to a space or volume that may be suitable to contain one or more species or conduct one or more reactions. A partition may be a physical compartment, such as a droplet or well. The partition may isolate space or volume from another space or volume. The droplet may be a first phase (e.g., aqueous phase) in a second phase (e.g., oil) immiscible with the first phase. The droplet may be a first phase in a second phase that does not phase separate from the first phase, such as, for example, a capsule or liposome in an aqueous phase. A partition may comprise one or more other (inner) partitions.

In some cases, a partition may be a virtual compartment that can be defined and identified by an index (e.g., indexed libraries) across multiple and/or remote physical compartments. For example, a physical compartment may comprise a plurality of virtual compartments.

[00119] The present disclosure provides methods, systems, and kits for processing multiple types of nucleic acid molecules. The methods, systems, and kits provided herein may facilitate sample preparation for sequencing of nucleic acid molecules included in cells, cell beads, or cell nuclei of interest. For example, the present disclosure provides methods for processing both deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) molecules included within a cell, cell bead, or cell nucleus. The methods may comprise performing Assay for Transposase Accessible Chromatin with high-throughput sequencing (ATAC-seq) and RNA sequencing (RNA-seq) assays in tandem. Partitioning and barcoding schemes may be utilized to facilitate identification of resultant sequencing reads with the cell, cell bead, or cell nucleus from which they are derived.

[00120] The present disclosure also provides methods, systems, and kits for processing biological samples comprising nucleic acid molecules. The method may comprise providing one or more nucleic acid molecules (e.g., deoxyribonucleic acid (DNA) molecules and/or ribonucleic acid (RNA) molecules) from a nucleic acid sample (e.g., a sample comprising a cell, cell bead, or cell nucleus) in a partition of a plurality of partitions (e.g., a plurality of droplets or wells). The one or more nucleic acid molecules may be one or more DNA molecules. The one or more DNA molecules may be transcribed to generate one or more RNA molecules, where the one or more RNA molecules may be reverse transcribed to generate one or more complementary DNA (cDNA) molecules. The one or more cDNA molecules, or derivatives thereof, may then be recovered from the partition of the plurality of partitions (e.g., by pooling the contents of the plurality of partitions). The one or more cDNA molecules, or derivatives thereof, may comprise one or more nucleic acid barcode sequences, or complements thereof, where the one or more nucleic acid barcode sequences, or complements thereof, may be incorporated into nucleic acid molecules during any processing step (e.g., during transcription of a DNA molecule, reverse transcription of an RNA molecule, etc.). The one or more nucleic acid barcode sequences, or complements thereof, may be used to identify sequencing reads (e.g., sequencing reads obtained using a nucleic acid sequencing assay) corresponding to the one or more cDNA molecules from a nucleic acid molecule of the one or more nucleic acid molecules from the nucleic acid sample.

Tandem DNA and RNA Barcoding

[00121] In an aspect, the present disclosure provides a method for processing nucleic acid molecules from a cell, cell bead, or cell nucleus. The method may comprise contacting a cell,

cell bead, or cell nucleus with a transposase-nucleic acid complex comprising a transposase molecule and one or more transposon end oligonucleotide molecules. The cell, cell bead or cell nucleus may be contacted with a transposase-nucleic acid complex in bulk solution, such that the cell, cell bead or cell nucleus undergoes "tagmentation" via a tagmentation reaction. Contacting the cell, cell bead, or cell nucleus with the transposase-nucleic acid complex may generate one or more template nucleic acid fragments (e.g., "tagmented fragments" or "tagged fragments"). The one or more template nucleic acid fragments may correspond to one or more target nucleic acid molecules (e.g., deoxyribonucleic acid (DNA) molecules) within the cell, cell bead, or cell nucleus. In parallel, the cell, cell bead, or cell nucleus may be contacted with a primer molecule (e.g., a primer molecule comprising a poly-T sequence) configured to interact with one or more additional target nucleic acid molecules (e.g., ribonucleic acid (RNA) molecules, such as messenger RNA (mRNA) molecules). The cell, cell bead, or cell nucleus may be contacted with a primer molecule in bulk solution. Alternatively or in addition to, the cell, cell bead, or cell nucleus may be contacted with a primer molecule within a partition. Interaction between these moieties may yield one or more additional template nucleic acid fragments (e.g., RNA fragments). For example, the primer molecule may have at least partial sequence complementarity to the one or more additional target nucleic acid molecules (e.g., mRNA molecules). The primer molecule may hybridize to a sequence of an additional target nucleic acid molecule of the one or more additional target nucleic acid molecules. The cell, cell bead, or cell nucleus may be partitioned (e.g., co-partitioned with one or more reagents) into a partition (e.g., of a plurality of partitions). The partition may be, for example, a droplet or a well. The partition may comprise one or more reagents, including, for example, one or more particles (e.g., beads) comprising one or more nucleic acid barcode molecules. The cell, cell bead, or cell nucleus may be lysed, permeabilized, fixed, cross-linked or otherwise manipulated to provide access to the one or more template nucleic acid fragments and the one or more additional template nucleic acid fragments therein. The one or more template nucleic acid fragments and the one or more additional template nucleic acid fragments therein may undergo one or more processing steps within the partition. For example, the one or more template nucleic acid fragments and/or the one or more additional template nucleic acid fragments may undergo a barcoding process, a ligation process, a reverse transcription process, a template switching process, a linear amplification process, and/or a gap filling process. The resultant one or more processed template nucleic acid fragments (e.g., tagmented fragments) and/or the one or more processed additional template nucleic acid fragments (e.g., RNA fragments) may each include a barcode sequence (e.g., a nucleic acid barcode sequence, as described herein). The one or more

processed template nucleic acid fragments and/or the one or more processed additional template nucleic acid fragments may be released from the partition (e.g., pooled with contents of other partitions of a plurality of partitions) and may undergo one or more additional processing steps in bulk. For example, the one or more processed template nucleic acid fragments and/or the one or more processed additional template nucleic acid fragments may undergo a gap filling process, a dA tailing process, a terminal-transferase process, a phosphorylation process, a ligation process, a nucleic acid amplification process, or a combination thereof. For example, the one or more processed template nucleic acid fragments and/or the one or more processed additional template nucleic acid fragments may be subjected to conditions sufficient to undergo one or more polymerase chain reactions (PCR, such as sequence independent PCR) to generate amplification products corresponding to the one or more processed template nucleic acid fragments (e.g., tagmented fragments) and/or the one or more processed additional template nucleic acid fragments (e.g., RNA fragments). Sequences of such amplification products can be detected using, for example, a nucleic acid sequencing assay and used to identify sequences of the one or more target nucleic acid molecules (e.g., DNA molecules) and the one or more additional target nucleic acid molecules (e.g., RNA molecules) of the cell, cell bead, or cell nucleus from which they derive.

[00122] A biological sample (e.g., a nucleic acid sample) may comprise one or more cells, cell beads, and/or cell nuclei. A biological sample may also comprise tissue, which tissue may comprise one or more cells, cell beads, and/or cell nuclei. In some cases, a biological sample may comprise a plurality of cells comprising a plurality of cell nuclei. In some cases, a biological sample may comprise a plurality of cell nuclei, which plurality of cell nuclei are not included within cells (e.g., other components of the cell have degraded, dissociated, dissolved, or otherwise been removed). A biological sample may comprise a plurality of cell-free nucleic acid molecules (e.g., nucleic acid molecules that are not included within cells). For example, a biological sample may comprise a plurality of cell-free fetal DNA (cffDNA) or circulating tumor DNA (ctDNA) or other cell-free nucleic acid molecules (e.g., deriving from degraded cells). Such a biological sample may be processed to separate such cell-free nucleic acid molecules from cells, cell beads, and/or cell nuclei, which cells, cell beads, and/or cell nuclei may be subjected to further processing (e.g., as described herein).

[00123] Nucleic acid molecules included within a biological sample may include, for example, DNA molecules and RNA molecules. For example, a biological sample may comprise genomic DNA comprising chromatin (e.g., within a cell, cell bead, or cell nucleus). A biological sample may comprise a plurality of RNA molecules, such as a plurality of pre-mRNA or mRNA

molecules (e.g., within a cell, cell bead, or cell nucleus). mRNA molecules and other RNA molecules may comprise a polyA sequence. At least a subset of a plurality of RNA molecules included in a cell or cell bead may be present in a cell nucleus.

[00124] A nucleic acid molecule may undergo one or more processing steps within a cell, cell bead, or cell nucleus. For example, chromatin within a cell, cell bead, or cell nucleus may be contacted with a transposase. A transposase may be included within a transposase-nucleic acid complex, which transposase-nucleic acid complex may comprise a transposase molecule and one or more transposon end oligonucleotide molecules. A transposase may be a Tn transposase, such as a Tn3, Tn5, Tn7, Tn10, Tn552, Tn903 transposase. Alternatively, a transposase may be a MuA transposase, a Vibhar transposase (e.g. from *Vibrio harveyi*), Ac-Ds, Ascot-1, Bs1, Cin4, Copia, En/Spm, F element, hobo, Hsmar1, Hsmar2, IN (HIV), IS1, IS2, IS3, IS4, IS5, IS6, IS10, IS21, IS30, IS50, IS51, IS150, IS256, IS407, IS427, IS630, IS903, IS911, IS982, IS1031, ISL2, L1, Mariner, P element, Tam3, Tc1, Tc3, Te1, THE-1, Tn/O, TnA, Tn3, Tn5, Tn7, Tn10, Tn552, Tn903, Tol1, Tol2, TnlO, Tyl, any prokaryotic transposase, or any transposase related to and/or derived from those listed above. For example, a transposase may be a Tn5 transposase or a mutated, hyperactive Tn5 transposase. A transposase related to and/or derived from a parent transposase may comprise a peptide fragment with at least about 50%, about 55%, about 60%, about 65%, about 70%, about 75%, about 80%, about 85%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% amino acid sequence homology to a corresponding peptide fragment of the parent transposase. The peptide fragment may be at least about 10, about 15, about 20, about 25, about 30, about 35, about 40, about 45, about 50, about 60, about 70, about 80, about 90, about 100, about 150, about 200, about 250, about 300, about 400, or about 500 amino acids in length. For example, a transposase derived from Tn5 may comprise a peptide fragment that is 50 amino acids in length and about 80% homologous to a corresponding fragment in a parent Tn5 transposase. Action of a transposase (e.g., insertion) may be facilitated and/or triggered by addition of one or more cations, such as one or more divalent cations (e.g., Ca^{2+}, Mg^{2+}, or Mn^{2+}).

[00125] A transposase-nucleic acid complex may comprise one or more nucleic acid molecules. For example, a transposase-nucleic acid complex may comprise one or more transposon end oligonucleotide molecules. A transposon end oligonucleotide molecule may comprise one or more adapter sequences (e.g., comprising one or more primer sequences) and/or one or more transposon end sequences. A transposon end sequence may be, for example, a Tn5 or modified Tn5 transposon end sequence or a Mu transposon end sequence. A transposon end sequence may have a sequence of, for example,

AGATGTGTATAAGAGACA (SEQ ID NO: 1).

A primer sequence of a transposon end oligonucleotide molecule may be a sequencing primer, such as an R1 or R2 sequencing primer, or a portion thereof. A sequencing primer may be, for example, a TrueSeq or Nextera sequencing primer. An R1 sequencing primer region may have a sequence of

TCTACACTCTTTCCCTACACGACGCTCTTCCGATCT (SEQ ID NO: 2),

or some portion thereof. An R1 sequencing primer region may have a sequence of

TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG (SEQ ID NO: 3),

or some portion thereof. A transposon end oligonucleotide molecule may comprise a partial R1 sequence. A partial R1 sequence may be

ACTACACGACGCTCTTCCGATCT (SEQ ID NO: 4).

A transposon end oligonucleotide molecule may comprise an R2 sequencing priming region. An R2 sequencing primer region may have a sequence of

GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT (SEQ ID NO: 5),

or some portion thereof. An R2 sequencing primer region may have a sequence of

GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG (SEQ ID NO: 6),

or some portion thereof. A transposon end oligonucleotide molecule may comprise a T7 promoter sequence. A T7 promoter sequence may be

TAATACGACTCACTATAG (SEQ ID NO: 7).

A transposon end oligonucleotide molecule may comprise a region at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to any one of SEQ ID NO: 1-7. A transposon end oligonucleotide molecule may comprise a P5 sequence and/or a P7 sequence. A transposon end oligonucleotide molecule may comprise a sample index sequence, such as a barcode sequence or unique molecular identifier sequence. One or more transposon end oligonucleotide molecules of a transposase-nucleic acid complex may be attached to a solid support (e.g., a solid or semi-solid particle such as a bead (e.g., gel bead)). A transposon end oligonucleotide molecule may be releasably coupled to a solid support (e.g., a bead). Examples of transposon end oligonucleotide molecules may be found in, for example, PCT Patent Publications Nos. WO2018/218226, WO2014/189957, US. Pat. Pub. 20180340171, and U.S. Pat. 10,059,989; each of which are herein incorporated by reference in their entireties.

[00126] **FIG. 9** includes an example of a transposase-nucleic acid complex for use in the methods provided herein. Transposase-nucleic acid complex **900** (e.g., comprising a transpose dimer) comprises partially double-stranded oligonucleotide **901** and partially double-stranded

oligonucleotide **905**. Partially double-stranded oligonucleotide **901** comprises transposon end sequence **903**, first primer sequence **902**, and a sequence **904** that is complementary to transposon end sequence **903**. Partially double-stranded oligonucleotide **905** comprises transposon end sequence **906**, first primer sequence **907**, and a sequence **908** that is complementary to transposon end sequence **906**. Primer sequences **902** and **907** may be the same or different. In some cases, primer sequence **902** may be designated "R1" and primer sequence **907** may be designated "R2". Transposon end sequences **903** and **906** may be the same or different. Transposon end sequences **903** and **906** may alternately be referred to as "mosaic end" or "ME" sequences, while their complementary sequences **904** and **908** may be referred to as "mosaic end reverse complement" or "MErc" sequences.

[00127] **FIG. 10** includes another example of a transposase-nucleic acid complex for use in the methods provided herein. Transposase-nucleic acid complex **1000** (e.g., comprising a transpose dimer) comprises forked adapters **1001** and **1006**, which forked adapters are partially double-stranded oligonucleotides. Partially double-stranded oligonucleotide **1001** comprises transposon end sequence **1003**, first primer sequence **1002**, second primer sequence **1005**, and a sequence **1004** that is complementary to transposon end sequence **1003**. Partially double-stranded oligonucleotide **1006** comprises transposon end sequence **1007**, first primer sequence **1008**, second primer sequence **1010**, and a sequence **1009** that is complementary to transposon end sequence **1007**. Primer sequences **1002, 1005, 1008**, and **1010** may be the same or different. In some cases, primer sequences **1002** and **1008** may be designated "R1" and primer sequences **1005** and **1010** may be designated "R2". Alternatively, primer sequences **1002** and **1010** may be designated "R1" and primer sequences **1005** and **1008** may be designated "R2". Alternatively, primer sequences **1002** and **1008** may be designated "R2" and primer sequences **1005** and **1010** may be designated "R1". Alternatively, primer sequences **1002** and **1010** may be designated "R2" and primer sequences **1005** and **1008** may be designated "R1". Transposon end sequences **1003** and **1007** may be the same or different. These sequences may alternately be referred to as "mosaic end" or "ME" sequences, while their complementary sequences **1004** and **1009** may be referred to as "mosaic end reverse complement" or "MErc" sequences.

[00128] **FIG. 11** shows transposase-nucleic acid complex **1100** (e.g., comprising a transpose dimer) comprising hairpin molecules **1101** and **1106**. Hairpin molecule **1101** comprises transposon end sequence **1103**, first hairpin sequence **1102**, second hairpin sequence **1105**, and a sequence **1104** that is complementary to transposon end sequence **1103**. Hairpin molecule **1106** comprises transposon end sequence **1107**, third hairpin sequence **1108**, fourth hairpin sequence **1110**, and a sequence **1109** that is complementary to transposon end sequence **1107**. Hairpin

sequences **1102, 1105, 1108**, and **1110** may be the same or different. For example, hairpin sequence **1105** may be the same or different as hairpin sequence **1110**, and/or hairpin sequence **1102** may be the same or different as hairpin sequence **1108**. Hairpin sequences **1102** and **1108** may be spacer sequences or adapter sequences. Hairpin sequences **1105** and **1110** may be a promoter sequence such as T7 recognition or promoter sequences and/or UMI sequences. Transposon end sequences **1103** and **1107** may be the same or different. Transposon end sequences **1103** and **1107** may alternately be referred to as "mosaic end" or "ME" sequences, while their complementary sequences **1104** and **1109** may be referred to as "mosaic end reverse complement" or "MErc" sequences. In some cases, sequence **1104** is a transposon end sequence and **1103** is a sequence complementary to sequence **1104**. In some cases, sequence **1109** is a transposon end sequence and **1107** is a sequence complementary to sequence **1109**.

[00129] Contacting a cell, cell bead, or cell nucleus comprising one or more target nucleic acid molecules (e.g., DNA molecules) with a transposase-nucleic acid complex may generate one or more template nucleic acid fragments (e.g., "tagmented fragments"). The one or more template nucleic acid fragments may each comprise a sequence of the one or more target nucleic acid molecules (e.g., a target sequence). The transposase-nucleic acid complex may be configured to target a specific region of the one or more target nucleic acid molecules to provide one or more template nucleic acid fragments comprising specific target sequences. The one or more template nucleic acid fragments may comprise target sequences corresponding to accessible chromatin. Generation of tagmented fragments may take place within a bulk solution. In other cases, generation of tagmented fragments may take place within a partition (e.g., a droplet or well). A template nucleic acid fragment (e.g., tagmented fragment) may comprise one or more gaps (e.g., between a transposon end sequence or complement thereof and a target sequence on one or both strands of a double-stranded fragment). Gaps may be filled via a gap filling process using, e.g., a polymerase (e.g., DNA polymerase), ligase, or reverse transcriptase. In some cases, a mixture of enzymes may be used to repair a partially double-stranded nucleic acid molecule and fill one or more gaps. Gap filling may not include strand displacement. Gaps may be filled within or outside of a partition.

[00130] Alternatively or in addition to, one or more additional nucleic acid molecules may be contacted with one or more capture nucleic acid molecules within a cell, cell bead, or cell nucleus to provide one or more additional template nucleic acid fragments. For example, an RNA molecule (e.g., an mRNA) molecule may be contacted with a primer molecule within a cell, cell bead, or cell nucleus. A primer molecule may comprise a primer sequence, which primer sequence may be a targeted primer sequence or a non-specific primer sequence (e.g.,

random N-mer). A targeted primer sequence may comprise, for example, a polyT sequence, which polyT sequence may interact with a polyA sequence of an RNA molecule. A primer nucleic acid molecule may also comprise one or more additional sequences, such as one or more sample index sequences, spacer or linker sequences, or one or more additional primer sequences. Generation of additional template nucleic acid fragments (e.g., RNA fragments) may take place within a bulk solution. In other cases, generation of additional template nucleic acid fragments may take place within a partition (e.g., a droplet or well).

[00131] Processing of nucleic acid molecules within a cell, cell bead, or cell nucleus (e.g., generation of template nucleic acid fragments using a transposase-nucleic acid complex and/or generation of additional template nucleic acid fragments using a capture nucleic acid molecule) may occur in a bulk solution comprising a plurality of cells, cell beads, and/or cell nuclei. In some cases, template nucleic acid fragments (e.g., tagmented fragments) may be generated in bulk solution and additional template nucleic acid fragments (e.g., RNA fragments) may be generated in a partition.

[00132] A plurality of cells, cell beads, and/or cell nuclei (e.g., a plurality of cells, cell beads, and/or cell nuclei that have undergone processing such as a tagmentation process) may be partitioned amongst a plurality of partitions. Partitions may be, for example, droplets or wells. Droplets (e.g., aqueous droplets) may be generated according to the methods provided herein. Partitioning may be performed according to the method provided herein. For example, partitioning a biological particle (e.g., cell, cell bead, or cell nucleus) and one or more reagents may comprise flowing a first phase comprising an aqueous fluid, the biological particle, and the one or more reagents and a second phase comprising a fluid that is immiscible with the aqueous fluid toward a junction. Upon interaction of the first and second phases, a discrete droplet of the first phase comprising the biological particle and the one or more reagents may be formed. The plurality of cells, cell beads, and/or cell nuclei may be partitioned amongst a plurality of partitions such that at least a subset of the plurality of partitions may comprise at most one cell, cell bead, or cell nucleus. Cells, cell beads, and/or cell nuclei may be co-partitioned with one or more reagents such that a partition of at least a subset of the plurality of partitions comprises a single cell, cell bead, or cell nucleus and one or more reagents. The one or more reagents may include, for example, enzymes (e.g., polymerases, reverse transcriptases, ligases, etc.), nucleic acid barcode molecules (e.g., nucleic acid barcode molecules comprising one or more barcode sequences, such as nucleic acid barcode molecules coupled to one or more beads), template switching oligonucleotides, deoxynucleotide triphosphates, buffers, lysis agents, primers, barcodes, detergents, reducing agents, chelating agents, oxidizing agents, nanoparticles, beads,

antibodies, or any other useful reagents. Enzymes may include, for example, temperature-sensitive enzymes, pH-sensitive enzymes, light-sensitive enzymes, reverse transcriptases, proteases, ligases, polymerases, kinases, restriction enzymes, nucleases, protease inhibitors, exonucleases, and nuclease inhibitors.

[00133] A reagent of the one or more reagents may be useful for lysing or permeabilizing a cell, cell bead, or cell nucleus, or otherwise providing access to nucleic acid molecules and/or template nucleic acid fragments therein. A cell may be lysed using a lysis agent such as a bioactive agent. A bioactive agent useful for lysing a cell may be, for example, an enzyme (e.g., as described herein). An enzyme used to lyse a cell may or may not be capable of carrying out additional actions such as degrading one or more RNA molecules. Alternatively, an ionic, zwitterionic, or non-ionic surfactant may be used to lyse a cell. Examples of surfactants include, but are not limited to, TritonX-100, Tween 20, sarcosyl, or sodium dodecyl sulfate. Cell lysis may also be achieved using a cellular disruption method such as an electroporation or a thermal, acoustic, or mechanical disruption method. Alternatively, a cell may be permeabilized to provide access to a plurality of nucleic acid molecules included therein. Permeabilization may involve partially or completely dissolving or disrupting a cell membrane or a portion thereof. Permeabilization may be achieved by, for example, contacting a cell membrane with an organic solvent or a detergent such as Triton X-100 or NP-40. By lysing or permeabilizing a cell, cell bead, or cell nucleus within a partition (e.g., droplet) to provide access to the plurality of nucleic acid molecules and/or template nucleic acid fragments therein, molecules originating from the same cell, cell bead, or cell nucleus may be isolated within the same partition.

[00134] A partition of a plurality of partitions (e.g., a partition comprising a cell, cell bead, and/or cell nucleus) may comprise one or more beads (e.g., gel beads). A bead may be a gel bead. A bead may comprise a plurality of nucleic acid barcode molecules (e.g., nucleic acid molecules each comprising one or more barcode sequences, as described herein). A bead may comprise at least 10,000 nucleic acid barcode molecules attached thereto. For example, the bead may comprise at least 100,000, 1,000,000, or 10,000,000 nucleic acid barcode molecules attached thereto. The plurality of nucleic acid barcode molecules may be releasably attached to the bead. The plurality of nucleic acid barcode molecules may be releasable from the bead upon application of a stimulus. Such a stimulus may be selected from the group consisting of a thermal stimulus, a photo stimulus, and a chemical stimulus. For example, the stimulus may be a reducing agent such as dithiothreitol Application of a stimulus may result in one or more of (i) cleavage of a linkage between nucleic acid barcode molecules of the plurality of nucleic acid

barcode molecules and the bead, and (ii) degradation or dissolution of the bead to release nucleic acid barcode molecules of the plurality of nucleic acid barcode molecules from the bead.

[00135] A plurality of nucleic acid barcode molecules attached (e.g., releasably attached) to a bead (e.g., gel bead) may be suitable for barcoding template nucleic acid fragments or additional template nucleic acid fragments deriving from DNA and/or RNA molecules of the plurality of cells, cell beads, and/or cell nuclei. For example, a nucleic acid barcode molecule of a plurality of nucleic acid barcode molecule may comprise a barcode sequence, unique molecular identifier (UMI) sequence, primer sequence, universal primer sequence, sequencing adapter or primer, flow cell adapter sequence, or any other useful feature. In an example, a nucleic acid barcode molecule of a plurality of nucleic acid barcode molecules attached to a bead may comprise a flow cell adapter sequence (e.g., a P5 or P7 sequence), a barcode sequence, a capture sequence, and a sequencing primer sequence or portion thereof (e.g., an R1 or R2 sequence or portion thereof), or a complement of any of these sequences. These sequences may be arranged in any useful order and may be linked or may include one or more spacer sequences disposed between them. For instance, the flow cell adapter sequence, where present, may be disposed near (e.g., proximal to) an end of the nucleic acid barcode molecule that is closest to the bead, while the sequencing primer or portion thereof may be disposed at an end of the nucleic acid barcode molecule that is furthest from (e.g., distal to) the bead (e.g., most available to template nucleic acid fragments for interaction). In another example, a nucleic acid barcode molecule of a plurality of nucleic acid barcode molecules attached to a bead may comprise a flow cell adapter sequence (e.g., a P5 or P7 sequence), a barcode sequence, a sequencing primer sequence or portion thereof (e.g., an R1 or R2 sequence or portion thereof), and a UMI sequence, or a complement of any of these sequences. The nucleic acid barcode molecule may further comprise a capture sequence, which capture sequence may be a targeted capture sequence or comprise a template switch sequence (e.g., comprising a polyC or poly G sequence). These sequences may be arranged in any useful order and may be linked or may include one or more spacer sequences disposed between them. For instance, the flow cell adapter sequence may be disposed near (e.g., proximal to) an end of the nucleic acid barcode molecule that is closest to the bead, while the capture sequence or template switch sequence may be disposed at an end of the nucleic acid barcode molecule that is furthest from the bead (e.g., most available to template nucleic acid fragments for interaction).

[00136] All of the nucleic acid barcode molecules attached (e.g., releasably attached) to a bead (e.g., gel bead) of a plurality of beads may be the same. For example, all of the nucleic acid barcode molecules attached to the bead may have the same nucleic acid sequence. In such

an instance, all of the nucleic acid barcode molecules attached to the bead may comprise the same flow cell adapter sequence, sequencing primer or portion thereof, and barcode sequence. The barcode sequence of a plurality of nucleic acid barcode molecules attached to a bead of a plurality of beads may be different from other barcode sequences of other nucleic acid barcode molecules attached to other beads of the plurality of beads. For example, a plurality of beads may comprise a plurality of barcode sequences, such that, for at least a subset of the plurality of beads, each bead comprises a different barcode sequence of the plurality of barcode sequences. This differentiation may permit template nucleic acid fragments (e.g., included within cells, cell beads, and/or cell nuclei) co-partitioned with a plurality of beads between a plurality of partitions to be differentially barcoded within their respective partitions, such that the template nucleic acid fragments or molecules derived therefrom may be identified with the partition (and thus the cell, cell bead, and/or cell nucleus) to which they correspond (e.g., using a nucleic acid sequencing assay, as described herein). A barcode sequence may comprise between 4-20 nucleotides. A barcode sequence may comprise one or more segments, which segments may range in size from 2-20 nucleotides, such as from 4-20 nucleotides. Such segments may be combined to form barcode sequences using a combinatorial assembly method, such as a split-pool method. Details of such methods can be found, for example, in PCT/US2018/061391, filed November 15, 2018, and US 20190249226, each of which are herein incorporated by reference in their entireties.

[00137] In some cases, nucleic acid barcode molecules attached to a bead may not be the same. For example, the plurality of nucleic acid barcode molecules attached to a bead may each comprise a UMI sequence, which UMI sequence varies across the plurality of nucleic acid barcode molecules. All other sequences of the plurality of nucleic acid barcode molecules attached to the bead may be the same.

[00138] In some cases, a bead may comprise multiple different nucleic acid barcode molecules attached thereto. For example, a bead may comprise a first plurality of nucleic acid barcode molecules and a second plurality of nucleic acid barcode molecules, which first plurality of nucleic acid barcode molecules is different than the second plurality of nucleic acid barcode molecules. The first plurality of nucleic acid barcode molecules and the second plurality of nucleic acid barcode molecules coupled to a bead may comprise one or more shared sequences. For example, each nucleic acid barcode molecule of the first plurality of nucleic acid barcode molecules and each nucleic acid barcode molecule of the second plurality of nucleic acid barcode molecules may comprise the same barcode sequence (e.g., as described herein). Such a barcode sequence may be prepared using a combinatorial assembly process (e.g., as described

herein). For example, barcode sequences may comprise identical barcode sequence segments. Similarly, each nucleic acid barcode molecule of the first plurality of nucleic acid barcode molecules coupled to a bead may comprise the same flow cell adapter sequence and/or sequencing primer or portion thereof as each nucleic acid barcode molecule of the second plurality of nucleic acid barcode molecules coupled to the bead. In an example, each nucleic acid barcode molecule of the first plurality of nucleic acid barcode molecules coupled to a bead comprises a sequencing primer, and each nucleic acid barcode molecule of the second plurality of nucleic acid barcode molecules coupled to the bead comprises a portion of the same sequencing primer. In some instances, each nucleic acid barcode molecule of the first plurality of nucleic acid barcode molecules coupled to a bead may comprise a first sequencing primer (e.g., a TruSeq R1 sequence) a barcode sequence, and a first functional sequence, and each nucleic acid barcode molecule of the second plurality of nucleic acid barcode molecules coupled to the bead may comprise a second sequencing primer (e.g., a Nextera R1 sequence, or a portion thereof), the barcode sequence, and a second functional sequence. Sequences shared between different sets of nucleic acid barcode molecules coupled to the same bead may be included in the same or different order and may be separated by the same or different sequences. Alternatively or in addition, the first plurality of nucleic acid barcode molecules and the second plurality of nucleic acid barcode molecules coupled to a bead may include one or more different sequences. For example, each nucleic acid barcode molecule of a first plurality of nucleic acid barcode molecules coupled to a bead of a plurality of beads may comprise one or more of a flow cell adapter sequence, a barcode sequence, UMI sequence, capture sequence, and a sequencing primer or portion thereof, while each nucleic acid barcode molecule of a second plurality of nucleic acid barcode molecules coupled to the bead may comprise one or more of a flow cell adapter sequence (e.g., the same flow cell adapter sequence), a barcode sequence (e.g., the same barcode sequence), UMI sequence, capture sequence, and a sequencing primer or portion thereof (e.g., the same sequencing primer or portion thereof). Nucleic acid barcode molecules of the first plurality of nucleic acid barcode molecules may not include a UMI sequence or capture sequence. A bead comprising multiple different populations of nucleic acid barcode molecules, such as a first plurality of nucleic acid molecules and a second plurality of nucleic acid molecules (e.g., as described above), may be referred to as a "multi-functional bead."

[00139] A cell, cell bead, or cell nucleus comprising template nucleic acid fragments (e.g., template nucleic acid fragments and additional template nucleic acid fragments deriving from DNA or RNA molecules included within the cell, cell bead, or cell nucleus) may be co-partitioned with one or more beads (e.g., as described herein). For example, a cell, cell bead, or

cell nucleus may be co-partitioned with a first bead (e.g., first gel bead) configured to interact with a first set of template nucleic acid fragments (e.g., template nucleic acid fragments deriving from DNA molecules, such as tagmented fragments) and a second bead (e.g., second gel bead) configured to interact with a second set of template nucleic acid fragments (e.g., additional template nucleic acid fragments deriving from RNA molecules). The first bead may comprise a first nucleic acid molecule comprising a flow cell adapter sequence, a barcode sequence, and a sequencing primer or portion thereof, which sequencing primer or portion thereof may be configured to interact with (e.g., anneal or hybridize to) a complementary sequence included in template nucleic acid fragments deriving from DNA molecules of the cell, cell bead, or cell nucleus, or derivatives thereof. The second bead may comprise a second nucleic acid molecule comprising the flow cell adapter sequence, the barcode sequence, the sequencing primer or a portion thereof, a UMI sequence, and a capture sequence, which capture sequence may be configured to interact with (e.g., anneal or hybridize to) a sequence of template nucleic acid fragments deriving from RNA molecules of the cell, cell bead, or cell nucleus, or derivatives thereof. In some cases, the capture sequence may be configured to interact with a sequence of a cDNA molecule generated upon reverse transcription of an RNA fragment. The first and second beads may be linked together (e.g., covalently or non-covalently). The first and second beads may each comprise a plurality of nucleic acid molecules. For example, the first bead may comprise a plurality of first nucleic acid molecules and the second bead may comprise a plurality of second nucleic acid molecules, where each first nucleic acid molecule of the plurality of first nucleic acid molecules comprises a first shared sequence and each second nucleic acid molecule of the plurality of second nucleic acid molecules comprises a second shared sequence. The first shared sequence and the second shared sequence may be the same or different. The first shared sequence and the second shared sequence may comprise one or more shared components, such as a shared barcode sequence or sequencing primer or portion thereof.

[00140] Alternatively, a cell, cell bead, or cell nucleus comprising template nucleic acid fragments (e.g., template nucleic acid fragments or additional template nucleic acid fragments deriving from DNA or RNA molecules included within the cell, cell bead, or cell nucleus) may be co-partitioned with a single bead (e.g., gel bead). For example, a cell, cell bead, or cell nucleus may be co-partitioned with a bead comprising (i) a first plurality of nucleic acid barcode molecules configured to interact with a first set of template nucleic acid fragments (e.g., template nucleic acid fragments deriving from DNA molecules, such as tagmented fragments), or derivatives thereof, and (ii) a second plurality of nucleic acid barcode molecules configured to interact with a second set of template nucleic acid fragments (e.g., additional template nucleic

acid fragments deriving from RNA molecules), or derivatives thereof (such as cDNA generated from an RNA fragment). A nucleic acid barcode molecule of the first plurality of nucleic acid barcode molecules may comprise a flow cell adapter sequence, a barcode sequence, and a sequencing primer or portion thereof, which sequencing primer or portion thereof may be configured to interact with (e.g., anneal or hybridize to) a complementary sequence included in template nucleic acid fragments deriving from DNA molecules of the cell, cell bead, or cell nucleus, or derivatives thereof. A nucleic acid barcode molecule of the second plurality of nucleic acid barcode molecules may comprise the flow cell adapter sequence, the barcode sequence, the sequencing primer or a portion thereof, a UMI sequence, and a capture sequence, which capture sequence may be configured to interact with (e.g., anneal or hybridize to) a sequence of template nucleic acid fragments deriving from RNA molecules of the cell, cell bead, or cell nucleus, or derivatives thereof, such as cDNA generated from an RNA fragment. The first plurality of nucleic acid barcode molecules may comprise approximately the same number of nucleic acid barcode molecules as the second plurality of nucleic acid barcode molecules. Alternatively, the first plurality of nucleic acid barcode molecules may comprise a greater number of nucleic acid barcode molecules than the second plurality of nucleic acid barcode molecules, or vice versa. The distribution of nucleic acid barcode molecules on a bead may be controlled by, for example, sequence control, concentration control, and or blocking methods during assembly of the nucleic acid barcode molecules on the bead. Details of such processes are provided in, for example, PCT/US2018/061391, filed November 15, 2018, and US 20190249226, each of which are incorporated by reference in their entireties.

[00141] **FIGs. 8A** and **8B** show examples of beads for use according to the method provided herein. **FIG. 8A** shows a first bead **801** and a second bead **811** that may be co-partitioned with a cell, cell bead, or cell nucleus into a partition of a plurality of partitions (e.g., droplets or wells). First bead **801** may comprise nucleic acid molecule **802**. Nucleic acid molecule **802** may comprise sequences **803**, **804**, and **805**. Sequence **803** may be, for example, a flow cell adapter sequence (e.g., a P5 or P7 sequence). Sequence **804** may be, for example, a barcode sequence. Sequence **805** may be, for example, a sequencing primer sequence or portion thereof (e.g., an R1 or R2 primer sequence, or portion thereof). Nucleic acid molecule **802** may also include additional sequences, such as a UMI sequence. First bead **801** may comprise a plurality of nucleic acid molecules **802**. Second bead **811** may comprise nucleic acid molecule **812**. Nucleic acid molecule **812** may comprise sequences **813**, **814**, and **815**. Sequence **813** may be, for example, a flow cell adapter sequence (e.g., a P5 or P7 sequence). Sequence **814** may be, for example, a barcode sequence. Sequence **815** may be, for example, a sequencing primer

sequence or portion thereof (e.g., an R1 or R2 primer sequence, or portion thereof). Nucleic acid molecule **812** may also include additional sequences, such as a UMI sequence and a capture sequence. Second bead **801** may comprise a plurality of nucleic acid molecules **812**.

[00142] **FIG. 8B** shows a bead **821** (e.g., a multifunctional bead having two or more species of nucleic acid barcode molecules attached or coupled thereto) that may be co-partitioned with a cell, cell bead, or cell nucleus into a partition of a plurality of partitions (e.g., droplets or wells). Bead **821** may comprise nucleic acid molecule **822** and nucleic acid molecule **826**. Nucleic acid molecule **822** may comprise sequences **823**, **824**, and **825**. Sequence **823** may be, for example, a flow cell adapter sequence (e.g., a P5 or P7 sequence). Sequence **824** may be, for example, a barcode sequence. Sequence **825** may be, for example, a sequencing primer or portion thereof (e.g., an R1 or R2 primer sequence, or portion thereof, such as a Nextera R1 sequence or portion thereof). In some instances, sequence **825** may also be, for example, a sequence configured to hybridize to a splint oligonucleotide as described elsewhere herein. Nucleic acid molecule **826** may comprise sequences **827**, **828**, and **829**. Sequence **827** may be, for example, a flow cell adapter sequence (e.g., a P5 or P7 sequence). Sequence **828** may be, for example, a barcode sequence (e.g., the same barcode sequence as sequence **824**). Sequence **829** may be, for example, a sequencing primer or portion thereof (e.g., an R1 or R2 primer sequence, or portion thereof). Sequence **827** may be, for example, a sequencing primer or portion thereof (e.g., an R1 or R2 primer sequence, or portion thereof, such as a TruSeq R1 sequence, or portion thereof). Sequence **828** may be, for example, a barcode sequence (e.g., the same barcode sequence as **824**). Sequence **829** may be, for example, a capture sequence (e.g., a poly-T sequence), such as a capture sequence that is configured to hybridize with a target nucleic acid molecule (e.g., mRNA molecule). Sequence **829** may be, for example, a template switching oligonucleotide (TSO) sequence configured to facilitate a template switching reaction with a target nucleic acid molecule (e.g., mRNA molecule). Sequence **823** and sequence **827** may be the same. Alternatively, sequence **823** and sequence **827** may be different. Sequence **824** and sequence **828** may be the same. Alternatively, sequence **824** and sequence **828** may be different. Sequence **825** and sequence **829** may be the same. Alternatively, sequence **825** and sequence **829** may be different. Nucleic acid molecules **822** and **826** may also include additional sequences, such as a UMI sequence and a capture sequence. Bead **821** may comprise a plurality of nucleic acid molecules **822** and a plurality of nucleic acid molecules **826**.

[00143] Within a partition (e.g., as described herein), an RNA fragment (e.g., a molecule comprising a sequence of an RNA molecule of a cell, cell bead, or cell nucleus that is hybridized to a primer molecule) may be processed to provide a barcoded molecule. The RNA fragment

may be reverse transcribed to generate a complementary cDNA strand, which cDNA strand may be barcoded. In some cases, template switching can be used to increase the length of a cDNA (e.g., via incorporation of one or more sequences, such as one or more barcode or unique molecular identifier sequences). In one example of template switching, cDNA can be generated from reverse transcription of a template (e.g., an mRNA molecule) where a reverse transcriptase with terminal transferase activity can add additional nucleotides, e.g., polyC, to the cDNA that are not encoded by the template, such, as at an end of the cDNA. Template switch oligonucleotides (e.g., switch oligos) can include sequences complementary to the additional nucleotides, e.g. polyG (such as poly-riboG). The additional nucleotides (e.g., polyC) on the cDNA can hybridize to the sequences complementary to the additional nucleotides (e.g., polyG) on the template switch oligonucleotide, whereby the template switch oligonucleotide can be used by the reverse transcriptase as template to further extend the cDNA. Template switch oligonucleotides may comprise deoxyribonucleic acids, ribonucleic acids, modified nucleic acids including locked nucleic acids (LNA), or any combination thereof. A template switch oligonucleotide may comprise one or more sequences including, for example, one or more sequences selected from the group consisting of a sequencing primer, a barcode sequence, a unique molecular identifier sequence, and a homopolymer sequence (e.g., a polyG sequence), or a complement of any of the preceding sequence.

[00144] In some cases, the length of a template switch oligonucleotide may be 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, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250 nucleotides or longer.

[00145] In some cases, an adapter and/or barcode sequence may be added to an RNA molecule via a method other than template switching. For example, one or more sequences may

be ligated to an end of an RNA molecule. Similarly, one or more sequences may be ligated to an end of a cDNA molecule generated via reverse transcription of an RNA molecule.

[00146] In an example, a cell, cell bead, or cell nucleus comprising chromatin and one or more RNA molecules is provided. The chromatin in the cell, cell bead, or cell nucleus may be processed to provide a first template nucleic acid fragment derived from the chromatin (e.g., a tagmented fragment, as described herein). The chromatin may be processed in bulk solution. An RNA molecule may be processed to provide a second template nucleic acid fragment derived from the RNA molecule (e.g., as described herein). The RNA molecule may be processed within a partition. The configuration of the first template nucleic acid fragment may be at least partially dependent on the structure of the transposase-nucleic acid complex used to generate the first template nucleic acid fragment. For example, a transposase-nucleic acid complex such as that shown in **FIG. 9** may be used to prepare the first template nucleic acid fragment. The first template nucleic acid fragment may be at least partially double-stranded. The first template nucleic acid fragment may comprise a double-stranded region comprising sequences of chromatin of the cell, cell bead, or cell nucleus. A first end of a first strand of the double-stranded region may be linked to a first transposon end sequence (e.g., mosaic end sequence), which first transposon end sequence may be linked to a first sequencing primer or portion thereof. A first end of the second strand of the double-stranded region, which end is opposite the first end of the first strand, may be linked to a second transposon end sequence (e.g., mosaic end sequence), which second transposon end sequence may be linked to a second sequencing primer or portion thereof. The second transposon end sequence may be the same as or different from the first transposon end sequence. The first sequencing primer or portion thereof may be the same as or different from the second sequencing primer or portion thereof. In some cases, the first sequencing primer or portion thereof may be an R1 sequence or portion thereof, and the second sequencing primer or portion thereof may be an R2 sequence or portion thereof. The first transposon end sequence may be hybridized to a first complementary sequence (e.g., mosaic end reverse complement sequence), which first complementary sequence may not be linked to a second end of the second strand of the double-stranded region of the first template nucleic acid fragment. Similarly, the second transposon end sequence may be hybridized to a second complementary sequence (e.g., mosaic end reverse complement sequence), which second complementary sequence may not be linked to a second end of the first strand of the double-stranded region of the first template nucleic acid fragment. In other words, the first template nucleic acid fragment may comprise one or more gaps. In some cases, the one or more gaps may be approximately 9 bp in length each. The second template nucleic acid fragment (e.g., an

additional template nucleic acid fragment) may comprise a sequence of an RNA molecule of the cell, cell bead, or cell nucleus and a sequence hybridized to a primer molecule (e.g., a capture nucleic acid molecule). For example, the second template nucleic acid fragment may comprise a sequence of an RNA molecule of the cell, cell bead, or cell nucleus and a polyA sequence hybridized to a polyT sequence of a primer molecule. The primer molecule may also comprise an additional primer sequence.

[00147] The cell, cell bead, or cell nucleus comprising the first template nucleic acid fragment (e.g., tagmented fragment) may be co-partitioned with one or more reagents into a partition of a plurality of partitions (e.g., as described herein). The partition may be, for example, a droplet or well. The partition may comprise one or more beads (e.g., as described herein). A bead of the one or more beads may comprise a first plurality of nucleic acid barcode molecules. A nucleic acid barcode molecule of the first plurality of nucleic acid barcode molecules may comprise one or more of a flow cell adapter sequence (e.g., P5 sequence), a barcode sequence, a sequencing primer or portion thereof (e.g., R1 sequence or portion thereof, or a complement thereof), and a sequence configured to hybridize to a splint oligonucleotide. The sequencing primer or portion thereof may be complementary to a sequence of the first template nucleic acid fragment. In some instances, a nucleic acid barcode molecule of the first plurality of nucleic acid barcode molecules may comprise a flow cell adapter sequence (e.g., P5 sequence), a barcode sequence, and a sequence configured to hybridize to a splint oligonucleotide as described elsewhere herein. A bead of the one or more beads may also comprise a second plurality of nucleic acid barcode molecules. A nucleic acid barcode molecule of the second plurality of nucleic acid barcode molecules may comprise one or more of a flow cell adapter sequence (e.g., P5 sequence), a barcode sequence, a sequencing primer or portion thereof (e.g., R1 sequence or portion thereof, or a complement thereof), and a sequence configured to hybridize to a splint oligonucleotide as described elsewhere herein. In some instances, a nucleic acid barcode molecule of the second plurality of nucleic acid barcode molecules may comprise a sequencing primer or portion thereof (e.g., R1 sequence or portion thereof, or a complement thereof), a barcode sequence, and a capture sequence (e.g., a poly T sequence) configured to hybridize to a nucleic acid molecule (e.g., RNA molecule). In some cases, the first plurality of nucleic acid barcode molecules and the second plurality of nucleic acid barcode molecules may be same.

[00148] Within the partition, the RNA molecule may be processed to provide the second template nucleic acid fragment (e.g., as described herein).

[00149] Within the partition, the cell, cell bead, or cell nucleus may be lysed or permeabilized to provide access to the first and/or second template nucleic acid fragments therein (e.g., as

described herein). The second template nucleic acid fragment may be generated after the cell, cell bead, or cell nucleus is lysed or permeabilized.

[00150] The first and second template nucleic acid fragments may undergo processing within the partition. Within the partition, the gaps in the first template nucleic acid molecule may be filled via a gap filling extension process (e.g., using a DNA polymerase or reverse transcriptase). The resultant double-stranded nucleic acid molecule may be denatured to provide a single strand comprising a chromatin sequence flanked by transposon end sequences and/or sequences complementary to transposon end sequences. Each transposon end sequence and/or sequence complementary to transposon end sequence may be linked to a sequencing primer or portion thereof, or a complement thereof (e.g., an R1 or R2 sequence or a portion thereof, or a complement thereof). A nucleic acid barcode molecule of the first plurality of nucleic acid barcode molecules may hybridize to a sequencing primer or portion thereof, or a complement thereof, of the single strand. A primer extension reaction may then be used to generate a complement of the single strand (e.g., using a DNA polymerase or reverse transcriptase). Such a process may amount to a linear amplification process. This process incorporates the barcode sequence of the nucleic acid barcode molecule of the first plurality of nucleic acid barcode molecules, or a complement thereof. The resultant double-stranded molecule may be denatured to provide a single strand comprising the flow cell adapter sequence, or complement thereof, of the nucleic acid barcode molecule of the first plurality of nucleic acid barcode molecules; barcode sequence, or complement thereof, of the nucleic acid barcode molecule of the first plurality of nucleic acid barcode molecules; sequencing primer or portion thereof, or complement thereof, of the nucleic acid barcode molecule of the first plurality of nucleic acid barcode molecules; transposon end sequences, and/or complements thereof; second sequencing primer or portion thereof, or complement thereof. An additional amplification process may or may not be performed within a partition. For example, exponential amplification may or may not be performed within a partition.

[00151] Within the partition, the second template nucleic acid fragment derived from the RNA molecule of the cell, cell bead, or cell nucleus may be reverse transcribed (e.g., using a reverse transcriptase) to provide a cDNA strand. The reverse transcription process may append a sequence to an end of a strand of the resultant double-stranded nucleic acid molecule comprising the RNA strand and the cDNA strand, such as a polyC sequence. A template switching oligonucleotide may comprise a sequence (e.g., a polyG sequence) that may hybridize to at least a portion of the double-stranded nucleic acid molecule (e.g., to the appended polyC sequence) and be used to further extend the strand of the double-stranded nucleic acid molecule

to provide an extended double-stranded nucleic acid molecule. Such a sequence may comprise ribobases. The template switching oligonucleotide may comprise a UMI sequence, or complement thereof, and a sequencing primer or portion thereof, or complement thereof. The extended double-stranded nucleic acid molecule comprising the template switching oligonucleotide and a complement thereof, and the prior double-stranded nucleic acid molecule may be denatured to provide a single strand comprising a sequencing primer or portion thereof, or complement thereof, of the nucleic acid barcode molecule of the second plurality of nucleic acid barcode molecules; the UMI sequence, or complement thereof; the poly(C) or poly(G) sequence; the sequence corresponding to the RNA molecule of the cell, cell bead, or cell nucleus, or complement thereof; and sequences of the capture nucleic acid molecule, or complements thereof. A nucleic acid barcode molecule of the second plurality of nucleic acid barcode molecules may hybridize to a sequencing primer or portion thereof, or a complement thereof, of the single strand. A primer extension reaction may then be used to generate a complement of the single strand (e.g., using a DNA polymerase). Such a process may amount to a linear amplification process. This process incorporates the barcode sequence of the nucleic acid barcode molecule of the second plurality of nucleic acid barcode molecules, or a complement thereof. The resultant double-stranded molecule may be denatured to provide a single strand comprising a flow cell adapter sequence, or complement thereof, of the nucleic acid barcode molecule of the second plurality of nucleic acid barcode molecules; a barcode sequence, or complement thereof, of the nucleic acid barcode molecule of the second plurality of nucleic acid barcode molecules; a sequencing primer or portion thereof, or complement thereof, of the nucleic acid barcode molecule of the second plurality of nucleic acid barcode molecules; the UMI sequence, or complement thereof; the poly(C) or poly(G) sequence; the sequence corresponding to the RNA molecule of the cell, cell bead, or cell nucleus, or complement thereof; and sequences of the capture nucleic acid molecule, or complements thereof. An additional amplification process may or may not be performed within a partition. For example, exponential amplification may or may not be performed within a partition.

[00152] The linear amplification products corresponding to the chromatin and the RNA molecule of the cell, cell bead, or cell nucleus included within the partition of the plurality of partitions may be recovered from the partition. For example, the contents of the plurality of partitions may be pooled to provide the linear amplification products in a bulk solution. The linear amplification product corresponding to the chromatin may then be subjected to conditions sufficient to undergo one or more nucleic acid amplification reactions (e.g., PCR) to generate one or more amplification products corresponding to the chromatin. A nucleic acid

amplification process may incorporate one or more additional sequences, such as one or more additional flow cell adapter sequences. The linear amplification product corresponding to the RNA molecule may be subjected to fragmentation, end repair, and dA tailing processes. An additional primer sequence (e.g., a sequencing primer or portion thereof, such as an R2 sequence) may then be ligated to the resultant molecule. A nucleic acid amplification reaction (e.g., PCR) may then be performed to generate one or more amplification products corresponding to the RNA molecule. A nucleic acid amplification process may incorporate one or more additional sequences, such as one or more additional flow cell adapter sequences (see, for example, **FIG. 12**).

[00153] In an RNA workflow, in-partition template switching may attach a sequencing primer (e.g., a TruSeq R1 or R2 sequence) to the 3' or 5' end of an RNA transcript. A bead (e.g., gel bead) carrying the sequencing primer, or portion thereof (e.g., partial TruSeq R1 or R2 sequence) may be also used for priming in a DNA (e.g., chromatin) workflow. This may allow for differential amplification of DNA (e.g., ATAC) and RNA libraries after removing materials from partitions (e.g., breaking emulsions) and sample splitting. Another advantage of this method is that the same enzyme (e.g. DNA polymerase or reverse transcriptase) may be used to barcode nucleic acid fragments derived from both DNA (e.g., chromatin) and RNA.

[00154] **FIG. 12** shows an example schematic corresponding to the preceding example. Panel **1200** shows a workflow corresponding to processing of chromatin from a cell, cell bead, or cell nucleus, and panel **1250** shows a workflow corresponding to processing of an mRNA molecule from the cell, cell bead, or cell nucleus. In the figure, two distinct beads (e.g., gel beads) are shown. However, the same bead (e.g., a bead that may be a multifunctional bead) may be used in each workflow.

[00155] As shown in panel **1200**, in bulk solution, chromatin included within a cell, cell bead, or cell nucleus is processed (e.g., as described herein) to provide a template nucleic acid fragment (e.g., tagmented fragment) **1204** comprising insert sequence **1208** (e.g., region of open chromatin) and a complement thereof, transposon end sequences **1206** and complements thereof, sequencing primer or portion thereof **1202** (e.g., an R1 sequence), sequencing primer or portion thereof **1210** (e.g., an R2 sequence), and gaps **1207**. Template nucleic acid fragment **1204** may then be partitioned within a partition (e.g., a droplet or well, as described herein). Within the partition, the cell, cell bead, or cell nucleus comprising template nucleic acid fragment **1204** may be lysed, permeabilized, or otherwise processed to provide access to template nucleic acid fragment **1204** (and one or more RNA molecules) therein. Gaps **1207** may be filled **1212** via a gap filling extension process (e.g., using a DNA polymerase). The partition may include a bead

(e.g., gel bead) **1216a** coupled to a nucleic acid barcode molecule **1218a**. Nucleic acid barcode molecule **1218a** may comprise a flow cell adapter sequence **1220a** (e.g., a P5 sequence), a barcode sequence **1222a**, and a sequencing primer or portion thereof or complement thereof **1202'**. Sequence **1202'** may hybridize to sequence **1202** of template nucleic acid fragment **1204**, or its complement, and undergo primer extension **1214** to yield a strand comprising sequences **1220a, 1222a, 1202', 1210**, and insert sequence **1208** or a complement thereof. The contents of the partition may then be recovered in bulk solution (e.g., a droplet may be broken) to provide the strand in bulk solution. This strand may undergo amplification (e.g., PCR) **1224** to provide a double-stranded amplification product **1226** that includes sequences of the nucleic acid barcode molecule **1218a**, the original chromatin molecule, and, optionally, an additional sequence **1228** that may be a flow cell adapter sequence (e.g., a P7 sequence).

[00156] In parallel to the chromatin workflow of panel **1200**, an RNA molecule deriving from the same cell, cell bead, or cell nucleus may be processed. As shown in panel **1250**, RNA molecule **1258** comprising RNA sequence **1260** and polyA sequence **1262** may be contacted **1264** with primer molecule **1252** comprising polyT sequence **1254** and additional primer sequence **1256**. RNA molecule **1258** may then be reverse transcribed **1266** off of polyT sequence **1254** using a reverse transcriptase with terminal transferase activity, which reverse transcriptase may append sequence **1268** to the resultant cDNA molecule comprising cDNA sequence **1270**. Sequence **1268** may be a polyC sequence. A template switch oligonucleotide **1272** comprising sequencing primer or portion thereof or complement thereof **1274**, unique molecule identifier sequence or complement thereof **1276**, and capture sequence (e.g., polyG sequence) **1278** may then hybridize **1280** to the cDNA molecule and template switching may take place. The partition may include a bead (e.g., gel bead) **1216b** coupled to a nucleic acid barcode molecule **1218b**. Nucleic acid barcode molecule **1218b** may comprise a flow cell adapter sequence **1220b** (e.g., a P5 sequence), a barcode sequence **1222b**, and a sequencing primer or portion thereof or complement thereof **1274'**. Bead (e.g., gel bead) **1216b** may be the same as bead **1216a** such that partition comprises a single bead (e.g., **1218a** and **1218b** are attached to a single bead). In such a case, nucleic acid barcode molecule **1218b** and nucleic acid barcode molecule **1218a** may have the same sequences. Sequence **1274'** may hybridize to sequence **1274** of the cDNA molecule, or its complement, and undergo primer extension **1282** to yield a strand comprising sequences **1220b, 1222b, 1274', 1276** or a complement thereof, **1268** or a complement thereof, and insert sequence **1270** or a complement thereof. The contents of the partition may then be recovered in bulk solution (e.g., a droplet may be broken) to provide the strand in bulk solution. This strand may undergo amplification (e.g., PCR) **1284** to provide a

double-stranded amplification product **1286** that includes sequences of the nucleic acid barcode molecule **1218b**, the original RNA molecule or cDNA corresponding thereto, and, optionally, an additional sequence **1288** that may comprise a sequencing primer or portion thereof (e.g., an R2 sequence) **1290,** a sample index sequence **1292**, and a flow cell adapter sequence (e.g., a P7 sequence) **1294**.

[00157] **FIG. 13** shows another example schematic corresponding to the preceding example. Panel **1300** shows a workflow corresponding to processing of chromatin from a cell, cell bead, or cell nucleus, and panel **1350** shows a workflow corresponding to processing of an mRNA molecule from the cell, cell bead, or cell nucleus. In the figure, two distinct beads (e.g., gel beads) are shown. However, the same bead (e.g., gel bead) may be used in each workflow.

[00158] As shown in panel **1300**, in bulk solution, chromatin included within a cell, cell bead, or cell nucleus is processed (e.g., as described herein) to provide a template nucleic acid fragment (e.g., tagmented fragment) **1304** comprising insert sequence **1308** (e.g., region of open chromatin) and a complement thereof, transposon end sequences **1306** and complements thereof, sequencing primer or portion thereof **1302** (e.g., an R1 sequence), sequencing primer or portion thereof **1310** (e.g., an R2 sequence), and gaps **1307**. Template nucleic acid fragment **1304** may then be partitioned within a partition (e.g., a droplet or well, as described herein). Within the partition, the cell, cell bead, or cell nucleus comprising template nucleic acid fragment **1304** may be lysed, permeabilized, or otherwise processed to provide access to template nucleic acid fragment **1304** (and one or more RNA molecules) therein. Gaps **1307** may be filled **1312** via a gap filling extension process (e.g., using a DNA polymerase). The partition may include a bead (e.g., gel bead) **1316a** coupled to a nucleic acid barcode molecule **1318a**. Nucleic acid barcode molecule **1318a** may comprise a flow cell adapter sequence **1320a** (e.g., a P5 sequence), a barcode sequence **1322a**, and a sequencing primer or portion thereof or complement thereof **1302'**. Sequence **1302'** may hybridize to sequence **1302** of template nucleic acid fragment **1304**, or its complement, and undergo primer extension **1314** to yield a strand comprising sequences **1320a, 1322a, 1302', 1310**, and insert sequence **1308** or a complement thereof. The contents of the partition may then be recovered in bulk solution (e.g., a droplet may be broken) to provide the strand in bulk solution. This strand may undergo amplification (e.g., PCR) **1324** to provide a double-stranded amplification product **1326** that includes sequences of the nucleic acid barcode molecule **1318a**, the original chromatin molecule, and, optionally, an additional sequence **1328** that may be a flow cell adapter sequence (e.g., a P7 sequence).

[00159] In parallel to the chromatin workflow of panel **1300**, an RNA molecule deriving from the same cell, cell bead, or cell nucleus may be processed. As shown in panel **1350**, RNA

molecule **1358** comprising RNA sequence **1360** and polyA sequence **1362** may be contacted with primer molecule **1352** comprising polyT sequence **1354,** UMI sequence **1355,** and sequencing primer or portion thereof (e.g., R1 sequence) **1356**. RNA molecule **1358** may be reverse transcribed **1364** off of polyT sequence **1354** using a reverse transcriptase with terminal transferase activity, which reverse transcriptase may append sequence **1366** (e.g., a polyC sequence) to the resultant cDNA molecule comprising cDNA sequence **1368**. A template switch oligonucleotide **1370** comprising additional primer sequence **1372** and a homopolymer sequence **1374** (e.g., a polyG) sequence that is complementary to sequence **1366** may then hybridize **1376** to the cDNA molecule and template switching may take place. The partition may include a bead (e.g., gel bead) **1316b** coupled to a nucleic acid barcode molecule **1318b**. Nucleic acid barcode molecule **1318b** may comprise a flow cell adapter sequence **1320b** (e.g., a P5 sequence), a barcode sequence **1322b**, and a sequencing primer or portion thereof or complement thereof **1356'**. Bead (e.g., gel bead) **1316b** may be the same as bead (e.g., gel bead) **1316a** such that partition comprises a single bead (i.e., **1318a** and **1318b** are attached to a single bead). In such a case, nucleic acid barcode molecule **1318b** and nucleic acid barcode molecule **1318a** may have the same sequences. Sequence **1356'** may hybridize to sequence **1356** of the cDNA molecule, or its complement, and undergo primer extension **1378** to yield a strand comprising sequences **1320b, 1322b, 1356', 1355** or a complement thereof, **1366** or a complement thereof, and insert sequence **1368** or a complement thereof. The contents of the partition may then be recovered in bulk solution (e.g., a droplet may be broken) to provide the strand in bulk solution. This strand may undergo amplification (e.g., PCR) **1380** to provide a double-stranded amplification product **1382** that includes sequences of the nucleic acid barcode molecule **1318b**, the original RNA molecule or cDNA corresponding thereto, and, optionally, an additional sequence **1384** that may comprise a sequencing primer or portion thereof (e.g., an R2 sequence) **1390,** a sample index sequence **1388**, and a flow cell adapter sequence (e.g., a P7 sequence) **1386**.

[00160] In another example, a cell, cell bead, or cell nucleus comprising chromatin and one or more RNA molecules is provided. The chromatin in the cell, cell bead, or cell nucleus may be processed to provide a first template nucleic acid fragment derived from the chromatin (e.g., a tagmented fragment, as described herein). The chromatin may be processed in bulk solution. An RNA molecule may be processed to provide a second template nucleic acid fragment derived from the RNA molecule (e.g., as described herein). The RNA molecule may be processed within a partition. The configuration of the first template nucleic acid fragment may be at least partially dependent on the structure of the transposase-nucleic acid complex used to generate the first template nucleic acid fragment. For example, a transposase-nucleic acid complex such as

that shown in **FIG. 9** may be used to prepare the first template nucleic acid fragment. The first template nucleic acid fragment may be at least partially double-stranded. The first template nucleic acid fragment may comprise a double-stranded region comprising sequences of chromatin of the cell, cell bead, or cell nucleus. A first end of a first strand of the double-stranded region may be linked to a first transposon end sequence (e.g., mosaic end sequence), which first transposon end sequence may be linked to a first sequencing primer or portion thereof. A first end of the second strand of the double-stranded region, which end is opposite the first end of the first strand, may be linked to a second transposon end sequence (e.g., mosaic end sequence), which second transposon end sequence may be linked to a second sequencing primer or portion thereof. The second transposon end sequence may be the same as or different from the first transposon end sequence. The first sequencing primer or portion thereof may be the same as or different from the second sequencing primer or portion thereof. In some cases, the first sequencing primer or portion thereof may be an R1 sequence or portion thereof, and the second sequencing primer or portion thereof may be an R2 sequence or portion thereof. The first transposon end sequence may be hybridized to a first complementary sequence (e.g., mosaic end reverse complement sequence), which first complementary sequence may not be linked to a second end of the second strand of the double-stranded region of the first template nucleic acid fragment. Similarly, the second transposon end sequence may be hybridized to a second complementary sequence (e.g., mosaic end reverse complement sequence), which second complementary sequence may not be linked to a second end of the first strand of the double-stranded region of the first template nucleic acid fragment. In other words, the first template nucleic acid fragment may comprise one or more gaps. In some cases, the one or more gaps may be approximately 9 bp in length each. For example, one or more gaps may be at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more bp in length. For example, one or more gaps may be at most about 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 bp in length. The second template nucleic acid fragment (e.g., an additional template nucleic acid fragment) may comprise a sequence of an RNA molecule of the cell, cell bead, or cell nucleus and a sequence hybridized to a primer molecule (e.g., a capture nucleic acid molecule). For example, the second template nucleic acid fragment may comprise a sequence of an RNA molecule of the cell, cell bead, or cell nucleus and a polyA sequence hybridized to a polyT sequence of a primer molecule. The primer molecule may also comprise an additional primer sequence.

[00161] The cell, cell bead, or cell nucleus comprising the first template nucleic acid fragment (e.g., tagmented fragment) may be co-partitioned with one or more reagents into a partition of a plurality of partitions (e.g.,The partition may be, for example, a droplet or well. The partition

may comprise one or more beads (e.g., as described herein). A bead (e.g., gel bead) of the one or more beads may comprise a first plurality of nucleic acid barcode molecules. A nucleic acid barcode molecule of the first plurality of nucleic acid barcode molecules may comprise a flow cell adapter sequence (e.g., P5 sequence), a barcode sequence, and a sequencing primer or portion thereof (e.g., R1 sequence or portion thereof, or a complement thereof). The sequencing primer or portion thereof may be complementary to a sequence of the first template nucleic acid fragment. The flow cell adapter sequence and/or barcode sequence may be hybridized to their complementary sequences. A bead (e.g., gel bead) of the one or more beads may also comprise a second plurality of nucleic acid barcode molecules. A nucleic acid barcode molecule of the second plurality of nucleic acid barcode molecules may comprise a flow cell adapter sequence (e.g., P5 sequence), a barcode sequence, a sequencing primer or portion thereof (e.g., R1 sequence or portion thereof, or a complement thereof), a UMI sequence, and a capture sequence (e.g., a polyG sequence, a polydT sequence or target specific sequence). In some cases, the first plurality of nucleic acid barcode molecules and the second plurality of nucleic acid barcode molecules may be coupled to the same bead, and the partition may comprise a single bead.

[00162] Within the partition, the RNA molecule may be processed to provide the second template nucleic acid fragment (e.g., as described herein).

[00163] Within the partition, the cell, cell bead, or cell nucleus may be lysed or permeabilized to provide access to the first and/or second template nucleic acid fragments therein. The second template nucleic acid fragment may be generated after the cell, cell bead, or cell nucleus is lysed or permeabilized.

[00164] The first and second template nucleic acid fragments may undergo processing within the partition. Within the partition, a sequencing primer or portion thereof of the first template nucleic acid fragment corresponding to the chromatin of the cell, cell bead, or cell nucleus may hybridize to a sequencing primer or portion thereof of the nucleic acid barcode molecule of the first plurality of nucleic acid barcode molecules. The sequencing primer or portion thereof of the nucleic acid barcode molecule may then be ligated (e.g., using a ligase) to a transposon end sequence of the first template nucleic acid fragment, or a complement thereof to provide a partially double-stranded nucleic acid molecule corresponding to the chromatin of the cell, cell bead, or cell nucleus.

[00165] Within the partition, the second template nucleic acid fragment derived from the RNA molecule of the cell, cell bead, or cell nucleus may be reverse transcribed (e.g., using a reverse transcriptase) to provide a cDNA strand. The reverse transcription process may append a sequence to an end of a strand of the resultant double-stranded nucleic acid molecule

comprising the RNA strand and the cDNA strand, such as a polyC sequence. The capture sequence of the nucleic acid barcode molecule of the second plurality of nucleic acid barcode molecules may hybridize to the appended sequence (e.g., polyC sequence) of the double-stranded nucleic acid molecule and a template switching process may take place to provide an extended double-stranded nucleic acid molecule. Such a sequence may comprise ribobases. The sequence of the nucleic acid barcode molecule of the second plurality of nucleic acid barcode molecules may be considered a template switching oligonucleotide. Accordingly, barcoding and template switching may take place contemporaneously to provide a barcoded cDNA molecule. The cDNA strand of the barcoded cDNA molecule (e.g., a molecule comprising both a cDNA strand and an RNA strand) may comprise the polyC sequence, a sequence complementary to the sequence of the template switch oligonucleotide or a portion thereof (e.g., sequences complementary to the sequencing primer, barcode sequence, and UMI sequence of the template switch oligonucleotide), the cDNA sequence, the polyT sequence, and the additional primer sequence of the primer molecule. The RNA strand of the barcoded cDNA molecule may comprise the sequence of the template switch oligonucleotide, the mRNA sequence, and a sequence complementary to the additional primer sequence of the primer molecule.

[00166] The partially double-stranded molecule corresponding to the chromatin of the cell, cell bead, or cell nucleus and the barcoded cDNA molecule corresponding to the RNA molecule of the cell, cell bead, or cell nucleus included within the partition (e.g., droplet or well) of the plurality of partitions may be recovered from the partition. For example, the contents of the plurality of partitions may be pooled to provide these products in a bulk solution.

[00167] Outside of the partition, the gaps in the partially double-stranded nucleic acid molecule corresponding to the chromatin may be filled using via a gap filling extension process (e.g., using a DNA polymerase or reverse transcriptase). The gap filling extension process may not include strand displacement. The resultant gap-filled double-stranded nucleic acid molecule may be denatured to provide a single strand, which single strand may be subjected to conditions sufficient to perform one or more nucleic acid amplification reactions (e.g., PCR) to generate amplification products corresponding to the chromatin of the cell, cell bead, or cell nucleus. A nucleic acid amplification process may incorporate one or more additional sequences, such as one or more additional flow cell adapter sequences.

[00168] Outside of the partition, the barcoded cDNA molecule corresponding to the RNA molecule may be subjected to fragmentation, end repair, a dA tailing process, tagmentation, or a combination thereof. An additional primer sequence (e.g., a sequencing primer or portion thereof, such as an R2 sequence) may be ligated to the resultant molecule. Alternatively or in

addition, a nucleic acid amplification reaction (e.g., PCR) may be performed to generate one or more amplification products corresponding to the RNA molecule or the cDNA molecule generated therefrom. A nucleic acid amplification process may incorporate one or more additional sequences, such as one or more additional flow cell adapter sequences.

[00169] **FIG. 14** shows an example schematic corresponding to the preceding example. Panel **1400** shows a workflow corresponding to processing of chromatin from a cell, cell bead, or cell nucleus, and panel **1450** shows a workflow corresponding to processing of an mRNA molecule from the cell, cell bead, or cell nucleus.

[00170] As shown in panel **1400**, in bulk solution, chromatin included within a cell, cell bead, or cell nucleus is processed (e.g., as described herein) to provide a template nucleic acid fragment (e.g., tagmented fragment) **1404** comprising insert sequence **1408** (e.g., region of open chromatin) and a complement thereof, transposon end sequences **1406** and complements thereof, sequencing primer or portion thereof **1402** (e.g., an R1 sequence), sequencing primer or portion thereof **1410** (e.g., an R2 sequence), and gaps **1407**. Template nucleic acid fragment **1404** may then be partitioned within a partition (e.g., a droplet or well, as described herein). Within the partition, the cell, cell bead, or cell nucleus comprising template nucleic acid fragment **1404** may be lysed, permeabilized, or otherwise processed to provide access to template nucleic acid fragment **1404** (and one or more RNA molecules) therein. The partition may include a bead (e.g., gel bead) **1416** coupled to nucleic acid barcode molecules **1418a** and **1418b**. Nucleic acid barcode molecule **1418a** may comprise a flow cell adapter sequence **1420a** (e.g., a P5 sequence), a barcode sequence **1422a**, and a sequencing primer or portion thereof or complement thereof **1402'**. Sequences **1420a** and **1422a** may be hybridized to complementary sequences **1420'** and **1422'**, respectively. Sequence **1402'** may hybridize to sequence **1402** of template nucleic acid fragment **1404**, or its complement, and sequence **1422'** may be ligated **1412** to sequence **1402** of template nucleic acid fragment **1404**. In some instances, template nucleic acid fragment **1404** may be phosphorylated using a suitable kinase enzyme (e.g., polynucleotide kinase (PNK), such as T4 PNK). In some instances, PNK and ATP may be added in bulk in the tagmentation (e.g., ATAC) reaction and/or prior to partitioning the cell, cell bead, or cell nucleus, or plurality thereof. 15U of PNK with 1mM of ATP may be spiked into the tagmentation reaction. For example, less than 95U of PNK may be spiked into the tagmentation reaction. The contents of the partition may then be recovered in bulk solution (e.g., a droplet may be broken) to provide the partially double-stranded nucleic acid molecule comprising nucleic acid barcode molecule **1418a** attached to template nucleic acid fragment **1404** in bulk solution. In bulk solution, gaps **1407** may be filled **1424** via a gap filling extension process (e.g., using a DNA polymerase) to

provide a double-stranded nucleic acid molecule. This molecule may undergo amplification (e.g., PCR) **1426** to provide a double-stranded amplification product **1428** that includes sequences of the nucleic acid barcode molecule **1418a**, the original chromatin molecule, and, optionally, an additional sequence **1430** that may be a flow cell adapter sequence (e.g., a P7 sequence). Gaps may be filled in the partition prior to bulk processing.

[00171] In parallel to the chromatin workflow of panel **1400**, an RNA molecule deriving from the same cell, cell bead, or cell nucleus may be processed. As shown in panel **1450**, RNA molecule **1458** comprising RNA sequence **1460** and polyA sequence **1462** may be contacted **1464** with primer molecule **1452** comprising polyT sequence **1454** and additional primer sequence **1456**. RNA molecule **1458** may then be reverse transcribed **1476** off of polyT sequence **1454** using a reverse transcriptase with terminal transferase activity, which reverse transcriptase may append sequence **1470** to the resultant cDNA molecule comprising cDNA sequence **1468**. Sequence **1470** may be a polyC sequence. Bead (e.g., gel bead) **1416** (e.g., the same bead described in panel **1400**) may be included within the partition and may be coupled to nucleic acid barcode molecule **1418b**. Nucleic acid barcode molecule **1418b** may comprise a flow cell adapter sequence **1420b** (e.g., a P5 sequence), a barcode sequence **1422b**, UMI sequence **1472**, and a sequence **1474** complementary to sequence **1470** (e.g., a polyG sequence). In some instances, nucleic acid barcode molecule **1418b** may comprise a sequencing primer sequence **1420b** (e.g., an R1 sequence or partial R1 sequence), a barcode sequence **1422b**, UMI sequence **1472**, and a template switching sequence **1474** complementary to sequence **1470** (e.g., a polyG sequence). Nucleic acid barcode molecule **1418b** may be used to perform template switching **1478**, which process may also result in the generation of a barcoded cDNA molecule. The contents of the partition may then be recovered in bulk solution (e.g., a droplet may be broken) to provide the barcoded cDNA molecule in bulk solution. The barcoded cDNA molecule may undergo amplification (e.g., PCR) **1480** to provide a double-stranded amplification product **1484** that includes sequences of the nucleic acid barcode molecule **1418b**, the original RNA molecule or cDNA corresponding thereto, a flow cell adapter sequence **1486**, and an additional sequence **1488** that may comprise a sequencing primer or portion thereof (e.g., an R2 sequence) **1490,** a sample index sequence **1492**, and a flow cell adapter sequence (e.g., a P7 sequence) **1494**. The barcoded cDNA molecule may also or alternatively undergo fragmentation, end repair, dA tailing, ligation of one or more adapter sequences, and/or nucleic acid amplification.

[00172] **FIG. 15** shows another example schematic corresponding to the preceding example. Panel **1500** shows a workflow corresponding to processing of chromatin from a cell, cell bead,

or cell nucleus, and panel **1550** shows a workflow corresponding to processing of an mRNA molecule from the cell, cell bead, or cell nucleus.

[00173] As shown in panel **1500**, in bulk solution, chromatin included within a cell, cell bead, or cell nucleus is processed (e.g., as described herein) to provide a template nucleic acid fragment (e.g., tagmented fragment) **1504** comprising insert sequence **1508** (e.g., region of open chromatin) and a complement thereof, transposon end sequences **1506** and complements thereof, sequencing primer or portion thereof **1502** (e.g., an R1 sequence), sequencing primer or portion thereof **1510** (e.g., an R2 sequence), and gaps **1507**. Template nucleic acid fragment **1504** may then be partitioned within a partition (e.g., a droplet or well, as described herein). Within the partition, the cell, cell bead, or cell nucleus comprising template nucleic acid fragment **1504** may be lysed, permeabilized, or otherwise processed to provide access to template nucleic acid fragment **1504** (and one or more RNA molecules) therein. The partition may include a bead (e.g., gel bead) **1516** coupled to nucleic acid barcode molecules **1518a** and **1518b**. Nucleic acid barcode molecule **1518a** may comprise a flow cell adapter sequence **1520a** (e.g., a P5 sequence), a barcode sequence **1522a**, and a sequencing primer or portion thereof or complement thereof **1502'**. Sequences **1520a** and **1522a** may be hybridized to complementary sequences **1520'** and **1522'**, respectively. Sequence **1502'** may hybridize to sequence **1502** of template nucleic acid fragment **1504**, or its complement, and sequence **1522'** may be ligated **1512** to sequence **1502** of template nucleic acid fragment **1504**. In some instances, template nucleic acid fragment **1504** may be phosphorylated using a suitable kinase enzyme (e.g., polynucleotide kinase (PNK), such as T4 PNK). In some instances, PNK and ATP may be added in bulk in the tagmentation (e.g., ATAC) reaction and/or prior to partitioning the cell, cell bead, or cell nucleus, or plurality thereof. 15U of PNK with 1mM of ATP may be spiked into the tagmentation reaction. For example, less than 95U of PNK may be spiked into the tagmentation reaction. The contents of the partition may then be recovered in bulk solution (e.g., a droplet may be broken) to provide the partially double-stranded nucleic acid molecule comprising nucleic acid barcode molecule **1518a** attached to template nucleic acid fragment **1504** in bulk solution. In bulk solution, gaps **1507** may be filled **1524** via a gap filling extension process (e.g., using a DNA polymerase) to provide a double-stranded nucleic acid molecule. This molecule may undergo amplification (e.g., PCR) **1526** to provide a double-stranded amplification product **1528** that includes sequences of the nucleic acid barcode molecule **1518a**, the original chromatin molecule, and, optionally, an additional sequence **1530** that may be a flow cell adapter sequence (e.g., a P7 sequence). Gaps may be filled in the partition prior to bulk processing.

[00174] In parallel to the chromatin workflow of panel **1500**, an RNA molecule deriving from the same cell, cell bead, or cell nucleus may be processed. As shown in panel **1550**, RNA molecule **1558** comprising RNA sequence **1560** and polyA sequence **1562** and bead (e.g., gel bead) **1516** may be provided within a partition. Bead (e.g., gel bead) **1516** (e.g., the same bead described in panel **1500**) may be included within the partition and may be coupled to nucleic acid barcode molecule **1518b**. Nucleic acid barcode molecule **1518b** may comprise a flow cell adapter sequence **1568** (e.g., a P5 sequence), a barcode sequence **1522b** (e.g., the same barcode sequence as barcode sequence **1522a**), UMI sequence **1566**, and a polyT sequence **1564** complementary to polyA sequence **1562**. In some instances, nucleic acid barcode molecule **1518b** may comprise a sequencing primer sequence **1568** (e.g., an R1 sequence or partial R1 sequence), a barcode sequence **1522b** (e.g., the same barcode sequence as barcode sequence **1522a**), UMI sequence **1566**, and a polyT sequence **1564** complementary to polyA sequence **1562**. PolyT sequence **1564** may hybridize to polyA sequence **1562** of RNA molecule **1558**. RNA molecule **1558** may be reverse transcribed **1570** off of polyT sequence **1564** to provide an cDNA molecule comprising cDNA sequence **1572**. The reverse transcription process may use a reverse transcriptase with terminal transferase activity, which reverse transcriptase may append sequence **1574** to the resultant cDNA molecule comprising cDNA sequence **1572**. Sequence **1574** may be a polyC sequence. A template switch oligonucleotide **1578** comprising a primer sequence **1580** and a sequence complementary to sequence **1574** (e.g., a polyG sequence) may hybridize to the cDNA molecule and facilitate a template switching reaction onto template switch oligonucleotide **1578**. The contents of the partition may then be recovered in bulk solution (e.g., a droplet may be broken) to provide the cDNA molecule in bulk solution. The cDNA molecule may undergo amplification (e.g., PCR) **1584**. Additional amplification (e.g., PCR) **1586** may to performed to provide a double-stranded amplification product **1588** that includes sequences of the nucleic acid barcode molecule **1518b**, the original RNA molecule or cDNA corresponding thereto, a flow cell adapter sequence **1598** (e.g., a P7 sequence), and an additional sequence **1590** that may comprise a sequencing primer or portion thereof (e.g., an R2 sequence) **1596,** a sample index sequence **1594**, and a flow cell adapter sequence (e.g., a P5 sequence) **1592**. The barcoded cDNA molecule may also or alternatively undergo fragmentation, end repair, dA tailing, ligation of one or more adapter sequences, and/or nucleic acid amplification.

[00175] In another example, a cell, cell bead, or cell nucleus comprising chromatin and one or more RNA molecules is provided. The chromatin in the cell, cell bead, or cell nucleus may be processed to provide a first template nucleic acid fragment derived from the chromatin (e.g., a

tagmented fragment, as described herein). The chromatin may be processed in bulk solution. An RNA molecule may be processed to provide a second template nucleic acid fragment derived from the RNA molecule (e.g., an additional nucleic acid fragment, as described herein). The RNA molecule may be processed within a partition. The second template nucleic acid fragment derived from the RNA molecule may be processed according to the preceding examples. The configuration of the first template nucleic acid fragment may be at least partially dependent on the structure of the transposase-nucleic acid complex used to generate the first template nucleic acid fragment. For example, a transposase-nucleic acid complex such as that shown in **FIG. 9** may be used to prepare the first template nucleic acid fragment. Relative to the preceding examples, the polarities of the transposase-nucleic acid may be reversed such that sequencing primers (e.g., R1 and R2 sequencing primers) are not directly linked to the chromatin (see, e.g., **FIG. 17**). The first template nucleic acid fragment may be at least partially double-stranded. The first template nucleic acid fragment may comprise a double-stranded region comprising sequences of chromatin of the cell, cell bead, or cell nucleus. A first end of a first strand of the double-stranded region may be linked to a first transposon end sequence (e.g., mosaic end sequence). A first end of the second strand of the double-stranded region, which end is opposite the first end of the first strand, may be linked to a second transposon end sequence (e.g., mosaic end sequence). The second transposon end sequence may be the same as or different from the first transposon end sequence. The first transposon end sequence may be hybridized to a first complementary sequence (e.g., mosaic end reverse complement sequence), which first complementary sequence may not be linked to a second end of the second strand of the double-stranded region of the first template nucleic acid fragment. The first complementary sequence may be linked to a first sequencing primer or portion thereof. Similarly, the second transposon end sequence may be hybridized to a second complementary sequence (e.g., mosaic end reverse complement sequence), which second complementary sequence may not be linked to a second end of the first strand of the double-stranded region of the first template nucleic acid fragment. The second complementary sequence may be linked to a second sequencing primer or portion thereof. In other words, the first template nucleic acid fragment may comprise one or more gaps. In some cases, the one or more gaps may be approximately 9 bp in length each. For example, one or more gaps may be at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more bp in length. For example, one or more gaps may be at most about 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 bp in length. The first sequencing primer or portion thereof may be the same as or different from the second sequencing primer or portion thereof. In some cases, the first sequencing primer or portion

thereof may be an R1 sequence or portion thereof, and the second sequencing primer or portion thereof may be an R2 sequence or portion thereof.

[00176] The cell, cell bead, or cell nucleus comprising the first template nucleic acid fragment (e.g., tagmented fragment) may be co-partitioned with one or more reagents into a partition of a plurality of partitions (e.g., as described herein). The partition may be, for example, a droplet or well. The partition may comprise one or more beads (e.g., as described herein). A bead (e.g., gel bead) of the one or more beads may comprise a first plurality of nucleic acid barcode molecules. A nucleic acid barcode molecule of the first plurality of nucleic acid barcode molecules may comprise a flow cell adapter sequence (e.g., P5 sequence), a barcode sequence, and a sequencing primer or portion thereof (e.g., R1 sequence or portion thereof, or a complement thereof). The sequencing primer or portion thereof may be complementary to a sequence of the first template nucleic acid fragment. The flow cell adapter sequence and/or barcode sequence may be hybridized to their complementary sequences. The same bead or another bead may comprise a second plurality of nucleic acid barcode molecules. A nucleic acid barcode molecule of the second plurality of nucleic acid barcode molecules may comprise a sequencing primer or portion thereof (e.g., an R1 sequence or portion thereof, or complement thereof), a barcode sequence, a unique molecular identifier sequence, and a capture sequence.

[00177] Within the partition, the RNA molecule may be processed to provide the second template nucleic acid fragment (e.g., as described herein). For example, the RNA molecule (e.g., mRNA molecule) may be contacted with a primer molecule comprising a first primer sequence (e.g., a polyT sequence) and an additional primer sequence.

[00178] Within the partition, the cell, cell bead, or cell nucleus may be lysed or permeabilized to provide access to the first and/or second template nucleic acid fragments therein (e.g., as described herein). The second template nucleic acid fragment may be generated after the cell, cell bead, or cell nucleus is lysed or permeabilized.

