Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 3138806 A1 2020/11/26

(21) 3 138 806

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) Date de dépôt PCT/PCT Filing Date: 2020/05/22
(87) Date publication PCT/PCT Publication Date: 2020/11/26
(85) Entrée phase nationale/National Entry: 2021/11/19
(86) N° demande PCT/PCT Application No.: US 2020/034404
(87) N° publication PCT/PCT Publication No.: 2020/237222
(30) Priorités/Priorities: 2019/05/22 (US62/851,448); 2019/09/23 (US62/904,374)

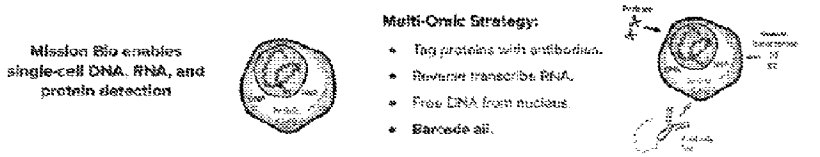
(51) Cl.Int./Int.Cl. *C12Q 1/6834* (2018.01)

(71) Demandeur/Applicant:
MISSION BIO, INC., US

(72) Inventeurs/Inventors:
DHINGRA, DALIA, US;
OOI, AIK, US;
MENDEZ, PEDRO, US;
RUFF, DAVID, US

(74) Agent: GOWLING WLG (CANADA) LLP

(54) Titre : METHODE ET APPAREIL DE SEQUENCAGE CIBLE SIMULTANE D'ADN, D'ARN ET DE PROTEINE
(54) Title: METHOD AND APPARATUS FOR SIMULTANEOUS TARGETED SEQUENCING OF DNA, RNA AND PROTEIN



The workflow is enabled by its two droplet steps (below). Cells are individually encapsulated in a first droplet (1A) where analyte preparation can occur, including aggressive digestion and subsequent protease heat inactivation (1B). Afterwards, sensitive enzymes that are incompatible with the earlier preparation can be added to make a new drop (2A) where barcoding and amplification proceed (2B).

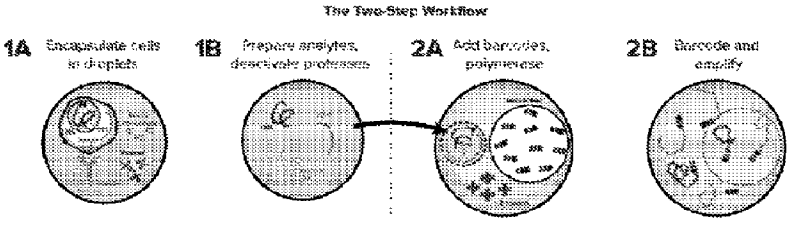


FIGURE 1

(57) Abrégé/Abstract:
Provided herein are methods and systems for the simultaneous targeted detection and sequencing of DNA, RNA, and Protein. In typical embodiments, the DNA, RNA, and Proteins are detected, characterized, and sequenced using just a single mammalian cell. One embodiment of detecting and characterizing DNA, RNA, or protein from a mammalian cell includes encapsulating a single cell in a drop and performing a protease digest on the encapsulated cell drop, performing a reverse transcriptase reaction; performing a droplet merger with barcoding PCR reagents and barcoding beads; performing a PCR reaction to attach the cell barcodes to the DNA targeted amplicons, RNA targeted amplicons, and protein tag amplicons, where all amplicons from the same emulsion contain the same cell barcode; and detecting and characterizing a DNA, RNA, or protein amplicon by sequencing the cell barcode incorporated into each amplicon.

Date Submitted: 2021/11/19

CA App. No.: 3138806

Abstract:

Provided herein are methods and systems for the simultaneous targeted detection and sequencing of DNA, RNA, and Protein. In typical embodiments, the DNA, RNA, and Proteins are detected, characterized, and sequenced using just a single mammalian cell. One embodiment of detecting and characterizing DNA, RNA, or protein from a mammalian cell includes encapsulating a single cell in a drop and performing a protease digest on the encapsulated cell drop, performing a reverse transcriptase reaction; performing a droplet merger with barcoding PCR reagents and barcoding beads; performing a PCR reaction to attach the cell barcodes to the DNA targeted amplicons, RNA targeted amplicons, and protein tag amplicons, where all amplicons from the same emulsion contain the same cell barcode; and detecting and characterizing a DNA, RNA, or protein amplicon by sequencing the cell barcode incorporated into each amplicon.

METHOD AND APPARATUS FOR SIMULTANEOUS TARGETED SEQUENCING OF DNA, RNA AND PROTEIN

FIELD

[0001] This invention relates generally to methods and systems for the simultaneous targeted detection and sequencing of DNA, RNA, and Protein, and more particularly performing this analysis from a single cell.

RELATED APPLICATIONS

[0002] This application takes priority to the following U.S. Provisional Application U.S.S.N. 62/851,448 filed May 22, 2019 by D. Dhingra et al., and entitled 'Method and Apparatus For Simultaneous Targeted Sequencing Of DNA, RNA And Protein'; and U.S.S.N., all incorporated by reference herein.

BACKGROUND

[0003] The development of multiomic approaches to studying analytes in a human cell holds many promises for increasing our understanding and for developing new therapies. However, these technologies have yet to fulfil this promise yet due to the complexity of such systems, limitations, road blocks and problems with current methods and systems, such as the low amount of biological sample available from small samples (e.g. cells, or a cell). There is a need for method, system and apparatus to provide high-throughput, single-cell analysis that incorporates targeted, DNA, RNA and protein detection and characterization. There is also a need for a system that can be customized for the detection of particular analytes. The inventions provided here address these unmet needs.

BRIEF SUMMARY

[0004] The inventions described and claimed herein have many attributes and embodiments including, but not limited to, those set forth or described or referenced in this Brief Summary. The inventions described and claimed herein are not limited to, or by, the features or embodiments identified in this Summary, which is included for purposes of illustration only and not restriction.

[0005] In a first aspect, embodiments of the invention are directed to methods for the simultaneous targeted detection and sequencing of DNA, RNA, and Protein. In preferred embodiments, the DNA, RNA, and proteins are detected, characterized, and sequenced using just a single cell. A multiomic detection and characterization method provided herein may utilize the following novel strategy: i) proteins are tagged with antibodies, ii) RNA is reverse transcribed, iii) DNA is released from the cell nucleus, and iv) each of the preceding is tagged with a cell identifier (e.g. a barcode) so that DNA, RNA, and proteins from the same cell will have the same identifier that is unique to that cell.

[0006] One embodiment of a multiomic detection and characterization method for detecting DNA, RNA, or protein from a single cell includes, independent of order, the following steps: encapsulating a cell

in a drop comprising a reaction mixture comprising a protease; performing a protease digest on the encapsulated cell drop with the protease to produce a cell lysate; providing a reverse transcriptase and performing a reverse transcription reaction; performing a droplet merger with barcoding PCR reagents and barcoding beads; performing a PCR reaction to attach the cell barcodes to the DNA targeted amplicons, RNA targeted amplicons, and protein tag amplicons, wherein all amplicons from the same emulsion contain the same cell barcode; performing a capture of DNA and RNA amplicons to a solid phase, wherein protein tag amplicons are separated from DNA and RNA amplicons; and detecting and characterizing a DNA, RNA, or protein amplicon by sequencing the cell barcode incorporated into each amplicon. In a preferred embodiment, DNA, RNA, and proteins are detected and characterized. Also, the DNA, RNA, and proteins are detected and characterized from a single cell in preferred embodiments.

[0007] In some embodiments, the reverse transcription reaction is performed in the same drop as the protease digest. In other embodiments, a droplet merger with the cell lysate and reverse transcriptase is performed before the reverse transcription reaction is performed. Typically, the protease and reverse transcriptase reactions are performed on a PCR thermocycler and later transferred to another instrument for processing and analysis. The preferred instrument for processing and analysis comprises a Mission Bio, Tapestri™ system and the droplet merger with barcoding PCR reagents and barcoding beads is performed on this instrument. In some embodiments, a nucleic acid concentration step is performed.

[0008] In one particular exemplary embodiment, a sample of cells is obtained and the cells are stained. The single cells are encapsulated with a buffer containing components for lysis, reverse transcription, and a protease treatment. The reactions are performed on a thermocycler and the encapsulated cells transferred to suitable platform, preferably a Mission Bio, Tapestri™ system, to perform a droplet merger with barcoding PCR reagents and barcoding beads. A PCR is performed which attaches the cell barcodes to the DNA targeted amplicons, RNA targeted amplicons, and protein tag amplicons. All amplicons from the same emulsion contain the same cell barcode. After emulsions are broken, an exonuclease reaction, and a SPRI cleanup may be performed. The supernatant is kept to process into protein libraries. The SPRI beads contain both the DNA and RNA amplicons that are then separated using a biotin capture oligo and streptavidin beads. The DNA and RNA amplicons are typically, but not always, processed separately to form sequencing libraries. This embodiment may use a two-droplet process where cells are individually encapsulated in a first droplet where analyte preparation can occur, including aggressive protease digestion and subsequent protease heat inactivation. Afterwards, additional reagents and enzymes can be merged into a second drop in which a barcoding and amplification reaction takes place.

[0009] In some embodiments, one or more DNA, RNA, and protein listed in Figure 9 is detected and characterized. In some embodiments, two or more DNA, RNA, and protein listed in Figure 9 are detected and characterized. In some embodiments, three or more DNA, RNA, and protein listed in Figure 9 are detected and characterized. In some embodiments, four or more DNA, RNA, and protein listed in Figure 9 are detected and characterized. In some embodiments, five or more DNA, RNA, and protein listed in Figure 9 are detected and characterized. In some embodiments, at least five DNA, RNA, and proteins

listed in Figure 9 are detected and characterized. In some embodiments, at least ten DNA, RNA, and proteins listed in Figure 9 are detected and characterized. In some embodiments, at least twenty DNA, RNA, and proteins listed in Figure 9 are detected and characterized.

[0010] In another aspect, a method for preparing a protein library and a DNA library which can be paired based on the cell barcode is provided.

[0011] In another aspect, a method for preparing a protein library and an RNA library which can be paired based on the cell barcode is provided. These embodiments do not typically utilize a protease treatment as part of the cell preparation.

[0012] In another aspect, a method for preparing a protein library, DNA library, and RNA library which can be paired based on the cell barcode is provided.

BRIEF DESCRIPTION OF THE DRAWINGS

[0013] Figure 1 is a schematic diagram of a multiomic strategy used in some embodiments to detect DNA, RNA, and protein from a single cell.

[0014] Figure 2 is a schematic diagram of the first droplet of an exemplary embodiment of the invention.

[0015] Figure 3 is a schematic diagram of the second droplet of an exemplary embodiment of the invention.

[0016] Figure 4 is a schematic diagram of the mutliomics workflow. Stained cells are input onto the Tapestri platform. The single cells are encapsulated with a buffer containing components for lysis, reverse transcription, and a protease treatment. These reactions are performed on a thermocycler and the encapsulated cells returned to the Tapestri platform to for droplet merger with barcoding PCR reagents and barcoding beads. A PCR is performed which attaches the cell barcodes to the DNA targeted amplicons, RNA targeted amplicons, and protein tag amplicons. All amplicons from the same emulsion would contain the same cell barcode. After emulsions are broken, an exonuclease reaction and a SPRI cleanup are performed. The supernatant is kept to process into protein libraries. The SPRI beads contain both the DNA and RNA amplicons that are then separated using a biotin capture oligo and streptavidin beads. The DNA and RNA amplicons are then processed separately to form sequencing libraries.

[0017] Figure 5 is Bioanalyzer trace of a DNA library from the untreated cells in tubes 4-8 from the multiomics experiment from Example I where 3 cell lines (Jurkat, K-562, and KCL-22) were untreated or treated 3 different doses of imatinib (10 uM, 100 uM, and 250 uM). 1 uL of undiluted library was loaded onto a HS DNA chip.

[0018] Figure 6 is Bioanalyzer trace of a protein library from the untreated cells in tubes 4-8 from the multiomics experiment from Example I where 3 cell lines (Jurkat, K-562, and KCL-22) were untreated

or treated 3 different doses of imatinib (10 uM, 100 uM, and 250 uM), 1 uL of undiluted library was loaded onto a HS DNA chip.