[00179] The first and second template nucleic acid fragments may undergo processing within the partition. Within the partition, a sequencing primer or portion thereof of the first template nucleic acid fragment corresponding to the chromatin of the cell, cell bead, or cell nucleus may hybridize to a sequencing primer or portion thereof of the nucleic acid barcode molecule of the first plurality of nucleic acid barcode molecules. The sequencing primer or portion thereof of the nucleic acid barcode molecule may then be ligated (e.g., using a ligase) to a transposon end sequence of the first template nucleic acid fragment, or a complement thereof to provide a partially double-stranded nucleic acid molecule corresponding to the chromatin of the cell, cell bead, or cell nucleus. The second template nucleic acid fragment corresponding to the RNA

molecule may be reverse transcribed using a reverse transcriptase with terminal transferase activity, which reverse transcriptase may append a sequence (e.g., a polyC sequence) to the cDNA strand of the resultant cDNA molecule. The cDNA molecule may then be contacted with a nucleic acid barcode molecule of the second plurality of nucleic acid barcode molecules that may be a template switch oligonucleotide. The nucleic acid barcode molecule may comprise a sequencing primer or portion thereof (e.g., an R1 sequence or portion thereof, or complement thereof), a barcode sequence, a unique molecular identifier sequence, and a capture sequence. The capture sequence may be a sequence that is complementary to the sequence appended to the cDNA strand (e.g., a polyG sequence). Template switching and barcoding may then take place to provide a barcoded cDNA molecule.

[00180] The partially double-stranded molecule corresponding to the chromatin of the cell, cell bead, or cell nucleus and the barcoded cDNA molecule corresponding to the RNA molecule (e.g., prepared as described above) of the cell, cell bead, or cell nucleus included within the partition of the plurality of partitions may be recovered from the partition. For example, the contents of the plurality of partitions may be pooled to provide the linear amplification products in a bulk solution.

[00181] Outside of the partition, the gaps in the partially double-stranded nucleic acid molecule corresponding to the chromatin may be filled using via a gap filling extension process (e.g., using a DNA polymerase). Gaps may be filled in the partition prior to bulk processing. The resultant gap-filled double-stranded nucleic acid molecule may be denatured to provide a single strand, which single strand may be subjected to conditions sufficient to perform one or more nucleic acid amplification reactions (e.g., PCR) to generate amplification products corresponding to the chromatin of the cell, cell bead, or cell nucleus. A nucleic acid amplification process may incorporate one or more additional sequences, such as one or more additional flow cell adapter sequences. The barcoded cDNA molecule corresponding to the RNA molecule may also be processed and amplified according to the preceding examples.

[00182] **FIG. 16** shows an example schematic corresponding to the preceding example. Panel **1600** shows a workflow corresponding to processing of chromatin from a cell, cell bead, or cell nucleus, and panel **1650** shows a workflow corresponding to processing of an mRNA molecule from the cell, cell bead, or cell nucleus.

[00183] As shown in panel **1600**, in bulk solution, chromatin included within a cell, cell bead, or cell nucleus is processed (e.g., as described herein) to provide a template nucleic acid fragment (e.g., tagmented fragment) **1604** comprising insert sequence **1608** (e.g., region of open chromatin) and a complement thereof, transposon end sequences **1606** and complements thereof,

sequencing primer or portion thereof **1602** (e.g., an R1 sequence), sequencing primer or portion thereof **1610** (e.g., an R2 sequence), and gaps **1607**. Template nucleic acid fragment **1604** may then be partitioned within a partition (e.g., a droplet or well, as described herein). Within the partition, the cell, cell bead, or cell nucleus comprising template nucleic acid fragment **1604** may be lysed, permeabilized, or otherwise processed to provide access to template nucleic acid fragment **1604** (and one or more RNA molecules) therein. The partition may include a bead (e.g., gel bead) **1616** coupled to a nucleic acid barcode molecules **1618a** and **1618b**. Nucleic acid barcode molecule **1618a** may comprise a flow cell adapter sequence **1620a** (e.g., a P5 sequence), a barcode sequence **1622a**, and a sequencing primer or portion thereof or complement thereof **1602'**. Sequence **1602'** may hybridize to sequence **1602** of template nucleic acid fragment **1604**, or its complement. Sequence **1602**' may then be ligated **1612** to a transposon end sequence **1606** of template nucleic acid fragment **1604**. In some instances, **1604** may be phosphorylated using a suitable kinase enzyme (e.g., polynucleotide kinase (PNK), such as T4 PNK). In some instances, PNK and ATP may be added in bulk in the tagmentation (e.g., ATAC) reaction and/or prior to partitioning the cell, cell beads or cell nucleus, or plurality thereof. 15U of PNK with 1mM of ATP may be spiked into the tagmentation reaction. For example, less than 95U of PNK may be spiked into the tagmentation reaction. The contents of the partition may then be recovered in bulk solution (e.g., a droplet may be broken) to provide the partially double-stranded nucleic acid molecule comprising nucleic acid barcode molecule **1618a** attached to template nucleic acid fragment **1604** in bulk solution. In bulk solution, gaps **1607** may be filled **1614** via a gap filling extension process (e.g., using a DNA polymerase) and the molecule extended from sequence **1602** to provide a double-stranded nucleic acid molecule. This molecule may undergo amplification (e.g., PCR) **1624** to provide a double-stranded amplification product **1626** that includes sequences of the nucleic acid barcode molecule **1618a**, the original chromatin molecule, and, optionally, an additional sequence **1628** that may be a flow cell adapter sequence (e.g., a P7 sequence). Gaps may be filled in the partition prior to bulk processing.

[00184] In parallel to the chromatin workflow of panel **1600**, an RNA molecule deriving from the same cell, cell bead, or cell nucleus may be processed. As shown in panel **1650**, RNA molecule **1658** comprising RNA sequence **1660** and polyA sequence **1662** may be contacted **1664** with primer molecule **1652** comprising polyT sequence **1654** and additional primer sequence **1656**. RNA molecule **1658** may then be reverse transcribed **1676** off of polyT sequence **1654** using a reverse transcriptase with terminal transferase activity, which reverse transcriptase may append sequence **1670** to the resultant cDNA molecule comprising cDNA

sequence **1668**. Sequence **1670** may be a polyC sequence. Bead (e.g., gel bead) **1616** (e.g., the same bead described in panel **1600**) may be included within the partition and may be coupled to nucleic acid barcode molecule **1618b**. Nucleic acid barcode molecule **1618b** may comprise a flow cell adapter sequence **1620b** (e.g., a P5 sequence), a barcode sequence **1622b**, UMI sequence **1672**, and a sequence **1674** complementary to sequence **1670** (e.g., a polyG sequence). In some instances, nucleic acid barcode molecule **1618b** may comprise a sequencing primer sequence **1620b** (e.g., an R1 sequence or partial R1 sequence), a barcode sequence **1622b**, UMI sequence **1672**, and a template switching sequence **1674** complementary to sequence **1670** (e.g., a polyG sequence). Nucleic acid barcode molecule **1618b** may be used to perform template switching **1678**, which process may also result in the generation of a barcoded cDNA molecule. The contents of the partition may then be recovered in bulk solution (e.g., a droplet may be broken) to provide the barcoded cDNA molecule in bulk solution. The barcoded cDNA molecule may undergo amplification (e.g., PCR) **1680** to provide a double-stranded amplification product **1684** that includes sequences of the nucleic acid barcode molecule **1618b**, the original RNA molecule or cDNA corresponding thereto, a flow cell adapter sequence **1686**, and an additional sequence **1688** that may comprise a sequencing primer or portion thereof (e.g., an R2 sequence) **1690,** a sample index sequence **1692**, and a flow cell adapter sequence (e.g., a P7 sequence) **1694**. The barcoded cDNA molecule may also or alternatively undergo fragmentation, end repair, dA tailing, ligation of one or more adapter sequences, and/or nucleic acid amplification.

[00185] **FIG. 17** shows another example schematic corresponding to the preceding example. Panel **1700** shows a workflow corresponding to processing of chromatin from a cell, cell bead, or cell nucleus, and panel **1750** shows a workflow corresponding to processing of an mRNA molecule from the cell, cell bead, or cell nucleus.

[00186] As shown in panel **1700**, in bulk solution, chromatin included within a cell, cell bead, or cell nucleus is processed (e.g., as described herein) to provide a template nucleic acid fragment (e.g., tagmented fragment) **1704** comprising insert sequence **1708** (e.g., region of open chromatin) and a complement thereof, transposon end sequences **1706** and complements thereof, sequencing primer or portion thereof **1702** (e.g., an R1 sequence), sequencing primer or portion thereof **1710** (e.g., an R2 sequence), and gaps **1707**. Template nucleic acid fragment **1704** may then be partitioned within a partition (e.g., a droplet or well, as described herein). Within the partition, the cell, cell bead, or cell nucleus comprising template nucleic acid fragment **1704** may be lysed, permeabilized, or otherwise processed to provide access to template nucleic acid fragment **1704** (and one or more RNA molecules) therein. The partition may include a bead

(e.g., gel bead) **1716** coupled to nucleic acid barcode molecules **1718a** and **1718b**. Nucleic acid barcode molecule **1718a** may comprise a flow cell adapter sequence **1720a** (e.g., a P5 sequence), a barcode sequence **1722a**, and a sequencing primer or portion thereof or complement thereof **1702'**. Sequence **1702'** may hybridize to sequence **1702** of template nucleic acid fragment **1704**, or its complement. Sequence **1702**' may then be ligated **1712** to a transposon end sequence **1706** of template nucleic acid fragment **1704**. In some instances, **1704** may be phosphorylated using a suitable kinase enzyme (e.g., polynucleotide kinase (PNK), such as T4 PNK). In some instances, PNK and ATP may be added in bulk in the tagmentation reaction (e.g., ATAC) and/or prior to partitioning the cell, cell bead, or cell nucleus, or plurality thereof. 15U of PNK with 1mM of ATP may be spiked into the tagmentation reaction. For example, less than 95U of PNK may be spiked into the tagmentation reaction. The contents of the partition may then be recovered in bulk solution (e.g., a droplet may be broken) to provide the partially double-stranded nucleic acid molecule comprising nucleic acid barcode molecule **1718a** attached to template nucleic acid fragment **1704** in bulk solution. In bulk solution, gaps **1707** may be filled **1714** via a gap filling extension process (e.g., using a DNA polymerase) and the molecule extended from sequence **1702** to provide a double-stranded nucleic acid molecule. This molecule may undergo amplification (e.g., PCR) **1724** to provide a double-stranded amplification product **1726** that includes sequences of the nucleic acid barcode molecule **1718a**, the original chromatin molecule, and, optionally, an additional sequence **1728** that may be a flow cell adapter sequence (e.g., a P7 sequence). Gaps may be filled in the partition prior to bulk processing.

[00187] In parallel to the chromatin workflow of panel **1700**, an RNA molecule deriving from the same cell, cell bead, or cell nucleus may be processed. As shown in panel **1750**, RNA molecule **1758** comprising RNA sequence **1760** and polyA sequence **1762** and bead (e.g., gel bead) **1716** may be provided within a partition. Bead (e.g., gel bead) **1716** (e.g., the same bead described in panel **1700**) may be included within the partition and may be coupled to nucleic acid barcode molecule **1718b**. Nucleic acid barcode molecule **1718b** may comprise a flow cell adapter sequence **1768** (e.g., a P5 sequence), a barcode sequence **1722b** (e.g., the same barcode sequence as barcode sequence **1722a**), UMI sequence **1766**, and a polyT sequence **1764** complementary to polyA sequence **1762**. In some entrances, nucleic acid barcode molecule **1718b** may comprise a sequencing primer sequence **1768** (e.g., an R1 sequence or partial R1 sequence), a barcode sequence **1722b** (e.g., the same barcode sequence as barcode sequence **1722a**), UMI sequence **1766**, and a polyT sequence **1764** complementary to polyA sequence **1762**. PolyT sequence **1764** may hybridize to polyA sequence **1762** of RNA molecule **1758**.

RNA molecule **1758** may be reverse transcribed **1770** off of polyT sequence **1764** to provide an cDNA molecule comprising cDNA sequence **1772**. The reverse transcription process may use a reverse transcriptase with terminal transferase activity, which reverse transcriptase may append sequence **1774** to the resultant cDNA molecule comprising cDNA sequence **1772**. Sequence **1774** may be a polyC sequence. A template switch oligonucleotide **1778** comprising a primer sequence **1780** and a sequence complementary to sequence **1774** (e.g., a polyG sequence) may hybridize to the cDNA molecule. The contents of the partition may then be recovered in bulk solution (e.g., a droplet may be broken) to provide the cDNA molecule in bulk solution. The cDNA molecule may undergo amplification (e.g., PCR) **1784**. Additional amplification (e.g., PCR) **1786** may to performed to provide a double-stranded amplification product **1788** that includes sequences of the nucleic acid barcode molecule **1718b**, the original RNA molecule or cDNA corresponding thereto, a flow cell adapter sequence **1798** (e.g., a P7 sequence), and an additional sequence **1790** that may comprise a sequencing primer or portion thereof (e.g., an R2 sequence) **1796,** a sample index sequence **1794**, and a flow cell adapter sequence (e.g., a P5 sequence) **1792**. The barcoded cDNA molecule may also or alternatively undergo fragmentation, end repair, dA tailing, ligation of one or more adapter sequences, and/or nucleic acid amplification.

[00188] In another example, a cell, cell bead, or cell nucleus comprising chromatin and one or more RNA molecules is provided. The chromatin in the cell, cell bead, or cell nucleus may be processed to provide a first template nucleic acid fragment derived from the chromatin (e.g., a tagmented fragment, as described herein). The chromatin may be processed in bulk solution. An RNA molecule may be processed to provide a second template nucleic acid fragment derived from an RNA molecule (e.g., as described herein). The RNA molecule may be processed within a partition. The configuration of the first template nucleic acid fragment may be at least partially dependent on the structure of the transposase-nucleic acid complex used to generate the first template nucleic acid fragment. For example, a transposase-nucleic acid complex such as that shown in **FIG. 9** may be used to prepare the first template nucleic acid fragment. The first template nucleic acid fragment may be at least partially double-stranded. The first template nucleic acid fragment may comprise a double-stranded region comprising sequences of chromatin of the cell, cell bead, or cell nucleus. A first end of a first strand of the double-stranded region may be linked to a first transposon end sequence (e.g., mosaic end sequence), which first transposon end sequence may be linked to a first sequencing primer or portion thereof. A first end of the second strand of the double-stranded region, which end is opposite the first end of the first strand, may be linked to a second transposon end sequence (e.g., mosaic end

sequence), which second transposon end sequence may be linked to a second sequencing primer or portion thereof. The second transposon end sequence may be the same as or different from the first transposon end sequence. The first sequencing primer or portion thereof may be the same as or different from the second sequencing primer or portion thereof. In some cases, the first sequencing primer or portion thereof may be an R1 sequence or portion thereof, and the second sequencing primer or portion thereof may be an R2 sequence or portion thereof. The first transposon end sequence may be hybridized to a first complementary sequence (e.g., mosaic end reverse complement sequence), which first complementary sequence may not be linked to a second end of the second strand of the double-stranded region of the first template nucleic acid fragment. Similarly, the second transposon end sequence may be hybridized to a second complementary sequence (e.g., mosaic end reverse complement sequence), which second complementary sequence may not be linked to a second end of the first strand of the double-stranded region of the first template nucleic acid fragment. In other words, the first template nucleic acid fragment may comprise one or more gaps. In some cases, the one or more gaps may be approximately 9 bp in length each. For example, one or more gaps may be at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more bp in length. For example, one or more gaps may be at most about 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 bp in length. The second template nucleic acid fragment (e.g., an additional template nucleic acid fragment) may comprise a sequence of an RNA molecule of the cell, cell bead, or cell nucleus and a sequence hybridized to a primer molecule (e.g., a capture nucleic acid molecule). For example, the second template nucleic acid fragment may comprise a sequence of an RNA molecule of the cell, cell bead, or cell nucleus and a polyA sequence hybridized to a polyT sequence of a primer molecule. The primer molecule may also comprise an additional primer sequence.

[00189] The cell, cell bead, or cell nucleus comprising the first template nucleic acid fragment (e.g., tagmented fragment) may be co-partitioned with one or more reagents into a partition of a plurality of partitions (e.g., as described herein). The partition may be, for example, a droplet or well. The partition may comprise one or more beads (e.g., as described herein). A bead (e.g., gel bead) of the one or more beads may comprise a first plurality of nucleic acid barcode molecules. A nucleic acid barcode molecule of the first plurality of nucleic acid barcode molecules may comprise a flow cell adapter sequence (e.g., P5 sequence), a barcode sequence, and an overhang sequence. The partition may also comprise a splint sequence comprising a sequence complementary to the overhang sequence and a sequencing primer or portion thereof that may be complementary to a sequence of the first template nucleic acid fragment. A bead of the one or more beads may also comprise a second plurality of nucleic acid barcode molecules.

A nucleic acid barcode molecule of the second plurality of nucleic acid barcode molecules may comprise a flow cell adapter sequence (e.g., P5 sequence), a barcode sequence, a sequencing primer or portion thereof (e.g., R1 sequence or portion thereof, or a complement thereof), a UMI sequence, and a capture sequence (e.g., a polyG sequence or a polydT sequence). In some cases, the first plurality of nucleic acid barcode molecules and the second plurality of nucleic acid barcode molecules may be coupled to the same bead, and the partition may comprise a single bead.

[00190] Within the partition, the RNA molecule may be processed to provide the second template nucleic acid fragment (e.g., as described herein).

[00191] Within the partition, the cell, cell bead, or cell nucleus may be lysed or permeabilized to provide access to the first and/or second template nucleic acid fragments therein (e.g., as described herein). The second template nucleic acid fragment may be generated after the cell, cell bead, or cell nucleus is lysed or permeabilized.

[00192] The first and second template nucleic acid fragments may undergo processing within the partition. Within the partition, a sequencing primer or portion thereof of the first template nucleic acid fragment corresponding to the chromatin of the cell, cell bead, or cell nucleus may hybridize to a complementary sequence of the sequencing primer or portion thereof in the splint sequence. The splint sequence may also hybridize to the overhang sequence of the nucleic acid barcode molecule of the first plurality of nucleic acid barcode molecules. The overhang sequence of the nucleic acid barcode molecule may then be ligated (e.g., using a ligase) to a sequencing primer or portion thereof of the first template nucleic acid fragment. The resultant partially double-stranded nucleic acid molecule may comprise the barcode sequence as well as one or more gaps.

[00193] Within the partition, the second template nucleic acid fragment derived from the RNA molecule of the cell, cell bead, or cell nucleus may be reverse transcribed (e.g., using a reverse transcriptase) to provide a cDNA strand. The reverse transcription process may append a sequence to an end of a strand of the resultant double-stranded nucleic acid molecule comprising the RNA strand and the cDNA strand, such as a polyC sequence. The capture sequence of the nucleic acid barcode molecule of the second plurality of nucleic acid barcode molecules may hybridize to the appended sequence (e.g., polyC sequence) of the double-stranded nucleic acid molecule and a template switching process may take place to provide a second double-stranded nucleic acid molecule. The sequence of the nucleic acid barcode molecule of the second plurality of nucleic acid barcode molecules may be considered a template switching oligonucleotide. The template switch process may result in a barcoded cDNA

molecule. The barcoded cDNA molecule may comprise the sequencing primer or portion thereof, or complement thereof, of the nucleic acid barcode molecule of the second plurality of nucleic acid barcode molecules; the barcode sequence, or complement thereof, of the nucleic acid barcode molecule of the second plurality of nucleic acid barcode molecules; the UMI sequence, or complement thereof, of the nucleic acid barcode molecule of the second plurality of nucleic acid barcode molecules; the capture sequence, or complement thereof, of the nucleic acid barcode molecule of the second plurality of nucleic acid barcode molecules; the poly(C) or poly(G) sequence; the sequence corresponding to the RNA molecule of the cell, cell bead, or cell nucleus, or complement thereof; and sequences of the capture nucleic acid molecule, or complements thereof.

[00194] The partially double-stranded nucleic acid molecule corresponding to the chromatin of the cell, cell bead, or cell nucleus and the barcoded cDNA molecule corresponding to the RNA molecule of the cell, cell bead, or cell nucleus included within the partition of the plurality of partitions may be recovered from the partition. For example, the contents of the plurality of partitions may be pooled to provide the partially double-stranded nucleic acid molecule and the barcoded cDNA molecule in a bulk solution.

[00195] Outside of the partition, the gaps in the partially double-stranded nucleic acid molecule corresponding to the chromatin may be filled using via a gap filling extension process (e.g., using a DNA polymerase or reverse transcriptase). The DNA polymerase may lack strand displacement activity. The resultant gap-filled double-stranded nucleic acid molecule may be denatured to provide a single strand, which single strand may be subjected to conditions sufficient to perform one or more nucleic acid amplification reactions (e.g., PCR) to generate amplification products corresponding to the chromatin of the cell, cell bead, or cell nucleus. A nucleic acid amplification process may incorporate one or more additional sequences, such as one or more additional flow cell adapter sequences.

[00196] Outside of the partition, the barcoded cDNA molecule corresponding to the RNA molecule may be subjected to fragmentation, end repair, a dA tailing process, tagmentation, or a combination thereof. An additional primer sequence (e.g., a sequencing primer or portion thereof, such as an R2 sequence) may then be ligated to the resultant molecule. A nucleic acid amplification reaction (e.g., PCR) may then be performed to generate one or more amplification products corresponding to the RNA molecule. A nucleic acid amplification process may incorporate one or more additional sequences, such as one or more additional flow cell adapter sequences.

[00197] **FIG. 18** shows an example schematic corresponding to the preceding example. Panel **1800** shows a workflow corresponding to processing of chromatin from a cell, cell bead, or cell nucleus, and panel **1850** shows a workflow corresponding to processing of an mRNA molecule from the cell, cell bead, or cell nucleus.

[00198] As shown in panel **1800**, in bulk solution, chromatin included within a cell, cell bead, or cell nucleus is processed (e.g., as described herein) to provide a template nucleic acid fragment (e.g., tagmented fragment) **1804** comprising insert sequence **1808** (e.g., region of open chromatin) and a complement thereof, transposon end sequences **1806** and complements thereof, sequencing primer or portion thereof **1802** (e.g., an R1 sequence), sequencing primer or portion thereof **1810** (e.g., an R2 sequence), and gaps **1807**. Template nucleic acid fragment **1804** may then be partitioned within a partition (e.g., a droplet or well, as described herein). Within the partition, the cell, cell bead, or cell nucleus comprising template nucleic acid fragment **1804** may be lysed, permeabilized, or otherwise processed to provide access to template nucleic acid fragment **1804** (and one or more RNA molecules) therein. The partition may comprise splint sequence **1812**, which splint sequence may comprise a first sequence **1802'** that is complementary to sequencing primer or portion thereof **1802** and a second sequence **1824**. Sequence **1824** may comprise a blocking group (e.g., a 3' blocking group), which blocking group may prevent extension by reverse transcription. The partition may also include a bead (e.g., gel bead) **1816** coupled to nucleic acid barcode molecules **1818a** and **1812b**. Nucleic acid barcode molecule **1818a** may comprise a flow cell adapter sequence **1820a** (e.g., a P5 sequence), a barcode sequence **1822a**, and an overhang sequence **1824'** that is complementary to sequence **1824** of the splint sequence. Sequence **1824** may hybridize to sequence **1824'** to provide a partially double-stranded nucleic acid molecule comprising the sequences of nucleic acid barcode molecule **1818a** and the template nucleic acid fragment **1804**. Sequence **1824'** of nucleic acid barcode molecule **1818a** may be ligated (e.g., using a ligase) **1826** to sequence **1802** of template nucleic acid fragment **1804**. In some instances, **1804** may be phosphorylated using a suitable kinase enzyme (e.g., polynucleotide kinase (PNK), such as T4 PNK). In some instances, PNK and ATP may be added in bulk in the tagmentation reaction (e.g., ATAC) and/or prior to partitioning the cell, cell bead, or cell nucleus, or plurality thereof. 15U of PNK with 1mM of ATP may be spiked into the tagmentation reaction. For example, less than 95U of PNK may be spiked into the tagmentation reaction. The contents of the partition may then be recovered in bulk solution (e.g., a droplet may be broken) to provide the partially double-stranded nucleic acid molecule comprising nucleic acid barcode molecule **1818a** attached to template nucleic acid fragment **1804** in bulk solution. In bulk solution, gaps **1807** may be filled **1828** via a gap filling

extension process (e.g., using a DNA polymerase) to provide a double-stranded nucleic acid molecule. This molecule may undergo amplification (e.g., PCR) **1830** to provide a double-stranded amplification product **1832** that includes sequences of the nucleic acid barcode molecule **1818a**, the original chromatin molecule, and, optionally, an additional sequence **1834** that may be a flow cell adapter sequence (e.g., a P7 sequence). Gaps may be filled in the partition prior to bulk processing.

[00199] In parallel to the chromatin workflow of panel **1800**, an RNA molecule deriving from the same cell, cell bead, or cell nucleus may be processed. As shown in panel **1850**, RNA molecule **1858** comprising RNA sequence **1860** and polyA sequence **1862** may be contacted **1864** with primer molecule **1852** comprising polyT sequence **1854** and additional primer sequence **1856**. RNA molecule **1858** may then be reverse transcribed **1876** off of polyT sequence **1854** using a reverse transcriptase with terminal transferase activity, which reverse transcriptase may append sequence **1870** to the resultant cDNA molecule comprising cDNA sequence **1868**. Sequence **1870** may be a polyC sequence. Bead (e.g., gel bead) **1816** (e.g., the same bead described in panel **1800**) may be included within the partition and may be coupled to nucleic acid barcode molecule **1818b**. Nucleic acid barcode molecule **1818b** may comprise a flow cell adapter sequence **1820b** (e.g., a P5 sequence), a barcode sequence **1822b**, UMI sequence **1872**, and a sequence **1874** complementary to sequence **1870** (e.g., a polyG sequence). In some instances, nucleic acid barcode molecule **1818b** may comprise a sequencing primer sequence **1820b** (e.g., an R1 sequence or partial R1 sequence), a barcode sequence **1822b**, UMI sequence **1872**, and a sequence **1874** complementary to sequence **1870** (e.g., a polyG sequence). Nucleic acid barcode molecule **1818b** may be used to perform template switching **1878**, which process may also result in the generation of a barcoded cDNA molecule. The contents of the partition may then be recovered in bulk solution (e.g., a droplet may be broken) to provide the barcoded cDNA molecule in bulk solution. The barcoded cDNA molecule may undergo amplification (e.g., PCR) **1880** to provide a double-stranded amplification product **1884** that includes sequences of the nucleic acid barcode molecule **1818b**, the original RNA molecule or cDNA corresponding thereto, a flow cell adapter sequence **1886**, and an additional sequence **1888** that may comprise a sequencing primer or portion thereof (e.g., an R2 sequence) **1890,** a sample index sequence **1892**, and a flow cell adapter sequence (e.g., a P7 sequence) **1894**. The barcoded cDNA molecule may also or alternatively undergo fragmentation, end repair, dA tailing, ligation of one or more adapter sequences, and/or nucleic acid amplification.

[00200] **FIG. 19** shows an example schematic corresponding to the preceding example. Panel **1900** shows a workflow corresponding to processing of chromatin from a cell, cell bead, or cell

nucleus, and panel **1950** shows a workflow corresponding to processing of an mRNA molecule from the cell, cell bead, or cell nucleus.

[00201] As shown in panel **1900**, in bulk solution, chromatin included within a cell, cell bead, or cell nucleus is processed (e.g., as described herein) to provide a template nucleic acid fragment (e.g., tagmented fragment) **1904** comprising insert sequence **1908** (e.g., region of open chromatin) and a complement thereof, transposon end sequences **1906** and complements thereof, sequencing primer or portion thereof **1902** (e.g., an R1 sequence), sequencing primer or portion thereof **1910** (e.g., an R2 sequence), and gaps **1907**. Template nucleic acid fragment **1904** may then be partitioned within a partition (e.g., a droplet or well, as described herein). Within the partition, the cell, cell bead, or cell nucleus comprising template nucleic acid fragment **1904** may be lysed, permeabilized, or otherwise processed to provide access to template nucleic acid fragment **1904** (and one or more RNA molecules) therein. The partition may comprise splint sequence **1912**, which splint sequence may comprise a first sequence **1902'** that is complementary to sequencing primer or portion thereof **1902** and a second sequence **1924**. Sequence **1924** may comprise a blocking group (e.g., a 3' blocking group), which blocking group may prevent extension by reverse transcription. The partition may also include a bead (e.g., gel bead) **1916** coupled to nucleic acid barcode molecules **1918a** and **1918b**. Nucleic acid barcode molecule **1918a** may comprise a flow cell adapter sequence **1920a** (e.g., a P5 sequence), a barcode sequence **1922a**, and an overhang sequence **1924'** that is complementary to sequence **1924** of the splint sequence. Sequence **1924** may hybridize to sequence **1924'** to provide a partially double-stranded nucleic acid molecule comprising the sequences of nucleic acid barcode molecule **1918a** and the template nucleic acid fragment **1904**. Sequence **1924'** of nucleic acid barcode molecule **1918a** may be ligated (e.g., using a ligase) **1926** to sequence **1902** of template nucleic acid fragment **1904**. In some instances, **1904** may be phosphorylated using a suitable kinase enzyme (e.g., polynucleotide kinase (PNK), such as T4 PNK). In some instances, PNK and ATP may be added in bulk in the tagmentation reaction (e.g., ATAC) and/or prior to partitioning the cell, cell bead, or cell nucleus, or plurality thereof. 15U of PNK with 1mM of ATP may be spiked into the tagmentation reaction. For example, less than 95U of PNK may be spiked into the tagmentation reaction. The contents of the partition may then be recovered in bulk solution (e.g., a droplet may be broken) to provide the partially double-stranded nucleic acid molecule comprising nucleic acid barcode molecule **1918a** attached to template nucleic acid fragment **1904** in bulk solution. In bulk solution, gaps **1907** may be filled **1928** via a gap filling extension process (e.g., using a DNA polymerase) to provide a double-stranded nucleic acid molecule. This molecule may undergo amplification (e.g., PCR) **1930** to provide a double-

stranded amplification product **1932** that includes sequences of the nucleic acid barcode molecule **1918a**, the original chromatin molecule, and, optionally, an additional sequence **1934** that may be a flow cell adapter sequence (e.g., a P7 sequence). Gaps may be filled in the partition prior to bulk processing.

[00202] In parallel to the chromatin workflow of panel **1900**, an RNA molecule deriving from the same cell, cell bead, or cell nucleus may be processed. As shown in panel **1950**, RNA molecule **1958** comprising RNA sequence **1960** and polyA sequence **1962** and bead (e.g., gel bead) **1916** may be provided within a partition. Bead (e.g., gel bead) **1916** (e.g., the same bead described in panel **1900**) may be included within the partition and may be coupled to nucleic acid barcode molecule **1918b**. Nucleic acid barcode molecule **1918b** may comprise a flow cell adapter sequence **1968** (e.g., a P5 sequence), a barcode sequence **1922b** (e.g., the same barcode sequence as barcode sequence **1922a**), UMI sequence **1966**, and a polyT sequence **1964** complementary to polyA sequence **1962**. In some instances, nucleic acid barcode molecule **1918b** may comprise a sequencing primer sequence **1968** (e.g., a R1 sequence or partial R1 sequence), a barcode sequence **1922b** (e.g., the same barcode sequence as barcode sequence **1922a**), UMI sequence **1966**, and a polyT sequence **1964** complementary to polyA sequence **1962**. PolyT sequence **1964** may hybridize to polyA sequence **1962** of RNA molecule **1958**. RNA molecule **1958** may be reverse transcribed **1970** off of polyT sequence **1964** to provide an cDNA molecule comprising cDNA sequence **1972**. The reverse transcription process may use a reverse transcriptase with terminal transferase activity, which reverse transcriptase may append sequence **1974** to the resultant cDNA molecule comprising cDNA sequence **1972**. Sequence **1974** may be a polyC sequence. A template switch oligonucleotide **1978** comprising a primer sequence **1980** and a sequence complementary to sequence **1974** (e.g., a polyG sequence) may hybridize to the cDNA molecule. The contents of the partition may then be recovered in bulk solution (e.g., a droplet may be broken) to provide the cDNA molecule in bulk solution. The cDNA molecule may undergo amplification (e.g., PCR) **1984**. Additional amplification (e.g., PCR) **1986** may to performed to provide a double-stranded amplification product **1988** that includes sequences of the nucleic acid barcode molecule **1918b**, the original RNA molecule or cDNA corresponding thereto, a flow cell adapter sequence **1998** (e.g., a P7 sequence), and an additional sequence **1990** that may comprise a sequencing primer or portion thereof (e.g., an R2 sequence) **1996,** a sample index sequence **1994**, and a flow cell adapter sequence (e.g., a P5 sequence) **1992**. The barcoded cDNA molecule may also or alternatively undergo fragmentation, end repair, dA tailing, ligation of one or more adapter sequences, and/or nucleic acid amplification.

[00203] In another aspect, the present disclosure provides a method for processing a biological sample (e.g., a nucleic acid sample), which method may comprise performing sequential transcription and reverse transcription processes within a partition. The method may comprise providing a partition (e.g., droplet or well) of a plurality of partitions comprising a nucleic acid molecule (e.g., DNA molecule) derived from a nucleic acid sample. The nucleic acid molecule may be transcribed (e.g., using a transcriptase) to provide an RNA molecule. The RNA molecule may then be reverse transcribed (e.g., using a reverse transcriptase) within the partition to generate a complementary DNA (cDNA) molecule. The cDNA molecule may undergo further processing within the partition to provide a derivative of the cDNA molecule. The cDNA molecule or derivative thereof may be recovered from the partition (e.g., by pooling the contents of the plurality of partitions). The partition may be a well among a plurality of wells. Alternatively, the partition may be a droplet among a plurality of droplets.

[00204] A nucleic acid molecule (e.g., DNA molecule) processed according to the method provided herein may derive from a cell, cell bead, or cell nucleus. In some cases, the nucleic acid molecule may be included within the cell, cell bead, or cell nucleus. The nucleic acid molecule may be chromatin. The cell, cell bead, or cell nucleus comprising the nucleic acid molecule may be included within the partition. For example, the cell, cell bead, or cell nucleus may be co-partitioned with one or more reagents (e.g., as described herein) into a partition (e.g., droplet or well). The cell, cell bead, or cell nucleus may be lysed or permeabilized (e.g., within a partition) to provide access to the nucleic acid molecule therein (e.g., as described herein).

[00205] A nucleic acid molecule processed according to the method provided herein may be a DNA molecule, such as chromatin. In some cases, the method may further comprise processing an open chromatin structure of the nucleic acid sample with a transposase (e.g., included within a transposase-nucleic acid complex) to provide the nucleic acid molecule. For example, a nucleic acid molecule (e.g., within a cell, cell bead, or cell nucleus) may be contacted with a transposase-nucleic acid complex (e.g., as described herein). A transposase used in such a process may be, for example, a Tn5 transposase. A transposase-nucleic acid complex may have a structure such as that of **FIG. 9** or **FIG. 10**. Alternatively, a transposase-nucleic acid complex may comprise one or more transposon end oligonucleotide molecules, which transposon end oligonucleotide molecules comprise hairpin molecules. An example of such a transposase-nucleic acid complex is shown in **FIG. 11**.

[00206] A nucleic acid molecule processed using a transposase-nucleic acid complex comprising one or more hairpin molecules may be a tagmented fragment comprising a double-stranded region comprising sequences corresponding to the nucleic acid molecule (e.g.,

chromatin) of the cell, cell bead, or cell nucleus from which it originates or is derived, as well as one or more hairpin molecules appended to either end of the double-stranded region. For example, the double-stranded region may comprise a first hairpin molecule at one end and a second hairpin molecule at a second end. Generally, only one end of a hairpin molecule may be attached to the double-stranded region, such that the tagmented fragment comprises a gap at either end. For example, a hairpin molecule may be attached to a 3' end of the double-stranded region. The hairpin molecule may comprise a promoter sequence, such as a T7 promoter sequence, and/or a UMI sequence.

[00207] Within the partition, the nucleic acid molecule (e.g., tagmented fragment) may undergo a gap filling process with a reverse transcriptase. The reverse transcriptase enzyme may be a mutant reverse transcriptase enzyme such as, but not limited to, Moloney Murine Leukemia Virus (MMLV) reverse transcriptase. In one aspect, the reverse transcriptase is a mutant MMLV reverse transcriptase such as, but not limited to, enzyme "42B" (see, US Patent Publication No. 20180312822). Enzyme 42B may demonstrated to reduce inhibition of reverse transcription of mRNAs from a single cell due to one or more unknown components present in cell lysate of the single cell when prepared in reaction volumes of less than 1 nanoliter (nL). Enzyme 42B as compared to a commercially available mutant MMLV RT enzyme (CA-MMLV) may show improved reverse transcriptase activity. Such a process may generate a double-stranded nucleic acid molecule comprising the double-stranded region corresponding to the nucleic acid molecule (e.g., chromatin) of the cell, cell bead, or cell nucleus from which it is derived, the sequences of the hairpin molecules at either end of the double-stranded region, and sequences complementary to the sequences of the hairpin molecules. The double-stranded nucleic acid molecule may then undergo transcription with a T7 polymerase, which process begins at an end of a T7 promoter sequences of a hairpin molecules. Both strands may be transcribed in this manner to provide two nucleic acid strands each comprising the T7 promoter sequence, and a complement thereof; one or more transposon end sequences, and one or more complements thereof; and a sequence of the original nucleic acid molecule of the cell, cell bead, or cell nucleus. The strands may also comprise one or more spacer, UMI, or other sequences (e.g., from the hairpin molecules). A strand may then undergo a self-priming process in which the transposon end sequence and complement thereof of a hairpin molecule hybridize to one another to regenerate a hairpin molecule at an end of the strand. The hairpin molecule may serve as the priming site for reverse transcription. A reverse transcriptase process may then be performed (e.g., using a reverse transcriptase). Before, during, or after this process, a sequence may be appended to the end of the molecule, which sequence may be a polyC sequence. A template switching oligonucleotide

comprising a sequence complementary to the appended sequence (e.g., a polyG sequence) may hybridize to the appended sequence. The template switching oligonucleotide may comprise a UMI sequence (e.g., a second UMI sequence that may index transcripts that undergo template switching), a barcode sequence, and/or a priming sequence such as a sequencing primer sequence or portion thereof (e.g., an R1 or R2 sequence, or portion thereof). The template switching oligonucleotide may be attached to a bead (e.g., a gel bead) included within the partition. For example, the template switching oligonucleotide may be a nucleic acid barcode molecule of a plurality of nucleic acid barcode molecules attached to the bead (e.g., as described herein). The resultant partially double-stranded nucleic acid molecule may comprise a hairpin moiety; sequences corresponding to the original nucleic acid molecule of the cell, cell bead, or cell nucleus; and the sequences of the template switching oligonucleotide, including a barcode sequence (see, e.g., **FIG. 20**).

[00208] The partially double-stranded nucleic acid molecule may be released from the partition (e.g., droplet or well). Releasing materials from the partition may comprise breaking or disrupting a droplet. The contents of multiple partitions of the plurality of partitions may be pooled together to provide a bulk solution for further processing. Nucleic acid molecules (e.g., partially double-stranded nucleic acid molecule) of the partitions of the plurality of partitions may each be differentially barcoded such that the nucleic acid molecule of each such partition comprises a different barcode sequence.

[00209] Outside of the partition, the partially double-stranded nucleic acid molecule may be partially denatured to provide a single-stranded molecule (e.g., a single-stranded cDNA molecule). An RNase treatment may be used to remove the hairpin molecule as well as the shorter strand (e.g., the RNA sequence) of the partially double-stranded nucleic acid molecule. The single-stranded molecule remaining may include the template switching oligonucleotide comprising the barcode sequence and, optionally, UMI sequence. A primer molecule comprising a priming sequence complementary to the priming sequence of the template switching oligonucleotide may be provided and may hybridize to the priming sequence of the template switching oligonucleotide. The priming sequence of the primer molecule may be a 5'-blocked priming sequence. A polymerase with dA tailing activity (e.g., a Klenow fragment having 5'→3' polymerase activity, such as an exo-Klenow fragment lacking exonuclease activity) may be used to generate a second nucleic acid strand. The resultant second strand may be dA tailed. The first strand may also be dA tailed. However, if a 5'-blocking priming sequence is used in the preceding processes, the dA tail appended to the first strand may not be available as a hybridization site for another moiety. Instead, a priming sequence comprising a

sequencing primer (e.g., an R1 sequence or complement thereof) and a flow cell adapter sequence (e.g., a P5 sequence or complement thereof) may hybridize to a complementary sequence of the double-stranded nucleic acid molecule. At the opposite end of the double-stranded nucleic acid molecule, the dA moiety appended to the end of the second strand may serve as a site for hybridization of a priming sequence comprising a dT moiety at an end, a sequencing primer (e.g., an R2 sequence or complement thereof), and a flow cell adapter sequence (e.g., a P7 sequence or complement thereof). The double-stranded nucleic acid molecule may then be subjected to conditions sufficient to perform one or more nucleic acid amplification reactions (e.g., PCR) to provide amplification products corresponding to the original nucleic acid molecule of the cell, cell bead, or cell nucleus. The amplification products may comprise flow cell adapter sequences (e.g., P5 and P7 sequences) at either end to facilitate sequencing (e.g., as described herein).

[00210] The method provided herein may overcome certain challenges of performing reverse transcription within partitions. For example, reverse transcriptase may have a DNA-dependent DNA polymerase activity, and/or terminal transferase activities. The latter may result in generation of variable overhangs under certain reaction conditions. In the methods provided herein, every insertion site may be provided a T7 promoter, averting losses that may otherwise be encountered via R1-R1 and R2-R2 interactions. Moreover, both mRNA and chromatin-derived fragments may be barcoded using the same biochemistry (e.g., RT template switching). Performance of linear amplification of both such strands of a nucleic acid molecule may provide strand awareness and introduce a new dimension for, e.g., ATAC-seq processes. Further, this method may enable isothermal linear amplification of transposase derived nucleic acid fragments within partitions. Notably, this method may be combined with any of the RNA workflows described elsewhere herein.

[00211] **FIG. 20** shows a workflow **2000** corresponding to the preceding example. Workflow **2000** may be performed in parallel with an RNA workflow, such as an RNA workflow of any of **FIGs. 12-19**. Multiple beads, each comprising nucleic acid barcode molecules configured for analysis of DNA or RNA molecules, may be included within a partition. Alternatively, a single bead (e.g., gel bead) comprising nucleic acid barcode molecules configured for analysis of both DNA and RNA molecules (e.g., as described herein) may be included within a partition. In some embodiments, the single bead (e.g., in a single partition) comprises a plurality of identical nucleic acid barcode molecules for both RNA and DNA analysis. The single bead (e.g., within a single partition) may comprise a first plurality of nucleic acid barcode molecules for DNA analysis and a second plurality of nucleic acid barcode molecules for RNA molecules, where the

first and second plurality of nucleic acid barcode molecules comprise a common barcode sequence.

[00212] Template nucleic acid fragment (e.g., tagmented fragment) **2002** may be prepared (e.g., using a transposase-nucleic acid complex such as that shown in **FIG. 11**) and provided in a partition (as described herein). Template nucleic acid fragment **2002** may comprise hairpin moieties **2003** and **2004** and target sequences **2005** and **2006**. Template nucleic acid fragment **2002** also comprises gaps **2007**. Gaps **2007** may be filled using a reverse transcriptase (e.g., a 42B enzyme), which process may result in the generation of a double-stranded nucleic acid molecule comprising the double-stranded region corresponding to the original nucleic acid molecule (e.g., chromatin) of the cell, cell bead, or cell nucleus comprising sequences **2005** and **2006** and sequences of the hairpin molecules **2003** and **2004**. The double-stranded nucleic acid molecule may comprise transposon end sequences **2008**, promoter (e.g., T7 promoter) sequences **2010**, and UMI sequences **2012**. The double-stranded nucleic acid molecule may then undergo transcription with a T7 polymerase, which process begins at an end of a T7 promoter sequences of a hairpin molecule. Both strands may be transcribed in this manner to provide two nucleic acid strands. **FIG. 20** shows one such strand comprising T7 promoter sequence **2010**, and a complement thereof; one or more transposon end sequences **2008**, and one or more complements thereof; UMI sequence **2012**, and a complement of a UMI sequence; and an RNA sequence **2006'** corresponding to sequence **2006** of the original nucleic acid molecule of the cell, cell bead, or cell nucleus. The strand may then undergo a self-priming process in which the transposon end sequence and complement thereof of hairpin molecule **2004** hybridize to one another to regenerate a hairpin molecule at an end of the strand. Regenerated hairpin molecule **2004** may serve as the priming site for reverse transcription. Reverse transcription and template switching may then be performed (e.g., using a reverse transcriptase). The reverse transcription process may append sequence **2014** (e.g., a polyC sequence) to the resultant cDNA molecule comprising cDNA sequence **2026** and sequences **2012'** and **2008'** that are complementary to sequences **2012** and **2008**, respectively. The template switching process may comprise the use of a template switch oligonucleotide coupled to bead (e.g., gel bead) **2016** included within the partition. Bead (e.g., gel bead) **2016** may be coupled to nucleic acid barcode molecule **2018** that is the template switch oligonucleotide that comprises sequencing primer or portion thereof **2020**, barcode sequence **2022**, UMI sequence **2024**, and a sequence **2014'** that is complementary to sequence **2014** (e.g., a polyG sequence). The resultant cDNA molecule may comprise a first strand comprising nucleic acid barcode molecule **2018** and RNA sequence **2006'** and a second

strand comprising cDNA sequence **2026**, appended sequence **2014**, and sequences **2020'**, **2022'**, and **2024'** that are complementary to sequences **2020**, **2022**, and **2024**, respectively.

[00213] The cDNA molecule may be released from the partition (e.g., droplet or well). Releasing materials from the partition may comprise breaking or disrupting a droplet. The contents of multiple partitions of the plurality of partitions may be pooled together to provide a bulk solution for further processing. Outside of the partition, the cDNA molecule may be treated with RNase to remove the hairpin molecule as well as the shorter strand (e.g., the RNA sequence) of the partially double-stranded nucleic acid molecule. The single-stranded molecule remaining may include sequences **2020'**, **2022'**, **2024'**, **2014**, **2012'**, **2008'**, and **2026**. Primer molecule **2028** may then hybridize to sequence **2020'**. Primer molecule **2028** may be a 5'-blocked priming sequence. A polymerase with dA tailing activity (e.g., a Klenow fragment having 5'→3' polymerase activity, such as an exo-Klenow fragment lacking exonuclease activity) may be used to generate a second nucleic acid strand comprising sequence **2026'** that is complementary to cDNA sequence **2026**. The resultant second strand may be dA tailed. The first strand may also be dA tailed at an end of sequence **2020'**. However, if a 5'-blocking priming sequence is used in the preceding processes, the dA tail appended to the first strand may not be available as a hybridization site for another moiety. A priming sequence **2030** comprising a dT moiety, a sequencing primer (e.g., an R2 sequence or complement thereof) **2032** and a flow cell adapter sequence (e.g., a P7 sequence or complement thereof) **2034** may hybridize to the dA moiety of the double-stranded nucleic acid molecule. A priming sequence **2036** comprising a sequencing primer (e.g., an R1 sequence or complement thereof) **2038** and a flow cell adapter sequence (e.g., a P5 sequence or complement thereof) **2040** may hybridize to sequence **2028** of the double-stranded nucleic acid molecule. The double-stranded nucleic acid molecule may then be amplified to provide amplified product **2042**, which amplification product may be subjected to further processing such as nucleic acid sequencing.

[00214] **FIG. 21** provides an overview of a workflow **2100** for processing a nucleic acid molecule (e.g., a nucleic acid molecule included within a cell, cell bead, or cell nucleus). The nucleic acid molecule (e.g., DNA molecule, such as chromatin) is tagmented (e.g., as described herein) to generate a tagmented fragment. The tagmented fragment then undergoes transcription, reverse transcription, and barcoding within a partition (e.g., as described herein). The resultant products are released from the partition and subjected to one of two processes, the first of which provides an ATAC library and the second of which provides a gene expression library. The first process may involve RNase treatment to remove RNA and provide cDNA, dA

tailing and ligation of a sequencing primer, and PCR. The second process may involve cDNA amplification; fragmentation, dA tailing, and ligation of a sequencing primer; and PCR.

[00215] The present disclosure also provides a method of processing a nucleic acid molecule of a cell, cell bead, or cell nucleus using a reverse transcriptase fill-in process coupled with a barcoding process. The nucleic acid molecule (e.g., DNA molecule) may derive from a cell, cell bead, or cell nucleus. In some cases, the nucleic acid molecule may be included within the cell, cell bead, or cell nucleus. The nucleic acid molecule may be chromatin. The cell, cell bead, or cell nucleus comprising the nucleic acid molecule may be included within the partition. For example, the cell, cell bead, or cell nucleus may be co-partitioned with one or more reagents (e.g., as described herein) into a partition (e.g., droplet or well). The cell, cell bead, or cell nucleus may be lysed or permeabilized (e.g., within a partition) to provide access to the nucleic acid molecule therein (e.g., as described herein).

[00216] A nucleic acid molecule processed according to the method provided herein may be a DNA molecule, such as chromatin. In some cases, the method may further comprise processing an open chromatin structure of the nucleic acid sample with a transposase (e.g., included within a transposase-nucleic acid complex) to provide the nucleic acid molecule. For example, a nucleic acid molecule (e.g., within a cell, cell bead, or cell nucleus) may be contacted with a transposase-nucleic acid complex (e.g., as described herein). A transposase used in such a process may be, for example, a Tn5 transposase. A transposase-nucleic acid complex may have a structure such as that of **FIG. 9**, **10**, or **11**. Subsequent to generation of a tagmented fragment (e.g., as described herein), the transposase of the transposase-nucleic acid complex may leave or be removed (e.g., displaced, for example, by an enzyme). Alternatively, the transposase may remain in place. The tagmented fragment may comprise sequences corresponding to the original nucleic acid molecule of the cell, cell bead, or cell nucleus; transposon end sequences and sequences complementary thereto; and one or more sequencing primers or portions thereof. A splint sequence comprising a sequence complementary to a sequencing primer or portion thereof the tagmented fragment may hybridize to the sequencing primer or portion thereof. The splint sequence may be ligated to a transposon end sequence or complement thereof of the tagmented fragment (e.g., using a ligase). Prior to or after hybridization and/or ligation of the splint sequence, the tagmented fragment may be partitioned into a partition of a plurality of partitions (e.g., droplets of wells). The tagmented fragment may be co-partitioned with one or more reagents. The tagmented fragment may be included within a cell, cell bead, or cell nucleus, which cell, cell bead, or cell nucleus may be lysed or permeabilized to provide access to the tagmented fragment therein (e.g., as described herein). A sequence of the splint sequence may

then hybridize to a nucleic acid barcode molecule (e.g., a nucleic acid barcode molecule coupled to a bead, as described herein). The bead (e.g., gel bead) may comprise a plurality of nucleic acid barcode molecules, where a nucleic acid barcode molecule of the plurality of nucleic acid barcode molecules may comprise, for example, a flow cell adapter sequence, a barcode sequence, and a UMI sequence. The nucleic acid barcode molecule may also comprise an overhang sequence having sequence complementarity to a sequence of the splint sequence. The overhang sequence may hybridize to the sequence of the splint sequence. A transposase reserved in the tagmented fragment may block gap filling during these processes. The splint sequence may then be extended within the partition (e.g., using a reverse transcriptase).

[00217] Subsequent to the barcoding/template switching and extension (e.g., reverse transcription) processes, the contents of the partition of the plurality of partitions may be released from the partition (e.g., as described herein). Prior or subsequent to release of the contents of the partition, the nucleic acid barcode molecule may be ligated to the sequencing primer of the processed tagmented fragment. Outside of the partition, the nucleic acid barcode molecule may hybridize to the sequencing primer or portion thereof of the template nucleic acid fragment. If a transposase is reserved in the tagmented fragment, the transposase may leave the processed tagmented fragment (e.g., via a strand displacing polymerase) and the remaining gaps may be filled to provide a double-stranded nucleic acid molecule. Alternatively, gaps may be filled as described elsewhere herein. The double-stranded nucleic acid molecule may then be subjected to a nucleic acid amplification process (e.g., PCR, as described herein). Amplification may comprise incorporation of one or more additional sequences, such as one or more flow cell adapter sequences (e.g., P7 sequences).

[00218] **FIG. 22** shows an example schematic corresponding to the preceding example. Panel **2200** shows a workflow corresponding to processing of chromatin from a cell, cell bead, or cell nucleus, and panel **2250** shows a workflow corresponding to processing of an mRNA molecule from the cell, cell bead, or cell nucleus. Multiple beads (e.g., gel beads), each comprising nucleic acid barcode molecules configured for analysis of DNA or RNA molecules, may be included within a partition. Alternatively, a single bead (e.g., gel bead) comprising nucleic acid barcode molecules configured for analysis of both DNA and RNA molecules (e.g., as described herein) may be included within a given partition.

[00219] As shown in panel **2200**, in bulk solution, chromatin included within a cell, cell bead, or cell nucleus is processed (e.g., as described herein) to provide a template nucleic acid fragment (e.g., tagmented fragment) **2204** comprising insert sequence **2208** (e.g., region of open chromatin) and a complement thereof, transposon end sequences **2206** and complements thereof,

sequencing primer or portion thereof **2202** (e.g., an R1 sequence), sequencing primer or portion thereof **2210** (e.g., an R2 sequence), and gaps **2207**. The cell, cell bead, or cell nucleus comprising template nucleic acid fragment **2204** may be lysed, permeabilized, or otherwise processed to provide access to template nucleic acid fragment **2204** (and one or more RNA molecules) therein. Template nucleic acid fragment **2204** may be contacted with splint sequence **2212**, which splint sequence may comprise a first sequence **2202'** that is complementary to sequencing primer or portion thereof **2202** and a second sequence **2224**. Sequence **2224** may comprise a blocking group (e.g., a 3' blocking group), which blocking group may prevent extension by reverse transcription. Sequence **2202'** may hybridize **2214** to sequence **2202** of template nucleic acid fragment **2204** to provide a partially double-stranded nucleic acid molecule comprising splint sequence **2212** and template nucleic acid fragment **2204**. Sequence **2202'** may be ligated **2226** to the complement of transposon end sequence **2206** of template nucleic acid fragment **2204** (e.g., using a ligase). Template nucleic acid fragment **2204** attached to splint sequence **2212** may then be partitioned within a partition (e.g., droplet or well) within a plurality of partitions (e.g., as described herein). The partition may also include a bead (e.g., gel bead) **2216** coupled to nucleic acid barcode molecules **2218a** and **2218b**. Nucleic acid barcode molecule **2218a** may comprise a flow cell adapter sequence **2220a** (e.g., a P5 sequence), a barcode sequence **2222a**, and an overhang sequence **2224'** that is complementary to sequence **2224** of the splint sequence **2212**. Sequence **2224** may hybridize **2228** to sequence **2224'**. Splint sequence **2212** may then be extended **2230** (e.g., using a reverse transcriptase or DNA polymerase) to provide sequences **2220a'** and **2222a'** that are complementary to sequences **2220a** and **2222a** of nucleic acid barcode molecule **2218a**. Alternatively, sequence **2224** may hybridize to sequence **2224'** to provide a partially double-stranded nucleic acid molecule and nucleic acid barcode molecule **2218a** may be ligated (e.g., using a ligase) to sequence **2202** of template nucleic acid fragment **2204**. The contents of the partition may then be recovered in bulk solution (e.g., a droplet may be broken) to provide the partially double-stranded nucleic acid molecule comprising nucleic acid barcode molecule **2218a** attached to splint sequence **2212** and template nucleic acid fragment **2204** in bulk solution. Sequence **2224'** of nucleic acid barcode molecule **2218a** may be ligated (e.g., using a ligase) **2232** to sequence **2202** of template nucleic acid fragment **2204**. In bulk solution, gaps **2207** may be filled **2234** via a gap filling extension process (e.g., using a DNA polymerase) to provide a double-stranded nucleic acid molecule. This molecule may also undergo amplification (e.g., PCR) to provide a double-stranded amplification product **2236** that includes sequences of the nucleic acid barcode molecule **2218a**, the original chromatin molecule, and, optionally, an additional sequence **2238** that may be a flow

cell adapter sequence (e.g., a P7 sequence). Gaps may be filled in the partition prior to bulk processing.

[00220] In parallel to the chromatin workflow of panel **2200**, an RNA molecule deriving from the same cell, cell bead, or cell nucleus may be processed. As shown in panel **2250**, RNA molecule **2258** comprising RNA sequence **2260** and polyA sequence **2262** may be contacted **2264** with primer molecule **2252** comprising polyT sequence **2254** and additional primer sequence **2256**. RNA molecule **2258** may then be reverse transcribed **2276** off of polyT sequence **2254** using a reverse transcriptase with terminal transferase activity, which reverse transcriptase may append sequence **2270** to the resultant cDNA molecule comprising cDNA sequence **2268**. Sequence **2270** may be a polyC sequence. Bead (e.g., gel bead) **2216** (e.g., the same bead described in panel **2200**) may be included within the partition and may be coupled to nucleic acid barcode molecule **2218b**. Nucleic acid barcode molecule **2218b** may comprise a flow cell adapter sequence **2220b** (e.g., a P5 sequence), a barcode sequence **2222b**, UMI sequence **2272**, and a sequence **2274** complementary to sequence **2270** (e.g., a polyG sequence). In some instances. nucleic acid barcode molecule **2218b** may comprise a sequencing primer sequence **2220b** (e.g., an R1 sequence or partial R1 sequence), a barcode sequence **2222b**, UMI sequence **2272**, and a sequence **2274** complementary to sequence **2270** (e.g., a polyG sequence). Nucleic acid barcode molecule **2218b** may be used to perform template switching **2278**, which process may also result in the generation of a barcoded cDNA molecule. The contents of the partition may then be recovered in bulk solution (e.g., a droplet may be broken) to provide the barcoded cDNA molecule in bulk solution. The barcoded cDNA molecule may undergo amplification (e.g., PCR) **2280** to provide a double-stranded amplification product **2284** that includes sequences of the nucleic acid barcode molecule **2218b**, the original RNA molecule or cDNA corresponding thereto, a flow cell adapter sequence **2286**, and an additional sequence **2288** that may comprise a sequencing primer or portion thereof (e.g., an R2 sequence) **2290,** a sample index sequence **2292**, and a flow cell adapter sequence (e.g., a P7 sequence) **2294**. The barcoded cDNA molecule may also or alternatively undergo fragmentation, end repair, dA tailing, ligation of one or more adapter sequences, and/or nucleic acid amplification.

[00221] **FIG. 23** shows another example schematic corresponding to the preceding example. Panel **2300** shows a workflow corresponding to processing of chromatin from a cell, cell bead, or cell nucleus, and panel **2350** shows a workflow corresponding to processing of an mRNA molecule from the cell, cell bead, or cell nucleus. Multiple beads (e.g., gel beads), each comprising nucleic acid barcode molecules configured for analysis of DNA or RNA molecules, may be included within a partition. Alternatively, a single bead (e.g., gel bead) comprising

nucleic acid barcode molecules configured for analysis of both DNA and RNA molecules (e.g., as described herein) may be included within a given partition.

[00222] As shown in panel **2300**, in bulk solution, chromatin included within a cell, cell bead, or cell nucleus is processed (e.g., as described herein) to provide a template nucleic acid fragment (e.g., tagmented fragment) **2304** comprising insert sequence **2308** (e.g., region of open chromatin) and a complement thereof, transposon end sequences **2306** and complements thereof, sequencing primer or portion thereof **2302** (e.g., an R1 sequence), sequencing primer or portion thereof **2310** (e.g., an R2 sequence), and gaps **2307**. The cell, cell bead, or cell nucleus comprising template nucleic acid fragment **2304** may be lysed, permeabilized, or otherwise processed to provide access to template nucleic acid fragment **2304** (and one or more RNA molecules) therein. Template nucleic acid fragment **2304** may be contacted with splint sequence **2312**, which splint sequence may comprise a first sequence **2302'** that is complementary to sequencing primer or portion thereof **2302** and a second sequence **2324**. Sequence **2324** may comprise a blocking group (e.g., a 3' blocking group), which blocking group may prevent extension by reverse transcription. Sequence **2302'** may hybridize **2314** to sequence **2302** of template nucleic acid fragment **2304** to provide a partially double-stranded nucleic acid molecule comprising splint sequence **2312** and template nucleic acid fragment **2304**. Sequence **2302'** may be ligated **2326** to the complement of transposon end sequence **2306** of template nucleic acid fragment **2304** (e.g., using a ligase). Template nucleic acid fragment **2304** attached to splint sequence **2312** may then be partitioned within a partition (e.g., droplet or well) within a plurality of partitions (e.g., as described herein). The partition may also include a bead (e.g., gel bead) **2316** coupled to nucleic acid barcode molecules **2318a** and **2318b**. Nucleic acid barcode molecule **2318a** may comprise a flow cell adapter sequence **2320a** (e.g., a P5 sequence), a barcode sequence **2322a**, and an overhang sequence **2324'** that is complementary to sequence **2324** of the splint sequence **2312**. Sequence **2324** may hybridize **2328** to sequence **2324'**. Splint sequence **2312** may then be extended **2330** (e.g., using a reverse transcriptase or DNA polymerase) to provide sequences **2320a'** and **2322a'** that are complementary to sequences **2320a** and **2322a** of nucleic acid barcode molecule **2318a**. Alternatively, sequence **2324** may hybridize to sequence **2324'** to provide a partially double-stranded nucleic acid molecule and nucleic acid barcode molecule **2318a** may be ligated (e.g., using a ligase) to sequence **2302** of template nucleic acid fragment **2304**. The contents of the partition may then be recovered in bulk solution (e.g., a droplet may be broken) to provide the partially double-stranded nucleic acid molecule comprising nucleic acid barcode molecule **2318a** attached to splint sequence **2312** and template nucleic acid fragment **2304** in bulk solution. Sequence **2324'** of nucleic acid

barcode molecule **2318a** may be ligated (e.g., using a ligase) **2332** to sequence **2302** of template nucleic acid fragment **2304**. In bulk solution, gaps **2307** may be filled **2334** via a gap filling extension process (e.g., using a DNA polymerase) to provide a double-stranded nucleic acid molecule. This molecule may also undergo amplification (e.g., PCR) to provide a double-stranded amplification product **2336** that includes sequences of the nucleic acid barcode molecule **2318a**, the original chromatin molecule, and, optionally, an additional sequence **2338** that may be a flow cell adapter sequence (e.g., a P7 sequence). Gaps may be filled in the partition prior to bulk processing.

[00223] In parallel to the chromatin workflow of panel **2300**, an RNA molecule deriving from the same cell, cell bead, or cell nucleus may be processed. As shown in panel **2350**, RNA molecule **2358** comprising RNA sequence **2360** and polyA sequence **2362** and gel bead **2316** may be provided within a partition. Bead (e.g., gel bead) **2316** (e.g., the same bead described in panel **2300**) may be included within the partition and may be coupled to nucleic acid barcode molecule **2318b**. Nucleic acid barcode molecule **2318b** may comprise a flow cell adapter sequence **2368** (e.g., a P5 sequence), a barcode sequence **2322b** (e.g., the same barcode sequence as barcode sequence **2322a**), UMI sequence **2366**, and a polyT sequence **2364** complementary to polyA sequence **2362**. In some instances, nucleic acid barcode molecule **2318b** may comprise a sequencing primer sequence **2368** (e.g., a R1 sequence or partial R1 sequence), a barcode sequence **2322b** (e.g., the same barcode sequence as barcode sequence **2322a**), UMI sequence **2366**, and a polyT sequence **2364** complementary to polyA sequence **2362**. PolyT sequence **2364** may hybridize to polyA sequence **2362** of RNA molecule **2358**. RNA molecule **2358** may be reverse transcribed **2370** off of polyT sequence **2364** to provide an cDNA molecule comprising cDNA sequence **2372**. The reverse transcription process may use a reverse transcriptase with terminal transferase activity, which reverse transcriptase may append sequence **2374** to the resultant cDNA molecule comprising cDNA sequence **2372**. Sequence **2374** may be a polyC sequence. A template switch oligonucleotide **2378** comprising a primer sequence **2380** and a sequence complementary to sequence **2374** (e.g., a polyG sequence) may hybridize to the cDNA molecule. The contents of the partition may then be recovered in bulk solution (e.g., a droplet may be broken) to provide the cDNA molecule in bulk solution. The cDNA molecule may undergo amplification (e.g., PCR) **2384**. Additional amplification (e.g., PCR) **2386** may to performed to provide a double-stranded amplification product **2388** that includes sequences of the nucleic acid barcode molecule **2318b**, the original RNA molecule or cDNA corresponding thereto, a flow cell adapter sequence **2398** (e.g., a P7 sequence), and an additional sequence **2390** that may comprise a sequencing primer or portion thereof (e.g., an R2

sequence) **2396,** a sample index sequence **2394**, and a flow cell adapter sequence (e.g., a P5 sequence) **2392**. The barcoded cDNA molecule may also or alternatively undergo fragmentation, end repair, dA tailing, ligation of one or more adapter sequences, and/or nucleic acid amplification.

Cell characterization

[00224] In an aspect, the present disclosure provides a method for characterizing cells and/or cell nuclei. For example, the present disclosure provides a method for characterizing a plurality of cells and/or cell nuclei as belonging to different classes of cells (e.g., cell types) and/or originating from different sources, such as from different tissues or organs. For example, the present disclosure provides a method for identifying tumor-associated cells and/or cell nuclei among a plurality of cells and/or cell nuclei.

[00225] A method for characterizing a cell or cell nucleus may comprise providing a partition (e.g., droplet) comprising the cell or cell nucleus and a particle (e.g., gel bead). The cell or cell nucleus may comprise a plurality of nucleic acid molecules, which plurality of nucleic acid molecules may comprise a plurality of ribonucleic acid (RNA) molecules and a plurality of deoxyribonucleic acid (DNA) molecules. The plurality of DNA molecules may comprise chromatin (e.g., as described herein). The particle may comprise a plurality of nucleic acid barcode molecules coupled thereto (e.g., as described herein). The plurality of nucleic acid barcode molecules may be releasably coupled to the particle and may be releasable from the particle upon application of a stimulus, such as a chemical stimulus (e.g., a reducing agent such as DTT). The plurality of nucleic acid barcode molecules may be coupled to the particle via a labile moiety (e.g., as described herein). The plurality of nucleic acid barcode molecules coupled to the particle may all be identical. Alternatively, the plurality of nucleic acid barcode molecules may comprise one or more different nucleic acid sequences. For example, each nucleic acid barcode molecule of the plurality of nucleic acid barcode molecules may comprise a unique molecular identifier sequence. One or more other sequences of the plurality of nucleic acid barcode molecules may be identical. For example, each nucleic acid barcode molecule of the plurality of nucleic acid barcode molecules may comprise an identical nucleic acid barcode sequence. Nucleic acid barcode molecules of the plurality of nucleic acid barcode molecules coupled to the particle and nucleic acid molecules of the plurality of nucleic acid molecules of the cell or cell nucleus may be used to generate a plurality of barcoded nucleic acid molecules (e.g., as described herein). Examples of workflows for generating barcoded nucleic acid molecules are described elsewhere herein and shown in, for example, **FIGs. 12-23**. The plurality of barcoded nucleic acid molecules may comprise (i) a first subset comprising

sequences corresponding to an RNA molecule of the plurality of RNA molecules and (ii) a second subset comprising sequences corresponding to a DNA molecule of the plurality of DNA molecules. Each barcoded nucleic acid molecule of the plurality of barcoded nucleic acid molecules may comprise a common nucleic acid barcode sequence. The plurality of barcoded nucleic acid barcode molecules, or derivatives thereof, may be processed to generate sequencing information corresponding to the DNA molecule and the RNA molecule. The sequencing information may be used to characterize the cell or cell nucleus. For example, the sequencing information may be used to identify a type of the cell or cell nucleus. The cell or cell nucleus may be identified as having a type selected from the group consisting of, for example, monocytes, natural killer cells, B cells, T cells, granulocytes, plasmacytoid dendritic cells, dendritic cells, and stromal cells. Specific subclasses of such cell types may also be identified. For example, the cell or cell nucleus may be identified as being a CD14 monocyte, CD16 monocyte, a myeloid dendritic cell, a plasmacytoid dendritic cell, replicating B cell, normal B cell, tumor B cell, naïve B cell, memory B cell, plasmablast B cells, lymphoplasmacytoid cells, B-1 cells, regulatory B cells, plasma B cells, IgM+ memory B cell, IgD+ memory B cell, regulatory B cells, plasma B cell, replicating T cell, normal T cell, helper T cells, cytotoxic T cells, memory T cells, regulatory T cells, natural killer T cells, mucosal associated invariant T (MAIT) cell, CD8+ MAIT cell, CD8 gamma delta T cells, effector T cell, CD4 memory T cell, naïve T cell, or another cell type.

[00226] A method for characterizing cells and/or cell nuclei may involve generating sequencing information corresponding to ribonucleic acid (RNA) molecules (e.g., as described herein) and sequencing information corresponding to deoxyribonucleic acid (DNA) molecules (e.g., as described herein). RNA sequencing information may include information relating to gene expression and is thus alternately referred to herein as "gene expression data." For example, RNA sequencing information may include information derived from messenger RNA (mRNA), which information may provide insight into proteins that may be translated from the mRNA. DNA sequencing information may include information relating to regions of accessible chromatin (e.g., tagmented fragments) and is thus alternately referred to herein as "chromatin data" or "ATAC data." For example, DNA sequencing information may include information derived from regions of accessible or open chromatin (e.g., regions of chromatin between nucleosomes). Generating RNA and DNA sequencing information may comprise preparation of barcoded nucleic acid molecules comprising nucleic acid barcode sequences corresponding to specific cells and/or cell nuclei. For example, a cell or cell nucleus may be provided in a partition and subjected to parallel workflows to process DNA molecules (e.g., chromatin) and

RNA molecules included therein. Examples of such workflows are shown in, e.g., **FIGs. 12-23**. The resultant barcoded nucleic acid molecules may include a first set of barcoded nucleic acid molecules corresponding to DNA molecules of the cell or cell nucleus and a second set of barcoded nucleic acid molecules corresponding to RNA molecules of the cell or cell nucleus. Each barcoded nucleic acid molecule may include a nucleic acid barcode sequence, such as a nucleic acid barcode sequence provided to the partition as a component of a nucleic acid barcode molecule coupled to a particle (e.g., gel bead). The nucleic acid barcode sequence of the barcoded nucleic acid molecules generated for a given cell or cell nucleus associated with a given partition (e.g., droplet or well) may be identical across the barcoded nucleic acid molecules generated for the given cell or cell nucleus. Therefore, barcoded nucleic acid molecules corresponding to both DNA and RNA molecules of the given cell or cell nucleus may comprise the same nucleic acid barcode sequence.

[00227] Where a plurality of cells or cell nuclei are processed (e.g., within a plurality of partitions, as described elsewhere herein), barcoded nucleic acid molecules associated with each cell or cell nucleus may comprise different nucleic acid barcode sequences (e.g., as described herein). Processing a plurality of cells and/or cell nuclei in this manner may generate a first data set comprising sequencing information corresponding to a first plurality of barcoded nucleic acid molecules derived from DNA molecules (e.g., chromatin or tagmented fragments) of the plurality of cells and cell nuclei and a second data set comprising sequencing information corresponding to a second plurality of barcoded nucleic acid molecules derived from RNA molecules of the plurality of cells and cell nuclei. The plurality of cells or cell nuclei may be derived from a sample comprising a tumor or suspected of comprising a tumor. The plurality of cells or cell nuclei may be derived from a sample obtained from a subject, such as a human subject. The subject may be known to have or have had a tumor and/or a proliferative disease (e.g., cancer). Alternatively, the subject may be suspected of having a tumor and/or a proliferative disease (e.g., cancer). The sample may be derived from a bodily fluid, such as blood and/or plasma. The sample may be derived from a biopsy, such as a tumor biopsy. A tumor may be a B cell lymphoma tumor.

[00228] **FIG. 24** schematically illustrates how cells of a plurality of cells can be grouped into cell types (e.g., modalities) using parallel DNA (e.g., chromatin) and RNA processing (e.g., as described herein). As described herein, nucleic acid barcode sequences of barcoded nucleic acid molecules analyzed via nucleic acid sequencing methods (e.g., as described herein) can be used to link DNA (e.g., open chromatin) and RNA (e.g., gene expression) information for particular cells of the plurality of cells.

[00229] **FIG. 25** schematically illustrates an example workflow for generating DNA (e.g., chromatin) and RNA (e.g., gene expression) information corresponding to a plurality of cells or cell nuclei. As shown in the left-most panel, a plurality of transposed nuclei and a plurality of particles (e.g., gel beads) coupled to a plurality of nucleic acid barcode molecules may be provided. Each particle of the plurality of particles may be coupled to nucleic acid barcode molecules of the plurality of nucleic acid barcode molecules comprising a common nucleic acid barcode sequence (e.g., as described herein). The plurality of transposed nuclei and the plurality of particles may be processed using a microfluidic device to generate a plurality of droplets (e.g., aqueous droplets in oil). At least a subset of the plurality of droplets may comprise both a particle of the plurality of particles and a transposed nucleus of the plurality of transposed nuclei. The plurality of droplets may also comprise one or more reagents for analyzing DNA and/or RNA molecules of the plurality of transposed nuclei (e.g., as described herein). The plurality of droplets may be subjected to conditions sufficient to process the DNA and/or RNA molecules of transposed nuclei of the plurality of transposed nuclei to generate a plurality of barcoded nucleic acid molecules (e.g., using ligation, reverse transcription, etc., as described elsewhere herein), where each barcoded nucleic acid molecule of the plurality of barcoded nucleic acid molecules includes a nucleic acid barcode sequence of a nucleic acid barcode molecule of the plurality of nucleic acid barcode molecules coupled to the plurality of particles. The plurality of barcoded nucleic acid molecules may be recovered from the plurality of particles and may undergo additional processing, including nucleic acid amplification to generate copies of the plurality of barcoded nucleic acid molecules and/or to append additional sequences such as sequencing primers and flow cell sequences to the plurality of barcoded nucleic acid molecules. The plurality of barcoded nucleic acid molecules may then be processed to generate sequencing libraries corresponding to accessible chromatin and gene expression for the transposed nuclei of the plurality of transposed nuclei. Note that this workflow may also be performed with cells or cell nuclei that have not yet been subjected to transposition (e.g., a transposition process may be performed within the partitions, as described elsewhere herein).

[00230] **FIG. 39** schematically illustrates the data generated via parallel DNA (e.g., chromatin) and RNA processing (e.g., as described herein). First data set **3900** corresponds to DNA molecules of a plurality of cells (here, four cells) and includes sequencing reads **3901, 3902, 3903, 3904, 3905, 3906, 3907**, and **3908**. Each sequencing read includes a barcode sequence **3911, 3912, 3913**, or **3914**. Barcode sequence **3911** is associated with a first cell ("Cell 1"). Barcode sequence **3912** is associated with a second cell ("Cell 2"). Barcode sequence **3913** is associated with a third cell ("Cell 3"). Barcode sequence **3914** is associated

with a third cell ("Cell 4"). Each sequencing read also includes a sequence corresponding to a sequence of a DNA molecule (e.g., chromatin) of a cell of the plurality of cells. For example, sequencing read **3901** includes sequence **3915**. Sequencing reads **3902, 3903, 3904, 3905, 3906, 3907**, and **3908** include sequences **3916, 3917, 3918, 3919, 3920, 3921**, and **3922**, respectively. Second data set **3930** corresponds to RNA molecules of the plurality of cells and includes sequencing reads **3931, 3932, 3933, 3934, 3935, 3936, 3937**, and **3938**. Each sequencing read includes a barcode sequence **3911, 3912, 3913**, or **3914**, indicating that the sequencing read is associated with a particular cell of the plurality of cells (e.g., Cell 1, Cell 2, Cell 3, or Cell 4). Each sequencing read also includes a sequence corresponding to a sequence of an RNA molecule of a cell of the plurality of cells. For example, sequencing read **3931** includes sequence **3945**. Sequencing reads **3932, 3933, 3934, 3935, 3936, 3937**, and **3938** include sequences **3946, 3947, 3948, 3949, 3950, 3951**, and **3952**, respectively. First data set **3900** and second data set 39**30** can be processed (e.g., using an algorithm to identify nucleic acid barcode sequences within sequencing reads and associate sequencing reads comprising common nucleic acid barcode sequences with one another) to generate third data set **3960**. Third data set **3960** includes both RNA and DNA sequencing information corresponding to cells of the plurality of cells. As shown in the figure, barcode sequences can be used to identify RNA and DNA sequencing reads as corresponding to particular cells of the plurality of cells. For example, sequencing reads **3901, 3902, 3931**, and **3932** can be identified as deriving from Cell 1 based on their shared nucleic acid barcode sequence **3911**. Similarly, sequencing reads **3903, 3904, 3933**, and **3934** can be identified as deriving from Cell 2 based on barcode sequence **3912**; sequencing reads **3905, 3906, 3935**, and **3936** can be identified as deriving from Cell 3 based on barcode sequence **3913**; and sequencing reads **3907, 3908, 3937**, and **3938** can be identified as deriving from Cell 4 based on barcode sequence **3914**. As described herein, the complementary RNA (e.g., gene expression) and DNA (e.g., accessible chromatin) information associated with the cells of the plurality of cells may be used to characterize the cells. For example, gene expression and/or accessible chromatin data for Cell 1 may be used to identify Cell 1 as a tumor B cell, while gene expression and/or accessible chromatin data for Cell 2 may be used to identify Cell 2 as a normal B cell.

[00231] Sequencing information obtained for DNA molecules (e.g., chromatin) and RNA molecules (e.g., as described herein) may be used to characterize cells and/or cell nuclei of a plurality of cells from which it derives. In some cases, gene expression information alone may be sufficient to identify a cell type for a given cell or cell nucleus or collection thereof of a plurality of cells and/or cell nuclei. In other cases, accessible chromatin information alone may

be sufficient to identify a cell type for a given cell or cell nucleus or collection thereof of a plurality of cells and/or cell nuclei. In some cases, gene expression and accessible chromatin information can be used to identify a cell type for a given cell or cell nucleus or collection thereof of a plurality of cells and/or cell nuclei. The combination of gene expression and accessible chromatin information may be especially useful for identifying cell types across a plurality of cells, such as for a plurality of cells comprising at least about 100, 200, 300, 400, 500, 600, 700, 800, 900, 1,000, 2,000, 3,000, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 15,000, 20,000, 25,000, 30,000, 40,000, 50,000, 60,000, 70,000, 80,000, 90,000, 100,000, or more cells.

[00232] **FIG. 26** shows gene expression profiling of 24,000 peripheral blood mononuclear cells (PBMCs), while **FIG. 27** shows ATAC profiling of the same cells. Cells may be clustered by cell type, as shown in the left and right panels of the figures. As shown, cells may be characterized at a coarse or granular level and multiple different cell types may be identified among the plurality of cells. In this example, cells may be identified at a more granular analysis using gene expression than by using ATAC information. However, the inverse may be true (e.g., for different cell samples). As shown in **FIGs. 26** and **27**, cells may be identified as having a type selected from, for example, monocytes, natural killer cells, B cells, T cells, granulocytes, dendritic cells, and stromal cells. B cells may be identified as, for example, normal B cells, replicating B cells, tumor B cells, naïve B cells, memory B cells, IgM+ memory B cells, IgD+ memory B cells, IgM+IgD+ memory B cells, and plasma B cells. T cells may be identified as, for example, replicating T cells, normal T cells, mucosal associated invariant T (MAIT) cells, CD8+ MAIT cells, CD8 effector T cells, CD4 memory T cells, and naïve T cells. Monocytes may be identified as, for example, CD14 monocytes and CD16 monocytes. Dendritic cells may be identified as, for example, myeloid dendritic cells and plasmacytoid dendritic cells. **FIGs. 28** and **29A-B** illustrate concordance between the gene expression and chromatin-based characterization of cells shown in **FIGs. 26** and **27**, respectively. The overlap shown in the central panel of **FIG. 28** is a result of the clusters in the gene expression and chromatin cell type representations being oriented differently. As shown in **FIG. 29A-B**, the cell type characterizations made using chromatin data may be subclassified into narrower cell classes upon correlation with gene expression data. For example, monocytes may be subdivided into CD14 monocyte and CD16 monocyte classes. Similarly, B cells may be subdivided into naïve/memory B cells, IgM+IgD+ memory B cells, etc. Reannotation of chromatin data using gene expression markers is illustrated in **FIG. 30**. **FIG. 31** compares different representations of cell clustering based on open chromatin analysis (upper left panel), gene expression analysis

(upper right panel), and open chromatin analysis annotated using gene expression markers (lower panel).

[00233] Sequencing information corresponding to DNA molecules (e.g., chromatin information) and sequencing information corresponding to RNA molecules (e.g., gene expression information) may be used in tandem to refine cell type classifications. For example, as described above, gene expression information may be used to subclassify (e.g., annotate) cell types identified using chromatin information. Similarly, chromatin information may be used to subclassify or disentangle cell types identified using gene expression information. An example of such a process is shown in **FIG. 32**. The upper panels of **FIG. 32** display cell classification based on chromatin information annotated with gene expression markers (upper left panel) and cell classification based on gene expression information alone (upper right panel). As shown in the classification based on gene expression information alone (upper right panel), naïve/memory B cells may be identified as a single cluster. However, in the classification based on chromatin information annotated with gene expression markers (upper left panel), naïve B cells and memory B cells are identified as two distinct clusters (e.g., they have distinct chromatin signatures). These subclusters are highlighted in the lower left panel as subclusters 1 and 2. The lower right panel illustrates that these subclusters may be obscured when gene expression alone is used to classify cell types. Gene expression analysis of subclusters 1 and 2 identifies subcluster 1 as prospective memory B cells based on the relatively higher Ig and relatively lower naïve B cell associated transcripts, while subcluster 2 may be identified as naïve B cells. Therefore, complementary gene expression and chromatin analyses may be useful in differentiating between a wide variety of cell types among a population of cells or cell nuclei that may not be distinguishable using a single analysis alone.

[00234] The methods described herein may be applied to analysis of cell populations including diseased cells such as tumor cells. The methods provided herein may comprise using sequencing information to identify a presence of a tumor cell or cell nucleus in a sample. Thus, the methods provided herein may be useful in diagnostic evaluations. The methods provided herein may also or alternatively comprise using sequencing information to identify a cell type, a cell state, a tumor-specific gene expression pattern, or a tumor-specific differentially accessible region of chromatin in a sample. Based at least in part on such analysis, a therapeutic regimen may be identified, e.g., for a subject from which a sample derives. A therapeutic regimen may comprise administration of a therapeutically effective amount of an agent targeting one or more targets identified in a tumor-specific gene expression pattern or a tumor-specific differentially accessible region of chromatin.

[00235] The methods provided herein may also be used to correlate genotypes and phenotypes at a single cell level. Such analyses may provide unique insights into various diseases and biological processes. For example, such analysis may provide insight into the development, diagnosis, and treatment of various cancers and other disorders in which somatic mutations may be key features. Gene expression analysis coupled with chromatin analysis may also be used to identify tumor-specific signaling pathways: gene expression analysis may identify active receptors for a particular cell type (e.g., tumor cells), while chromatin analysis may provide insight into transcription factor activities.

[00236] A method of characterizing a plurality of cells or cell nuclei may comprise providing a first data set corresponding to regions of accessible chromatin of a plurality of DNA molecules (e.g., chromatin) of a plurality of cells or cell nuclei and a second data set corresponding to a plurality of RNA molecules of the plurality of cells or cell nuclei. The first data set may comprise sequencing information (e.g., a first plurality of sequencing reads) corresponding to sequences of the regions of accessible chromatin and a plurality of nucleic acid barcode sequences. The second data set may comprise sequencing information (e.g., a second plurality of sequencing reads) corresponding to sequences of the plurality of RNA molecules and the plurality of nucleic acid barcode sequences. A cell or cell nucleus of the plurality of cells or cell nuclei may correspond to a nucleic acid barcode sequences of the plurality of nucleic acid barcode sequences. For example, a particle comprising a plurality of nucleic acid barcode molecules comprising a common nucleic acid barcode sequence of the plurality of nucleic acid barcode sequences may be used to process the plurality of DNA molecules and RNA molecules of a cell or cell nucleus of the plurality of cells or cell nuclei (e.g., within a partition) (as described herein). The plurality of nucleic acid barcode sequences of the first data set and the second data set may be used to identify first sequencing reads of the first plurality of sequencing reads and second sequencing reads of the second plurality of sequencing reads as corresponding to cells or cell nuclei of the plurality of cells or cell nuclei, thereby generating a third data set comprising sequence information corresponding to regions of accessible chromatin and RNA molecules associated with cells or cell nuclei of the plurality of cells or cell nuclei. Cell types of the cells or cell nuclei may be identified using the sequence information of the third data set.