[0019] Figure 7 is a Bioanalyzer trace of a RNA library from the untreated cells in tubes 4-8 from the multiomics experiment from Example I where 3 cell lines (Jurkat, K-562, and KCL-22) were untreated or treated 3 different doses of imatinib (10 uM, 100 uM, and 250 uM). 1 uL of a 1:5 library dilution was loaded onto a HS DNA chip.

[0020] Figure 8 shows the results from the multiomics experiment from Example 1 where 3 cell lines (Jurkat, K-562, and KCL-22) were untreated or treated 3 different doses of imatinib (10 uM, 100 uM, and 250 uM). The DNA variants are used to make a heat map where the Y-axis are the cell barcodes are shown in Fig. 9. This heat map separates into 3 clusters (green, red, and black, which alternately can be represented as grey shades) which corresponds to each of the cell lines. Figure 8 also shows the clustering with t-SNE and umap where the colors are the same as in the heat map.

[0021] Figure 9 depicts the results from the multiomics experiment from Example 1 where 3 cell lines (Jurkat, K-562, and KCL-22) were untreated or treated 3 different doses of imatinib (10 uM, 100 uM, and 250 uM). These cells were stained then input onto the Tapestri for the multiomics workflow. DNA libraries, RNA libraries, and protein libraries from single cells were produced where the libraries produced from the same cell shared a cell barcode. By looking at the mutations found in the DNA libraries, the 3 cell lines were identified. Each cell barcode also has corresponding RNA reads and protein reads. Figure 9 also shows those same cells clustered with umap based on SNV, CNV, protein, and RNA where the color of each datapoint corresponds to the cell identified by the mutations found in DNA. The doses of imatinib were differentiated by the tags on the cell hashing antibodies resulting in more clusters for the protein expression.

[0022] Figure 10 depicts the results from the multiomics experiment from Example II where PBMCs were untreated or treated with PHA. A spike-in of the Jurkat cell line was also included. DNA libraries, RNA libraries, and protein libraries from single cells were produced where the libraries produced from the same cell shared a cell barcode. By looking at the mutations found in the DNA libraries, Jurkat cells, cell doublets, and merged droplets were identified. Each cell barcode also has corresponding RNA reads and protein reads (top). Figure 11 shows the fraction of PBMCs that had low RNA expression and high RNA expression. The cells were identified as being treated by PHA or untreated by the protein tags used for cell hashing.

DETAILED DESCRIPTION

[0023] Various aspects of the invention will now be described with reference to the following section which will be understood to be provided by way of illustration only and not to constitute a limitation on the scope of the invention.

[0024] "Complementarity" refers to the ability of a nucleic acid to form hydrogen bond(s) or hybridize with another nucleic acid sequence by either traditional Watson-Crick or other non-traditional types. As used herein "hybridization," refers to the binding, duplexing, or hybridizing of a molecule only

to a particular nucleotide sequence under low, medium, or highly stringent conditions, including when that sequence is present in a complex mixture (e.g., total cellular) DNA or RNA. See e.g. Ausubel, et al., Current Protocols In Molecular Biology, John Wiley & Sons, New York, N.Y., 1993. If a nucleotide at a certain position of a polynucleotide is capable of forming a Watson-Crick pairing with a nucleotide at the same position in an anti-parallel DNA or RNA strand, then the polynucleotide and the DNA or RNA molecule are complementary to each other at that position. The polynucleotide and the DNA or RNA molecule are "substantially complementary" to each other when a sufficient number of corresponding positions in each molecule are occupied by nucleotides that can hybridize or anneal with each other in order to affect the desired process. A complementary sequence is a sequence capable of annealing under stringent conditions to provide a 3'-terminal serving as the origin of synthesis of complementary chain.

[0025] "Identity," as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as determined by the match between strings of such sequences. "Identity" and "similarity" can be readily calculated by known methods, including, but not limited to, those described in Computational Molecular Biology, Lesk, A. M., ed., Oxford University Press, New York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D. W., ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part I, Griffin, A. M., and Griffin, H. G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., Siam J. Applied Math., 48:1073 (1988). In addition, values for percentage identity can be obtained from amino acid and nucleotide sequence alignments generated using the default settings for the AlignX component of Vector NTI Suite 8.0 (Informax, Frederick, Md.). Preferred methods to determine identity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, the GCG program package (Devereux, J., et al., Nucleic Acids Research 12(1): 387 (1984)), BLASTP, BLASTN, and FASTA (Atschul, S. F. et al., J. Molec. Biol. 215:403-410 (1990)). The BLAST X program is publicly available from NCBI and other sources (BLAST Manual, Altschul, S., et al., NCBINLM NIH Bethesda, Md. 20894: Altschul, S., et al., J. Mol. Biol. 215:403-410 (1990). The well-known Smith Waterman algorithm may also be used to determine identity.

[0026] The terms "amplify", "amplifying", "amplification reaction" and their variants, refer generally to any action or process whereby at least a portion of a nucleic acid molecule (referred to as a template nucleic acid molecule) is replicated or copied into at least one additional nucleic acid molecule. The additional nucleic acid molecule optionally includes sequence that is substantially identical or substantially complementary to at least some portion of the template nucleic acid molecule. The template nucleic acid molecule can be single-stranded or double-stranded and the additional nucleic acid molecule can independently be single-stranded or double-stranded. In some embodiments, amplification includes a

template-dependent in vitro enzyme-catalyzed reaction for the production of at least one copy of at least some portion of the nucleic acid molecule or the production of at least one copy of a nucleic acid sequence that is complementary to at least some portion of the nucleic acid molecule. Amplification optionally includes linear or exponential replication of a nucleic acid molecule. In some embodiments, such amplification is performed using isothermal conditions; in other embodiments, such amplification can include thermocycling. In some embodiments, the amplification is a multiplex amplification that includes the simultaneous amplification of a plurality of target sequences in a single amplification reaction. At least some of the target sequences can be situated, on the same nucleic acid molecule or on different target nucleic acid molecules included in the single amplification reaction. In some embodiments, "amplification" includes amplification of at least some portion of DNA- and RNA-based nucleic acids alone, or in combination. The amplification reaction can include single or double-stranded nucleic acid substrates and can further including any of the amplification processes known to one of ordinary skill in the art. In some embodiments, the amplification reaction includes polymerase chain reaction (PCR). In the present invention, the terms "synthesis" and "amplification" of nucleic acid are used. The synthesis of nucleic acid in the present invention means the elongation or extension of nucleic acid from an oligonucleotide serving as the origin of synthesis. If not only this synthesis but also the formation of other nucleic acid and the elongation or extension reaction of this formed nucleic acid occur continuously, a series of these reactions is comprehensively called amplification. The polynucleic acid produced by the amplification technology employed is generically referred to as an "amplicon" or "amplification product."

[0027] A number of nucleic acid polymerases can be used in the amplification reactions utilized in certain embodiments provided herein, including any enzyme that can catalyze the polymerization of nucleotides (including analogs thereof) into a nucleic acid strand. Such nucleotide polymerization can occur in a template-dependent fashion. Such polymerases can include without limitation naturally occurring polymerases and any subunits and truncations thereof, mutant polymerases, variant polymerases, recombinant, fusion or otherwise engineered polymerases, chemically modified polymerases, synthetic molecules or assemblies, and any analogs, derivatives or fragments thereof that retain the ability to catalyze such polymerization. Optionally, the polymerase can be a mutant polymerase comprising one or more mutations involving the replacement of one or more amino acids with other amino acids, the insertion or deletion of one or more amino acids from the polymerase, or the linkage of parts of two or more polymerases. Typically, the polymerase comprises one or more active sites at which nucleotide binding and/or catalysis of nucleotide polymerization can occur. Some exemplary polymerases include without limitation DNA polymerases and RNA polymerases. The term "polymerase" and its variants, as used herein, also includes fusion proteins comprising at least two portions linked to each other, where the first portion comprises a peptide that can catalyze the polymerization of nucleotides into a nucleic acid strand and is linked to a second portion that comprises a second polypeptide. In some embodiments, the second polypeptide can include a reporter enzyme or a processivity-enhancing domain. Optionally, the polymerase can possess 5' exonuclease activity or terminal transferase activity. In some embodiments, the polymerase can be optionally reactivated, for example through the use of heat, chemicals or re-addition of new amounts

of polymerase into a reaction mixture. In some embodiments, the polymerase can include a hot-start polymerase or an aptamer-based polymerase that optionally can be reactivated.

[0028] The terms "target primer" or "target-specific primer" and variations thereof refer to primers that are complementary to a binding site sequence. Target primers are generally a single stranded or double-stranded polynucleotide, typically an oligonucleotide, that includes at least one sequence that is at least partially complementary to a target nucleic acid sequence.

[0029] "Forward primer binding site" and "reverse primer binding site" refers to the regions on the template DNA and/or the amplicon to which the forward and reverse primers bind. The primers act to delimit the region of the original template polynucleotide which is exponentially amplified during amplification. In some embodiments, additional primers may bind to the region 5' of the forward primer and/or reverse primers. Where such additional primers are used, the forward primer binding site and/or the reverse primer binding site may encompass the binding regions of these additional primers as well as the binding regions of the primers themselves. For example, in some embodiments, the method may use one or more additional primers which bind to a region that lies 5' of the forward and/or reverse primer binding region. Such a method was disclosed, for example, in WO0028082 which discloses the use of "displacement primers" or "outer primers".

[0030] A 'barcode' nucleic acid identification sequence can be incorporated into a nucleic acid primer or linked to a primer to enable independent sequencing and identification to be associated with one another via a barcode which relates information and identification that originated from molecules that existed within the same sample. There are numerous techniques that can be used to attach barcodes to the nucleic acids within a discrete entity. For example, the target nucleic acids may or may not be first amplified and fragmented into shorter pieces. The molecules can be combined with discrete entities, e.g., droplets, containing the barcodes. The barcodes can then be attached to the molecules using, for example, splicing by overlap extension. In this approach, the initial target molecules can have "adaptor" sequences added, which are molecules of a known sequence to which primers can be synthesized. When combined with the barcodes, primers can be used that are complementary to the adaptor sequences and the barcode sequences, such that the product amplicons of both target nucleic acids and barcodes can anneal to one another and, via an extension reaction such as DNA polymerization, be extended onto one another, generating a double-stranded product including the target nucleic acids attached to the barcode sequence. Alternatively, the primers that amplify that target can themselves be barcoded so that, upon annealing and extending onto the target, the amplicon produced has the barcode sequence incorporated into it. This can be applied with a number of amplification strategies, including specific amplification with PCR or non-specific amplification with, for example, MDA. An alternative enzymatic reaction that can be used to attach barcodes to nucleic acids is ligation, including blunt or sticky end ligation. In this approach, the DNA barcodes are incubated with the nucleic acid targets and ligase enzyme, resulting in the ligation of the barcode to the targets. The ends of the nucleic acids can be modified as needed for ligation by a number of techniques, including by using adaptors introduced with ligase or fragments to enable greater control over the number of barcodes added to the end of the molecule.

[0031] A barcode sequence can additionally be incorporated into microfluidic beads to decorate the bead with identical sequence tags. Such tagged beads can be inserted into microfluidic droplets and via droplet PCR amplification, tag each target amplicon with the unique bead barcode. Such barcodes can be used to identify specific droplets upon a population of amplicons originated from. This scheme can be utilized when combining a microfluidic droplet containing single individual cell with another microfluidic droplet containing a tagged bead. Upon collection and combination of many microfluidic droplets, amplicon sequencing results allow for assignment of each product to unique microfluidic droplets. In a typical implementation, we use barcodes on the Mission Bio Tapestri™ beads to tag and then later identify each droplet's amplicon content. The use of barcodes is described in US Patent Application Serial No. 15/940,850 filed March 29, 2018 by Abate, A. et al., entitled 'Sequencing of Nucleic Acids via Barcoding in Discrete Entities', incorporated by reference herein.

[0032] In some embodiments, it may be advantageous to introduce barcodes into discrete entities, e.g., microdroplets, on the surface of a bead, such as a solid polymer bead or a hydrogel bead. These beads can be synthesized using a variety of techniques. For example, using a mix-split technique, beads with many copies of the same, random barcode sequence can be synthesized. This can be accomplished by, for example, creating a plurality of beads including sites on which DNA can be synthesized. The beads can be divided into four collections and each mixed with a buffer that will add a base to it, such as an A, T, G, or C. By dividing the population into four subpopulations, each subpopulation can have one of the bases added to its surface. This reaction can be accomplished in such a way that only a single base is added and no further bases are added. The beads from all four subpopulations can be combined and mixed together, and divided into four populations a second time. In this division step, the beads from the previous four populations may be mixed together randomly. They can then be added to the four different solutions, adding another, random base on the surface of each bead. This process can be repeated to generate sequences on the surface of the bead of a length approximately equal to the number of times that the population is split and mixed. If this was done 10 times, for example, the result would be a population of beads in which each bead has many copies of the same random 10-base sequence synthesized on its surface. The sequence on each bead would be determined by the particular sequence of reactors it ended up in through each mix-spit cycle.

[0033] A barcode may further comprise a 'unique identification sequence' (UMI). A UMI is a nucleic acid having a sequence which can be used to identify and/or distinguish one or more first molecules to which the UMI is conjugated from one or more second molecules. UMIs are typically short, e.g., about 5 to 20 bases in length, and may be conjugated to one or more target molecules of interest or amplification products thereof. UMIs may be single or double stranded. In some embodiments, both a nucleic acid barcode sequence and a UMI are incorporated into a nucleic acid target molecule or an amplification product thereof. Generally, a UMI is used to distinguish between molecules of a similar type within a population or group, whereas a nucleic acid barcode sequence is used to distinguish between populations or groups of molecules. In some embodiments, where both a UMI and a nucleic acid barcode sequence are utilized, the UMI is shorter in sequence length than the nucleic acid barcode sequence.

[0034] The terms "identity" and "identical" and their variants, as used herein, when used in reference to two or more nucleic acid sequences, refer to similarity in sequence of the two or more sequences (e.g., nucleotide or polypeptide sequences). In the context of two or more homologous sequences, the percent identity or homology of the sequences or subsequences thereof indicates the percentage of all monomeric units (e.g., nucleotides or amino acids) that are the same (i.e., about 70% identity, preferably 75%, 80%, 85%, 90%, 95%, 97%, 98% or 99% identity). The percent identity can be over a specified region, when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using a BLAST or BLAST 2.0 sequence comparison algorithms with default parameters described below, or by manual alignment and visual inspection. Sequences are said to be "substantially identical" when there is at least 85% identity at the amino acid level or at the nucleotide level. Preferably, the identity exists over a region that is at least about 25, 50, or 100 residues in length, or across the entire length of at least one compared sequence. A typical algorithm for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul et al, Nuc. Acids Res. 25:3389-3402 (1977). Other methods include the algorithms of Smith & Waterman, Adv. Appl. Math. 2:482 (1981), and Needleman & Wunsch, J. Mol. Biol. 48:443 (1970), etc. Another indication that two nucleic acid sequences are substantially identical is that the two molecules or their complements hybridize to each other under stringent hybridization conditions.

[0035] The terms "nucleic acid," "polynucleotides," and "oligonucleotides" refers to biopolymers of nucleotides and, unless the context indicates otherwise, includes modified and unmodified nucleotides, and both DNA and RNA, and modified nucleic acid backbones. For example, in certain embodiments, the nucleic acid is a peptide nucleic acid (PNA) or a locked nucleic acid (LNA). Typically, the methods as described herein are performed using DNA as the nucleic acid template for amplification. However, nucleic acid whose nucleotide is replaced by an artificial derivative or modified nucleic acid from natural DNA or RNA is also included in the nucleic acid of the present invention insofar as it functions as a template for synthesis of complementary chain. The nucleic acid of the present invention is generally contained in a biological sample. The biological sample includes animal, plant or microbial tissues, cells, cultures and excretions, or extracts therefrom. In certain aspects, the biological sample includes intracellular parasitic genomic DNA or RNA such as virus or mycoplasma. The nucleic acid may be derived from nucleic acid contained in said biological sample. For example, genomic DNA, or cDNA synthesized from mRNA, or nucleic acid amplified on the basis of nucleic acid derived from the biological sample, are preferably used in the described methods. Unless denoted otherwise, whenever a oligonucleotide sequence is represented, it will be understood that the nucleotides are in 5' to 3' order from left to right and that "A" denotes deoxyadenosine, "C" denotes deoxycytidine, "G" denotes deoxyguanosine, "T" denotes thymidine, and "U' denotes deoxyuridine. Oligonucleotides are said to have "5' ends" and "3' ends" because mononucleotides are typically reacted to form oligonucleotides via attachment of the 5' phosphate or equivalent group of one nucleotide to the 3' hydroxyl or equivalent group of its neighboring nucleotide, optionally via a phosphodiester or other suitable linkage.

[0036] A template nucleic acid is a nucleic acid serving as a template for synthesizing a complementary chain in a nucleic acid amplification technique. A complementary chain having a nucleotide sequence complementary to the template has a meaning as a chain corresponding to the template, but the relationship between the two is merely relative. That is, according to the methods described herein a chain synthesized as the complementary chain can function again as a template. That is, the complementary chain can become a template. In certain embodiments, the template is derived from a biological sample, e.g., plant, animal, virus, micro-organism, bacteria, fungus, etc. In certain embodiments, the animal is a mammal, e.g., a human patient. A template nucleic acid typically comprises one or more target nucleic acid. A target nucleic acid in exemplary embodiments may comprise any single or double-stranded nucleic acid sequence that can be amplified or synthesized according to the disclosure, including any nucleic acid sequence suspected or expected to be present in a sample.

[0037] Primers and oligonucleotides used in embodiments herein comprise nucleotides. A nucleotide comprises any compound, including without limitation any naturally occurring nucleotide or analog thereof, which can bind selectively to, or can be polymerized by, a polymerase. Typically, but not necessarily, selective binding of the nucleotide to the polymerase is followed by polymerization of the nucleotide into a nucleic acid strand by the polymerase; occasionally however the nucleotide may dissociate from the polymerase without becoming incorporated into the nucleic acid strand, an event referred to herein as a "non-productive" event. Such nucleotides include not only naturally occurring nucleotides but also any analogs, regardless of their structure, that can bind selectively to, or can be polymerized by, a polymerase. While naturally occurring nucleotides typically comprise base, sugar and phosphate moieties, the nucleotides of the present disclosure can include compounds lacking any one, some or all of such moieties. For example, the nucleotide can optionally include a chain of phosphorus atoms comprising three, four, five, six, seven, eight, nine, ten or more phosphorus atoms. In some embodiments, the phosphorus chain can be attached to any carbon of a sugar ring, such as the 5' carbon. The phosphorus chain can be linked to the sugar with an intervening O or S. In one embodiment, one or more phosphorus atoms in the chain can be part of a phosphate group having P and O. In another embodiment, the phosphorus atoms in the chain can be linked together with intervening O, NH, S, methylene, substituted methylene, ethylene, substituted ethylene, CNH_2, C(O), $C(CH_2)$, CH_2CH_2, or $C(OH)CH_2R$ (where R can be a 4-pyridine or 1-imidazole). In one embodiment, the phosphorus atoms in the chain can have side groups having O, BH3, or S. In the phosphorus chain, a phosphorus atom with a side group other than O can be a substituted phosphate group. In the phosphorus chain, phosphorus atoms with an intervening atom other than O can be a substituted phosphate group. Some examples of nucleotide analogs are described in Xu, U.S. Pat. No. 7,405,281.

[0038] In some embodiments, the nucleotide comprises a label and referred to herein as a "labeled nucleotide"; the label of the labeled nucleotide is referred to herein as a "nucleotide label". In some embodiments, the label can be in the form of a fluorescent moiety (e.g. dye), luminescent moiety, or the like attached to the terminal phosphate group, i.e., the phosphate group most distal from the sugar. Some examples of nucleotides that can be used in the disclosed methods and compositions include, but are not limited to, ribonucleotides, deoxyribonucleotides, modified ribonucleotides, modified

deoxyribonucleotides, ribonucleotide polyphosphates, deoxyribonucleotide polyphosphates, modified ribonucleotide polyphosphates, modified deoxyribonucleotide polyphosphates, peptide nucleotides, modified peptide nucleotides, metallonucleosides, phosphonate nucleosides, and modified phosphate-sugar backbone nucleotides, analogs, derivatives, or variants of the foregoing compounds, and the like. In some embodiments, the nucleotide can comprise non-oxygen moieties such as, for example, thio- or borano-moieties, in place of the oxygen moiety bridging the alpha phosphate and the sugar of the nucleotide, or the alpha and beta phosphates of the nucleotide, or the beta and gamma phosphates of the nucleotide, or between any other two phosphates of the nucleotide, or any combination thereof. "Nucleotide 5'-triphosphate" refers to a nucleotide with a triphosphate ester group at the 5' position, and are sometimes denoted as "NTP", or "dNTP" and "ddNTP" to particularly point out the structural features of the ribose sugar. The triphosphate ester group can include sulfur substitutions for the various oxygens, e.g. α-thio-nucleotide 5'-triphosphates. For a review of nucleic acid chemistry, see: Shabarova, Z. and Bogdanov, A. Advanced Organic Chemistry of Nucleic Acids, VCH, New York, 1994.

[0039] Any nucleic acid amplification method may be utilized, such as a PCR-based assay, e.g., quantitative PCR (qPCR), or an isothermal amplification may be used to detect the presence of certain nucleic acids, e.g., genes, of interest, present in discrete entities or one or more components thereof, e.g., cells encapsulated therein. Such assays can be applied to discrete entities within a microfluidic device or a portion thereof or any other suitable location. The conditions of such amplification or PCR-based assays may include detecting nucleic acid amplification over time and may vary in one or more ways.

[0040] The number of amplification/PCR primers that may be added to a microdroplet may vary. The number of amplification or PCR primers that may be added to a microdroplet may range from about 1 to about 500 or more, e.g., about 2 to 100 primers, about 2 to 10 primers, about 10 to 20 primers, about 20 to 30 primers, about 30 to 40 primers, about 40 to 50 primers, about 50 to 60 primers, about 60 to 70 primers, about 70 to 80 primers, about 80 to 90 primers, about 90 to 100 primers, about 100 to 150 primers, about 150 to 200 primers, about 200 to 250 primers, about 250 to 300 primers, about 300 to 350 primers, about 350 to 400 primers, about 400 to 450 primers, about 450 to 500 primers, or about 500 primers or more.

[0041] One or both primer of a primer set may also be attached or conjugated to an affinity reagent that may comprise anything that binds to a target molecule or moiety. Nonlimiting examples of affinity reagent include ligands, receptors, antibodies and binding fragments thereof, peptide, nucleic acid, and fusions of the preceding and other small molecule that specifically binds to a larger target molecule in order to identify, track, capture, or influence its activity. Affinity reagents may also be attached to solid supports, beads, discrete entities, or the like, and are still referenced as affinity reagents herein.