[00237] In various embodiments, an example data analysis workflow can include one or more of the following analysis operations: a gene expression data processing operation, an ATAC data processing operation, a joint cell calling operation, a gene expression analysis operation, an ATAC analysis operation, and an ATAC and RNA analysis operation, or any combinations thereof. It is understood that, certain operations within the disclosure can be used either

independently or in combination with other operations within the disclosure, while certain other operations within the disclosure can only be used in combination with certain other operations within the disclosure. Further, one or more of the operations or filters described below, presumably defaulted to be utilized as part of the computational pipeline for analyzing both the gene expression sequencing data and the single cell ATAC sequencing data, can also not be utilized per user input. It is understood that the reverse is also contemplated. It is further understood that additional operations for analyzing the sequencing data generated by the single cell sequencing workflow are also contemplated as part of the computational pipeline within the disclosure.

Gene Expression Data Processing

[00238] The gene expression data processing operation can comprise processing the barcodes in the single cell sequencing data set for fixing the occasional sequencing error in the barcodes so that the sequenced fragments can be associated with the original barcodes, thus improving the data quality.

[00239] The barcode processing operation can include checking each barcode sequence against a "whitelist" of correct barcode sequences. The barcode processing operation can further include counting the frequency of each whitelist barcode. The barcode processing operation can also include various barcode correction operations as part of the various embodiments disclosed herein. For example, one may attempt to correct the barcodes that are not included on the whitelist by finding all the whitelisted barcodes that are within 2 differences (Hamming distance <= 2) of the observed sequence, and then scoring them based on the abundance of that barcode in the read data and quality value of the incorrect bases. As another example, an observed barcode that is not present in the whitelist can be corrected to a whitelist barcode if it has > 90% probability of being the real barcode based.

[00240] The gene expression data processing can further comprise aligning the read sequences (also referred to as the "reads") to a reference sequence. In the alignment operation of the various embodiments herein, a reference-based analysis is performed by aligning the read sequences (also referred to as the "reads") to a reference sequence. The reference sequence for the various embodiments herein can include a reference transcriptome sequence (including genes and introns) and its associated genome annotations, which include gene and transcript coordinates. The reference transcriptome sequence and annotations of various embodiments herein can be obtained from reputable, well-established consortia, including but not limited to NCBI, GENCODE, Ensembl, and ENCODE. In various embodiments, the reference sequence can include single species and/or multi-species reference sequences. In various embodiments,

systems and methods within the disclosure can also provide pre-built single and multi-species reference sequences. In various embodiments, the pre-built reference sequences can include information and files related to regulatory regions including, but not limited to, annotation of promoter, enhancer, CTCF binding sites, and DNase hypersensitivity sites. In various embodiments, systems and methods within the disclosure can also provide building custom reference sequences that are not pre-built.

[00241] Various embodiments herein can be configured to correct for sequencing errors in the UMI sequences, before UMI counting. Reads that were confidently mapped to the transcriptome can be placed into groups that share the same barcode, UMI, and gene annotation. If two groups of reads have the same barcode and gene, but their UMIs differ by a single base (i.e., are Hamming distance 1 apart), then one of the UMIs was likely introduced by a substitution error in sequencing. In this case, the UMI of the less-supported read group is corrected to the UMI with higher support.

[00242] After grouping the reads by barcode, UMI (possibly corrected), and gene annotation, if two or more groups of reads have the same barcode and UMI, but different gene annotations, the gene annotation with the most supporting reads is kept for UMI counting, and the other read groups can be discarded. In case of a tie for maximal read support, all read groups can be discarded, as the gene cannot be confidently assigned.

[00243] After these two filtering operations, each observed barcode, UMI, gene combination is recorded as a UMI count in an unfiltered feature-barcode matrix, which contains every barcode from fixed list of known-good barcode sequences. This includes background and cell associated barcodes. The number of reads supporting each counted UMI is also recorded in the molecule info file.

[00244] The gene expression data processing can further comprise annotating the individual cDNA fragment reads as exonic, intronic, intergenic, and by whether they align to the reference genome with high confidence. In various embodiments, a fragment read is annotated as exonic if at least a portion of the fragment intersects an exon. In various embodiments, a fragment read is annotated as intronic if it is non-exonic and intersects an intron. The annotation process can be determined by the alignment method and its parameters/settings as performed, for example, using the STAR aligner.

[00245] The gene expression data processing can further comprise unique molecule processing to better identify certain subpopulations such as for example, low RNA content cells, a unique molecule processing operation can be performed prior to cell calling. For low RNA content cells, such a operation is important, particularly when low RNA content cells are mixed

into a population of high RNA content cells. The unique molecule processing can include a high content (e.g., RNA content) capture operation and a low content capture operation.

ATAC Data Processing

[00246] The ATAC data processing operation can comprise processing the barcodes in the single cell ATAC sequencing data for fixing the occasional sequencing error in the barcodes so that the sequenced fragments can be associated with the original barcodes, thus improving the data quality.

[00247] The barcode processing operation can include checking each barcode sequence against a "whitelist" of correct barcode sequences. The barcode processing operation can further include counting the frequency of each whitelist barcode. The barcode processing operation can also include various barcode correction operations as part of the various embodiments disclosed herein. For example, one may attempt to correct the barcodes that are not included on the whitelist by finding all the whitelisted barcodes that are within 2 differences (Hamming distance <= 2) of the observed sequence, and then scoring them based on the abundance of that barcode in the read data and quality value of the incorrect bases. As another example, an observed barcode that is not present in the whitelist can be corrected to a whitelist barcode if it has > 90% probability of being the real barcode based.

[00248] The ATAC data processing operation can further comprise aligning the read sequences (also referred to as the "reads") to a reference sequence. One of more sub-operations can be utilized for trimming off adapter sequences, primer oligo sequences, or both in the read sequence before the read sequence is aligned to the reference genome.

[00249] The ATAC data processing operation can further comprise marking sequencing and PCR duplicates and outputting high quality de-duplicated fragments. One or more sub-operations can be employed for identifying duplicate reads, such as sorting aligned reads by 5' position to account for transposition event and identifying groups of read-pairs and original read-pair. The process may further include filters that, when activated in various embodiments herein, can determine whether a fragment is mapped with MAPQ > 30 on both reads (i.e., includes a barcode overlap for reads with mapping quality below 30), not mitochondrial, and not chimerically mapped.

[00250] The ATAC data processing operation can comprise a peak calling analysis that includes counting cut sites in a window around each base-pair of the genome and thresholding it to find regions enriched for open chromatin. Peaks are regions in the genome enriched for accessibility to transposase enzymes. Only open chromatin regions that are not bound by nucleosomes and regulatory DNA-binding proteins (e.g., transcription factors) are accessible by

transposase enzymes for ATAC sequencing. Therefore, the ends of each sequenced fragment of the various embodiments herein can be considered to be indicative of a region of open chromatin. Accordingly, the combined signal from these fragments can be analyzed in accordance with various embodiments herein to determine regions of the genome enriched for open chromatin, and thereby, to understand the regulatory and functional significance of such regions. Therefore, using the sites as determined by the ends of the fragments in the position-sorted fragment file (e.g., the fragments.tsv.gz file) described above, the number of transposition events at each base-pair along the genome can be counted. In one embodiment within the disclosure, the cut sites in a window around each base-pair of the genome is counted.

Joint Cell Calling analysis

[00251] The joint cell calling analysis operation can comprise a cell calling analysis that includes associating a subset of barcodes observed in both the single cell gene expression library and the single cell ATAC library to the cells loaded from the sample. Identification of these cell barcodes can allow one to then analyze the variation and quantification in data at a single cell resolution.

[00252] The process may further include correction of gel bead artifacts, such as gel bead multiples (where a cell shares more than one barcoded gel bead) and barcode multiplets (which occurs when a cell associated gel bead has more than one barcode). In some embodiments, the operations associated with cell calling and correction of gel bead artifacts are utilized together for performing the necessary analysis as part of the various embodiments herein.

[00253] In accordance with various embodiments, the record of mapped high-quality fragments that passed all the filters of the various embodiments disclosed in the operations above and were indicated as a fragment in the fragment file (e.g., the fragments.tsv file), are recorded. With the peaks determined in the peak calling operation disclosed herein, the number of fragments that overlap any peak regions, for each barcode, can be utilized to separate the signal from noise, i.e., to separate barcodes associated with cells from non-cell barcodes. It is to be understood that such method of separation of signal from noise works better in practice as compared to naively using the number of fragments per barcode.

[00254] Various methods, in accordance with various embodiments herein, can be employed for joint cell calling. In various embodiments, the joint cell calling can be performed in at least two operations. In the first operation of cell calling of the various embodiments herein, the barcodes that have fraction of fragments overlapping called peaks lower than the fraction of genome in peaks are identified. When this first operation is employed in the cell calling process

of the various embodiments herein, the peaks are padded by 2000 bp on both sides so as to account for the fragment length for this calculation.

Gene Expression Analysis

[00255] The gene expression analysis operation can comprise generating a feature-barcode matrix that summarizes that gene expression counts per each cell. The feature-barcode matrix can include only detected cellular barcodes. The generation of the feature-barcode matrix can involve compiling the valid non-filtered UMI counts per gene (e.g., output from the 'Unique Molecule Processing' operation discussed herein) from each cell-associated barcode (e.g., output from the 'Cell Calling operation discussed above) together into the final output count matrix, which can then be used for downstream analysis operations.

[00256] The gene expression analysis operation can comprise various dimensionality reduction, clustering, t-SNE and UMAP projection tools. Dimensionality reduction tools of the various embodiments herein are utilized to reduce the number of random variables under consideration by obtaining a set of principal variables. In accordance with various embodiments herein, clustering tools can be utilized to assign objects of the various embodiments herein to homogeneous groups (called clusters) while ensuring that objects in different groups are not similar. t-SNE and UMAP projection tools of the various embodiments herein can include an algorithm for visualization of the data of the various embodiments herein. In accordance with various embodiments, systems and methods within the disclosure can further include dimensionality reduction, clustering and t-SNE and UMAP projection tools. In some embodiments, the analysis associated with dimensionality reduction, clustering, and t-SNE and UMAP projection for visualization are utilized together for performing the necessary analysis as part of the various embodiments herein. Various analysis tools for dimensionality reduction include Principal Component Analysis (PCA), Latent Semantic Analysis (LSA), and Probabilistic Latent Semantic Analysis (PLSA), clustering, and t-SNE and UMAP projection for visualization that allow one to group and compare a population of cells with another.

[00257] In some embodiments, the systems and methods within the disclosure are directed to identifying differential gene expression. As the data is sparse at single cell resolution, dimensionality reduction in accordance with various embodiments herein can be performed to cast the data into a lower dimensional space.

[00258] In accordance with various embodiments, the gene expression analysis operation can comprise a differential expression analysis that performs differential analysis to identify genes whose expression is specific to each cluster, Cell Ranger tests, for each gene and each cluster, whether the in-cluster mean differs from the out-of-cluster mean.

ATAC Analysis

[00259] The ATAC analysis can comprise determining the peak-barcode matrix. In accordance with various embodiments, in the ATAC analysis operation, a raw peak-barcode matrix can be generated first, which is a count matrix consisting of the counts of fragment ends (or cut sites) within each peak region for each barcode. This raw peak-barcode matrix captures the enrichment of open chromatin per barcode. The raw matrix can then be filtered to consist only of cell barcodes by filtering out the non-cell barcodes from the raw peak-barcode matrix, which can then be used in the various dimensionality reduction, clustering and visualization operations of the various embodiments herein.

[00260] The ATAC analysis operation can comprise various dimensionality reduction, clustering and t-SNE projection tools, similar to as described above in the gene expression analysis operation.

[00261] The ATAC analysis operation can comprise annotating the peaks by performing gene annotations and discovering transcription factor-motif matches on each peak. It is contemplated that peak annotation can be employed with subsequent differential analysis operations within various embodiments of the disclosure. Various peak annotation procedures and parameters are contemplated and are discussed in detail below.

[00262] Peaks are regions enriched for open chromatin, and thus have potential for regulatory function. It is therefore understood that observing the location of peaks with respect to genes can be insightful. Various embodiments herein, e.g., bedtools closest -D=b, can be used to associate each peak with genes based on closest transcription start sites (TSS) that are packaged within the reference. In accordance with some embodiments within the disclosure, a peak is associated with a gene if the peak is within 600 bases upstream or 100 bases downstream of the TSS. Additionally, in accordance with some embodiments within the disclosure, genes can be associated to putative distal peaks that are much further from the TSS and are less than 100kb upstream or downstream from the ends of the transcript. This association can be adopted by companion visualization software of the various embodiments herein, e.g., Loupe Cell Browser. In another embodiment, this association can be used to construct and visualize derived features such as promoter-sums that can pool together counts from peaks associated with a gene.

[00263] The ATAC analysis operation can further comprise a transcription factor (TF) motif enrichment analysis. TF motif enrichment analysis includes generating a TF-barcode matrix consisting of the peak-barcode matrix (i.e., pooled cut-site counts for peaks) having a TF motif match, for each motif and each barcode. It is contemplated that the TF motif enrichment can then be utilized for subsequent analysis operations, such as differential accessibility analysis,

within various embodiments of the disclosure. Detail related to TF motif enrichment analysis is provided below.

[00264] The ATAC analysis operation can further comprise a differential accessibility analysis that performs differential analysis of TF binding motifs and peaks for identifying differential gene expression between different cells or groups of cells. Various algorithms and statistical models within the disclosure, such as a Negative Binomial (NB2) generalized linear model (GLM), can be employed for the differential accessibility analysis.

ATAC and RNA Feature Linkage Analysis

[00265] The ATAC and RNA analysis operation can comprise a feature linkage analysis for detecting correlations between pairs of genomic features detected in each of a plurality of cells, for example, between open chromatin regions and genes from single cell datasets. Such correlations can be denoted as feature linkages or linkage correlations and can be used for inferring enhancer-gene targeting relationships and constructing transcriptional networks. More details for feature linkage analysis is provided in **FIG. 43**.

[00266] In various embodiments, joint data from the joint cell calling operation can be further processed by the ATAC and RNA analysis operation to identify correlations and a significance of the correlations between the single cell gene expression library and the single cell ATAC library. The features with strong linkage correlations can be considered to be "co-expressed" and enrich for a shared regulatory mechanism. For example, the accessibility of an enhancer and the expression of its target gene can display a very synchronized differential pattern across a heterogeneous population of cells. A highly accessible enhancer leads to an elevated level of transcription factor (TF) binding, which in turn leads to elevated (or repressed) gene expression. On the other hand, when the enhancer is inaccessible, no TF can bind to the enhancer, and thus transcription activation is at minimum, which leads to reduced target gene expression.

[00267] In some embodiments, the analysis associated with gene expression or chromatin accessibility can also comprise random forest tree, random tree, naïve bayes classifier, k-means clustering, hierarchical clustering, predictions tree, classification Tree, c4.5 classifier, regression trees, neural network, affinity propagation, agglomerative clustering, birch DBSCAN clustering, mini-batch k-means, mean shift, spectral clustering, mixture of Gaussians, or xGBoost.

[00268] In an aspect, the present disclosure provides a method for identifying a genetic feature such as a cis-regulatory element corresponding to an expressed protein. The method may comprise providing DNA (e.g., chromatin) sequencing information and RNA (e.g., gene expression) sequencing information (e.g., as described herein) and associating such information with cells and/or cell nuclei from which they derive (e.g., as described herein). The cells and/or

cell nuclei may be characterized as corresponding to specific cell types (e.g., as described herein). Chromatin data and gene expression data may be used in tandem to identify cell types for the cells and/or cell nuclei. Alternatively, only one data set may be used to identify cell types for the cells and/or cell nuclei. Gene expression sequencing information may be used to identify an expressed protein of interest, such as a protein that is differentially expressed between different cell types. For example, gene expression data may be used to identify a protein such as a receptor that is over-expressed in a first cell type relative to a second cell type, where the first cell type may be associated with a disease state (e.g., a tumor cell) and the second cell type may be associated with a healthy state (e.g., a normal cell). Alternatively, gene expression data may be used to identify a protein such as a receptor that is under-expressed in a first cell type relative to a second cell type, where the first cell type may be associated with a disease state (e.g., a tumor cell) and the second cell type may be associated with a healthy state (e.g., a normal cell). The differentially expressed protein may be, for example, a cytokine, such as an interleukin (e.g., IL-2, IL-4, IL-10, IL-13, etc.) receptor. Chromatin information may be used to identify a genetic feature that may be relevant to the differentially expressed protein. The genetic feature may be, for example, a cis-regulatory element or trans-regulatory element.

[00269] A method of identifying a genetic feature (e.g., regulatory region) corresponding to an expressed protein may comprise providing a first data set corresponding to regions of accessible chromatin of a plurality of DNA molecules (e.g., chromatin) of a plurality of cells or cell nuclei and a second data set corresponding to a plurality of RNA molecules of the plurality of cells or cell nuclei. The first data set may comprise sequencing information (e.g., a first plurality of sequencing reads) corresponding to sequences of the regions of accessible chromatin and a plurality of nucleic acid barcode sequences. The second data set may comprise sequencing information (e.g., a second plurality of sequencing reads) corresponding to sequences of the plurality of RNA molecules and the plurality of nucleic acid barcode sequences. A cell or cell nucleus of the plurality of cells or cell nuclei may correspond to a nucleic acid barcode sequences of the plurality of nucleic acid barcode sequences. For example, a particle comprising a plurality of nucleic acid barcode molecules comprising a common nucleic acid barcode sequence of the plurality of nucleic acid barcode sequences may be used to process the plurality of DNA molecules and RNA molecules of a cell or cell nucleus of the plurality of cells or cell nuclei (e.g., within a partition) (as described herein). The plurality of nucleic acid barcode sequences of the first data set and the second data set may be used to identify first sequencing reads of the first plurality of sequencing reads and second sequencing reads of the second plurality of sequencing reads as corresponding to cells or cell nuclei of the plurality of cells or

cell nuclei, thereby generating a third data set comprising sequence information corresponding to regions of accessible chromatin and RNA molecules associated with individual cells or cell nuclei of the plurality of cells or cell nuclei. Cell types of the cells or cell nuclei may be identified using the sequence information of the third data set. The sequence information corresponding to the RNA molecules may be used to identify an expressed protein of a cell type of the identified cell types. The sequence information corresponding to the regions of accessible chromatin may be used to identify a genetic feature such as a cis-regulatory element corresponding to the expressed protein.

[00270] In some embodiments, the sequence information comprised by the first data set is generated from a plurality of tagged DNA fragments generated and processed as described herein. In some examples, a plurality of barcoded nucleic acid molecules are generated, including a first subset comprising sequences corresponding to regions of accessible chromatin of said plurality of deoxyribonucleic acid (DNA) molecules of a cell or cell nucleus of said sample and a second subset comprising sequences corresponding to said ribonucleic acid (RNA) molecules of the same cell or cell nucleus. In some cases, the plurality of barcoded nucleic acid molecules from the same cell or cell nucleus is generated within the same partition. The barcoded nucleic acid molecules generated in the same partition may share the same barcode sequence. In some embodiments, the method includes sequencing the barcoded nucleic acid molecules.

Feature Linkage and Linked Signatures

[00271] In some embodiments, the provided methods include: generating a first data set comprising sequencing information corresponding to regions of accessible chromatin of a plurality of deoxyribonucleic acid (DNA) molecules of a cell or cell nuclei (e.g., including ATAC data processing and/or analysis), generating a second data set comprising sequencing information corresponding to a plurality of ribonucleic acid (RNA) molecules of said cell or cell nuclei (e.g., including gene expression data processing and/or analysis), and generating a linked signature of said cell or cell nuclei using said first and second data sets (e.g., including ATAC and RNA feature linkage analysis). In some embodiments, the linked signature is generated using a feature linkage analysis process and workflow or information therefrom. In some embodiments, generating the linked feature comprises linking at least one pair of genomic features. In other cases, the linked feature comprises linking at least two, three, four, five, six, seven, eight, nine, of ten genomic features. In some embodiments, a genomic feature can comprise regions of accessible chromatin or gene expression levels. In other embodiments, a genomic feature can comprise transcriptional factor accessibility, nucleosome occupancy,

sequence identity, secondary structure of a region, tertiary structure of a region, location in the genome or physical location in the nucleus (e.g., perinuclear DNA positioning or nuclear pore complex-associated positioning). A location in the genome can be calculated based on physical distance or recombination frequency between two genetic loci. In some embodiments, at least two genomic features can be linked by a genetic locus. A genetic locus can comprise a functional unit of a gene. A functional unit of a gene can comprise sequences required to express or synthesize at least a nucleic acid or polypeptide product. In some cases, the chromatin accessibility of a promoter sequence and the level of expressed RNA driven or regulated by the promoter sequence can be linked. A nucleic acid product can an RNA. An RNA can comprise coding or non-coding RNA. An RNA can comprise an mRNA, tRNA, rRNA, snRNA, snoRNA, long non-coding RNA, miRNA, small rDNA-derived RNA (srRNA), tRNA-derived small RNA (tsRNA), antisense RNA, siRNA, or eRNA. An RNA can also comprise any precursor of the RNA herein and thereof.

[00272] A linked signature for a cell or cell nucleus may link two or more distinct data sets corresponding to genomic features or other information about the cell or cell nucleus. A linked signature may comprise a combined or correlated data set of genomic features or other information. For example, a linked signature for a cell or cell nucleus may link a first data set comprising sequencing information corresponding to regions of accessible chromatin of a plurality of DNA molecules of the cell or cell nucleus and a second data set comprising sequence information corresponding to a plurality of RNA molecules, or gene expression data thereof. A linked signature can comprise manual classification of a plurality distinct data sets corresponding to genomic features or other information about the cell or cell nucleus. A linked signature can also comprise computational classification of a plurality of distinct data sets corresponding to genomic features or other information about the cell or cell nucleus. A linked signature can comprise addition, subtraction, multiplication, division, or exponentiation of a plurality of distinct data sets corresponding to genomic features or other information about the cell or cell nucleus.

[00273] In some embodiments, feature linkages comprise correlated signals across cells or cell nuclei for peaks identified in the ATAC library and transcript levels identified in the gene expression library. In some embodiments, feature linkages can be positively or negatively correlated (**FIG. 51**). For example, an open enhancer region may have a positive correlation with gene expression of its associated transcript, while the binding of a repressor would result in a negatively correlated feature linkage. In some embodiments, the significance of a feature linkage can be determined. In some cases, the greater the correlation between open chromatin signal and

gene expression, the higher the significance of the feature linkage. In some embodiments, a signal of a cell nuclei for peaks identified in the ATAC library, in transcript levels identified in the gene expression library, and/or correlated feature linkage (e.g., linked signature) of the two can be compared between a test sample and a control sample (or between a test and control cell or cell nuclei thereof).

[00274] The linked signature for a cell or cell nucleus may correlate distinct pieces of information about the cell or cell nucleus to determine a characteristic or a condition of a sample or subject from which the cell or cell nucleus is described. The linked signature for a cell or cell nucleus may be unique to the cell or cell nucleus. The linked signature for a cell or cell nucleus may be unique to a group of cells or cell nuclei sharing a similar signature. In some cases, respective linked signatures of two or more cells may be compared or otherwise processed against each other to determine similarities or differences between the two or more cells. For example, a linked signature for a reference or control cell known to have a condition may be compared or processed against a linked signature for an analyte cell, and if one or more aspects of the two linked signatures are similar, which one or more aspects are indicative of the known condition, it may be determined that the analyte cell, or the sample or subject form which the analyte cell is derived from, has or suffers from the same known condition. Similarly, if one or more aspects of the two linked signatures diverge, which one or more aspects are indicative of the known condition, it may be determined that the analyte cell, or the sample or subject form which the analyte cell is derived from, does not have or does not suffer from the same known condition as the reference or control cell. In some embodiments, similarity or divergence of a plurality of linked signatures can be determined by manual classification or computational classification of the plurality of linked signatures. Similarity or divergence of a plurality of linked signatures can also be determined by qualitative or quantitative classification of the plurality of linked signatures. A qualitative classification can be based on the presence or absence of a genomic feature or a plurality of genomic features. A quantitative classification of the plurality of linked signatures can comprise correlating at least two linked signatures or computing the similarity of a plurality of linked signatures. The correlating can comprise carrying out a correlation test between a plurality of linked signatures. A correlation test can comprise, for example, a Pearson, Spearman's rank, Kendall's rank, biweight midcorrelation, distance correlation, percentage bend, Shepherd's Pi, Blomqvist, Hoeffding's D, gamma, Gaussian rank, point-biserial and biserial, polychoric, tetrachoric, partial, or multilevel correlation test, or other type of correlation test. In some instances, a linked signature or a plurality of linked signatures may be analyzed using one or more machine learning algorithms.

For example, a machine learning algorithm can comprise a statistical classification model or a clustering model. A classification model can comprise a k-nearest neighbor classifier, Naïve Bayes classifier, support vector machine, or neural network. A clustering model can comprise hierarchical clustering, categorical clustering, or k-means clustering. The similarity of a plurality of linked signatures can be computed based on a similarity score. A similarity score can comprise a distance measurement. A distance measurement can comprise the Euclidean distance or Manhattan distance.

[00275] In some embodiments, the methods provided herein to generate and process the linked signature of the cell or cell nuclei can determine a condition. In some embodiments, a linked signature can indicate a condition if the signature is similar to a control linked signature with the condition. In other cases, a linked signature can indicate a condition if the signature is divergent from a control linked signature without the condition. In some embodiments, a linked signature can indicate a condition if the signature is classified as the same class of a control linked signature with the condition. In some embodiments, a linked signature can indicate a condition if the signature is not classified as the same class of a control linked signature without the condition. In some embodiments, a linked signature can indicate a condition if the signature is identified as the same cluster of a control linked signature with the condition. In some embodiments, a linked signature can indicate a condition if the signature is not identified as the same cluster of a control linked signature without the condition.

[00276] A control linked signature can comprise a linked signature from a control cell or nucleus of a control cell sample. In some embodiments, the control linked signature profile can be obtained from a population of samples or subjects. A control cell or nucleus of a control cell sample can comprise a cell or cell nucleus with at least one known alteration, condition, variation, characteristic or property.

[00277] In some embodiments, a cell or cell nucleus can be derived from a bodily fluid. In some cases, a bodily fluid can comprise blood, saliva, excreta, body tissue, mucus, semen, urine, amniotic fluid, aqueous humour, bile, breast milk, cerebrospinal fluid, cerumen, chyle, exudates, gastric juice, lymph, pericardial fluid, peritoneal fluid, pleural fluid, pus, sebum, serous fluid, sputum, synovial fluid, tears, vomit, or tissue fluid. In other cases, a cell or cell nucleus can be derived from a biopsy. A biopsy can comprise a sample of tissue or cells for determining the presence or extent of a condition. A biopsy can comprise cells or tissues from lymph nodes. Lymph nodes can comprise lymph node clusters from the head, neck, thorax, abdomen, arm, or lower limbs.

[00278] A condition can comprise a disease state or a risk to a disease state. A disease state can mean a subject suffering from a disease or a symptom. In some embodiments, a subject may or may not be diagnosed with the disease state. In some embodiments, a subject may not suffer from the disease but can be at risk of developing the disease state. In some embodiments, the disease state or the risk to the disease state can comprise diseases related to cell growth, division, differentiation, migration, replication, or proliferation. In some embodiments, a condition can be a tumor, cancer, malignancy, neoplasm, or other proliferative disease or disorder. In some embodiments, the disease state can comprise non-Hodgkin lymphomas. In other cases, the disease state can comprise a B cell malignancy or a T cell malignancy. In some embodiments, the disease state can comprise, a B cell lymphoma. A B cell lymphoma can comprise Diffuse large B-cell lymphoma (DLBCL), Follicular lymphoma, Chronic lymphocytic leukemia (CLL), small lymphocytic lymphoma (SLL), Mantle cell lymphoma (MCL), Marginal zone lymphoma, Burkitt lymphoma, Lymphoplasmacytic lymphoma (Waldenstrom macroglobulinemia), Hairy cell leukemia, Nodal marginal zone B cell lymphoma, Splenic marginal zone lymphoma, Primary effusion lymphoma, Lymphomatoid granulomatosis, Primary central nervous system lymphoma, ALK+ large B-cell lymphoma, Plasmablastic lymphoma, or Primary intraocular lymphoma.

[00279] The methods provided herein to determine a condition can comprise diagnostic evaluation, prognostic evaluation, monitoring, and/or management of the condition. The method can comprise measuring the gene expression and/or chromatin accessibility, such as to determine a state of the subject prior to, during, or subsequent to treatment. In some embodiments, a level of one or more regions of accessible chromatin or said one and/or more genes expressed determined from a linked signature indicative of a condition can be analyzed in a sample obtained from an individual suspected of having said condition. In some cases, to analyze a sample obtained from an individual suspected of having a condition, both the one or more regions of accessible chromatin and the one and/or more genes expressed determined from a linked signature indicative of a condition are analyzed for individual cells. In some instances, to analyze a sample obtained from an individual suspected of having a condition, the analysis includes detecting cells exhibiting the linked signature indicative of the condition. In some embodiments, gene expression data can be used to identify a functional gene set for upregulated genes associated with a condition and/or a particular cell type, e.g., using a tool as described by Chen et al., "Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool" BMC Bioinformatics. 2013;128(14); Kuleshov et al., "Enrichr: a comprehensive gene set enrichment analysis web server 2016 update" Nucleic Acids Research. 2016; gkw377.

[00280] In some embodiments, subjects who can undergo diagnostic evaluation, prognostic evaluation, monitoring, or management of the condition are subjects who have increased growth or metastasis of B cell lymphoma tumor or metastasis, or subjects suspected of having B cell lymphoma. In some of any such embodiments, the diagnostic evaluation, prognostic evaluation, monitoring, or management of the condition is performed on an *in vitro* biological sample. In some embodiments, the subject is identified as having B cell lymphoma with a surgically excised tissue biopsy. An excision biopsy of a lymph node or other tissues, depending on the types of B cell lymphomas, can be tested with immunocytochemistry, flow cytometry, fluorescent in situ hybridization, or DNA/RNA, including, but not limited to, QPCR, digital PCR, DNA sequencing, or RNA sequencing. For example, DLBCL cells ae CD45 positive and express CD19, CD20, CD22, CD79a, and only express one of κ or λ immunoglobin light chains. A subject being evaluated, monitored, or managed with B cell lymphoma can also comprise receiving a CT scan. The CT scan can be used to measure tumor masses and frequencies of disease-related symptoms. B cell lymphoma-related symptoms can comprise night sweats, fever, weight loss, fatigue, appetite loss, redouble breathing, abdominal pain, swollen abdomen, chest pain, coughing, swollen lymph nodes, severe itching, or any combination thereof. In other embodiments, the diagnostic evaluation, prognostic evaluation, monitoring, or management of B cell lymphoma can be based on the B cell lymphoma tumor growth, B cell lymphoma lesion number, B cell lymphoma cell number, or any combination thereof. The subject may suffer from, or be at risk of, any condition.

[00281] Provided herein are methods for analyzing a biological sample comprising determining one or more linked signatures (e.g., feature linkages) of a cell or cell nucleus in the biological sample, wherein the linked signature comprises a correlation of (i) sequencing information corresponding to regions of accessible chromatin of a plurality of deoxyribonucleic acid (DNA) molecules of the cell or cell nucleus, and (ii) sequencing information corresponding to a plurality of ribonucleic acid (RNA) molecules (or derivatives thereof) of the cell or cell nucleus, wherein the linked signature of at least one region of accessible chromatin and expression of at least one RNA at a significant level within the biological sample is indicative of a condition of the cell or cell nucleus.

[00282] Provided herein are *in vitro* methods for preparing a biological sample comprising: processing open chromatin structures of the T-cells and/or B-cells from the biological sample with a transposase to provide a plurality of DNA molecules; generating a first plurality of barcoded nucleic acid molecules comprising the plurality of DNA molecules; generating a second plurality of barcoded nucleic acid molecules comprising a plurality of

nucleic acids comprising mRNA sequences or derivatives thereof from the T-cells and/or B-cells from the biological sample; and generating a first and second sequencing library from the first and second plurality of barcoded nucleic acid molecules, respectively, to determine a linked signature of a cell of the T-cells and/or B-cells. In some embodiments, the feature linkages (e.g., linked signatures) can be used to determine correlated gene expression and open chromatin regions across the genome. In some cases, the feature linkages (e.g., linked signatures) can be used to determine gene regulatory networks. In some cases, a significance level for the linked signature can be determined. In some embodiments, a biological sample isolated and obtained from an individual is provided and processed *ex vivo*. In some aspects, the method further comprises obtaining the biological sample from the individual. In some aspects, the method may include, but not require, the further step of obtaining the biological sample from the individual. In some embodiments, the preparing and processing of the biological sample is performed *ex vivo*.

[00283] In some of any such embodiments, the generation of the first and/or second plurality of barcoded nucleic acid molecules can be performed within a plurality of partitions. In some aspects, the method includes reverse transcribing the plurality of mRNA sequences from the T-cells and/or B-cells from the biological sample to provide a plurality of complementary DNA (cDNA) molecules, and the second plurality of barcoded nucleic acid molecules comprises the cDNA molecules. In some aspects, the method includes barcoding 3' ends of the mRNA. In some embodiments, the method comprises encapsulating single nuclei of the T-cells and/or B-cells in droplets. In some cases, the droplets can be formed prior to generating the first and/or second plurality of barcoded nucleic acid molecules. In some embodiments, generating of the first and second plurality of barcoded nucleic acid molecules is performed at the same time.

[00284] In some embodiments, the method further includes determining from the first and second sequencing libraries a presence, absence, and/or level of the one or more linked signatures correlated with a condition. In some cases, the condition is a tumor, cancer, malignancy, neoplasm, or other proliferative disease or disorder. In some instances, the condition is a B cell malignancy (e.g., B cell lymphoma). In some embodiments, one or more biomarkers (e.g., genes, transcription factors) can be identified or selected using the feature linkages (e.g., linked signatures). In some embodiments, one or more biomarkers for diagnostic evaluation, prognostic evaluation, monitoring, or management of a condition (e.g., B cell lymphoma) in subjects can be selected from **FIG. 50**. For example, the one or more biomarkers (e.g., genes) identified can be differentially expressed between B cells and tumor B cells. In some of any such embodiments, a linked signature of a cell or cell nucleus from a subject (e.g.,

T-cells and/or B-cells) can be compared to a control linked signature of a control cell or cell nucleus of a control sample.

[00285] An agent of a given dosage may be administered to a subject suffering from, or at risk of, a condition. In the case where the subject's status does improve, as determined from monitoring of the subject's status, upon the administrator (e.g., doctor)'s discretion, the administration of the agent may be given continuously. Alternatively, the dose of the agent may be temporarily reduced or temporarily suspended for a certain length of time (e.g., a "drug holiday"). The length of the drug holiday can be on the order of hours, days, months, and years. The dose reduction during such drug holiday may be by any amount. Once improvement or maintenance of the subject's condition has occurred, the subject may be administered a maintenance dosage, for example, at the discretion of the administrator. Subsequently, the dosage or frequency of administration, or both, can be reduced. Subjects may receive intermittent treatment upon any recurrence of the condition.

[00286] Subjects undergoing diagnostic evaluation, prognostic evaluation, monitoring, or management of a disease can be administered a therapeutically effective amount of an agent.

[00287] The amount of a given agent that is administered to a subject may correspond to and vary depending upon factors such as the particular agent, the severity of the disease, the identity (e.g., weight) of the subject or host in need of treatment, but can nevertheless be routinely determined in a manner known in the art according to the particular circumstances surrounding the case, including, e.g., the specific agent being administered, the route of administration, and the subject or host being treated. The terms "therapeutically effective amount" and "effective amount" of a compound generally refer to an amount sufficient to provide a therapeutic benefit in the treatment, prevention and/or management of a disease, to delay or minimize one or more symptoms associated with the disease or disorder to be treated. The terms "therapeutically effective amount" and "effective amount" can encompass an amount that improves overall therapy, reduces or avoids symptoms or causes of disease or disorder, or enhances the therapeutic efficacy of another therapeutic agent. The desired dose(s) may conveniently be presented in a single dose or as divided doses administered simultaneously (or over a short period of time) or at appropriate intervals, for example as two, three, four or more sub-doses per day.

[00288] In some cases, a therapeutically effective amount of an agent can be determined as an amount or dose of agent that can alter a linked signature of a condition or a risk, as determined from a sample(s) of the subject, to a linked signature of not having the disease or the risk of the disease, such as determined from a control sample, when administered to the subject.

In some embodiments, the therapeutically effective amount of an agent can comprise the amount of the agent that changes a similarity of the linked signature of the subject and that of a reference linked signature (e.g., known to have the condition, or known not to have the condition) by at least about 5%, 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 99%, or more. In some cases, the therapeutically effective amount of an agent can decrease a tumor size or the frequency of disease-related symptoms after administering the dose of the agent by at least 5%, 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 99%, or more. In some embodiments, the therapeutically effective amount can decrease at least 5%, 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 99%, or more in B cell lymphoma tumor growth, B cell lymphoma lesion number, B cell lymphoma cell number, or any combination thereof after administering the dose of the agent.

[00289] The methods provided herein and thereof can be used to identify one or more regions of accessible chromatin and/or one or more genes expressed as a target or targets for a therapeutically effective agent. In some embodiments, the method can identify a target region or regions of accessible chromatin having the same accessibility in the cells with the same condition. In some embodiments, the method can identify a target gene or genes expressed by the cells with the same condition. In some embodiments, the method can identify a target gene or genes not expressed by the cells with the same condition. In some cases, the method can identify a target gene or genes expressed to about the same extent by the cells with the same condition. A therapeutically effective agent can comprise a small molecule, a nucleic acid, a polypeptide, a radiation, or a probiotic. In some embodiments, the chromatin accessibility of a target region can be modified by a therapeutically effective agent. In other cases, the chromatin accessibility of a target region can be modified by a therapeutically effective agent to diverge from that of a cell with a known condition. In other embodiments, the chromatin accessibility of a target region can be modified by a therapeutically effective agent to mimic that of a cell without a condition. In some embodiments, the expression of a target gene can be modified by a therapeutically effective agent. In other cases, the expression of a target gene can be modified by a therapeutically effective agent to diverge from that of a cell with a condition. In other embodiments, the expression of a target gene can be modified by a therapeutically effective agent to mimic that of a cell without a condition.

[00290] A therapeutically effective agent can be formulated to be compatible with its intended route of administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or

other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

[00291] Administration or application of a therapeutically effective agent disclosed herein can be performed for a treatment duration of at least about at least about 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, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100 days consecutive or nonconsecutive days. In some cases, a treatment duration can be from about 1 to about 30 days, from about 2 to about 30 days, from about 3 to about 30 days, from about 4 to about 30 days, from about 5 to about 30 days, from about 6 to about 30 days, from about 7 to about 30 days, from about 8 to about 30 days, from about 9 to about 30 days, from about 10 to about 30 days, from about 11 to about 30 days, from about 12 to about 30 days, from about 13 to about 30 days, from about 14 to about 30 days, from about 15 to about 30 days, from about 16 to about 30 days, from about 17 to about 30 days, from about 18 to about 30 days, from about 19 to about 30 days, from about 20 to about 30 days, from about 21 to about 30 days, from about 22 to about 30 days, from about 23 to about 30 days, from about 24 to about 30 days, from about 25 to about 30 days, from about 26 to about 30 days, from about 27 to about 30 days, from about 28 to about 30 days, or from about 29 to about 30 days. Administration or application of a composition disclosed herein can be performed for a treatment duration of at least about 1 week, at least about 1 month, at least about 1 year, at least about 2 years, at least about 3 years, at least about 4 years, at least about 5 years, at least about 6 years, at least about 7 years, at least about 8 years, at least about 9 years, at least about 10 years, at least about 15 years, at least about 20 years, or more. Administration can be performed repeatedly over a lifetime of a subject, such as once a month or once a year for the lifetime of a subject. Administration can be performed repeatedly over a substantial portion of a subject's life, such as once a month or once a year for at least about 1 year, 5 years, 10 years, 15 years, 20 years, 25 years, 30 years, or more.

[00292] Administration or application of a therapeutically effective agent disclosed herein can be performed at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, or 24 times a day. In some cases, administration or application of composition disclosed

herein can be performed at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or 21 times a week. In some cases, administration or application of composition disclosed herein can be performed at least 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, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, or 90 times a month. In some cases, a therapeutically effective agent can be administered/applied as a single dose or as divided doses. In some cases, the agents described herein can be administered at a first time point and a second time point. In some cases, an agent can be administered such that a first administration is administered before the other with a difference in administration time of 1 hour, 2 hours, 4 hours, 8 hours, 12 hours, 16 hours, 20 hours, 1 day, 2 days, 4 days, 7 days, 2 weeks, 4 weeks, 2 months, 3 months, 4 months, 5 months, 6 months, 7 months, 8 months, 9 months, 10 months, 11 months, 1 year or more.

[00293] In some embodiments, agents disclosed herein can be in unit dose forms or multiple-dose forms. For example, a therapeutically effective agent described herein can be in unit dose form. Unit dose forms, as used herein, refer to physically discrete units suitable for administration to human or non-human subjects (e.g., pets, livestock, non-human primates, and the like) and packaged individually. Each unit dose can contain a predetermined quantity of an active ingredient(s) that can be sufficient to produce the desired therapeutic effect in association with pharmaceutical carriers, diluents, excipients, or any combination thereof. Examples of unit dose forms can include ampules, syringes, and individually packaged tablets and capsules. In some instances, a unit dose form can be comprised in a food. In some instances, unit-dosage forms can be administered in fractions or multiples thereof. A multiple-dose form can be a plurality of identical unit dose forms packaged in a single container, which can be administered in segregated a unit dose form. Examples of a multiple-dose form can include vials, bottles of tablets or capsules, bottles of gummies, or bottles of pints or gallons. In some instances, a multiple-dose form can comprise different pharmaceutically active agents. In some embodiments, a unit dose form can be a serving. In some cases, a multiple-dose form can have more than about: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100 or 200 servings. In some embodiments, a multiple-dose form can have less than about: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100 or 200 servings. In some instances, a multiple-dose form can have from about: 1 serving to about 200 servings, 1 serving to about 20 servings, 5 servings to about 50 servings, 10 servings to about 100 servings, or about 30 servings to about 150 servings.

[00294] The methods provided herein and thereof to determine a condition can be used to monitor a condition prior to, during, or subsequent to the administration of a therapeutically effective agent. In some embodiments, the therapeutically effective dose, dosing regimens, administration route of an agent can be modified based on the monitoring of a condition. Such modification can decrease or increase therapeutically effective dose, dosing regimens, administration route of an agent. In other cases, the monitoring of a condition can comprise switching, adding, or removing a therapeutically effective agent.

[00295] In accordance with various embodiments, a general schematic workflow **4300** is provided in **FIG. 43** to illustrate a non-limiting example process of a feature linkage analysis workflow for feature linkage analysis. The workflow **4300** can include various combinations of features, whether it be more or less features than that illustrated in **FIG. 43**. As such, **FIG. 43** simply illustrates one example of a possible workflow for conducting feature linkage analysis.

[00296] **FIG. 43** provides a schematic workflow **4300** for conducting feature linkage analysis. It should be appreciated that the methodologies described in the workflow **4300** of **FIG. 43** and accompanying descriptions can be implemented independently of the methodologies for generating single cell gene expression sequencing data or single cell ATAC sequencing data described in general. Therefore, **FIG. 43** can be implemented independently of a sequencing data generating workflow as long as it is capable of sufficiently analyzing single cell sequencing data sets for feature linkage analysis.

[00297] Moreover, the data analysis workflow can include one or more of the analysis operations illustrated in **FIG. 43**. Not all the operations within the disclosure of **FIG. 43** need to be utilized as a group. Therefore, some of the operations within **FIG. 43** are capable of independently performing the necessary data analysis as part of the various embodiments disclosed herein. Accordingly, it is understood that, certain operations within the disclosure can be used either independently or in combination with other operations within the disclosure, while certain other operations within the disclosure can only be used in combination with certain other operations within the disclosure. Further, one or more of the operations described below, presumably defaulted to be utilized as part of the computational pipeline, can also not be utilized per user input. It is understood that the reverse is also contemplated. It is further understood that additional operations for analyzing the generated sequencing data are also contemplated as part of the computational pipeline within the disclosure.

Joint Feature-Barcode Matrix

[00298] In operation **4310**, a joint feature-barcode matrix may be generated and received. The joint feature-barcode matrix can be generated by gene expression data processing and

ATAC data processing. For example, the joint cell barcode matrix may comprise the counts of fragment ends (cut sites) within each peak region for each barcode and the counts of UMIs for each barcode.

Matrix Normalization

[00299] In operation **4320**, the joint feature-barcode matrix may be normalized to generate a normalized matrix. The normalization may reduce the bias introduced by the variance of totals signals per single cell. The total signals per cell, alternatively referred as depth, can be the sum of unique molecular identifiers (UMIs) for gene expression or the sum of total cut sites in ATAC.

[00300] Previous methods for normalization will create a strong artifact for feature linkage analysis, so a depth-adaptive negative binomial distribution model may be used to overcome this defect. Normalization may comprise selecting genomic features detected in each of a plurality of cells within a pre-set size of genomic window, for example, 100 kb, 200 kb, 300 kb, 400 kb, 500 kb, 600 kb, 700 kb, 800 kb, 900 kb, 1 Mb, 1.5 Mb, 2 Mb, or any intermediate ranges or values therefrom.

[00301] Normalization may further comprise using a depth adaptive negative binomial distribution model to model molecular counts of the joint feature-barcode matrix, in which the mean of the distribution for every genomic feature is assumed to vary linearly with the library size for each cell. The negative binomial distribution is a probability distribution that is used with discrete random variables. This type of distribution concerns the number of trials that must occur in order to have a predetermined number of successes. In various embodiments, the depth adaptive negative binomial distribution model may be applied to at least two data types, including, but not limited to, both gene expression data and ATAC data. For example, normalized matrix count $\hat{x}_{ij}$ is a standardized value of raw count x_{ij} based on a non-limiting exemplary formula as shown below:

$$l_j = \frac{x_{ij}}{\sum_i \sum_j x_{ij}}$$
$$\hat{\mu}_{ij} = \frac{x_{ij}}{l_j}$$

$$\hat{r}_i = MLE(x_{i\cdot})$$

$$E[x_{ij}] = \hat{\mu}_{ij}$$
$$Var[x_{ij}] = \hat{\mu}_{ij} + \frac{\hat{\mu}_{ij}^2}{\hat{r}_i}$$
$$\widehat{x_{ij}} = \frac{x_{ij} - E[x_{ij}]}{\sqrt{Var[x_{ij}]}}$$

Where x_{ij} is entry of the feature-barcode matrix for feature i and cell j and $\hat{x}_{ij}$ is the normalized value for feature i and cell j. "μ hat" and "r hat" represent the negative binomial mean and dispersion.

Matrix Smoothing

[00302] In operation **4330**, the joint feature-barcode matrix may be smoothed by K-nearest neighbors (KNN) distance and Gaussian kernel to generate a cell-cell similarity matrix.

[00303] Due to the sparsity of single-cell data, especially the cut sites count in peaks, it is most likely that one will not detect signals of a peak and a gene at the same time in one cell when both are expected to have high expression levels. As a result, direct computation of correlation or other measures of dependence on the raw counts between two genomic features detected in each of a plurality of cells may fail to generate any meaningful value that distinguishes highly co-expressed features from the rest.

[00304] To overcome this barrier, smoothing may be performed so the value of a feature in a given cell is enhanced by "borrowing" values of the same feature from "neighboring" cells. Here, neighboring cells describe a population of cells whose gene expression profile or ATAC profile share a high similarity, i.e., a low distance. For example, the distance is an Euclidean distance. The Euclidean distance or Euclidean metric is the "ordinary" straight-line distance between two points in Euclidean space.

[00305] The high similarity may be determined by applying a K-nearest neighbor algorithm called "Ball-Tree" on the principal component analysis (PCA)-reduced dimension. For example, the ball tree nearest-neighbor algorithm examines nodes in depth-first order, starting at the root. During the search, the algorithm maintains a max-first priority queue (often implemented with a

heap), denoted Q here, of the K nearest points encountered so far. Principal component analysis (PCA) refers to a main linear technique for dimensionality reduction and performs a linear mapping of the data to a lower-dimensional space in such a way that the variance of the data in the low-dimensional representation is maximized.

[00306] Smoothing comprises "borrowing" information from neighboring cells. In various embodiments, the information "borrowing" can be achieved by a weighted summation of signals of all a predetermined number of neighboring cells using K-nearest neighbors distance (for example, K=30). K may be selected to be 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 or any intermediate ranges or values, depending on how many cells are in a given dataset. For example, if more than 10,000 cells are available, a larger K value (K=50) might be selected.

[00307] In various embodiments, the cell-to-cell similarity matrix may determine the smoothing weights. The smoothing weights may be determined as the Euclidean distance based on the gene expression principal components, such that weight Wij is only positive if cells i and j are neighbors and there are no self-edges.

[00308] Additionally and alternatively, to avoid over-smoothing, raw distances can be normalized using a Gaussian kernel:

$$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma^2}\right)$$

[00309] In certain embodiments, based on the use of a Gaussian kernel, smoothing weights are high only when two cells have a highly similar gene expression profile and quickly decays to zero when the similarity between cells decreases. The 'kernel' for smoothing, defines the shape of the function that is used to take the average of the neighboring points. A Gaussian kernel is a kernel with the shape of a Gaussian (normal distribution) curve.

[00310] After smoothing, putative co-expressed features showed a much strong correlation pattern compared to randomly selected pairs of features.

Smoothed Matrix

[00311] In operation **4340**, a smoothed matrix may be generated by the normalized matrix from operation **4320** and the cell-cell similarity matrix from operation **4330**. For example, the smoothed matrix may be generated by multiplying the normalized matrix with the cell-cell similarity matrix.

Feature Linkage Correlations

[00312] In operation **4350**, feature linkage correlations may be generated. Linkage correlation is the direct measure of the strength of the linkage, with the value bound by [-1, 1]. The sign of the correlation indicates a positive or negative association. It provides a very interpretable measure of the linkage strength.

[00313] For example, feature linkage correlations may be generated by computing a Pearson correlation coefficient between two genomic features detected in each of a plurality of cells as the linkage correlation after smoothing.

[00314] The Pearson's correlation coefficient r_{xy} for vector X and Y of the same length, referred as Pearson correlation, may be computed as:

$$r_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}}.$$

Where {(x1, y1), (x2, y2), …, (xn, yn)} are the paired data of X and Y, i is a cell number (1, 2, 3, ..., N), and n is the sample size.

[00315] The workflow **4300** can comprise, at operation **4370**, generating feature linkage significances. In various embodiments, feature linkage significances may be generated as a probability score.

[00316] Significance of feature linkage provides measures of statistical uncertainty for feature linkage inference and offers more contrast of strong linkages relative to weak linkages. Significance may be generated by determining a local correlation value for a linkage between at least two genomic features detected in each of a plurality of cells and transform the value to a Gaussian random variable. This method allows for hypothesis testing.

[00317] For example, linkage significance is computed using a modified algorithm based on improvements and extensions of local correlation from Hotspot (DeTomaso et al., DeTomaso, D., & Yosef, N. (2020). *Identifying Informative Gene Modules Across Modalities of Single Cell Genomics*. BioRxiv, 2020.02.06.937805).

$$H_{xy} = w_{ij}(x_i y_j + y_i x_j)$$

$$E(H_{xy}) = 0$$

$$E(H_{xy}^2) = (\sum_{i}^{N} \sum_{j \in N(i)} w_{ij} x_j)^2$$

$$\widehat{Z_{xy}} = \frac{H_{ij} - E(H)}{\sqrt{E(H^2) - (E(H))^2}} = \frac{w_{ij}(x_i y_j + y_i x_j)}{(\sum_{i}^{N} \sum_{j \in N(i)} w_{ij} x_j)^2}$$

[00318] Specifically, computation of Hxy and E(Hxy2) was significantly accelerated by converting a loop-based procedure in DeTomaso et al. into a matrix multiplication-based procedure. In this matrix multiplication, local correlation of a N pairs of features (for example, 10,000 pairs of features), denoted as the Z score "Zxy hat", can be generated in one operation of operation instead of a loop of N operations (for example, 10,000 operations).

[00319] Additionally and alternatively, the local correlation Z score may be extended to a hypothesis-testing framework to generate a probability score. Because the Z score follows a Gaussian distribution of mean 0 and variance 1 based on the normalization operation as described above, it can be converted to a probability score and subject to multiple testing correction.

[00320] The resulting value is a false discovery rate for whether a given pair of features x and y are significantly correlated.

Sparsity Generation

[00321] The workflow **4300** can comprise, at operation **4370**, sparsity generation. A sparse statistical model is one in which only a relatively small number of parameters (or predictors) play an important role. Because the number of computable linkages is quadratic of the number of features, and it is expected that majority of computable linkages are not biologically significant, it is natural to expect sparsity in the inference of feature linkage.

[00322] Since most feature linkages are not significant, a subset of linkages that have a significance lower than a pre-set threshold may be filtered out and the sparsified linkage matrix may be used for better interpretation. Thresholding may be selected based on feature significance. For a particular example, a thresholding method may be used, in which linkages with significance < 5 are removed from the linkage matrix. The threshold can be determined based on an analysis of continuously down-sampling reads and comparing of the decay of linkage significance and correlation. For example, significance = 5 may have the best balance of

linkage strength and stability against down-sampling. In various embodiments, thresholding may use a feature significance threshold, such as significance more than or equal to 4, 4.5, 5, 5.5, 6 or any intermediate ranges or values derived therefrom for selecting for feature linkages. In additional and alternative embodiments, thresholding may be set using a value of correlation, for example, feature linkages with a correlation value more than 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5 or any intermediate values or ranges may be selected and set as a threshold for selecting for feature linkages.

[00323] Several sparsity-generating strategies may be used. For example, the sparsity-generation may use thresholding, which is to exclude linkages with a pre-set threshold on correlation or significance. Thresholding may be a particular example of sparsity-generating strategy based on its simplicity, interpretability, and good consistency against differential expression.

[00324] In additional and alternative embodiments, the sparsity-generation may use Gaussian graphical models (GGM). A GGM is an undirected graph in which each edge represents the pairwise correlation between two variables conditioned against the correlations with all other variables (also denoted as partial correlation coefficients). GGMs have a simple interpretation in terms of linear regression techniques. When regressing two random variables X and Y on the remaining variables in the data set, the partial correlation coefficient between X and Y can be determined by the Pearson correlation of the residuals from both regressions. Intuitively speaking, we remove the (linear) effects of all other variables on X and Y and compare the remaining signals. If the variables are still correlated, the correlation is directly determined by the association of X and Y and not mediated by the other variables.

[00325] Several flavors of GGM-based methods have been tested and can be used, including, but not being limited to, graphical lasso, relaxed graphical lasso, sparse estimation of covariance, and sparse Steinian covariance estimation. The benefit of GGM is that it has a strong statistics framework and allows linkage-specific regularization. However, GGM based on optimizing the precision matrix creates false negative, in which strong linkages can be erroneously determined to be zero. GGM that optimizing the covariance matrix may need to be used to improve GGM-based sparsity generation.

Feature Linkage Matrix

[00326] The workflow **4300** can comprise, at operation **4380**, generating a feature linkage matrix after sparsity generation for downstream analysis.

FEATURE LINKAGE ANALYSIS METHODS

[00327] In various embodiments, methods are provided for feature linkage analysis. The methods can be implemented via computer software or hardware. The methods can also be implemented on a computing device/system that can include a combination of engines for feature linkage analysis. In various embodiments, the computing device/system can be communicatively connected to one or more of a data source, sample analyzer (e.g., a genomic sequence analyzer), and display device via a direct connection or through an internet connection.

[00328] Referring now to **FIG. 44**, a flowchart illustrating a non-limiting example method **4400** for feature linkage analysis is disclosed, in accordance with various embodiments. The method can comprise, at operation **4402**, receiving a data matrix comprising at least two genomic features detected in each of a plurality of cells. For example, at least two genomic features can be gene expression features (such as genes and mRNA) and assay for transposase-accessible chromatin (ATAC) features (such as open chromatin regions or accessible chromatin regions). For example, the data matrix may be a joint feature-barcode matrix that comprises data of both cut sites and UMIs for each barcode. In additional and alternative embodiments, the data matrix may be generated from single-cell sequencing as discussed above, sci-CAR or SNARE-seq, or a combination thereof.

[00329] The method can comprise, at operation **4404**, smoothing the data matrix to generate a smoothed matrix, wherein smoothing the data matrix comprises normalizing the first genomic feature and the second genomic feature identified for each cell in the data matrix with the first and second genomic features from a subset of neighboring cells. Normalizing the data matrix may comprise using a depth-adaptive negative binomial distribution model to model molecular counts of the data matrix, such as joint feature-barcode matrix.

[00330] The method can comprise, at operation **4406**, generating linkage correlations between the first genomic feature and second genomic feature identified for each of the plurality of cells in the data matrix. For example, feature linkage correlations may be generated by computing a Pearson correlation coefficient between two genomic features as the linkage correlation after smoothing.

[00331] The method can comprise, at operation **4408**, generating linkage significances of the linkage correlations of pairs of the first and second genomic features identified for each of the plurality of cells in the data matrix. In various embodiments, feature linkage significances may be generated as a probability score. For example, the feature linkage significances can be generated by using multiplication of a plurality of linkage matrixes. Each linkage matrix can

comprise linkage correlations of pairs of the first and second genomic features identified for each of the plurality of cells in the data matrix.

[00332] In additional and alternative embodiments, the feature linkage significances may be generated using a matrix multiplication. In this matrix multiplication, local correlation of a N pairs of features (for example, 10,000 pairs of features), denoted as the Z score "Zxy hat", can be generated in one operation of operation instead of a loop of N operations (for example, 10,000 operations).

[00333] The method can comprise, at operation **4410**, outputting the linkage correlations and linkage significances.

[00334] In some aspects, the feature linkages, correlations, and/or significances can be used as a linked signature for a biological particle (e.g., a cell and/or cell nucleus). In some cases, feature linkages, correlations, and/or significances can be used to generate a plurality of linked signatures for each of a plurality of biological particles (e.g., cells and/or cell nuclei).

[00335] **FIG. 40** illustrates an example workflow for a method of identifying a genetic feature corresponding to an expressed protein. In process **4010**, chromatin sequencing information and gene expression sequencing information is associated with cells and/or cell nuclei. In process **4020**, the cells and/or cell nuclei are characterizing as corresponding to specific cell types. In process **4030**, gene expression sequencing information is used to identify an expressed protein. In process **4040**, chromatin sequencing information is used to identify a genetic feature such as a cis-regulatory element (e.g., a promoter or enhancer) corresponding to the expressed protein.

[00336] A plurality of cells or cell nuclei subjected to analysis provided herein may comprise at least 500 cells. For example, the plurality of cells or cell nuclei may comprise at least 500, 1,000, 2,000, 5,000, 7,500, 10,000, or more cells or cell nuclei.

[00337] A genetic feature that may be relevant to the differentially expressed protein may be upstream of a nearby gene. Alternatively, the genetic feature may be downstream of a nearby gene. In some cases, the genetic feature may be disposed substantially far from the gene that it may influence.

[00338] A cis-regulatory element (CRE) may be a region of non-coding DNA that regulates transcription of a nearby gene. A cis-regulatory element may be upstream of a transcription site. Alternatively, a cis-regulatory element such as an enhancer may be downstream of a gene it regulates. A cis-regulatory element may be, for example, a promoter, enhancer, operator, or silencer. A promoter may occur at a site where transcription is initiated. A promoter may include one or more of a TATA box, a transcription factor II B (TFIIB) recognition site, an initiator, and a core promoter element. An enhancer may influence (e.g., enhance) the

transcription of a gene. A cis-regulatory element that modulates a gene may be one of a plurality of cis-regulatory elements (e.g., promoters and enhancers) that influences transcription of the gene. A cis-regulatory element may alternatively be a silencer or operator that may repress transcription of a gene.

[00339] A genetic feature such as a cis-regulatory element may be correlated with epigenetic markers, such as epigenetic markers included in databases such as the curated GeneHancer database. An epigenetic marker may comprise a differentially methylated region. An example of an epigenetic modification is H3K4me1, which includes mono-methylation at the 4^{th} lysine residue of the histone H3 protein. H3K4me1 is an enhancer signature for various genes.

[00340] The methods provided herein may facilitate prediction of relative protein expression across various cell types. In some instances, promoter peaks measurable in accessible chromatin data across a group of cell types corresponding to a given protein may not predictive of protein expression. However, correlation of the promoter peaks with an additional genetic feature (e.g., a cis-regulatory element, as described above) may enable prediction of relative protein expression across the group of cell types. This information may be useful in developing diagnostic methods and therapies for disease types including various cancers.

[00341] In addition to identifying genetic features, chromatin information may also be used to analyze activation and/or de-activation of other proteins including signal transducer (e.g., signal transducer and activator of transcription (STAT)) proteins. For example, activation of latent proteins such as STAT proteins may be analyzed. Such proteins are latent cytoplasmic proteins, making expression a poor proxy for function. Chromatin information may be used to identify accessibility of a given binding motif for such a protein and thus activation of the protein for a given cell type (e.g., for a tumor cell type).

[00342] The methods described herein may be applied to analysis of cell populations including diseased cells such as tumor cells. The methods provided herein may comprise using sequencing information to identify a presence of a tumor cell or cell nucleus in a sample. Thus, the methods provided herein may be useful in diagnostic evaluations. The methods provided herein may also or alternatively comprise using sequencing information to identify a cell type, a cell state, a tumor-specific gene expression pattern, or a tumor-specific differentially accessible region of chromatin in a sample. Based at least in part on such analysis, a therapeutic regimen may be identified, e.g., for a subject from which a sample derives. A therapeutic regimen may comprise administration of a therapeutically effective amount of an agent targeting one or more targets identified in a tumor-specific gene expression pattern or a tumor-specific differentially accessible region of chromatin.

[00343] The methods provided herein may also be applied to identify single nucleotide polymorphisms (SNPs) associated to various diseases. Other methods for associating SNPs with specific diseases may rely on analysis of regions of chromatin that are readily linked to expressed proteins. The present disclosure improves upon such methods by providing a mechanism for identifying genetic features (e.g., regulatory regions) that may be correlated to gene expression by specific cell types, including diseased cells. Thus, in another aspect, the present disclosure provides a method for identifying a single nucleotide polymorphism or other genetic feature associated with a disease. The method may comprise providing a first data set corresponding to regions of accessible chromatin of a plurality of DNA molecules (e.g., chromatin) of a plurality of cells or cell nuclei and a second data set corresponding to a plurality of RNA molecules of the plurality of cells or cell nuclei. The first data set may comprise sequencing information (e.g., a first plurality of sequencing reads) corresponding to sequences of the regions of accessible chromatin and a plurality of nucleic acid barcode sequences. The second data set may comprise sequencing information (e.g., a second plurality of sequencing reads) corresponding to sequences of the plurality of RNA molecules and the plurality of nucleic acid barcode sequences. A cell or cell nucleus of the plurality of cells or cell nuclei may correspond to a nucleic acid barcode sequences of the plurality of nucleic acid barcode sequences. For example, a particle comprising a plurality of nucleic acid barcode molecules comprising a common nucleic acid barcode sequence of the plurality of nucleic acid barcode sequences may be used to process the plurality of DNA molecules and RNA molecules of a cell or cell nucleus of the plurality of cells or cell nuclei (e.g., within a partition) (as described herein). The plurality of nucleic acid barcode sequences of the first data set and the second data set may be used to identify first sequencing reads of the first plurality of sequencing reads and second sequencing reads of the second plurality of sequencing reads as corresponding to cells or cell nuclei of the plurality of cells or cell nuclei, thereby generating a third data set comprising sequence information corresponding to regions of accessible chromatin and RNA molecules associated with cells or cell nuclei of the plurality of cells or cell nuclei. Cell types of the cells or cell nuclei may be identified using the sequence information of the third data set. The sequence information corresponding to the RNA molecules may be used to identify an expressed protein of a cell type of the identified cell types, such as a protein that is differentially expressed for a diseased cell state. The sequence information corresponding to the regions of accessible chromatin may be used to identify an SNP or other genetic feature such as a cis-regulatory element corresponding to the expressed protein, thereby identifying an SNP or other genetic feature associated with a diseased cell state.

TRANSCRIPTIONAL NETWORK CONSTRUCTION AND ANALYSIS

[00344] In some embodiments, a feature linkage matrix generated from operation **4380** of the workflow **4300** can be used for further downstream analysis, e.g., to construct a transcriptional network or transcription factor (TF) network. For example, an exemplary workflow for constructing a TF networks is depicted in **FIG. 47**. In some aspects, starting from peak-to-gene feature linkages, peaks and genes can be filtered using motif enrichment and differential expression analysis, respectively. The remaining peaks can be further mapped to motifs. In some embodiments, a 3-step method for constructing a transcriptional network can be used. Firstly, differentially expressed gene in tumor cells are used to generate a list of genes upregulated in a particular condition (e.g., in tumor cells) by setting the threshold of p-value (e.g., < 10^-20). Next, the gene list can be intersected with the inferred feature linkages to identify the peaks linked to the condition (e.g., tumor-specific genes). Finally, motif enrichment analysis can be used to identify transcription factors with enriched motif occurrence in tumor cells. In some embodiments, the enriched transcription factors with linked target genes can be used to define the edges of the TF regulatory network.

[00345] In some embodiments, the method comprises performing motif enrichment analysis. In some examples, transcription factor motifs with increased accessibility in a condition can be identified (e.g., in tumor B cells over normal B cells). In some cases, global enrichment is performed comprising estimating enrichment using all peaks between two populations (e.g., two cell or nucleic populations). In some aspects, TF deviation z scores can be evaluated using chromVAR and two-sample t tests for every motif between the two populations (e.g., between tumor cells and normal B cells). In some embodiments, the inferred mean difference between the two populations is further z scored as the enrichment score. In some embodiments, the method comprises identifying motifs enrichment only in the peaks with linkage to the top tumor-upregulated genes. In some cases, context-dependent enrichment is performed comprising enrichment estimated under the tumor context by comparing potential tumor-specific enhancers with a set of background peaks sharing the GC and accessibility profile in tumor cells. For example, a set of background peaks can be computed using a similar strategy as chromVAR, with modification of matching GC content and peak size, instead of the GC content and cut sites per cell. In some embodiments, the motif enrichment is determined by hypergeometric tests of motif occurrences in enhancer peaks and background peaks. In some embodiments, one or more enrichment motifs are identified (e.g., top ranked motifs in global analysis and the top ranked motifs in context-specific analysis). In some cases, redundancy of TF motif sequences can be

removed and further processed, e.g., by grouping the identified motifs into families based on motif clustering (Fornes et al., (2020) Nucleic Acids Research, 48(D1):D87–D92).

Systems and methods for sample compartmentalization

[00346] In an aspect, the systems and methods described herein provide for the compartmentalization, depositing, or partitioning of one or more particles (e.g., biological particles, macromolecular constituents of biological particles, beads, reagents, etc.) into discrete compartments or partitions (referred to interchangeably herein as partitions), where each partition maintains separation of its own contents from the contents of other partitions. The partition can be a droplet in an emulsion or a well. A partition may comprise one or more other partitions.

[00347] A partition may include one or more particles. A partition may include one or more types of particles. For example, a partition of the present disclosure may comprise one or more biological particles and/or macromolecular constituents thereof. A partition may comprise one or more beads. A partition may comprise one or more gel beads. A partition may comprise one or more cell beads. A partition may include a single gel bead, a single cell bead, or both a single cell bead and single gel bead. A partition may include one or more reagents. Alternatively, a partition may be unoccupied. For example, a partition may not comprise a bead. A cell bead can be an biological particle and/or one or more of its macromolecular constituents encased inside of a gel or polymer matrix, such as via polymerization of a droplet containing the biological particle and precursors capable of being polymerized or gelled. Unique identifiers, such as barcodes, may be injected into the droplets previous to, subsequent to, or concurrently with droplet generation, such as via a microcapsule (e.g., bead), as described elsewhere herein.

[00348] The methods and systems of the present disclosure may comprise methods and systems for generating one or more partitions such as droplets. The droplets may comprise a plurality of droplets in an emulsion. In some examples, the droplets may comprise droplets in a colloid. In some cases, the emulsion may comprise a microemulsion or a nanoemulsion. In some examples, the droplets may be generated with aid of a microfluidic device and/or by subjecting a mixture of immiscible phases to agitation (e.g., in a container). In some cases, a combination of the mentioned methods may be used for droplet and/or emulsion formation.

[00349] Droplets can be formed by creating an emulsion by mixing and/or agitating immiscible phases. Mixing or agitation may comprise various agitation techniques, such as vortexing, pipetting, tube flicking, or other agitation techniques. In some cases, mixing or agitation may be performed without using a microfluidic device. In some examples, the droplets may be formed by exposing a mixture to ultrasound or sonication. Systems and methods for

droplet and/or emulsion generation by agitation are described in International Application No. PCT/US20/17785, which is entirely incorporated herein by reference for all purposes.

[00350] Microfluidic devices or platforms comprising microfluidic channel networks (e.g., on a chip) can be utilized to generate partitions such as droplets and/or emulsions as described herein. Methods and systems for generating partitions such as droplets, methods of encapsulating biological particlemethods of increasing the throughput of droplet generation, and various geometries, architectures, and configurations of microfluidic devices and channels are described in U.S. Patent Publication Nos. 2019/0367997 and 2019/0064173, each of which is entirely incorporated herein by reference for all purposes.

[00351] In some examples, individual particles can be partitioned to discrete partitions by introducing a flowing stream of particles in an aqueous fluid into a flowing stream or reservoir of a non-aqueous fluid, such that droplets may be generated at the junction of the two streams/reservoir, such as at the junction of a microfluidic device provided elsewhere herein.

[00352] The methods of the present disclosure may comprise generating partitions and/or encapsulating particles, such as analyte carriers or analyte carriers, in some cases, individual analyte carriers such as single cells. In some examples, reagents may be encapsulated and/or partitioned (e.g., co-partitioned with analyte carriers) in the partitions. Various mechanisms may be employed in the partitioning of individual particles. An example may comprise porous membranes through which aqueous mixtures of cells may be extruded into fluids (e.g., non-aqueous fluids).

[00353] The partitions can be flowable within fluid streams. The partitions may comprise, for example, micro-vesicles that have an outer barrier surrounding an inner fluid center or core. In some cases, the partitions may comprise a porous matrix that is capable of entraining and/or retaining materials within its matrix. The partitions can be droplets of a first phase within a second phase, wherein the first and second phases are immiscible. For example, the partitions can be droplets of aqueous fluid within a non-aqueous continuous phase (e.g., oil phase). In another example, the partitions can be droplets of a non-aqueous fluid within an aqueous phase. In some examples, the partitions may be provided in a water-in-oil emulsion or oil-in-water emulsion. A variety of different vessels are described in, for example, U.S. Patent Application Publication No. 2014/0155295, which is entirely incorporated herein by reference for all purposes. Emulsion systems for creating stable droplets in non-aqueous or oil continuous phases are described in, for example, U.S. Patent Application Publication No. 2010/0105112, which is entirely incorporated herein by reference for all purposes.

[00354] Fluid properties (e.g., fluid flow rates, fluid viscosities, etc.), particle properties (e.g., volume fraction, particle size, particle concentration, etc.), microfluidic architectures (e.g., channel geometry, etc.), and other parameters may be adjusted to control the occupancy of the resulting partitions (e.g., number of biological particles per partition, number of beads per partition, etc.). For example, partition occupancy can be controlled by providing the aqueous stream at a certain concentration and/or flow rate of particles. To generate single biological particle partitions, the relative flow rates of the immiscible fluids can be selected such that, on average, the partitions may contain less than one biological particle per partition in order to ensure that those partitions that are occupied are primarily singly occupied. In some cases, partitions among a plurality of partitions may contain at most one biological particle (e.g., bead, DNA, cell or cellular material). In some embodiments, the various parameters (e.g., fluid properties, particle properties, microfluidic architectures, etc.) may be selected or adjusted such that a majority of partitions are occupied, for example, allowing for only a small percentage of unoccupied partitions. The flows and channel architectures can be controlled as to ensure a given number of singly occupied partitions, less than a certain level of unoccupied partitions and/or less than a certain level of multiply occupied partitions.

[00355] **FIG. 1** shows an example of a microfluidic channel structure **100** for partitioning individual biological particles. The channel structure **100** can include channel segments **102**, **104**, **106** and **108** communicating at a channel junction **110**. In operation, a first aqueous fluid **112** that includes suspended biological particles (or cells) **114** may be transported along channel segment **102** into junction **110**, while a second fluid **116** that is immiscible with the aqueous fluid **112** is delivered to the junction **110** from each of channel segments **104** and **106** to create discrete droplets **118**, **120** of the first aqueous fluid **112** flowing into channel segment **108**, and flowing away from junction **110**. The channel segment **108** may be fluidically coupled to an outlet reservoir where the discrete droplets can be stored and/or harvested. A discrete droplet generated may include an individual biological particle **114** (such as droplets **118**). A discrete droplet generated may include more than one individual biological particle **114** (not shown in **FIG. 1**). A discrete droplet may contain no biological particle **114** (such as droplet **120**). Each discrete partition may maintain separation of its own contents (e.g., individual biological particle **114**) from the contents of other partitions.

[00356] The second fluid **116** can comprise an oil, such as a fluorinated oil, that includes a fluorosurfactant for stabilizing the resulting droplets, for example, inhibiting subsequent coalescence of the resulting droplets **118, 120**. Examples of particularly useful partitioning

fluids and fluorosurfactants are described, for example, in U.S. Patent Application Publication No. 2010/0105112, which is entirely incorporated herein by reference for all purposes.

[00357] As will be appreciated, the channel segments described herein may be coupled to any of a variety of different fluid sources or receiving components, including reservoirs, tubing, manifolds, or fluidic components of other systems. As will be appreciated, the microfluidic channel structure **100** may have other geometries. For example, a microfluidic channel structure can have more than one channel junction. For example, a microfluidic channel structure can have 2, 3, 4, or 5 channel segments each carrying particles (e.g., biological particles, cell beads, and/or gel beads) that meet at a channel junction. Fluid may be directed to flow along one or more channels or reservoirs via one or more fluid flow units. A fluid flow unit can comprise compressors (e.g., providing positive pressure), pumps (e.g., providing negative pressure), actuators, and the like to control flow of the fluid. Fluid may also or otherwise be controlled via applied pressure differentials, centrifugal force, electrokinetic pumping, vacuum, capillary or gravity flow, or the like.

[00358] The generated droplets may comprise two subsets of droplets: (1) occupied droplets **118**, containing one or more biological particles **114**, and (2) unoccupied droplets **120**, not containing any biological particles **114**. Occupied droplets **118** may comprise singly occupied droplets (having one biological particle) and multiply occupied droplets (having more than one biological particle). As described elsewhere herein, in some cases, the majority of occupied partitions can include no more than one biological particle per occupied partition and some of the generated partitions can be unoccupied (of any biological particle). In some cases, though, some of the occupied partitions may include more than one biological particle. In some cases, the partitioning process may be controlled such that fewer than about 25% of the occupied partitions contain more than one biological particle, and in many cases, fewer than about 20% of the occupied partitions have more than one biological particle, while in some cases, fewer than about 10% or even fewer than about 5% of the occupied partitions include more than one biological particle per partition.

[00359] In some cases, it may be desirable to minimize the creation of excessive numbers of empty partitions, such as to reduce costs and/or increase efficiency. While this minimization may be achieved by providing a sufficient number of biological particles (e.g., biological particles **114**) at the partitioning junction **110**, such as to ensure that at least one biological particle is encapsulated in a partition, the Poissonian distribution may expectedly increase the number of partitions that include multiple biological particles. As such, where singly occupied partitions are to be obtained, at most about 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%,

50%, 45%, 40%, 35%, 30%, 25%, 20%, 15%, 10%, 5% or less of the generated partitions can be unoccupied.

[00360] In some cases, the flow of one or more of the biological particles (e.g., in channel segment **102**), or other fluids directed into the partitioning junction (e.g., in channel segments **104, 106**) can be controlled such that, in many cases, no more than about 50% of the generated partitions, no more than about 25% of the generated partitions, or no more than about 10% of the generated partitions are unoccupied. These flows can be controlled so as to present a non-Poissonian distribution of single-occupied partitions while providing lower levels of unoccupied partitions. The above noted ranges of unoccupied partitions can be achieved while still providing any of the single occupancy rates described above. For example, in many cases, the use of the systems and methods described herein can create resulting partitions that have multiple occupancy rates of less than about 25%, less than about 20%, less than about 15%, less than about 10%, and in many cases, less than about 5%, while having unoccupied partitions of less than about 50%, less than about 40%, less than about 30%, less than about 20%, less than about 10%, less than about 5%, or less.

[00361] As will be appreciated, the above-described occupancy rates are also applicable to partitions that include both biological particles and additional reagents, including, but not limited to, microcapsules or beads (e.g., gel beads) carrying barcoded nucleic acid molecules (e.g., oligonucleotides) (described in relation to **FIG. 2**). The occupied partitions (e.g., at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or 99% of the occupied partitions) can include both a microcapsule (e.g., bead) comprising barcoded nucleic acid molecules and a biological particle.

[00362] In another aspect, in addition to or as an alternative to droplet based partitioning, biological particles may be encapsulated within a microcapsule that comprises an outer shell, layer or porous matrix in which is entrained one or more individual biological particles or small groups of biological particles. The microcapsule may include other reagents. Encapsulation of biological particles may be performed by a variety of processes. Such processes may combine an aqueous fluid containing the biological particles with a polymeric precursor material that may be capable of being formed into a gel or other solid or semi-solid matrix upon application of a particular stimulus to the polymer precursor. Such stimuli can include, for example, thermal stimuli (e.g., either heating or cooling), photo-stimuli (e.g., through photo-curing), chemical stimuli (e.g., through crosslinking, polymerization initiation of the precursor (e.g., through added initiators)), mechanical stimuli, or a combination thereof.

[00363] Preparation of microcapsules comprising biological particles may be performed by a variety of methods. For example, air knife droplet or aerosol generators may be used to dispense droplets of precursor fluids into gelling solutions in order to form microcapsules that include individual biological particles or small groups of biological particles. Likewise, membrane based encapsulation systems may be used to generate microcapsules comprising encapsulated biological particles as described herein. Microfluidic systems of the present disclosure, such as that shown in **FIG. 1**, may be readily used in encapsulating cells as described herein. In particular, and with reference to **FIG. 1**, the aqueous fluid **112** comprising (i) the biological particles **114** and (ii) the polymer precursor material (not shown) is flowed into channel junction **110**, where it is partitioned into droplets **118**, **120** through the flow of non-aqueous fluid **116**. In the case of encapsulation methods, non-aqueous fluid **116** may also include an initiator (not shown) to cause polymerization and/or crosslinking of the polymer precursor to form the microcapsule that includes the entrained biological particles. Examples of polymer precursor/initiator pairs include those described in U.S. Patent Application Publication No. 2014/0378345, which is entirely incorporated herein by reference for all purposes.

[00364] For example, in the case where the polymer precursor material comprises a linear polymer material, such as a linear polyacrylamide, PEG, or other linear polymeric material, the activation agent may comprise a cross-linking agent, or a chemical that activates a cross-linking agent within the formed droplets. Likewise, for polymer precursors that comprise polymerizable monomers, the activation agent may comprise a polymerization initiator. For example, in certain cases, where the polymer precursor comprises a mixture of acrylamide monomer with a N,N'-bis-(acryloyl)cystamine (BAC) comonomer, an agent such as tetraethylmethylenediamine (TEMED) may be provided within the second fluid streams **116** in channel segments **104** and **106**, which can initiate the copolymerization of the acrylamide and BAC into a cross-linked polymer network, or hydrogel.

[00365] Upon contact of the second fluid stream **116** with the first fluid stream **112** at junction **110**, during formation of droplets, the TEMED may diffuse from the second fluid **116** into the aqueous fluid **112** comprising the linear polyacrylamide, which will activate the crosslinking of the polyacrylamide within the droplets **118, 120**, resulting in the formation of gel (e.g., hydrogel) microcapsules, as solid or semi-solid beads or particles entraining the cells **114**. Although described in terms of polyacrylamide encapsulation, other 'activatable' encapsulation compositions may also be employed in the context of the methods and compositions described herein. For example, formation of alginate droplets followed by exposure to divalent metal ions (e.g., Ca^{2+} ions), can be used as an encapsulation process using the described processes.

Likewise, agarose droplets may also be transformed into capsules through temperature based gelling (e.g., upon cooling, etc.).

[00366] In some cases, encapsulated biological particles can be selectively releasable from the microcapsule, such as through passage of time or upon application of a particular stimulus, that degrades the microcapsule sufficiently to allow the biological particles (e.g., cell), or its other contents to be released from the microcapsule, such as into a partition (e.g., droplet). For example, in the case of the polyacrylamide polymer described above, degradation of the microcapsule may be accomplished through the introduction of an appropriate reducing agent, such as DTT or the like, to cleave disulfide bonds that cross-link the polymer matrix. See, for example, U.S. Patent Application Publication No. 2014/0378345, which is entirely incorporated herein by reference for all purposes.

[00367] The biological particle can be subjected to other conditions sufficient to polymerize or gel the precursors. The conditions sufficient to polymerize or gel the precursors may comprise exposure to heating, cooling, electromagnetic radiation, and/or light. The conditions sufficient to polymerize or gel the precursors may comprise any conditions sufficient to polymerize or gel the precursors. Following polymerization or gelling, a polymer or gel may be formed around the biological particle. The polymer or gel may be diffusively permeable to chemical or biochemical reagents. The polymer or gel may be diffusively impermeable to macromolecular constituents of the biological particle. In this manner, the polymer or gel may act to allow the biological particle to be subjected to chemical or biochemical operations while spatially confining the macromolecular constituents to a region of the droplet defined by the polymer or gel. The polymer or gel may include one or more of disulfide cross-linked polyacrylamide, agarose, alginate, polyvinyl alcohol, polyethylene glycol (PEG)-diacrylate, PEG-acrylate, PEG-thiol, PEG-azide, PEG-alkyne, other acrylates, chitosan, hyaluronic acid, collagen, fibrin, gelatin, or elastin. The polymer or gel may comprise any other polymer or gel.

[00368] The polymer or gel may be functionalized to bind to targeted analytes, such as nucleic acids, proteins, carbohydrates, lipids or other analytes. The polymer or gel may be polymerized or gelled via a passive mechanism. The polymer or gel may be stable in alkaline conditions or at elevated temperature. The polymer or gel may have mechanical properties similar to the mechanical properties of the bead. For instance, the polymer or gel may be of a similar size to the bead. The polymer or gel may have a mechanical strength (e.g. tensile strength) similar to that of the bead. The polymer or gel may be of a lower density than an oil. The polymer or gel may be of a density that is roughly similar to that of a buffer. The polymer or gel may have a tunable pore size. The pore size may be chosen to, for instance, retain

denatured nucleic acids. The pore size may be chosen to maintain diffusive permeability to exogenous chemicals such as sodium hydroxide (NaOH) and/or endogenous chemicals such as inhibitors. The polymer or gel may be biocompatible. The polymer or gel may maintain or enhance cell viability. The polymer or gel may be biochemically compatible. The polymer or gel may be polymerized and/or depolymerized thermally, chemically, enzymatically, and/or optically.

[00369] The polymer may comprise poly(acrylamide-co-acrylic acid) crosslinked with disulfide linkages. The preparation of the polymer may comprise a two-step reaction. In the first activation step, poly(acrylamide-co-acrylic acid) may be exposed to an acylating agent to convert carboxylic acids to esters. For instance, the poly(acrylamide-co-acrylic acid) may be exposed to 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMTMM). The polyacrylamide-co-acrylic acid may be exposed to other salts of 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium. In the second cross-linking step, the ester formed in the first step may be exposed to a disulfide crosslinking agent. For instance, the ester may be exposed to cystamine (2,2'-dithiobis(ethylamine)). Following the two steps, the biological particle may be surrounded by polyacrylamide strands linked together by disulfide bridges. In this manner, the biological particle may be encased inside of or comprise a gel or matrix (e.g., polymer matrix) to form a "cell bead."

[00370] A cell bead can contain biological particles (e.g., a cell) or macromolecular constituents (e.g., RNA, DNA, proteins, etc.) of biological particles. A cell bead may include a single cell or multiple cells, or a derivative of the single cell or multiple cells. For example after lysing and washing the cells, inhibitory components from cell lysates can be washed away and the macromolecular constituents can be bound as cell beads. Systems and methods disclosed herein can be applicable to both cell beads (and/or droplets or other partitions) containing biological particles and cell beads (and/or droplets or other partitions) containing macromolecular constituents of biological particles. Cell beads may be or include a cell, cell derivative, cellular material and/or material derived from the cell in, within, or encased in a matrix, such as a polymeric matrix. In some cases, a cell bead may comprise a live cell. In some instances, the live cell may be capable of being cultured when enclosed in a gel or polymer matrix, or of being cultured when comprising a gel or polymer matrix. In some instances, the polymer or gel may be diffusively permeable to certain components and diffusively impermeable to other components (e.g., macromolecular constituents).