[0042] One or both primers of a primer set may comprise a barcode sequence described herein. In some embodiments, individual cells, for example, are isolated in discrete entities, e.g., droplets. These cells may be lysed and their nucleic acids barcoded. This process can be performed on a large number of single cells in discrete entities with unique barcode sequences enabling subsequent deconvolution of mixed sequence reads by barcode to obtain single cell information. This approach provides a way to group together

nucleic acids originating from large numbers of single cells. Additionally, affinity reagents such as antibodies can be conjugated with nucleic acid labels, e.g., oligonucleotides including barcodes, which can be used to identify antibody type, e.g., the target specificity of an antibody. These reagents can then be used to bind to the proteins within or on cells, thereby associating the nucleic acids carried by the affinity reagents to the cells to which they are bound. These cells can then be processed through a barcoding workflow as described herein to attach barcodes to the nucleic acid labels on the affinity reagents. Techniques of library preparation, sequencing, and bioinformatics may then be used to group the sequences according to cell/discrete entity barcodes. Any suitable affinity reagent that can bind to or recognize a biological sample or portion or component thereof, such as a protein, a molecule, or complexes thereof, may be utilized in connection with these methods. The affinity reagents may be labeled with nucleic acid sequences that relates their identity, e.g., the target specificity of the antibodies, permitting their detection and quantitation using the barcoding and sequencing methods described herein. Exemplary affinity reagents can include, for example, antibodies, antibody fragments, Fabs, scFvs, peptides, drugs, etc. or combinations thereof. The affinity reagents, e.g., antibodies, can be expressed by one or more organisms or provided using a biological synthesis technique, such as phage, mRNA, or ribosome display. The affinity reagents may also be generated via chemical or biochemical means, such as by chemical linkage using N-Hydroxysuccinimide (NETS), click chemistry, or streptavidin-biotin interaction, for example. The oligo-affinity reagent conjugates can also be generated by attaching oligos to affinity reagents and hybridizing, ligating, and/or extending via polymerase, etc., additional oligos to the previously conjugated oligos. An advantage of affinity reagent labeling with nucleic acids is that it permits highly multiplexed analysis of biological samples. For example, large mixtures of antibodies or binding reagents recognizing a variety of targets in a sample can be mixed together, each labeled with its own nucleic acid sequence. This cocktail can then be reacted to the sample and subjected to a barcoding workflow as described herein to recover information about which reagents bound, their quantity, and how this varies among the different entities in the sample, such as among single cells. The above approach can be applied to a variety of molecular targets, including samples including one or more of cells, peptides, proteins, macromolecules, macromolecular complexes, etc. The sample can be subjected to conventional processing for analysis, such as fixation and permeabilization, aiding binding of the affinity reagents. To obtain highly accurate quantitation, the unique molecular identifier (UMI) techniques described herein can also be used so that affinity reagent molecules are counted accurately. This can be accomplished in a number of ways, including by synthesizing UMIs onto the labels attached to each affinity reagent before, during, or after conjugation, or by attaching the UMIs microfluidically when the reagents are used. Similar methods of generating the barcodes, for example, using combinatorial barcode techniques as applied to single cell sequencing and described herein, are applicable to the affinity reagent technique. These techniques enable the analysis of proteins and/or epitopes in a variety of biological samples to perform, for example, mapping of epitopes or post translational modifications in proteins and other entities or performing single cell proteomics. For example, using the methods described herein, it is possible to generate a library of labeled affinity reagents that detect an epitope in all proteins in the proteome of an organism, label those epitopes with the reagents, and apply

the barcoding and sequencing techniques described herein to detect and accurately quantitate the labels associated with these epitopes.

[0043] Primers may contain primers for one or more nucleic acid of interest, e.g. one or more genes of interest. The number of primers for genes of interest that are added may be from about one to 500, e.g., about 1 to 10 primers, about 10 to 20 primers, about 20 to 30 primers, about 30 to 40 primers, about 40 to 50 primers, about 50 to 60 primers, about 60 to 70 primers, about 70 to 80 primers, about 80 to 90 primers, about 90 to 100 primers, about 100 to 150 primers, about 150 to 200 primers, about 200 to 250 primers, about 250 to 300 primers, about 300 to 350 primers, about 350 to 400 primers, about 400 to 450 primers, about 450 to 500 primers, or about 500 primers or more. Primers and/or reagents may be added to a discrete entity, e.g., a microdroplet, in one step, or in more than one step. For instance, the primers may be added in two or more steps, three or more steps, four or more steps, or five or more steps. Regardless of whether the primers are added in one step or in more than one step, they may be added after the addition of a lysing agent, prior to the addition of a lysing agent, or concomitantly with the addition of a lysing agent. When added before or after the addition of a lysing agent, the PCR primers may be added in a separate step from the addition of a lysing agent. In some embodiments, the discrete entity, e.g., a microdroplet, may be subjected to a dilution step and/or enzyme inactivation step prior to the addition of the PCR reagents. Exemplary embodiments of such methods are described in PCT Publication No. WO 2014/028378, the disclosure of which is incorporated by reference herein in its entirety and for all purposes.

[0044] A primer set for the amplification of a target nucleic acid typically includes a forward primer and a reverse primer that are complementary to a target nucleic acid or the complement thereof. In some embodiments, amplification can be performed using multiple target-specific primer pairs in a single amplification reaction, wherein each primer pair includes a forward target-specific primer and a reverse target-specific primer, where each includes at least one sequence that substantially complementary or substantially identical to a corresponding target sequence in the sample, and each primer pair having a different corresponding target sequence. Accordingly, certain methods herein are used to detect or identify multiple target sequences from a single cell sample.

[0045] In some implementations, solid supports, beads, and the like are coated with affinity reagents. Affinity reagents include, without limitation, antigens, antibodies or aptamers with specific binding affinity for a target molecule. The affinity reagents bind to one or more targets within the single cell entities. Affinity reagents are often detectably labeled (e.g., with a fluorophore). Affinity reagents are sometimes labeled with unique barcodes, oligonucleotide sequences, or UMI's.

[0046] In some implementations, a RT/PCR polymerase reaction and amplification reaction are performed, for example in the same reaction mixture, as an addition to the reaction mixture, or added to a portion of the reaction mixture.

[0047] In one particular implementation, a solid support contains a plurality of affinity reagents, each specific for a different target molecule but containing a common sequence to be used to identify the unique solid support. Affinity reagents that bind a specific target molecule are collectively labeled with the same oligonucleotide sequence such that affinity molecules with different binding affinities for different

targets are labeled with different oligonucleotide sequences. In this way, target molecules within a single target entity are differentially labeled in these implements to determine which target entity they are from but contain a common sequence to identify them from the same solid support.

Targeted detection and sequencing of DNA, RNA, and Protein

[0048] A first objective of some implementations is to provide methods for the simultaneous targeted detection and sequencing of DNA, RNA, and Protein. In preferred embodiments, the DNA, RNA, and Protein is detected, characterized, and sequenced using single cells. Single-cell analysis is conducted herein by sequencing either genomic DNA targets, RNA/cDNA transcripts or protein detection. Genomic DNA can be assessed by first amplifying whole genomic DNA or targeted approaches.

[0049] One such example is amplifying whole genomic DNA or targeted portions of the DNA in a Mission Bio's Tapestri® platform. In some embodiments using this approach, RNA/cDNA transcripts are often accessed by first priming with 3' mRNA end oligo dT extension strategies or random primers. Targeted cDNA extension can be conducted using transcript specific primers and is amenable to these workflows. Surface protein markers are readily detected using dye-labelled antibodies and a fluorescence-activated cell sorting platform (FACS). In preferred embodiments, DNA barcoded antibodies can be employed in cell staining and readout by next generation sequencing (NGS). We have combined three different methodologies to simultaneously detect, by targeted sequencing, DNA, RNA, and surface protein markers as our so-called 'triomic' methodology. Our novel approach represents a great advancement in single-cell analysis.

[0050] Typically, methods are optimized to efficiently detect and amplify genomic DNA amplicon targets. Additionally, some implementations use reverse transcriptase enzyme with RNA-specific primers to extend RNA targets and generate cDNA templates for efficient RNA detection. Additionally, an antibody-oligonucleotide tagging system for surface protein detection on single-cells using microfluidic droplets has been developed. In certain embodiments provided herein, a Mission Bio Tapestri® workflow scheme has been developed that applies central concepts from all three 'omics' to thereby provide for a triomic interrogation process for single-cells. While the disclosed embodiments relate to using the Tapestri® workflow, it should be noted that the disclosed principles are not limited thereto and may be applied to other instrumentations and/or workflow.

[0051] Certain methods provided herein utilize specific antibodies to detect epitopes of interest. In some embodiments, antibodies are labeled with sequence tags that can be read out with microfluidic barcoding and DNA sequencing. This and related implementations are used herein to characterize cell surface proteins of different cell types at the single-cell level.

[0052] In some embodiments, a barcode identity is encoded by its full nucleobase sequence and thus confers a combinatorial tag space far exceeding what is possible with conventional approaches using fluorescence. A modest tag length of ten bases provides over a million unique sequences, sufficient to label an antibody against every epitope in the human proteome. Indeed, with this approach, the limit to

multiplexing is not the availability of unique tag sequences but, rather, that of specific antibodies that can detect the epitopes of interest in a multiplexed reaction.

[0053] In some implementations, cells are bound with antibodies against the different target epitopes, as in conventional immunostaining, except that the antibodies are labeled with barcodes.

[0054] In practice, when an antibody binds its target the DNA barcode tag is carried with it and thus allows the presence of the target to be inferred based on the presence of the barcode. In some implementations, counting barcode tags provides an estimate of the different epitopes present in the cell.

[0055] Other embodiments implementations are used to distinguish particular cells by their protein expression profiles. Some embodiments of DNA-tagged antibodies provided herein have multiple advantages for profiling proteins in single cells.

[0056] A primary advantage of these implementations is the ability to amplify low-abundance tags to make them detectable with sequencing. Another advantage in some implementations is the capability of using molecular indices for quantitative results. Some implementations also have essentially limitless multiplexing capabilities.

[0057] Some embodiments utilize solid beads having an alternate chemistry where the primers to be used are in solution and contain a PCR annealing sequence embedded, or 'handle', that allows hybridization to primers. In some implementations, the handle is a specific tail 5' upstream of the target sequence and this handle is complimentary to bead barcoded oligo and serves as a PCR extension bridge to link the target amplicon to the bead barcode library primer sequence. The solid beads may contain primers that can anneal to the PCR handle on the primers.

[0058] Some embodiments are used to detect and characterize cell surface proteins, DNA, and RNA. Such a workflow can begin with an antibody-oligonucleotide staining and washing of a single-cell suspension. The stained cells are loaded onto a Mission, Bio Tapestri™ system, cells are lysed, RNA is converted to cDNA by reverse transcriptase, PCR cycling is used to amplify antibody-oligonucleotides, targeted cDNA species, and targeted genomic DNA regions. All three omic libraries are purified and quantified, and sequencing is conducted to determine the identity of an analyte. This workflow is only exemplary, and it is understood that certain steps can be removed and other steps can be added.

[0059] Other aspects of the invention may be described in the follow embodiments:

1. An apparatus or system for performing a method described herein.
2. A composition or reaction mixture for performing a method described herein.
3. An antibody library generated by methods described herein.
4. A genomic library generated by methods described herein.
5. A transcriptome library generated according to a method described herein.

6. An antibody library, genomic, and transcriptome library generated according to a method described herein.

7. A kit for performing a method described herein.

8. A cell population selected by the methods described herein.

9. A method of determining and characterizing the protein expression pattern of a single cell, the method comprising the steps of:

a) conjugating barcode sequences flanked by PCR priming sites onto antibodies, wherein a barcode sequence is specific to an antibody;

b) performing a cell identification step using the barcode conjugated antibodies;

c) partitioning or separating individual cells and encapsulating one or more individual cell(s) in a reaction mixture comprising a protease;

d) incubating the encapsulated cell with the protease in the drop to produce a cell lysate;

e) performing a reverse transcriptase reaction, wherein a reverse transcriptase is in the reaction mixture or added to the reaction mixture

f)providing one or more nucleic acid amplification primer sets targeting nucleic acids present in a cell, wherein one or more primer of a primer set includes a barcode identification sequence associated with an antibody;

g) providing one or more nucleic acid amplification primer sets targeting nucleic acids present in a cell, wherein one or more primer of a primer set includes a barcode identification sequence unique to each cell;

h) performing a nucleic acid amplification reaction to produce one or more amplicons;

i)providing an affinity reagent that comprises a nucleic acid sequence complementary to the identification barcode sequence of one of more nucleic acid primer of a primer set, wherein said affinity reagent comprising said nucleic acid sequence complementary to the identification barcode sequence is capable of binding to a nucleic acid amplification primer set comprising a barcode identification sequence;

j) contacting an affinity reagent to the amplification product comprising amplicons of one or more target nucleic acid sequence under conditions sufficient for binding of the affinity reagent to the target nucleic acid to form an affinity reagent bound target nucleic acid; and

k) determining the identity and characterizing one or more protein by sequencing a barcode of an amplicon.

10. A method for adding a barcode identification sequence linked to an antibody, the method comprising the steps:

i) performing a barcoding PCR reaction of a target gDNA using a) a primer containing a cell barcode sequence and a PCR handle; b) a primer containing sequence complementary to the target genomic DNA and a PCR handle that is complementary to the primer containing the cell barcode and c) a reverse primer comprising a sequence complementary to the target genomic DNA, an antibody tag sequence, a

second PCR handle, and could include a unique molecular tag , to produce an amplicon comprising a cell barcode, a target DNA sequence, an antibody tag with a PCR handle on both the 5' end and 3' end; and

ii) performing a library creation PCR reaction using a first primers comprising sequencing adapters , sample indexes, and sequences complementary to the two PCR handles produced on the amplicon to produce library comprising sequencing adapters, dual or single sample indexes, a cell barcode, a target DNA sequence, an antibody tag, and could include a unique molecular tag.

11. A method for adding a barcode identification sequence linked to an antibody, the method comprising the steps:

i) performing a barcoding PCR reaction of a target gDNA using a) a primer containing a cell barcode sequence and a PCR handle; b) a primer containing sequence complementary to the target genomic DNA and a PCR handle that is complementary to the primer containing the cell barcode and c) a reverse primer comprising a sequence complementary to the target genomic DNA, an antibody tag sequence, a second PCR handle, and could include a unique molecular tag , to produce an amplicon comprising a cell barcode, a target DNA sequence, an antibody tag with a PCR handle on both the 5' end and 3' end, a first read sequence a first cell barcode, a constant region 1, a second cell bar code, a constant region 2, the forward primer sequence, an insert sequence of length 'n', a reverse primer comprising a sequence complementary to the target genomic DNA, a unique molecular identifier, an antibody tag sequence, to a second unique molecular identifier; a second read sequence; and

ii) performing a library creation PCR reaction using a first primers comprising sequencing adapters , sample indexes, and sequences complementary to the two PCR handles produced on the amplicon comprising a P5 sequence and a second read sequence and a second primer comprising a second read sequence, and index sequence, and a P7 sequence to produce library comprising sequencing adapters, dual or single sample indexes, a cell barcode, a target DNA sequence, an antibody tag, and could include a unique molecular tag.

[0060] The following Examples are included for illustration and not limitation.

Example I

[0061] In this Example, we provide an embodiment for a single workflow for the simultaneous detection of DNA, RNA, and protein as described in Figures 1. Fresh Jurkat cells, K-562 cells, and KCL-22 cells were treated with imatinib at doses of 10 uM, 100 uM, and 250 uM or left untreated. The cells from each dose were then mixed in equal ratios and stained with the antibodies listed in Table 1 as well as two antibodies for cell hashing, B2M and CD298.

CD14	CD117	CD45RA
CD33	CD123	Mouse IgG

CD44	HLA-DR	Annexin V
CD38	CD90	
CD4	CD34	
CD3	CD7	
EPCAM	CD45	

Table 1: The 17 protein targets used in the multi-omics experiment from Example I.

[0062] After staining, ~100,000 cells were loaded onto a Tapestri™ instrument for each encapsulation. The encapsulation partitioned single cells along with reagents for cell lysis, reverse transcription, and a protease treatment. The reverse transcription targets are listed in Table 2. Gene specific priming was used for reverse transcription with a reverse transcriptase, such as SuperScript® IV First-Strand Synthesis System. After protease treatment, the encapsulation droplets were then reloaded onto the Tapestri™ instrument for barcoding PCR. The gDNA from each cell were targeted by the 88 amplicons listed in Table 3. Barcoding PCR was performed using the reagents from the Tapestri® Single-Cell DNA Sequencing V2 kit.

[0063] After barcoding PCR, sequencing libraries were made separating out the protein libraries, RNA libraries, and DNA libraries as described in Figure 4. The bioanalyzer traces from the untreated cells are shown in Figures 5 (DNA), 6 (protein), and 7 (RNA). These libraries were pooled and sequenced. Each sequencing read with the same cell barcode sequence was identified as being from the same cell. Reads from DNA, RNA, and protein were then combined for each cell as seen in Figure 9. The cell lines were identified using their SNVs as seen in Figure 8.

CCND3	CREB5	HIF1A	NFKB1	SIRT1	SRF
CD44	CREB1	HSPB1	MYC	NCL	TP53
CCND1	ELK1	IKBKG	PIK3CB	RHOA	CASP9
CD33	FOS	IRF9	PIM1	MCM4	CASP3
CDK6	FHL1	BCL2	PIAS1	NASP	CASP8
CDK4	FASLG	BCL2L11	PRKCB	SOS1	UBB
CDKN1B	GNG12	MAP2K1	PTEN	TCL1B	MRPL16

CREB3L4	GSK3B	MAPK1	HSPA1A	SOCS3	MRPL21
CDKN1A	BAD	BCL2L1	HSPA2	SOCS2	FAM32A
CREBBP	FOXO4	MYB	IL2RB	STAT4	ABCB7
CREB3L1	FOXO1	NF1	IL2RA	STAT6	PCBP1

Table 2: The 66 genes targeted by 66 RNA amplicons in the multi-omics experiment from Example I.

EPS15	PIK3CA	WRN	ARHGEF12	BUB1B	METTL23	EP300
NRAS	MAP3K13	JAK2	KRAS	PALB2	SRSF2	SSX1
RPS27A	NSD1	GATA3	COL2A1	FANCA	MFSD11	
AFF3	PTPRK	DKK1	KMT2D	NCOR1	DNM2	
PAX3	CARD11	POLA2	CLIP1	ERBB2	CIC	
CMTM6	EGFR	CCND1	FLT3	KAT2A	BCR	
RHOA	EZH2	ATM	BRCA2	RAB5C	MYH9	

Table 3: The 44 genes targeted by 88 DNA amplicons in the multi-omics experiment from Example I.

Example II

[0064] In this Example, we provide an embodiment for a single workflow for the simultaneous detection of DNA, RNA, and protein with PBMCs. PBMCs were untreated or treated with PHA. The cells were combined with a small percentage of Jurkat cells then stained with tagged antibodies. These stained cells were then encapsulated on the Tapestri™ instrument with reagents for cell lysis, reverse transcription and protease treatment. Following these reactions, the droplets were then merged with PCR reagents on the Tapestri™ instrument. This reaction attached the cell barcodes onto each amplicon for DNA, RNA, and protein. These amplicons were then used to produce separate sequencing libraries using the workflow described in Figure 4. Following sequencing, the cell barcodes were used to link the reads from DNA, RNA, and protein to the same cells as seen in Figure 10. Shown in Figure 11, these reads could then be used for cluster analysis and expression level analysis.

[0065] In an exemplary embodiment, the workflow consists of three discrete processes to access, tag and amplify PCR amplicons for each omic. The general outlay of the exemplary process is summarized below (and reference to Figures 1A and 1B):

[0066] A bulk suspension of cells (typically from a mammalian source such as a cultured cells, blood-derived and bone marrow-derived cells) are incubated with oligonucleotide-tagged antibodies. The oligonucleotide is attached to the antibody by use of click chemistry. The 5'-end nucleotide sequence comprise a "handle" sequence for hybridization to a "forward" PCR primer. The mid-section of the oligonucleotide contains an antibody-specific barcode and random nucleotide sequence. The 3'-end contains a sequence for the "reverse" PCR primer.

[0067] Typically, 0.1-50 nM of oligo-tagged antibody is incubated with 1,000-1,000,000 cells in a volume of 1-1000 uL. After 0.1-20 hours incubation at 4-37C temperature, the cells are wash with 0.1-100 mL of wash buffer. After 2-8 washes, the cells are resuspended in 10-1000 uL for loading onto a Tapestri™ cartridge.

[0068] The single cells are combined into the first droplet merger containing reagents for reverse transcription and a protease lysis buffer. After a brief reverse transcription and protease incubation, the protease is heat inactivated at 70-95C for 2-60 minutes.

[0069] The first droplet is then merged with a second droplet containing the cell barcoded bead, PCR hot start enzyme and complete buffer, multiplex forward and reverse primers for both targeted DNA and RNA,. After a 1-60 minute incubation to prime and extend RNA templates, the droplets are heated to 80-95C 1-30 minutes to activate the thermostable PCR enzyme.

[0070] PCR amplification of antibody, RNA/cDNA and DNA targets is commenced by the standard Mission Bio v2 cycling conditions, with tolerances added.

[0071] After the first round of PCR, the emulsion is eliminated and the PCR product aqueous phase carried forward.

[0072] A biotin-labeled oligonucleotide directed to hybridize to the antibody reverse primer sequence region and thereby selectively isolates the antibody tags away from the RNA and DNA amplicons.

[0073] Both the DNA, RNA and protein amplicon fractions are selectively purified by AMPure bead process (beads and instructions available from Beckman Coulter, Fullerton CA).

[0074] A second round of PCR with Illumina compatible sequencing primers is commenced on the single-cell libraries. Different sample indexes can be used for each analyte.

[0075] The DNA, RNA and Protein libraries are combined together at a range varying from 1:100 to 100:1 ratio, depending on the desired read distribution of DNA to RNA to protein. The mixture is loaded onto an Illumina sequencing platform.

[0076] All patents, publications, scientific articles, web sites, and other documents and materials referenced or mentioned herein are indicative of the levels of skill of those skilled in the art to which the invention pertains, and each such referenced document and material is hereby incorporated by reference to the same extent as if it had been incorporated by reference in its entirety individually or set forth herein in its entirety. Applicants reserve the right to physically incorporate into this specification any and all

materials and information from any such patents, publications, scientific articles, web sites, electronically available information, and other referenced materials or documents.

[0077] The specific methods and compositions described herein are representative of preferred embodiments and are exemplary and not intended as limitations on the scope of the invention. Other objects, aspects, and embodiments will occur to those skilled in the art upon consideration of this specification, and are encompassed within the spirit of the invention as defined by the scope of the claims. It will be readily apparent to one skilled in the art that varying substitutions and modifications may be made to the invention disclosed herein without departing from the scope and spirit of the invention. The invention illustratively described herein suitably may be practiced in the absence of any element or elements, or limitation or limitations, which is not specifically disclosed herein as essential. Thus, for example, in each instance herein, in embodiments or examples of the present invention, any of the terms "comprising", "consisting essentially of", and "consisting of" may be replaced with either of the other two terms in the specification. Also, the terms "comprising", "including", containing", etc. are to be read expansively and without limitation. The methods and processes illustratively described herein suitably may be practiced in differing orders of steps, and that they are not necessarily restricted to the orders of steps indicated herein or in the claims. It is also that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Under no circumstances may the patent be interpreted to be limited to the specific examples or embodiments or methods specifically disclosed herein. Under no circumstances may the patent be interpreted to be limited by any statement made by any Examiner or any other official or employee of the Patent and Trademark Office unless such statement is specifically and without qualification or reservation expressly adopted in a responsive writing by Applicants.

[0078] The terms and expressions that have been employed are used as terms of description and not of limitation, and there is no intent in the use of such terms and expressions to exclude any equivalent of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the invention as claimed. Thus, it will be understood that although the present invention has been specifically disclosed by preferred embodiments and optional features, modification and variation of the concepts herein disclosed may be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of this invention as defined by the appended claims.

[0079] The invention has been described broadly and generically herein. Each of the narrower species and subgeneric groupings falling within the generic disclosure also form part of the invention. This includes the generic description of the invention with a proviso or negative limitation removing any subject matter from the genus, regardless of whether or not the excised material is specifically recited herein.

[0080] Other embodiments are within the following claims. In addition, where features or aspects of the invention are described in terms of Markush groups, those skilled in the art will recognize that the invention is also thereby described in terms of any individual member or subgroup of members of the Markush group.

What is claimed is:

1. A multiomic detection and characterization method for detecting DNA, RNA, or protein from a cell, comprising

encapsulating a cell in a drop comprising a reaction mixture comprising a protease;

performing a protease digest on the encapsulated cell drop with the protease to produce a cell lysate;

providing a reverse transcriptase and performing a reverse transcription reaction;

performing a droplet merger with barcoding PCR reagents and barcoding beads;

performing a PCR reaction to attach the cell barcodes to the DNA targeted amplicons, RNA targeted amplicons, and protein tag amplicons, wherein all amplicons from the same emulsion contain the same cell barcode; and

detecting and characterizing a DNA, RNA, or protein amplicon by sequencing the cell barcode incorporated into each amplicon.