[00371] Encapsulated biological particles can provide certain potential advantages of being more storable and more portable than droplet-based partitioned biological particles.

Furthermore, in some cases, it may be desirable to allow biological particles to incubate for a select period of time before analysis, such as in order to characterize changes in such biological particles over time, either in the presence or absence of different stimuli (or reagents). In such cases, encapsulation may allow for longer incubation than partitioning in emulsion droplets, although in some cases, droplet partitioned biological particles may also be incubated for different periods of time, e.g., at least 10 seconds, at least 30 seconds, at least 1 minute, at least 5 minutes, at least 10 minutes, at least 30 minutes, at least 1 hour, at least 2 hours, at least 5 hours, or at least 10 hours or more. The encapsulation of biological particles may constitute the partitioning of the biological particles into which other reagents are co-partitioned. Alternatively or in addition, encapsulated biological particles may be readily deposited into other partitions (e.g., droplets) as described above.

Wells

[00372] As described herein, one or more processes may be performed in a partition, which may be a well. The well may be a well of a plurality of wells of a substrate, such as a microwell of a microwell array or plate, or the well may be a microwell or microchamber of a device (e.g., microfluidic device) comprising a substrate. The well may be a well of a well array or plate, or the well may be a well or chamber of a device (e.g., fluidic device). Accordingly, the wells or microwells may assume an "open" configuration, in which the wells or microwells are exposed to the environment (e.g., contain an open surface) and are accessible on one planar face of the substrate, or the wells or microwells may assume a "closed" or "sealed" configuration, in which the microwells are not accessible on a planar face of the substrate. In some instances, the wells or microwells may be configured to toggle between "open" and "closed" configurations. For instance, an "open" microwell or set of microwells may be "closed" or "sealed" using a membrane (e.g., semi-permeable membrane), an oil (e.g., fluorinated oil to cover an aqueous solution), or a lid, as described elsewhere herein.

[00373] The well may have a volume of less than 1 milliliter (mL). For instance, the well may be configured to hold a volume of at most 1000 microliters (μL), at most 100 μL, at most 10 μL, at most 1 μL, at most 100 nanoliters (nL), at most 10 nL, at most 1 nL, at most 100 picoliters (pL), at most 10 (pL), or less. The well may be configured to hold a volume of about 1000 μL, about 100 μL, about 10 μL, about 1 μL, about 100 nL, about 10 nL, about 1 nL, about 100 pL, about 10 pL, etc. The well may be configured to hold a volume of at least 10 pL, at least 100 pL, at least 1 nL, at least 10 nL, at least 100 nL, at least 1 μL, at least 10 μL, at least 100 μL, at least 1000 μL, or more. The well may be configured to hold a volume in a range of volumes listed herein, for example, from about 5 nL to about 20 nL, from about 1 nL to about 100 nL, from

about 500 pL to about 100 μL, etc. The well may be of a plurality of wells that have varying volumes and may be configured to hold a volume appropriate to accommodate any of the partition volumes described herein.

[00374] In some instances, a microwell array or plate comprises a single variety of microwells. In some instances, a microwell array or plate comprises a variety of microwells. For instance, the microwell array or plate may comprise one or more types of microwells within a single microwell array or plate. The types of microwells may have different dimensions (e.g., length, width, diameter, depth, cross-sectional area, etc.), shapes (e.g., circular, triangular, square, rectangular, pentagonal, hexagonal, heptagonal, octagonal, nonagonal, decagonal, etc.), aspect ratios, or other physical characteristics. The microwell array or plate may comprise any number of different types of microwells. For example, the microwell array or plate may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 or more different types of microwells. A well may have any dimension (e.g., length, width, diameter, depth, cross-sectional area, volume, etc.), shape (e.g., circular, triangular, square, rectangular, pentagonal, hexagonal, heptagonal, octagonal, nonagonal, decagonal, other polygonal, etc.), aspect ratios, or other physical characteristics described herein with respect to any well.

[00375] In certain instances, the microwell array or plate comprises different types of microwells that are located adjacent to one another within the array or plate. For instance, a microwell with one set of dimensions may be located adjacent to and in contact with another microwell with a different set of dimensions. Similarly, microwells of different geometries may be placed adjacent to or in contact with one another. The adjacent microwells may be configured to hold different articles; for example, one microwell may be used to contain a cell, cell bead, or other sample (e.g., cellular components, nucleic acid molecules, etc.) while the adjacent microwell may be used to contain a microcapsule, droplet, bead, or other reagent. In some cases, the adjacent microwells may be configured to merge the contents held within, e.g., upon application of a stimulus, or spontaneously, upon contact of the articles in each microwell.

[00376] As is described elsewhere herein, a plurality of partitions may be used in the systems, compositions, and methods described herein. For example, any suitable number of partitions (e.g., wells or droplets) can be generated or otherwise provided. For example, in the case when wells are used, at least about 1,000 wells, at least about 5,000 wells, at least about 10,000 wells, at least about 50,000 wells, at least about 100,000 wells, at least about 500,000 wells, at least about 1,000,000 wells, at least about 5,000,000 wells at least about 10,000,000 wells, at least about 50,000,000 wells, at least about 100,000,000 wells, at least about 500,000,000 wells, at

least about 1,000,000,000 wells, or more wells can be generated or otherwise provided. Moreover, the plurality of wells may comprise both unoccupied wells (e.g., empty wells) and occupied wells.

[00377] A well may comprise any of the reagents described herein, or combinations thereof. These reagents may include, for example, barcode molecules, enzymes, adapters, and combinations thereof. The reagents may be physically separated from a sample (e.g., a cell, cell bead, or cellular components, e.g., proteins, nucleic acid molecules, etc.) that is placed in the well. This physical separation may be accomplished by containing the reagents within, or coupling to, a microcapsule or bead that is placed within a well. The physical separation may also be accomplished by dispensing the reagents in the well and overlaying the reagents with a layer that is, for example, dissolvable, meltable, or permeable prior to introducing the polynucleotide sample into the well. This layer may be, for example, an oil, wax, membrane (e.g., semi-permeable membrane), or the like. The well may be sealed at any point, for example, after addition of the microcapsule or bead, after addition of the reagents, or after addition of either of these components. The sealing of the well may be useful for a variety of purposes, including preventing escape of beads or loaded reagents from the well, permitting select delivery of certain reagents (e.g., via the use of a semi-permeable membrane), for storage of the well prior to or following further processing, etc.

[00378] A well may comprise free reagents and/or reagents encapsulated in, or otherwise coupled to or associated with, microcapsules, beads, or droplets. Any of the reagents described in this disclosure may be encapsulated in, or otherwise coupled to, a microcapsule, droplet, or bead, with any chemicals, particles, and elements suitable for sample processing reactions involving biomolecules, such as, but not limited to, nucleic acid molecules and proteins. For example, a bead or droplet used in a sample preparation reaction for DNA sequencing may comprise one or more of the following reagents: enzymes, restriction enzymes (e.g., multiple cutters), ligase, polymerase, fluorophores, oligonucleotide barcodes, adapters, buffers, nucleotides (e.g., dNTPs, ddNTPs) and the like.

[00379] Additional examples of reagents include, but are not limited to: buffers, acidic solution, basic solution, temperature-sensitive enzymes, pH-sensitive enzymes, light-sensitive enzymes, metals, metal ions, magnesium chloride, sodium chloride, manganese, aqueous buffer, mild buffer, ionic buffer, inhibitor, enzyme, protein, polynucleotide, antibodies, saccharides, lipid, oil, salt, ion, detergents, ionic detergents, non-ionic detergents, oligonucleotides, nucleotides, deoxyribonucleotide triphosphates (dNTPs), dideoxyribonucleotide triphosphates (ddNTPs), DNA, RNA, peptide polynucleotides, complementary DNA (cDNA), double stranded

DNA (dsDNA), single stranded DNA (ssDNA), plasmid DNA, cosmid DNA, chromosomal DNA, genomic DNA, viral DNA, bacterial DNA, mtDNA (mitochondrial DNA), mRNA, rRNA, tRNA, nRNA, siRNA, snRNA, snoRNA, scaRNA, microRNA, dsRNA, ribozyme, riboswitch and viral RNA, polymerase, ligase, restriction enzymes, proteases, nucleases, protease inhibitors, nuclease inhibitors, chelating agents, reducing agents, oxidizing agents, fluorophores, probes, chromophores, dyes, organics, emulsifiers, surfactants, stabilizers, polymers, water, small molecules, pharmaceuticals, radioactive molecules, preservatives, antibiotics, aptamers, and pharmaceutical drug compounds. As described herein, one or more reagents in the well may be used to perform one or more reactions, including but not limited to: cell lysis, cell fixation, permeabilization, nucleic acid reactions, e.g., nucleic acid extension reactions, amplification, reverse transcription, transposase reactions (e.g., tagmentation), etc.

[00380] The wells may be provided as a part of a kit. For example, a kit may comprise instructions for use, a microwell array or device, and reagents (e.g., beads). The kit may comprise any useful reagents for performing the processes described herein, e.g., nucleic acid reactions, barcoding of nucleic acid molecules, sample processing (e.g., for cell lysis, fixation, and/or permeabilization).

[00381] In some cases, a well comprises a microcapsule, bead, or droplet that comprises a set of reagents that has a similar attribute (e.g., a set of enzymes, a set of minerals, a set of oligonucleotides, a mixture of different barcode molecules, a mixture of identical barcode molecules). In other cases, a microcapsule, bead, or droplet comprises a heterogeneous mixture of reagents. In some cases, the heterogeneous mixture of reagents can comprise all components necessary to perform a reaction. In some cases, such mixture can comprise all components necessary to perform a reaction, except for 1, 2, 3, 4, 5, or more components necessary to perform a reaction. In some cases, such additional components are contained within, or otherwise coupled to, a different microcapsule, droplet, or bead, or within a solution within a partition (e.g., microwell) of the system.

[00382] **FIG. 5** schematically illustrates an example of a microwell array. The array can be contained within a substrate **500**. The substrate **500** comprises a plurality of wells **502**. The wells **502** may be of any size or shape, and the spacing between the wells, the number of wells per substrate, as well as the density of the wells on the substrate **500** can be modified, depending on the particular application. In one such example application, a sample molecule **506**, which may comprise a cell or cellular components (e.g., nucleic acid molecules) is co-partitioned with a bead **504**, which may comprise a nucleic acid barcode molecule coupled thereto. The wells **502** may be loaded using gravity or other loading technique (e.g., centrifugation, liquid handler,

acoustic loading, optoelectronic, etc.). In some instances, at least one of the wells **502** contains a single sample molecule **506** (e.g., cell) and a single bead **504**.

[00383] Reagents may be loaded into a well either sequentially or concurrently. In some cases, reagents are introduced to the device either before or after a particular operation. In some cases, reagents (which may be provided, in certain instances, in microcapsules, droplets, or beads) are introduced sequentially such that different reactions or operations occur at different steps. The reagents (or microcapsules, droplets, or beads) may also be loaded at operations interspersed with a reaction or operation step. For example, microcapsules (or droplets or beads) comprising reagents for fragmenting polynucleotides (e.g., restriction enzymes) and/or other enzymes (e.g., transposases, ligases, polymerases, etc.) may be loaded into the well or plurality of wells, followed by loading of microcapsules, droplets, or beads comprising reagents for attaching nucleic acid barcode molecules to a sample nucleic acid molecule. Reagents may be provided concurrently or sequentially with a sample, e.g., a cell or cellular components (e.g., organelles, proteins, nucleic acid molecules, carbohydrates, lipids, etc.). Accordingly, use of wells may be useful in performing multi-step operations or reactions.

[00384] As described elsewhere herein, the nucleic acid barcode molecules and other reagents may be contained within a microcapsule, bead, or droplet. These microcapsules, beads, or droplets may be loaded into a partition (e.g., a microwell) before, after, or concurrently with the loading of a cell, such that each cell is contacted with a different microcapsule, bead, or droplet. This technique may be used to attach a unique nucleic acid barcode molecule to nucleic acid molecules obtained from each cell. Alternatively or in addition to, the sample nucleic acid molecules may be attached to a support. For instance, the partition (e.g., microwell) may comprise a bead which has coupled thereto a plurality of nucleic acid barcode molecules. The sample nucleic acid molecules, or derivatives thereof, may couple or attach to the nucleic acid barcode molecules on the support. The resulting barcoded nucleic acid molecules may then be removed from the partition, and in some instances, pooled and sequenced. In such cases, the nucleic acid barcode sequences may be used to trace the origin of the sample nucleic acid molecule. For example, polynucleotides with identical barcodes may be determined to originate from the same cell or partition, while polynucleotides with different barcodes may be determined to originate from different cells or partitions.

[00385] The samples or reagents may be loaded in the wells or microwells using a variety of approaches. The samples (e.g., a cell, cell bead, or cellular component) or reagents (as described herein) may be loaded into the well or microwell using an external force, e.g., gravitational force, electrical force, magnetic force, or using mechanisms to drive the sample or reagents into

the well, e.g., via pressure-driven flow, centrifugation, optoelectronics, acoustic loading, electrokinetic pumping, vacuum, capillary flow, etc. In certain cases, a fluid handling system may be used to load the samples or reagents into the well. The loading of the samples or reagents may follow a Poissonian distribution or a non-Poissonian distribution, e.g., super Poisson or sub-Poisson. The geometry, spacing between wells, density, and size of the microwells may be modified to accommodate a useful sample or reagent distribution; for instance, the size and spacing of the microwells may be adjusted such that the sample or reagents may be distributed in a super-Poissonian fashion.

[00386] In one particular non-limiting example, the microwell array or plate comprises pairs of microwells, in which each pair of microwells is configured to hold a droplet (e.g., comprising a single cell) and a single bead (such as those described herein, which may, in some instances, also be encapsulated in a droplet). The droplet and the bead (or droplet containing the bead) may be loaded simultaneously or sequentially, and the droplet and the bead may be merged, e.g., upon contact of the droplet and the bead, or upon application of a stimulus (e.g., external force, agitation, heat, light, magnetic or electric force, etc.). In some cases, the loading of the droplet and the bead is super-Poissonian. In other examples of pairs of microwells, the wells are configured to hold two droplets comprising different reagents and/or samples, which are merged upon contact or upon application of a stimulus. In such instances, the droplet of one microwell of the pair can comprise reagents that may react with an agent in the droplet of the other microwell of the pair. For instance, one droplet can comprise reagents that are configured to release the nucleic acid barcode molecules of a bead contained in another droplet, located in the adjacent microwell. Upon merging of the droplets, the nucleic acid barcode molecules may be released from the bead into the partition (e.g., the microwell or microwell pair that are in contact), and further processing may be performed (e.g., barcoding, nucleic acid reactions, etc.). In cases where intact or live cells are loaded in the microwells, one of the droplets may comprise lysis reagents for lysing the cell upon droplet merging.

[00387] A droplet or microcapsule may be partitioned into a well. The droplets may be selected or subjected to pre-processing prior to loading into a well. For instance, the droplets may comprise cells, and only certain droplets, such as those containing a single cell (or at least one cell), may be selected for use in loading of the wells. Such a pre-selection process may be useful in efficient loading of single cells, such as to obtain a non-Poissonian distribution, or to pre-filter cells for a selected characteristic prior to further partitioning in the wells. Additionally, the technique may be useful in obtaining or preventing cell doublet or multiplet formation prior to or during loading of the microwell.

[00388] In some instances, the wells can comprise nucleic acid barcode molecules attached thereto. The nucleic acid barcode molecules may be attached to a surface of the well (e.g., a wall of the well). The nucleic acid barcode molecule (e.g., a partition barcode sequence) of one well may differ from the nucleic acid barcode molecule of another well, which can permit identification of the contents contained with a single partition or well. In some cases, the nucleic acid barcode molecule can comprise a spatial barcode sequence that can identify a spatial coordinate of a well, such as within the well array or well plate. In some cases, the nucleic acid barcode molecule can comprise a unique molecular identifier for individual molecule identification. In some instances, the nucleic acid barcode molecules may be configured to attach to or capture a nucleic acid molecule within a sample or cell distributed in the well. For example, the nucleic acid barcode molecules may comprise a capture sequence that may be used to capture or hybridize to a nucleic acid molecule (e.g., RNA, DNA) within the sample. In some instances, the nucleic acid barcode molecules may be releasable from the microwell. For instance, the nucleic acid barcode molecules may comprise a chemical cross-linker which may be cleaved upon application of a stimulus (e.g., photo-, magnetic, chemical, biological, stimulus). The released nucleic acid barcode molecules, which may be hybridized or configured to hybridize to a sample nucleic acid molecule, may be collected and pooled for further processing, which can include nucleic acid processing (e.g., amplification, extension, reverse transcription, etc.) and/or characterization (e.g., sequencing). In such cases, the unique partition barcode sequences may be used to identify the cell or partition from which a nucleic acid molecule originated.

[00389] Characterization of samples within a well may be performed. Such characterization can include, in non-limiting examples, imaging of the sample (e.g., cell, cell bead, or cellular components) or derivatives thereof. Characterization techniques such as microscopy or imaging may be useful in measuring sample profiles in fixed spatial locations. For instance, when cells are partitioned, optionally with beads, imaging of each microwell and the contents contained therein may provide useful information on cell doublet formation (e.g., frequency, spatial locations, etc.), cell-bead pair efficiency, cell viability, cell size, cell morphology, expression level of a biomarker (e.g., a surface marker, a fluorescently labeled molecule therein, etc.), cell or bead loading rate, number of cell-bead pairs, etc. In some instances, imaging may be used to characterize live cells in the wells, including, but not limited to: dynamic live-cell tracking, cell-cell interactions (when two or more cells are co-partitioned), cell proliferation, etc. Alternatively or in addition to, imaging may be used to characterize a quantity of amplification products in the well.

[00390] In operation, a well may be loaded with a sample and reagents, simultaneously or sequentially. When cells or cell beads are loaded, the well may be subjected to washing, e.g., to remove excess cells from the well, microwell array, or plate. Similarly, washing may be performed to remove excess beads or other reagents from the well, microwell array, or plate. In the instances where live cells are used, the cells may be lysed in the individual partitions to release the intracellular components or cellular analytes. Alternatively, the cells may be fixed or permeabilized in the individual partitions. The intracellular components or cellular analytes may couple to a support, e.g., on a surface of the microwell, on a solid support (e.g., bead), or they may be collected for further downstream processing. For instance, after cell lysis, the intracellular components or cellular analytes may be transferred to individual droplets or other partitions for barcoding. Alternatively, or in addition to, the intracellular components or cellular analytes (e.g., nucleic acid molecules) may couple to a bead comprising a nucleic acid barcode molecule; subsequently, the bead may be collected and further processed, e.g., subjected to nucleic acid reaction such as reverse transcription, amplification, or extension, and the nucleic acid molecules thereon may be further characterized, e.g., via sequencing. Alternatively, or in addition to, the intracellular components or cellular analytes may be barcoded in the well (e.g., using a bead comprising nucleic acid barcode molecules that are releasable or on a surface of the microwell comprising nucleic acid barcode molecules). The barcoded nucleic acid molecules or analytes may be further processed in the well, or the barcoded nucleic acid molecules or analytes may be collected from the individual partitions and subjected to further processing outside the partition. Further processing can include nucleic acid processing (e.g., performing an amplification, extension) or characterization (e.g., fluorescence monitoring of amplified molecules, sequencing). At any convenient or useful step, the well (or microwell array or plate) may be sealed (e.g., using an oil, membrane, wax, etc.), which enables storage of the assay or selective introduction of additional reagents.

[00391] **FIG. 6** schematically shows an example workflow for processing nucleic acid molecules within a sample. A substrate **600** comprising a plurality of microwells **602** may be provided. A sample **606** which may comprise a cell, cell bead, cellular components or analytes (e.g., proteins and/or nucleic acid molecules) can be co-partitioned, in a plurality of microwells **602**, with a plurality of beads **604** comprising nucleic acid barcode molecules. During process **610**, the sample **606** may be processed within the partition. For instance, in the case of live cells, the cell may be subjected to conditions sufficient to lyse the cells and release the analytes contained therein. In process **620**, the bead **604** may be further processed. By way of example,

processes **620a** and **620b** schematically illustrate different workflows, depending on the properties of the bead **604**.

[00392] In **620a**, the bead comprises nucleic acid barcode molecules that are attached thereto, and sample nucleic acid molecules (e.g., RNA, DNA) may attach, e.g., via hybridization of ligation, to the nucleic acid barcode molecules. Such attachment may occur on the bead. In process **630**, the beads **604** from multiple wells **602** may be collected and pooled. Further processing may be performed in process **640**. For example, one or more nucleic acid reactions may be performed, such as reverse transcription, nucleic acid extension, amplification, ligation, transposition, etc. In some instances, adapter sequences are ligated to the nucleic acid molecules, or derivatives thereof, as described elsewhere herein. For instance, sequencing primer sequences may be appended to each end of the nucleic acid molecule. In process **650**, further characterization, such as sequencing may be performed to generate sequencing reads. The sequencing reads may yield information on individual cells or populations of cells, which may be represented visually or graphically, e.g., in a plot **655**.

[00393] In **620b**, the bead comprises nucleic acid barcode molecules that are releasably attached thereto, as described below. The bead may degrade or otherwise release the nucleic acid barcode molecules into the well **602**; the nucleic acid barcode molecules may then be used to barcode nucleic acid molecules within the well **602**. Further processing may be performed either inside the partition or outside the partition. For example, one or more nucleic acid reactions may be performed, such as reverse transcription, nucleic acid extension, amplification, ligation, transposition, etc. In some instances, adapter sequences are ligated to the nucleic acid molecules, or derivatives thereof, as described elsewhere herein. For instance, sequencing primer sequences may be appended to each end of the nucleic acid molecule. In process **650**, further characterization, such as sequencing may be performed to generate sequencing reads. The sequencing reads may yield information on individual cells or populations of cells, which may be represented visually or graphically, e.g., in a plot **655**.

Beads

[00394] Nucleic acid barcode molecules may be delivered to a partition (e.g., a droplet or well) via a solid support or carrier (e.g., a bead). In some cases, nucleic acid barcode molecules are initially associated with the solid support and then released from the solid support upon application of a stimulus, which allows the nucleic acid barcode molecules to dissociate or to be released from the solid support. In specific examples, nucleic acid barcode molecules are initially associated with the solid support (e.g., bead) and then released from the solid support

upon application of a biological stimulus, a chemical stimulus, a thermal stimulus, an electrical stimulus, a magnetic stimulus, and/or a photo stimulus.

[00395] A nucleic acid barcode molecule may contain a barcode sequence and a functional sequence, such as a nucleic acid primer sequence or a template switch oligonucleotide (TSO) sequence.

[00396] The solid support may be a bead. A solid support, e.g., a bead, may be porous, non-porous, hollow (e.g., a microcapsule), solid, semi-solid, and/or a combination thereof. Beads may be solid, semi-solid, semi-fluidic, fluidic, and/or a combination thereof. In some instances, a solid support, e.g., a bead, may be at least partially dissolvable, disruptable, and/or degradable. In some cases, a solid support, e.g., a bead, may not be degradable. In some cases, the solid support, e.g., a bead, may be a gel bead. A gel bead may be a hydrogel bead. A gel bead may be formed from molecular precursors, such as a polymeric or monomeric species. A semi-solid support, e.g., a bead, may be a liposomal bead. Solid supports, e.g., beads, may comprise metals including iron oxide, gold, and silver. In some cases, the solid support, e.g., the bead, may be a silica bead. In some cases, the solid support, e.g., a bead, can be rigid. In other cases, the solid support, e.g., a bead, may be flexible and/or compressible.

[00397] A partition may comprise one or more unique identifiers, such as barcodes. Barcodes may be previously, subsequently or concurrently delivered to the partitions that hold the compartmentalized or partitioned biological particle. For example, barcodes may be injected into droplets or deposited in microwells previous to, subsequent to, or concurrently with droplet generation or providing of reagents in the microwells, respectively. The delivery of the barcodes to a particular partition allows for the later attribution of the characteristics of the individual biological particle to the particular partition. Barcodes may be delivered, for example on a nucleic acid molecule (e.g., an oligonucleotide), to a partition via any suitable mechanism. Barcoded nucleic acid molecules can be delivered to a partition via a microcapsule. A microcapsule, in some instances, can comprise a bead. Beads are described in further detail below.

[00398] In some cases, barcoded nucleic acid molecules can be initially associated with the microcapsule and then released from the microcapsule. Release of the barcoded nucleic acid molecules can be passive (e.g., by diffusion out of the microcapsule). In addition or alternatively, release from the microcapsule can be upon application of a stimulus which allows the barcoded nucleic acid nucleic acid molecules to dissociate or to be released from the microcapsule. Such stimulus may disrupt the microcapsule, an interaction that couples the barcoded nucleic acid molecules to or within the microcapsule, or both. Such stimulus can

include, for example, a thermal stimulus, photo-stimulus, chemical stimulus (e.g., change in pH or use of a reducing agent(s)), a mechanical stimulus, a radiation stimulus; a biological stimulus (e.g., enzyme), or any combination thereof. Methods and systems for partitioning barcode carrying beads into droplets are provided in US. Patent Publication Nos. 2019/0367997 and 2019/0064173, and International Application No. PCT/US20/17785, each of which is herein entirely incorporated by reference for all purposes.

[00399] In some examples, beads, analyte carriers, and droplets may flow along channels (e.g., the channels of a microfluidic device), in some cases at substantially regular flow profiles (e.g., at regular flow rates). Such regular flow profiles may permit a droplet to include a single bead and a single biological particle. Such regular flow profiles may permit the droplets to have an occupancy (e.g., droplets having beads and biological particles) greater than 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 95%. Such regular flow profiles and devices that may be used to provide such regular flow profiles are provided in, for example, U.S. Patent Publication No. 2015/0292988, which is entirely incorporated herein by reference.

[00400] A bead may be porous, non-porous, solid, semi-solid, semi-fluidic, fluidic, and/or a combination thereof. In some instances, a bead may be dissolvable, disruptable, and/or degradable. In some cases, a bead may not be degradable. In some cases, the bead may be a gel bead. A gel bead may be a hydrogel bead. A gel bead may be formed from molecular precursors, such as a polymeric or monomeric species. A semi-solid bead may be a liposomal bead. Solid beads may comprise metals including iron oxide, gold, and silver. In some cases, the bead may be a silica bead. In some cases, the bead can be rigid. In other cases, the bead may be flexible and/or compressible.

[00401] A bead may be of any suitable shape. Examples of bead shapes include, but are not limited to, spherical, non-spherical, oval, oblong, amorphous, circular, cylindrical, and variations thereof.

[00402] Beads may be of uniform size or heterogeneous size. In some cases, the diameter of a bead may be at least about 10 nanometers (nm), 100 nm, 500 nm, 1 micrometer (μm), 5μm, 10μm, 20μm, 30μm, 40μm, 50μm, 60μm, 70μm, 80μm, 90μm, 100μm, 250μm, 500μm, 1mm, or greater. In some cases, a bead may have a diameter of less than about 10 nm, 100 nm, 500 nm, 1μm, 5μm, 10μm, 20μm, 30μm, 40μm, 50μm, 60μm, 70μm, 80μm, 90μm, 100μm, 250μm, 500μm, 1mm, or less. In some cases, a bead may have a diameter in the range of about 40-75μm, 30-75μm, 20-75μm, 40-85μm, 40-95μm, 20-100μm, 10-100μm, 1-100μm, 20-250μm, or 20-500μm.

[00403] In certain aspects, beads can be provided as a population or plurality of beads having a relatively monodisperse size distribution. Where it may be desirable to provide relatively consistent amounts of reagents within partitions, maintaining relatively consistent bead characteristics, such as size, can contribute to the overall consistency. In particular, the beads described herein may have size distributions that have a coefficient of variation in their cross-sectional dimensions of less than 50%, less than 40%, less than 30%, less than 20%, and in some cases less than 15%, less than 10%, less than 5%, or less.

[00404] A bead may comprise natural and/or synthetic materials. For example, a bead can comprise a natural polymer, a synthetic polymer or both natural and synthetic polymers. Examples of natural polymers include proteins and sugars such as deoxyribonucleic acid, rubber, cellulose, starch (e.g., amylose, amylopectin), proteins, enzymes, polysaccharides, silks, polyhydroxyalkanoates, chitosan, dextran, collagen, carrageenan, ispaghula, acacia, agar, gelatin, shellac, sterculia gum, xanthan gum, Corn sugar gum, guar gum, gum karaya, agarose, alginic acid, alginate, or natural polymers thereof. Examples of synthetic polymers include acrylics, nylons, silicones, spandex, viscose rayon, polycarboxylic acids, polyvinyl acetate, polyacrylamide, polyacrylate, polyethylene glycol, polyurethanes, polylactic acid, silica, polystyrene, polyacrylonitrile, polybutadiene, polycarbonate, polyethylene, polyethylene terephthalate, poly(chlorotrifluoroethylene), poly(ethylene oxide), poly(ethylene terephthalate), polyethylene, polyisobutylene, poly(methyl methacrylate), poly(oxymethylene), polyformaldehyde, polypropylene, polystyrene, poly(tetrafluoroethylene), poly(vinyl acetate), poly(vinyl alcohol), poly(vinyl chloride), poly(vinylidene dichloride), poly(vinylidene difluoride), poly(vinyl fluoride) and/or combinations (e.g., co-polymers) thereof. Beads may also be formed from materials other than polymers, including lipids, micelles, ceramics, glass-ceramics, material composites, metals, other inorganic materials, and others.

[00405] In some instances, the bead may contain molecular precursors (e.g., monomers or polymers), which may form a polymer network via polymerization of the molecular precursors. In some cases, a precursor may be an already polymerized species capable of undergoing further polymerization via, for example, a chemical cross-linkage. In some cases, a precursor can comprise one or more of an acrylamide or a methacrylamide monomer, oligomer, or polymer. In some cases, the bead may comprise prepolymers, which are oligomers capable of further polymerization. For example, polyurethane beads may be prepared using prepolymers. In some cases, the bead may contain individual polymers that may be further polymerized together. In some cases, beads may be generated via polymerization of different precursors, such that they comprise mixed polymers, co-polymers, and/or block co-polymers. In some cases, the bead may

comprise covalent or ionic bonds between polymeric precursors (e.g., monomers, oligomers, linear polymers), nucleic acid molecules (e.g., oligonucleotides), primers, and other entities. In some cases, the covalent bonds can be carbon-carbon bonds, thioether bonds, or carbon-heteroatom bonds.

[00406] Cross-linking may be permanent or reversible, depending upon the particular cross-linker used. Reversible cross-linking may allow for the polymer to linearize or dissociate under appropriate conditions. In some cases, reversible cross-linking may also allow for reversible attachment of a material bound to the surface of a bead. In some cases, a cross-linker may form disulfide linkages. In some cases, the chemical cross-linker forming disulfide linkages may be cystamine or a modified cystamine.

[00407] In some cases, disulfide linkages can be formed between molecular precursor units (e.g., monomers, oligomers, or linear polymers) or precursors incorporated into a bead and nucleic acid molecules (e.g., oligonucleotides). Cystamine (including modified cystamines), for example, is an organic agent comprising a disulfide bond that may be used as a crosslinker agent between individual monomeric or polymeric precursors of a bead. Polyacrylamide may be polymerized in the presence of cystamine or a species comprising cystamine (e.g., a modified cystamine) to generate polyacrylamide gel beads comprising disulfide linkages (e.g., chemically degradable beads comprising chemically-reducible cross-linkers). The disulfide linkages may permit the bead to be degraded (or dissolved) upon exposure of the bead to a reducing agent.

[00408] In some cases, chitosan, a linear polysaccharide polymer, may be crosslinked with glutaraldehyde via hydrophilic chains to form a bead. Crosslinking of chitosan polymers may be achieved by chemical reactions that are initiated by heat, pressure, change in pH, and/or radiation.

[00409] In some cases, a bead may comprise an acrydite moiety, which in certain aspects may be used to attach one or more nucleic acid molecules (e.g., barcode sequence, barcoded nucleic acid molecule, barcoded oligonucleotide, primer, or other oligonucleotide) to the bead. In some cases, an acrydite moiety can refer to an acrydite analogue generated from the reaction of acrydite with one or more species, such as, the reaction of acrydite with other monomers and cross-linkers during a polymerization reaction. Acrydite moieties may be modified to form chemical bonds with a species to be attached, such as a nucleic acid molecule (e.g., barcode sequence, barcoded nucleic acid molecule, barcoded oligonucleotide, primer, or other oligonucleotide). Acrydite moieties may be modified with thiol groups capable of forming a disulfide bond or may be modified with groups already comprising a disulfide bond. The thiol or disulfide (via disulfide exchange) may be used as an anchor point for a species to be attached

or another part of the acrydite moiety may be used for attachment. In some cases, attachment can be reversible, such that when the disulfide bond is broken (e.g., in the presence of a reducing agent), the attached species is released from the bead. In other cases, an acrydite moiety can comprise a reactive hydroxyl group that may be used for attachment.

[00410] Functionalization of beads for attachment of nucleic acid molecules (e.g., oligonucleotides) may be achieved through a wide range of different approaches, including activation of chemical groups within a polymer, incorporation of active or activatable functional groups in the polymer structure, or attachment at the pre-polymer or monomer stage in bead production.

[00411] For example, precursors (e.g., monomers, cross-linkers) that are polymerized to form a bead may comprise acrydite moieties, such that when a bead is generated, the bead also comprises acrydite moieties. The acrydite moieties can be attached to a nucleic acid molecule (e.g., oligonucleotide) that comprises one or more functional sequences, such as a TSO sequence or a primer sequence (e.g., a poly T sequence, or a nucleic acid primer sequence complementary to a target nucleic acid sequence and/or for amplifying a target nucleic acid sequence, a random primer, primer sequence for messenger RNA) and/or one or more barcode sequences. The one more barcode sequences may include sequences that are the same for all nucleic acid molecules coupled to a bead and/or sequences that are different across all nucleic acid molecules coupled to the bead. The nucleic acid molecule may be incorporated into the bead.

[00412] In some cases, the nucleic acid molecule can comprise a functional sequence, for example, for attachment to a sequencing flow cell, such as, for example, a P5 sequence (or a portion thereof) for Illumina® sequencing. In some cases, the nucleic acid molecule or derivative thereof (e.g., oligonucleotide or polynucleotide generated from the nucleic acid molecule) can comprise another functional sequence, such as, for example, a P7 sequence (or a portion thereof) for attachment to a sequencing flow cell for Illumina sequencing. In some cases, the nucleic acid molecule can comprise a barcode sequence. In some cases, the nucleic acid molecule can further comprise a unique molecular identifier (UMI). In some cases, the nucleic acid molecule can comprise an R1 primer sequence for Illumina sequencing. In some cases, the nucleic acid molecule can comprise an R2 primer sequence for Illumina sequencing. Examples of such nucleic acid molecules (e.g., oligonucleotides, polynucleotides, etc.) and uses thereof, as may be used with compositions, devices, methods and systems of the present disclosure, are provided in U.S. Patent Pub. Nos. 2014/0378345 and 2015/0376609, each of which is entirely incorporated herein by reference.

[00413] In some cases, the nucleic acid molecule can comprise one or more functional sequences. For example, a functional sequence can comprise a sequence for attachment to a sequencing flow cell, such as, for example, a P5 sequence for Illumina® sequencing. In some cases, the nucleic acid molecule or derivative thereof (e.g., oligonucleotide or polynucleotide generated from the nucleic acid molecule) can comprise another functional sequence, such as, for example, a P7 sequence for attachment to a sequencing flow cell for Illumina sequencing. In some cases, the functional sequence can comprise a barcode sequence or multiple barcode sequences. In some cases, the functional sequence can comprise a unique molecular identifier (UMI). In some cases, the functional sequence can comprise a primer sequence (e.g., an R1 primer sequence for Illumina sequencing, an R2 primer sequence for Illumina sequencing, etc.). In some cases, a functional sequence can comprise a partial sequence, such as a partial barcode sequence, partial anchoring sequence, partial sequencing primer sequence (e.g., partial R1 sequence, partial R2 sequence, etc.), a partial sequence configured to attach to the flow cell of a sequencer (e.g., partial P5 sequence, partial P7 sequence, etc.), or a partial sequence of any other type of sequence described elsewhere herein. A partial sequence may contain a contiguous or continuous portion or segment, but not all, of a full sequence, for example. In some cases, a downstream procedure may extend the partial sequence, or derivative thereof, to achieve a full sequence of the partial sequence, or derivative thereof.

[00414] Examples of such nucleic acid molecules (e.g., oligonucleotides, polynucleotides, etc.) and uses thereof, as may be used with compositions, devices, methods and systems of the present disclosure, are provided in U.S. Patent Pub. Nos. 2014/0378345 and 2015/0376609, each of which is entirely incorporated herein by reference.

[00415] **FIG. 3** illustrates an example of a barcode carrying bead. A nucleic acid molecule **302**, such as an oligonucleotide, can be coupled to a bead **304** by a releasable linkage **306**, such as, for example, a disulfide linker. The same bead **304** may be coupled (e.g., via releasable linkage) to one or more other nucleic acid molecules **318, 320**. The nucleic acid molecule **302** may be or comprise a barcode. As noted elsewhere herein, the structure of the barcode may comprise a number of sequence elements. The nucleic acid molecule **302** may comprise a functional sequence **308** that may be used in subsequent processing. For example, the functional sequence **308** may include one or more of a sequencer specific flow cell attachment sequence (e.g., a P5 sequence for Illumina® sequencing systems) and a sequencing primer sequence (e.g., a R1 primer for Illumina® sequencing systems), or partial sequence(s) thereof. The nucleic acid molecule **302** may comprise a barcode sequence **310** for use in barcoding the sample (e.g., DNA, RNA, protein, etc.). In some cases, the barcode sequence **310** can be bead-specific such that the

barcode sequence **310** is common to all nucleic acid molecules (e.g., including nucleic acid molecule **302**) coupled to the same bead **304**. Alternatively or in addition, the barcode sequence **310** can be partition-specific such that the barcode sequence **310** is common to all nucleic acid molecules coupled to one or more beads that are partitioned into the same partition. The nucleic acid molecule **302** may comprise a specific priming sequence **312**, such as an mRNA specific priming sequence (e.g., poly-T sequence), a targeted priming sequence, and/or a random priming sequence. The nucleic acid molecule **302** may comprise an anchoring sequence **314** to ensure that the specific priming sequence **312** hybridizes at the sequence end (e.g., of the mRNA). For example, the anchoring sequence **314** can include a random short sequence of nucleotides, such as a 1-mer, 2-mer, 3-mer or longer sequence, which can ensure that a poly-T segment is more likely to hybridize at the sequence end of the poly-A tail of the mRNA.

[00416] The nucleic acid molecule **302** may comprise a unique molecular identifying sequence **316** (e.g., unique molecular identifier (UMI)). In some cases, the unique molecular identifying sequence **316** may comprise from about 5 to about 8 nucleotides. Alternatively, the unique molecular identifying sequence **316** may compress less than about 5 or more than about 8 nucleotides. The unique molecular identifying sequence **316** may be a unique sequence that varies across individual nucleic acid molecules (e.g., **302, 318, 320**, etc.) coupled to a single bead (e.g., bead **304**). In some cases, the unique molecular identifying sequence **316** may be a random sequence (e.g., such as a random N-mer sequence). For example, the UMI may provide a unique identifier of the starting mRNA molecule that was captured, in order to allow quantitation of the number of original expressed RNA. As will be appreciated, although **FIG. 3** shows three nucleic acid molecules **302, 318, 320** coupled to the surface of the bead **304**, an individual bead may be coupled to any number of individual nucleic acid molecules, for example, from one to tens to hundreds of thousands or even millions of individual nucleic acid molecules. The respective barcodes for the individual nucleic acid molecules can comprise both common sequence segments or relatively common sequence segments (e.g., **308**, **310, 312**, etc.) and variable or unique sequence segments (e.g., **316**) between different individual nucleic acid molecules coupled to the same bead.

[00417] In operation, a biological particle (e.g., cell, DNA, RNA, etc.) can be co-partitioned along with a barcode bearing bead **304**. The nucleic acid barcode molecules **302, 318, 320** can be released from the bead **304** in the partition. By way of example, in the context of analyzing sample RNA, the poly-T segment (e.g., **312**) of one of the released nucleic acid molecules (e.g., **302**) can hybridize to the poly-A tail of a mRNA molecule. Reverse transcription may result in a cDNA transcript of the mRNA, but which transcript includes each of the sequence segments

308, 310, 316 of the nucleic acid molecule **302**. Because the nucleic acid molecule **302** comprises an anchoring sequence **314**, it will more likely hybridize to and prime reverse transcription at the sequence end of the poly-A tail of the mRNA. Within any given partition, all of the cDNA transcripts of the individual mRNA molecules may include a common barcode sequence segment **310**. However, the transcripts made from the different mRNA molecules within a given partition may vary at the unique molecular identifying sequence **312** segment (e.g., UMI segment). Beneficially, even following any subsequent amplification of the contents of a given partition, the number of different UMIs can be indicative of the quantity of mRNA originating from a given partition, and thus from the biological particle (e.g., cell). As noted above, the transcripts can be amplified, cleaned up and sequenced to identify the sequence of the cDNA transcript of the mRNA, as well as to sequence the barcode segment and the UMI segment. While a poly-T primer sequence is described, other targeted or random priming sequences may also be used in priming the reverse transcription reaction. Likewise, although described as releasing the barcoded oligonucleotides into the partition, in some cases, the nucleic acid molecules bound to the bead (e.g., gel bead) may be used to hybridize and capture the mRNA on the solid phase of the bead, for example, in order to facilitate the separation of the RNA from other cell contents. In such cases, further processing may be performed, in the partitions or outside the partitions (e.g., in bulk). For instance, the RNA molecules on the beads may be subjected to reverse transcription or other nucleic acid processing, additional adapter sequences may be added to the barcoded nucleic acid molecules, or other nucleic acid reactions (e.g., amplification, nucleic acid extension) may be performed. The beads or products thereof (e.g., barcoded nucleic acid molecules) may be collected from the partitions, and/or pooled together and subsequently subjected to clean up and further characterization (e.g., sequencing).

[00418] The operations described herein may be performed at any useful or convenient step. For instance, the beads comprising nucleic acid barcode molecules may be introduced into a partition (e.g., well or droplet) prior to, during, or following introduction of a sample into the partition. The nucleic acid molecules of a sample may be subjected to barcoding, which may occur on the bead (in cases where the nucleic acid molecules remain coupled to the bead) or following release of the nucleic acid barcode molecules into the partition. In cases where the nucleic acid molecules from the sample remain attached to the bead, the beads from various partitions may be collected, pooled, and subjected to further processing (e.g., reverse transcription, adapter attachment, amplification, clean up, sequencing). In other instances, the processing may occur in the partition. For example, conditions sufficient for barcoding, adapter

attachment, reverse transcription, or other nucleic acid processing operations may be provided in the partition and performed prior to clean up and sequencing.

[00419] In some instances, a bead may comprise a capture sequence or binding sequence configured to bind to a corresponding capture sequence or binding sequence. In some instances, a bead may comprise a plurality of different capture sequences or binding sequences configured to bind to different respective corresponding capture sequences or binding sequences. For example, a bead may comprise a first subset of one or more capture sequences each configured to bind to a first corresponding capture sequence, a second subset of one or more capture sequences each configured to bind to a second corresponding capture sequence, a third subset of one or more capture sequences each configured to bind to a third corresponding capture sequence, and etc. A bead may comprise any number of different capture sequences. In some instances, a bead may comprise at least 2, 3, 4, 5, 6, 7, 8, 9, 10 or more different capture sequences or binding sequences configured to bind to different respective capture sequences or binding sequences, respectively. Alternatively or in addition, a bead may comprise at most about 10, 9, 8, 7, 6, 5, 4, 3, or 2 different capture sequences or binding sequences configured to bind to different respective capture sequences or binding sequences. In some instances, the different capture sequences or binding sequences may be configured to facilitate analysis of a same type of analyte. In some instances, the different capture sequences or binding sequences may be configured to facilitate analysis of different types of analytes (with the same bead). The capture sequence may be designed to attach to a corresponding capture sequence. Beneficially, such corresponding capture sequence may be introduced to, or otherwise induced in, a biological particle (e.g., cell, cell bead, etc.) for performing different assays in various formats (e.g., barcoded antibodies comprising the corresponding capture sequence, barcoded MHC dextramers comprising the corresponding capture sequence, barcoded guide RNA molecules comprising the corresponding capture sequence, etc.), such that the corresponding capture sequence may later interact with the capture sequence associated with the bead. In some instances, a capture sequence coupled to a bead (or other support) may be configured to attach to a linker molecule, such as a splint molecule, wherein the linker molecule is configured to couple the bead (or other support) to other molecules through the linker molecule, such as to one or more analytes or one or more other linker molecules.

[00420] **FIG. 4** illustrates another example of a barcode carrying bead. A nucleic acid molecule **405**, such as an oligonucleotide, can be coupled to a bead **404** by a releasable linkage **406**, such as, for example, a disulfide linker. The nucleic acid molecule **405** may comprise a first capture sequence **460**. The same bead **404** may be coupled (e.g., via releasable linkage) to

one or more other nucleic acid molecules **403, 407** comprising other capture sequences. The nucleic acid molecule **405** may be or comprise a barcode. As noted elsewhere herein, the structure of the barcode may comprise a number of sequence elements, such as a functional sequence **408** (e.g., flow cell attachment sequence, sequencing primer sequence, etc.), a barcode sequence **410** (e.g., bead-specific sequence common to bead, partition-specific sequence common to partition, etc.), and a unique molecular identifier **412** (e.g., unique sequence within different molecules attached to the bead), or partial sequences thereof. The capture sequence **460** may be configured to attach to a corresponding capture sequence **465**. In some instances, the corresponding capture sequence **465** may be coupled to another molecule that may be an analyte or an intermediary carrier. For example, as illustrated in **FIG. 4**, the corresponding capture sequence **465** is coupled to a guide RNA molecule **462** comprising a target sequence **464**, wherein the target sequence **464** is configured to attach to the analyte. Another oligonucleotide molecule **407** attached to the bead **404** comprises a second capture sequence **480** which is configured to attach to a second corresponding capture sequence **485**. As illustrated in **FIG. 4**, the second corresponding capture sequence **485** is coupled to an antibody **482**. In some cases, the antibody **482** may have binding specificity to an analyte (e.g., surface protein). Alternatively, the antibody **482** may not have binding specificity. Another oligonucleotide molecule **403** attached to the bead **404** comprises a third capture sequence **470** which is configured to attach to a second corresponding capture sequence **475**. As illustrated in **FIG. 4**, the third corresponding capture sequence **475** is coupled to a molecule **472**. The molecule **472** may or may not be configured to target an analyte. The other oligonucleotide molecules **403**, **407** may comprise the other sequences (e.g., functional sequence, barcode sequence, UMI, etc.) described with respect to oligonucleotide molecule **405**. While a single oligonucleotide molecule comprising each capture sequence is illustrated in **4**, it will be appreciated that, for each capture sequence, the bead may comprise a set of one or more oligonucleotide molecules each comprising the capture sequence. For example, the bead may comprise any number of sets of one or more different capture sequences. Alternatively or in addition, the bead **404** may comprise other capture sequences. Alternatively or in addition, the bead **404** may comprise fewer types of capture sequences (e.g., two capture sequences). Alternatively or in addition, the bead **404** may comprise oligonucleotide molecule(s) comprising a priming sequence, such as a specific priming sequence such as an mRNA specific priming sequence (e.g., poly-T sequence), a targeted priming sequence, and/or a random priming sequence, for example, to facilitate an assay for gene expression.

[00421] In operation, the barcoded oligonucleotides may be released (e.g., in a partition), as described elsewhere herein. Alternatively, the nucleic acid molecules bound to the bead (e.g., gel bead) may be used to hybridize and capture analytes (e.g., one or more types of analytes) on the solid phase of the bead.

[00422] In some cases, precursors comprising a functional group that is reactive or capable of being activated such that it becomes reactive can be polymerized with other precursors to generate gel beads comprising the activated or activatable functional group. The functional group may then be used to attach additional species (e.g., disulfide linkers, primers, other oligonucleotides, etc.) to the gel beads. For example, some precursors comprising a carboxylic acid (COOH) group can co-polymerize with other precursors to form a gel bead that also comprises a COOH functional group. In some cases, acrylic acid (a species comprising free COOH groups), acrylamide, and bis(acryloyl)cystamine can be co-polymerized together to generate a gel bead comprising free COOH groups. The COOH groups of the gel bead can be activated (e.g., via 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) and N-Hydroxysuccinimide (NHS) or 4-(4,6-Dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMTMM)) such that they are reactive (e.g., reactive to amine functional groups where EDC/NHS or DMTMM are used for activation). The activated COOH groups can then react with an appropriate species (e.g., a species comprising an amine functional group where the carboxylic acid groups are activated to be reactive with an amine functional group) comprising a moiety to be linked to the bead.

[00423] Beads comprising disulfide linkages in their polymeric network may be functionalized with additional species via reduction of some of the disulfide linkages to free thiols. The disulfide linkages may be reduced via, for example, the action of a reducing agent (e.g., DTT, TCEP, etc.) to generate free thiol groups, without dissolution of the bead. Free thiols of the beads can then react with free thiols of a species or a species comprising another disulfide bond (e.g., via thiol-disulfide exchange) such that the species can be linked to the beads (e.g., via a generated disulfide bond). In some cases, free thiols of the beads may react with any other suitable group. For example, free thiols of the beads may react with species comprising an acrydite moiety. The free thiol groups of the beads can react with the acrydite via Michael addition chemistry, such that the species comprising the acrydite is linked to the bead. In some cases, uncontrolled reactions can be prevented by inclusion of a thiol capping agent such as N-ethylmalieamide or iodoacetate.

[00424] Activation of disulfide linkages within a bead can be controlled such that only a small number of disulfide linkages are activated. Control may be exerted, for example, by controlling

the concentration of a reducing agent used to generate free thiol groups and/or concentration of reagents used to form disulfide bonds in bead polymerization. In some cases, a low concentration (e.g., molecules of reducing agent:gel bead ratios of less than or equal to about 1:100,000,000,000, less than or equal to about 1:10,000,000,000, less than or equal to about 1:1,000,000,000, less than or equal to about 1:100,000,000, less than or equal to about 1:10,000,000, less than or equal to about 1:1,000,000, less than or equal to about 1:100,000, less than or equal to about 1:10,000) of reducing agent may be used for reduction. Controlling the number of disulfide linkages that are reduced to free thiols may be useful in ensuring bead structural integrity during functionalization. In some cases, optically-active agents, such as fluorescent dyes may be coupled to beads via free thiol groups of the beads and used to quantify the number of free thiols present in a bead and/or track a bead.

[00425] In some cases, addition of moieties to a gel bead after gel bead formation may be advantageous. For example, addition of an oligonucleotide (e.g., barcoded oligonucleotide) after gel bead formation may avoid loss of the species during chain transfer termination that can occur during polymerization. Moreover, smaller precursors (e.g., monomers or cross linkers that do not comprise side chain groups and linked moieties) may be used for polymerization and can be minimally hindered from growing chain ends due to viscous effects. In some cases, functionalization after gel bead synthesis can minimize exposure of species (e.g., oligonucleotides) to be loaded with potentially damaging agents (e.g., free radicals) and/or chemical environments. In some cases, the generated gel may possess an upper critical solution temperature (UCST) that can permit temperature driven swelling and collapse of a bead. Such functionality may aid in oligonucleotide (e.g., a primer) infiltration into the bead during subsequent functionalization of the bead with the oligonucleotide. Post-production functionalization may also be useful in controlling loading ratios of species in beads, such that, for example, the variability in loading ratio is minimized. Species loading may also be performed in a batch process such that a plurality of beads can be functionalized with the species in a single batch.

[00426] A bead injected or otherwise introduced into a partition may comprise releasably, cleavably, or reversibly attached barcodes. A bead injected or otherwise introduced into a partition may comprise activatable barcodes. A bead injected or otherwise introduced into a partition may be degradable, disruptable, or dissolvable beads.

[00427] Barcodes can be releasably, cleavably or reversibly attached to the beads such that barcodes can be released or be releasable through cleavage of a linkage between the barcode molecule and the bead, or released through degradation of the underlying bead itself, allowing

the barcodes to be accessed or be accessible by other reagents, or both. In non-limiting examples, cleavage may be achieved through reduction of di-sulfide bonds, use of restriction enzymes, photo-activated cleavage, or cleavage via other types of stimuli (e.g., chemical, thermal, pH, enzymatic, etc.) and/or reactions, such as described elsewhere herein. Releasable barcodes may sometimes be referred to as being activatable, in that they are available for reaction once released. Thus, for example, an activatable barcode may be activated by releasing the barcode from a bead (or other suitable type of partition described herein). Other activatable configurations are also envisioned in the context of the described methods and systems.

[00428] In addition to, or as an alternative to the cleavable linkages between the beads and the associated molecules, such as barcode containing nucleic acid molecules (e.g., barcoded oligonucleotides), the beads may be degradable, disruptable, or dissolvable spontaneously or upon exposure to one or more stimuli (e.g., temperature changes, pH changes, exposure to particular chemical species or phase, exposure to light, reducing agent, etc.). In some cases, a bead may be dissolvable, such that material components of the beads are solubilized when exposed to a particular chemical species or an environmental change, such as a change temperature or a change in pH. In some cases, a gel bead can be degraded or dissolved at elevated temperature and/or in basic conditions. In some cases, a bead may be thermally degradable such that when the bead is exposed to an appropriate change in temperature (e.g., heat), the bead degrades. Degradation or dissolution of a bead bound to a species (e.g., a nucleic acid molecule, e.g., barcoded oligonucleotide) may result in release of the species from the bead.

[00429] As will be appreciated from the above disclosure, the degradation of a bead may refer to the disassociation of a bound or entrained species from a bead, both with and without structurally degrading the physical bead itself. For example, the degradation of the bead may involve cleavage of a cleavable linkage via one or more species and/or methods described elsewhere herein. In another example, entrained species may be released from beads through osmotic pressure differences due to, for example, changing chemical environments. By way of example, alteration of bead pore sizes due to osmotic pressure differences can generally occur without structural degradation of the bead itself. In some cases, an increase in pore size due to osmotic swelling of a bead can permit the release of entrained species within the bead. In other cases, osmotic shrinking of a bead may cause a bead to better retain an entrained species due to pore size contraction.

[00430] A degradable bead may be introduced into a partition, such as a droplet of an emulsion or a well, such that the bead degrades within the partition and any associated species (e.g., oligonucleotides) are released within the droplet when the appropriate stimulus is applied.

The free species (e.g., oligonucleotides, nucleic acid molecules) may interact with other reagents contained in the partition. For example, a polyacrylamide bead comprising cystamine and linked, via a disulfide bond, to a barcode sequence, may be combined with a reducing agent within a droplet of a water-in-oil emulsion. Within the droplet, the reducing agent can break the various disulfide bonds, resulting in bead degradation and release of the barcode sequence into the aqueous, inner environment of the droplet. In another example, heating of a droplet comprising a bead-bound barcode sequence in basic solution may also result in bead degradation and release of the attached barcode sequence into the aqueous, inner environment of the droplet.

[00431] Any suitable number of molecular tag molecules (e.g., primer, barcoded oligonucleotide) can be associated with a bead such that, upon release from the bead, the molecular tag molecules (e.g., primer, e.g., barcoded oligonucleotide) are present in the partition at a pre-defined concentration. Such pre-defined concentration may be selected to facilitate certain reactions for generating a sequencing library, e.g., amplification, within the partition. In some cases, the pre-defined concentration of the primer can be limited by the process of producing nucleic acid molecule (e.g., oligonucleotide) bearing beads.

[00432] In some cases, beads can be non-covalently loaded with one or more reagents. The beads can be non-covalently loaded by, for instance, subjecting the beads to conditions sufficient to swell the beads, allowing sufficient time for the reagents to diffuse into the interiors of the beads, and subjecting the beads to conditions sufficient to de-swell the beads. The swelling of the beads may be accomplished, for instance, by placing the beads in a thermodynamically favorable solvent, subjecting the beads to a higher or lower temperature, subjecting the beads to a higher or lower ion concentration, and/or subjecting the beads to an electric field. The swelling of the beads may be accomplished by various swelling methods. The de-swelling of the beads may be accomplished, for instance, by transferring the beads in a thermodynamically unfavorable solvent, subjecting the beads to lower or high temperatures, subjecting the beads to a lower or higher ion concentration, and/or removing an electric field. The de-swelling of the beads may be accomplished by various de-swelling methods. Transferring the beads may cause pores in the bead to shrink. The shrinking may then hinder reagents within the beads from diffusing out of the interiors of the beads. The hindrance may be due to steric interactions between the reagents and the interiors of the beads. The transfer may be accomplished microfluidically. For instance, the transfer may be achieved by moving the beads from one co-flowing solvent stream to a different co-flowing solvent stream. The swellability and/or pore size of the beads may be adjusted by changing the polymer composition of the bead.

[00433] In some cases, an acrydite moiety linked to a precursor, another species linked to a precursor, or a precursor itself can comprise a labile bond, such as chemically, thermally, or photo-sensitive bond e.g., disulfide bond, UV sensitive bond, or the like. Once acrydite moieties or other moieties comprising a labile bond are incorporated into a bead, the bead may also comprise the labile bond. The labile bond may be, for example, useful in reversibly linking (e.g., covalently linking) species (e.g., barcodes, primers, etc.) to a bead. In some cases, a thermally labile bond may include a nucleic acid hybridization based attachment, e.g., where an oligonucleotide is hybridized to a complementary sequence that is attached to the bead, such that thermal melting of the hybrid releases the oligonucleotide, e.g., a barcode containing sequence, from the bead or microcapsule.

[00434] The addition of multiple types of labile bonds to a gel bead may result in the generation of a bead capable of responding to varied stimuli. Each type of labile bond may be sensitive to an associated stimulus (e.g., chemical stimulus, light, temperature, enzymatic, etc.) such that release of species attached to a bead via each labile bond may be controlled by the application of the appropriate stimulus. Such functionality may be useful in controlled release of species from a gel bead. In some cases, another species comprising a labile bond may be linked to a gel bead after gel bead formation via, for example, an activated functional group of the gel bead as described above. As will be appreciated, barcodes that are releasably, cleavably or reversibly attached to the beads described herein include barcodes that are released or releasable through cleavage of a linkage between the barcode molecule and the bead, or that are released through degradation of the underlying bead itself, allowing the barcodes to be accessed or accessible by other reagents, or both.

[00435] In some cases, a species (e.g., oligonucleotide molecules comprising barcodes) that are attached to a solid support (e.g., a bead) may comprise a U-excising element that allows the species to release from the bead. In some cases, the U-excising element may comprise a single-stranded DNA (ssDNA) sequence that contains at least one uracil. The species may be attached to a solid support via the ssDNA sequence containing the at least one uracil. The species may be released by a combination of uracil-DNA glycosylase (e.g., to remove the uracil) and an endonuclease (e.g., to induce an ssDNA break). If the endonuclease generates a 5' phosphate group from the cleavage, then additional enzyme treatment may be included in downstream processing to eliminate the phosphate group, e.g., prior to ligation of additional sequencing handle elements, e.g., Illumina full P5 sequence, partial P5 sequence, full R1 sequence, and/or partial R1 sequence.

[00436] The barcodes that are releasable as described herein may sometimes be referred to as being activatable, in that they are available for reaction once released. Thus, for example, an activatable barcode may be activated by releasing the barcode from a bead (or other suitable type of partition described herein). Other activatable configurations are also envisioned in the context of the described methods and systems.

[00437] In addition to thermally cleavable bonds, disulfide bonds and UV sensitive bonds, other non-limiting examples of labile bonds that may be coupled to a precursor or bead include an ester linkage (e.g., cleavable with an acid, a base, or hydroxylamine), a vicinal diol linkage (e.g., cleavable via sodium periodate), a Diels-Alder linkage (e.g., cleavable via heat), a sulfone linkage (e.g., cleavable via a base), a silyl ether linkage (e.g., cleavable via an acid), a glycosidic linkage (e.g., cleavable via an amylase), a peptide linkage (e.g., cleavable via a protease), or a phosphodiester linkage (e.g., cleavable via a nuclease (e.g., DNAase)). A bond may be cleavable via other nucleic acid molecule targeting enzymes, such as restriction enzymes (e.g., restriction endonucleases), as described further below.

[00438] Species may be encapsulated in beads during bead generation (e.g., during polymerization of precursors). Such species may or may not participate in polymerization. Such species may be entered into polymerization reaction mixtures such that generated beads comprise the species upon bead formation. In some cases, such species may be added to the gel beads after formation. Such species may include, for example, nucleic acid molecules (e.g., oligonucleotides), reagents for a nucleic acid amplification reaction (e.g., primers, polymerases, dNTPs, co-factors (e.g., ionic co-factors), buffers) including those described herein, reagents for enzymatic reactions (e.g., enzymes, co-factors, substrates, buffers), reagents for nucleic acid modification reactions such as polymerization, ligation, or digestion, and/or reagents for template preparation (e.g., tagmentation) for one or more sequencing platforms (e.g., Nextera® for Illumina®). Such species may include one or more enzymes described herein, including without limitation, polymerase, reverse transcriptase, restriction enzymes (e.g., endonuclease), transposase, ligase, proteinase K, DNAse, etc. Such species may include one or more reagents described elsewhere herein (e.g., lysis agents, inhibitors, inactivating agents, chelating agents, stimulus). Trapping of such species may be controlled by the polymer network density generated during polymerization of precursors, control of ionic charge within the gel bead (e.g., via ionic species linked to polymerized species), or by the release of other species. Encapsulated species may be released from a bead upon bead degradation and/or by application of a stimulus capable of releasing the species from the bead. Alternatively or in addition, species may be partitioned in a partition (e.g., droplet) during or subsequent to partition formation. Such species

may include, without limitation, the abovementioned species that may also be encapsulated in a bead.

[00439] A degradable bead may comprise one or more species with a labile bond such that, when the bead/species is exposed to the appropriate stimuli, the bond is broken and the bead degrades. The labile bond may be a chemical bond (e.g., covalent bond, ionic bond) or may be another type of physical interaction (e.g., van der Waals interactions, dipole-dipole interactions, etc.). In some cases, a crosslinker used to generate a bead may comprise a labile bond. Upon exposure to the appropriate conditions, the labile bond can be broken and the bead degraded. For example, upon exposure of a polyacrylamide gel bead comprising cystamine crosslinkers to a reducing agent, the disulfide bonds of the cystamine can be broken and the bead degraded.

[00440] A degradable bead may be useful in more quickly releasing an attached species (e.g., a nucleic acid molecule, a barcode sequence, a primer, etc.) from the bead when the appropriate stimulus is applied to the bead as compared to a bead that does not degrade. For example, for a species bound to an inner surface of a porous bead or in the case of an encapsulated species, the species may have greater mobility and accessibility to other species in solution upon degradation of the bead. In some cases, a species may also be attached to a degradable bead via a degradable linker (e.g., disulfide linker). The degradable linker may respond to the same stimuli as the degradable bead or the two degradable species may respond to different stimuli. For example, a barcode sequence may be attached, via a disulfide bond, to a polyacrylamide bead comprising cystamine. Upon exposure of the barcoded-bead to a reducing agent, the bead degrades and the barcode sequence is released upon breakage of both the disulfide linkage between the barcode sequence and the bead and the disulfide linkages of the cystamine in the bead.

[00441] As will be appreciated from the above disclosure, while referred to as degradation of a bead, in many instances as noted above, that degradation may refer to the disassociation of a bound or entrained species from a bead, both with and without structurally degrading the physical bead itself. For example, entrained species may be released from beads through osmotic pressure differences due to, for example, changing chemical environments. By way of example, alteration of bead pore sizes due to osmotic pressure differences can generally occur without structural degradation of the bead itself. In some cases, an increase in pore size due to osmotic swelling of a bead can permit the release of entrained species within the bead. In other cases, osmotic shrinking of a bead may cause a bead to better retain an entrained species due to pore size contraction.

[00442] Where degradable beads are provided, it may be beneficial to avoid exposing such beads to the stimulus or stimuli that cause such degradation prior to a given time, in order to, for

example, avoid premature bead degradation and issues that arise from such degradation, including for example poor flow characteristics and aggregation. By way of example, where beads comprise reducible cross-linking groups, such as disulfide groups, it will be desirable to avoid contacting such beads with reducing agents, e.g., DTT or other disulfide cleaving reagents. In such cases, treatment to the beads described herein will, in some cases be provided free of reducing agents, such as DTT. Because reducing agents are often provided in commercial enzyme preparations, it may be desirable to provide reducing agent free (or DTT free) enzyme preparations in treating the beads described herein. Examples of such enzymes include, e.g., polymerase enzyme preparations, reverse transcriptase enzyme preparations, ligase enzyme preparations, as well as many other enzyme preparations that may be used to treat the beads described herein. The terms "reducing agent free" or "DTT free" preparations can refer to a preparation having less than about 1/10th, less than about 1/50th, or even less than about 1/100th of the lower ranges for such materials used in degrading the beads. For example, for DTT, the reducing agent free preparation can have less than about 0.01 millimolar (mM), 0.005 mM, 0.001 mM DTT, 0.0005 mM DTT, or even less than about 0.0001 mM DTT. In many cases, the amount of DTT can be undetectable.

[00443] Numerous chemical triggers may be used to trigger the degradation of beads. Examples of these chemical changes may comprise pH-mediated changes to the integrity of a component within the bead, degradation of a component of a bead via cleavage of cross-linked bonds, and depolymerization of a component of a bead.

[00444] In some embodiments, a bead may be formed from materials that comprise degradable chemical crosslinkers, such as BAC or cystamine. Degradation of such degradable crosslinkers may be accomplished through a number of mechanisms. In some examples, a bead may be contacted with a chemical degrading agent that may induce oxidation, reduction or other chemical changes. For example, a chemical degrading agent may be a reducing agent, such as dithiothreitol (DTT). Additional examples of reducing agents may include β-mercaptoethanol, (2S)-2-amino-1,4-dimercaptobutane (dithiobutylamine or DTBA), tris(2-carboxyethyl) phosphine (TCEP), or combinations thereof. A reducing agent may degrade the disulfide bonds formed between gel precursors forming the bead, and thus, degrade the bead. In other cases, a change in pH of a solution, such as an increase in pH, may trigger degradation of a bead. In other cases, exposure to an aqueous solution, such as water, may trigger hydrolytic degradation, and thus degradation of the bead. In some cases, any combination of stimuli may trigger degradation of a bead. For example, a change in pH may enable a chemical agent (e.g., DTT) to become an effective reducing agent.

[00445] Beads may also be induced to release their contents upon the application of a thermal stimulus. A change in temperature can cause a variety of changes to a bead. For example, heat can cause a solid bead to liquefy. A change in heat may cause melting of a bead such that a portion of the bead degrades. In other cases, heat may increase the internal pressure of the bead components such that the bead ruptures or explodes. Heat may also act upon heat-sensitive polymers used as materials to construct beads.

[00446] Any suitable agent may degrade beads. In some embodiments, changes in temperature or pH may be used to degrade thermo-sensitive or pH-sensitive bonds within beads. In some embodiments, chemical degrading agents may be used to degrade chemical bonds within beads by oxidation, reduction or other chemical changes. For example, a chemical degrading agent may be a reducing agent, such as DTT, wherein DTT may degrade the disulfide bonds formed between a crosslinker and gel precursors, thus degrading the bead. In some embodiments, a reducing agent may be added to degrade the bead, which may or may not cause the bead to release its contents. Examples of reducing agents may include dithiothreitol (DTT), β-mercaptoethanol, (2S)-2-amino-1,4-dimercaptobutane (dithiobutylamine or DTBA), tris(2-carboxyethyl) phosphine (TCEP), or combinations thereof. The reducing agent may be present at a concentration of about 0.1mM, 0.5mM, 1mM, 5mM, 10mM. The reducing agent may be present at a concentration of at least about 0.1mM, 0.5mM, 1mM, 5mM, 10mM, or greater than 10 mM. The reducing agent may be present at concentration of at most about 10mM, 5mM, 1mM, 0.5mM, 0.1mM, or less.

[00447] Any suitable number of molecular tag molecules (e.g., primer, barcoded oligonucleotide) can be associated with a bead such that, upon release from the bead, the molecular tag molecules (e.g., primer, e.g., barcoded oligonucleotide) are present in the partition at a pre-defined concentration. Such pre-defined concentration may be selected to facilitate certain reactions for generating a sequencing library, e.g., amplification, within the partition. In some cases, the pre-defined concentration of the primer can be limited by the process of producing oligonucleotide bearing beads.

[00448] In some examples, a partition of the plurality of partitions may comprise a single biological particle or analyte carrier (e.g., a single cell or a single nucleus of a cell). In some examples, a partition of the plurality of partitions may comprise multiple biological particles or analyte carriers. Such partitions may be referred to as multiply occupied partitions, and may comprise, for example, two, three, four or more cells and/or microcapsules (e.g., beads) comprising barcoded nucleic acid molecules (e.g., oligonucleotides) within a single partition. Accordingly, as noted above, the flow characteristics of the biological particle and/or bead

containing fluids and partitioning fluids may be controlled to provide for such multiply occupied partitions. In particular, the flow parameters may be controlled to provide a given occupancy rate at greater than about 50% of the partitions, greater than about 75%, and in some cases greater than about 80%, 90%, 95%, or higher.

[00449] In some cases, additional microcapsules can be used to deliver additional reagents to a partition. In such cases, it may be advantageous to introduce different beads into a common channel or droplet generation junction, from different bead sources (e.g., containing different associated reagents) through different channel inlets into such common channel or droplet generation junction. In such cases, the flow and frequency of the different beads into the channel or junction may be controlled to provide for a certain ratio of microcapsules from each source, while ensuring a given pairing or combination of such beads into a partition with a given number of biological particles (e.g., one biological particle and one bead per partition).

[00450] The partitions described herein may comprise small volumes, for example, less than about 10 microliters (μL), 5μL, 1μL, 900 picoliters (pL), 800 pL, 700 pL, 600 pL, 500 pL, 400pL, 300 pL, 200 pL, 100pL, 50 pL, 20 pL, 10 pL, 1 pL, 500 nanoliters (nL), 100 nL, 50 nL, or less.

[00451] For example, in the case of droplet based partitions, the droplets may have overall volumes that are less than about 1000 pL, 900 pL, 800 pL, 700 pL, 600 pL, 500 pL, 400pL, 300 pL, 200 pL, 100pL, 50 pL, 20 pL, 10 pL, 1 pL, or less. Where co-partitioned with microcapsules, it will be appreciated that the sample fluid volume, e.g., including co-partitioned biological particles and/or beads, within the partitions may be less than about 90% of the above described volumes, less than about 80%, less than about 70%, less than about 60%, less than about 50%, less than about 40%, less than about 30%, less than about 20%, or less than about 10% of the above described volumes.

[00452] As is described elsewhere herein, partitioning species may generate a population or plurality of partitions. In such cases, any suitable number of partitions can be generated or otherwise provided. For example, at least about 1,000 partitions, at least about 5,000 partitions, at least about 10,000 partitions, at least about 50,000 partitions, at least about 100,000 partitions, at least about 500,000 partitions, at least about 1,000,000 partitions, at least about 5,000,000 partitions at least about 10,000,000 partitions, at least about 50,000,000 partitions, at least about 100,000,000 partitions, at least about 500,000,000 partitions, at least about 1,000,000,000 partitions, or more partitions can be generated or otherwise provided. Moreover, the plurality of partitions may comprise both unoccupied partitions (e.g., empty partitions) and occupied partitions.

Flow sorting

[00453] A sample may derive from any useful source including any subject, such as a human subject. A sample may comprise material (e.g., one or more analyte carriers) from one or more different sources, such as one or more different subjects. Multiple samples, such as multiple samples from a single subject (e.g., multiple samples obtained in the same or different manners from the same or different bodily locations, and/or obtained at the same or different times (e.g., seconds, minutes, hours, days, weeks, months, or years apparat)), or multiple samples from different subjects, may be obtained for analysis as described herein. For example, a first sample may be obtained from a subject at a first time and a second sample may be obtained from the subject at a second time later than the first time. The first time may be before a subject undergoes a treatment regimen or procedure (e.g., to address a disease or condition), and the second time may be during or after the subject undergoes the treatment regimen or procedure. In another example, a first sample may be obtained from a first bodily location or system of a subject (e.g., using a first collection technique) and a second sample may be obtained from a second bodily location or system of the subject (e.g., using a second collection technique), which second bodily location or system may be different than the first bodily location or system. In another example, multiple samples may be obtained from a subject at a same time from the same or different bodily locations. Different samples, such as different subjects collected from different bodily locations of a same subject, at different times, from multiple different subjects, and/or using different collection techniques, may undergo the same or different processing (e.g., as described herein). For example, a first sample may undergo a first processing protocol and a second sample may undergo a second processing protocol.

[00454] A sample may be a biological sample, such as a cell sample (e.g., as described herein). A sample may include one or more analyte carriers, such as one or more cells and/or cellular constituents, such as one or more cell nuclei. For example, a sample may comprise a plurality of analyte carriers, such as a plurality of cells and/or cellular constituents. Analyte carriers (e.g., cells or cellular constituents, such as cell nuclei) of a sample may be of a single type or a plurality of different types. For example, cells of a sample may include one or more different types or blood cells.

[00455] Cells and cellular constituents of a sample may be of any type. For example, a cell or cellular constituent may be a mammalian, fungal, plant, bacterial, or other cell type. In some cases, the cell is a mammalian cell, such as a human cell. The cell may be, for example, a stem cell, liver cell, nerve cell, bone cell, blood cell, reproductive cell, skin cell, skeletal muscle cell, cardiac muscle cell, smooth muscle cell, hair cell, hormone-secreting cell, or glandular cell. The

cell may be, for example, an erythrocyte (e.g., red blood cell), a megakaryocyte (e.g., platelet precursor), a monocyte (e.g., white blood cell), a leukocyte, a B cell, a T cell (such as a helper, suppressor, cytotoxic, or natural killer T cell), an osteoclast, a dendritic cell, a connective tissue macrophage, an epidermal Langerhans cell, a microglial cell, a granulocyte, a hybridoma cell, a mast cell, a natural killer cell, a reticulocyte, a hematopoietic stem cell, a myoepithelial cell, a myeloid-derived suppressor cell, a platelet, a thymocyte, a satellite cell, an epithelial cell, an endothelial cell, an epididymal cell, a kidney cell, a liver cell, an adipocyte, a lipocyte, or a neuron cell. In some cases, the cell may be associated with a cancer, tumor, or neoplasm. In some cases, the cell may be associated with a fetus. In some cases, the cell may be a Jurkat cell.

[00456] A cell of a biological sample may have any feature or dimension. For example, a cell may have a first dimension, a second dimension, and a third dimension, where the first, second, and third dimensions are approximately the same. In other cases, the first and second dimensions may be approximately the same, and the third dimension may be different, or the first, second, and third dimensions may all be different. In some cases, a cell may comprise a dimension (e.g., a diameter) of at least about 1 μm. For example, a cell may comprise a dimension of at least about 1 micrometer (μm), 2 μm, 3 μm, 4 μm, 5 μm, 6 μm, 7 μm, 8 μm, 9 μm, 10 μm, 11 μm, 12 μm, 13 μm, 14 μm, 15 μm, 16 μm, 17 μm, 18 μm, 19 μm, 20 μm, 25 μm, 30 μm, 35 μm, 40 μm, 45 μm, 50 μm, 55 μm, 60 μm, 65 μm, 70 μm, 75 μm, 80 μm, 85 μm, 90 μm, 100 μm, 120 μm, 140 μm, 160 μm, 180 μm, 200 μm, 250 μm, 300 μm, 350 μm, 400 μm, 450 μm, 500 μm, 550 μm, 600 μm, 650 μm, 700 μm, 750 μm, 800 μm, 850 μm, 900 μm, 950 μm, 1 millimeter (mm), or greater. In some cases, the cell may comprise a dimension of between about 1 μm and 500 μm, such as between about 1 μm and 100 μm, between about 100 μm and 200 μm, between about 200 μm and 300 μm, between about 300 μm and 400 μm, or between about 400 μm and 500 μm. For example, a cell may comprise a dimension of between about 1 μm and 100 μm. Any or all dimensions of a cell may be variable. For example, the dimensions of a substantially fluid cell may vary over a rapid timescale. Dimensions of a more rigid cell may be fixed or may vary with lesser amplitude. Accordingly, the dimensions provided herein may represent averages rather than fixed values. The volume of a cell may be at least about 1 μm^3. In some cases, the volume of a cell may be at least about 10 μm^3. For example, the volume of the cell may be at least 1 μm^3, 2 μm^3, 3 μm^3, 4 μm^3, 5 μm^3, 6 μm^3, 7 μm^3, 8 μm^3, 9 μm^3, 10 μm^3, 12 μm^3, 14 μm^3, 16 μm^3, 18 μm^3, 20 μm^3, 25 μm^3, 30 μm^3, 35 μm^3, 40 μm^3, 45 μm^3, 50 μm^3, 55 μm^3, 60 μm^3, 65 μm^3, 70 μm^3, 75 μm^3, 80 μm^3, 85 μm^3, 90 μm^3, 95 μm^3, 100 μm^3, 125 μm^3, 150 μm^3, 175 μm^3, 200 μm^3, 250 μm^3, 300 μm^3, 350 μm^3, 400 μm^3, 450 μm^3, μm^3, 500 μm^3, 550 μm^3, 600 μm^3, 650 μm^3, 700 μm^3, 750 μm^3, 800 μm^3, 850 μm^3,

900 μm^3, 950 μm^3, 1000 μm^3, 1200 μm^3, 1400 μm^3, 1600 μm^3, 1800 μm^3, 2000 μm^3, 2200 μm^3, 2400 μm^3, 2600 μm^3, 2800 μm^3, 3000 μm^3, or greater. In some cases, a cell may comprise a volume of between about 1 μm^3 and 100 μm^3, such as between about 1 μm^3 and 10 μm^3, between about10 μm^3 and 50 μm^3, or between about 50 μm^3 and 100 μm^3. In some cases, a cell may comprise a volume of between about 100 μm^3 and 1000 μm^3, such as between about 100 μm^3 and 500 μm^3 or between about 500 μm^3 and 1000 μm^3. In some cases, a cell may comprise a volume between about 1000 μm^3 and 3000 μm^3, such as between about 1000 μm^3 and 2000 μm^3 or between about 2000 μm^3 and 3000 μm^3. In some cases, a cell may comprise a volume between about 1 μm^3 and 3000 μm^3, such as between about 1 μm^3 and 2000 μm^3, between about 1 μm^3 and 1000 μm^3, between about 1 μm^3 and 500 μm^3, or between about 1 μm^3 and 250 μm^3.

[00457] A cell of a biological sample may comprise one or more cross-sections that may be the same or different. In some cases, a cell may have a first cross-section that is different from a second cross-section. a cell may have a first cross-section that is at least about 1 μm. For example, a cell may comprise a cross-section (e.g., a first cross-section) of at least about 1 micrometer (μm), 2 μm, 3 μm, 4 μm, 5 μm, 6 μm, 7 μm, 8 μm, 9 μm, 10 μm, 11 μm, 12 μm, 13 μm, 14 μm, 15 μm, 16 μm, 17 μm, 18 μm, 19 μm, 20 μm, 25 μm, 30 μm, 35 μm, 40 μm, 45 μm, 50 μm, 55 μm, 60 μm, 65 μm, 70 μm, 75 μm, 80 μm, 85 μm, 90 μm, 100 μm, 120 μm, 140 μm, 160 μm, 180 μm, 200 μm, 250 μm, 300 μm, 350 μm, 400 μm, 450 μm, 500 μm, 550 μm, 600 μm, 650 μm, 700 μm, 750 μm, 800 μm, 850 μm, 900 μm, 950 μm, 1 millimeter (mm), or greater. In some cases, a cell may comprise a cross-section (e.g., a first cross-section) of between about 1 μm and 500 μm, such as between about 1 μm and 100 μm, between about 100 μm and 200 μm, between about 200 μm and 300 μm, between about 300 μm and 400 μm, or between about 400 μm and 500 μm. For example, a cell may comprise a cross-section (e.g., a first cross-section) of between about 1 μm and 100 μm. In some cases, the cell may have a second cross-section that is at least about 1 μm. For example, the cell may comprise a second cross-section of at least about 1 micrometer (μm), 2 μm, 3 μm, 4 μm, 5 μm, 6 μm, 7 μm, 8 μm, 9 μm, 10 μm, 11 μm, 12 μm, 13 μm, 14 μm, 15 μm, 16 μm, 17 μm, 18 μm, 19 μm, 20 μm, 25 μm, 30 μm, 35 μm, 40 μm, 45 μm, 50 μm, 55 μm, 60 μm, 65 μm, 70 μm, 75 μm, 80 μm, 85 μm, 90 μm, 100 μm, 120 μm, 140 μm, 160 μm, 180 μm, 200 μm, 250 μm, 300 μm, 350 μm, 400 μm, 450 μm, 500 μm, 550 μm, 600 μm, 650 μm, 700 μm, 750 μm, 800 μm, 850 μm, 900 μm, 950 μm, 1 millimeter (mm), or greater. In some cases, a cell may comprise a second cross-section of between about 1 μm and 500 μm, such as between about 1 μm and 100 μm, between about 100 μm and 200 μm, between about 200 μm and 300 μm, between about 300 μm and 400 μm, or between about 400 μm and 500 μm. For example, a cell may comprise a second cross-section of between about 1 μm and 100 μm.

[00458] A cross section (e.g., a first cross-section) may correspond to a diameter of a cell. In some cases, a cell may be approximately spherical. In such cases, the first cross-section may correspond to the diameter of the cell. In other cases, the cell may be approximately cylindrical. In such cases, the first cross-section may correspond to a diameter, length, or width along the approximately cylindrical cell. In some cases, the cell may comprise a surface. A cell surface may comprise one or more features. For example, a cell may comprise a dendritic receiver, flagella, roughed border, or other feature.

[00459] A characteristic or set of characteristics of a cell may be changed by one or more conditions. A condition suitable for changing a characteristic or set of characteristics of a cell may be, for example, a temperature, a pH, an ion or salt concentration, a pressure, or another condition. For example, a cell may be exposed to a chemical species that may bring about a change in one or more characteristics of the cell. In some cases, a stimulus may be used to change one or more characteristics of a cell. For example, upon application of the stimulus, one or more characteristics of a cell may be changed. The stimulus may be, for example, a thermal stimulus, a photo stimulus, a chemical stimulus, or another stimulus. In some cases, conditions sufficient to change the one or more characteristics of a cell may comprise one or more different conditions, such as a temperature and a pressure, a pH and a salt concentration, a chemical species and a temperature, or any other combination of conditions. A temperature sufficient for changing one or more characteristics of the cell may be, for example, at least about 0 degrees Celsius (°C), 1 °C, 2 °C, 3 °C, 4 °C, 5 °C, 10 °C, or higher. For example, the temperature may be about 4 °C. In other cases, a temperature sufficient for changing one or more characteristics of the cell may be, for example, at least about 25 °C, 30 °C, 35 °C, 37 °C, 40 °C, 45 °C, 50 °C, or higher. For example, the temperature may be about 37 °C. A pH sufficient for changing one or more characteristics of the cell may be, for example, between about 5 and 8, such as between about 6 and 7.

[00460] A biological sample may include a plurality of cells having different dimensions and features. In some cases, processing of the biological sample, such as cell separation and sorting (e.g., as described herein), may affect the distribution of dimensions and cellular features included in the sample by depleting cells having certain features and dimensions and/or isolating cells having certain features and dimensions.

[00461] A sample may undergo one or more processes in preparation for analysis (e.g., as described herein), including, but not limited to, filtration, selective precipitation, purification, centrifugation, permeabilization, isolation, agitation, heating, and/or other processes. For example, a sample may be filtered to remove a contaminant or other materials. In an example, a

filtration process may comprise the use of microfluidics (e.g., to separate analyte carriers of different sizes, types, charges, or other features).

[00462] In an example, a sample comprising one or more cells may be processed to separate the one or more cells from other materials in the sample (e.g., using centrifugation and/or another process). In some cases, cells and/or cellular constituents of a sample may be processed to separate and/or sort groups of cells and/or cellular constituents, such as to separate and/or sort cells and/or cellular constituents of different types. Examples of cell separation include, but are not limited to, separation of white blood cells or immune cells from other blood cells and components, separation of circulating tumor cells from blood, and separation of bacteria from bodily cells and/or environmental materials. A separation process may comprise a positive selection process (e.g., targeting of a cell type of interest for retention for subsequent downstream analysis, such as by use of a monoclonal antibody that targets a surface marker of the cell type of interest), a negative selection process (e.g., removal of one or more cell types and retention of one or more other cell types of interest), and/or a depletion process (e.g., removal of a single cell type from a sample, such as removal of red blood cells from peripheral blood mononuclear cells).