2. A method according to claim 1, wherein DNA, RNA, and proteins are detected and characterized

3. A method according to claim 1, wherein DNA, RNA, and proteins are detected and characterized from a single cell.

4. A method according to claim 1, wherein the reverse transcription reaction is performed in the same drop as the protease digest.

5. A method according to claim 1, further comprising performing a capture of DNA and RNA amplicons to a solid phase, wherein protein tag amplicons are separated from DNA and RNA amplicons.

6. A method according to claim 1, wherein a droplet merger with the cell lysate and reverse transcription is performed before the PCR reaction is performed.

7. A method according to claim 1, wherein the protease and reverse transcriptase reactions are performed on a PCR thermocycler and later transferred to another instrument for processing and analysis.

8. A method according to claim 4, wherein the instrument for processing and analysis comprises a Mission Bio, Tapestri™ system and the droplet merger with barcoding PCR reagents and barcoding beads is performed on this instrument.

9. A method according to claim 1, further comprising performing a nucleic acid concentration step.

10. A method according to claim 1, further comprising making a protein library.

11. A method according to claim 1, further comprising making a DNA library.

12. A method according to claim 1, further comprising making a RNA library.

13. A method according to claim 1, further comprising making a DNA, protein, and RNA library.

14. A method according to claim 1, wherein a reverse transcription reaction to produce a reverse transcription product is performed before a nucleic acid amplification step.

15. A method according to claim 1, wherein in the step, of performing a capture of DNA and RNA amplicons to a solid phase, the solid phase is a bead.

16. The method of claim 1, wherein one or more DNA, RNA, and protein listed in Figure 9 is detected and characterized.

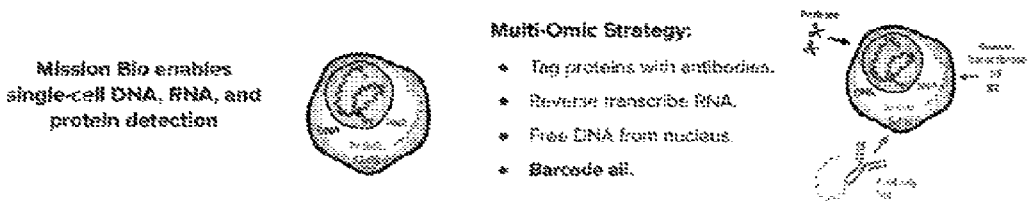
17. The method of claim 1, wherein two or more DNA, RNA, and protein listed in Figure 9 are detected and characterized.

18. The method of claim 1, wherein three or more DNA, RNA, and protein listed in Figure 9 are detected and characterized.

19. The method of claim 1, wherein four or more DNA, RNA, and protein listed in Figure 9 are detected and characterized.

20. The method of claim 1, wherein five or more DNA, RNA, and protein listed in Figure 9 are detected and characterized.

FIGURE 1



The workflow is enabled by its two droplet steps (below). Cells are individually encapsulated in a first droplet (1A) where analyte preparation can occur, including aggressive digestion and subsequent protease heat inactivation (1B). Afterwards, sensitive enzymes that are incompatible with the earlier preparation can be added to make a new drop (2A) where barcoding and amplification proceed (2B).

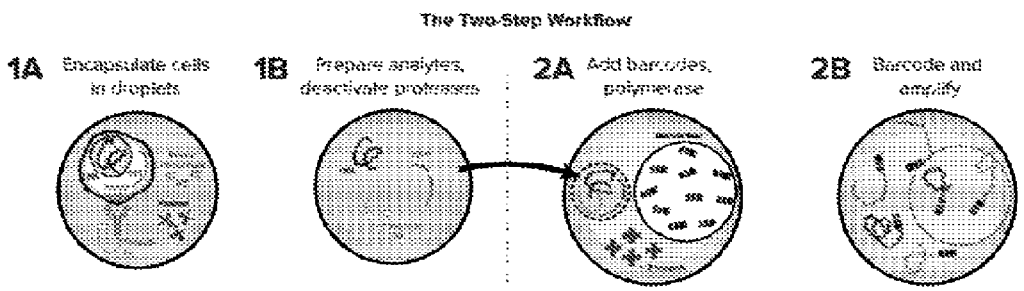


Figure 2

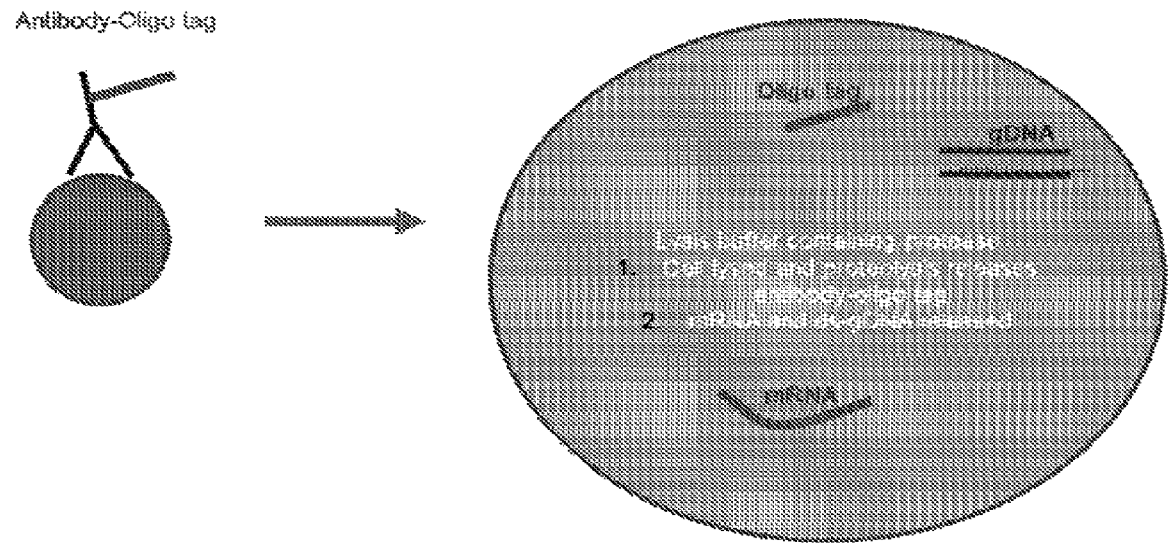
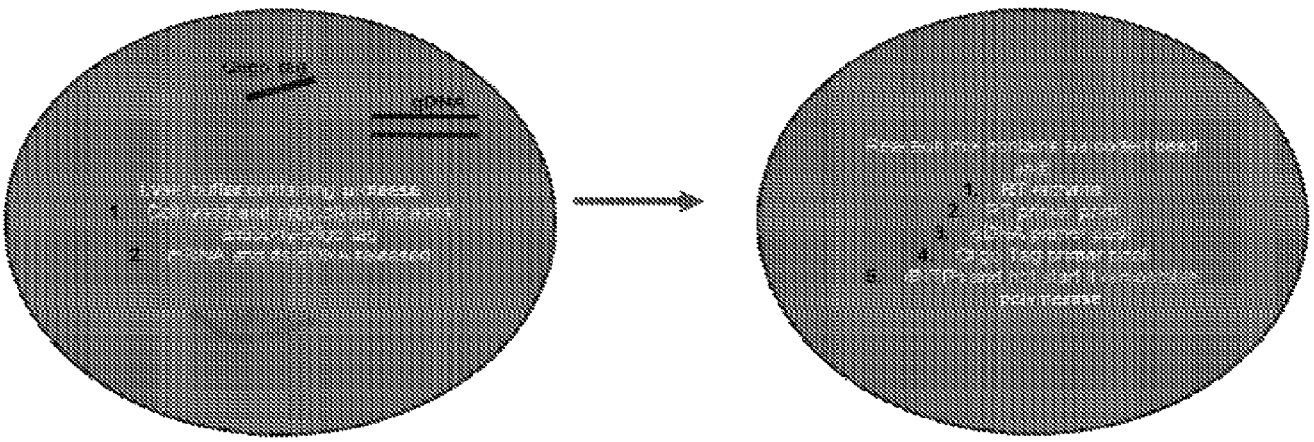


Figure 3



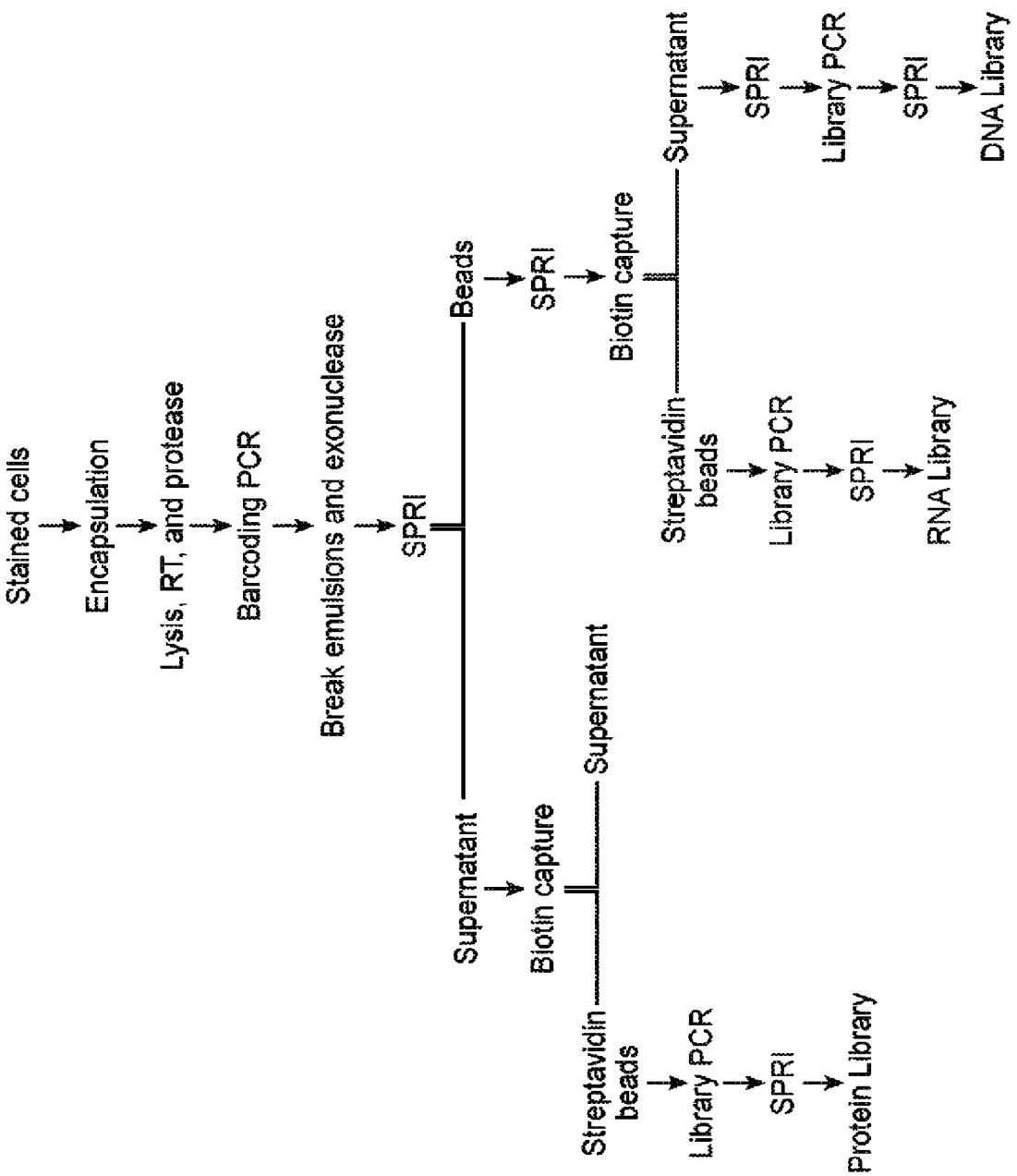
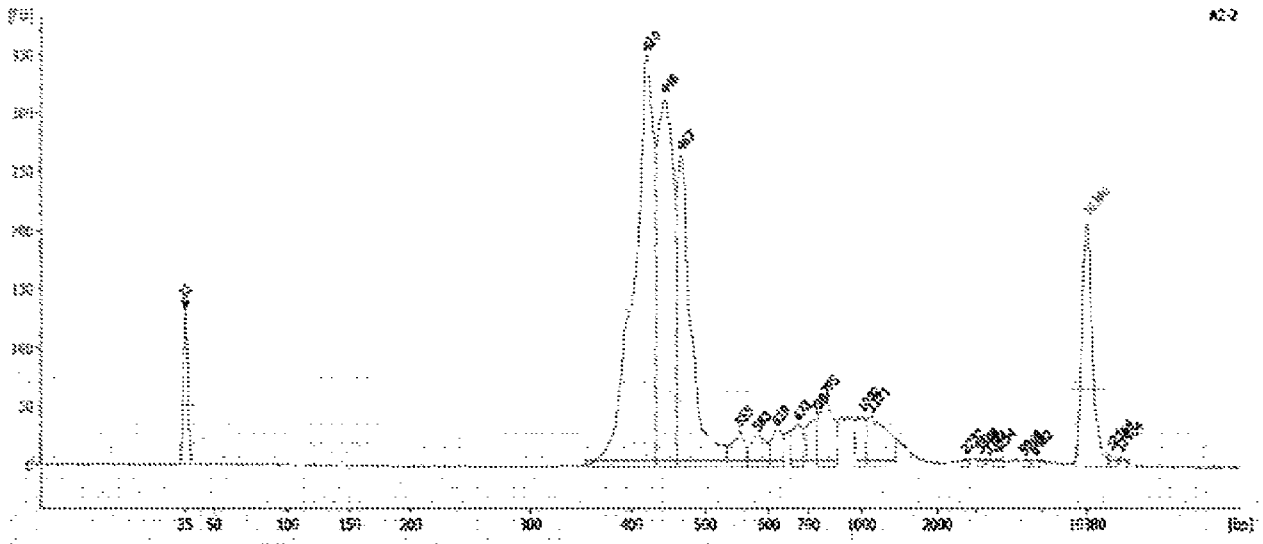


FIG. 4

FIGURE 5



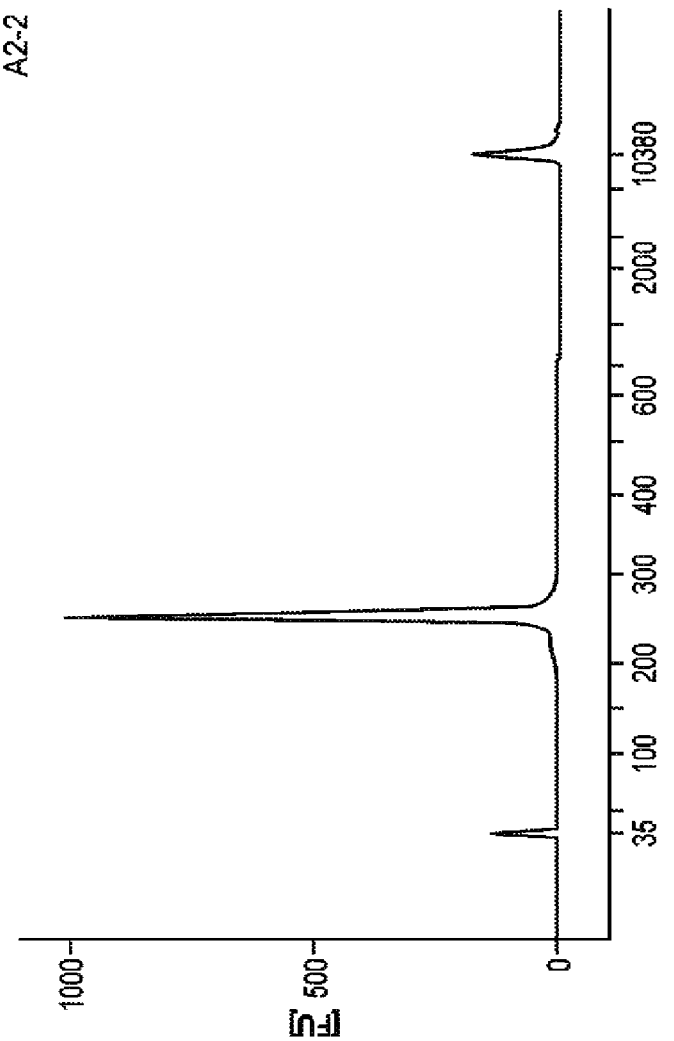


FIG. 6

FIGURE 7

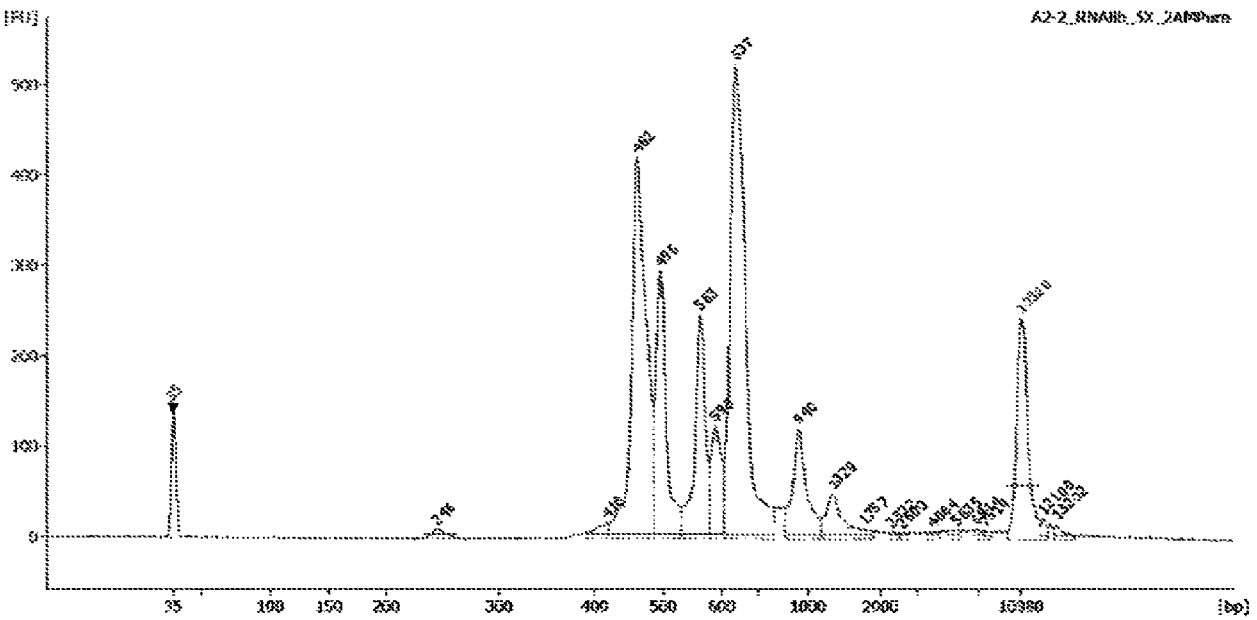

[FU]
A2-2_RNAlib_SX_2AMPure
500
400
300
200
100
0
35
100
150
200
300
400
500
600
1000
2000
10380
[bp]