[00463] Separation of one or more different types of cells may comprise, for example, centrifugation, filtration, microfluidic-based sorting, flow cytometry, fluorescence-activated cell sorting (FACS), magnetic-activated cell sorting (MACS), buoyancy-activated cell sorting (BACS), or any other useful method. For example, a flow cytometry method may be used to detect cells and/or cellular constituents based on a parameter such as a size, morphology, or protein expression. Flow cytometry-based cell sorting may comprise injecting a sample into a sheath fluid that conveys the cells and/or cellular constituents of the sample into a measurement region one at a time. In the measurement region, a light source such as a laser may interrogate the cells and/or cellular constituents and scattered light and/or fluorescence may be detected and converted into digital signals. A nozzle system (e.g., a vibrating nozzle system) may be used to generate droplets (e.g., aqueous droplets) comprising individual cells and/or cellular constituents. Droplets including cells and/or cellular constituents of interest (e.g., as determined via optical detection) may be labeled with an electric charge (e.g., using an electrical charging ring), which charge may be used to separate such droplets from droplets including other cells and/or cellular constituents. For example, FACS may comprise labeling cells and/or cellular constituents with fluorescent markers (e.g., using internal and/or external biomarkers). Cells and/or cellular constituents may then be measured and identified one by one and sorted based on the emitted fluorescence of the marker or absence thereof. MACS may use micro- or nano-scale

magnetic particles to bind to cells and/or cellular constituents (e.g., via an antibody interaction with cell surface markers) to facilitate magnetic isolation of cells and/or cellular constituents of interest from other components of a sample (e.g., using a column-based analysis). BACS may use microbubbles (e.g., glass microbubbles) labeled with antibodies to target cells of interest. Cells and/or cellular components coupled to microbubbles may float to a surface of a solution, thereby separating target cells and/or cellular components from other components of a sample. Cell separation techniques may be used to enrich for populations of cells of interest (e.g., prior to partitioning, as described herein). For example, a sample comprising a plurality of cells including a plurality of cells of a given type may be subjected to a positive separation process. The plurality of cells of the given type may be labeled with a fluorescent marker (e.g., based on an expressed cell surface marker or another marker) and subjected to a FACS process to separate these cells from other cells of the plurality of cells. The selected cells may then be subjected to subsequent partition-based analysis (e.g., as described herein) or other downstream analysis. The fluorescent marker may be removed prior to such analysis or may be retained. The fluorescent marker may comprise an identifying feature, such as a nucleic acid barcode sequence and/or unique molecular identifier.

[00464] In another example, a first sample comprising a first plurality of cells including a first plurality of cells of a given type (e.g., immune cells expressing a particular marker or combination of markers) and a second sample comprising a second plurality of cells including a second plurality of cells of the given type may be subjected to a positive separation process. The first and second samples may be collected from the same or different subjects, at the same or different types, from the same or different bodily locations or systems, using the same or different collection techniques. For example, the first sample may be from a first subject and the second sample may be from a second subject different than the first subject. The first plurality of cells of the first sample may be provided a first plurality of fluorescent markers configured to label the first plurality of cells of the given type. The second plurality of cells of the second sample may be provided a second plurality of fluorescent markers configured to label the second plurality of cells of the given type. The first plurality of fluorescent markers may include a first identifying feature, such as a first barcode, while the second plurality of fluorescent markers may include a second identifying feature, such as a second barcode, that is different than the first identifying feature. The first plurality of fluorescent markers and the second plurality of fluorescent markers may fluoresce at the same intensities and over the same range of wavelengths upon excitation with a same excitation source (e.g., light source, such as a laser). The first and second samples may then be combined and subjected to a FACS process to

separate cells of the given type from other cells based on the first plurality of fluorescent markers labeling the first plurality of cells of the given type and the second plurality of fluorescent markers labeling the second plurality of cells of the given type. Alternatively, the first and second samples may undergo separate FACS processes and the positively selected cells of the given type from the first sample and the positively selected cells of the given type from the second sample may then be combined for subsequent analysis. The encoded identifying features of the different fluorescent markers may be used to identify cells originating from the first sample and cells originating from the second sample. For example, the first and second identifying features may be configured to interact (e.g., in partitions, as described herein) with nucleic acid barcode molecules (e.g., as described herein) to generate barcoded nucleic acid products detectable using, e.g., nucleic acid sequencing.

Multiplexing

[00465] The present disclosures provides methods and systems for multiplexing, and otherwise increasing throughput in, analysis. For example, a single or integrated process workflow may permit the processing, identification, and/or analysis of more or multiple analytes, more or multiple types of analytes, and/or more or multiple types of analyte characterizations. For example, in the methods and systems described herein, one or more labelling agents capable of binding to or otherwise coupling to one or more cell features may be used to characterize analyte carriers and/or cell features. In some instances, cell features include cell surface features. Cell surface features may include, but are not limited to, a receptor, an antigen, a surface protein, a transmembrane protein, a cluster of differentiation protein, a protein channel, a protein pump, a carrier protein, a phospholipid, a glycoprotein, a glycolipid, a cell-cell interaction protein complex, an antigen-presenting complex, a major histocompatibility complex, an engineered T-cell receptor, a T-cell receptor, a B-cell receptor, a chimeric antigen receptor, a gap junction, an adherens junction, or any combination thereof. In some instances, cell features may include intracellular analytes, such as proteins, protein modifications (e.g., phosphorylation status or other post-translational modifications), nuclear proteins, nuclear membrane proteins, or any combination thereof. A labelling agent may include, but is not limited to, a protein, a peptide, an antibody (or an epitope binding fragment thereof), a lipophilic moiety (such as cholesterol), a cell surface receptor binding molecule, a receptor ligand, a small molecule, a bi-specific antibody, a bi-specific T-cell engager, a T-cell receptor engager, a B-cell receptor engager, a pro-body, an aptamer, a monobody, an affimer, a darpin, and a protein scaffold, or any combination thereof. The labelling agents can include (e.g., are attached to) a reporter oligonucleotide that is indicative of the cell surface feature to which the binding group binds.

For example, the reporter oligonucleotide may comprise a barcode sequence that permits identification of the labelling agent. For example, a labelling agent that is specific to one type of cell feature (e.g., a first cell surface feature) may have a first reporter oligonucleotide coupled thereto, while a labelling agent that is specific to a different cell feature (e.g., a second cell surface feature) may have a different reporter oligonucleotide coupled thereto. For a description of exemplary labelling agents, reporter oligonucleotides, and methods of use, see, e.g., U.S. Pat. 10,550,429; U.S. Pat. Pub. 20190177800; and 20190367969, each of which is herein entirely incorporated by reference for all purposes.

[00466] In a particular example, a library of potential cell feature labelling agents may be provided, where the respective cell feature labelling agents are associated with nucleic acid reporter molecules, such that a different reporter oligonucleotide sequence is associated with each labelling agent capable of binding to a specific cell feature. In some aspects, different members of the library may be characterized by the presence of a different oligonucleotide sequence label. For example, an antibody capable of binding to a first protein may have associated with it a first reporter oligonucleotide sequence, while an antibody capable of binding to a second protein may have a different reporter oligonucleotide sequence associated with it. The presence of the particular oligonucleotide sequence may be indicative of the presence of a particular antibody or cell feature which may be recognized or bound by the particular antibody.

[00467] Labelling agents capable of binding to or otherwise coupling to one or more analyte carriers may be used to characterize an analyte carrier as belonging to a particular set of analyte carriers. For example, labeling agents may be used to label a sample of cells or a group of cells. In this way, a group of cells may be labeled as different from another group of cells. In an example, a first group of cells may originate from a first sample and a second group of cells may originate from a second sample. Labelling agents may allow the first group and second group to have a different labeling agent (or reporter oligonucleotide associated with the labeling agent). This may, for example, facilitate multiplexing, where cells of the first group and cells of the second group may be labeled separately and then pooled together for downstream analysis. The downstream detection of a label may indicate analytes as belonging to a particular group.

[00468] For example, a reporter oligonucleotide may be linked to an antibody or an epitope binding fragment thereof, and labeling an analyte carrier may comprise subjecting the antibody-linked barcode molecule or the epitope binding fragment-linked barcode molecule to conditions suitable for binding the antibody to a molecule present on a surface of the analyte carrier. The binding affinity between the antibody or the epitope binding fragment thereof and the molecule present on the surface may be within a desired range to ensure that the antibody or the epitope

binding fragment thereof remains bound to the molecule. For example, the binding affinity may be within a desired range to ensure that the antibody or the epitope binding fragment thereof remains bound to the molecule during various sample processing steps, such as partitioning and/or nucleic acid amplification or extension. A dissociation constant (Kd) between the antibody or an epitope binding fragment thereof and the molecule to which it binds may be less than about 100 μM, 90 μM, 80 μM, 70 μM, 60 μM, 50 μM, 40 μM, 30 μM, 20 μM, 10 μM, 9 μM, 8 μM, 7 μM, 6 μM, 5 μM, 4 μM, 3 μM, 2 μM, 1 μM, 900 nM, 800 nM, 700 nM, 600 nM, 500 nM, 400 nM, 300 nM, 200 nM, 100 nM, 90 nM, 80 nM, 70 nM, 60 nM, 50 nM, 40 nM, 30 nM, 20 nM, 10 nM, 9 nM, 8 nM, 7 nM, 6 nM, 5 nM, 4 nM, 3 nM, 2 nM, 1 nM, 900 pM, 800 pM, 700 pM, 600 pM, 500 pM, 400 pM, 300 pM, 200 pM, 100 pM, 90 pM, 80 pM, 70 pM, 60 pM, 50 pM, 40 pM, 30 pM, 20 pM, 10 pM, 9 pM, 8 pM, 7 pM, 6 pM, 5 pM, 4 pM, 3 pM, 2 pM, or 1 pM. For example, the dissociation constant may be less than about 10 μM.

[00469] In another example, a reporter oligonucleotide may be coupled to a cell-penetrating peptide (CPP), and labeling cells may comprise delivering the CPP coupled reporter oligonucleotide into an analyte carrier. Labeling analyte carriers may comprise delivering the CPP conjugated oligonucleotide into a cell and/or cell bead by the cell-penetrating peptide. A cell-penetrating peptide that can be used in the methods provided herein can comprise at least one non-functional cysteine residue, which may be either free or derivatized to form a disulfide link with an oligonucleotide that has been modified for such linkage. Non-limiting examples of cell-penetrating peptides that can be used in embodiments herein include penetratin, transportan, plsl, TAT(48-60), pVEC, MTS, and MAP. Cell-penetrating peptides useful in the methods provided herein can have the capability of inducing cell penetration for at least about 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% of cells of a cell population. The cell-penetrating peptide may be an arginine-rich peptide transporter. The cell-penetrating peptide may be Penetratin or the Tat peptide.

[00470] In another example, a reporter oligonucleotide may be coupled to a fluorophore or dye, and labeling cells may comprise subjecting the fluorophore-linked barcode molecule to conditions suitable for binding the fluorophore to the surface of the analyte carrier. In some instances, fluorophores can interact strongly with lipid bilayers and labeling analyte carriers may comprise subjecting the fluorophore-linked barcode molecule to conditions such that the fluorophore binds to or is inserted into a membrane of the analyte carrier. In some cases, the fluorophore is a water-soluble, organic fluorophore. In some instances, the fluorophore is Alexa 532 maleimide, tetramethylrhodamine-5-maleimide (TMR maleimide), BODIPY-TMR maleimide, Sulfo-Cy3 maleimide, Alexa 546 carboxylic acid/succinimidyl ester, Atto 550

maleimide, Cy3 carboxylic acid/succinimidyl ester, Cy3B carboxylic acid/succinimidyl ester, Atto 565 biotin, Sulforhodamine B, Alexa 594 maleimide, Texas Red maleimide, Alexa 633 maleimide, Abberior STAR 635P azide, Atto 647N maleimide, Atto 647 SE, or Sulfo-Cy5 maleimide. See, e.g., Hughes L D, et al. PLoS One. 2014 Feb. 4; 9(2):e87649, which is hereby incorporated by reference in its entirety for all purposes, for a description of organic fluorophores.

[00471] A reporter oligonucleotide may be coupled to a lipophilic molecule, and labeling analyte carriers may comprise delivering the nucleic acid barcode molecule to a membrane of the analyte carrier or a nuclear membrane by the lipophilic molecule. Lipophilic molecules can associate with and/or insert into lipid membranes such as cell membranes and nuclear membranes. In some cases, the insertion can be reversible. In some cases, the association between the lipophilic molecule and analyte carrier may be such that the analyte carrier retains the lipophilic molecule (e.g., and associated components, such as nucleic acid barcode molecules, thereof) during subsequent processing (e.g., partitioning, cell permeabilization, amplification, pooling, etc.). The reporter nucleotide may enter into the intracellular space and/or a cell nucleus.

[00472] A reporter oligonucleotide may be part of a nucleic acid molecule comprising any number of functional sequences, as described elsewhere herein, such as a target capture sequence, a random primer sequence, and the like, and coupled to another nucleic acid molecule that is, or is derived from, the analyte.

[00473] Prior to partitioning, the cells may be incubated with the library of labelling agents, that may be labelling agents to a broad panel of different cell features, e.g., receptors, proteins, etc., and which include their associated reporter oligonucleotides. Unbound labelling agents may be washed from the cells, and the cells may then be co-partitioned (e.g., into droplets or wells) along with partition-specific barcode oligonucleotides (e.g., attached to a support, such as a bead or gel bead) as described elsewhere herein. As a result, the partitions may include the cell or cells, as well as the bound labelling agents and their known, associated reporter oligonucleotides.

[00474] In other instances, e.g., to facilitate sample multiplexing, a labelling agent that is specific to a particular cell feature may have a first plurality of the labelling agent (e.g., an antibody or lipophilic moiety) coupled to a first reporter oligonucleotide and a second plurality of the labelling agent coupled to a second reporter oligonucleotide. For example, the first plurality of the labeling agent and second plurality of the labeling agent may interact with different cells, cell populations or samples, allowing a particular report oligonucleotide to indicate a particular cell population (or cell or sample) and cell feature. In this way, different

samples or groups can be independently processed and subsequently combined together for pooled analysis (e.g., partition-based barcoding as described elsewhere herein). See, e.g., U.S. Pat. Pub. 20190323088, which is hereby entirely incorporated by reference for all purposes.

[00475] As described elsewhere herein, libraries of labelling agents may be associated with a particular cell feature as well as be used to identify analytes as originating from a particular analyte carrier, population, or sample. The analyte carriers may be incubated with a plurality of libraries and a given analyte carrier may comprise multiple labelling agents. For example, a cell may comprise coupled thereto a lipophilic labeling agent and an antibody. The lipophilic labeling agent may indicate that the cell is a member of a particular cell sample, whereas the antibody may indicate that the cell comprises a particular analyte. In this manner, the reporter oligonucleotides and labelling agents may allow multi-analyte, multiplexed analyses to be performed.

[00476] In some instances, these reporter oligonucleotides may comprise nucleic acid barcode sequences that permit identification of the labelling agent which the reporter oligonucleotide is coupled to. The use of oligonucleotides as the reporter may provide advantages of being able to generate significant diversity in terms of sequence, while also being readily attachable to most biomolecules, e.g., antibodies, etc., as well as being readily detected, e.g., using sequencing or array technologies.

[00477] Attachment (coupling) of the reporter oligonucleotides to the labelling agents may be achieved through any of a variety of direct or indirect, covalent or non-covalent associations or attachments. For example, oligonucleotides may be covalently attached to a portion of a labelling agent (such a protein, e.g., an antibody or antibody fragment) using chemical conjugation techniques (e.g., Lightning-Link® antibody labelling kits available from Innova Biosciences), as well as other non-covalent attachment mechanisms, e.g., using biotinylated antibodies and oligonucleotides (or beads that include one or more biotinylated linker, coupled to oligonucleotides) with an avidin or streptavidin linker. Antibody and oligonucleotide biotinylation techniques are available. See, e.g., Fang, et al., "Fluoride-Cleavable Biotinylation Phosphoramidite for 5′-end-Labelling and Affinity Purification of Synthetic Oligonucleotides," Nucleic Acids Res. Jan. 15, 2003; 31(2):708-715, which is entirely incorporated herein by reference for all purposes. Likewise, protein and peptide biotinylation techniques have been developed and are readily available. See, e.g., U.S. Pat. No. 6,265,552, which is entirely incorporated herein by reference for all purposes. Furthermore, click reaction chemistry such as a Methyltetrazine-PEG5-NHS Ester reaction, a TCO-PEG4-NHS Ester reaction, or the like, may be used to couple reporter oligonucleotides to labelling agents. Commercially available kits,

such as those from Thunderlink and Abcam, and techniques common in the art may be used to couple reporter oligonucleotides to labelling agents as appropriate. In another example, a labelling agent is indirectly (e.g., via hybridization) coupled to a reporter oligonucleotide comprising a barcode sequence that identifies the label agent. For instance, the labelling agent may be directly coupled (e.g., covalently bound) to a hybridization oligonucleotide that comprises a sequence that hybridizes with a sequence of the reporter oligonucleotide. Hybridization of the hybridization oligonucleotide to the reporter oligonucleotide couples the labelling agent to the reporter oligonucleotide. In some embodiments, the reporter oligonucleotides are releasable from the labelling agent, such as upon application of a stimulus. For example, the reporter oligonucleotide may be attached to the labeling agent through a labile bond (e.g., chemically labile, photolabile, thermally labile, etc.) as generally described for releasing molecules from supports elsewhere herein. In some instances, the reporter oligonucleotides described herein may include one or more functional sequences that can be used in subsequent processing, such as an adapter sequence, a unique molecular identifier (UMI) sequence, a sequencer specific flow cell attachment sequence (such as an P5, P7, or partial P5 or P7 sequence), a primer or primer binding sequence, a sequencing primer or primer biding sequence (such as an R1, R2, or partial R1 or R2 sequence).

[00478] In some cases, the labelling agent can comprise a reporter oligonucleotide and a label. A label can be fluorophore, a radioisotope, a molecule capable of a colorimetric reaction, a magnetic particle, or any other suitable molecule or compound capable of detection. The label can be conjugated to a labelling agent (or reporter oligonucleotide) either directly or indirectly (e.g., the label can be conjugated to a molecule that can bind to the labelling agent or reporter oligonucleotide). In some cases, a label is conjugated to an oligonucleotide that is complementary to a sequence of the reporter oligonucleotide, and the oligonucleotide may be allowed to hybridize to the reporter oligonucleotide. **FIG. 11** describes exemplary labelling agents (**1110**, **1120**, **1130**) comprising reporter oligonucleotides (**1140**) attached thereto. Labelling agent **1110** (e.g., any of the labelling agents described herein) is attached (either directly, e.g., covalently attached, or indirectly) to reporter oligonucleotide **1140**. Reporter oligonucleotide **1140** may comprise barcode sequence **1142** that identifies labelling agent **1110**. Reporter oligonucleotide **1140** may also comprise one or more functional sequences that can be used in subsequent processing, such as an adapter sequence, a unique molecular identifier (UMI) sequence, a sequencer specific flow cell attachment sequence (such as an P5, P7, or partial P5 or P7 sequence), a primer or primer binding sequence, or a sequencing primer or primer biding sequence (such as an R1, R2, or partial R1 or R2 sequence).

[00479] Referring to **FIG. 11**, in some instances, reporter oligonucleotide **1140** conjugated to a labelling agent (e.g., **1110**, **1120**, **1130**) comprises a primer sequence **1141**, a barcode sequence that identifies the labelling agent (e.g., **1110**, **1120**, **1130**), and functional sequence **1143**. Functional sequence **1143** may be configured to hybridize to a complementary sequence, such as a complementary sequence present on a nucleic acid barcode molecule **1190** (not shown), such as those described elsewhere herein. In some instances, nucleic acid barcode molecule **1190** is attached to a support (e.g., a bead, such as a gel bead), such as those described elsewhere herein. For example, nucleic acid barcode molecule **1190** may be attached to the support via a releasable linkage (e.g., comprising a labile bond), such as those described elsewhere herein. In some instances, reporter oligonucleotide **1140** comprises one or more additional functional sequences, such as those described above.

[00480] In some instances, the labelling agent **1110** is a protein or polypeptide (e.g., an antigen or prospective antigen) comprising reporter oligonucleotide **1140**. Reporter oligonucleotide **1140** comprises barcode sequence **1142** that identifies polypeptide **1110** and can be used to infer the presence of an analyte, e.g., a binding partner of polypeptide **1110** (i.e., a molecule or compound to which polypeptide **1110** can bind). In some instances, the labelling agent **1110** is a lipophilic moiety (e.g., cholesterol) comprising reporter oligonucleotide **1140,** where the lipophilic moiety is selected such that labelling agent **1110** integrates into a membrane of a cell or nucleus. Reporter oligonucleotide **1140** comprises barcode sequence **1142** that identifies lipophilic moiety **1110** which in some instances is used to tag cells (e.g., groups of cells, cell samples, etc.) and may be used for multiplex analyses as described elsewhere herein. In some instances, the labelling agent is an antibody **1120** (or an epitope binding fragment thereof) comprising reporter oligonucleotide **1140**. Reporter oligonucleotide **1140** comprises barcode sequence **1142** that identifies antibody **1120** and can be used to infer the presence of, e.g., a target of antibody **1120** (i.e., a molecule or compound to which antibody **1120** binds). In other embodiments, labelling agent **1130** comprises an MHC molecule **1131** comprising peptide **1132** and reporter oligonucleotide **1140** that identifies peptide **1132**. In some instances, the MHC molecule is coupled to a support **1133**. In some instances, support **1133** may be a polypeptide, such as streptavidin, or a polysaccharide, such as dextran. In some instances, reporter oligonucleotide **1140** may be directly or indirectly coupled to MHC labelling agent **1130** in any suitable manner. For example, reporter oligonucleotide **1140** may be coupled to MHC molecule **1131**, support **1133**, or peptide **1132**. In some embodiments, labelling agent **1130** comprises a plurality of MHC molecules, (e.g. is an MHC multimer, which may be coupled to a support (e.g., **1133**)). There are many possible configurations of Class I and/or

Class II MHC multimers that can be utilized with the compositions, methods, and systems disclosed herein, e.g., MHC tetramers, MHC pentamers (MHC assembled via a coiled-coil domain, e.g., Pro5® MHC Class I Pentamers, (ProImmune, Ltd.), MHC octamers, MHC dodecamers, MHC decorated dextran molecules (e.g., MHC Dextramer® (Immudex)), etc. For a description of exemplary labelling agents, including antibody and MHC-based labelling agents, reporter oligonucleotides, and methods of use, see, e.g., U.S. Pat. 10,550,429 and U.S. Pat. Pub. 20190367969, each of which is herein entirely incorporated by reference for all purposes.

[00481] **FIG. 13** illustrates another example of a barcode carrying bead. In some embodiments, analysis of multiple analytes (e.g., RNA and one or more analytes using labelling agents described herein) may comprise nucleic acid barcode molecules as generally depicted in **FIG. 13**. In some embodiments, nucleic acid barcode molecules **1310** and **1320** are attached to support **1330** via a releasable linkage **1340** (e.g., comprising a labile bond) as described elsewhere herein. Nucleic acid barcode molecule **1310** may comprise adapter sequence **1311**, barcode sequence **1312** and adapter sequence **1313**. Nucleic acid barcode molecule **1320** may comprise adapter sequence **1321**, barcode sequence **1312**, and adapter sequence **1323**, wherein adapter sequence **1323** comprises a different sequence than adapter sequence **1313**. In some instances, adapter **1311** and adapter **1321** comprise the same sequence. In some instances, adapter **1311** and adapter **1321** comprise different sequences. Although support **1330** is shown comprising nucleic acid barcode molecules **1310** and **1320**, any suitable number of barcode molecules comprising common barcode sequence **1312** are contemplated herein. For example, in some embodiments, support **1330** further comprises nucleic acid barcode molecule **1350**. Nucleic acid barcode molecule **1350** may comprise adapter sequence **1351**, barcode sequence **1312** and adapter sequence **1353**, wherein adapter sequence **1353** comprises a different sequence than adapter sequence **1313** and **1323**. In some instances, nucleic acid barcode molecules (e.g., **1310**, **1320**, **1350)** comprise one or more additional functional sequences, such as a UMI or other sequences described herein. The nucleic acid barcode molecules **1310**, **1320** or **1350** may interact with analytes as described elsewhere herein, for example, as depicted in **FIGs. 12A-C**.

[00482] Referring to **FIG. 12A**, in an instance where cells are labelled with labeling agents, sequence **1223** may be complementary to an adapter sequence of a reporter oligonucleotide. Cells may be contacted with one or more reporter oligonucleotide **1220** conjugated labelling agents **1210** (e.g., polypeptide, antibody, or others described elsewhere herein). In some cases, the cells may be further processed prior to barcoding. For example, such processing steps may include one or more washing and/or cell sorting steps. In some instances, a cell that is bound to

labelling agent **1210** which is conjugated to oligonucleotide **1220** and support **1230** (e.g., a bead, such as a gel bead) comprising nucleic acid barcode molecule **1290** is partitioned into a partition amongst a plurality of partitions (e.g., a droplet of a droplet emulsion or a well of a microwell array). In some instances, the partition comprises at most a single cell bound to labelling agent **1210**. In some instances, reporter oligonucleotide **1220** conjugated to labelling agent **1210** (e.g., polypeptide, an antibody, pMHC molecule such as an MHC multimer, etc.) comprises a first adapter sequence **1211** (e.g., a primer sequence), a barcode sequence **1212** that identifies the labelling agent **1210** (e.g., the polypeptide, antibody, or peptide of a pMHC molecule or complex), and an adapter sequence **1213**. Adapter sequence **1213** may be configured to hybridize to a complementary sequence, such as sequence **1223** present on a nucleic acid barcode molecule **1290**. In some instances, oligonucleotide **1220** comprises one or more additional functional sequences, such as those described elsewhere herein.

[00483] Barcoded nucleic may be generated (e.g., via a nucleic acid reaction, such as nucleic acid extension or ligation) from the constructs described in **FIGs. 12A-C**. For example, sequence **1213** may then be hybridized to complementary sequence **1223** to generate (e.g., via a nucleic acid reaction, such as nucleic acid extension or ligation) a barcoded nucleic acid molecule comprising cell (e.g., partition specific) barcode sequence **1222** (or a reverse complement thereof) and reporter barcode sequence **1212** (or a reverse complement thereof). Barcoded nucleic acid molecules can then be optionally processed as described elsewhere herein, e.g., to amplify the molecules and/or append sequencing platform specific sequences to the fragments. See, e.g., U.S. Pat. Pub. 2018/0105808, which is hereby entirely incorporated by reference for all purposes. Barcoded nucleic acid molecules, or derivatives generated therefrom, can then be sequenced on a suitable sequencing platform.

[00484] In some instances, analysis of multiple analytes (e.g., nucleic acids and one or more analytes using labelling agents described herein) may be performed. For example, the workflow may comprise a workflow as generally depicted in any of **FIGs. 12A-C**, or a combination of workflows for an individual analyte, as described elsewhere herein. For example, by using a combination of the workflows as generally depicted in **FIGs. 12A-C,** multiple analytes can be analyzed.

[00485] In some instances, analysis of an analyte (e.g. a nucleic acid, a polypeptides, a carbohydrate, a lipid, etc.) comprises a workflow as generally depicted in **FIGs. 12A**. A nucleic acid barcode molecule **1290** may be co-partitioned with the one or more analytes. In some instances, nucleic acid barcode molecule **1290** is attached to a support **1230** (e.g., a bead, such as a gel bead), such as those described elsewhere herein. For example, nucleic acid barcode

molecule **1290** may be attached to support **1230** via a releasable linkage **1240** (e.g., comprising a labile bond), such as those described elsewhere herein. Nucleic acid barcode molecule **1290** may comprise a barcode sequence **1221** and optionally comprise other additional sequences, for example, a UMI sequence **1222** (or other functional sequences described elsewhere herein). The nucleic acid barcode molecule **1290** may comprise a sequence **1223** that may be complementary to another nucleic acid sequence, such that it may hybridize to a particular sequence.

[00486] For example, sequence **1223** may comprise a poly-T sequence and may be used to hybridize to mRNA. Referring to **FIG. 12C**, in some embodiments, nucleic acid barcode molecule **1290** comprises sequence **1223** complementary to a sequence of RNA molecule **1260** from a cell. In some instances, sequence **1223** comprises a sequence specific for an RNA molecule. Sequence **1223** may comprise a known or targeted sequence or a random sequence. In some instances, a nucleic acid extension reaction may be performed, thereby generating a barcoded nucleic acid product comprising sequence **1223**, the barcode sequence **1221**, UMI sequence **1222**, any other functional sequence, and a sequence corresponding to the RNA molecule **1260**.

[00487] In another example, sequence **1223** may be complementary to an overhang sequence or an adapter sequence that has been appended to an analyte. For example, referring to **FIG. 12B**, panel **1201**, in some embodiments, primer **1250** comprises a sequence complementary to a sequence of nucleic acid molecule **1260** (such as an RNA encoding for a BCR sequence) from an analyte carrier. In some instances, primer **1250** comprises one or more sequences **1251** that are not complementary to RNA molecule **1260**. Sequence **1251** may be a functional sequence as described elsewhere herein, for example, an adapter sequence, a sequencing primer sequence, or a sequence the facilitates coupling to a flow cell of a sequencer. In some instances, primer **1250** comprises a poly-T sequence. In some instances, primer **1250** comprises a sequence complementary to a target sequence in an RNA molecule. In some instances, primer **1250** comprises a sequence complementary to a region of an immune molecule, such as the constant region of a TCR or BCR sequence. Primer **1250** is hybridized to nucleic acid molecule **1260** and complementary molecule **1270** is generated (*see* Panel **1202**). For example, complementary molecule **1270** may be cDNA generated in a reverse transcription reaction. In some instances, an additional sequence may be appended to complementary molecule **1270.** For example, the reverse transcriptase enzyme may be selected such that several non-templated bases **1280** (e.g., a poly-C sequence) are appended to the cDNA. In another example, a terminal transferase may also be used to append the additional sequence. Nucleic acid barcode molecule **1290** comprises a sequence **1224** complementary to the non-templated bases, and the reverse transcriptase

performs a template switching reaction onto nucleic acid barcode molecule **1290** to generate a barcoded nucleic acid molecule comprising cell (e.g., partition specific) barcode sequence **1222** (or a reverse complement thereof) and a sequence of complementary molecule **1270** (or a portion thereof). In some instances, sequence **1223** comprises a sequence complementary to a region of an immune molecule, such as the constant region of a TCR or BCR sequence. Sequence **1223** is hybridized to nucleic acid molecule **1260** and a complementary molecule **1270** is generated. For example complementary molecule **1270** may be generated in a reverse transcription reaction generating a barcoded nucleic acid molecule comprising cell (e.g., partition specific) barcode sequence **1222** (or a reverse complement thereof) and a sequence of complementary molecule **1270** (or a portion thereof). Additional methods and compositions suitable for barcoding cDNA generated from mRNA transcripts including those encoding V(D)J regions of an immune cell receptor and/or barcoding methods and composition including a template switch oligonucleotide are described in International Patent Application WO2018/075693, U.S. Patent Publication No. 2018/0105808, U.S. Patent Publication No. 2015/0376609, filed June 26, 2015, and U.S. Patent Publication No. 2019/0367969, , each of which applications is herein entirely incorporated by reference for all purposes.

Reagents

[00488] In accordance with certain aspects, biological particles may be partitioned along with lysis reagents in order to release the contents of the biological particles within the partition. In such cases, the lysis agents can be contacted with the biological particle suspension concurrently with, or immediately prior to, the introduction of the biological particles into the partitioning junction/droplet generation zone (e.g., junction **210**), such as through an additional channel or channels upstream of the channel junction. In accordance with other aspects, additionally or alternatively, biological particles may be partitioned along with other reagents, as will be described further below.

[00489] The methods and systems of the present disclosure may comprise microfluidic devices and methods of use thereof, which may be used for co-partitioning analyte carriers or analyte carriers with reagents. Such systems and methods are described in U.S. Patent Publication No. US/20190367997, which is herein incorporated by reference in its entirety for all purposes.

[00490] Beneficially, when lysis reagents and biological particles are co-partitioned, the lysis reagents can facilitate the release of the contents of the biological particles within the partition. The contents released in a partition may remain discrete from the contents of other partitions.

[00491] As will be appreciated, the channel segments of the microfluidic devices described elsewhere herein may be coupled to any of a variety of different fluid sources or receiving components, including reservoirs, tubing, manifolds, or fluidic components of other systems. As will be appreciated, the microfluidic channel structures may have various geometries and/or configurations. For example, a microfluidic channel structure can have more than two channel junctions. For example, a microfluidic channel structure can have 2, 3, 4, 5 channel segments or more each carrying the same or different types of beads, reagents, and/or biological particles that meet at a channel junction. Fluid flow in each channel segment may be controlled to control the partitioning of the different elements into droplets. Fluid may be directed flow along one or more channels or reservoirs via one or more fluid flow units. A fluid flow unit can comprise compressors (e.g., providing positive pressure), pumps (e.g., providing negative pressure), actuators, and the like to control flow of the fluid. Fluid may also or otherwise be controlled via applied pressure differentials, centrifugal force, electrokinetic pumping, vacuum, capillary or gravity flow, or the like.

[00492] Examples of lysis agents include bioactive reagents, such as lysis enzymes that are used for lysis of different cell types, e.g., gram positive or negative bacteria, plants, yeast, mammalian, etc., such as lysozymes, achromopeptidase, lysostaphin, labiase, kitalase, lyticase, and a variety of other lysis enzymes available from, e.g., Sigma-Aldrich, Inc. (St Louis, MO), as well as other commercially available lysis enzymes. Other lysis agents may additionally or alternatively be co-partitioned with the biological particles to cause the release of the biological particle's contents into the partitions. For example, in some cases, surfactant-based lysis solutions may be used to lyse cells, although these may be less desirable for emulsion based systems where the surfactants can interfere with stable emulsions. In some cases, lysis solutions may include non-ionic surfactants such as, for example, TritonX-100 and Tween 20. In some cases, lysis solutions may include ionic surfactants such as, for example, sarcosyl and sodium dodecyl sulfate (SDS). Electroporation, thermal, acoustic or mechanical cellular disruption may also be used in certain cases, e.g., non-emulsion based partitioning such as encapsulation of biological particles that may be in addition to or in place of droplet partitioning, where any pore size of the encapsulate is sufficiently small to retain nucleic acid fragments of a given size, following cellular disruption.

[00493] Alternatively or in addition to the lysis agents co-partitioned with the analyte carriers described above, other reagents can also be co-partitioned with the analyte carriers, including, for example, DNase and RNase inactivating agents or inhibitors, such as proteinase K, chelating agents, such as EDTA, and other reagents employed in removing or otherwise reducing negative

activity or impact of different cell lysate components on subsequent processing of nucleic acids. In addition, in the case of encapsulated analyte carriers (e.g., a cell or a nucleus in a polymer matrix), the analyte carriers may be exposed to an appropriate stimulus to release the analyte carriers or their contents from a co-partitioned microcapsule. For example, in some cases, a chemical stimulus may be co-partitioned along with an encapsulated analyte carrier to allow for the degradation of the microcapsule and release of the cell or its contents into the larger partition. In some cases, this stimulus may be the same as the stimulus described elsewhere herein for release of nucleic acid molecules (e.g., oligonucleotides) from their respective microcapsule (e.g., bead). In alternative examples, this may be a different and non-overlapping stimulus, in order to allow an encapsulated analyte carrier to be released into a partition at a different time from the release of nucleic acid molecules into the same partition. For a description of methods, compositions, and systems for encapsulating cells (also referred to as a "cell bead"), see, e.g., U.S. Pat. 10,428,326 and U.S. Pat. Pub. 20190100632, which are each incorporated by reference in their entirety.

[00494] Additional reagents may also be co-partitioned with the biological particles, such as endonucleases to fragment a biological particle's DNA, DNA polymerase enzymes and dNTPs used to amplify the biological particle's nucleic acid fragments and to attach the barcode molecular tags to the amplified fragments. Other enzymes may be co-partitioned, including without limitation, polymerase, transposase, ligase, proteinase K, DNAse, etc. Additional reagents may also include reverse transcriptase enzymes, including enzymes with terminal transferase activity, primers and oligonucleotides, and switch oligonucleotides (also referred to herein as "switch oligos" or "template switching oligonucleotides") which can be used for template switching. In some cases, template switching can be used to increase the length of a cDNA. In some cases, template switching can be used to append a predefined nucleic acid sequence to the cDNA. In an example of template switching, cDNA can be generated from reverse transcription of a template, e.g., cellular mRNA, where a reverse transcriptase with terminal transferase activity can add additional nucleotides, e.g., polyC, to the cDNA in a template independent manner. Switch oligos can include sequences complementary to the additional nucleotides, e.g., polyG. The additional nucleotides (e.g., polyC) on the cDNA can hybridize to the additional nucleotides (e.g., polyG) on the switch oligo, whereby the switch oligo can be used by the reverse transcriptase as template to further extend the cDNA. Template switching oligonucleotides may comprise a hybridization region and a template region. The hybridization region can comprise any sequence capable of hybridizing to the target. In some cases, as previously described, the hybridization region comprises a series of G bases to

complement the overhanging C bases at the 3' end of a cDNA molecule. The series of G bases may comprise 1 G base, 2 G bases, 3 G bases, 4 G bases, 5 G bases or more than 5 G bases. The template sequence can comprise any sequence to be incorporated into the cDNA. In some cases, the template region comprises at least 1 (e.g., at least 2, 3, 4, 5 or more) tag sequences and/or functional sequences. Switch oligos may comprise deoxyribonucleic acids; ribonucleic acids; modified nucleic acids including 2-Aminopurine, 2,6-Diaminopurine (2-Amino-dA), inverted dT, 5-Methyl dC, 2'-deoxyInosine, Super T (5-hydroxybutynl-2'-deoxyuridine), Super G (8-aza-7-deazaguanosine), locked nucleic acids (LNAs), unlocked nucleic acids (UNAs, e.g., UNA-A, UNA-U, UNA-C, UNA-G), Iso-dG, Iso-dC, 2' Fluoro bases (e.g., Fluoro C, Fluoro U, Fluoro A, and Fluoro G), or any combination.

[00495] In some cases, the length of a switch oligo may be at least about 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, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197 , 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249 or 250 nucleotides or longer.

[00496] In some cases, the length of a switch oligo may be at most about 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, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197 , 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223,

224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249 or 250 nucleotides.

[00497] Once the contents of the cells are released into their respective partitions, the macromolecular components (e.g., macromolecular constituents of biological particles, such as RNA, DNA, or proteins) contained therein may be further processed within the partitions. In accordance with the methods and systems described herein, the macromolecular component contents of individual biological particles can be provided with unique identifiers such that, upon characterization of those macromolecular components they may be attributed as having been derived from the same biological particle or particles. The ability to attribute characteristics to individual biological particles or groups of biological particles is provided by the assignment of unique identifiers specifically to an individual biological particle or groups of biological particles. Unique identifiers, e.g., in the form of nucleic acid barcodes can be assigned or associated with individual biological particles or populations of biological particles, in order to tag or label the biological particle's macromolecular components (and as a result, its characteristics) with the unique identifiers. These unique identifiers can then be used to attribute the biological particle's components and characteristics to an individual biological particle or group of biological particles.

[00498] In some aspects, this is performed by co-partitioning the individual biological particle or groups of biological particles with the unique identifiers, such as described above (with reference to **FIG. 2**). In some aspects, the unique identifiers are provided in the form of nucleic acid molecules (e.g., oligonucleotides) that comprise nucleic acid barcode sequences that may be attached to or otherwise associated with the nucleic acid contents of individual biological particle, or to other components of the biological particle, and particularly to fragments of those nucleic acids. The nucleic acid molecules are partitioned such that as between nucleic acid molecules in a given partition, the nucleic acid barcode sequences contained therein are the same, but as between different partitions, the nucleic acid molecule can, and do have differing barcode sequences, or at least represent a large number of different barcode sequences across all of the partitions in a given analysis. In some aspects, only one nucleic acid barcode sequence can be associated with a given partition, although in some cases, two or more different barcode sequences may be present.

[00499] The nucleic acid barcode sequences can include from about 6 to about 20 or more nucleotides within the sequence of the nucleic acid molecules (e.g., oligonucleotides). The nucleic acid barcode sequences can include from about 6 to about 20, 30, 40, 50, 60, 70, 80, 90, 100 or more nucleotides. In some cases, the length of a barcode sequence may be about 6, 7, 8,

9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 nucleotides or longer. In some cases, the length of a barcode sequence may be at least about 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 nucleotides or longer. In some cases, the length of a barcode sequence may be at most about 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 nucleotides or shorter. These nucleotides may be completely contiguous, i.e., in a single stretch of adjacent nucleotides, or they may be separated into two or more separate subsequences that are separated by 1 or more nucleotides. In some cases, separated barcode subsequences can be from about 4 to about 16 nucleotides in length. In some cases, the barcode subsequence may be about 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 nucleotides or longer. In some cases, the barcode subsequence may be at least about 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 nucleotides or longer. In some cases, the barcode subsequence may be at most about 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 nucleotides or shorter.

[00500] The co-partitioned nucleic acid molecules can also comprise other functional sequences useful in the processing of the nucleic acids from the co-partitioned biological particles. These sequences include, e.g., targeted or random/universal amplification primer sequences for amplifying nucleic acids (e.g., mRNA, the genomic DNA) from the individual biological particles within the partitions while attaching the associated barcode sequences, sequencing primers or primer recognition sites, hybridization or probing sequences, e.g., for identification of presence of the sequences or for pulling down barcoded nucleic acids, or any of a number of other potential functional sequences. Other mechanisms of co-partitioning oligonucleotides may also be employed, including, e.g., coalescence of two or more droplets, where one droplet contains oligonucleotides, or microdispensing of oligonucleotides (e.g., attached to a bead) into partitions, e.g., droplets within microfluidic systems.

[00501] In an example, microcapsules, such as beads, are provided that each include large numbers of the above described barcoded nucleic acid molecules (e.g., barcoded oligonucleotides) releasably attached to the beads, where all of the nucleic acid molecules attached to a particular bead will include the same nucleic acid barcode sequence, but where a large number of diverse barcode sequences are represented across the population of beads used. In some embodiments, hydrogel beads, e.g., comprising polyacrylamide polymer matrices, are used as a solid support and delivery vehicle for the nucleic acid molecules into the partitions, as they are capable of carrying large numbers of nucleic acid molecules, and may be configured to release those nucleic acid molecules upon exposure to a particular stimulus, as described elsewhere herein. In some cases, the population of beads provides a diverse barcode sequence library that includes at least about 1,000 different barcode sequences, at least about 5,000 different barcode sequences, at least about 10,000 different barcode sequences, at least about

50,000 different barcode sequences, at least about 100,000 different barcode sequences, at least about 1,000,000 different barcode sequences, at least about 5,000,000 different barcode sequences, or at least about 10,000,000 different barcode sequences, or more. Additionally, each bead can be provided with large numbers of nucleic acid (e.g., oligonucleotide) molecules attached. In particular, the number of molecules of nucleic acid molecules including the barcode sequence on an individual bead can be at least about 1,000 nucleic acid molecules, at least about 5,000 nucleic acid molecules, at least about 10,000 nucleic acid molecules, at least about 50,000 nucleic acid molecules, at least about 100,000 nucleic acid molecules, at least about 500,000 nucleic acids, at least about 1,000,000 nucleic acid molecules, at least about 5,000,000 nucleic acid molecules, at least about 10,000,000 nucleic acid molecules, at least about 50,000,000 nucleic acid molecules, at least about 100,000,000 nucleic acid molecules, at least about 250,000,000 nucleic acid molecules and in some cases at least about 1 billion nucleic acid molecules, or more. Nucleic acid molecules of a given bead can include identical (or common) barcode sequences, different barcode sequences, or a combination of both. Nucleic acid molecules of a given bead can include multiple sets of nucleic acid molecules. Nucleic acid molecules of a given set can include identical barcode sequences. The identical barcode sequences can be different from barcode sequences of nucleic acid molecules of another set.

[00502] Moreover, when the population of beads is partitioned, the resulting population of partitions can also include a diverse barcode library that includes at least about 1,000 different barcode sequences, at least about 5,000 different barcode sequences, at least about 10,000 different barcode sequences, at least at least about 50,000 different barcode sequences, at least about 100,000 different barcode sequences, at least about 1,000,000 different barcode sequences, at least about 5,000,000 different barcode sequences, or at least about 10,000,000 different barcode sequences. Additionally, each partition of the population can include at least about 1,000 nucleic acid molecules, at least about 5,000 nucleic acid molecules, at least about 10,000 nucleic acid molecules, at least about 50,000 nucleic acid molecules, at least about 100,000 nucleic acid molecules, at least about 500,000 nucleic acids, at least about 1,000,000 nucleic acid molecules, at least about 5,000,000 nucleic acid molecules, at least about 10,000,000 nucleic acid molecules, at least about 50,000,000 nucleic acid molecules, at least about 100,000,000 nucleic acid molecules, at least about 250,000,000 nucleic acid molecules and in some cases at least about 1 billion nucleic acid molecules.

[00503] In some cases, it may be desirable to incorporate multiple different barcodes within a given partition, either attached to a single or multiple beads within the partition. For example, in some cases, a mixed, but known set of barcode sequences may provide greater assurance of

identification in the subsequent processing, e.g., by providing a stronger address or attribution of the barcodes to a given partition, as a duplicate or independent confirmation of the output from a given partition.

[00504] The nucleic acid molecules (e.g., oligonucleotides) are releasable from the beads upon the application of a particular stimulus to the beads. In some cases, the stimulus may be a photo-stimulus, e.g., through cleavage of a photo-labile linkage that releases the nucleic acid molecules. In other cases, a thermal stimulus may be used, where elevation of the temperature of the beads environment will result in cleavage of a linkage or other release of the nucleic acid molecules from the beads. In still other cases, a chemical stimulus can be used that cleaves a linkage of the nucleic acid molecules to the beads, or otherwise results in release of the nucleic acid molecules from the beads. In one case, such compositions include the polyacrylamide matrices described above for encapsulation of biological particles, and may be degraded for release of the attached nucleic acid molecules through exposure to a reducing agent, such as DTT.

[00505] In some aspects, provided are systems and methods for controlled partitioning. Droplet size may be controlled by adjusting certain geometric features in channel architecture (e.g., microfluidics channel architecture). For example, an expansion angle, width, and/or length of a channel may be adjusted to control droplet size.

[00506] **FIG. 2** shows an example of a microfluidic channel structure for the controlled partitioning of beads into discrete droplets. A channel structure **200** can include a channel segment **202** communicating at a channel junction **206** (or intersection) with a reservoir **204**. The reservoir **204** can be a chamber. Any reference to "reservoir," as used herein, can also refer to a "chamber." In operation, an aqueous fluid **208** that includes suspended beads **212** may be transported along the channel segment **202** into the junction **206** to meet a second fluid **210** that is immiscible with the aqueous fluid **208** in the reservoir **204** to create droplets **216, 218** of the aqueous fluid **208** flowing into the reservoir **204**. At the junction **206** where the aqueous fluid **208** and the second fluid **210** meet, droplets can form based on factors such as the hydrodynamic forces at the junction **206**, flow rates of the two fluids **208, 210**, fluid properties, and certain geometric parameters (e.g., w, h_0, α, etc.) of the channel structure **200**. A plurality of droplets can be collected in the reservoir **204** by continuously injecting the aqueous fluid **208** from the channel segment **202** through the junction **206**.

[00507] A discrete droplet generated may include a bead (e.g., as in occupied droplets **216**). Alternatively, a discrete droplet generated may include more than one bead. Alternatively, a discrete droplet generated may not include any beads (e.g., as in unoccupied droplet **218**). In

some instances, a discrete droplet generated may contain one or more analyte carriers, as described elsewhere herein. In some instances, a discrete droplet generated may comprise one or more reagents, as described elsewhere herein.

[00508] In some instances, the aqueous fluid **208** can have a substantially uniform concentration or frequency of beads **212**. The beads **212** can be introduced into the channel segment **202** from a separate channel (not shown in **FIG. 2**). The frequency of beads **212** in the channel segment **202** may be controlled by controlling the frequency in which the beads **212** are introduced into the channel segment **202** and/or the relative flow rates of the fluids in the channel segment **202** and the separate channel. In some instances, the beads can be introduced into the channel segment **202** from a plurality of different channels, and the frequency controlled accordingly.

[00509] In some instances, the aqueous fluid **208** in the channel segment **202** can comprise biological particles. In some instances, the aqueous fluid **208** can have a substantially uniform concentration or frequency of biological particles. As with the beads, the biological particles can be introduced into the channel segment **202** from a separate channel. The frequency or concentration of the biological particles in the aqueous fluid **208** in the channel segment **202** may be controlled by controlling the frequency in which the biological particles are introduced into the channel segment **202** and/or the relative flow rates of the fluids in the channel segment **202** and the separate channel. In some instances, the biological particles can be introduced into the channel segment **202** from a plurality of different channels, and the frequency controlled accordingly. In some instances, a first separate channel can introduce beads and a second separate channel can introduce biological particles into the channel segment **202**. The first separate channel introducing the beads may be upstream or downstream of the second separate channel introducing the biological particles.

[00510] The second fluid **210** can comprise an oil, such as a fluorinated oil, that includes a fluorosurfactant for stabilizing the resulting droplets, for example, inhibiting subsequent coalescence of the resulting droplets.

[00511] In some instances, the second fluid **210** may not be subjected to and/or directed to any flow in or out of the reservoir **204**. For example, the second fluid **210** may be substantially stationary in the reservoir **204**. In some instances, the second fluid **210** may be subjected to flow within the reservoir **204**, but not in or out of the reservoir **204**, such as via application of pressure to the reservoir **204** and/or as affected by the incoming flow of the aqueous fluid **208** at the junction **206**. Alternatively, the second fluid **210** may be subjected and/or directed to flow in or

out of the reservoir **204**. For example, the reservoir **204** can be a channel directing the second fluid **210** from upstream to downstream, transporting the generated droplets.

[00512] The channel structure **200** at or near the junction **206** may have certain geometric features that at least partly determine the sizes of the droplets formed by the channel structure **200**. The channel segment **202** can have a height, h_0 and width, w, at or near the junction **206**. By way of example, the channel segment **202** can comprise a rectangular cross-section that leads to a reservoir **204** having a wider cross-section (such as in width or diameter). Alternatively, the cross-section of the channel segment **202** can be other shapes, such as a circular shape, trapezoidal shape, polygonal shape, or any other shapes. The top and bottom walls of the reservoir **204** at or near the junction **206** can be inclined at an expansion angle, α. The expansion angle, α, allows the tongue (portion of the aqueous fluid **208** leaving channel segment **202** at junction **206** and entering the reservoir **204** before droplet formation) to increase in depth and facilitate decrease in curvature of the intermediately formed droplet. Droplet size may decrease with increasing expansion angle. The resulting droplet radius, R_d, may be predicted by the following equation for the aforementioned geometric parameters of h_0, w, and α:

$$R_d \approx 0.44\left(1 + 2.2\sqrt{\tan\alpha}\frac{w}{h_0}\right)\frac{h_0}{\sqrt{\tan\alpha}}$$

[00513] By way of example, for a channel structure with w = 21 μm, h = 21 μm, and α = 3°, the predicted droplet size is 121 μm. In another example, for a channel structure with w = 25 μm, h = 25 μm, and α = 5°, the predicted droplet size is 123 μm. In another example, for a channel structure with w = 28 μm, h = 28 μm, and α = 7°, the predicted droplet size is 124 μm.

[00514] In some instances, the expansion angle, α, may be between a range of from about 0.5° to about 4°, from about 0.1° to about 10°, or from about 0° to about 90°. For example, the expansion angle can be at least about 0.01°, 0.1°, 0.2°, 0.3°, 0.4°, 0.5°, 0.6°, 0.7°, 0.8°, 0.9°, 1°, 2°, 3°, 4°, 5°, 6°, 7°, 8°, 9°, 10°, 15°, 20°, 25°, 30°, 35°, 40°, 45°, 50°, 55°, 60°, 65°, 70°, 75°, 80°, 85°, or higher. In some instances, the expansion angle can be at most about 89°, 88°, 87°, 86°, 85°, 84°, 83°, 82°, 81°, 80°, 75°, 70°, 65°, 60°, 55°, 50°, 45°, 40°, 35°, 30°, 25°, 20°, 15°, 10°, 9°, 8°, 7°, 6°, 5°, 4°, 3°, 2°, 1°, 0.1°, 0.01°, or less. In some instances, the width, w, can be between a range of from about 100 micrometers (μm) to about 500 μm. In some instances, the width, w, can be between a range of from about 10 μm to about 200 μm. Alternatively, the width can be less than about 10 μm. Alternatively, the width can be greater than about 500 μm. In some instances, the flow rate of the aqueous fluid **208** entering the junction **206** can be between about 0.04 microliters (μL)/minute (min) and about 40 μL/min. In some instances, the flow rate of the aqueous fluid **208** entering the junction **206** can be between about 0.01 microliters (μL)/minute (min) and about 100 μL/min. Alternatively, the flow rate of the aqueous

fluid **208** entering the junction **206** can be less than about 0.01 μL/min. Alternatively, the flow rate of the aqueous fluid **208** entering the junction **206** can be greater than about 40 μL/min, such as 45 μL/min, 50 μL/min, 55 μL/min, 60 μL/min, 65 μL/min, 70 μL/min, 75 μL/min, 80 μL/min, 85 μL/min, 90 μL/min, 95 μL/min, 100 μL/min, 110 μL/min , 120 μL/min , 130 μL/min , 140 μL/min , 150 μL/min, or greater. At lower flow rates, such as flow rates of about less than or equal to 10 microliters/minute, the droplet radius may not be dependent on the flow rate of the aqueous fluid **208** entering the junction **206**.

[00515] In some instances, at least about 50% of the droplets generated can have uniform size. In some instances, at least about 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or greater of the droplets generated can have uniform size. Alternatively, less than about 50% of the droplets generated can have uniform size.

[00516] The throughput of droplet generation can be increased by increasing the points of generation, such as increasing the number of junctions (e.g., junction **206**) between aqueous fluid **208** channel segments (e.g., channel segment **202**) and the reservoir **204**. Alternatively or in addition, the throughput of droplet generation can be increased by increasing the flow rate of the aqueous fluid **208** in the channel segment **202**.

[00517] The methods and systems described herein may be used to greatly increase the efficiency of single cell applications and/or other applications receiving droplet-based input. For example, following the sorting of occupied cells and/or appropriately-sized cells, subsequent operations that can be performed can include generation of amplification products, purification (e.g., via solid phase reversible immobilization (SPRI)), further processing (e.g., shearing, ligation of functional sequences, and subsequent amplification (e.g., via PCR)). These operations may occur in bulk (e.g., outside the partition). In the case where a partition is a droplet in an emulsion, the emulsion can be broken and the contents of the droplet pooled for additional operations. Additional reagents that may be co-partitioned along with the barcode bearing bead may include oligonucleotides to block ribosomal RNA (rRNA) and nucleases to digest genomic DNA from cells. Alternatively, rRNA removal agents may be applied during additional processing operations. The configuration of the constructs generated by such a method can help minimize (or avoid) sequencing of the poly-T sequence during sequencing and/or sequence the 5' end of a polynucleotide sequence. The amplification products, for example, first amplification products and/or second amplification products, may be subject to sequencing for sequence analysis. In some cases, amplification may be performed using the Partial Hairpin Amplification for Sequencing (PHASE) method.

[00518] A variety of applications require the evaluation of the presence and quantification of different biological particle or organism types within a population of biological particles, including, for example, microbiome analysis and characterization, environmental testing, food safety testing, epidemiological analysis, e.g., in tracing contamination or the like.

Computer systems

[00519] The present disclosure provides computer systems that are programmed to implement methods of the disclosure. **FIG. 7** shows a computer system **7** that is programmed or otherwise configured to implement one or more of the methods described herein. For example, computer system **701** may be programmed or otherwise configured to control a microfluidics system (e.g., fluid flow), (ii) sort occupied droplets from unoccupied droplets, (iii) polymerize droplets, (iv) perform sequencing applications, and/or (v) generate and maintain sequencing libraries. The computer system **701** can regulate various aspects of the present disclosure, such as, for example, regulating fluid flow rate in one or more channels in a microfluidic structure, regulating polymerization application units, etc. The computer system **701** can be an electronic device of a user or a computer system that is remotely located with respect to the electronic device. The electronic device can be a mobile electronic device.

[00520] The computer system **701** includes a central processing unit (CPU, also "processor" and "computer processor" herein) **705**, which can be a single core or multi core processor, or a plurality of processors for parallel processing. The computer system **701** also includes memory or memory location **710** (e.g., random-access memory, read-only memory, flash memory), electronic storage unit **715** (e.g., hard disk), communication interface **720** (e.g., network adapter) for communicating with one or more other systems, and peripheral devices **725**, such as cache, other memory, data storage and/or electronic display adapters. The memory **710**, storage unit **715**, interface **720** and peripheral devices **725** are in communication with the CPU **705** through a communication bus (solid lines), such as a motherboard. The storage unit **715** can be a data storage unit (or data repository) for storing data. The computer system **701** can be operatively coupled to a computer network ("network") **730** with the aid of the communication interface **720**. The network **730** can be the Internet, an internet and/or extranet, or an intranet and/or extranet that is in communication with the Internet. The network **730** in some cases is a telecommunication and/or data network. The network **730** can include one or more computer servers, which can enable distributed computing, such as cloud computing. The network **730**, in some cases with the aid of the computer system **701**, can implement a peer-to-peer network, which may enable devices coupled to the computer system **701** to behave as a client or a server.

[00521] The CPU **705** can execute a sequence of machine-readable instructions, which can be embodied in a program or software. The instructions may be stored in a memory location, such as the memory **710**. The instructions can be directed to the CPU **705**, which can subsequently program or otherwise configure the CPU **705** to implement methods of the present disclosure. Examples of operations performed by the CPU **705** can include fetch, decode, execute, and writeback.

[00522] The CPU **705** can be part of a circuit, such as an integrated circuit. One or more other components of the system **701** can be included in the circuit. In some cases, the circuit is an application specific integrated circuit (ASIC).

[00523] The storage unit **715** can store files, such as drivers, libraries and saved programs. The storage unit **715** can store user data, e.g., user preferences and user programs. The computer system **701** in some cases can include one or more additional data storage units that are external to the computer system **701**, such as located on a remote server that is in communication with the computer system **701** through an intranet or the Internet.

[00524] The computer system **701** can communicate with one or more remote computer systems through the network **730**. For instance, the computer system **701** can communicate with a remote computer system of a user (e.g., operator). Examples of remote computer systems include personal computers (e.g., portable PC), slate or tablet PC's (e.g., Apple® iPad, Samsung® Galaxy Tab), telephones, Smart phones (e.g., Apple® iPhone, Android-enabled device, Blackberry®), or personal digital assistants. The user can access the computer system **701** via the network **730**.

[00525] Methods as described herein can be implemented by way of machine (e.g., computer processor) executable code stored on an electronic storage location of the computer system **701**, such as, for example, on the memory **710** or electronic storage unit **715**. The machine executable or machine readable code can be provided in the form of software. During use, the code can be executed by the processor **705**. In some cases, the code can be retrieved from the storage unit **715** and stored on the memory **710** for ready access by the processor **705**. In some situations, the electronic storage unit **715** can be precluded, and machine-executable instructions are stored on memory **710**.

[00526] The code can be pre-compiled and configured for use with a machine having a processor adapted to execute the code, or can be compiled during runtime. The code can be supplied in a programming language that can be selected to enable the code to execute in a pre-compiled or as-compiled fashion.

[00527] Aspects of the systems and methods provided herein, such as the computer system **701**, can be embodied in programming. Various aspects of the technology may be thought of as "products" or "articles of manufacture" typically in the form of machine (or processor) executable code and/or associated data that is carried on or embodied in a type of machine readable medium. Machine-executable code can be stored on an electronic storage unit, such as memory (e.g., read-only memory, random-access memory, flash memory) or a hard disk. "Storage" type media can include any or all of the tangible memory of the computers, processors or the like, or associated modules thereof, such as various semiconductor memories, tape drives, disk drives and the like, which may provide non-transitory storage at any time for the software programming. All or portions of the software may at times be communicated through the Internet or various other telecommunication networks. Such communications, for example, may enable loading of the software from one computer or processor into another, for example, from a management server or host computer into the computer platform of an application server. Thus, another type of media that may bear the software elements includes optical, electrical and electromagnetic waves, such as used across physical interfaces between local devices, through wired and optical landline networks and over various air-links. The physical elements that carry such waves, such as wired or wireless links, optical links or the like, also may be considered as media bearing the software. As used herein, unless restricted to non-transitory, tangible "storage" media, terms such as computer or machine "readable medium" refer to any medium that participates in providing instructions to a processor for execution.

[00528] Hence, a machine readable medium, such as computer-executable code, may take many forms, including but not limited to, a tangible storage medium, a carrier wave medium or physical transmission medium. Non-volatile storage media include, for example, optical or magnetic disks, such as any of the storage devices in any computer(s) or the like, such as may be used to implement the databases, etc. shown in the drawings. Volatile storage media include dynamic memory, such as main memory of such a computer platform. Tangible transmission media include coaxial cables; copper wire and fiber optics, including the wires that comprise a bus within a computer system. Carrier-wave transmission media may take the form of electric or electromagnetic signals, or acoustic or light waves such as those generated during radio frequency (RF) and infrared (IR) data communications. Common forms of computer-readable media therefore include for example: a floppy disk, a flexible disk, hard disk, magnetic tape, any other magnetic medium, a CD-ROM, DVD or DVD-ROM, any other optical medium, punch cards paper tape, any other physical storage medium with patterns of holes, a RAM, a ROM, a PROM and EPROM, a FLASH-EPROM, any other memory chip or cartridge, a carrier wave

transporting data or instructions, cables or links transporting such a carrier wave, or any other medium from which a computer may read programming code and/or data. Many of these forms of computer readable media may be involved in carrying one or more sequences of one or more instructions to a processor for execution.

[00529] The computer system **701** can include or be in communication with an electronic display **735** that comprises a user interface (UI) **740** for providing, for example, results of sequencing analysis, etc. Examples of UIs include, without limitation, a graphical user interface (GUI) and web-based user interface.

[00530] Methods and systems of the present disclosure can be implemented by way of one or more algorithms. An algorithm can be implemented by way of software upon execution by the central processing unit **705**. The algorithm can, for example, perform a nucleic acid sequencing assay, etc..

[00531] Devices, systems, compositions and methods of the present disclosure may be used for various applications, such as, for example, processing a single analyte (e.g., RNA, DNA, or protein) or multiple analytes (e.g., DNA and RNA, DNA and protein, RNA and protein, or RNA, DNA and protein) from a single cell. For example, a biological particle (e.g., a cell or cell bead) is partitioned in a partition (e.g., droplet), and multiple analytes from the biological particle are processed for subsequent processing. The multiple analytes may be from the single cell. This may enable, for example, simultaneous proteomic, transcriptomic and genomic analysis of the cell.

System for Characterizing Cells

[00532] Disclosed herein, in some embodiments, are systems for characterizing cells. In an aspect, the present disclosure provides a system for characterizing cells, comprising: a plurality of partitions comprising a plurality of cells or cell nuclei and a plurality of particles, wherein a partition of the plurality of partitions comprises a cell or cell nucleus of the plurality of cells or cell nuclei and a particle of the plurality of particles, wherein (i) the plurality of cells or cell nuclei comprises a plurality of nucleic acid molecules, wherein the plurality of nucleic acid molecules comprises a plurality of RNA molecules and a plurality of DNA molecules; and (ii) the plurality of particles comprises a plurality of nucleic acid barcode molecules coupled thereto, wherein a nucleic acid barcode molecule of the plurality of nucleic acid barcode molecules comprises a nucleic acid barcode sequence of a plurality of nucleic acid barcode sequences, and wherein the particle comprises a unique nucleic acid barcode sequence of the plurality of nucleic acid barcode sequences; and one or more computer processors, individually or collectively, programmed to; (a) process a plurality of barcoded nucleic acid molecules, generated in the

plurality of partitions using the plurality of nucleic acid barcode molecules and the plurality of nucleic acid molecules, or derivatives thereof, to generate sequence information corresponding to the RNA molecules and the DNA molecules; and (b) use the sequence information to identify characteristics of the plurality of cells or cell nuclei.

[00533] In some embodiments, the characteristics of the plurality of cells or cell nuclei comprise cell types. In some embodiments, the cell types are selected from the group consisting of monocytes, natural killer cells, B cells, T cells, granulocytes, dendritic cells, and stromal cells. In some embodiments, the B cells are selected from the group consisting of replicating B cells, normal B cells, and tumor B cells. In some embodiments, the B cells are selected from the group consisting of naïve B cells, memory B cells, plasmablast B cells, lymphoplasmacytoid cells, B-1 cells, regulatory B cells, and plasma B cells. In some embodiments, the T cells are selected from the group consisting of replicating T cells and normal T cells. In some embodiments, the T cells are selected from the group consisting of helper T cells, cytotoxic T cells, memory T cells, regulatory T cells, natural killer T cells, mucosal associated invariant T (MAIT) cells, gamma delta T cells, effector T cells, and naïve T cells. In some embodiments, the monocytes are selected from the group consisting of monocytes characterized by high level expression of CD14 cell surface receptors and monocytes characterized by high level expression of CD16 cell surface receptors. In some embodiments, the dendritic cells are selected from conventional dendritic cells and plasmacytoid dendritic cells. In some embodiments, the sequences corresponding to the DNA molecules of the plurality of DNA molecules correspond to regions of accessible chromatin. In some embodiments, the RNA molecules of the plurality of RNA molecules comprise messenger RNA (mRNA) molecules. In some embodiments, the sequence information comprises a first plurality of sequencing reads corresponding to the DNA molecules and a second plurality of sequencing reads corresponding to the RNA molecules. In some embodiments, the sequence information comprises a plurality of sequencing reads associated with individual cells or cell nuclei of the plurality of cells or cell nuclei.

[00534] In some embodiments, wherein in (b) the one or more computer processors are individually or collectively programmed to: determine a linked signature of the cell or cell nucleus of the plurality of cells or cell nuclei using the sequence information, which linked signature of the cell or cell nucleus links a first data set comprising sequence information corresponding to DNA molecules of the cell or cell nucleus and a second data set comprising sequence information corresponding to RNA molecules of the cell or cell nucleus. In some embodiments, wherein in (b) the one or more computer processors are individually or collectively programmed to: use the sequence information to cluster cells or cell nuclei of the

plurality of cells or cell nuclei by gene expression signatures and/or by regions of accessible chromatin signatures. In some embodiments, wherein in (b) the one or more computer processors are individually or collectively programmed to: (i) use the sequence information to cluster cells or cell nuclei of the plurality of cells or cell nuclei by regions of accessible chromatin signatures, (ii) use the sequence information to cluster cells or cell nuclei of the plurality of cells or cell nuclei by gene expression signatures, and (iii) use the sequence information and the cells or cell nuclei clustered by the gene expression signatures to further characterize the cells or cell nuclei clustered by the regions of accessible chromatin. In some embodiments, wherein in (b) the one or more computer processors are individually or collectively programmed to: (i) use the sequence information to cluster cells or cell nuclei of the plurality of cells or cell nuclei by regions of accessible chromatin signatures, (ii) use the sequence information to cluster cells or cell nuclei of the plurality of cells or cell nuclei by gene expression signatures, and (iii) use the sequence information and the cells or cell nuclei clustered by the regions of accessible chromatin signatures to further characterize the cells or cell nuclei clustered by the gene expression signatures.

[00535] In some embodiments, the plurality of cells or cell nuclei is derived from a sample comprising a tumor or suspected of comprising a tumor. In some embodiments, the one or more computer processors are further, individually or collectively, programmed to process the sequence information corresponding to the RNA molecules and the DNA molecules with sequence information generated from a control sample. In some embodiments, the sample is derived from a bodily fluid. In some embodiments, the sample is derived from a biopsy. In some embodiments, the tumor is a B cell lymphoma tumor. In some embodiments, the one or more computer processors are further, individually or collectively, programmed to use the sequence information to identify a presence of a tumor cell or cell nucleus in the sample. In some embodiments, the one or more computer processors are further, individually or collectively, programmed to: (c) use the sequence information to identify a cell type, a cell state, a tumor-specific gene expression pattern, or a tumor-specific differentially accessible region of chromatin in the sample. In some embodiments, the one or more computer processors are further, individually or collectively, programmed to, based at least in part on (c), identify a therapeutic regimen for treatment of a subject from which the sample derives. In some embodiments, the therapeutic regimen comprises administration of a therapeutically effective amount of an agent targeting one or more targets identified in the tumor-specific gene expression pattern or the tumor-specific differentially accessible region of chromatin.

[00536] In some embodiments, the plurality of partitions comprises a plurality of droplets. In some embodiments, the plurality of cells or cell nuclei comprises a plurality of transposed nuclei. In some embodiments, the plurality of particles comprises a plurality of gel beads. In some embodiments, the plurality of nucleic acid barcode molecules is releasably coupled to the plurality of particles. In some embodiments, the nucleic acid barcode molecules of the plurality of nucleic acid barcode molecules are releasable from the particles of the plurality of particles upon application of a stimulus. In some embodiments, the stimulus is a chemical stimulus. In some embodiments, the stimulus comprises a reducing agent. In some embodiments, the plurality of nucleic acid barcode molecules is coupled to the plurality of particles via a plurality of labile moieties. In some embodiments, the system further comprises a microfluidic device that generates the plurality of partitions.

Systems for Determining a Condition of a Sample

[00537] Disclosed herein, in some embodiments, are systems for determining a condition of a sample. In an aspect, a system for determining a condition of a sample, comprises: one or more databases comprising (i) a first data set comprising sequencing information corresponding to regions of accessible chromatin of a plurality of deoxyribonucleic acid (DNA) molecules of a cell or cell nucleus of the sample, (ii) a second data set comprising sequencing information corresponding to a plurality of ribonucleic acid (RNA) molecules of the cell or cell nucleus, and (iii) a linked signature of the cell or cell nucleus using the first data set and the second data set; one or more computer processors operably coupled to the one or more databases, wherein the one or more computer processors are individually or collectively programmed to use the linked signature of the cell or cell nucleus and a control linked signature of a control cell or cell nucleus of a control sample to determine one or more regions of accessible chromatin of the plurality of DNA molecules or one or more genes expressed from the plurality of RNA molecules indicative of the condition.

[00538] In some embodiments, the one or more computer processors are individually or collectively programmed to determine a level of the one or more regions of accessible chromatin and/or the one or more genes expressed indicative of the condition in one or more samples of an individual suspected of having the condition. In some embodiments, the one or more computer processors are individually or collectively programmed to generate an output that relates to providing a diagnostic evaluation of the condition, a prognostic evaluation of the condition, monitoring of the condition, and/or management of the condition. In some embodiments, the one or more computer processors are individually or collectively configured to identify a gene

associated with the one or more regions of accessible chromatin and/or the one or more genes expressed as a target of a therapeutic regimen for treatment of the condition.

[00539] In some embodiments, the one or more computer processors are individually or collectively programmed to generate an output that relates to determining an administration regimen of a therapeutically effective amount of an agent to a subject targeting the target, wherein the sample is derived from the subject. In some embodiments, the one or more computer processors are individually or collectively programmed to generate an output that relates to determining an efficacy of an agent targeting the target when administered to a subject, wherein the sample is derived from the subject. In some embodiments, the one or more computer processors are individually or collectively programmed to generate an output that relates to detecting a presence or absence of a response to the agent by the subject, wherein the response comprises a quantity, degree, or extent of response following administration of a first dose or a subsequent dose of the agent.

[00540] In some embodiments, the response comprises a differential in gene expression and/or chromatin accessibility of the target between prior to and after administration of the agent. In some embodiments, the sample is from a subject having a tumor or suspected of having a tumor. In some embodiments, the condition is a tumor, cancer, malignancy, neoplasm, or other proliferative disease or disorder. In some embodiments, the condition is a B cell malignancy. In some embodiments, the B cell malignancy is B cell lymphoma. In some embodiments, the sample is derived from a bodily fluid. In some embodiments, the sample is derived from a biopsy.

[00541] In some embodiments, the first data set and the second data set comprise a plurality of sequencing reads corresponding to sequences of the plurality of DNA molecules and the plurality of RNA molecules, wherein the sequencing reads each correspond to the cell or cell nucleus via a nucleic acid barcode sequence. In some embodiments, the one or more computer processors are individually or collectively programmed to cluster a plurality of cells or cell nuclei of the sample by respective regions of accessible chromatin signatures, by respective genes expressed, and/or by respective linked signatures of the plurality of cells or cell nuclei. In some embodiments, the plurality of cells or cell nuclei are clustered by cell types selected from the group consisting of monocytes, natural killer cells, B cells, T cells, granulocytes, dendritic cells, and stromal cells. In some embodiments, the B cells are selected from the group consisting of replicating B cells, normal B cells, and tumor B cells. In some embodiments, the B cells are selected from the group consisting of naïve B cells, memory B cells, plasmablast B cells, lymphoplasmacytoid cells, B-1 cells, regulatory B cells, and plasma B cells. In some

embodiments, the T cells are selected from the group consisting of replicating T cells and normal T cells. In some embodiments, the T cells are selected from the group consisting of helper T cells, cytotoxic T cells, memory T cells, regulatory T cells, natural killer T cells, mucosal associated invariant T (MAIT) cells, gamma delta T cells, effector T cells, and naïve T cells.

[00542] In some embodiments, the one or more computer processors are individually or collectively programmed to monitor a level of the one or more regions of accessible chromatin and/or the one or more genes expressed indicative of the condition in the individual.

[00543] In some embodiments, the plurality of DNA fragments is tagged. In some embodiments, the first data set comprising sequencing information corresponding to regions of accessible chromatin of a plurality of DNA molecules of a cell or cell nucleus of the sample and the second data set comprising sequencing information corresponding to a plurality of RNA molecules of the cell or cell nucleus are barcoded with barcoded nucleic acid sequences. In some embodiments, the first data set comprising sequencing information corresponding to regions of accessible chromatin of a plurality of DNA molecules of a cell or cell nucleus of the sample and the second data set comprising sequencing information corresponding to a plurality of RNA molecules of the cell or cell nucleus are barcoded with barcoded nucleic acid sequences within a plurality of partitions.

[00544] In some embodiments, the system further comprises a device or a sequencer configured to sequence the plurality of barcoded nucleic acid sequences. In some embodiments, the first data set is generated by sequencing a first plurality of barcoded nucleic acid sequences comprising sequences corresponding to regions of accessible chromatin of the plurality of deoxyribonucleic acid (DNA) molecules of the cell or cell nucleus of the sample; and the second data set is generated by sequencing a second plurality of barcoded nucleic acid sequences comprising sequences of the ribonucleic acid (RNA) molecules of the cell or cell nucleus.

[00545] In some embodiments, the one or more computer processors are individually or collectively programmed to process the first data set and/or the second data set to generate a filtered first data set and/or a filtered second data set. In some embodiments, the filtered first data set is filtered using motif enrichment. In some embodiments, the filtered second data set is filtered using differential expression analysis. In some embodiments, the one or more computer processors are individually or collectively programmed to process the first data set and/or the second data set to generate a linkage significance. In some embodiments, the one or more computer processors are individually or collectively programmed to process the first filtered data set and/or the second filtered data set to generate an enrichment score. In some embodiments, the filtered first data set and the filtered second data set is used to generate a transcription factor-

target gene network. In some embodiments, wherein a gene from the transcription factor-target gene network is identified as a target of a therapeutic regimen for treatment of the condition. In some embodiments, the target is a transcription factor.

EXAMPLES

Example 1: Functional Characterization of Small B Cell Lymphoma and Its Signaling Pathways

[00546] A case study using B cell lymphoma is performed. **FIG. 33** summarizes details of the case study. Gene expression and chromatin data are generated from 9158 single cells from a CD20+ diffuse small cell lymphoma sample (e.g., as described herein). Gene expression markers are used to annotate immune cell types included in the sample (e.g., as described herein). **FIG. 34** shows cell type annotation for cells based on gene expression data (left panel) and chromatin data (right panel). As shown in the figure, the gene expression analysis initially identifies replicating T cells, T cells, monocytes, replicating B cells, and B cells as separate classes, while the chromatin analysis initially identifies only T cells, monocytes, and B cells as separate classes. B cells include both tumor B cells and normal B cells. Tumor B cells and normal B cells may be disentangled using gene expression information, chromatin information, or a combination thereof. Orthogonal lines of evidence may be used to differentiate between tumor B cells and normal B cells.

[00547] **FIG. 35** shows an example orthogonal method for parsing out tumor B cells from normal B cells using gene expression data. Mutation load may be mapped using single nucleotide variants (SNVs) from publicly available mutation data from the TCGA-DLLC project on diffuse large B cell lymphomas. This data is filtered to retain only 279 SNVs that are predicted to have a deleterious phenotype and present in the dbSNP database. These retained cells are shown as dark dots in the left panel of **FIG. 35**. As the tumor B cells are known to be CD20 positive (shown in **FIG. 41**), positive correlation between CD20 positive cells and SNVs is used to identify tumor B cells among total B cells. The BANK1 (B cell scaffold protein with akyrin repeats 1) pathway may also be used to identify tumor B cells among B cells. BANK1 modulates B cell antigen receptor (BCR)-induced calcium mobilization and weakens CD40-mediated Akt activation to prevent B cell hyperactivation. Therefore, depressed BANK1 expression is indicative of tumor B cells. Along this axis, CD40 expression may be enhanced in tumor B cells relative to normal B cells. Negative correlation between BANK1 expression and CD40 expression is thus indicative of tumor B cells. **FIG. 41** shows an example method for parsing out tumor B cells from normal B cells using these orthogonal lines of evidence using gene expression data.

[00548] **FIG. 36** shows cluster representations of cell types generated from gene expression data (left panel) and chromatin data (right panel). The cluster representation generated from gene expression data is annotated to show the separate tumor B cell and normal B cell populations. The cluster representation generated from chromatin data is annotated based on gene expression analysis. In this representation, normal B cells are included in a cluster that is distinct from the tumor B cell cluster. This indicates that normal B cells and tumor B cells have distinct chromatin signatures.

[00549] Gene expression data is also used to classify the tumor. Using gene expression data, the tumor is classified as a gastric MALT lymphoma.

[00550] Upon separation of normal B cells and tumor B cells, differential gene expression analysis may be performed to identify differential expression of various genes between these two cell populations. As shown in **FIG. 37**, proteins including the IL-4 receptor ("IL4R") are over expressed in tumor B cells relative to normal B cells, while proteins including BANK1 is under expressed in tumor B cells relative to normal B cells. The observed over-expression of IL4R is consistent with the hyperactivation of cytokines associated with many cancers.

[00551] The gene expression outlined above may provide insight into diagnostic decisions and may inform potential therapeutic options. Gene expression analysis coupled with chromatin analysis may also be used to identify tumor-specific signaling pathways. Dysregulation of the JAK-STAT signaling pathway is observed in many primary human tumors. Signal transducer and activator of transcription (STAT) proteins are critical mediators of cytokine signaling. However, STATs are latent cytoplasmic proteins, making their expression a poor proxy for function. The upper right panel of **FIG. 38A** shows relatively similar expression of the Stat3 and Stat6 proteins across all cell types including normal B cells and tumor B cells, indicating that gene expression alone is not a good proxy for function. Instead, activity may be assessed via analysis of chromatin data. Upon JAK-mediated phosphorylation, an activated STAT translocates to the nucleus and binds to its DNA-recognition motifs in the promoters of cytokine-inducible genes. This activity is illustrated in the center-left panels of **FIG. 38A**. Among the seven STAT proteins, Stat6 is activated by IL-4 and IL-13. As shown in the lower right panel of **FIG. 38A**, the Stat6 motif is enhanced in tumor B cells relative to normal B cells. This assessment provides insight into the transcription factors that may be activated in tumor cells. An expanded view of this analysis is included in **FIG. 42**. Bcl-2 and IgE are known targets of Stat6. Stat6 activation promotes immunoglobulin class switching to IgE and prevents apoptosis through the induction of anti-apoptotic genes such as Bcl-2. As shown in the lower right panel of **FIG. 42**, these targets are enhanced for tumor B cells.

[00552] The combination of gene expression and chromatin analysis may facilitate the identification of tumor-specific differentially accessible chromatin regions. **FIG. 38B** shows chromatin data associated with IL4R, which is enhanced in tumor B cells relative to normal B cells (e.g., as described above), for tumor B cells, normal B cells, monocytes, and T cells. Each of these cell classes includes a peak to the left of the middle of the plot corresponding to an IL4R promoter. However, unlike other cell types, tumor B cells include an area of enhanced intensity upstream of this peak (circled). **FIG. 38C** compares this chromatin data (upper left panel) to publicly available genetic data (lower left panel). The gray highlighted region encompassing the area of enhanced intensity for the tumor B cells overlaps with intensity for H3K4Me1, a differentially methylated region associated with enhancers. This so-called "linked peak" may be a useful predictor for IL4R expression. The right panels of **FIG. 38C** show the proportion of cells having signal for various features for different cell types. While IL4R expression analyzed with gene expression data is differentially expressed in tumor B cells compared to other cell types, the IL4R promoter peak analyzed with chromatin data is not as predictive of cell type. However, the "linked peak" (e.g., the enhancer identified using chromatin data) is strongly associated with tumor B cells compared to other cell types and may therefore be predictive of IL4R expression. This assessment indicates that enhancers observable in chromatin data may be better predictors of gene expression than promoters.

Example 2: Further Characterization of Small B Cell Lymphoma and Linkage Analysis

[00553] The underlying libraries from the B cell lymphoma case study of Example 1 were sequenced to further depth and processed to determine feature linkages. The data as shown in **FIGs. 45A** and **45B** confirmed the findings of the Example 1 and identified regions of open chromatin driving expression of IL4R in tumor B cells that were not observed in other cells (e.g., normal B cells, T cells, and monocytes/dendritic cells). In **FIG. 45B,** the outlined box highlights the region of open chromatin driving expression of IL4R in tumor B cells that were not observed in other cells.

Example 3: Cell Type and Heterogeneity of Different B Cell Lymphomas

[00554] A case study using B cell lymphoma is performed. Linked single-cell gene expression and chromatin signature is used to identify the cell type heterogeneity and malignancy of B cell lymphoma. Single-cell gene expression and chromatin accessibility profiling are performed on lymph node-derived lymphocytes from healthy individual and individuals suspected of suffering from a type of B cell lymphoma. Single-cell suspensions of

lymphoid tissue from each group of individuals are assayed to generate single cell gene expression sequencing data and single ATAC sequencing data.

[00555] Lymph node-derived cells from these samples are analyzed by staining with different surface and intracellular antibodies and flow cytometry. For example, staining for the expression of immunoglobin light chain κ or λ is used to distinguish between malignant and malignant B cells. The malignant B cells are distinguished from the non-malignant B cells because they only express one of κ or λ immunoglobin light chains. For example, the malignant B cells only express the κ light chain.

The single-cell gene expression and chromatin accessibility data is analyzed using dimension reduction techniques, including but not limited to t-distributed stochastic neighbor embedding (tSNE), Principle component Analysis (PCA), or Uniform Manifold Approximation and Projection (UMAP). All data points for the gene expression and chromatin profiling for each single cell are reduced to low dimensions such as two dimensions and visualized in a two-dimensional scatter plot. Marker expression, sample origin, lymphoma subtype, malignancy, and other data can be imposed on the scatter plot to help identifying different cell types. The relative frequency of each cell type in each type of lymphoma is compared to reveal how the cell heterogeneity changes based on different criteria including but not limited to malignancy of the lymphoma or the subtype of the lymphoma.

Example 4: Diagnosis of B Cell Lymphoma Subtype and Malignancy Using Machine Learning

[00556] The subtype or malignancy of B cell lymphoma is diagnosed using linked gene expression and chromatin accessibility signature. Single-cell gene expression and chromatin accessibility are profiled for each group of B cell lymphoma based on cancer subtype and malignancy. A subset of this data set, which at least encompasses every group of B cell lymphoma based on the cancer subtype and malignancy, is used as a training set using a machine learning algorithm such as but not limited to random forest tree or k-means clustering. The trained model is tested and verified with the remaining subset of the data. The parameters are optimized for the algorithm based on the accuracy and sensitivity of the prediction result. Accuracy or sensitivity is at least 70%.

[00557] Once a model is set up, it is used to diagnose the subtype or malignancy of B cell lymphoma in individuals with uncharacterized group of B cell lymphoma based on subtype and malignancy. Different algorithms return different types of prediction. For example, k-means clustering identifies each tested sample to one specific group of B cell lymphoma while random

forest tree provides a probability of each group of B cell lymphoma the tested sample belongs to. Other tests, such as marker expression, are also used to identify B cell lymphoma groups.

Example 5: Similarity of B Cell Lymphoma Subtype and Malignancy

[00558] The similarity of different groups of B cell lymphoma based on subtype and malignancy are identified using the similarity of linked gene expression and chromatin accessibility signature. Single-cell gene expression and chromatin accessibility are profiled for each group of B cell lymphoma based on subtype and malignancy. A panel of these profiles is generated. To diagnose an individual with uncharacterized group of B cell lymphoma based on subtype or malignancy, a single-cell gene expression and chromatin accessibility of the lymphocyte of this individual is generated. A similarity score is calculated using every data point of the gene expression and chromatin accessibility profiling. The similarity score is calculated using a distance measurement such as Euclidean distance or Manhattan distance. Before calculating the similarity, each group of data point (such as the expression or chromatin accessibility of each gene) is normalized by normalization method such as but not limited to Z-score. Two most similar profiles have the smallest distance among all tested profile-pairs. Such analysis is used to guide treatment option. For example, effective treatment for one profile based on B cell lymphoma subtype and malignancy suggests that the same treatment is also effective for the most similar profiles.

Example 6: Analysis of Expression in Tumors vs Normal Identified BCR Activation Signature

[00559] An analysis was performed on 14,000 cells collected from a human B cell lymphoma sample from a tumor classified as a diffuse small B cell lymphoma tissue. While the cells were from the same biopsy sample as described in Example 1 and 2, this sample was processed and analyzed separately. Isolated nuclei from the flash-frozen intra-abdominal lymph node tumor were flow sorted, permeabilized, and transposed in bulk before single nuclei encapsulation in droplets, where DNA fragments and the 3' ends of mRNA were barcoded. Paired ATAC and gene expression libraries were generated from 14,000 total nuclei. The major cell types were analyzed and annotated based on known gene expression markers of immune and stromal cells in the lymph node, including B cells, T cells, monocytes/dendritic cells, fibroblasts and other stromal cell types. There were two major clusters expressing B cell marker CD19 and MS4A1. One of the two clusters has high expression of CD40, a known marker for malignant cells in B cell lymphoma. Additionally, BANK1 was observed to be strongly repressed in this B cell

cluster. BANK1 is a known attenuator of BCR activation pathways and frequently repressed in lymphoma tumorigenesis. Therefore, this B cell cluster is annotated as tumor B cells. Expression of mitotic cell markers such as MKI67 were also examined and annotated distinct clusters for cycling tumor B cells and T cells.

[00560] A heatmap of mean expression in normal B cells, tumor B cells and cycling tumor B cells for selected top differentially expressed immune genes, transcription factors and cell cycle genes is shown in **FIG. 46A**. **FIG. 46B** shows an enriched functional gene set for up-regulated genes in tumor B cells, computed using EnrichR (Chen et al., "Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool" BMC Bioinformatics. 2013;128(14); Kuleshov et al., "Enrichr: a comprehensive gene set enrichment analysis web server 2016 update" Nucleic Acids Research. 2016; gkw377).

Example 7: Transcription Factor (TF) Network Analysis

[00561] A 3-step strategy for constructing a transcriptional network was designed and used to analyze the data from Example 6. Using this approach, tumor versus non-tumor cells were distinguished in a heterogenous sample and a cell type-specific gene regulatory network was reconstructed. Firstly, the differentially expressed gene in tumor cells were identified. By setting the threshold of p-value < 10^-20, a list of 198 genes upregulated in tumor cells were identified. Next, the gene list was intersected with the inferred feature linkages to find the peaks linked to the tumor-specific genes. Finally, motif enrichment analysis was performed to identify transcription factors with enriched motif occurrence in tumor cells. The enriched transcription factors with linked target genes defined the edges of the TF regulatory network. An analysis workflow of TF-gene network construction is depicted in **FIG. 47** showing the start from peak-to-gene feature linkages, peaks and genes were filtered using motif enrichment and differential expression analysis, respectively. The remaining peaks were further mapped to motifs, which connected the top differentially expressed genes as the inferred TF-target gene regulatory network.

[00562] Since mRNA and ATAC data were generated from the same cells, cell-type annotations can be transferred from one modality to the other. In addition to the identification of B cells, monocytes, and T-cell subtypes using canonical cell markers like the B-cell marker MS4A1, tumor B cells were distinguishable from normal B cells based on upregulated CD40 expression and reduced BANK1. PAX5 was significantly upregulated in tumor B cells relative to normal B cells, and has previously been identified as a core regulator of chronic lymphocytic leukemia (CLL) (Ott et al., (2018). *Cancer Cell*, *34*(6), 982-995.e7).

Super Enhancer Analysis

[00563] It has been reported that super enhancers are essential in CLL B cell tumorigenesis (Ott et al., (2018). *Cancer Cell*, *34*(6), 982-995.e7), in which PAX5 itself is regulated by a proximal super enhancer. The tumor-enriched linkages were compared with the annotated super enhancer in CLL and a strong enrichment of high significance linkage for super enhancers was observed (**FIG. 48A**). The top 3 most significant linkages that overlaps with a CLL super enhancer is PAX5, consistent with the observation that the PAX5 locus is the dominant super-enhancer in tumor B cells. **FIG. 48A** shows linkage significance distribution for tumor-enriched feature linkages, separated by the overlap of CLL-annotated super-enhancers and that the top 3 most significant linkages were PAX5 and ZCCHC7, which is the nearby gene of PAX5. **FIG. 48B** shows ATAC cut site coverage and inferred feature linkages at the PAX5 locus on the left. Linkages with correlation < 0.8 were filtered out to improve visualization and the top significant linkage is highlighted in the dashed box in **FIG. 48B** and per-cell type expression of PAX5 and linked peaks are shown on the right. Positively correlated feature linkages are denoted by arcs at top. Highlighted by the dotted box is a highly significant feature linkage between *PAX5* and a CLL super-enhancer (Ott et al., (2018). *Cancer Cell*, *34*(6), 982-995.e7). Below the illustrated feature linkages are open chromatin peaks identified for each cell cluster across a 0.3 Mb region. On the right are plots showing the expression level of *PAX5* across all cell clusters and the peak height (average number of cut sites per cell) for the selected feature across cell clusters. Tumor B cells, in contrast to normal B cells, have elevated *PAX5* expression and open chromatin at this super-enhancer (at the location of the dotted box in **FIG. 48B**).

Motif Enrichment Analysis

[00564] Motif enrichment analysis was performed in two different ways. The first way was aimed to find transcription factor motifs with increased accessibility in tumor B cells over normal B cells. In this global enrichment method, the enrichment is estimated using all peaks between two cell populations. To this end, TF deviation z scores were computed using chromVAR and two-sample t tests for every motif between tumor cells and normal B cells. The inferred mean difference between two populations is further z scored as the enrichment score. The second and alternative method is to identify motifs enrichment only in the peaks with linkage to the top tumor-upregulated genes. In this context-dependent enrichment, the enrichment is estimated under the tumor context by comparing potential tumor-specific enhancers with a set of background peaks sharing the GC and accessibility profile in tumor cells. To this end, a set of background peaks using a similar strategy as chromVAR was computed, with modification of matching GC content and peak size, instead of the GC content and cut sites

per cell. The motif enrichment is determined by hypergeometric tests of motif occurrences in enhancer peaks and background peaks.

[00565] The final list of enrichment motifs were defined as the union of the top 10 ranked motifs in global analysis and the top 10 ranked motifs in context-specific analysis. To improve interpretability and overcome the redundancy of TF motif sequences, TF motifs were grouped into TF families based on JASPAR 2020 vertebrate motif clustering result (Fornes et al., (2020) Nucleic Acids Research, 48(D1):D87–D92). The PAX5 motif was re-annotated to the PAX/CUX/ONECUT family, which includes PAX1, PAX2, PAX3, PAX4, PAX6, PAX7 and PAX9 because PAX5 is a well characterized transcription factor in B cells and shares high orthology with other PAX transcription factors.

[00566] Global enrichment analysis identified TCF3/TCF4 motifs as the transcription factor family with the most upregulated accessibility. This is consistent with the well characterized function of TCF3/4 in B cell lymphoma tumorigenesis (Basso et al., (2015). *Nature Reviews Immunology*, *15*(3), 172–184). Moreover, TCF4 as also the most abundant transcription factor detected in tumor cells. Several members of MYC family transcription factors were identified with strong enrichment in tumor vs normal as well as enhancer vs background. The top abundant MYC family TFs (basic helix-loop-helix and leucine zipper, or bHLHZ motifs) include HIF1A, MAX and MYC.

[00567] Paired box (PAX) family of transcription factors was identified in the analysis as the top 1 enriched motifs in tumor-specific enhancers over other accessible regions in tumor cells, with PAX5 as the most abundant PAX family TF expressed in tumor cells. Interestingly, PAX motifs were not observed to be differentially accessible in tumor cells compared to normal cells, which indicate a specific regulatory mode for PAX. The analysis suggests that the global accessibility of PAX TFs is maintained at a steady level between tumor and normal B cells. However, PAX TFs are more likely to bind tumor-specific enhancers. A similar enrichment pattern was reported in CLL patients where PAX5 is similarly enriched in both normal and tumor B cells in CLL patients, but in the context of tumor cells, PAX5 is the central regulatory transcription factor (Ott et al., (2018). *Cancer Cell*, *34*(6), 982-995.e7). As shown in **FIG. 49A**, the motif enrichment analysis workflow generated enriched motifs as the converged list from two analysis strategies. **FIG. 49B** shows motif enrichment scores of all motifs (left) and the top enriched hits (right).

Transcription factor network in tumor cells

[00568] Several key self and cross regulations for MYC, PAX and TCF transcription factors were identified. In **FIG. 50,** TF-target gene regulation is plotted in the heatmap color

scaled based on the feature linkage significance. Target genes are columns and are grouped based on gene ontology annotation. Transcription factor genes are rows and are grouped as TF families. Target gene differential expression p-values are plotted as the additional column annotation as well as total UMIs detected in tumor cells. Similarly, TF motif enrichment scores (ES_peaks for context-specific analysis and ES cells for global analysis) are plotted in the line plot as well as tumor UMIs for row annotations. Specifically, both PAX and TCF genes were observed to be regulated by the same TF family, such as TCF4-to-TCF4, PAX5-to-PAX5 and TCF4-to-PAX5. Other known TFs in B cell development and tumorigenesis were also among the target genes of MYC, PAX and TCF transcription factors, such as TP63, LEF1, IRF8 and MEF2B. Other potential target genes also include several key members of BCR activation pathway, such as BCL2, IGLC1, IL4R and SYK.

[00569] To identify tumor B cell–specific gene regulatory networks, feature linkages were filtered by genes upregulated in tumor B cells and motifs occurring in open chromatin peaks associated with these feature linkages to motifs in all peaks in tumor B cells were compared. Using this method, PAX1 motif was observed to be the most enriched (**FIG. 50**). PAX1 and PAX5 motifs were highly similar, however PAX1 was not expressed in tumor B cells, while PAX5 was highly expressed. Therefore, it is likely the PAX5 transcription factor is binding the identified PAX1 motif. At the PAX5 locus, which is differentially expressed between B cells and tumor B cells, and known to be a critical regulator of B-cell lymphoma, expression of PAX5 was highly correlated with open PAX5 motif sites in a previously identified super-enhancer, suggesting autoregulation (**FIG. 48B**, dashed box). Looking genome-wide, the presence of significant feature linkages suggests PAX5 may also regulate the immune transcription factors NFATC1, TCF4, IKZF1, and IRF8 (**FIG. 50**), suggesting PAX5 regulates a tumor B cell–specific network. Knockout of 147 different transcription factors in a CLL cell line showed that loss of PAX5 had the greatest effect on cell proliferation, confirming its importance (Ott et al., (2018). *Cancer Cell*, *34*(6), 982-995.e7). While confirmation of individual links in the predicted gene regulatory network may require functional tests, the confidence in regulatory connections is greatly increased by joint measurement of mRNA and ATAC data. In this example, feature linkages helped build putative gene regulatory networks by providing correlated gene expression and open chromatin regions across the genome.

[00570] While preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. It is not intended that the invention be limited by the specific

examples provided within the specification. While the invention has been described with reference to the aforementioned specification, the descriptions and illustrations of the embodiments herein are not meant to be construed in a limiting sense. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. Furthermore, it shall be understood that all aspects of the invention are not limited to the specific depictions, configurations or relative proportions set forth herein which depend upon a variety of conditions and variables. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is therefore contemplated that the invention shall also cover any such alternatives, modifications, variations or equivalents. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

CLAIMS

WHAT IS CLAIMED IS:

1. A method for characterizing cells or cell nuclei, comprising:

(a) providing a plurality of partitions comprising a plurality of cells or cell nuclei and a plurality of particles, wherein a partition of said plurality of partitions comprises a cell or cell nucleus of said plurality of cells or cell nuclei and a particle of said plurality of particles, wherein (i) said plurality of cells or cell nuclei comprises a plurality of nucleic acid molecules, wherein said plurality of nucleic acid molecules comprises a plurality of ribonucleic acid (RNA) molecules and a plurality of deoxyribonucleic acid (DNA) molecules; and (ii) said plurality of particles comprises a plurality of nucleic acid barcode molecules coupled thereto, wherein a nucleic acid barcode molecule of said plurality of nucleic acid barcode molecules comprises a nucleic acid barcode sequence of a plurality of nucleic acid barcode sequences, and wherein said particle comprises a unique nucleic acid barcode sequence of said plurality of nucleic acid barcode sequences;

(b) within said plurality of partitions, using nucleic acid barcode molecules of said plurality of nucleic acid barcode molecules and nucleic acid molecules of said plurality of nucleic acid molecules to generate a plurality of barcoded nucleic acid molecules, wherein said plurality of barcoded nucleic acid molecules comprises (i) a first subset comprising sequences corresponding to RNA molecules of said plurality of RNA molecules and (ii) a second subset comprising sequences corresponding to DNA molecules of said plurality of DNA molecules, wherein a barcoded nucleic acid molecule of said plurality of barcoded nucleic acid molecules comprises a sequence corresponding to a nucleic acid barcode sequence of said plurality of nucleic acid barcode sequences;

(c) processing said plurality of barcoded nucleic acid molecules, or derivatives thereof, to generate sequence information corresponding to said RNA molecules and said DNA molecules; and

(d) using said sequence information to identify characteristics of said plurality of cells or cell nuclei.

2. The method of claim 1, wherein said characteristics of said plurality of cells or cell nuclei comprise cell types.

3. The method of claim 2, wherein said cell types are selected from the group consisting of monocytes, natural killer cells, B cells, T cells, granulocytes, dendritic cells, and stromal cells.
4. The method of claim 3, wherein said B cells are selected from the group consisting of replicating B cells, normal B cells, and tumor B cells.
5. The method of claim 3 or 4, wherein said B cells are selected from the group consisting of naïve B cells, memory B cells, plasmablast B cells, lymphoplasmacytoid cells, B-1 cells, regulatory B cells, and plasma B cells.
6. The method of any one of claims 3-5, wherein said T cells are selected from the group consisting of replicating T cells and normal T cells.
7. The method of any one of claims 3-6, wherein said T cells are selected from the group consisting of helper T cells, cytotoxic T cells, memory T cells, regulatory T cells, natural killer T cells, mucosal associated invariant T (MAIT) cells, gamma delta T cells, effector T cells, and naïve T cells.
8. The method of any one of claims 3-7, wherein said monocytes are selected from the group consisting of monocytes characterized by high level expression of CD14 cell surface receptors and monocytes characterized by high level expression of CD16 cell surface receptors.
9. The method of any one of claims 3-8, wherein said dendritic cells are selected from conventional dendritic cells and plasmacytoid dendritic cells.
10. The method of any one of claims 1-9, wherein said sequences corresponding to said DNA molecules of said plurality of DNA molecules correspond to regions of accessible chromatin.
11. The method of any one of claims 1-10, wherein said RNA molecules of said plurality of RNA molecules comprise messenger RNA (mRNA) molecules.
12. The method of any one of claims 1-11, wherein said sequence information comprises a first plurality of sequencing reads corresponding to said DNA molecules and a second plurality of sequencing reads corresponding to said RNA molecules.
13. The method of any one of claims 1-12, wherein said sequence information comprises a plurality of sequencing reads associated with individual cells or cell nuclei of said plurality of cells or cell nuclei.
14. The method of any one of claims 1-13, wherein (d) comprises determining a linked signature of said cell or cell nucleus of said plurality of cells or cell nuclei using said sequence information, which linked signature of said cell or cell nucleus links a first data

set comprising sequence information corresponding to DNA molecules of said cell or cell nucleus and a second data set comprising sequence information corresponding to RNA molecules of said cell or cell nucleus.

15. The method of any one of claims 1-14, wherein (d) comprises using said sequence information to cluster cells or cell nuclei of said plurality of cells or cell nuclei by gene expression signatures and/or by regions of accessible chromatin signatures.

16. The method of any one of claims 1-13, wherein (d) comprises (i) using said sequence information to cluster cells or cell nuclei of said plurality of cells or cell nuclei by regions of accessible chromatin signatures, (ii) using said sequence information to cluster cells or cell nuclei of said plurality of cells or cell nuclei by gene expression signatures, and (iii) using said sequence information and said cells or cell nuclei clustered by said gene expression signatures to further characterize said cells or cell nuclei clustered by said regions of accessible chromatin.

17. The method of any one of claims 1-13, wherein (d) comprises (i) using said sequence information to cluster cells or cell nuclei of said plurality of cells or cell nuclei by regions of accessible chromatin signatures, (ii) using said sequence information to cluster cells or cell nuclei of said plurality of cells or cell nuclei by gene expression signatures, and (iii) using said sequence information and said cells or cell nuclei clustered by said regions of accessible chromatin signatures to further characterize said cells or cell nuclei clustered by said gene expression signatures.

18. The method of any one of claims 1-17, wherein said plurality of cells or cell nuclei is derived from a sample comprising a tumor or suspected of comprising a tumor.

19. The method of claim 18, further comprising processing said sequence information corresponding to said RNA molecules and said DNA molecules with sequence information generated from a control sample.

20. The method of claim 18, wherein said sample is derived from a bodily fluid.

21. The method of claim 18, wherein said sample is derived from a biopsy.

22. The method of any one of claims 18-21, wherein said tumor is a B cell lymphoma tumor.

23. The method of any one of claims 18-22, further comprising using said sequence information to identify a presence of a tumor cell or cell nucleus in said sample.

24. The method of any one of claims 18-23, further comprising (e) using said sequence information to identify a cell type, a cell state, a tumor-specific gene expression pattern, or a tumor-specific differentially accessible region of chromatin in said sample.

25. The method of claim 24, further comprising, based at least in part on (e), identifying a therapeutic regimen for treatment of a subject from which said sample derives.
26. The method of claim 25, wherein said therapeutic regimen comprises administration of a therapeutically effective amount of an agent targeting one or more targets identified in said tumor-specific gene expression pattern or said tumor-specific differentially accessible region of chromatin.
27. The method of any one of claims 1-26, wherein said plurality of partitions comprises a plurality of droplets.
28. The method of any one of claims 1-27, wherein said plurality of cells or cell nuclei comprises a plurality of transposed nuclei.
29. The method of any one of claims 1-28, wherein said plurality of particles comprises a plurality of gel beads.
30. The method of claim 29, wherein said plurality of nucleic acid barcode molecules are releasably coupled to said plurality of particles.
31. The method of claim 30, wherein nucleic acid barcode molecules of said plurality of nucleic acid barcode molecules are releasable from said particles of said plurality of particles upon application of a stimulus.
32. The method of claim 31, wherein said stimulus is a chemical stimulus.
33. The method of claim 32, wherein said stimulus comprises a reducing agent.
34. The method of claim 29, wherein said plurality of nucleic acid barcode molecules are coupled to said plurality of particles via a plurality of labile moieties.
35. The method of any one of claims 1-34, further comprising generating said plurality of partitions using a microfluidic device.
36. The method of any one of claims 1-35, further comprising recovering said plurality of barcoded nucleic acid molecules from said at least said subset of said plurality of partitions.
37. The method of any one of claims 1-36, further comprising, prior to (b), lysing or permeabilizing said plurality of cells or cell nuclei to provide access to said plurality of nucleic acid molecules therein.
38. The method of any one of claims 1-37, further comprising, prior to (a), processing open chromatin structures of said plurality of cells or cell nuclei with a transposase to provide said plurality of DNA molecules.

39. The method of any one of claims 1-38, further comprising, within said at least said subset of said plurality of partitions, reverse transcribing said plurality of RNA molecules to provide a plurality of complementary DNA (cDNA) molecules.

40. A method for identifying a genetic feature, comprising:

(a) providing a first data set corresponding to regions of accessible chromatin of a plurality of deoxyribonucleic acid (DNA) molecules of a plurality of cells or cell nuclei and a second data set corresponding to a plurality of ribonucleic acid (RNA) molecules of said plurality of cells or cell nuclei, wherein said first data set comprises a first plurality of sequencing reads corresponding to sequences of said regions of accessible chromatin and a plurality of nucleic acid barcode sequences, and wherein said second data set comprises a second plurality of sequencing reads corresponding to sequences of said plurality of RNA molecules and said plurality of nucleic acid barcode sequences, wherein a cell or cell nucleus of said plurality of cells or cell nuclei corresponds to a nucleic acid barcode sequence of said plurality of nucleic acid barcode sequences;

(b) using said plurality of nucleic acid barcode sequences of said first data set and said second data set to identify first sequencing reads of said first plurality of sequencing reads and second sequencing reads of said second plurality of sequencing reads as corresponding to cells or cell nuclei of said plurality of cells or cell nuclei, thereby generating a third data set comprising sequence information corresponding to regions of accessible chromatin and RNA molecules associated with cells or cell nuclei of said plurality of cells or cell nuclei;

(c) using said sequence information to identify cell types of said cells or cell nuclei;

(d) using said sequence information corresponding to said RNA molecules to identify an expressed protein of a cell type of said cell types; and

(e) using said sequence information corresponding to said regions of accessible chromatin to identify a genetic feature corresponding to said expressed protein.

41. The method of claim 40, wherein said cell types are selected from the group consisting of monocytes, natural killer cells, B cells, T cells, granulocytes, dendritic cells, and stromal cells.

42. The method of claim 41, wherein said B cells are selected from the group consisting of replicating B cells, normal B cells, and tumor B cells.

43. The method of claim 41 or 42, wherein said B cells are selected from the group consisting of naïve B cells, memory B cells, plasmablast B cells, lymphoplasmacytoid cells, B-1 cells, regulatory B cells, and plasma B cells.

44. The method of any one of claims 41-43, wherein said T cells are selected from the group consisting of replicating T cells and normal T cells.

45. The method of any one of claims 41-44, wherein said T cells are selected from the group consisting of helper T cells, cytotoxic T cells, memory T cells, regulatory T cells, natural killer T cells, mucosal associated invariant T (MAIT) cells, gamma delta T cells, effector T cells, and naïve T cells.

46. The method of any one of claims 41-45, wherein said monocytes are selected from the group consisting of monocytes characterized by high level expression of CD14 cell surface receptors and monocytes characterized by high level expression of CD16 cell surface receptors.

47. The method of any one of claims 41-46, wherein said dendritic cells are selected from conventional dendritic cells and plasmacytoid dendritic cells.

48. The method of any one of claims 40-47, wherein said plurality of cells or cell nuclei comprises at least 500 cells or cell nuclei.

49. The method of claim 48, wherein said plurality of cells or cell nuclei comprises at least 1,000 cells or cell nuclei.

50. The method of claim 49, wherein said plurality of cells or cell nuclei comprises at least 10,000 cells or cell nuclei.

51. The method of any one of claims 40-50, wherein said genetic feature is a cis-regulatory element.

52. The method of claim 51, wherein said cis-regulatory element is a promoter.

53. The method of claim 51, wherein said cis-regulatory element is an enhancer.

54. The method of any one of claims 40-53, wherein said expressed protein is a cytokine.

55. The method of any one of claims 40-54, wherein said plurality of RNA molecules comprises a plurality of messenger RNA (mRNA) molecules.

56. The method of any one of claims 40-55, further comprising determining a linked signature of said cell or cell nucleus of said plurality of cells or cell nuclei using said first data set and said second data set, which linked signature of said cell or cell nucleus links a fourth data set comprising sequence information corresponding to DNA molecules of said cell or cell nucleus and a fifth data set comprising sequence information corresponding to RNA molecules of said cell or cell nucleus.

57. The method of any one of claims 40-56, wherein (c) comprises using said sequence information to cluster cells or cell nuclei of said plurality of cells or cell nuclei by regions of accessible chromatin signatures.

58. The method of any one of claims 40-56, wherein (c) comprises using said sequence information to cluster cells or cell nuclei of said plurality of cells or cell nuclei by gene expression signatures.

59. The method of any one of claims 40-56, wherein (c) comprises (i) using said sequence information to cluster cells or cell nuclei of said plurality of cells or cell nuclei by regions of accessible chromatin signatures, (ii) using said sequence information to cluster cells or cell nuclei of said plurality of cells or cell nuclei by gene expression signatures, and (iii) using said sequence information and said cells or cell nuclei clustered by said gene expression signatures to further characterize said cells or cell nuclei clustered by said regions of accessible chromatin.

60. The method of any one of claims 40-56, wherein (c) comprises (i) using said sequence information to cluster cells or cell nuclei of said plurality of cells or cell nuclei by regions of accessible chromatin signatures, (ii) using said sequence information to cluster cells or cell nuclei of said plurality of cells or cell nuclei by gene expression signatures, and (iii) using said sequence information and said cells or cell nuclei clustered by said regions of accessible chromatin signatures to further characterize said cells or cell nuclei clustered by said gene expression signatures.

61. The method of any one of claims 40-60, wherein said plurality of cells or cell nuclei is derived from a sample comprising a tumor or suspected of comprising a tumor.

62. The method of claim 61, wherein said sample is derived from a bodily fluid.

63. The method of claim 61, wherein said sample is derived from a biopsy.

64. The method of any one of claims 61-63, wherein said tumor is a B cell lymphoma tumor.

65. The method of any one of claims 61-64, further comprising using said sequence information to identify a presence of a tumor cell or cell nucleus in said sample.

66. The method of any one of claims 61-65, further comprising (f) using said sequence information to identify a cell type, a cell state, a tumor-specific gene expression pattern, or a tumor-specific differentially accessible region of chromatin in said sample.

67. The method of claim 66, further comprising, based at least in part on (f), identifying a therapeutic regimen for treatment of a subject from which said sample derives.

68. The method of claim 67, wherein said therapeutic regimen comprises administration of a therapeutically effective amount of an agent targeting one or more targets identified in

said tumor-specific gene expression pattern or said tumor-specific differentially accessible region of chromatin.

69. A system for identifying a genetic feature, comprising:

one or more databases comprising a first data set corresponding to regions of accessible chromatin of a plurality of deoxyribonucleic acid (DNA) molecules of a plurality of cells or cell nuclei and a second data set corresponding to a plurality of ribonucleic acid (RNA) molecules of said plurality of cells or cell nuclei, wherein said first data set comprises a first plurality of sequencing reads corresponding to sequences of said regions of accessible chromatin and a plurality of nucleic acid barcode sequences, and wherein said second data set comprises a second plurality of sequencing reads corresponding to sequences of said plurality of RNA molecules and said plurality of nucleic acid barcode sequences, wherein a cell or cell nucleus of said plurality of cells or cell nuclei corresponds to a nucleic acid barcode sequence of said plurality of nucleic acid barcode sequences; and

one or more computer processors operably coupled to said one or more databases, wherein said one or more computer processors are individually or collectively programmed to:

(i) use said plurality of nucleic acid barcode sequences of said first data set and said second data set to identify first sequencing reads of said first plurality of sequencing reads and second sequencing reads of said second plurality of sequencing reads as corresponding to cells or cell nuclei of said plurality of cells or cell nuclei, thereby generating a third data set comprising sequence information corresponding to regions of accessible chromatin and RNA molecules associated with cells or cell nuclei of said plurality of cells or cell nuclei;

(ii) use said sequence information to identify cell types of said cells or cell nuclei;

(iii) use said sequence information corresponding to said RNA molecules to identify an expressed protein of a cell type of said cell types; and

(iv) use said sequence information corresponding to said regions of accessible chromatin to identify a genetic feature corresponding to said expressed protein.

70. The system of claim 69, wherein said cell types are selected from the group consisting of monocytes, natural killer cells, B cells, T cells, granulocytes, dendritic cells, and stromal cells.

71. The system of claim 70, wherein said B cells are selected from the group consisting of replicating B cells, normal B cells, and tumor B cells.

72. The system of claim 70 or 71, wherein said B cells are selected from the group consisting of naïve B cells, memory B cells, plasmablast B cells, lymphoplasmacytoid cells, B-1 cells, regulatory B cells, and plasma B cells.

73. The system of any one of claims 70, wherein said T cells are selected from the group consisting of replicating T cells and normal T cells.

74. The system of claim 70 or claim 73, wherein said T cells are selected from the group consisting of helper T cells, cytotoxic T cells, memory T cells, regulatory T cells, natural killer T cells, mucosal associated invariant T (MAIT) cells, gamma delta T cells, effector T cells, and naïve T cells.

75. The system of any one of claims 70-74, wherein said monocytes are selected from the group consisting of monocytes characterized by high level expression of CD14 cell surface receptors and monocytes characterized by high level expression of CD16 cell surface receptors.

76. The system of any one of claims 70-75, wherein said dendritic cells are selected from conventional dendritic cells and plasmacytoid dendritic cells.

77. The system of any one of claims 69-76, wherein said plurality of cells or cell nuclei comprises at least 500 cells or cell nuclei.

78. The system of claim 77, wherein said plurality of cells or cell nuclei comprises at least 1,000 cells or cell nuclei.

79. The system of claim 78, wherein said plurality of cells or cell nuclei comprises at least 10,000 cells or cell nuclei.

80. The system of any one of claims 69-79, wherein said genetic feature is a cis-regulatory element.

81. The system of claim 80, wherein said cis-regulatory element is a promoter.

82. The system of claim 80, wherein said cis-regulatory element is an enhancer.

83. The system of any one of claims 69-82, wherein said expressed protein is a cytokine.

84. The system of any one of claims 69-83, wherein said plurality of RNA molecules comprises a plurality of messenger RNA (mRNA) molecules.

85. The system of any one of claims 69-84, wherein said one or more computer processors are individually or collectively programmed to determine a linked signature of said cell or cell nucleus of said plurality of cells or cell nuclei using said first data set and said second data set, which linked signature of said cell or cell nucleus links a fourth data set

comprising sequence information corresponding to DNA molecules of said cell or cell nucleus and a fifth data set comprising sequence information corresponding to RNA molecules of said cell or cell nucleus.

86. The system of any one of claims 69-85, wherein said one or more computer processors are individually or collectively programmed to use said sequence information in (ii) to cluster cells or cell nuclei of said plurality of cells or cell nuclei by regions of accessible chromatin signatures.

87. The system of any one of claims 69-85, wherein said one or more computer processors are individually or collectively programmed to use said sequence information in (ii) to cluster cells or cell nuclei of said plurality of cells or cell nuclei by gene expression signatures.

88. The system of any one of claims 69-85, wherein in (ii) said one or more computer processors are individually or collectively programmed to use said sequence information (1) to cluster cells or cell nuclei of said plurality of cells or cell nuclei by regions of accessible chromatin signatures, (2) to cluster cells or cell nuclei of said plurality of cells or cell nuclei by gene expression signatures, and (3) with said cells or cell nuclei clustered by said gene expression signatures to further characterize said cells or cell nuclei clustered by said regions of accessible chromatin.

89. The system of any one of claims 69-85, wherein in (ii) said one or more computer processors are individually or collectively programmed to use said sequence information (1) to cluster cells or cell nuclei of said plurality of cells or cell nuclei by regions of accessible chromatin signatures, (2) to cluster cells or cell nuclei of said plurality of cells or cell nuclei by gene expression signatures, and (3) with said cells or cell nuclei clustered by said regions of accessible chromatin signatures to further characterize said cells or cell nuclei clustered by said gene expression signatures.

90. The system of any one of claims 69-89, wherein said one or more computer processors are individually or collectively further programmed to generate an output that relates to detecting a disease or condition in said sample, said output comprising said regions of accessible chromatin signatures and gene expression signatures, presence or absence of said disease or condition; or a level of progression of said disease or condition.

91. The system of any one of claims 69-90, wherein said plurality of cells or cell nuclei is derived from a sample comprising a tumor or suspected of comprising a tumor.

92. The system of claim 91, wherein said sample is derived from a bodily fluid.

93. The system of claim 91, wherein said sample is derived from a biopsy.

94. The system of any one of claims 91, wherein said tumor is a B cell lymphoma tumor.

95. The system of any one of claims 90-94, wherein said one or more computer processors are individually or collectively further programmed to use said sequence information to identify a presence of a tumor cell or cell nucleus in said sample.

96. The system of any one of claims 90-95, wherein said one or more computer processors are individually or collectively further programmed to compare said sequence information to sequence information from a control sample.

97. The system of any one of claims 90-96, wherein said one or more computer processors are individually or collectively further programmed to use said sequence information to identify a cell type, a cell state, a tumor-specific gene expression pattern, or a tumor-specific differentially accessible region of chromatin in said sample.

98. The system of any one of claims 90-97, wherein said one or more computer processors are individually or collectively further programmed to, based at least in part on said use of said sequence information to identify said cell type, said cell state, said tumor-specific gene expression pattern, or said tumor-specific differentially accessible region of chromatin in said sample, identify a therapeutic regimen for treatment of a subject from which said sample derives.

99. The system of claim 98, wherein said therapeutic regimen comprises administration of a therapeutically effective amount of an agent targeting one or more targets identified in said tumor-specific gene expression pattern or said tumor-specific differentially accessible region of chromatin.

100. The system of any one of claims 69-99, wherein said system is used to monitor efficacy of said therapeutic regimen for treatment.

101. A method for determining a condition of a sample comprising:

a) generating (i) a first data set comprising sequencing information corresponding to regions of accessible chromatin of a plurality of deoxyribonucleic acid (DNA) molecules of a cell or cell nucleus of said sample, (ii) a second data set comprising sequencing information corresponding to a plurality of ribonucleic acid (RNA) molecules of said cell or cell nucleus, and (iii) a linked signature of said cell or cell nucleus using said first data set and said second data set;

b) using said linked signature of said cell or cell nucleus and a control linked signature of a control cell or cell nucleus of a control sample to determine one or more regions of accessible chromatin of said plurality of DNA molecules or one or more genes expressed from said plurality of RNA molecules indicative of said condition.

102. The method of claim 101, further comprising c) determining a level of said one or more regions of accessible chromatin and/or said one or more genes expressed determined in b) indicative of said condition in one or more samples of an individual suspected of having said condition.

103. The method of claim 102, further comprising providing a diagnostic evaluation of said condition, a prognostic evaluation of said condition, monitoring of said condition, and/or management of said condition.

104. The method of claim 102 or claim 103, wherein a gene associated with said one or more regions of accessible chromatin and/or one or more genes expressed determined in b) is identified as a target of a therapeutic regimen for treatment of said condition.

105. The method of claim 104, further comprising administering a therapeutically effective amount of an agent to a subject targeting said target, wherein said sample is derived from said subject.

106. The method of claim 105, further comprising determining an efficacy of said agent in said subject.

107. The method of claim 106, wherein determining said efficacy comprises detecting a presence or absence of a response to said agent by said subject, wherein said response comprises a quantity, degree, or extent of response following administration of a first dose or a subsequent dose of said agent.

108. The method of claim 107, wherein said response comprises a differential in gene expression and/or chromatin accessibility of said target between prior to and after administration of said agent.

109. The method of any one of claims 101-108, wherein said sample is from a subject having a tumor or suspected of having a tumor.

110. The method of any one of claims 101-109, wherein said condition is a tumor, cancer, malignancy, neoplasm, or other proliferative disease or disorder.

111. The method of claim 110, wherein said condition is a B cell malignancy.

112. The method of claim 111, wherein said B cell malignancy is B cell lymphoma.

113. The method of any one of claims 101-112, wherein said sample is derived from a bodily fluid.

114. The method of any one of claims 101-112, wherein said sample is derived from a biopsy.

115. The method of any one of claims 101-114, wherein a) comprises providing a plurality of sequencing reads corresponding to sequences of said plurality of DNA molecules and

said plurality of RNA molecules, wherein said sequencing reads each correspond to said cell or cell nucleus via a nucleic acid barcode sequence.

116. The method of any one of claims 101-114, further comprising, prior to b), clustering a plurality of cells or cell nuclei of said sample by respective regions of accessible chromatin signatures, by respective genes expressed, and/or by respective linked signatures of said plurality of cells or cell nuclei.

117. The method of claim 116, wherein said plurality of cells or cell nuclei are clustered by cell types selected from the group consisting of monocytes, natural killer cells, B cells, T cells, granulocytes, dendritic cells, and stromal cells.

118. The method of claim 117, wherein said B cells are selected from the group consisting of replicating B cells, normal B cells, and tumor B cells.

119. The method of claim 117 or claim 118, wherein said B cells are selected from the group consisting of naïve B cells, memory B cells, plasmablast B cells, lymphoplasmacytoid cells, B-1 cells, regulatory B cells, and plasma B cells.

120. The method of any one of claims 117, wherein said T cells are selected from the group consisting of replicating T cells and normal T cells.

121. The method of claim 117 or claim 120, wherein said T cells are selected from the group consisting of helper T cells, cytotoxic T cells, memory T cells, regulatory T cells, natural killer T cells, mucosal associated invariant T (MAIT) cells, gamma delta T cells, effector T cells, and naïve T cells.

122. The method of any one of claims 102-121, further comprising c) monitoring a level of said one or more regions of accessible chromatin and/or said one or more genes expressed determined in b) indicative of said condition in the individual.

123. The method of any one of claims 101-122, further comprising, prior to a), generating a plurality of tagged DNA fragments.

124. The method of any one of claims 101-122, further comprising, prior to a), generating a plurality of barcoded nucleic acid molecules, wherein said plurality of barcoded nucleic acid molecules comprises (i) a first subset comprising sequences corresponding to regions of accessible chromatin of said plurality of deoxyribonucleic acid (DNA) molecules of said cell or cell nucleus of said sample and (ii) a second subset comprising sequences corresponding to said ribonucleic acid (RNA) molecules of said cell or cell nucleus.

125. The method of claim 124, wherein said generating is performed within a plurality of partitions.

126. The method of claim 125, further comprising sequencing said plurality of barcoded nucleic acid molecules.

127. The method of any one of claims 101-126, wherein:

said first data set is generated by sequencing a first plurality of barcoded nucleic acid molecules comprising sequences corresponding to regions of accessible chromatin of said plurality of deoxyribonucleic acid (DNA) molecules of said cell or cell nucleus of said sample; and

said second data set is generated by sequencing a second plurality of barcoded nucleic acid molecules comprising sequences of said ribonucleic acid (RNA) molecules of said cell or cell nucleus.

128. The method of any one of claims 101-127, further comprising processing said first data set and/or said second data set to generate a filtered first data set and/or a filtered second data set.

129. The method of claim 128, wherein said filtered first data set is filtered using motif enrichment.

130. The method of claim 128 or claim 129, wherein said filtered second data set is filtered using differential expression analysis.

131. The method of any one of claims 101-130, further comprising processing said first data set and/or said second data set to generate a linkage significance.

132. The method of any one of claims 128-131, further comprising processing said first filtered data set and/or said second filtered data set to generate an enrichment score.

133. The method of any one of claims 128-132, wherein said filtered first data set and said filtered second data set is used to generate a transcription factor-target gene network.

134. The method of claim 133, wherein a gene from said transcription factor-target gene network is identified as a target of a therapeutic regimen for treatment of said condition.

135. The method of claim 134, wherein said target is a transcription factor.

136. An *in vitro* method of preparing a biological sample, comprising:

(a) processing open chromatin structures of T-cells and/or B-cells from said biological sample with a transposase to provide a plurality of DNA molecules;

(b) generating a first plurality of barcoded nucleic acid molecules comprising said plurality of DNA molecules processed in (a);

(c) generating a second plurality of barcoded nucleic acid molecules comprising a plurality of nucleic acids comprising mRNA sequences or derivatives thereof from said T-cells and/or B-cells from said biological sample; and

(d) generating a first and second sequencing library from said first and second plurality of barcoded nucleic acid molecules, respectively, to determine a linked signature of a cell of said T-cells and/or B-cells.

137. The method of claim 136, wherein step (b) and/or step (c) is performed within a plurality of partitions.

138. The method of claim 136 or claim 137, further comprising determining a significance level for said linked signature determined in step (d).

139. The method of any one of claims 136-138, wherein step (c) comprises reverse transcribing said plurality of mRNA sequences from said T-cells and/or B-cells from said biological sample to provide a plurality of complementary DNA (cDNA) molecules, and said second plurality of barcoded nucleic acid molecules comprises said cDNA molecules.

140. The method of any one of claims 136-138, wherein step (c) comprises barcoding 3' ends of said mRNA were barcoded.

141. The method of any one of claims 136-139, further comprising encapsulating single nuclei of said T-cells and/or B-cells in droplets prior to step (b).

142. The method of any one of claims 136-141, further comprising (e) determining from said first and second sequencing libraries a presence, absence, and/or level of said one or more linked signatures correlated with a condition.

143. The method of claim 142, wherein said condition is a tumor, cancer, malignancy, neoplasm, or other proliferative disease or disorder.

144. The method of claim 143, wherein said condition is a B cell malignancy.

145. The method of claim 144, wherein said B cell malignancy is B cell lymphoma.

146. The method of any one of claims 136-145, wherein said linked signature of a cell or cell nucleus of said T-cells and/or B-cells is compared to a control linked signature of a control cell or cell nucleus of a control sample.

147. The method of any one of claims136-146, further comprising providing said biological sample isolated and obtained from an individual prior to step (a).

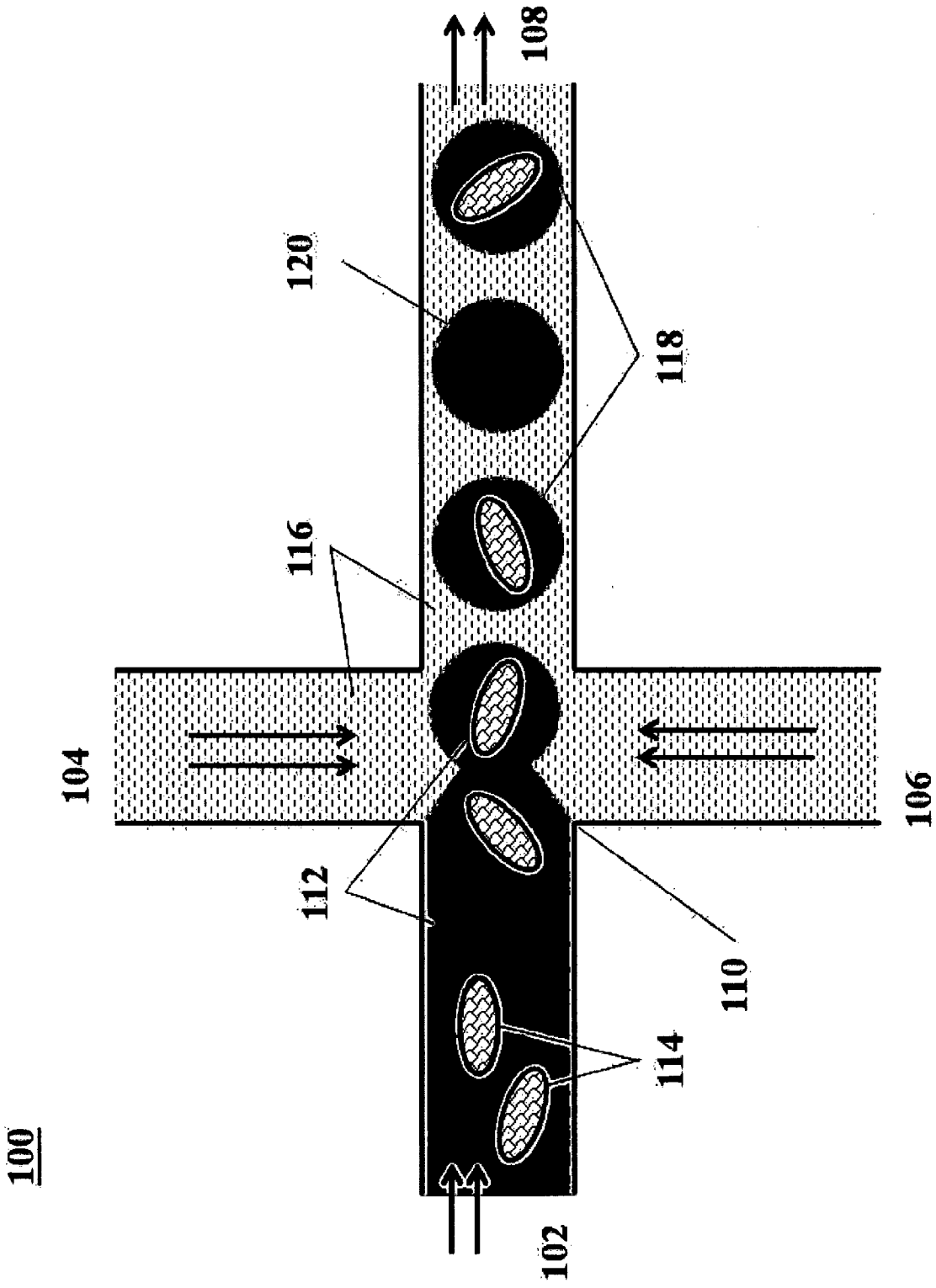


FIG. 1

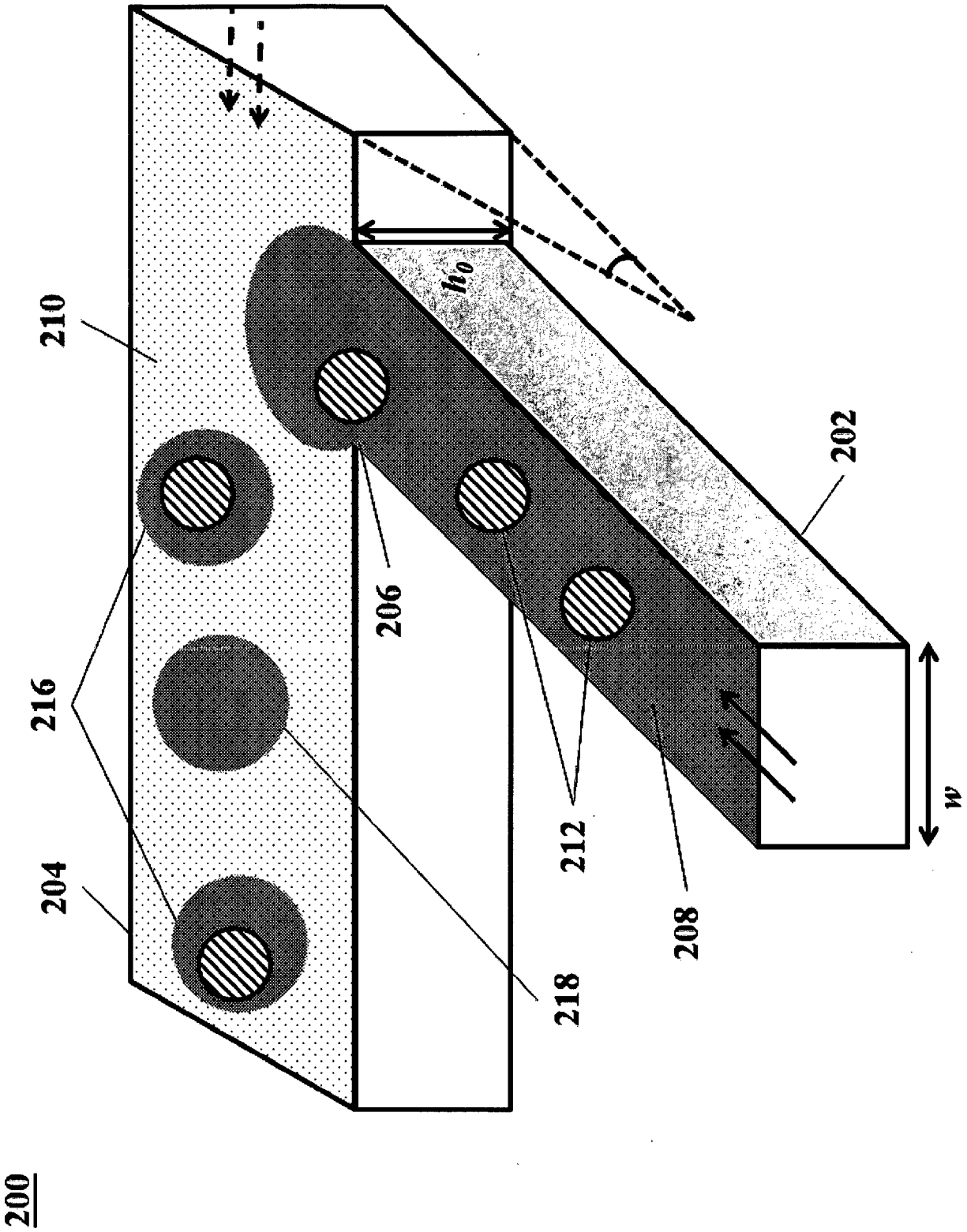


FIG. 2

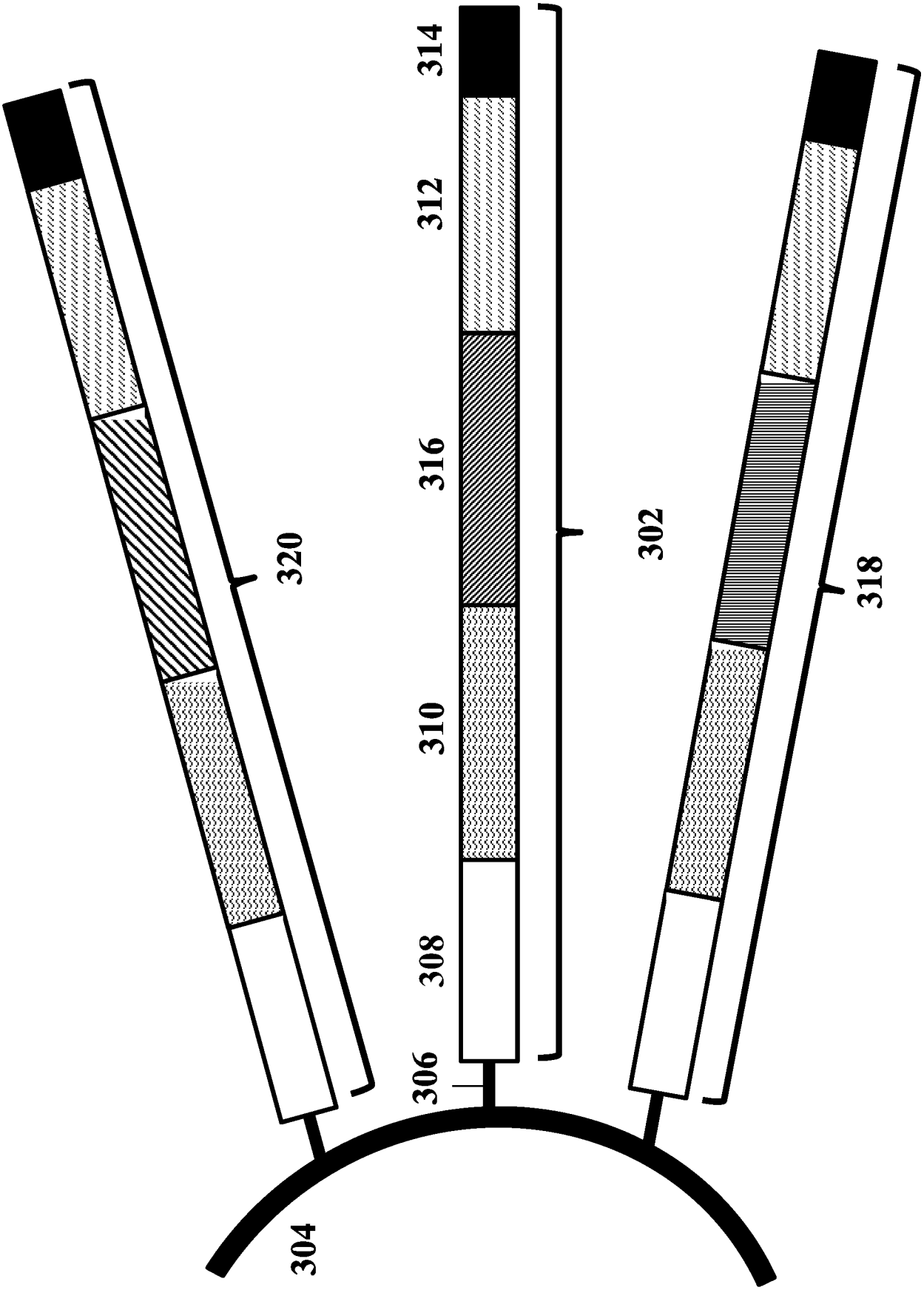


FIG. 3

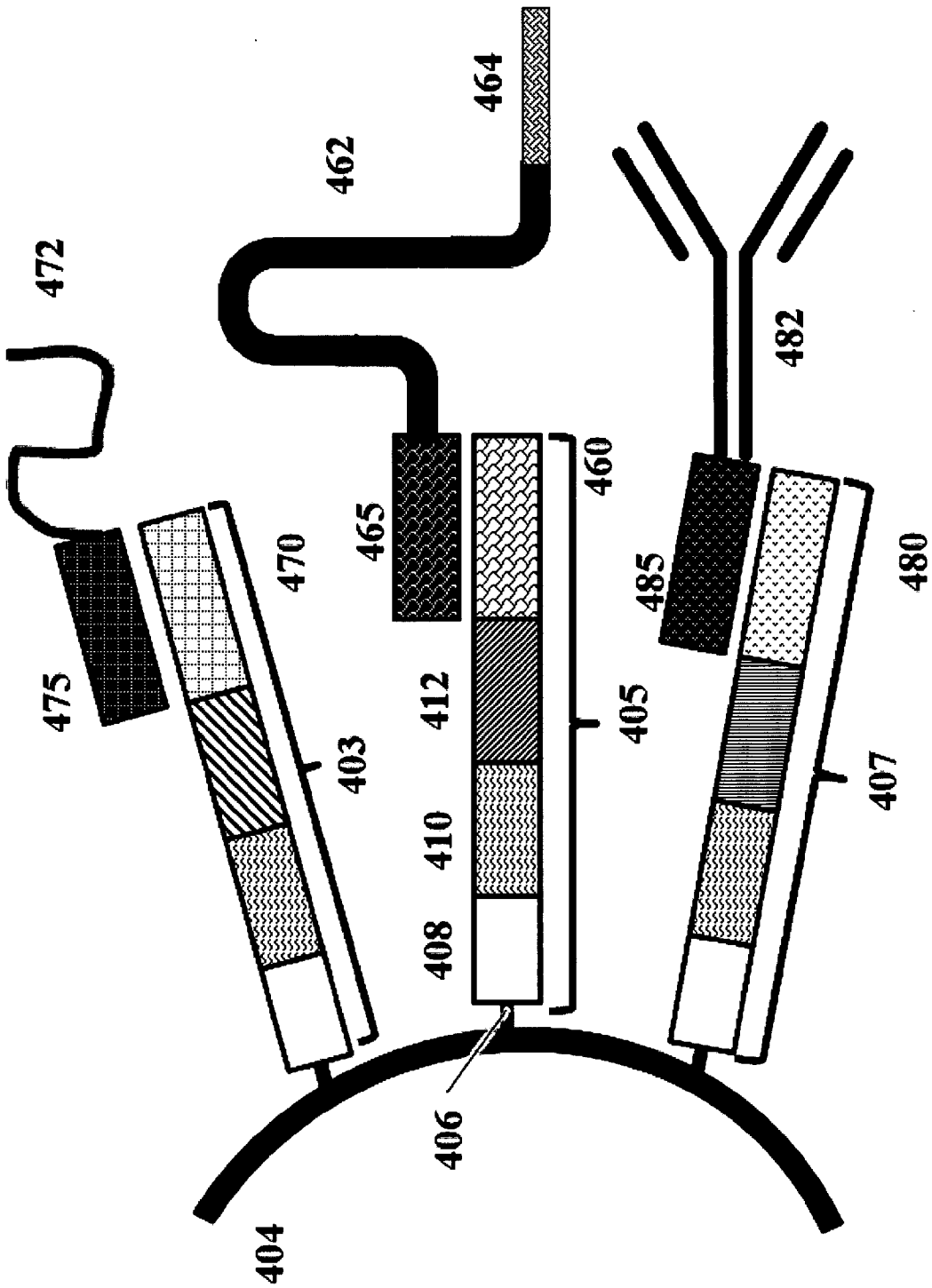

472
462
464
482
460
465
485
480
470
475
403
412
405
407
410
408
406
404

FIG. 4

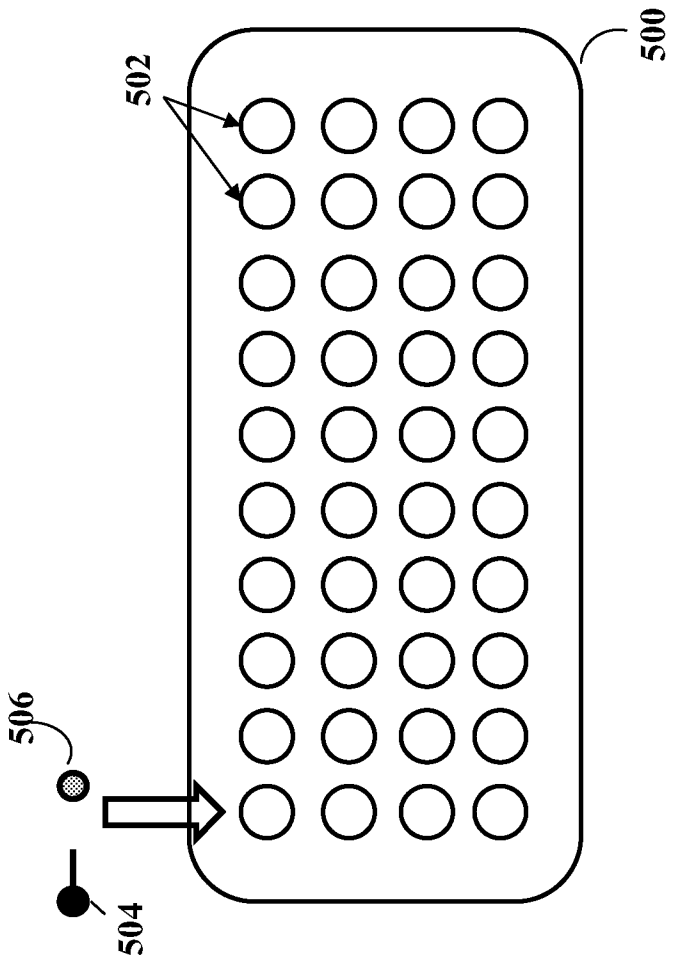


FIG. 5

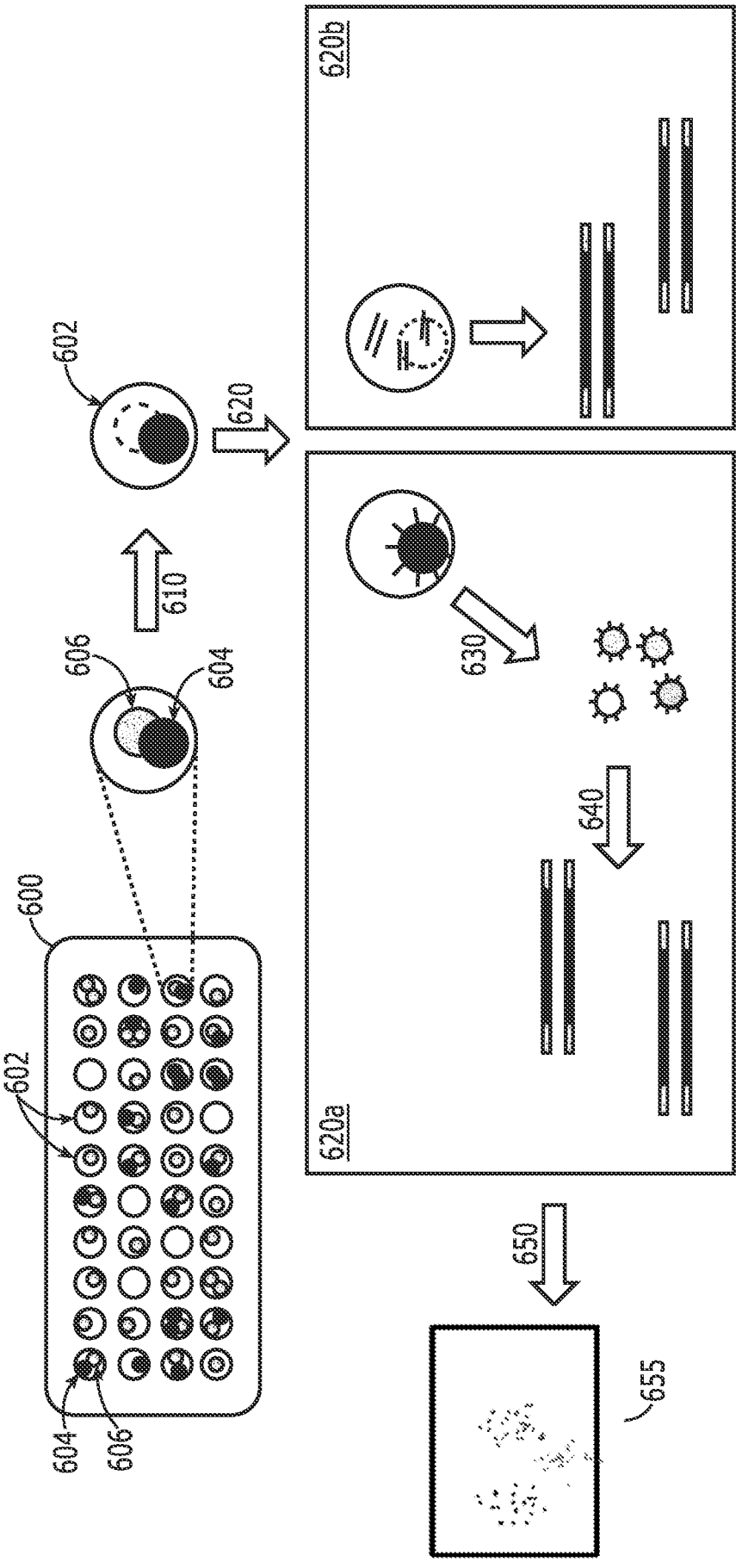


FIG. 6

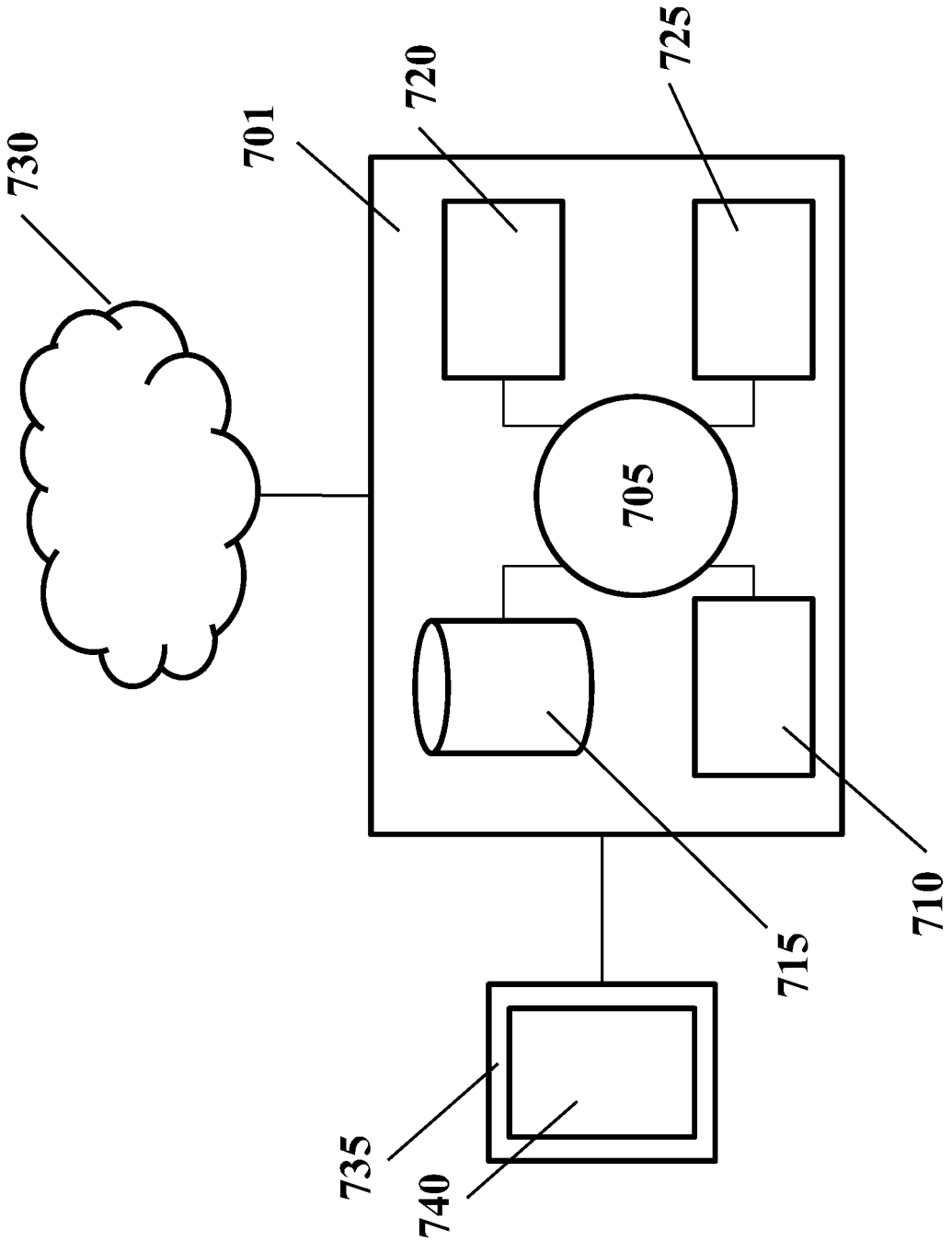


FIG. 7

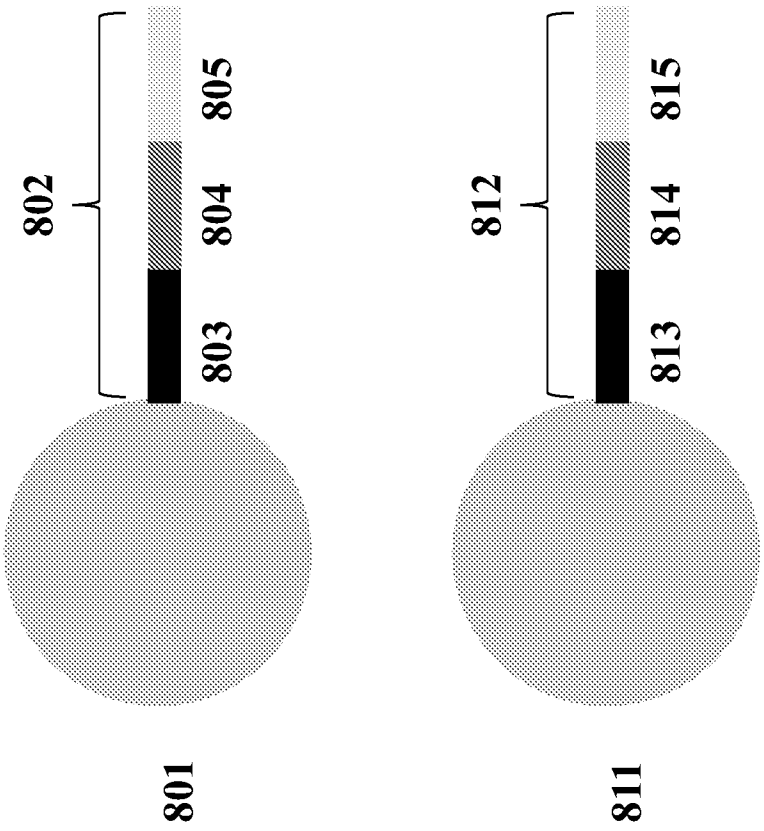

802
803
804
805
801
812
813
814
815
811

FIG. 8A

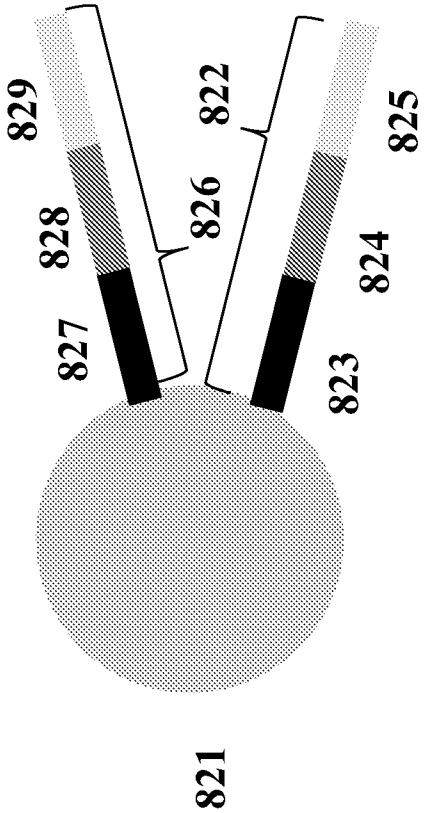

829
828
827
826
822
825
824
823
821

FIG. 8B

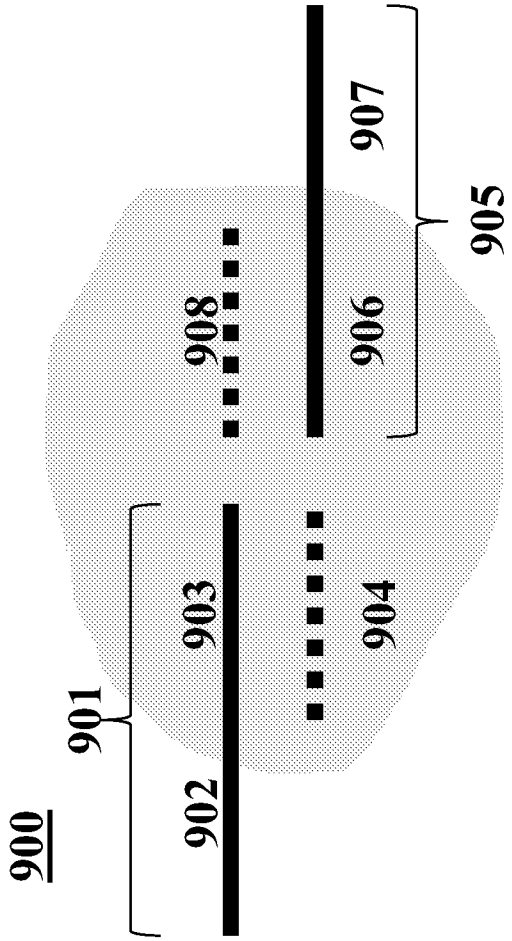


FIG. 9

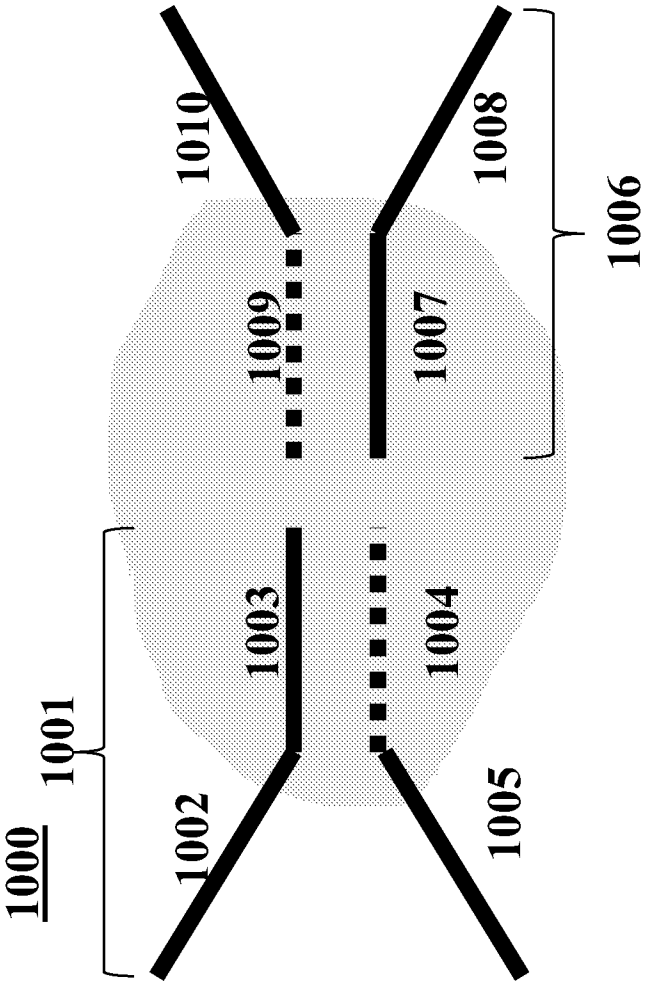


FIG. 10

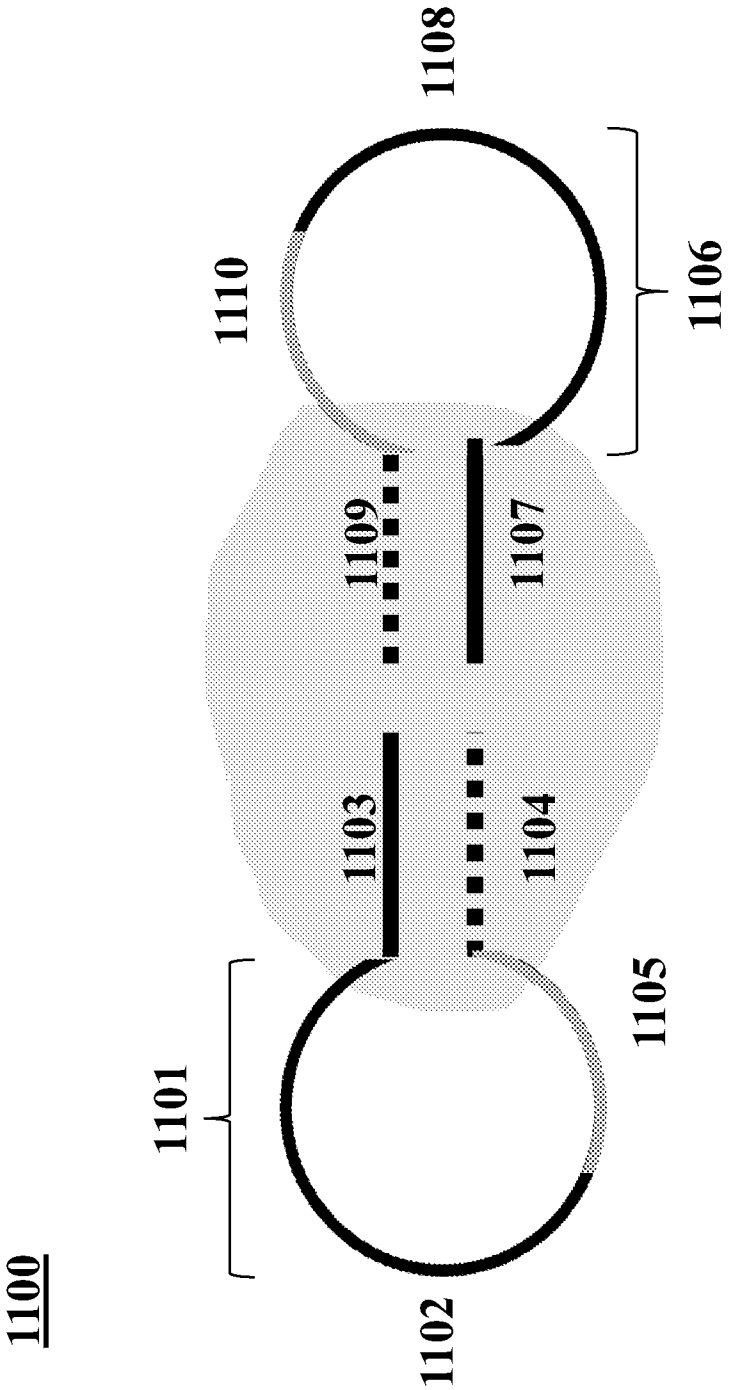


FIG. 11

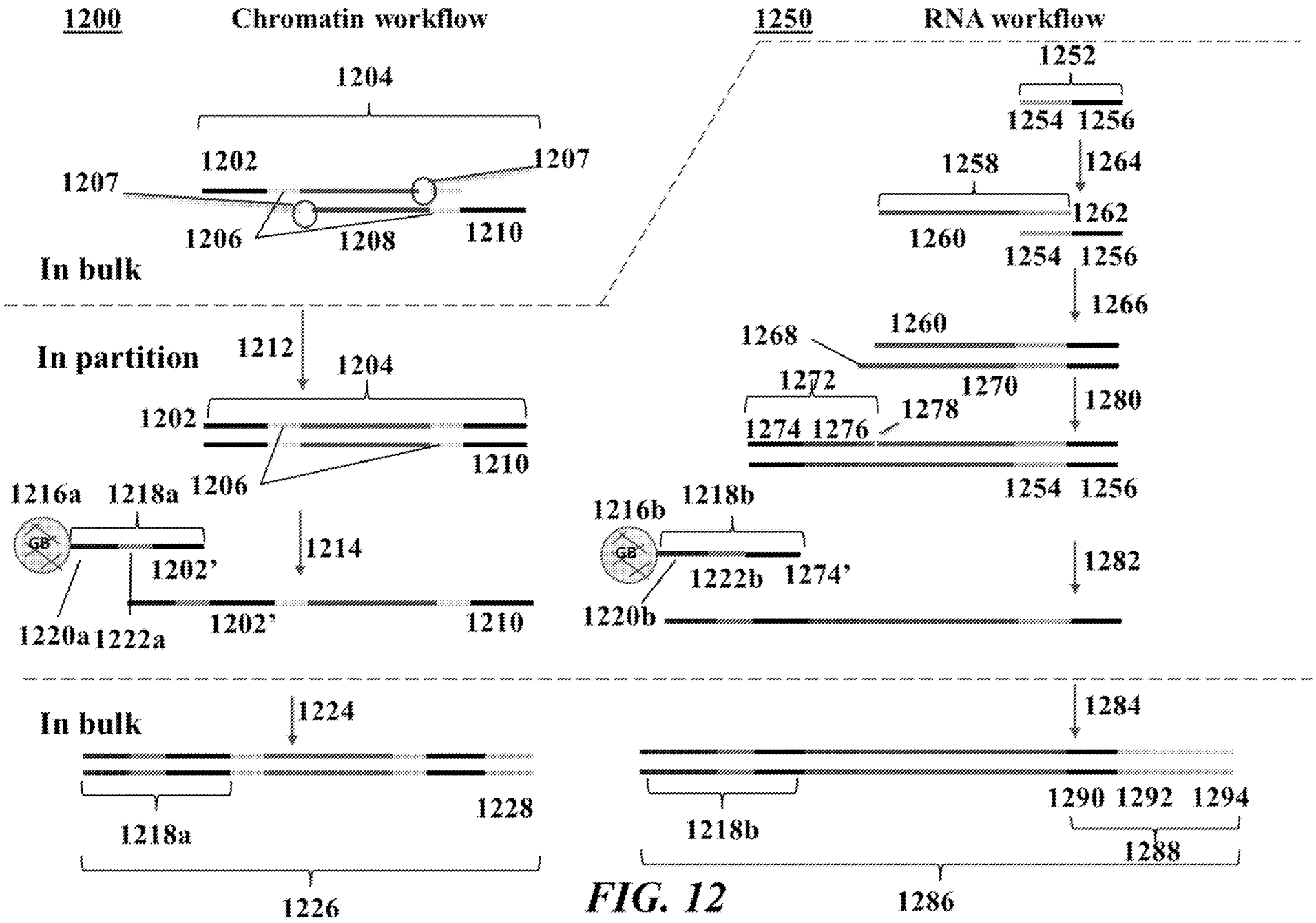

1200
Chromatin workflow
1204
1202
1207
1207
1206
1208
1210
In bulk
In partition
1212
1204
1202
1210
1206
1216a
1218a
GB
1214
1202'
1220a
1222a
1202'
1210
1250
RNA workflow
1252
1254
1256
1258
1264
1262
1260
1254
1256
1266
1268
1260
1272
1270
1280
1278
1274
1276
1254
1256
1216b
1218b
GB
1282
1222b
1274'
1220b
In bulk
1224
1228
1218a
1226
1284
1290
1292
1294
1218b
1288
1286

FIG. 12

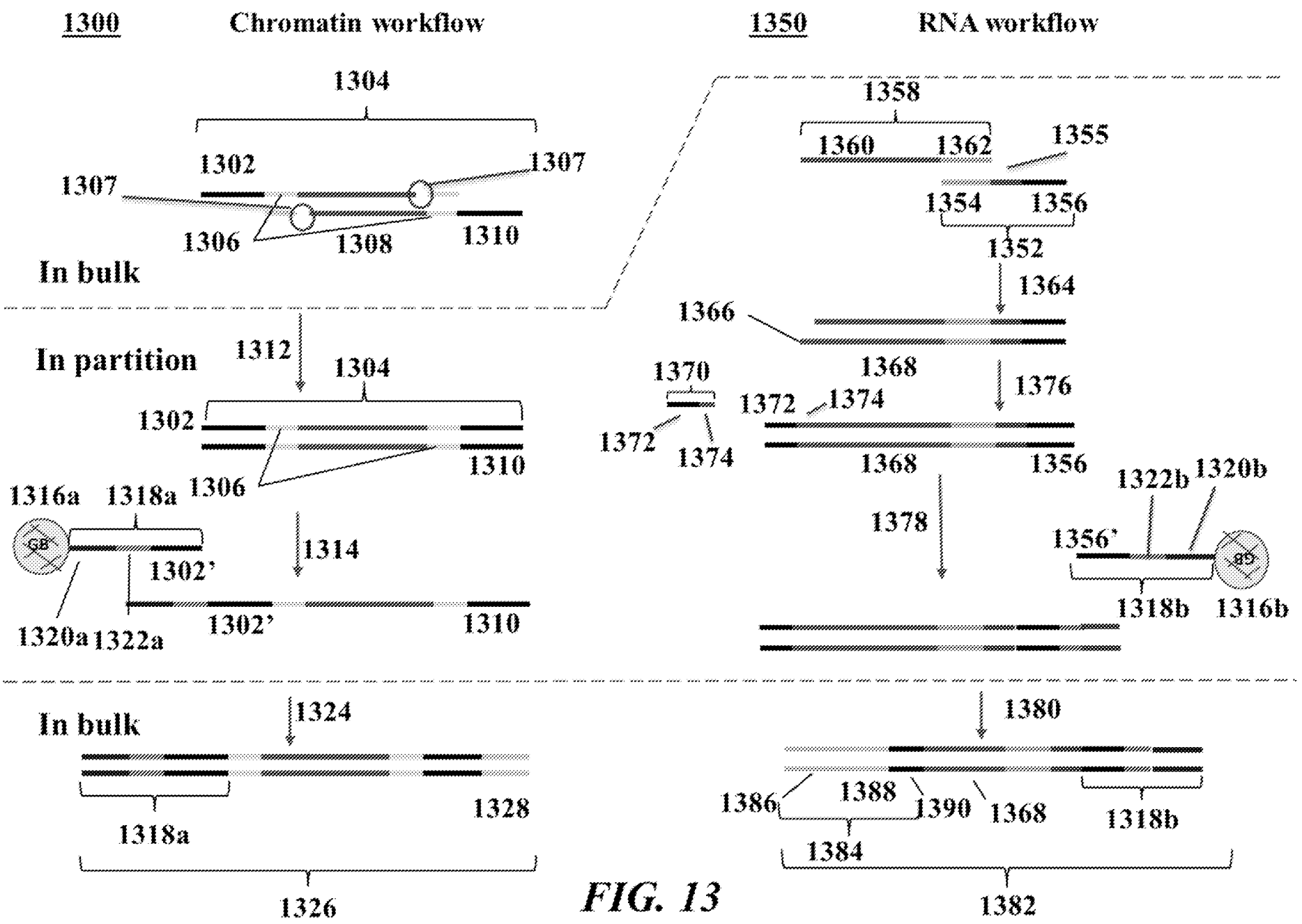

1300
Chromatin workflow
1304
1302
1307
1307
1306
1308
1310
In bulk
In partition
1312
1304
1302
1310
1306
1316a
1318a
GB
1314
1302'
1320a
1322a
1302'
1310
1350
RNA workflow
1358
1360
1362
1355
1354
1356
1352
1364
1366
1368
1370
1376
1372
1374
1372
1374
1368
1356
1322b
1320b
1378
1356'
BC
1318b
1316b
In bulk
1324
1328
1318a
1326
1380
1386
1388
1390
1368
1318b
1384
1382