FIGURE 8

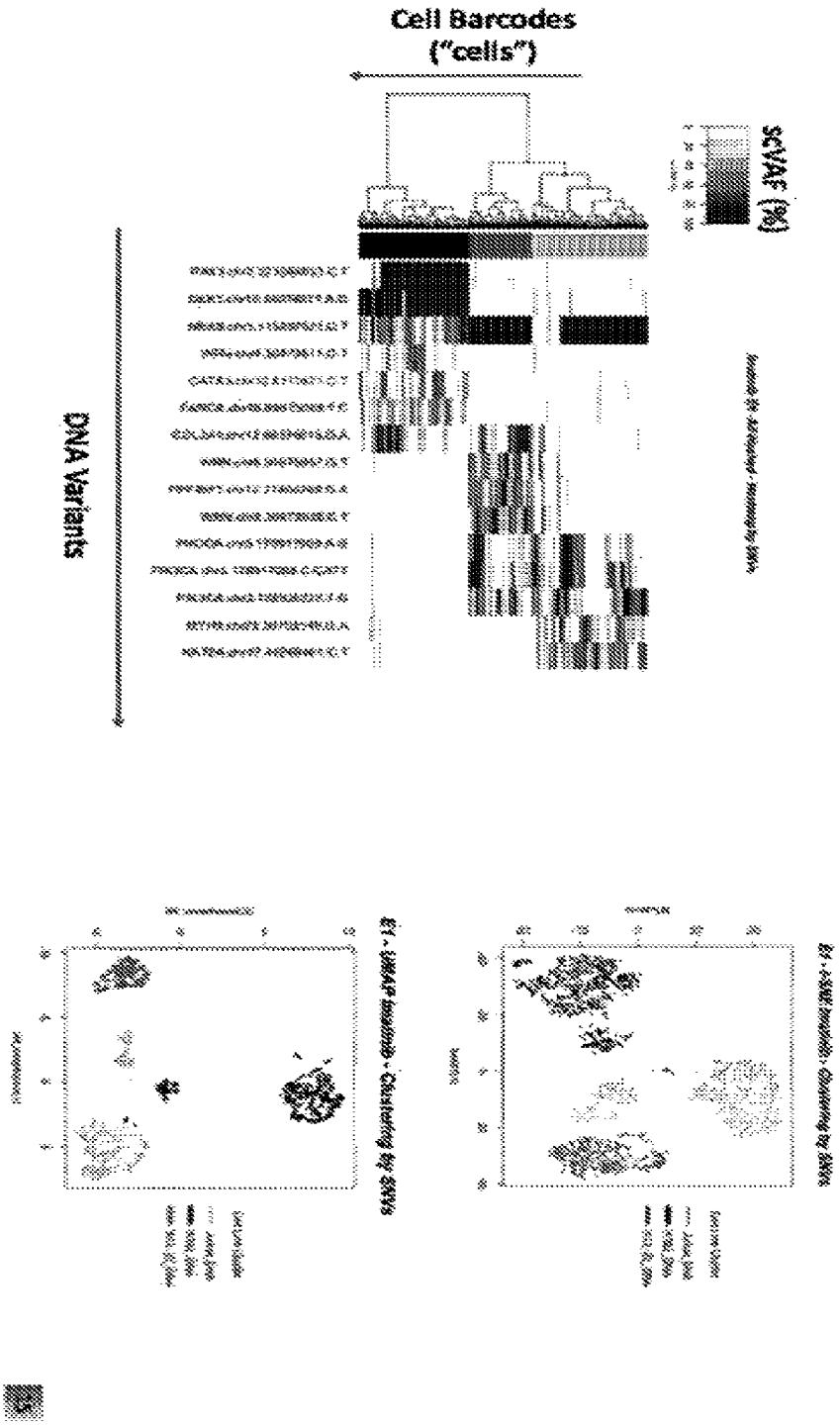


FIGURE 9

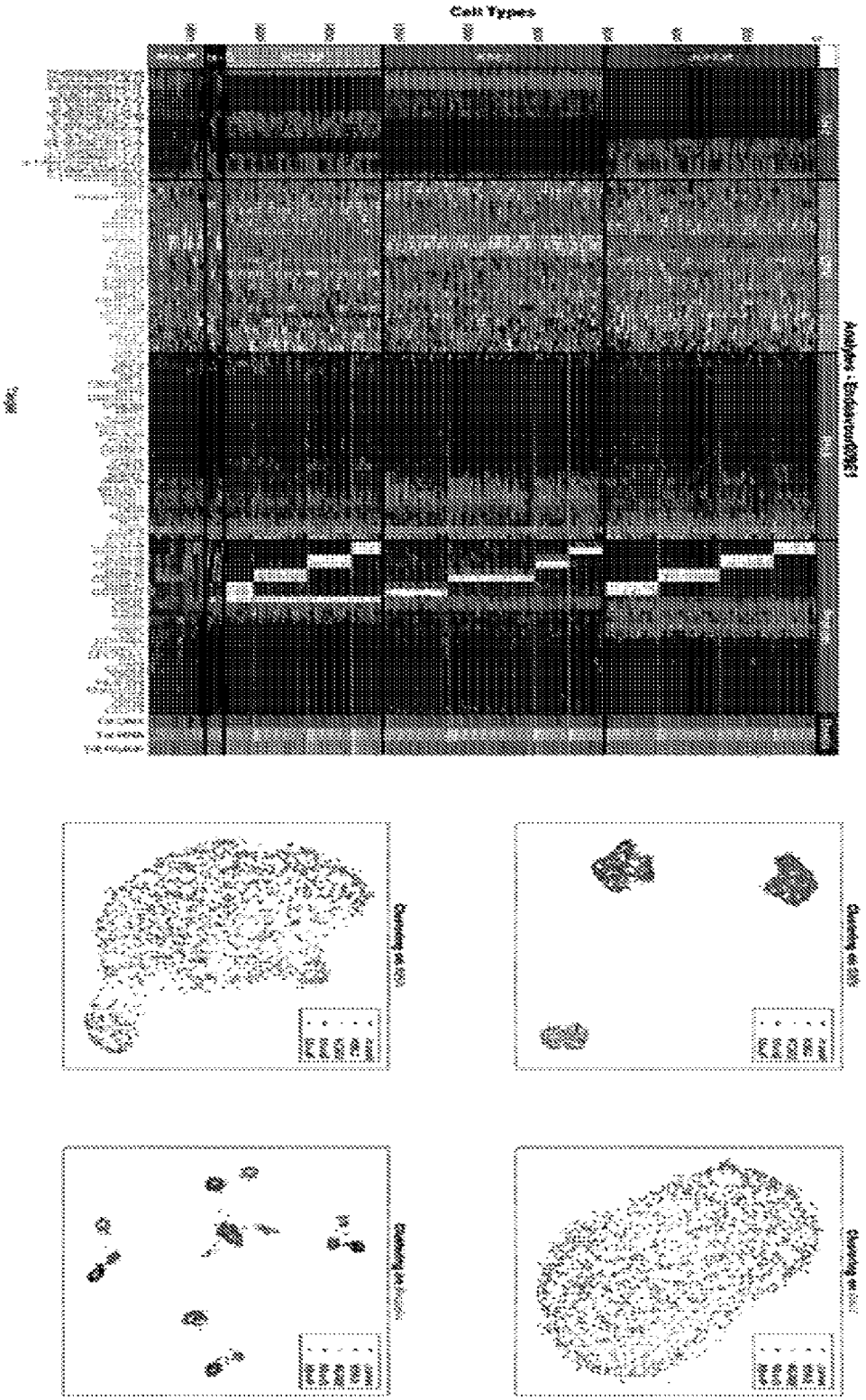


FIGURE 10

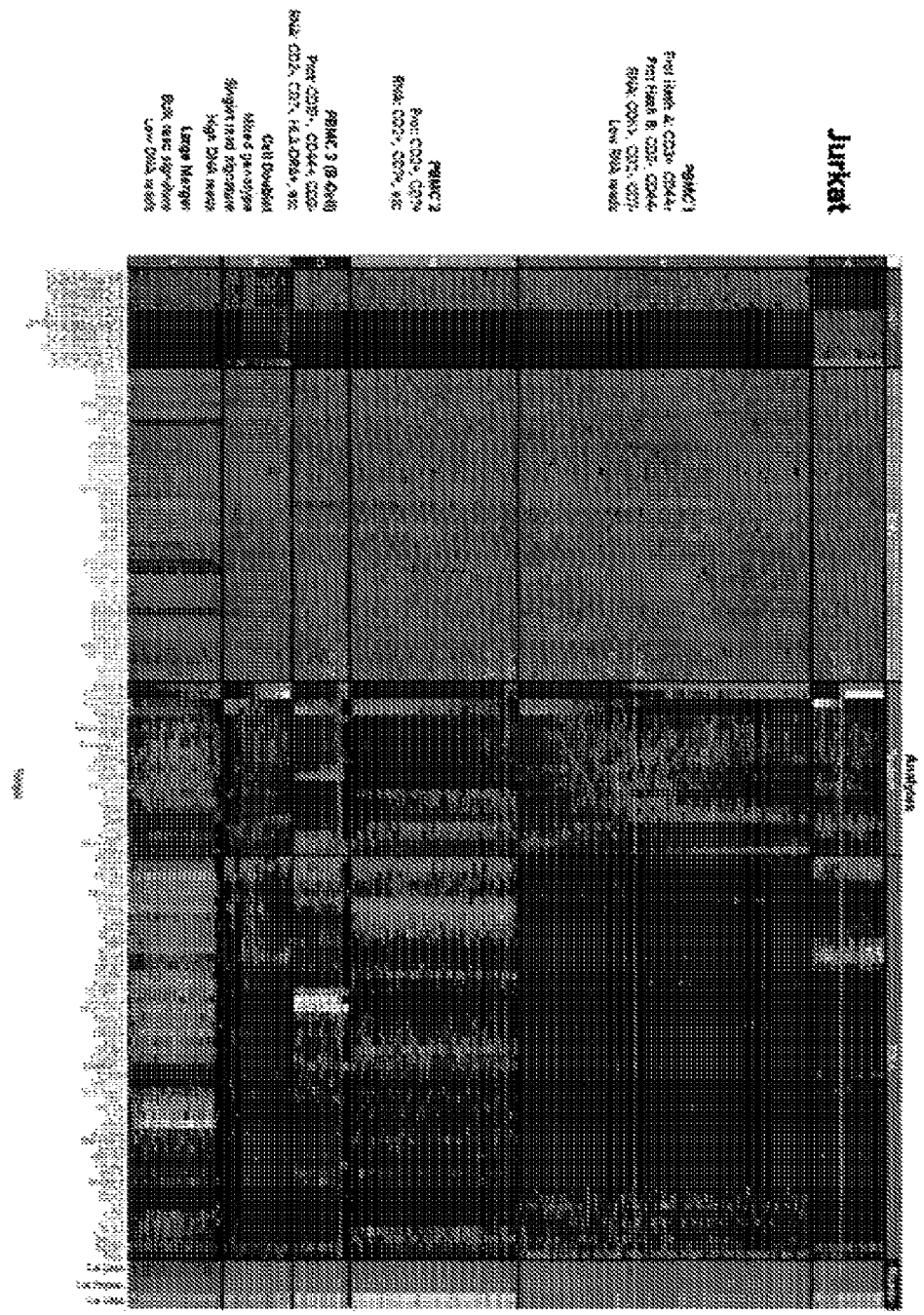
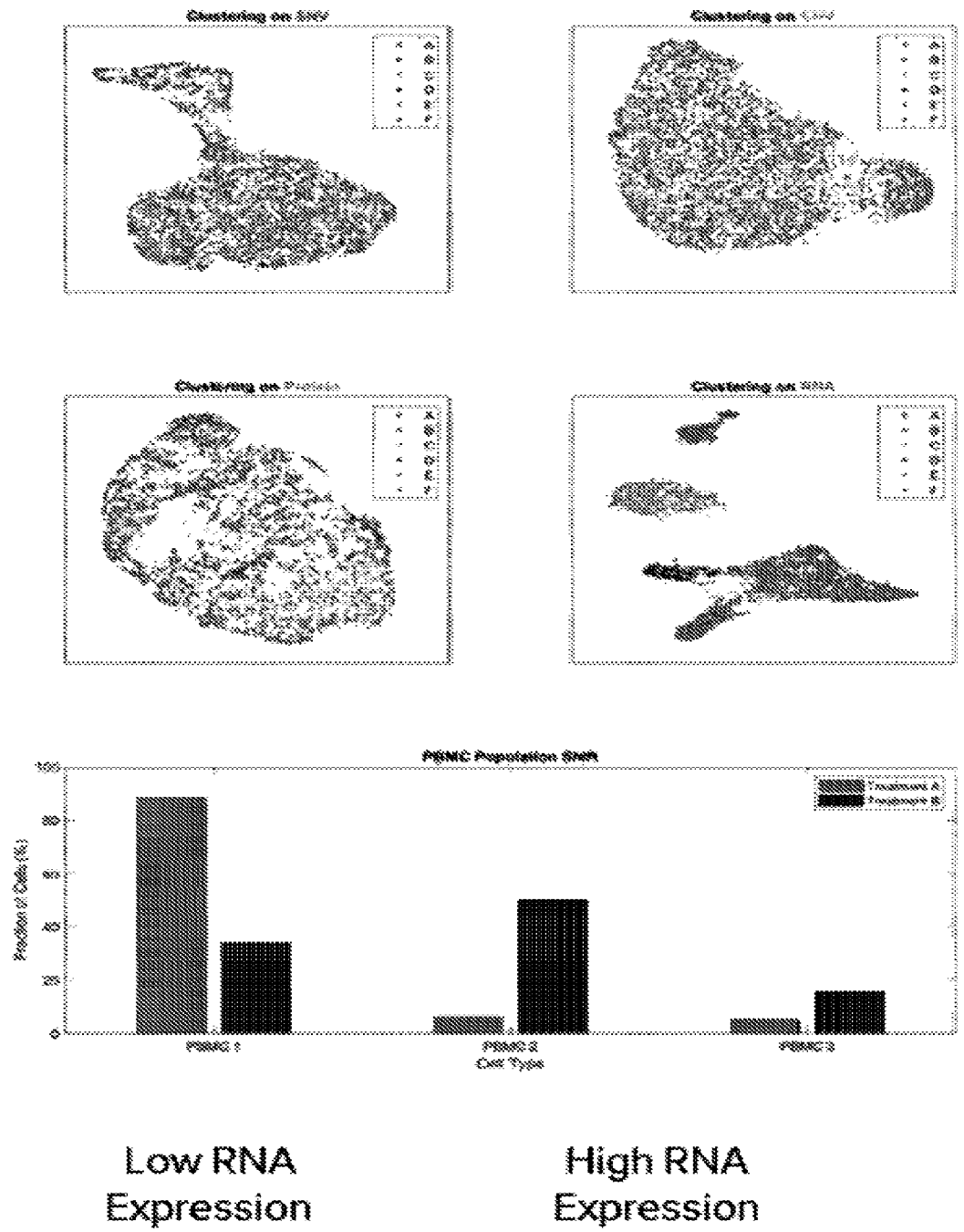
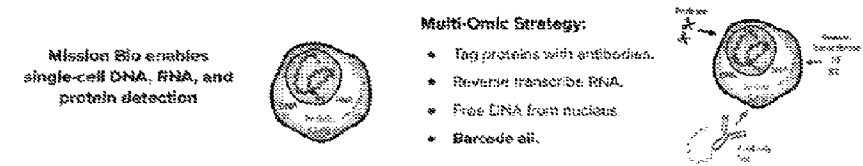


FIGURE 11


PBMC Population Shift
Cell Type
Low RNA Expression
High RNA Expression



The workflow is enabled by its two droplet steps (below). Cells are individually encapsulated in a first droplet (1A) where analyte preparation can occur, including aggressive digestion and subsequent protease heat inactivation (1B). Afterwards, sensitive enzymes that are incompatible with the earlier preparation can be added to make a new drop (2A) where barcoding and amplification proceed (2B).

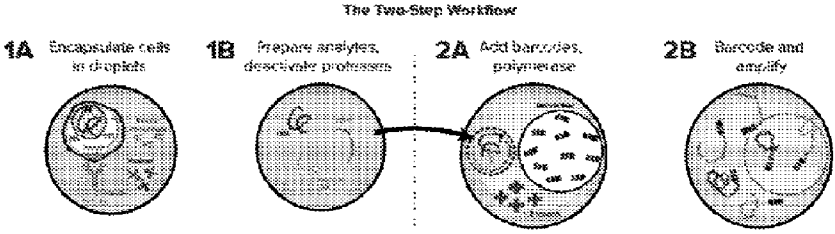


FIGURE 1