FIG. 13

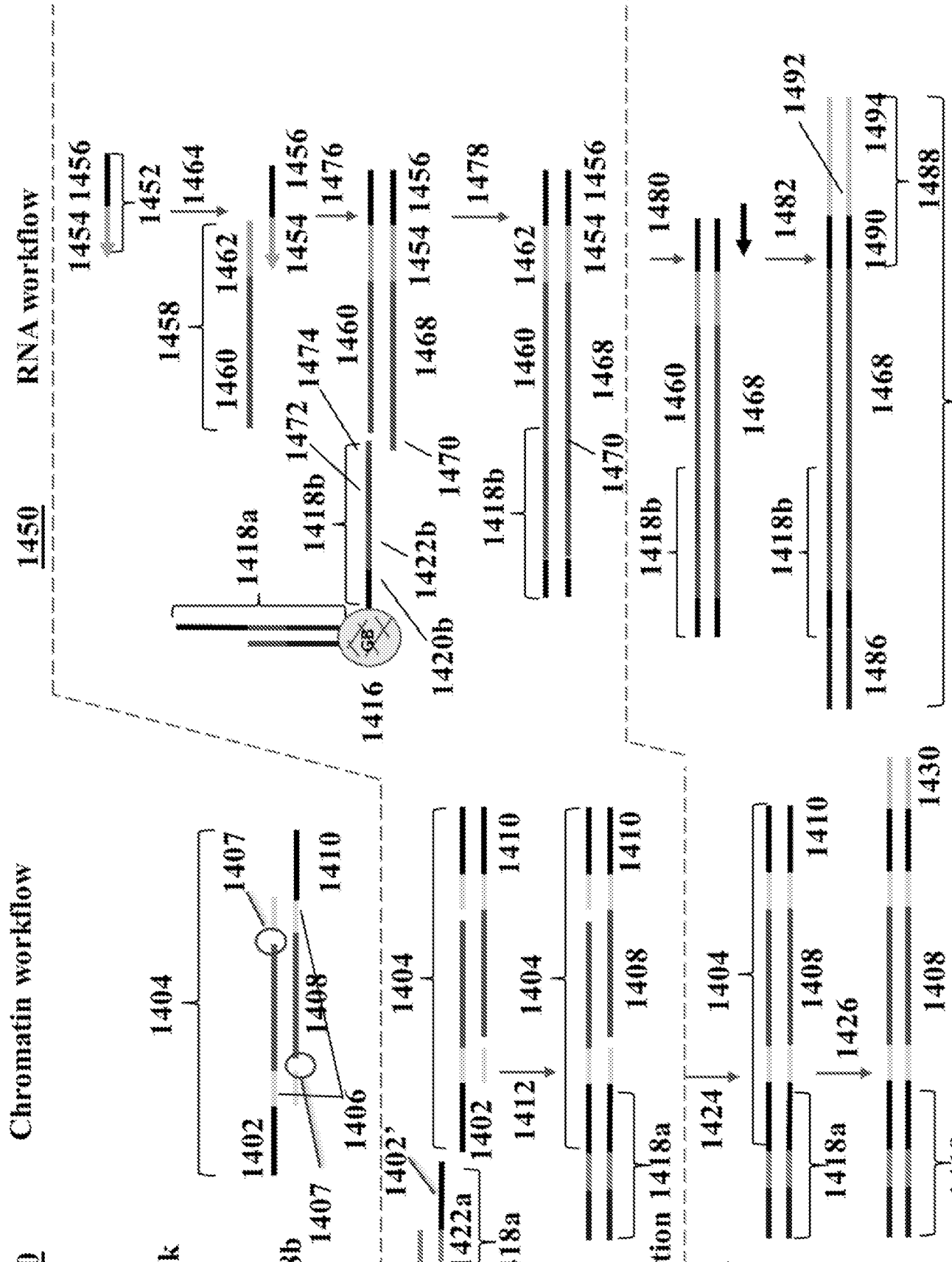

1400
Chromatin workflow
In bulk
1404
1407
1402
1410
1408
1406
1418b
1407
1402'
1404
1422a
1402
1410
1418a
1412
1404
1416
1420a
In partition 1418a
1408
1410
1424
1404
In bulk
1408
1410
1426
1418a
1430
1408
1428
1418a
1450
RNA workflow
1454 1456
1452
1464
1458
1462
1460
1454 1456
1476
1474
1460
1472
1468
1454 1456
1418b
1470
1422b
1418a
1420b
1416
1478
1462
1460
1454 1456
1418b
1468
1470
1480
1460
1418b
1468
1482
1492
1490
1494
1418b
1488
1468
1486
1484

FIG. 14

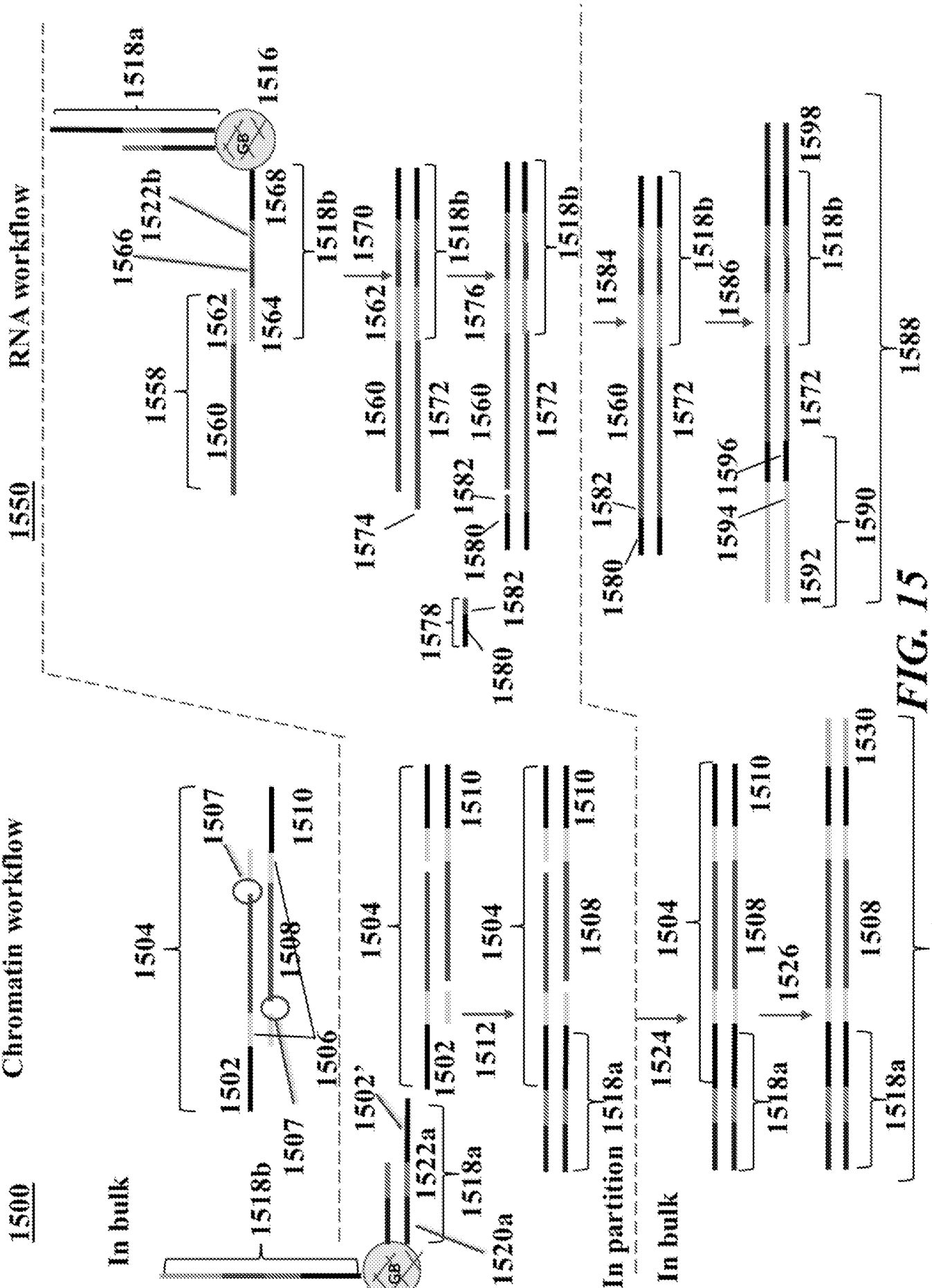

1500
Chromatin workflow
In bulk
In partition 1518a
In bulk
1550
RNA workflow
GB
1504
1507
1510
1508
1506
1502
1502'
1518b
1522a
1518a
1520a
1512
1524
1526
1528
1530
1516
1518a
1522b
1566
1568
1564
1562
1558
1560
1570
1572
1574
1576
1578
1580
1582
1584
1586
1588
1590
1592
1594
1596
1598

FIG. 15

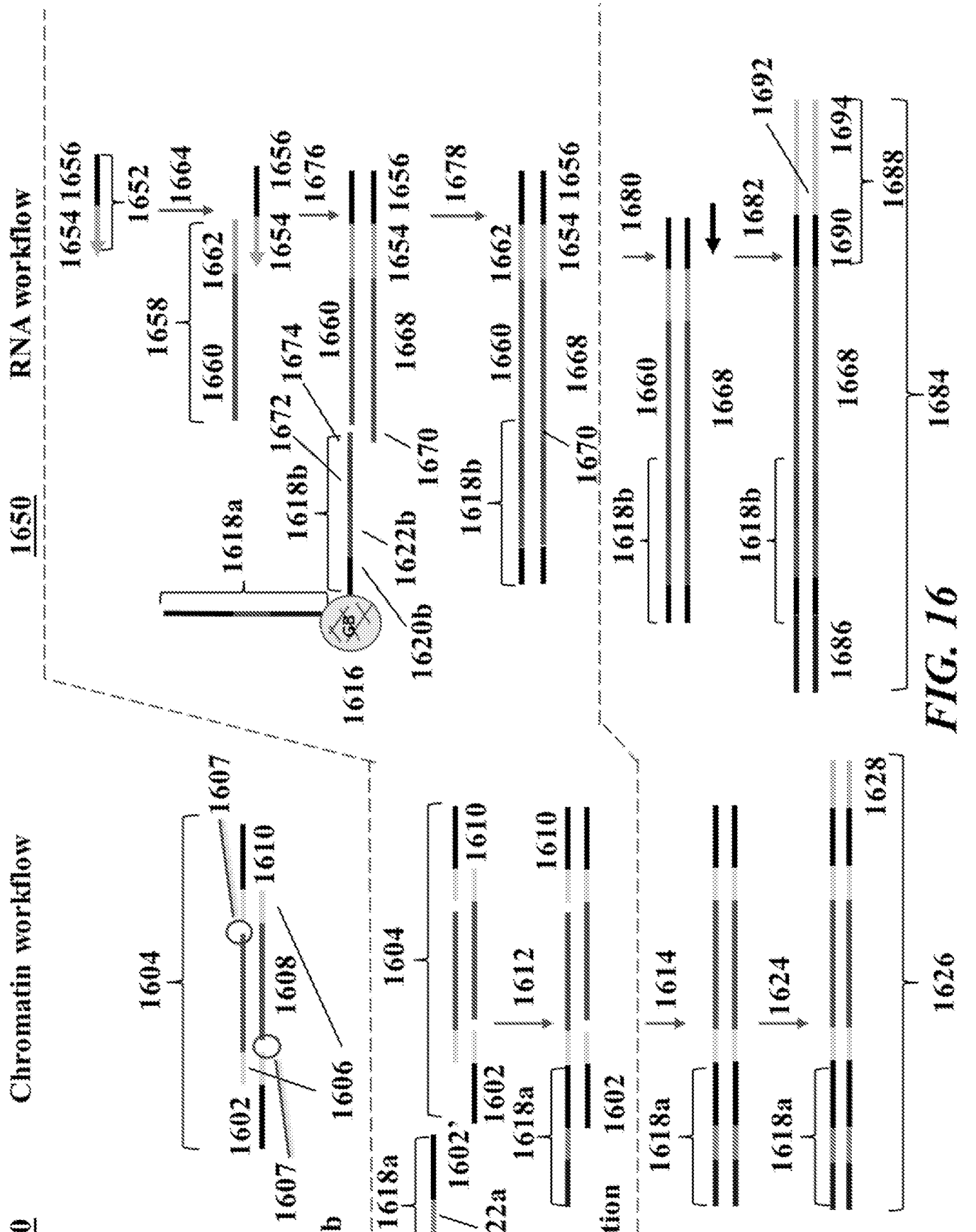

1600
Chromatin workflow
In bulk
1604
1607
1610
1608
1606
1602
1607
1618b
1618a
1602'
1622a
1620a
1616
GB
1604
1610
1602
1612
1610
In partition
1602
1618a
1614
1618a
1624
In bulk
1618a
1628
1626
1650
RNA workflow
1654 1656
1652
1664
1658
1662
1660
1660
1654 1656
1676
1674
1672
1618b
1618a
1622b
1620b
1616
GB
1670
1668
1654 1656
1678
1660
1662
1618b
1670
1668
1654 1656
1680
1660
1618b
1668
1682
1692
1618b
1686
1668
1690
1694
1688
1684

FIG. 16

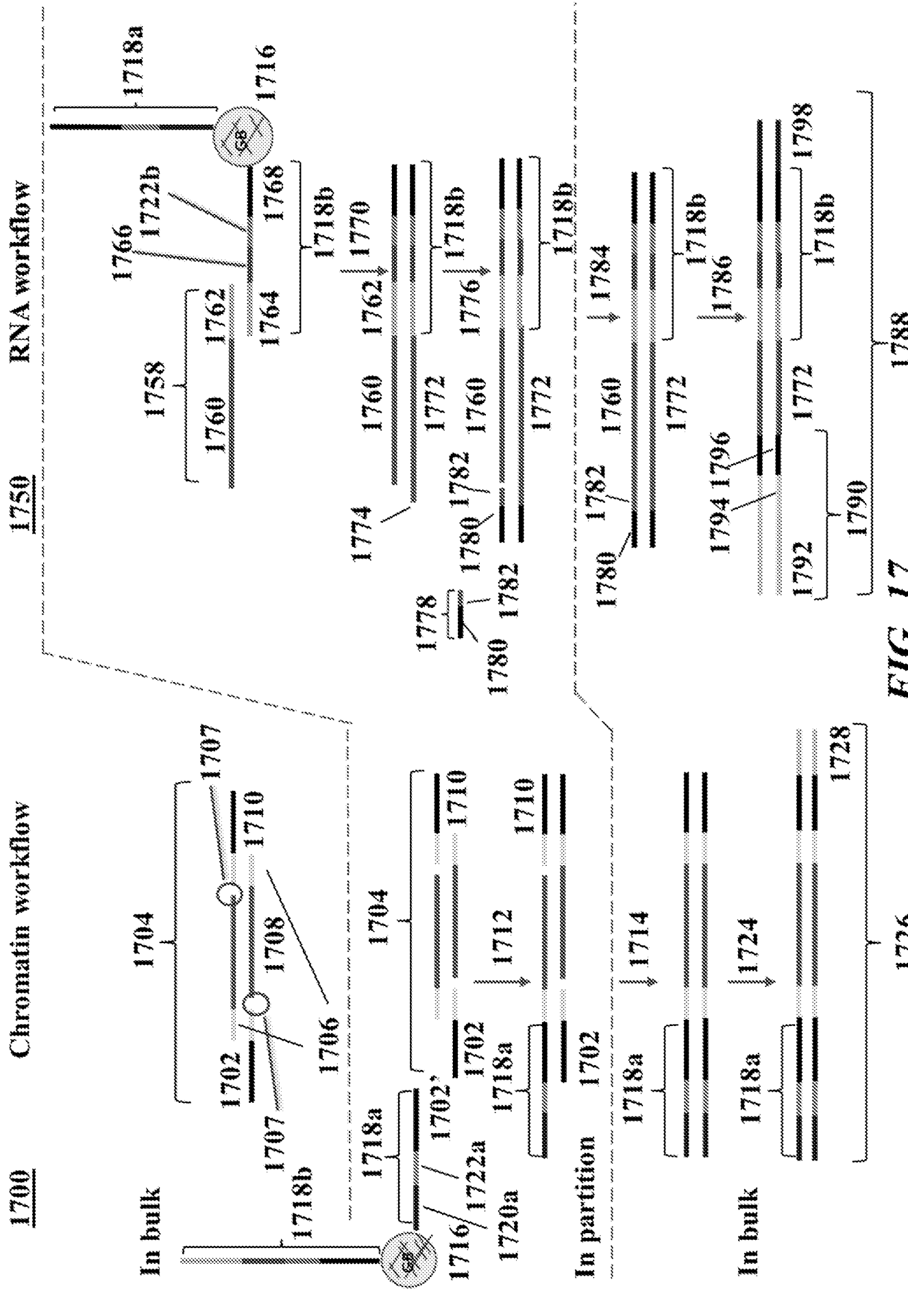

1700
Chromatin workflow
In bulk
In partition
In bulk
1750
RNA workflow
1704
1707
1710
1708
1706
1702
1707
1718b
1716
1722a
1720a
1718a
1702'
1704
1710
1712
1702
1718a
1702
1718a
1714
1718a
1724
1726
1728
1716
1718a
1722b
1766
1768
1764
1758
1760
1762
1718b
1770
1774
1760
1762
1772
1718b
1776
1778
1780
1782
1760
1772
1718b
1784
1780
1782
1760
1772
1718b
1786
1794
1796
1772
1798
1718b
1792
1790
1788

FIG. 17

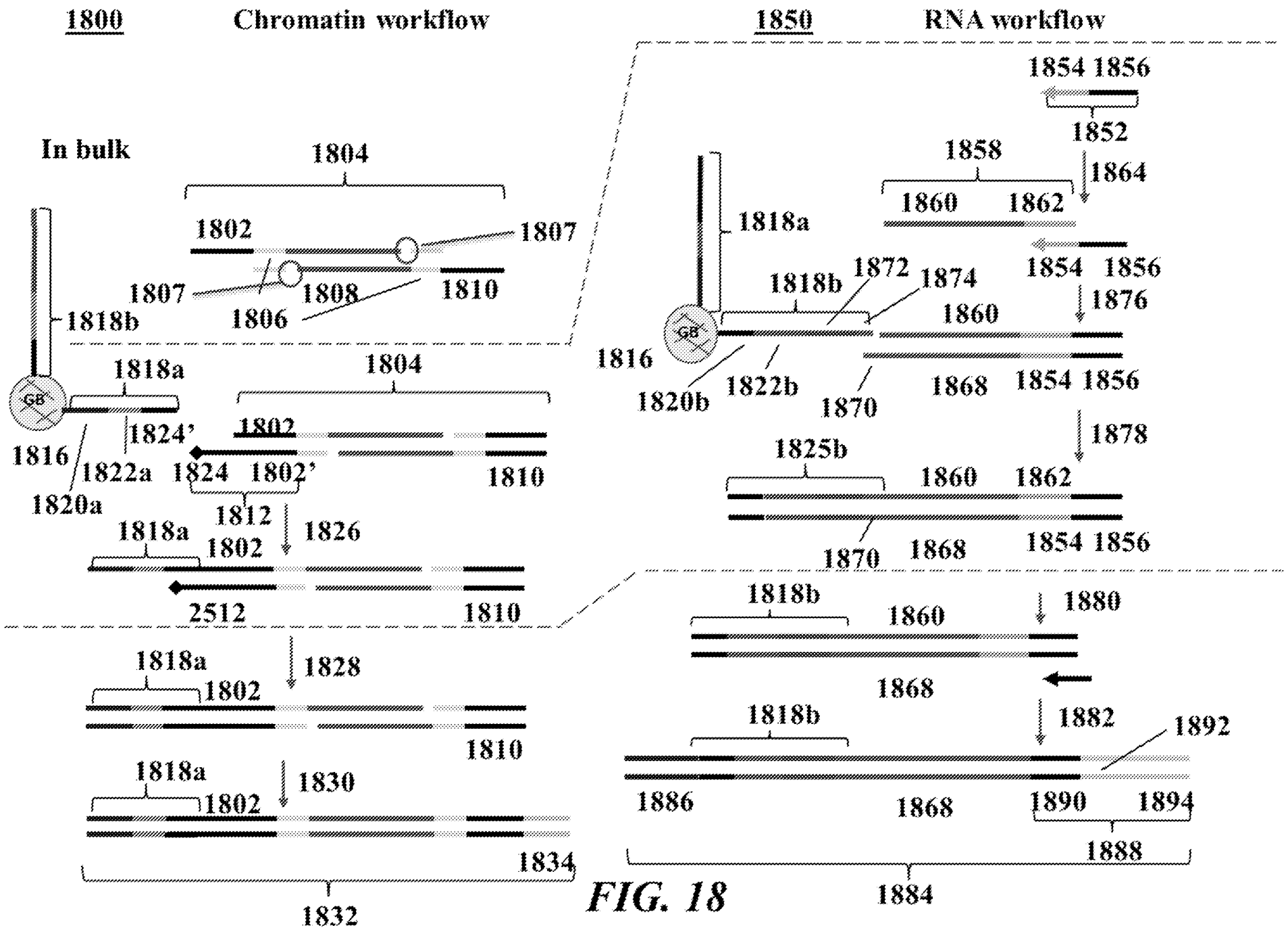

1800
Chromatin workflow
In bulk
1804
1802
1807
1807
1808
1810
1806
1818b
1818a
GB
1804
1816
1824'
1802
1822a
1824
1802'
1810
1820a
1812
1826
1818a
1802
2512
1810
1818a
1828
1802
1810
1818a
1830
1802
1834
1832
1850
RNA workflow
1854 1856
1852
1858
1864
1860
1862
1818a
1854
1856
1818b
1872
1874
1876
1860
GB
1816
1820b
1822b
1870
1868
1854 1856
1878
1825b
1860
1862
1870
1868
1854 1856
1818b
1880
1860
1868
1818b
1882
1892
1886
1868
1890
1894
1888
1884

FIG. 18

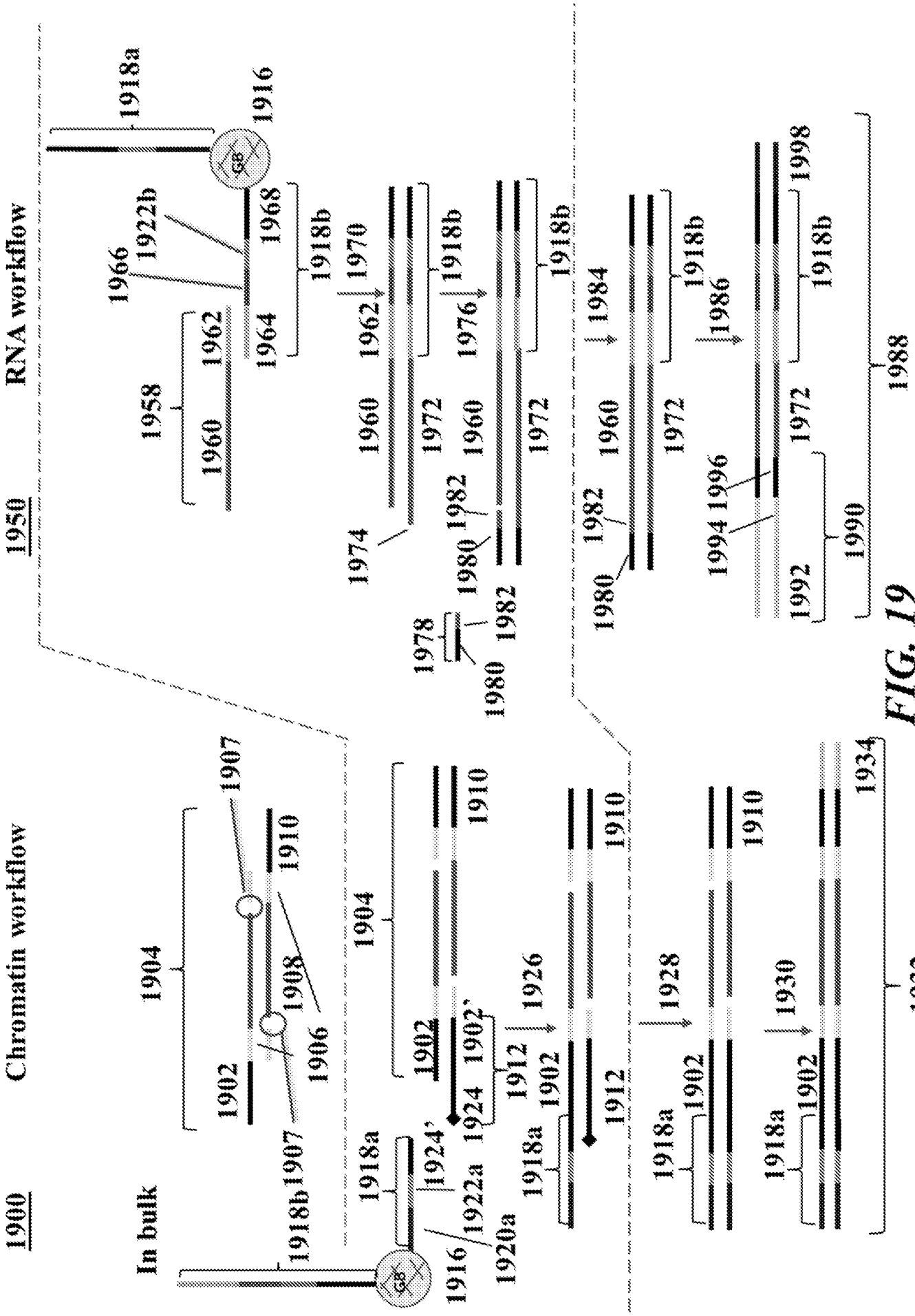

1900
In bulk
Chromatin workflow
1904
1907
1902
1910
1908
1906
1907
1918b
1916
1918a
1924'
1922a
1920a
1902
1904
1910
1924
1902'
1912
1926
1902
1918a
1912
1910
1928
1902
1918a
1910
1930
1902
1918a
1934
1932
1950
RNA workflow
1918a
1916
1966
1922b
1968
1964
1962
1958
1960
1918b
1970
1962
1960
1972
1974
1978
1980
1982
1976
1960
1972
1918b
1980
1982
1984
1960
1972
1986
1918b
1994
1996
1998
1972
1992
1990
1918b
1988

FIG. 19

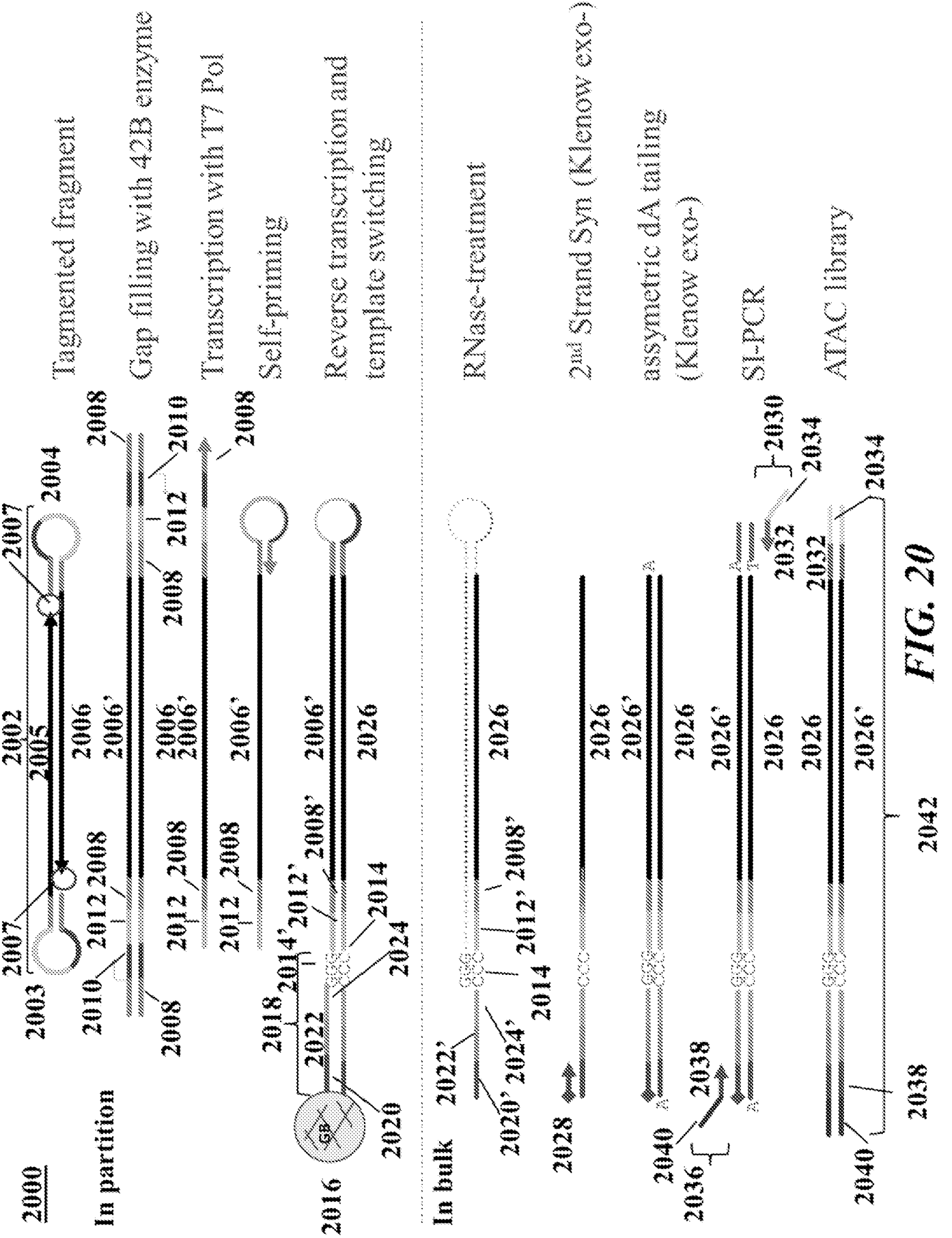

2000
In partition
In bulk
Tagmented fragment
Gap filling with 42B enzyme
Transcription with T7 Pol
Self-priming
Reverse transcription and template switching
RNase-treatment
2nd Strand Syn (Klenow exo-)
assymetric dA tailing (Klenow exo-)
SI-PCR
ATAC library
2002
2003
2004
2005
2006
2006'
2007
2008
2008'
2010
2012
2012'
2014
2014'
2016
2018
2020
2020'
2022
2022'
2024
2024'
2026
2026'
2028
2030
2032
2034
2036
2038
2040
2042
GB

FIG. 20

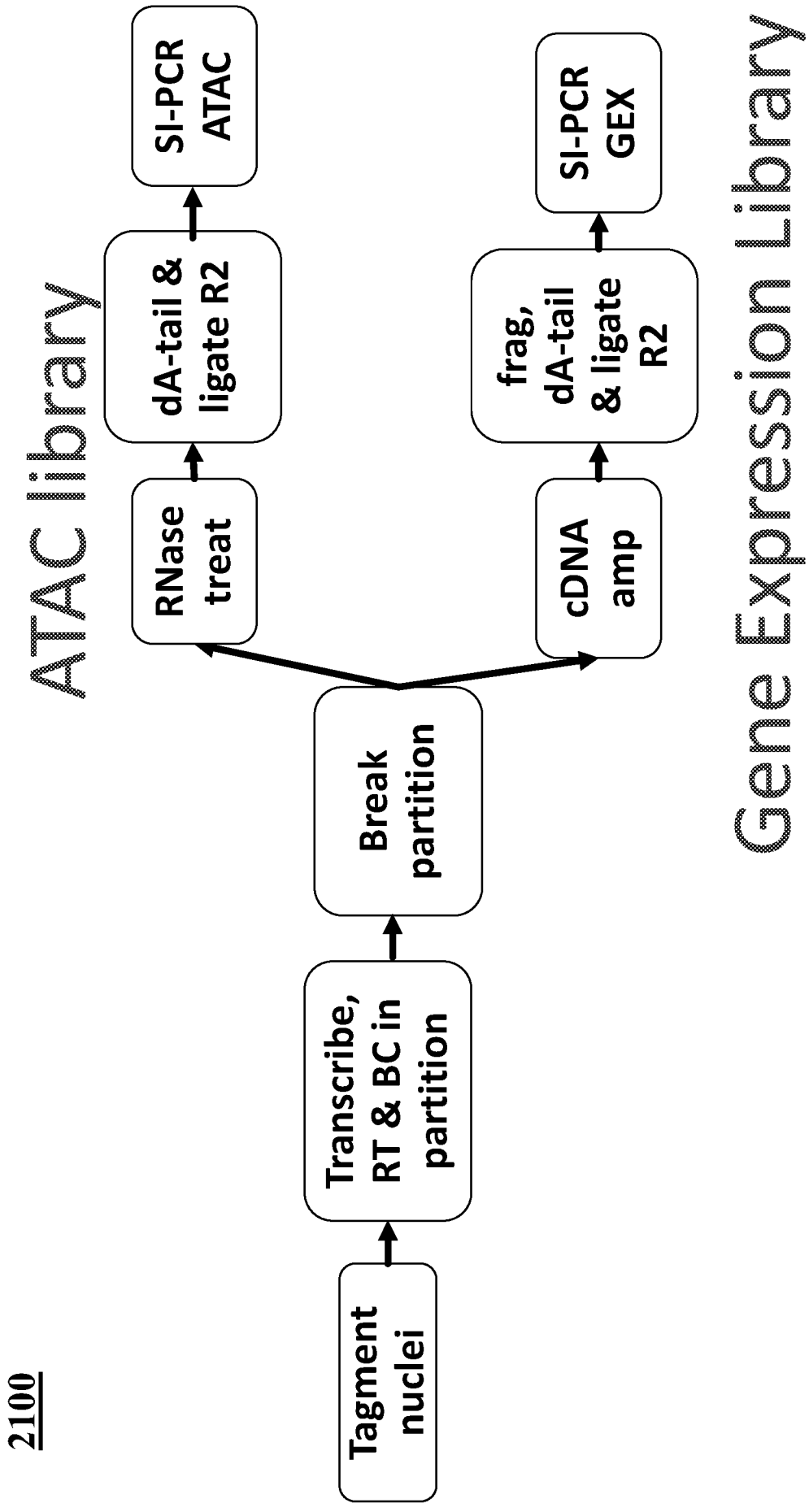

2100
Tagment nuclei
Transcribe, RT & BC in partition
Break partition
RNase treat
dA-tail & ligate R2
SI-PCR ATAC
ATAC library
cDNA amp
frag, dA-tail & ligate R2
SI-PCR GEX
Gene Expression Library

FIG. 21

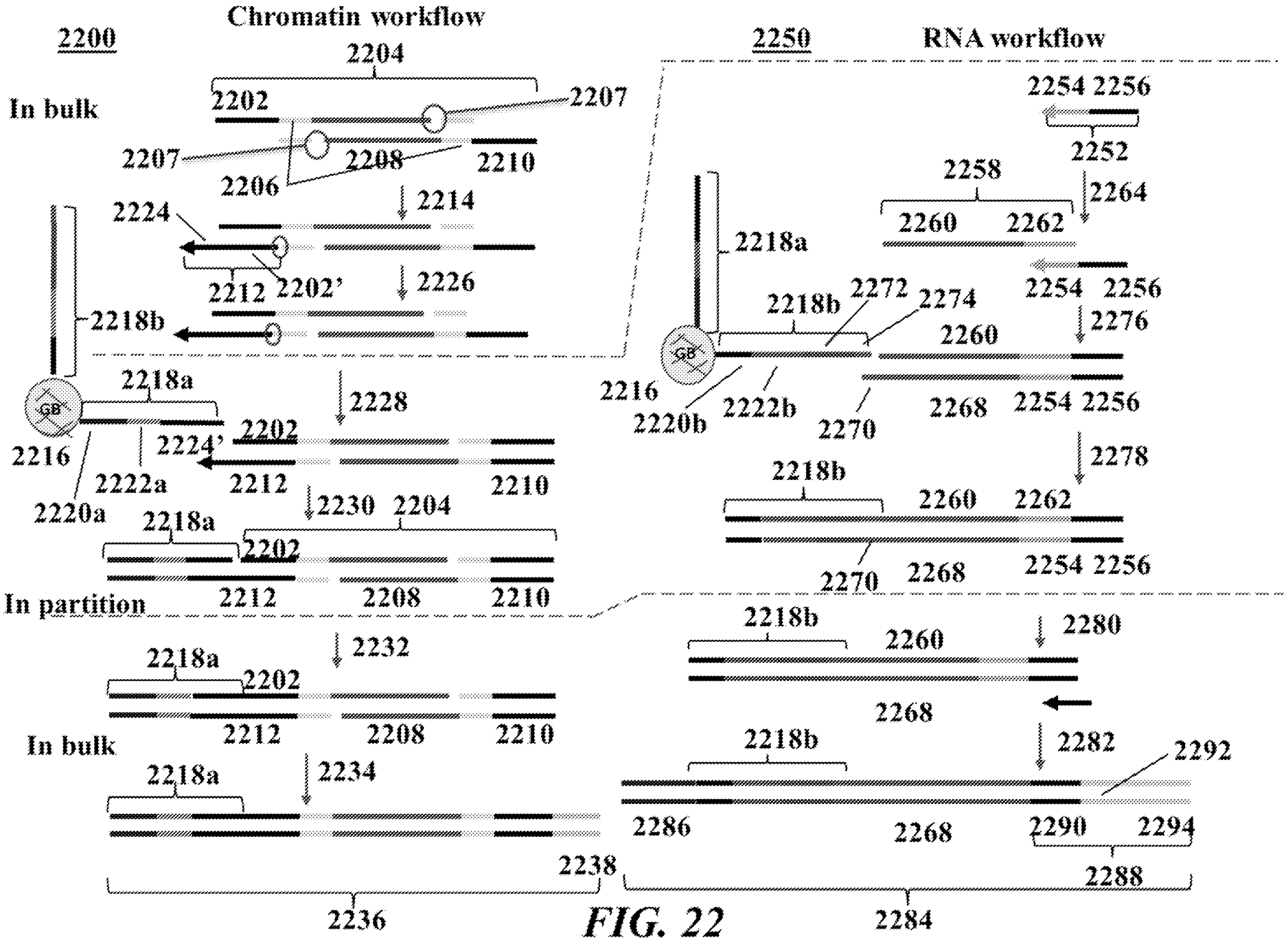

2200
Chromatin workflow
2250
RNA workflow
In bulk
In partition
In bulk
2204
2202
2207
2206
2208
2210
2214
2224
2212
2202'
2226
2218b
2218a
2228
2224'
2216
2222a
2220a
2230
2232
2234
2236
2238
2254 2256
2252
2258
2264
2260
2262
2272
2274
2276
2222b
2220b
2270
2268
2278
2280
2282
2292
2286
2290
2294
2288
2284
GB

FIG. 22

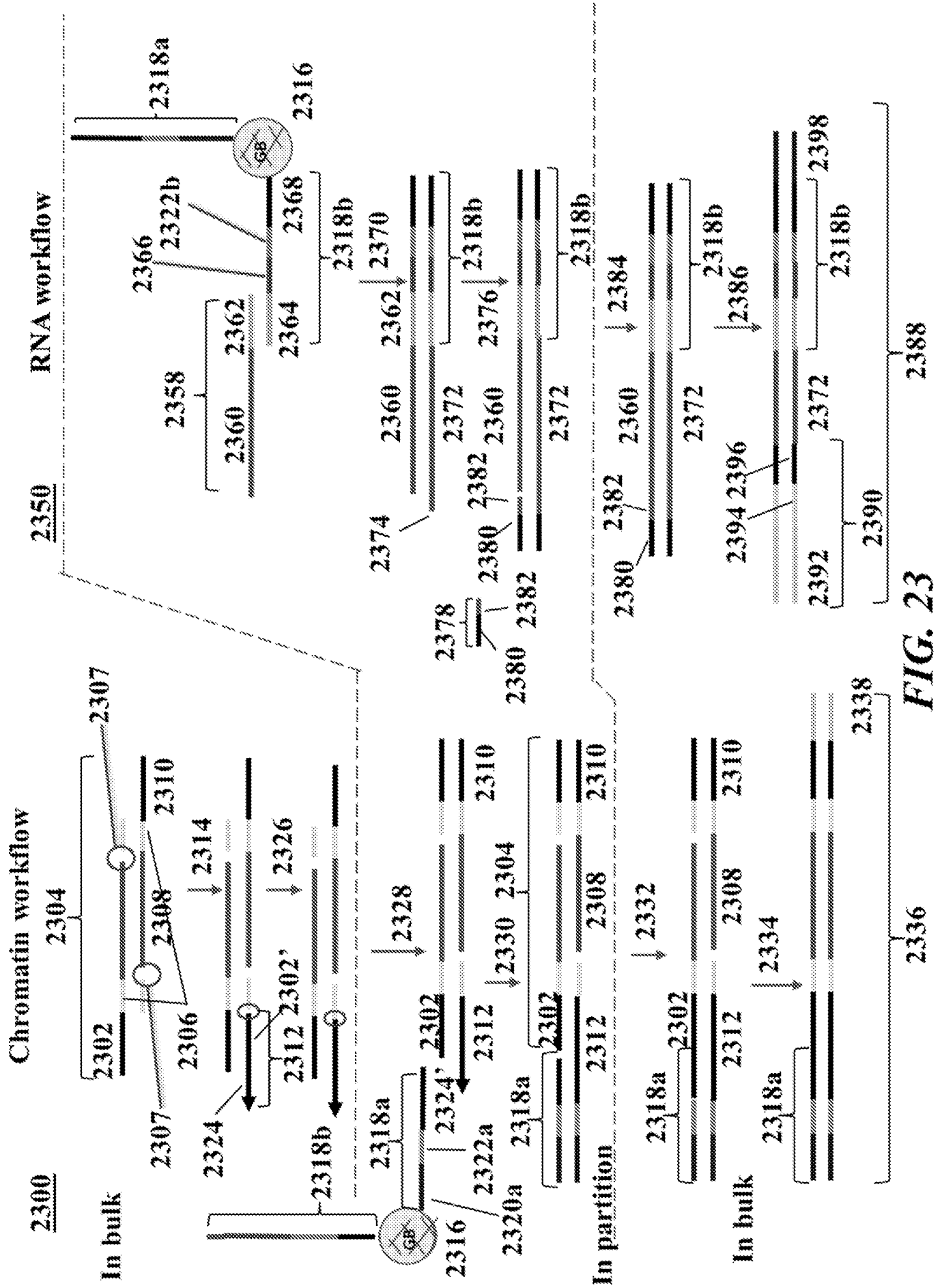

2300
Chromatin workflow
In bulk
2304
2307
2310
2308
2306
2314
2326
2302'
2312
2324
2318b
2318a
2316
2322a
2320a
2328
2302
2324'
2330
2304
In partition
2332
2334
2336
2338
2350
RNA workflow
2318a
2316
2366
2322b
2368
2318b
2370
2358
2362
2360
2364
2372
2374
2376
2378
2380
2382
2384
2386
2388
2390
2392
2394
2396
2398

FIG. 23

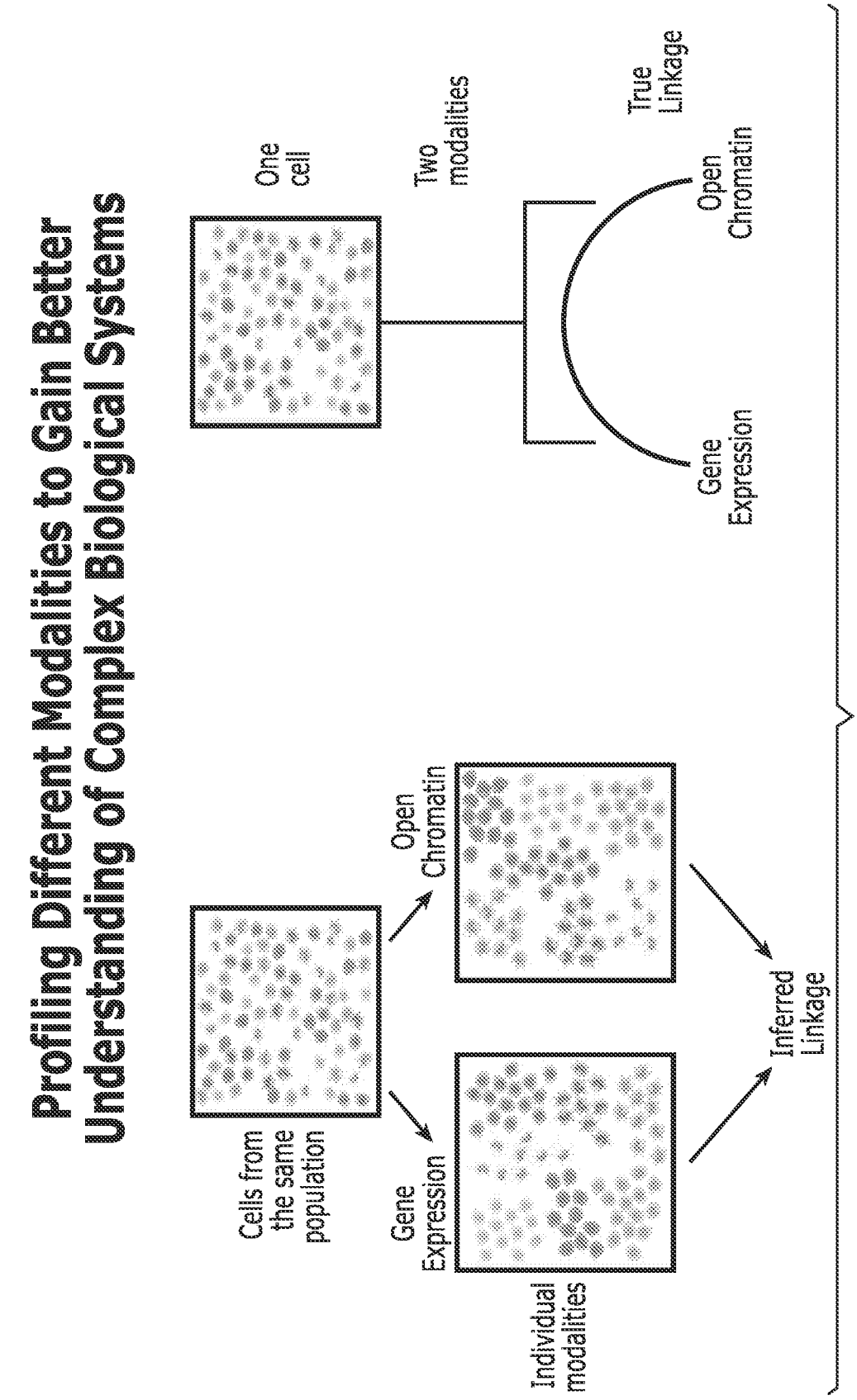

Profiling Different Modalities to Gain Better Understanding of Complex Biological Systems
Cells from the same population
Gene Expression
Open Chromatin
Individual modalities
Inferred Linkage
One cell
Two modalities
Gene Expression
Open Chromatin
True Linkage

FIG. 24

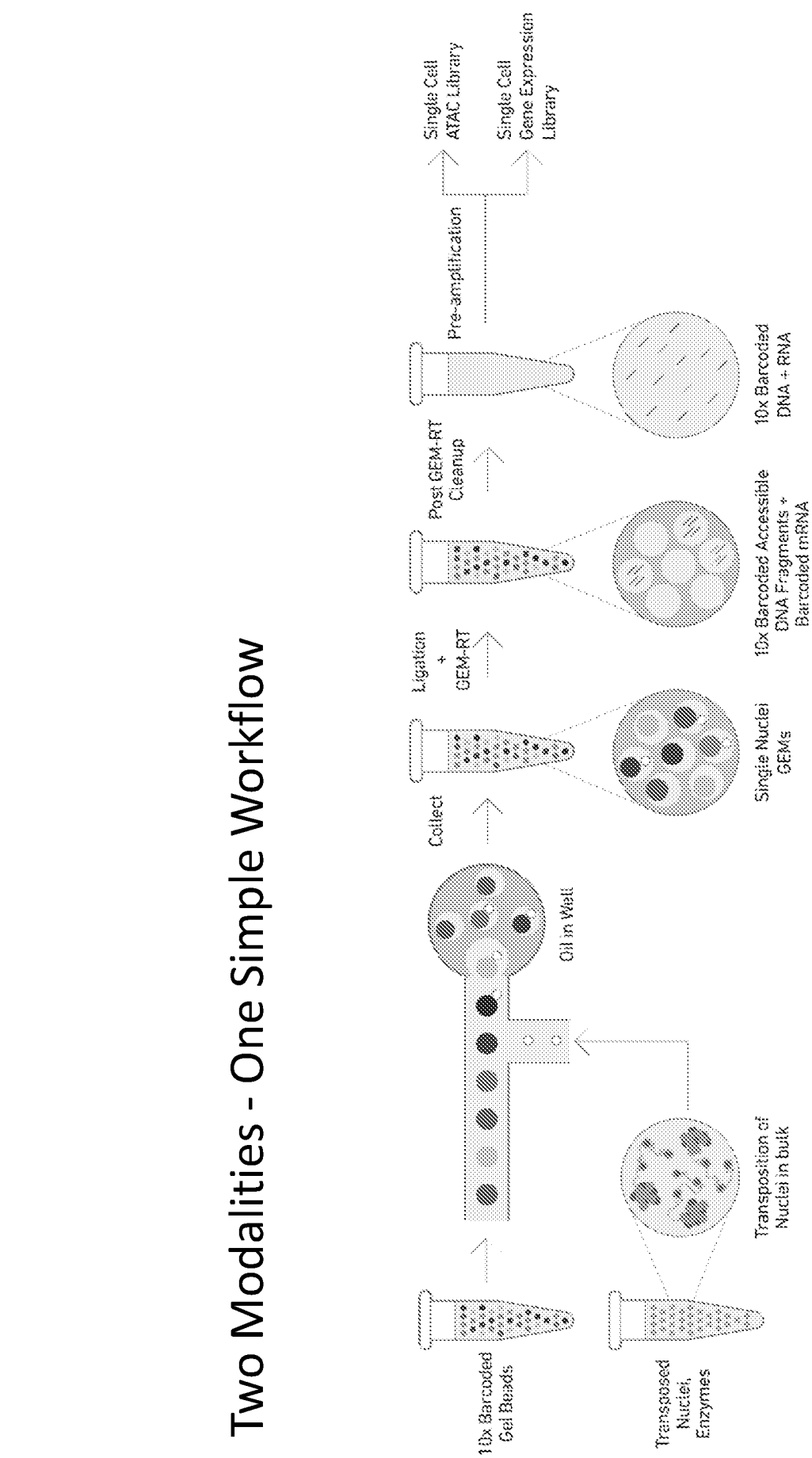

Two Modalities - One Simple Workflow
10x Barcoded Gel Beads
Transposed Nuclei, Enzymes
Transposition of Nuclei in bulk
Oil in Well
Collect
Singie Nuclei GEMs
Ligation + GEM-RT
10x Barcoded Accessible DNA Fragments + Barcoded mRNA
Post GEM-RT Cleanup
10x Barcoded DNA + RNA
Pre-amplification
Single Cell ATAC Library
Single Cell Gene Expression Library

FIG. 25

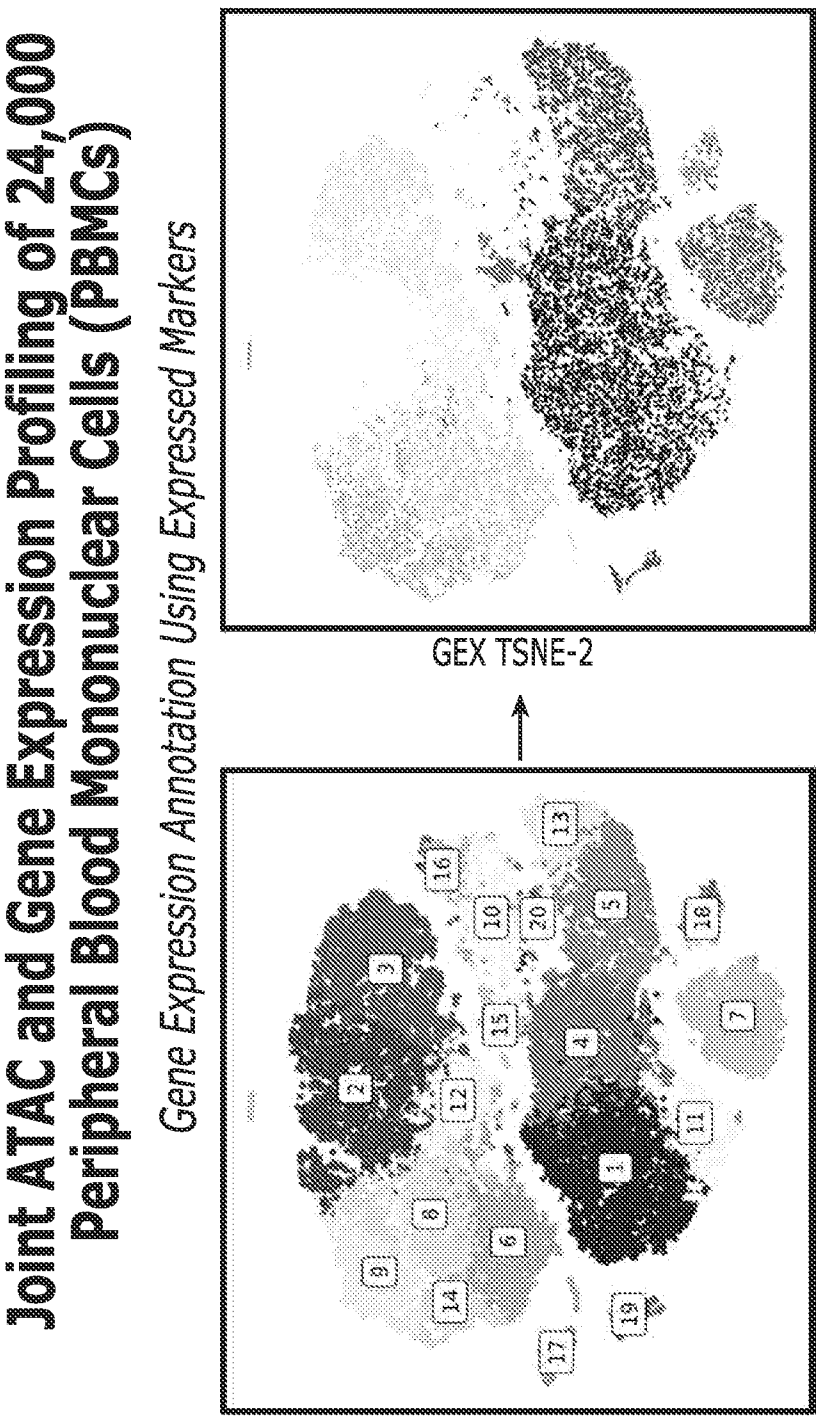

Joint ATAC and Gene Expression Profiling of 24,000 Peripheral Blood Mononuclear Cells (PBMCs)
Gene Expression Annotation Using Expressed Markers
CD14 monocytes
CD16 monocytes
myeloid DC
plasmacytoid DC
granulocytes
NK cells
CD8 effector T cells
naive T cells
naive/memory B cells
IgM+IgD+memory B cells
plasma B cells
NA
GEX TSNE-2
GEX TSNE-1

FIG. 26

Joint ATAC and Gene Expression Profiling of PBMCs
Open Chromatin Annotation Using TF Accessibility
monocytes
NK/CD8 + T cells
naive/memory T cells
B cells
NA

FIG. 27

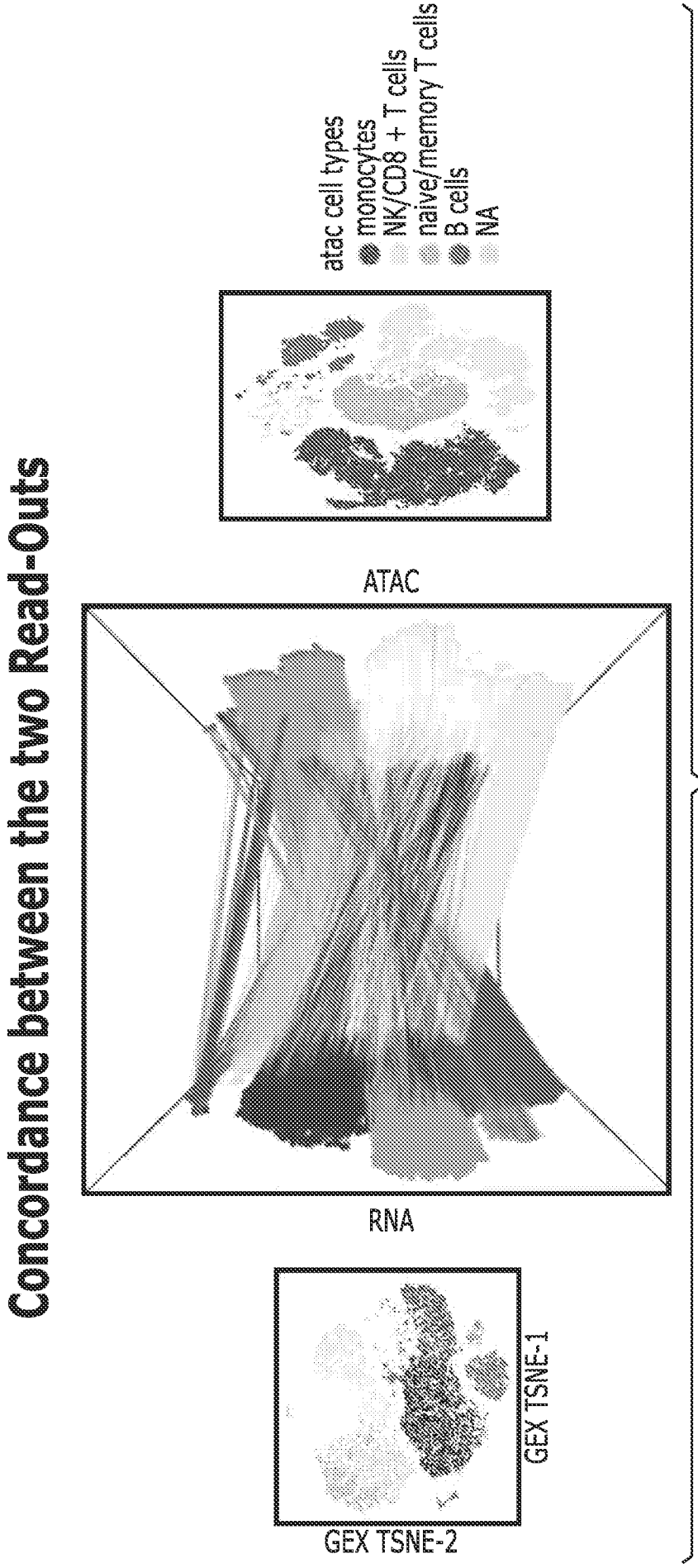

Concordance between the two Read-Outs
atac cell types
monocytes
NK/CD8 + T cells
naive/memory T cells
B cells
NA
ATAC
RNA
GEX TSNE-1
GEX TSNE-2

FIG. 28

Concordance between the two Read-Outs

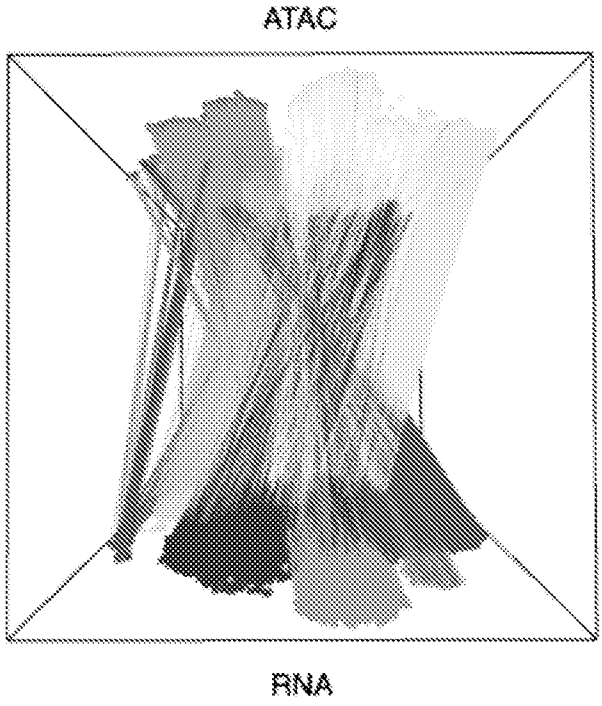


FIG. 29A

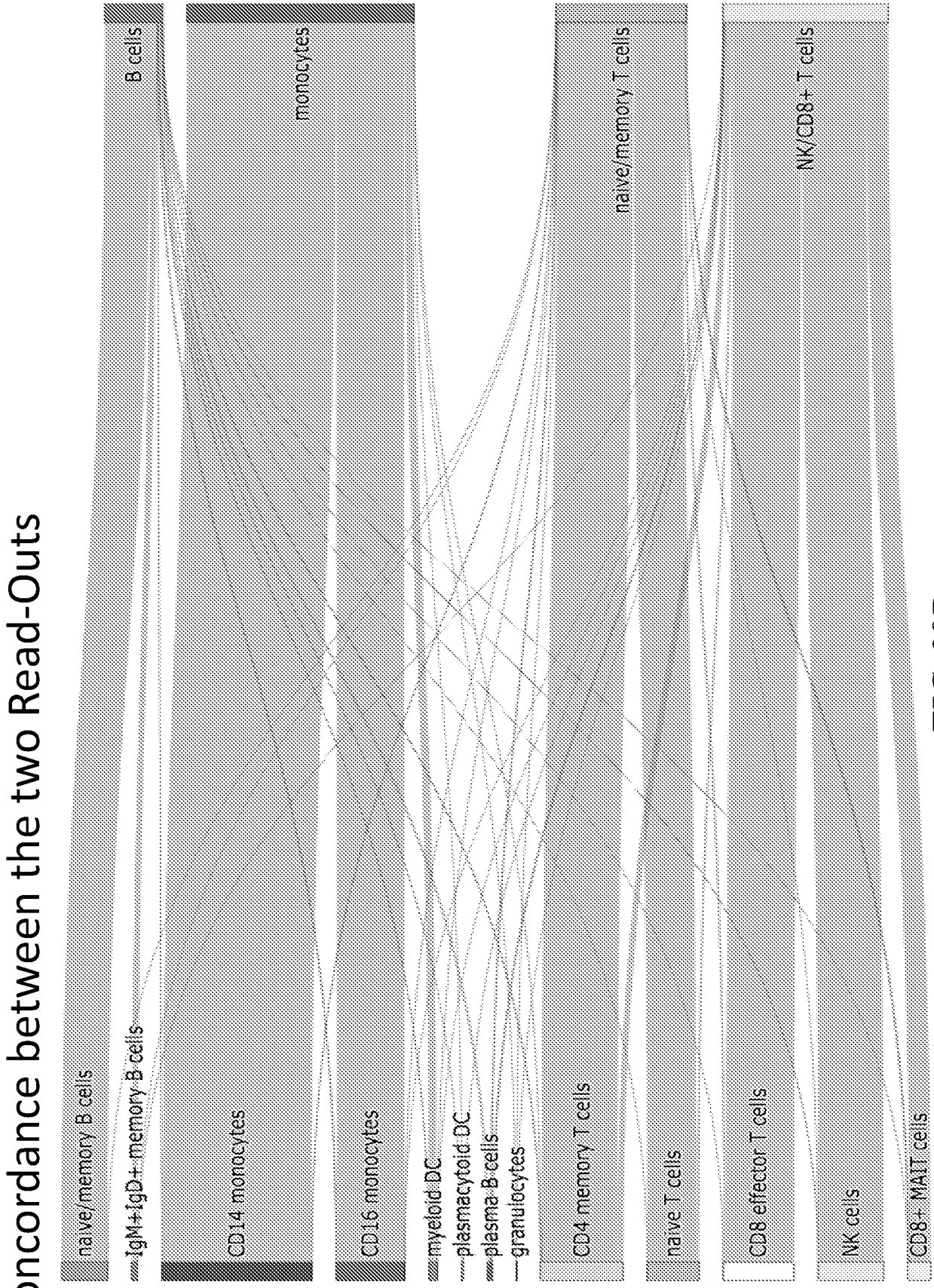

Concordance between the two Read-Outs
naive/memory B cells
IgM+IgD+ memory B cells
CD14 monocytes
CD16 monocytes
myeloid DC
plasmacytoid DC
plasma B cells
granulocytes
CD4 memory T cells
naive T cells
CD8 effector T cells
NK cells
CD8+ MAIT cells
B cells
monocytes
naive/memory T cells
NK/CD8+ T cells

FIG. 29B

Further Annotate your ATAC Populations
Transfer Gene Expression Marker-derived Annotation into ATAC Populations
ATAC
RNA
Open chromatin annotation using Transcription Factor accessibility
monocytes
NK/CD8 + T cells
naive/memory T cells
B cells
NA
Open chromatin annotation using Gene Expression markers
CD14 monocytes
CD16 monocytes
myeloid DC
plasmacytoid DC
granulocytes
NK cells
CD8 effector T cells
naive T cells
naive/memory B cells
IgM+IgD+memory B cells
plasma B cells
NA
ATAC TSNE-2
ATAC TSNE-1

FIG. 30

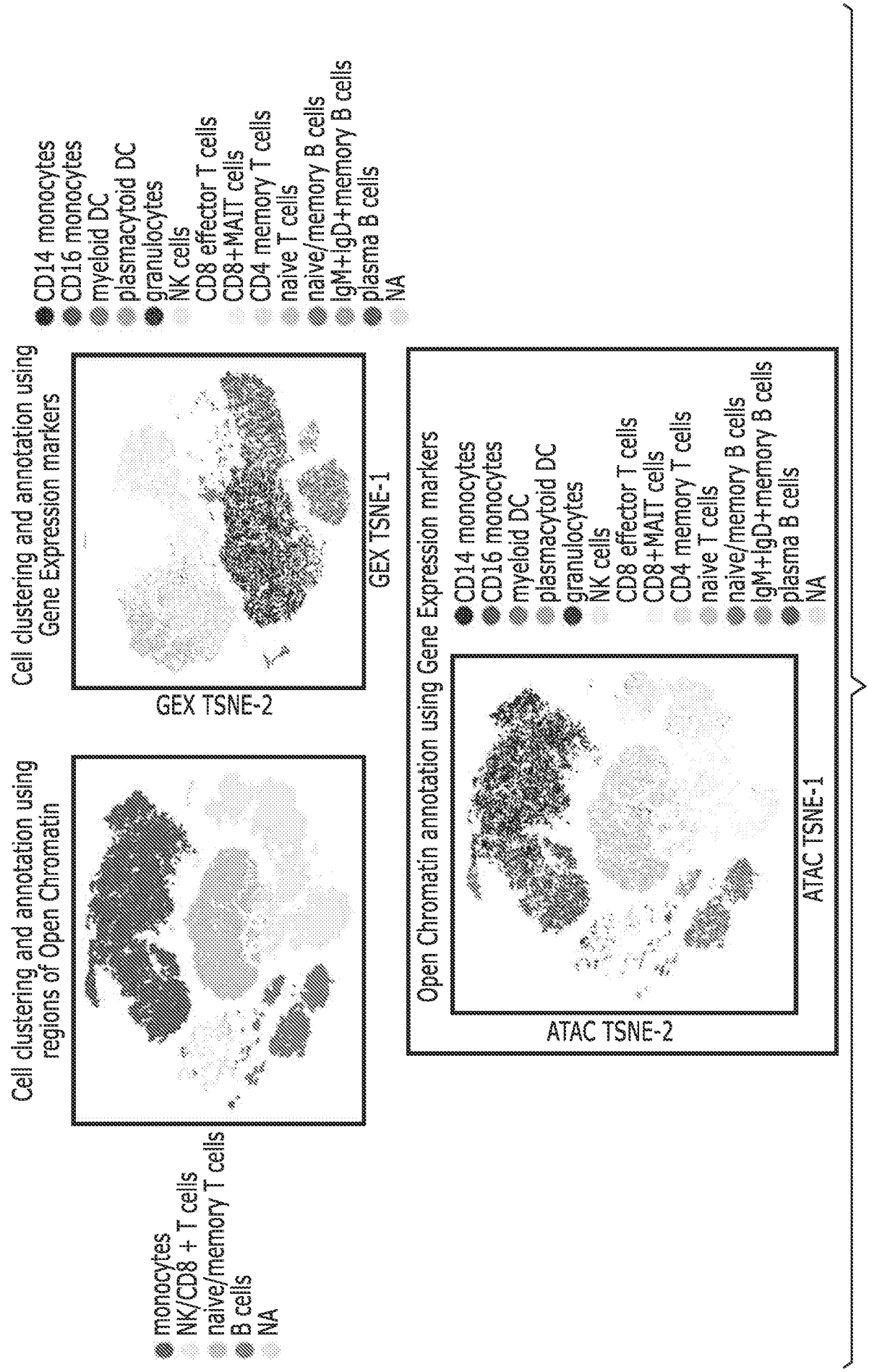

Cell clustering and annotation using Gene Expression markers
GEX TSNE-2
GEX TSNE-1
CD14 monocytes
CD16 monocytes
myeloid DC
plasmacytoid DC
granulocytes
NK cells
CD8 effector T cells
CD8+MAIT cells
CD4 memory T cells
naive T cells
naive/memory B cells
IgM+IgD+memory B cells
plasma B cells
NA
Cell clustering and annotation using regions of Open Chromatin
monocytes
NK/CD8 + T cells
naive/memory T cells
B cells
NA
Open Chromatin annotation using Gene Expression markers
ATAC TSNE-2
ATAC TSNE-1
CD14 monocytes
CD16 monocytes
myeloid DC
plasmacytoid DC
granulocytes
NK cells
CD8 effector T cells
CD8+MAIT cells
CD4 memory T cells
naive T cells
naive/memory B cells
IgM+IgD+memory B cells
plasma B cells
NA

FIG. 31

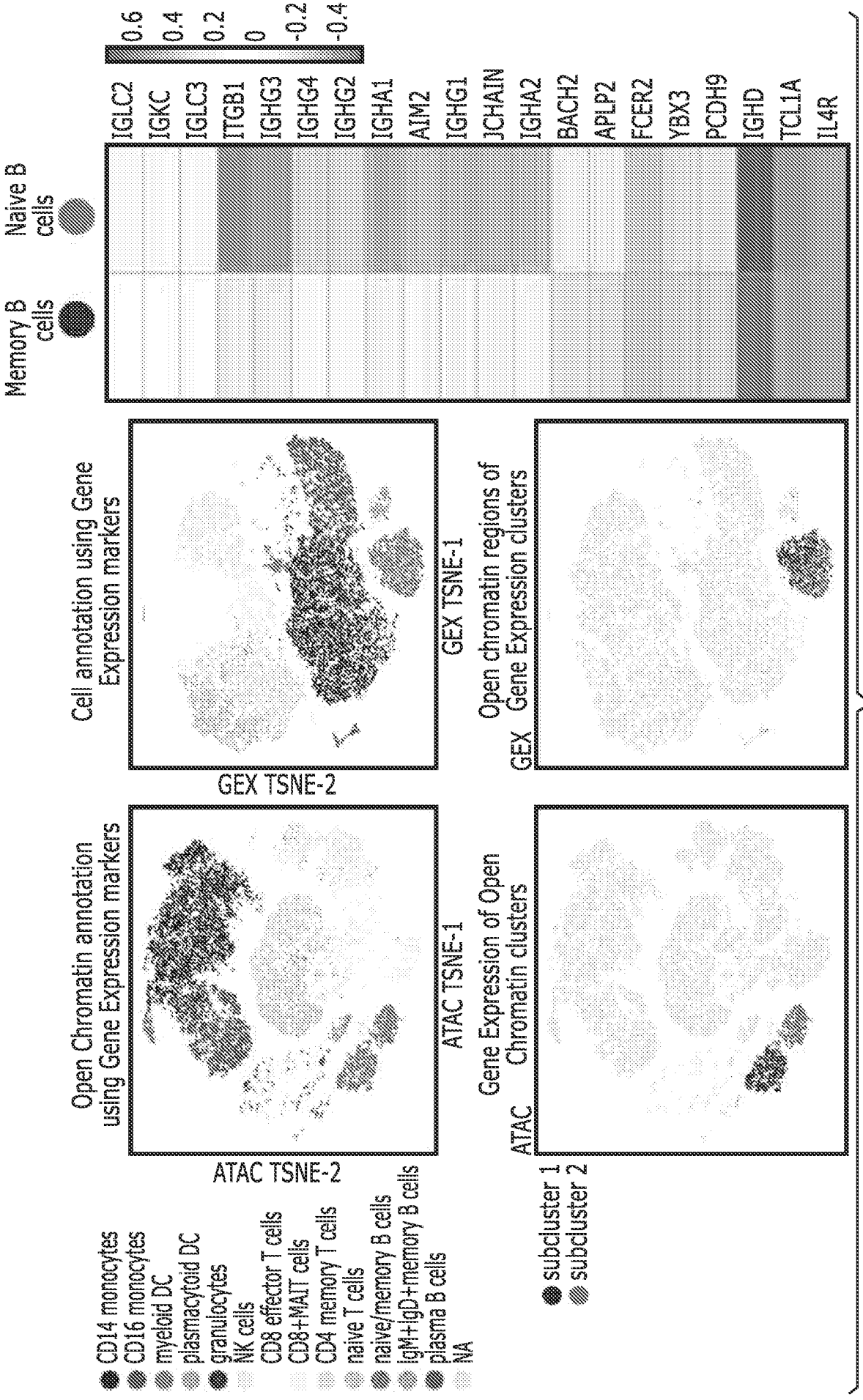

Memory B cells
Naive B cells
0.6
0.4
0.2
0
-0.2
-0.4
IGLC2
IGKC
IGLC3
ITGB1
IGHG3
IGHG4
IGHG2
IGHA1
AIM2
IGHG1
JCHAIN
IGHA2
BACH2
APLP2
FCER2
YBX3
PCDH9
IGHD
TCL1A
IL4R
Cell annotation using Gene Expression markers
GEX TSNE-2
GEX TSNE-1
Open chromatin regions of Gene Expression clusters
GEX
Open Chromatin annotation using Gene Expression markers
ATAC TSNE-2
ATAC TSNE-1
Gene Expression of Open Chromatin clusters
ATAC
CD14 monocytes
CD16 monocytes
myeloid DC
plasmacytoid DC
granulocytes
NK cells
CD8 effector T cells
CD8+MAIT cells
CD4 memory T cells
naive T cells
naive/memory B cells
IgM+IgD+memory B cells
plasma B cells
NA
subcluster 1
subcluster 2

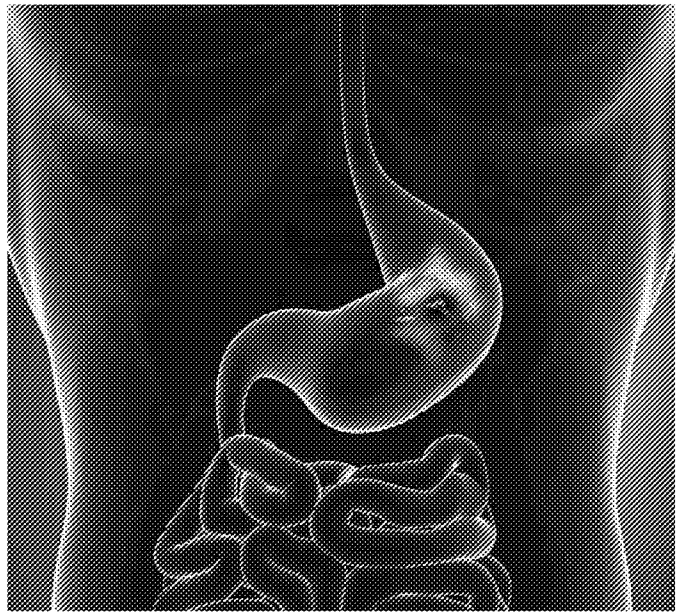
FIG. 32

Case Study - B Cell Lymphoma

Functional characterization of a small B cell lymphoma and its signaling pathways

Pathology report

Lymph node tumor
Diagnosis: Diffuse small lymphocytic lymphoma of the lymph node
Tumor location: Lymph node, Intra-abdominal
Diagnosis details: IHC: CD20(+), CD3(-)
Comments: Malignant lymphoma, small B cell, diffuse type



Profile ~9,000 nuclei through Single Cell ATAC + Gene Expression analysis

FIG. 33

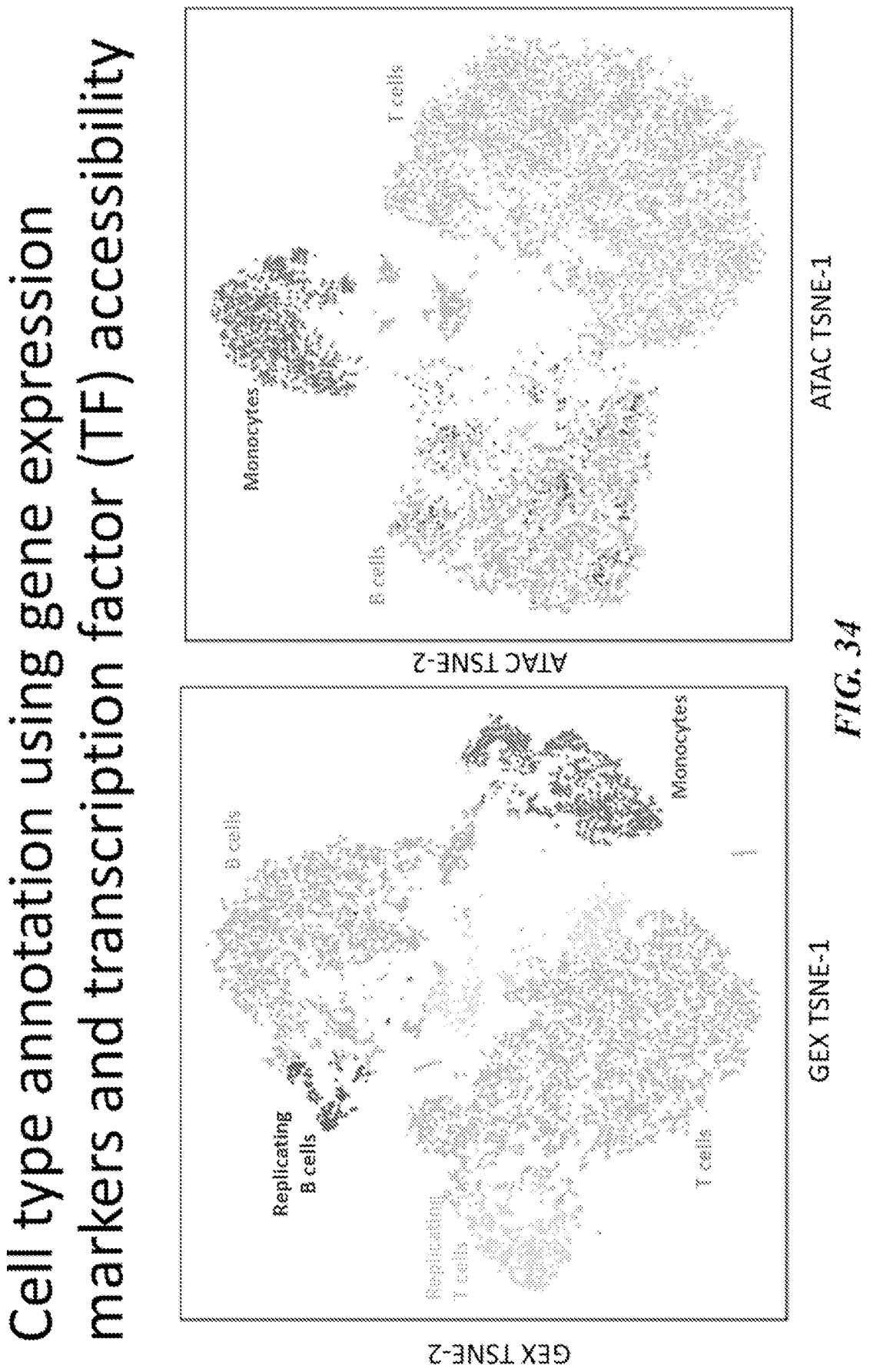

Cell type annotation using gene expression markers and transcription factor (TF) accessibility
Monocytes
T cells
B cells
ATAC TSNE-2
ATAC TSNE-1
B cells
Replicating B cells
Monocytes
T cells
GEX TSNE-2
GEX TSNE-1

FIG. 34

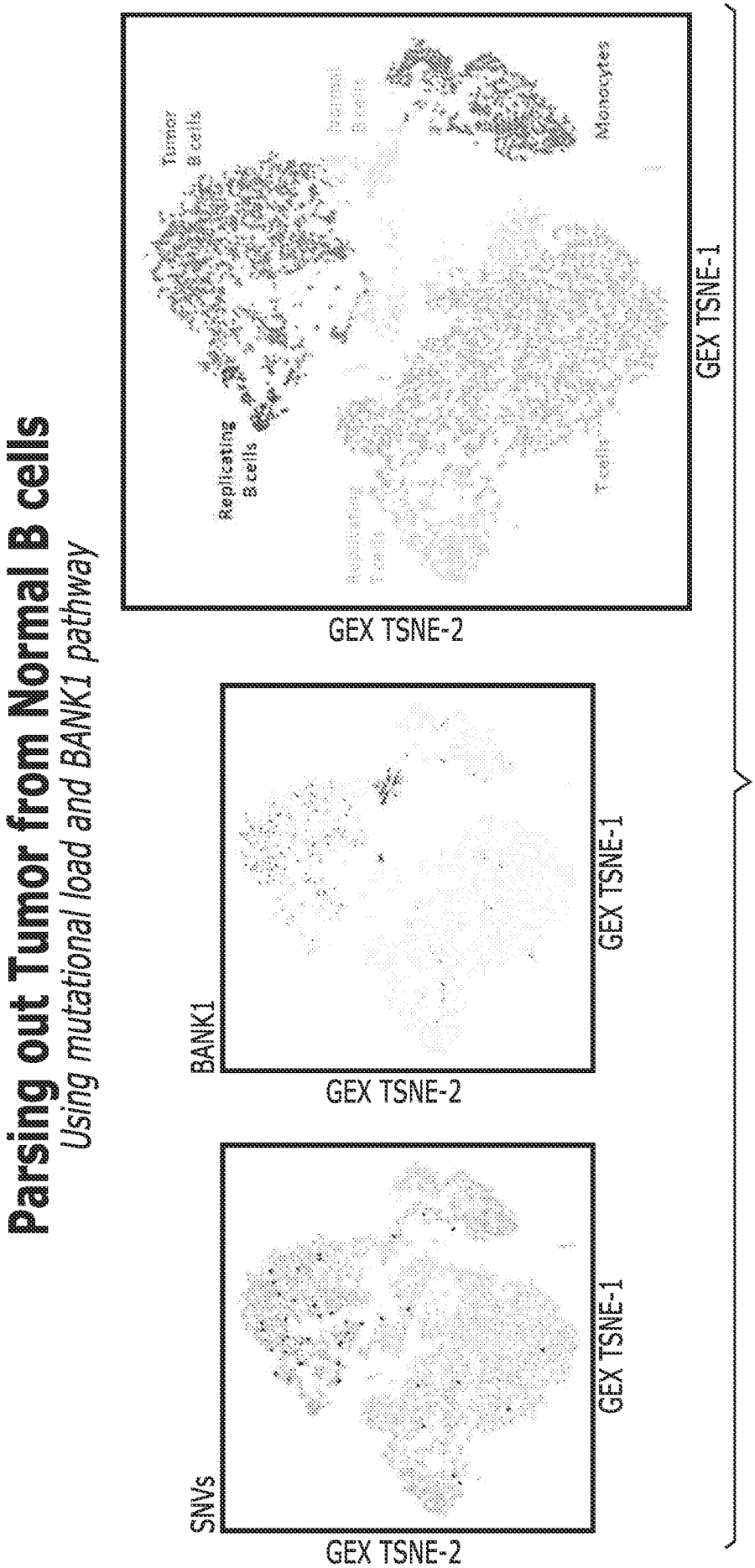

Parsing out Tumor from Normal B cells
Using mutational load and BANK1 pathway
Tumor B cells
Normal B cells
Monocytes
Replicating B cells
T cells
GEX TSNE-2
GEX TSNE-1
BANK1
GEX TSNE-2
GEX TSNE-1
SNVs
GEX TSNE-2
GEX TSNE-1

FIG. 35

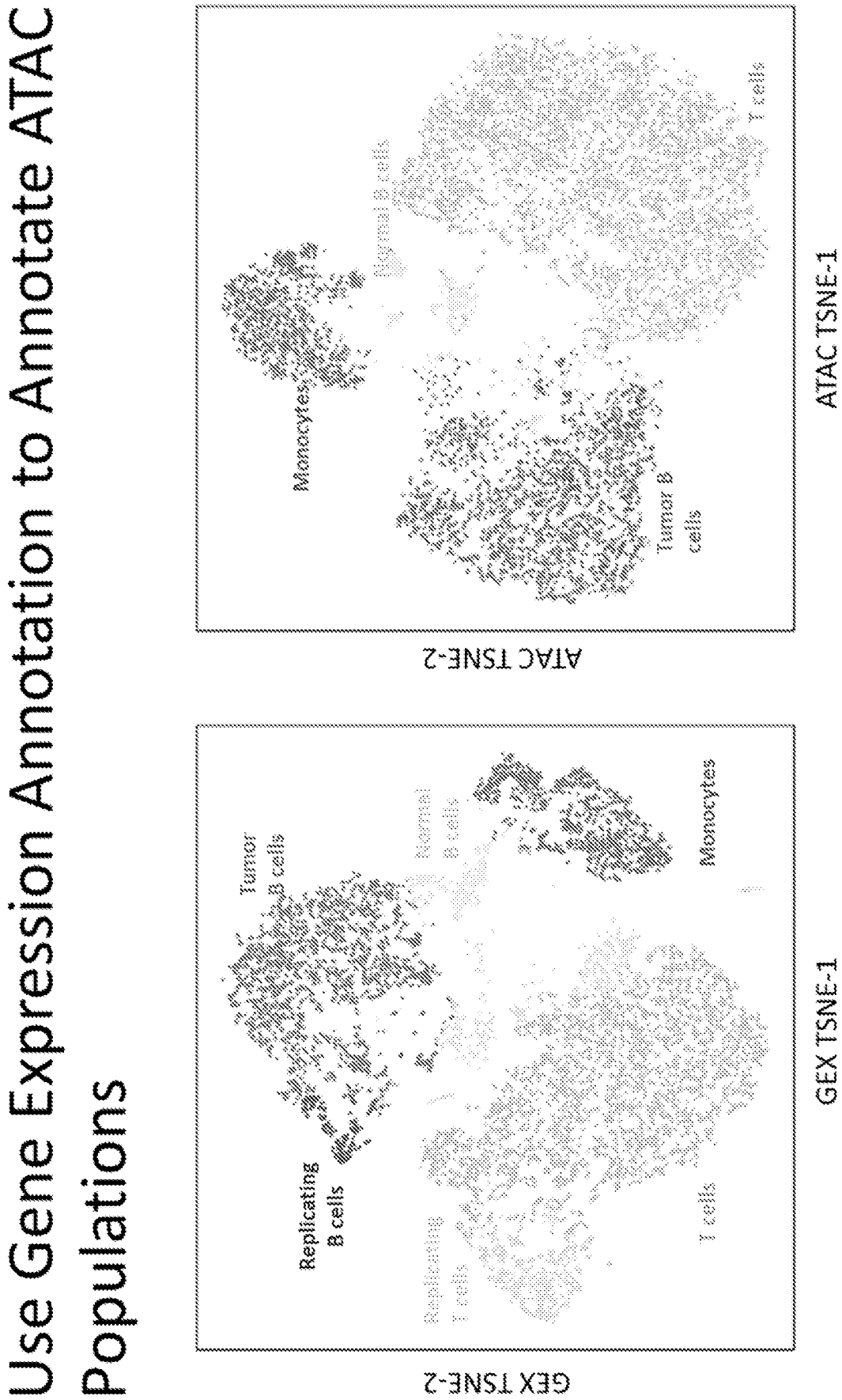

Use Gene Expression Annotation to Annotate ATAC Populations
Tumor B cells
Normal B cells
Monocytes
Replicating B cells
T cells
GEX TSNE-1
GEX TSNE-2
Monocytes
Normal B cells
T cells
Tumor B cells
ATAC TSNE-1
ATAC TSNE-2

FIG. 36

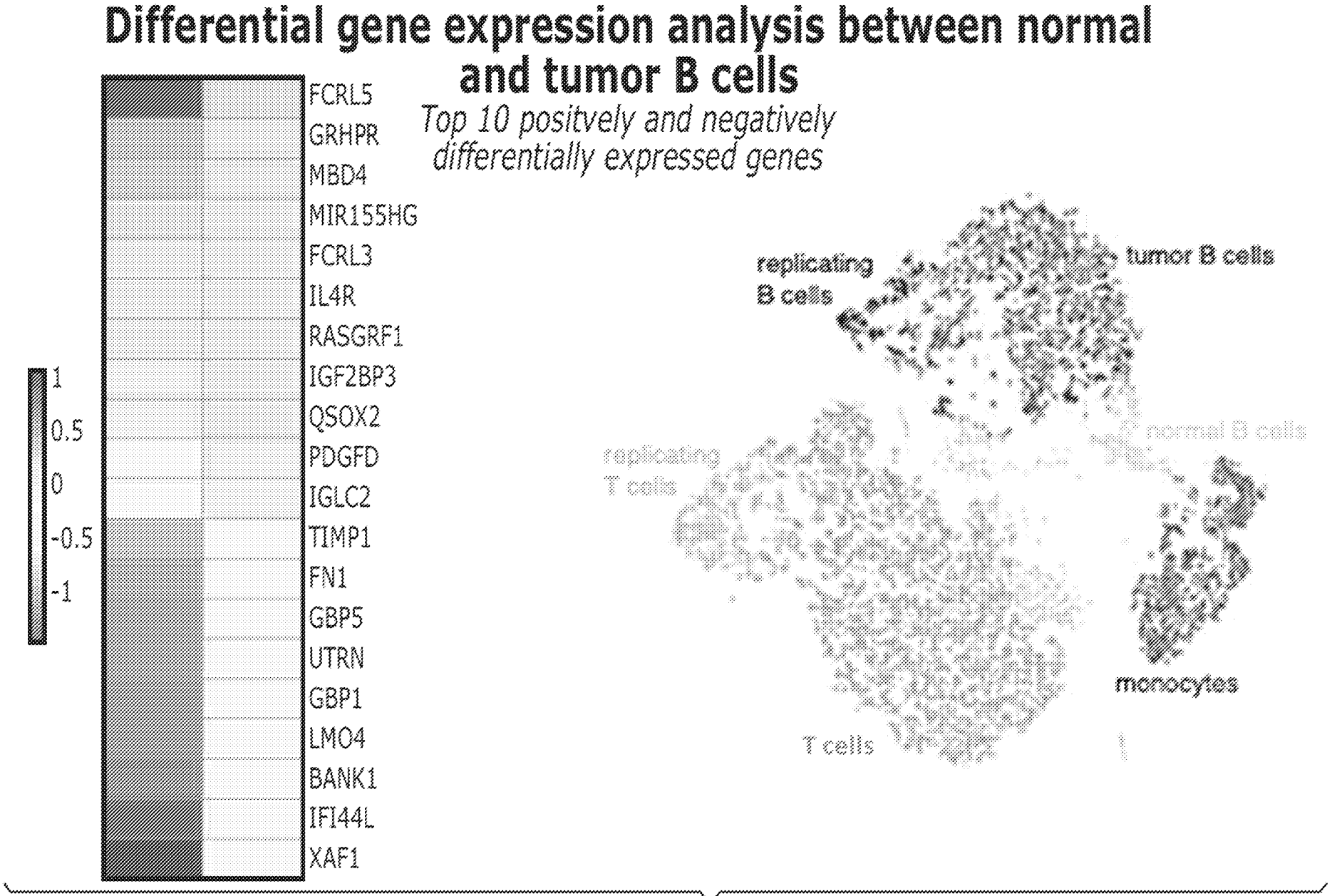

Differential gene expression analysis between normal and tumor B cells
Top 10 positvely and negatively differentially expressed genes
FCRL5
GRHPR
MBD4
MIR155HG
FCRL3
IL4R
RASGRF1
IGF2BP3
QSOX2
PDGFD
IGLC2
TIMP1
FN1
GBP5
UTRN
GBP1
LMO4
BANK1
IFI44L
XAF1
1
0.5
0
-0.5
-1
replicating B cells
tumor B cells
normal B cells
replicating T cells
monocytes
T cells

FIG. 37

When Gene Expression doesn't tell the whole story
STAT proteins accumulate in cytoplasm but are only active upon nuclear translocations
Expression
Accessibility
Stat3
Stat6
Stat3 MOTIF
Stat6 MOTIF
replicating tumor B cells
macrophage

FIG. 38A

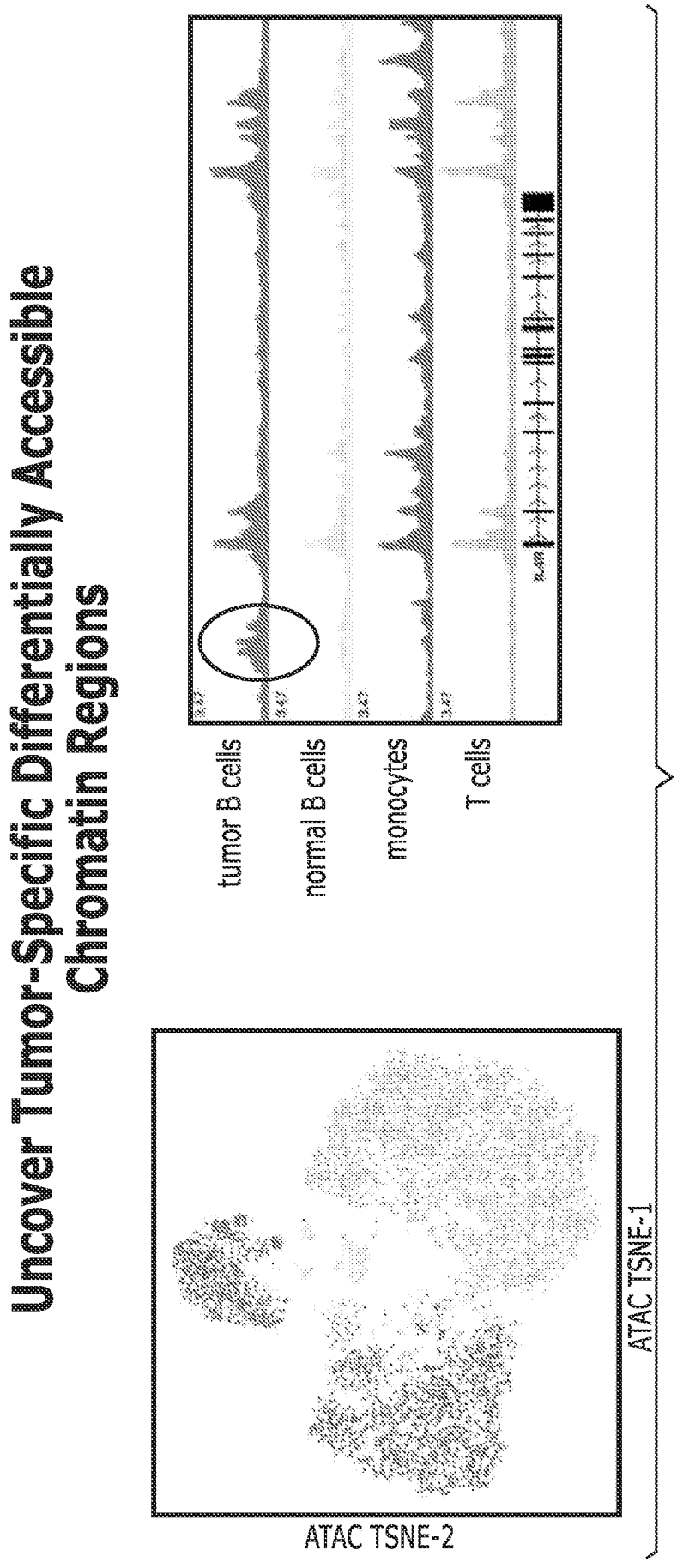

Uncover Tumor-Specific Differentially Accessible Chromatin Regions
tumor B cells
normal B cells
monocytes
T cells
ATAC TSNE-1
ATAC TSNE-2

FIG. 38B

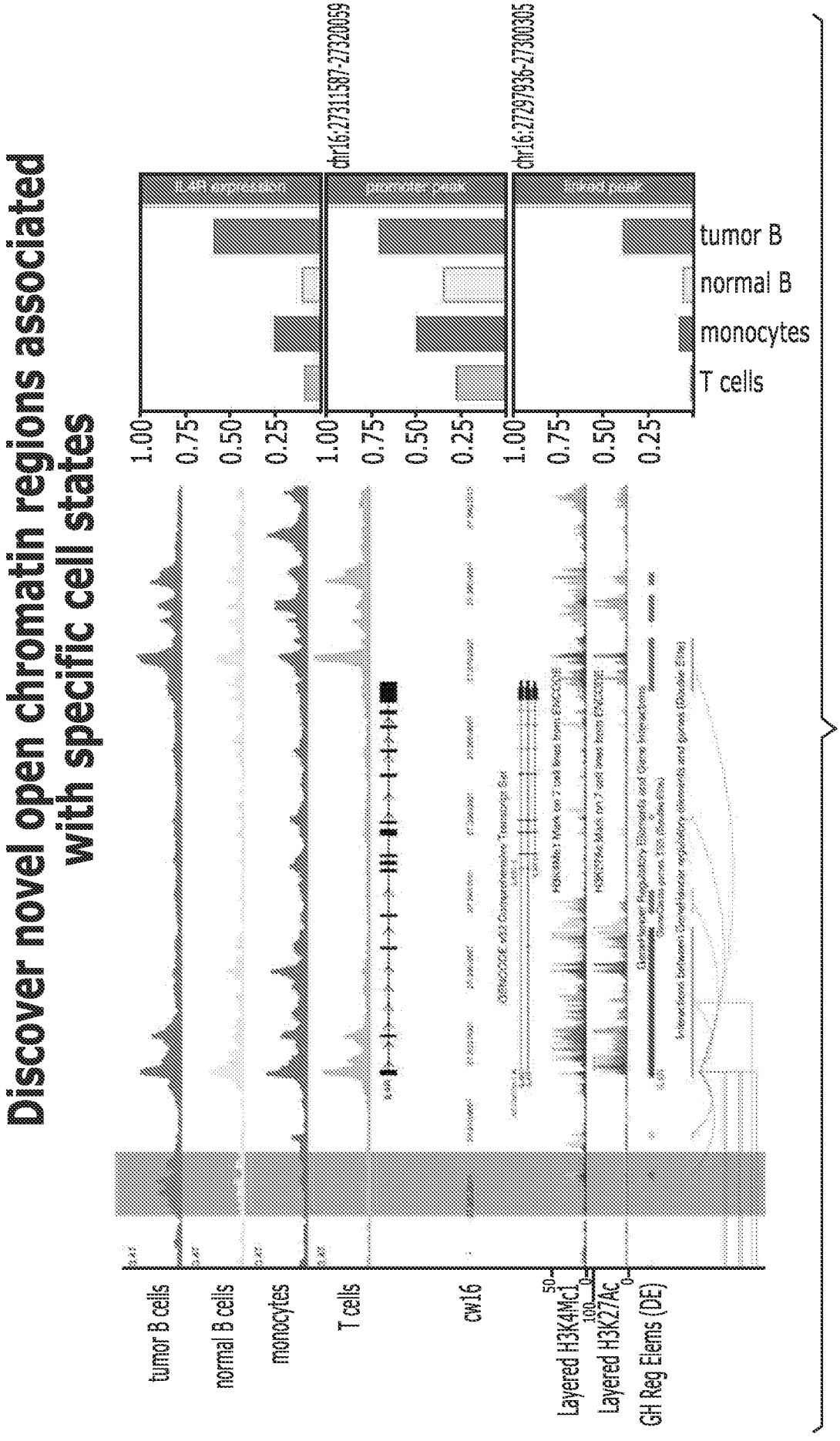

Discover novel open chromatin regions associated with specific cell states
chr16:27311587-27320059
chr16:27297936-27300305
tumor B
normal B
monocytes
T cells
1.00
0.75
0.50
0.25
tumor B cells
normal B cells
monocytes
T cells
cw16
Layered H3K4Me1
Layered H3K27Ac
GH Reg Elems (DE)

FIG. 38C

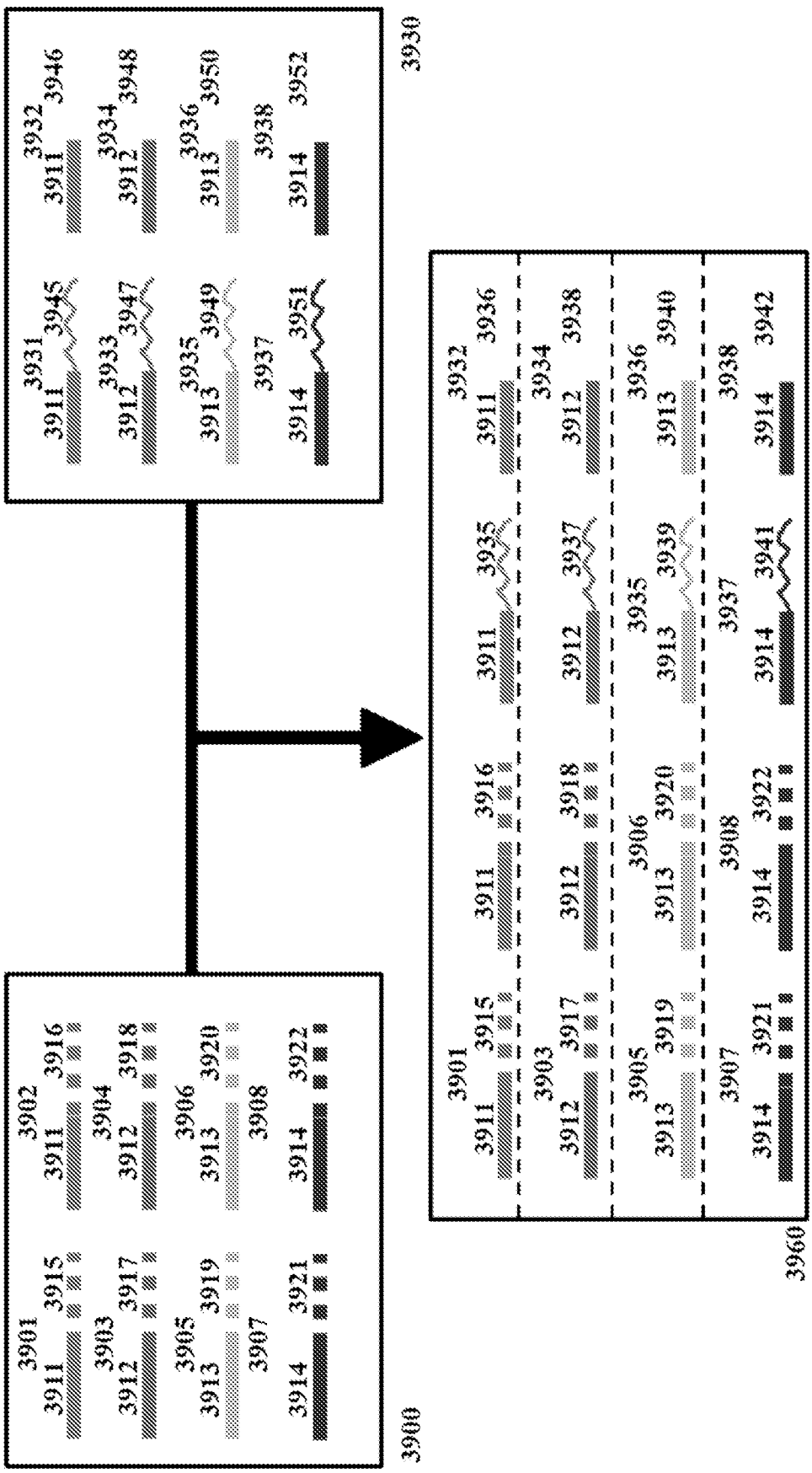

3900
3901
3911
3915
3902
3911
3916
3903
3912
3917
3904
3912
3918
3905
3913
3919
3906
3913
3920
3907
3908
3914
3921
3914
3922
3930
3931
3911
3945
3932
3911
3946
3933
3912
3947
3934
3912
3948
3935
3913
3949
3936
3913
3950
3937
3938
3914
3951
3914
3952
3960
3901
3911
3915
3911
3916
3911
3935
3932
3911
3936
3903
3912
3917
3912
3918
3912
3937
3934
3912
3938
3905
3913
3919
3913
3920
3913
3939
3935
3913
3940
3907
3914
3921
3914
3922
3914
3941
3937
3914
3942
3908
3906
3936
3938

FIG. 39

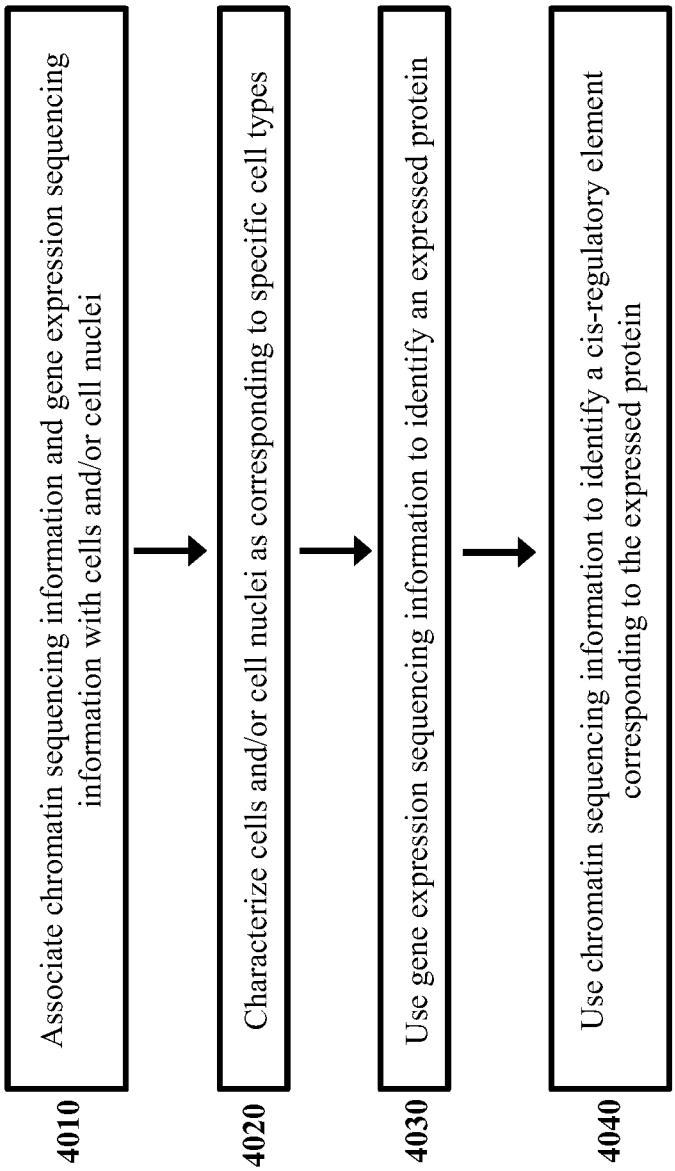

4010
Associate chromatin sequencing information and gene expression sequencing information with cells and/or cell nuclei
4020
Characterize cells and/or cell nuclei as corresponding to specific cell types
4030
Use gene expression sequencing information to identify an expressed protein
4040
Use chromatin sequencing information to identify a cis-regulatory element corresponding to the expressed protein

FIG. 40

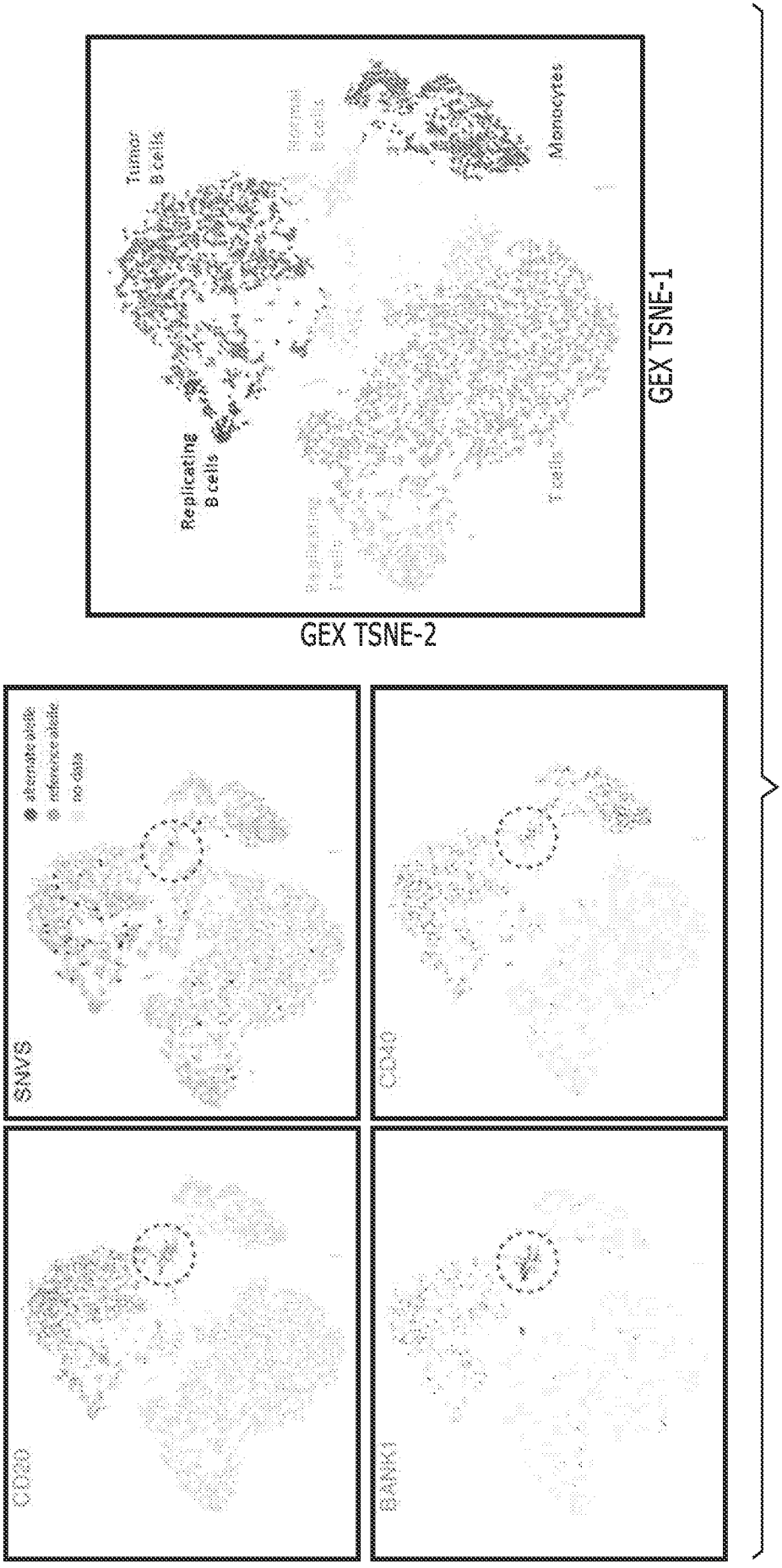
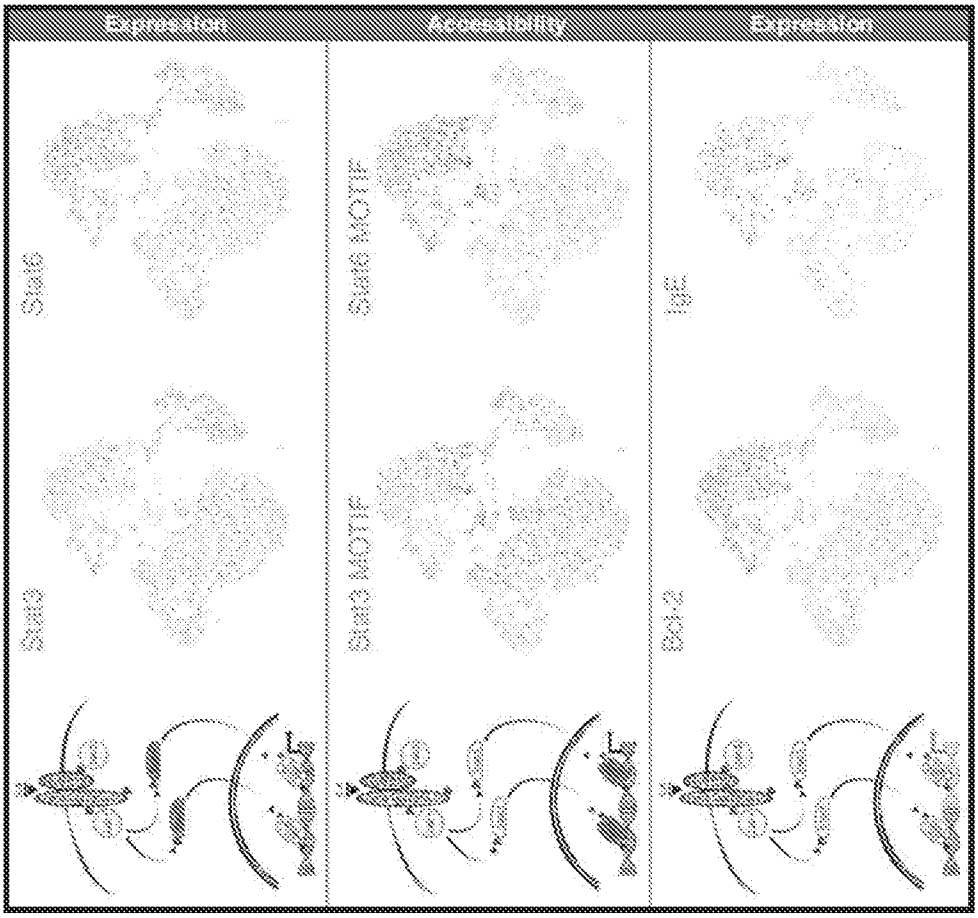

Tumor B cells
Monocytes
Replicating B cells
GEX TSNE-1
GEX TSNE-2
SNVS
CD40
CD20
BAFFR1

FIG. 41


Expression
Accessibility
Expression
Stat6
Stat6 MOTIF
Igf
Stat3
Stat3 MOTIF
Bcl-2

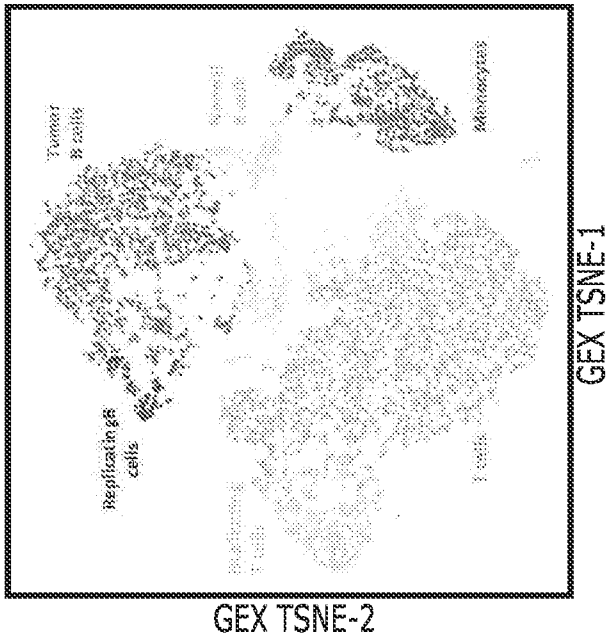

Tumor B cells
Monocytes
Replicating B cells
GEX TSNE-1
GEX TSNE-2

FIG. 42

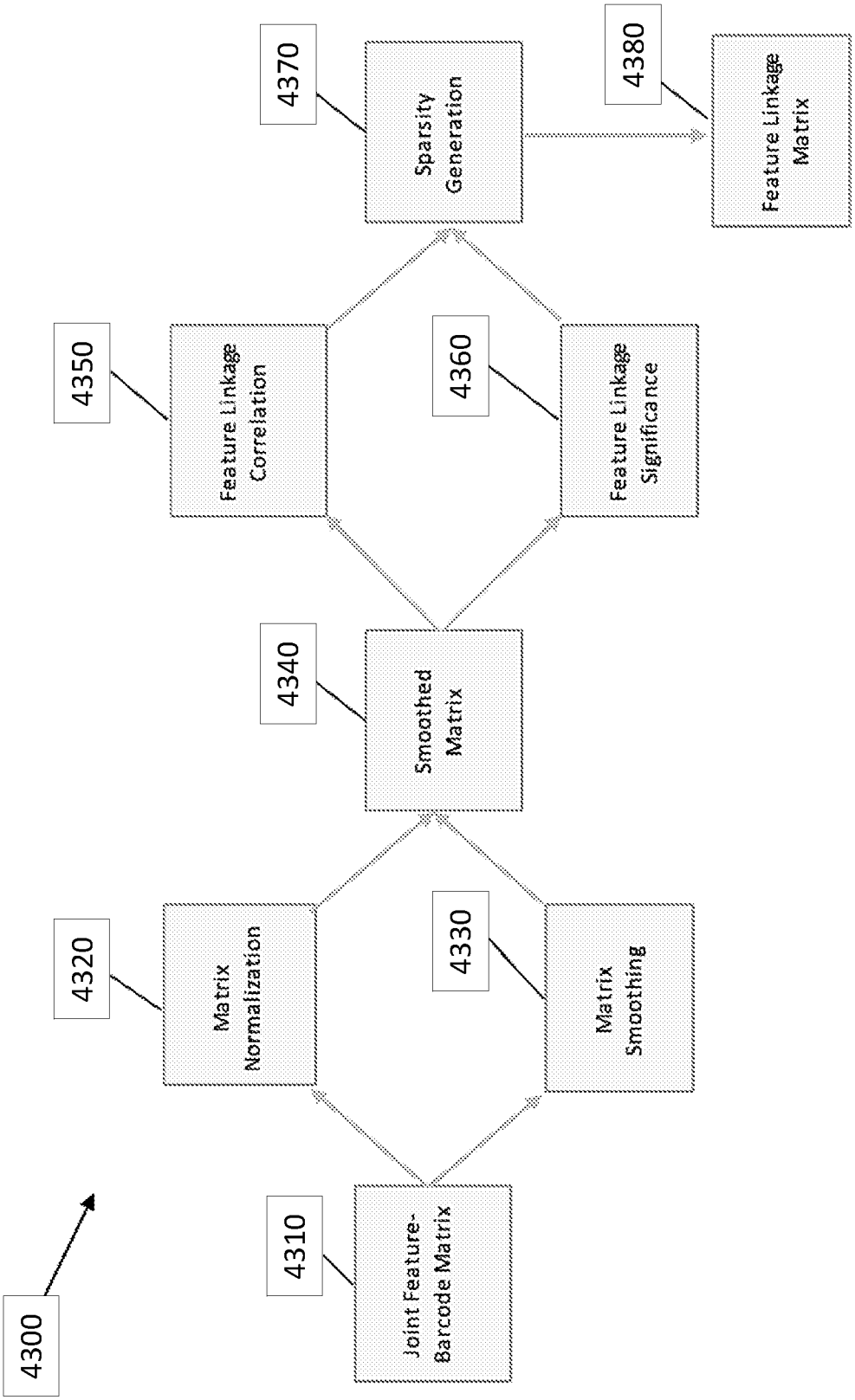

4300
4310
Joint Feature-Barcode Matrix
4320
Matrix Normalization
4330
Matrix Smoothing
4340
Smoothed Matrix
4350
Feature Linkage Correlation
4360
Feature Linkage Significance
4370
Sparsity Generation
4380
Feature Linkage Matrix

FIG. 43

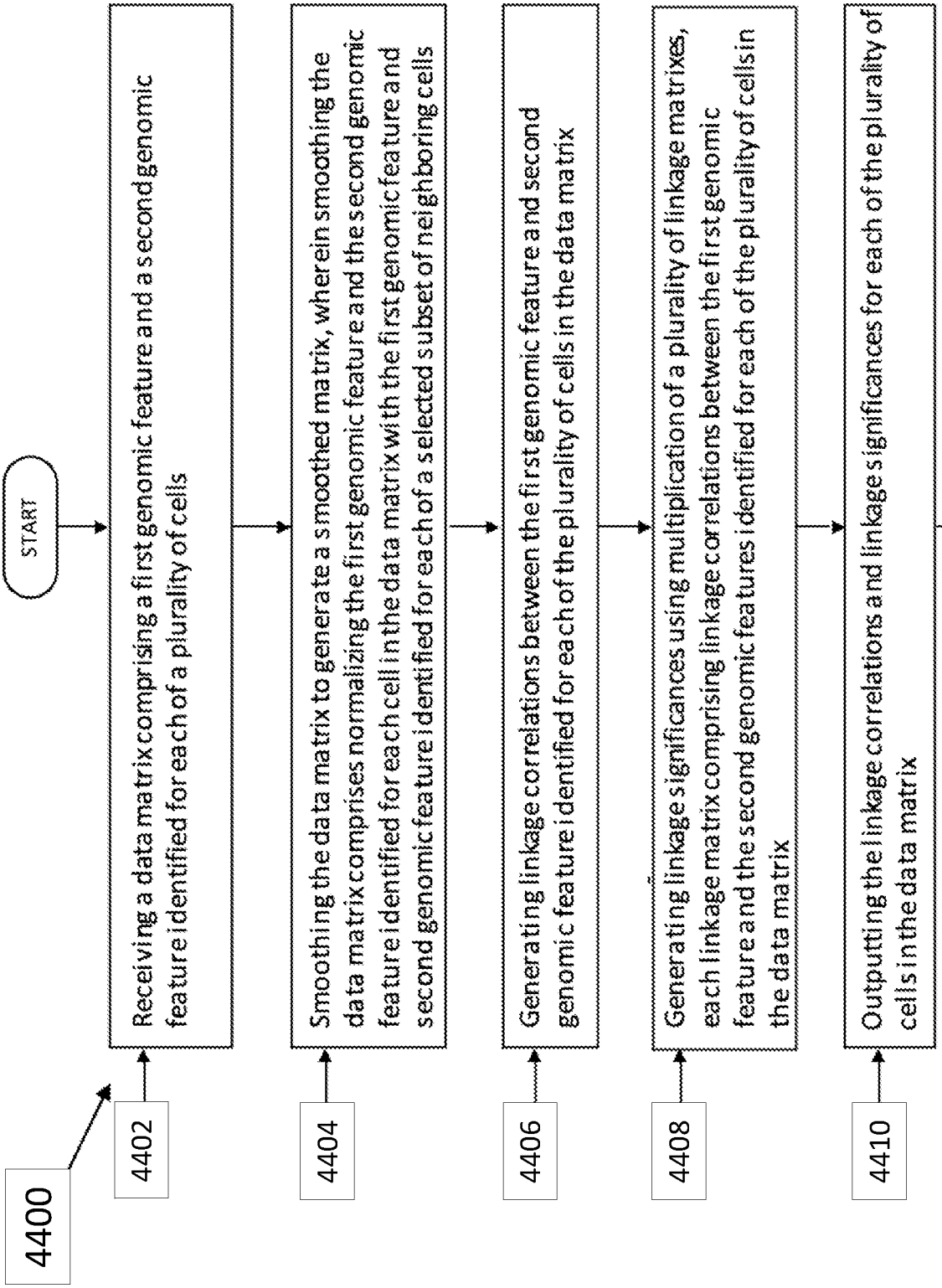

4400
START
4402
Receiving a data matrix comprising a first genomic feature and a second genomic feature identified for each of a plurality of cells
4404
Smoothing the data matrix to generate a smoothed matrix, wherein smoothing the data matrix comprises normalizing the first genomic feature and the second genomic feature identified for each cell in the data matrix with the first genomic feature and second genomic feature identified for each of a selected subset of neighboring cells
4406
Generating linkage correlations between the first genomic feature and second genomic feature identified for each of the plurality of cells in the data matrix
4408
Generating linkage significances using multiplication of a plurality of linkage matrixes, each linkage matrix comprising linkage correlations between the first genomic feature and the second genomic features identified for each of the plurality of cells in the data matrix
4410
Outputting the linkage correlations and linkage significances for each of the plurality of cells in the data matrix
END

FIG. 44

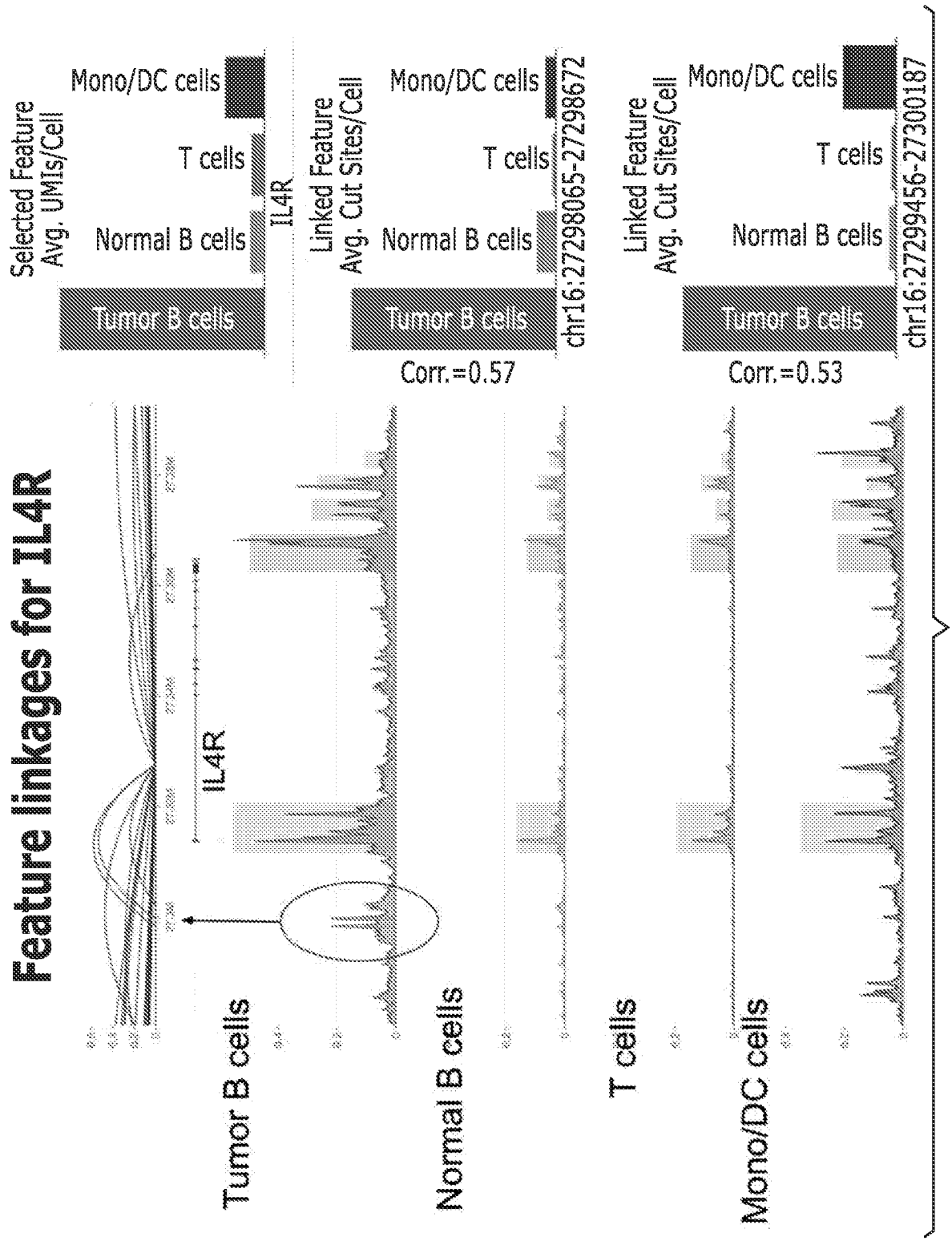

Feature linkages for IL4R
IL4R
Tumor B cells
Normal B cells
T cells
Mono/DC cells
Selected Feature
Avg. UMIs/Cell
Mono/DC cells
T cells
Normal B cells
Tumor B cells
IL4R
Linked Feature
Avg. Cut Sites/Cell
Mono/DC cells
T cells
Normal B cells
Tumor B cells
chr16:27298065-27298672
Corr.=0.57
Linked Feature
Avg. Cut Sites/Cell
Mono/DC cells
T cells
Normal B cells
Tumor B cells
chr16:27299456-27300187
Corr.=0.53

FIG. 45A

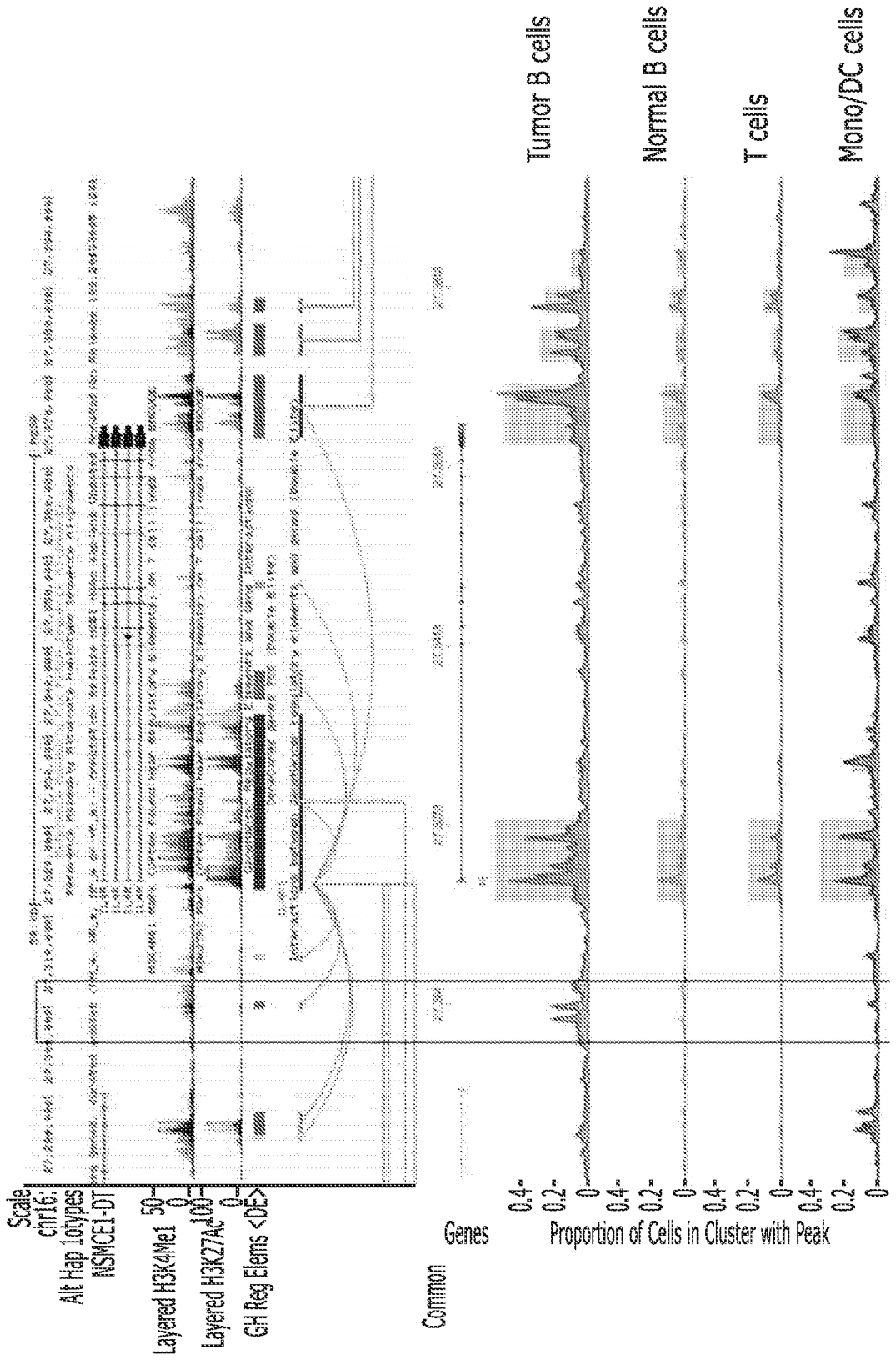

Scale
chr16:
Alt Hap 1otypes
NSMCE1-DT
Layered H3K4Me1
Layered H3K27Ac
GH Reg Elems
Common
Genes
Tumor B cells
Normal B cells
T cells
Mono/DC cells
Proportion of Cells in Cluster with Peak
0.4
0.2
0

FIG. 45B

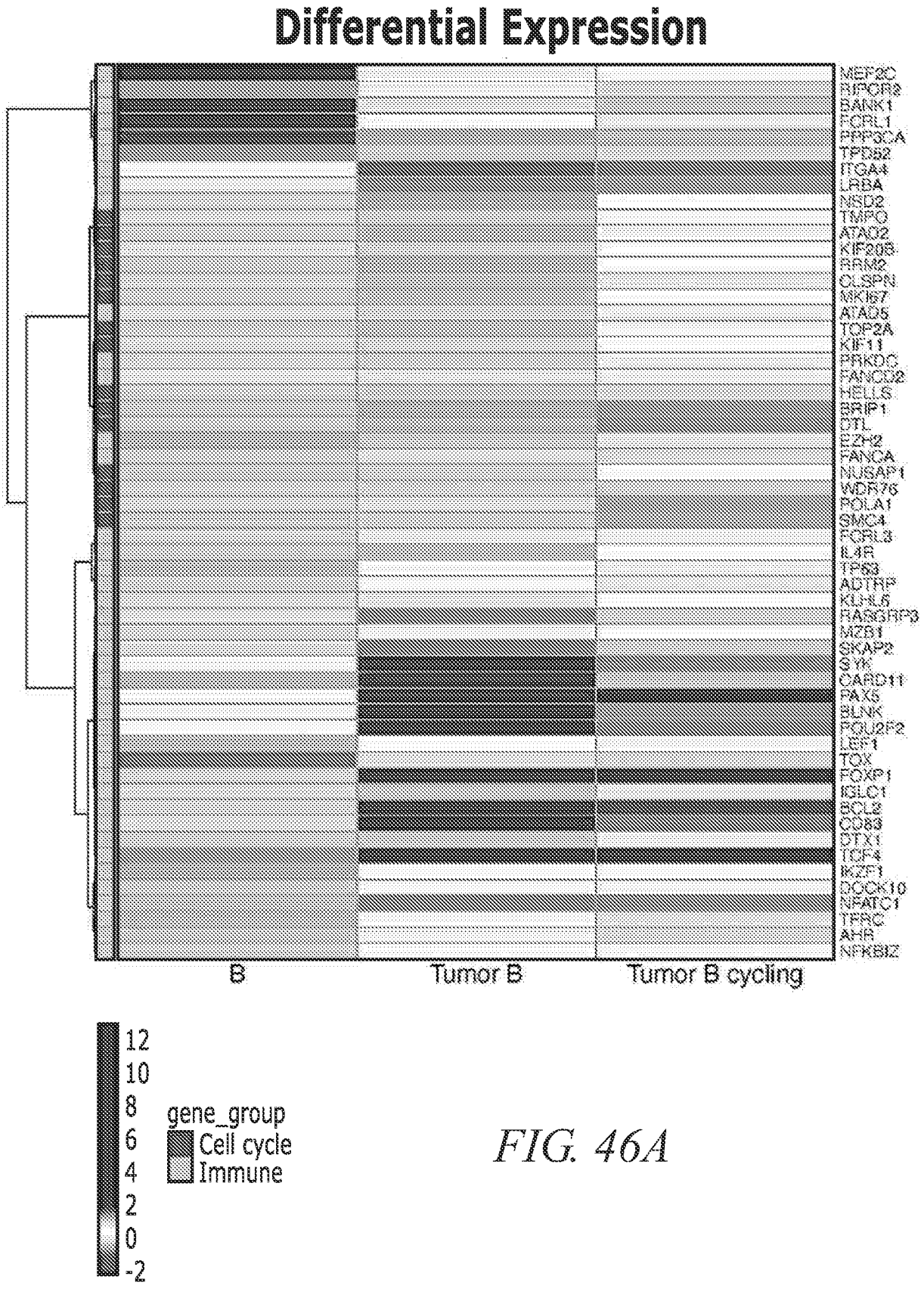

Differential Expression
MEF2C
RIPOR2
BANK1
FCRL1
PPP3CA
TPD52
ITGA4
LRBA
NSD2
TMPO
ATAD2
KIF20B
RRM2
OLSPN
MKI67
ATAD5
TOP2A
KIF11
PRKDC
FANCD2
HELLS
BRIP1
DTL
EZH2
FANCA
NUSAP1
WDR76
POLA1
SMC4
FCRL3
IL4R
TP53
ADTRP
KLHL6
RASGRP3
MZB1
SKAP2
SYK
CARD11
PAX5
BLNK
POU2F2
LEF1
TOX
FOXP1
IGLC1
BCL2
CD83
DTX1
TCF4
IKZF1
DOCK10
NFATC1
TFRC
AHR
NFKBIZ
B
Tumor B
Tumor B cycling
12
10
8
6
4
2
0
-2
gene_group
Cell cycle
Immune

FIG. 46A

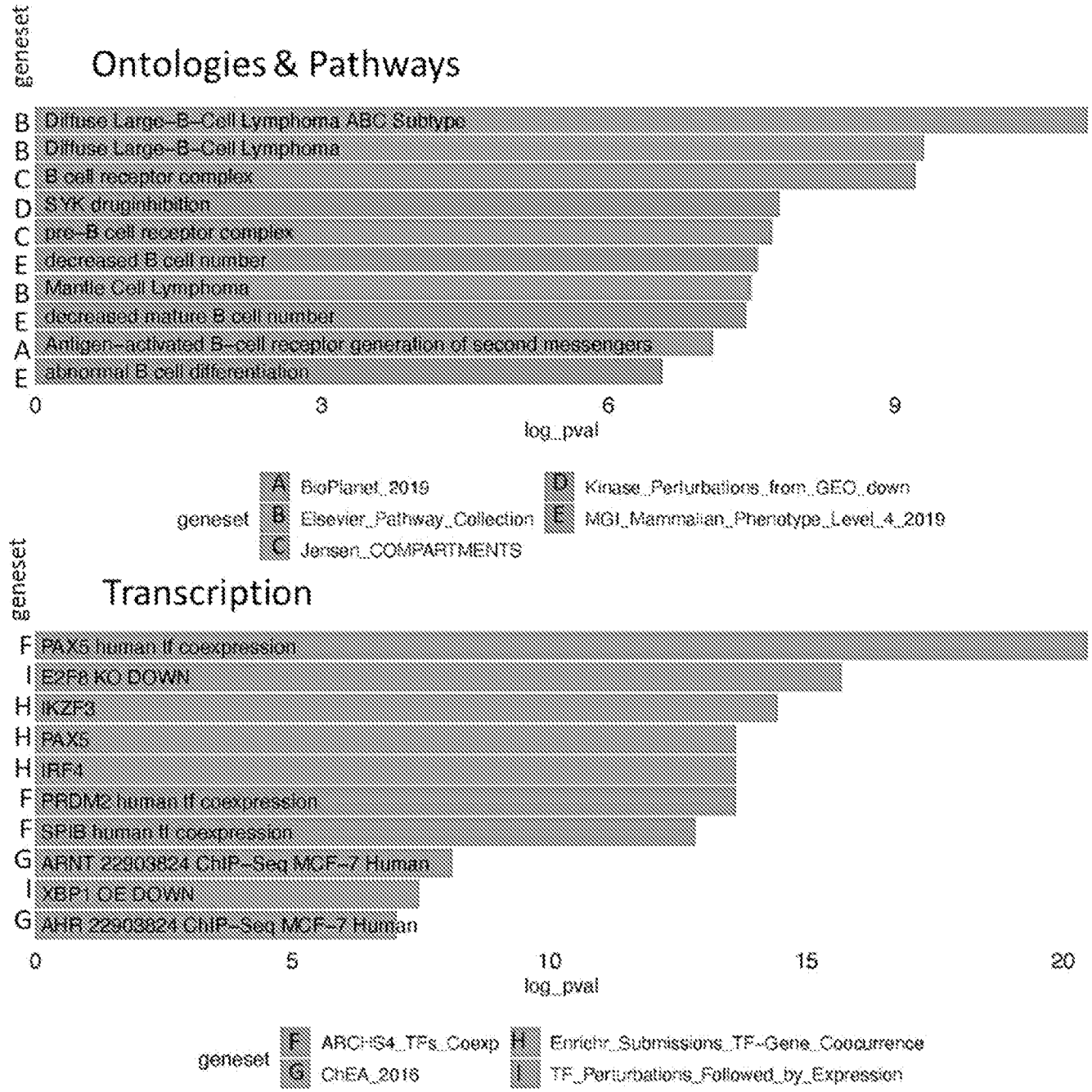

Ontologies & Pathways
geneset
B Diffuse Large−B−Cell Lymphoma ABC Subtype
B Diffuse Large−B−Cell Lymphoma
C B cell receptor complex
D SYK drug inhibition
C pre−B cell receptor complex
E decreased B cell number
B Mantle Cell Lymphoma
E decreased mature B cell number
A Antigen−activated B−cell receptor generation of second messengers
E abnormal B cell differentiation
0
3
6
9
log_pval
geneset
A BioPlanet_2019
B Elsevier_Pathway_Collection
C Jensen_COMPARTMENTS
D Kinase_Perturbations_from_GEO_down
E MGI_Mammalian_Phenotype_Level_4_2019
Transcription
geneset
F PAX5 human tf coexpression
I E2F8 KO DOWN
H IKZF3
H PAX5
H IRF4
F PRDM2 human tf coexpression
F SPIB human tf coexpression
G ARNT 22903824 ChIP−Seq MCF−7 Human
I XBP1 CE DOWN
G AHR 22903824 ChIP−Seq MCF−7 Human
0
5
10
15
20
log_pval
geneset
F ARCHS4_TFs_Coexp
G ChEA_2016
H Enrichr_Submissions_TF−Gene_Coocurrence
I TF_Perturbations_Followed_by_Expression

FIG. 46B

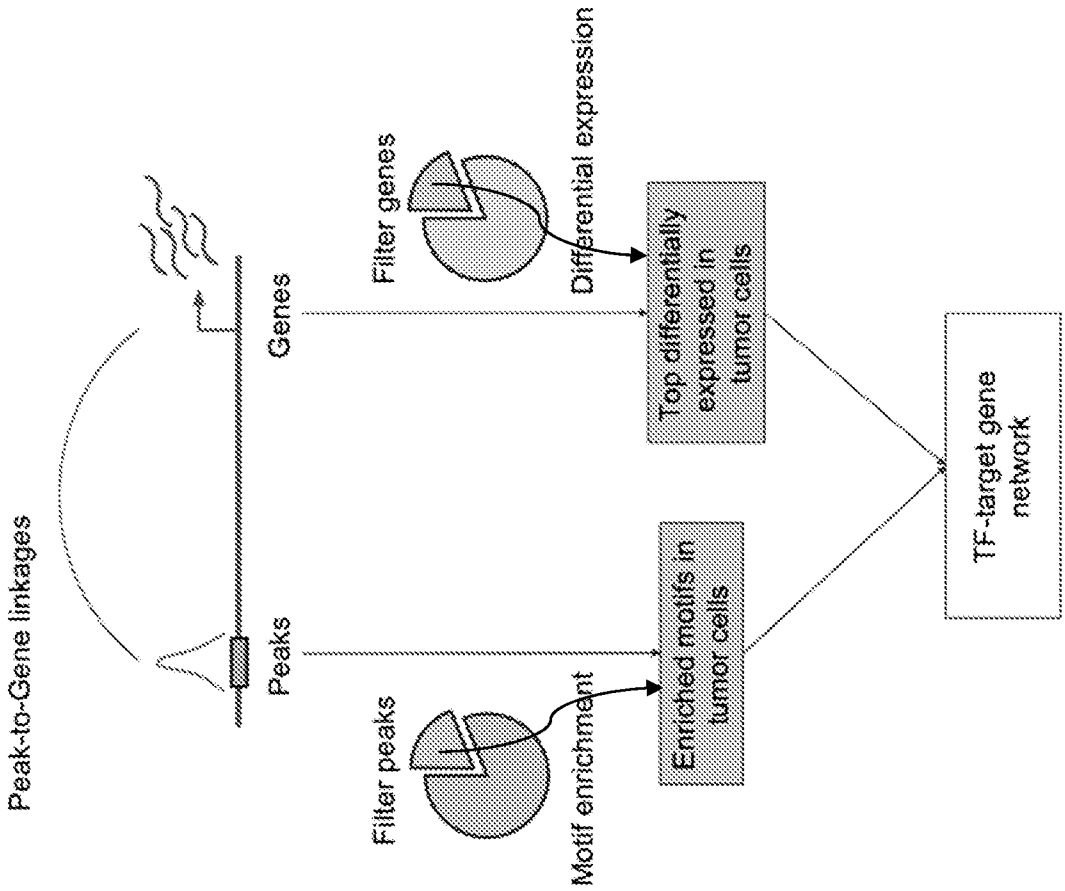

Peak-to-Gene linkages
Genes
Peaks
Filter genes
Differential expression
Top differentially expressed in tumor cells
Filter peaks
Motif enrichment
Enriched motifs in tumor cells
TF-target gene network

FIG. 47

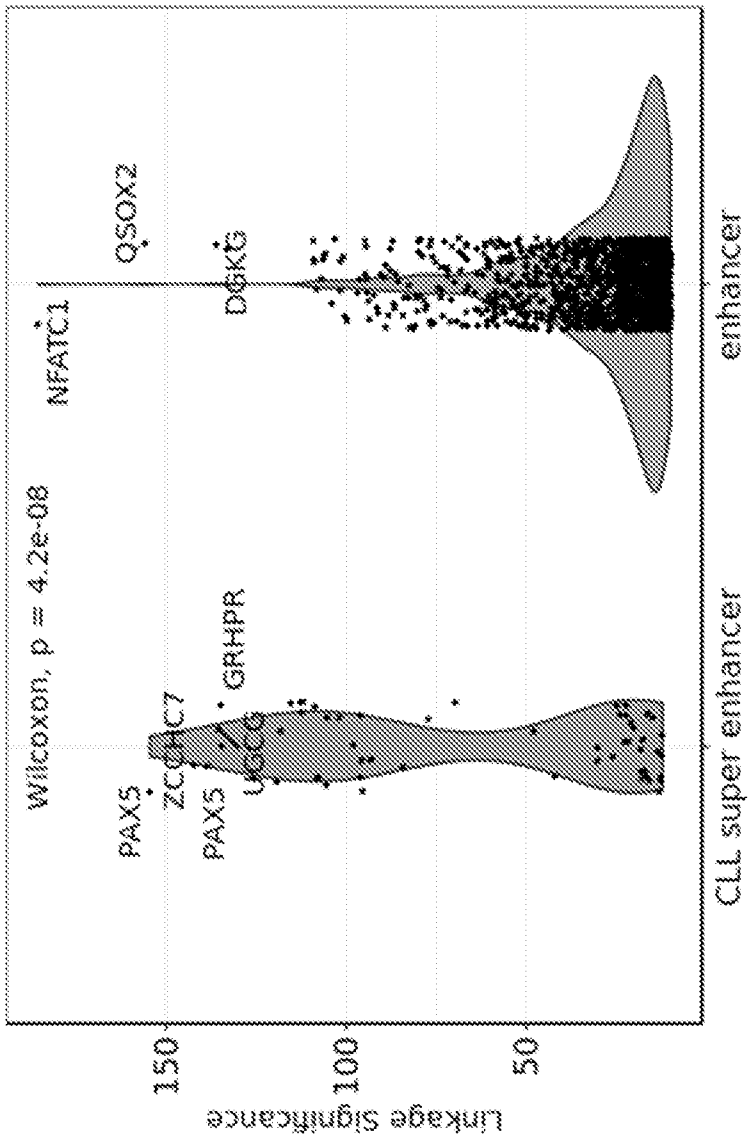

Linkage Significance
150
100
50
CLL super enhancer
enhancer
Wilcoxon, p = 4.2e-08
PAX5
ZCCHC7
PAX5
GRHPR
NFATC1
QSOX2
DGKG

FIG. 48A

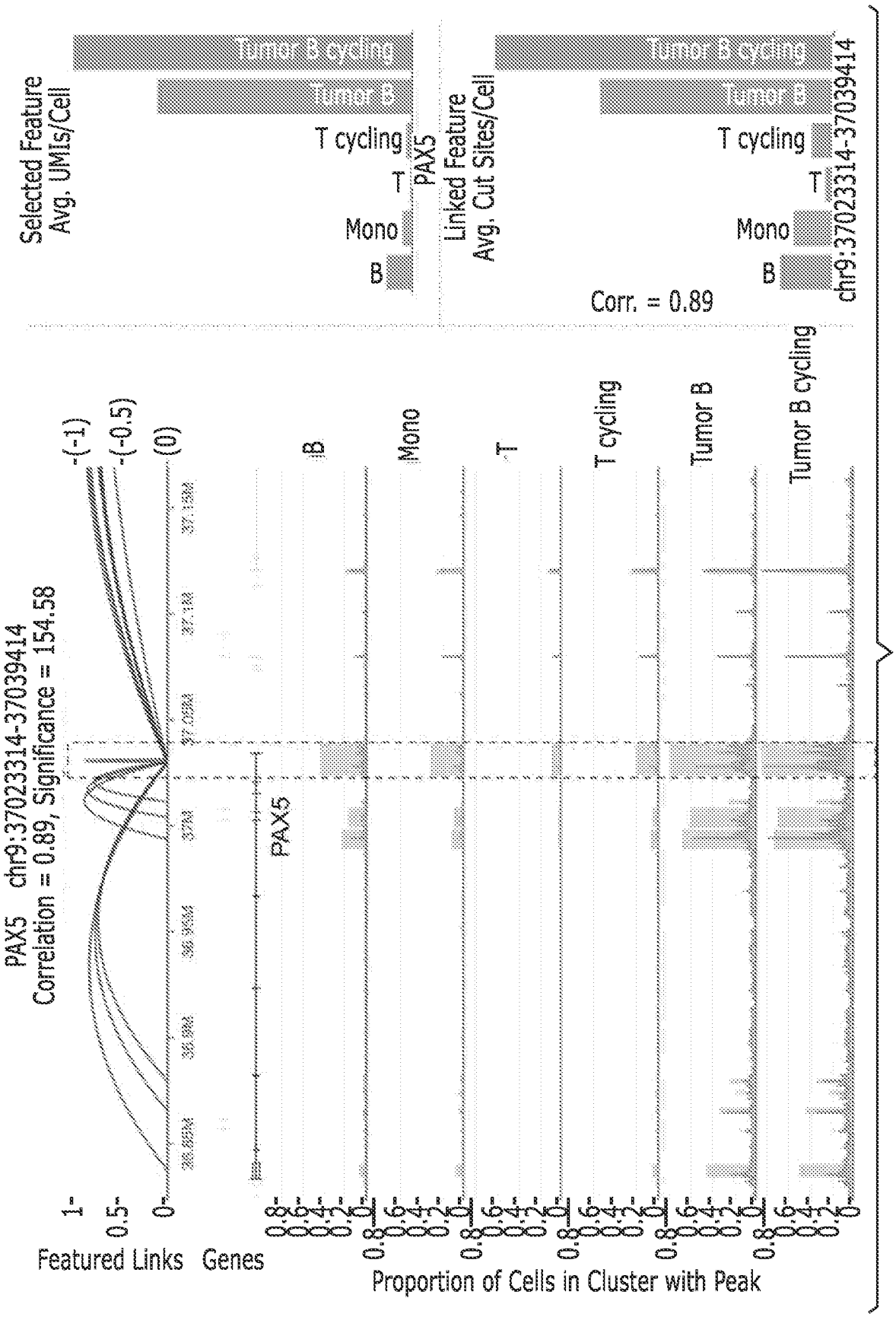

PAX5 chr9:37023314-37039414
Correlation = 0.89, Significance = 154.58
-(-1)
-(-0.5)
(0)
PAX5
Featured Links
Genes
B
Mono
T
T cycling
Tumor B
Tumor B cycling
Proportion of Cells in Cluster with Peak
Selected Feature
Avg. UMIs/Cell
Linked Feature
Avg. Cut Sites/Cell
Corr. = 0.89
chr9:37023314-37039414

FIG. 48B

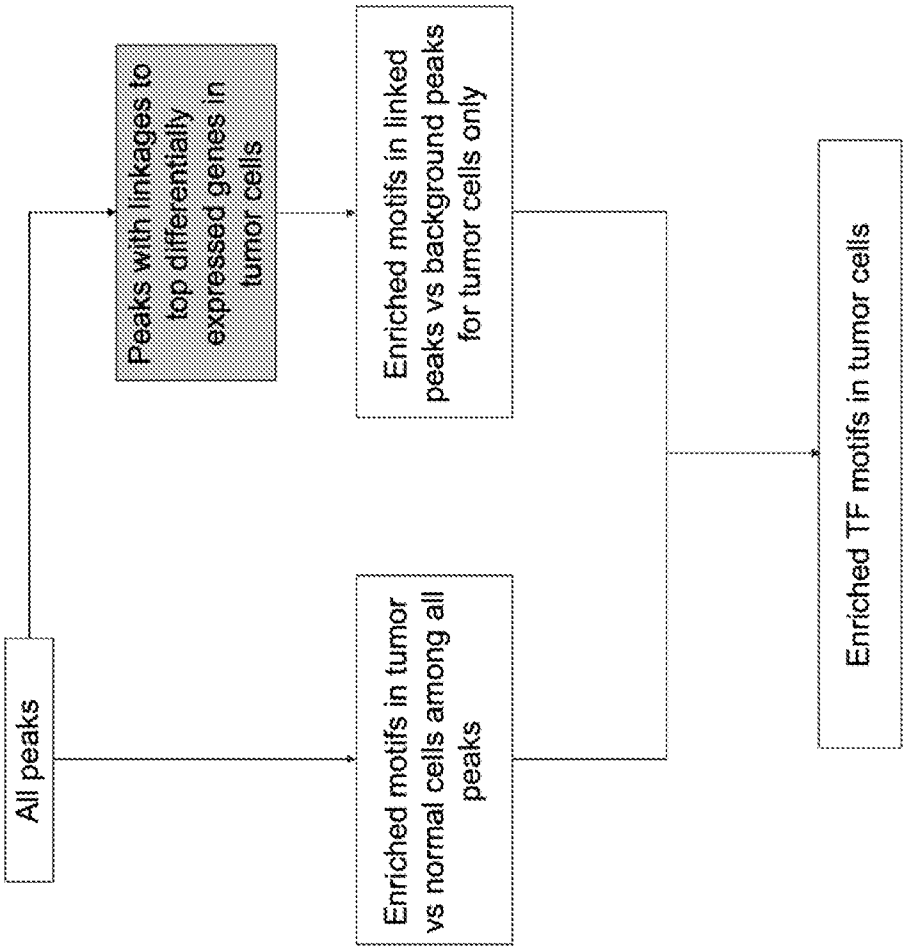

All peaks
Peaks with linkages to top differentially expressed genes in tumor cells
Enriched motifs in tumor vs normal cells among all peaks
Enriched motifs in linked peaks vs background peaks for tumor cells only
Enriched TF motifs in tumor cells

FIG. 49A

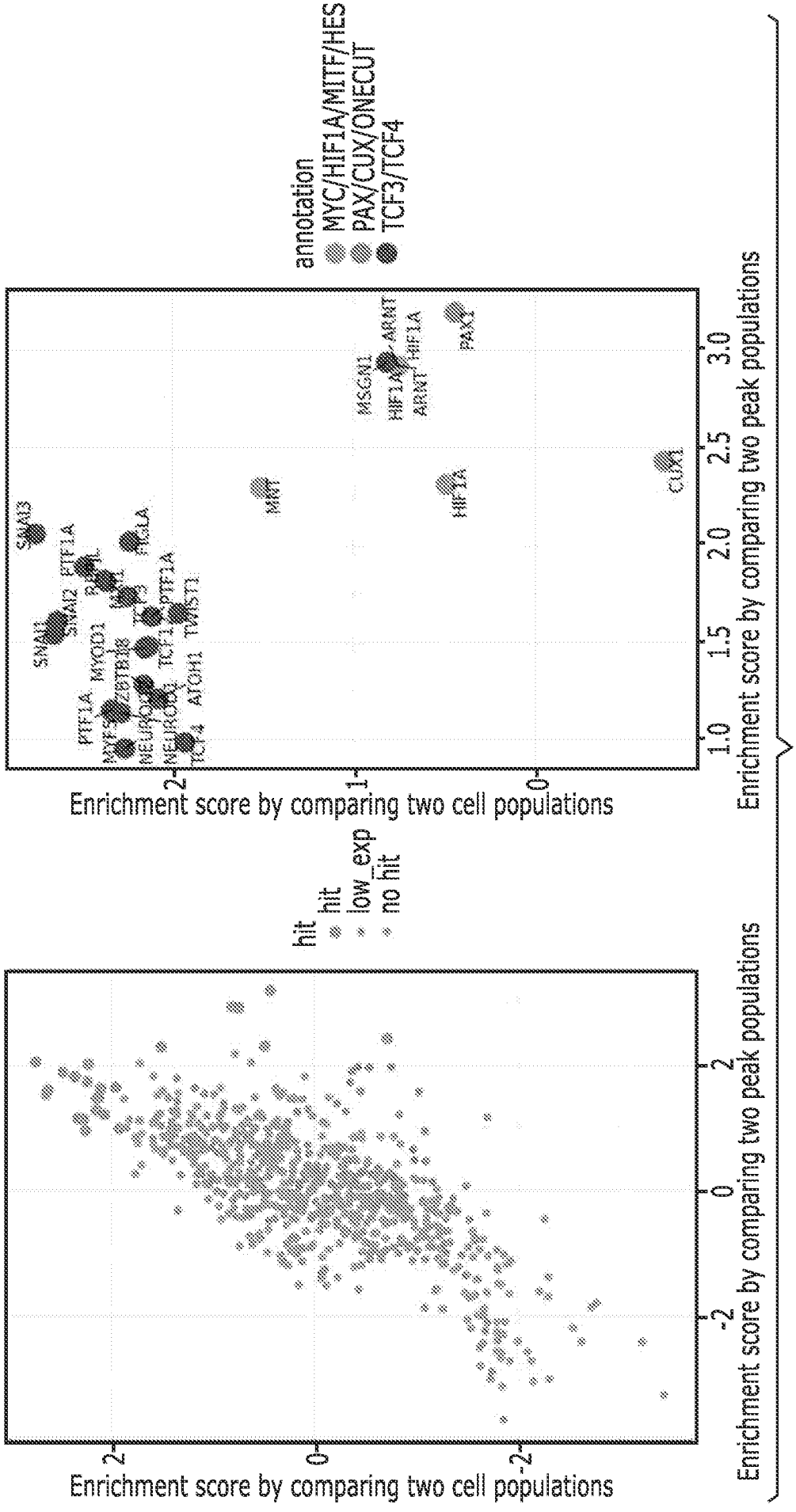

annotation
MYC/HIF1A/MITF/HES
PAX/CUX/ONECUT
TCF3/TCF4
SNAI3
SNAI2
FTF1A
MYCD1
ZBTB18
TCF4
ATOH1
TWIST1
MITF
MSGN1
HIF1A
ARNT
PAX7
CUX1
Enrichment score by comparing two cell populations
Enrichment score by comparing two peak populations
hit
hit
low_exp
no hit

FIG. 49B

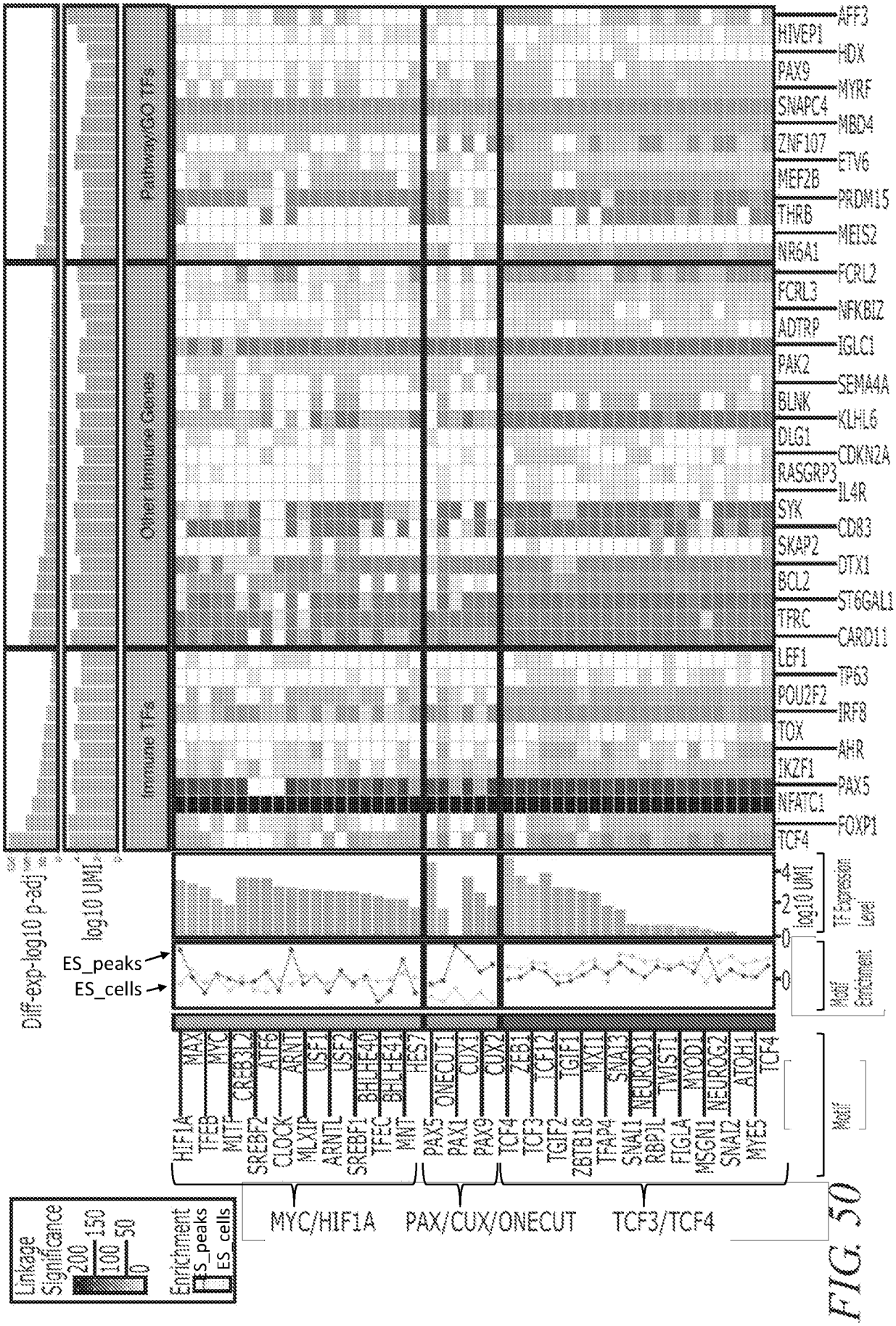

Diff-exp-log10 p-adj
log10 UMI
Pathway/GO TFs
Other Immune Genes
Immune TFs
AFF3
HIVEP1
HDX
PAX9
MYRF
SNAPC4
MBD4
ZNF107
ETV6
MEF2B
PRDM15
THRB
MEIS2
NR6A1
FCRL2
FCRL3
NFKBIZ
ADTRP
IGLC1
PAK2
SEMA4A
BLNK
KLHL6
DLG1
CDKN2A
RASGRP3
IL4R
SYK
CD83
SKAP2
OTX1
BCL2
ST6GAL1
TFRC
CARD11
LEF1
TP63
POU2F2
IRF8
TOX
AHR
IKZF1
PAX5
NFATC1
FOXP1
TCF4
TF Expression log10 UMI Level
ES_peaks
ES_cells
Motif Enrichment
MAX
MYC
CR3B3L2
ATF6
ARNT
USF1
USF2
BHLHE40
BHLHE41
HES7
MNT
ONECUT1
CUX1
CUX2
ZEB1
TCF12
TGIF1
MXI1
SNAI3
NEUROD1
TWIST1
MYOD1
NEUROG2
ATOH1
TCF4
HIF1A
TFEB
MITF
SREBF2
CLOCK
MLXIP
ARNTL
SREBF1
TFEC
MNT
PAX5
PAX1
PAX9
TCF4
TCF3
TGIF2
ZBTB18
TFAP4
SNAI1
RBPJL
FIGLA
MSGN1
SNAI2
MYF5
Motif
MYC/HIF1A
PAX/CUX/ONECUT
TCF3/TCF4
Linkage Significance
200
150
100
50
0
Enrichment
ES_peaks
ES_cells

FIG. 50

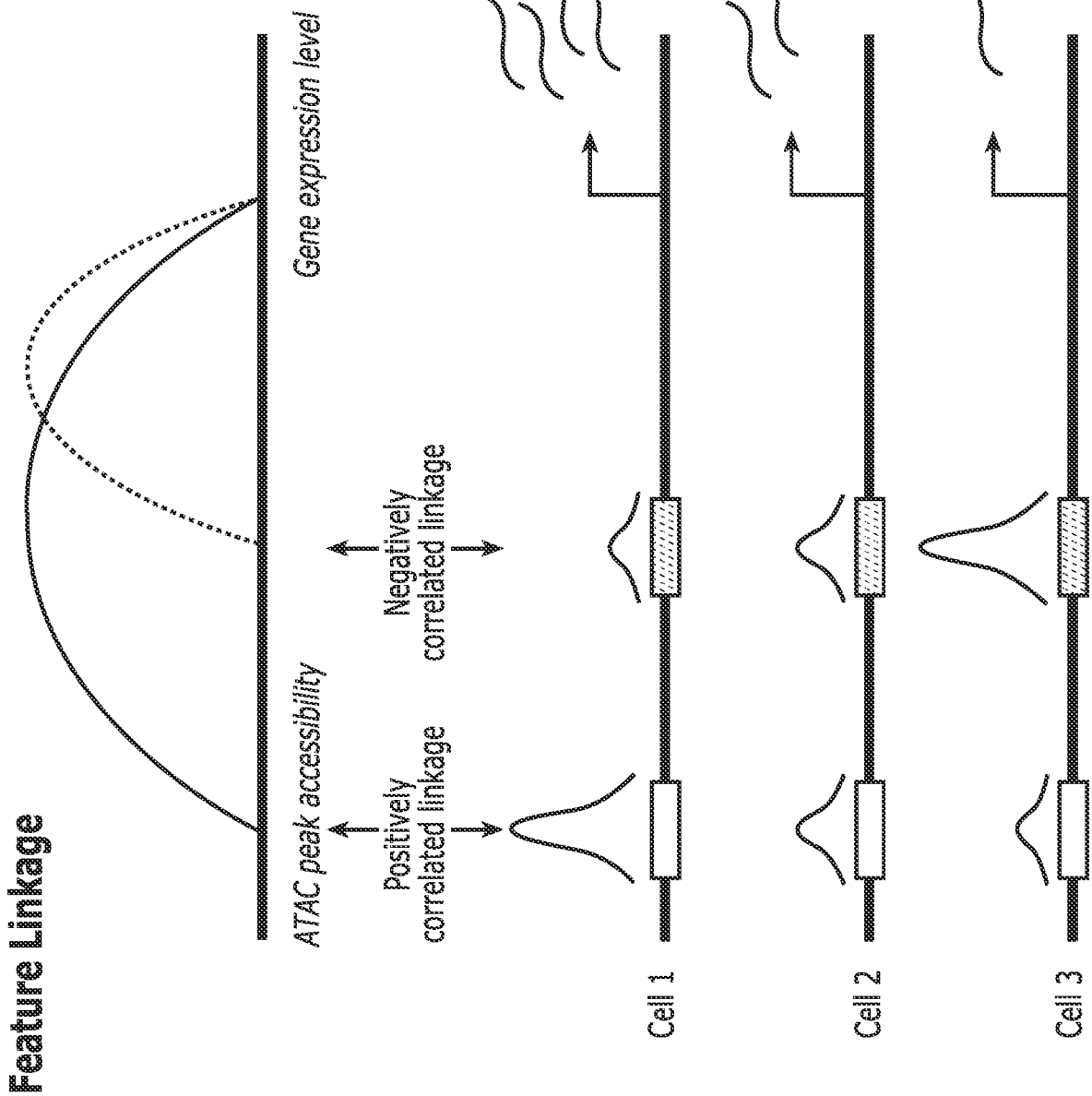

Feature Linkage
ATAC peak accessibility
Gene expression level
Positively correlated linkage
Negatively correlated linkage
Cell 1
Cell 2
Cell 3

FIG. 51