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

(19) **World Intellectual Property Organization**
International Bureau

(43) **International Publication Date**
13 August 2020 (13.08.2020)

(10) **International Publication Number**
WO 2020/163789 A1

(51) **International Patent Classification:**
G01N 33/532 (2006.01) *G01N 33/566* (2006.01)

(21) **International Application Number:**
PCT/US2020/017327

(22) **International Filing Date:**
07 February 2020 (07.02.2020)

(25) **Filing Language:** English

(26) **Publication Language:** English

(30) **Priority Data:**
62/803,067 08 February 2019 (08.02.2019) US
62/945,533 09 December 2019 (09.12.2019) US
62/962,788 17 January 2020 (17.01.2020) US

(71) **Applicant: THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK** [US/US]; 412 Low Memorial Library, 535 West 116th Street, New York, NY 10027 (US).

(72) **Inventors: CHAVEZ, Alejandro**; 100 Riverside Blvd., Unit 9K, New York, NY 10069 (US). **SHENG, Jiemin**; 611 W 171, Apt 1B, New York, NY 10032 (US).

(74) **Agent: TYNER, Lisa D.** et al.; McCarter & English, LLP, 265 Franklin Street, Boston, MA 02110 (US).

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

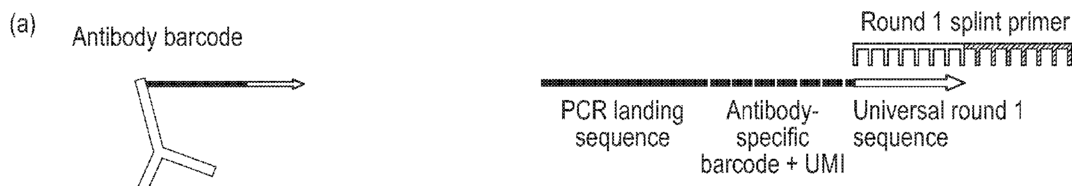
(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR,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

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

Published:
— *with international search report (Art. 21(3))*
— *with sequence listing part of description (Rule 5.2(a))*

(54) **Title:** SINGLE CELL/EXOSOME/VESICLE PROTEIN PROFILING

Figure 1



(57) **Abstract:** The present disclosure is directed, at least in part, to methods and systems for quantifying the levels of multiple, e.g., over a hundred or more, target molecules, e.g., proteins, in, or on the surface of, single entities, including single exosomes, single cells or single vesicles.

WO 2020/163789 A1

SINGLE CELL/EXOSOME/VESICLE PROTEIN PROFILING

RELATED APPLICATIONS

[0001] This application claims the benefit of priority to U.S. Provisional Patent Application No. 62/803,067, filed February 8, 2019, U.S. Provisional Patent Application No. 62/945,533, filed December 9, 2019, and U.S. Provisional Patent Application No. 62/962,788, filed January 17, 2020. The content of these applications is incorporated herein by reference in their entirety.

SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on February 6, 2020, is named 123756-01220_SL.txt and is 1,794 bytes in size.

BACKGROUND

[0003] The treatment of disease begins with an accurate and timely diagnosis of a patient. Many diagnoses are performed by examining well-known surrogate markers that can be assayed for in routine laboratory tests, such as white blood cell counts or serum analyte levels. Often, test results combined with experience and physical symptoms may be enough for the physician to make a confident diagnosis. However, there are still many instances where biopsies are used to confirm these diagnoses. Biopsies are, by nature, physically invasive and time consuming to perform. Even after diagnosis and treatment, physicians often have to revisit their hypotheses based on a patient's response to treatment. Follow-up exams, much like primary diagnostics, may rely heavily upon surrogate markers that do not provide adequate biological insight into the body's current state. Now with the increased realization that the course of disease and response to treatment is not uniform, which has prompted the adoption of so-called personalized medicine, real-time tissue health monitoring are needed. However, there is still a lack of non-invasive technologies that enable near real-time readout of tissue state.

[0004] Recent single-cell RNA-seq (scRNA-seq) methodology has improved the ability of researchers to study cellular composition and heterogeneity due to its unbiased and high-

throughput nature. New scRNA-seq based methods (e.g., CITE-seq, REAP-seq) can produce digital, sequencing-based readout for protein levels by conjugating antibodies to oligonucleotides that contain a barcode for antibody identification. However, these methods require access and familiarity with specialized equipment (e.g., mass cytometry (CyTOF), immunofluorescence microscopy, and/or flow cytometry) for single-cell library generation and do not use unique molecular identifiers (UMIs).

[0005] Thus, there is a need for noninvasive and near real-time readout to quantify the abundance and state of proteins in a cellular or exosomal sample isolated from a patient before and after treatment using technologies that do not require in-house specialized equipment and have the ability to scale to hundreds of protein targets in a patient sample.

SUMMARY OF THE DISCLOSURE

[0006] The present disclosure is directed, at least in part, to methods and systems for accurately quantifying the levels of multiple, e.g., over a hundred or more, target molecules, e.g., proteins, in, or on the surface of, single entities, including single exosomes, single cells or single vesicles. In some embodiments, the disclosure provides methods for high throughput detection of a plurality of target molecules within pooled single exosomes, single cells or single vesicles using DNA-barcoded antibodies (DBAs), split-pool sequencing, and optionally unique molecular identifiers (UMIs). The methods and systems of the disclosure require no specialized equipment to scale to hundreds of protein targets, thereby providing cost-effective, modular and easy-to-use quantitative single cell protein profiling platform.

[0007] The methods and systems of the disclosure can be applied to early stage discovery pipelines to identify pathways of interest (e.g., in a preclinical setting) and in clinical settings for diagnostic and/or therapeutic applications for a variety of diseases and disorders. For example, the methods and systems of the disclosure can be applied to identify which patients with, for example, cancer, autoimmune disease or inflammatory disease are more likely to respond to any of a diverse set of therapeutics, e.g., immune regulators, such as, for example, anti-PD1 antibodies).

[0008] Accordingly, in some aspects, the present disclosure is directed to methods for detecting a plurality of target molecules in, or on the surface of, entities such as exosomes, cells or vesicles in a sample, the method comprising: (a) contacting the entities such as exosomes, cells or vesicles with a plurality of target molecule-binding agents, wherein each target molecule-binding agent comprises a nucleic acid barcode and optionally a unique

molecular identifier (UMI), and wherein target molecule-binding agents that are specific to an identical target molecule share an identical nucleic acid barcode.

[0009] In some aspects, the present disclosure is directed to methods for diagnosing a disease or disorder, e.g., cancer, autoimmune disease or inflammatory disease, determining a patient's response to a therapy, and/or monitoring the patient for undesired side effects to the therapy, the method comprising: (a) contacting exosomes, cells or vesicles with a plurality of target molecule-binding agents, wherein each target molecule-binding agent comprises a nucleic acid barcode and optionally a unique molecular identifier (UMI), wherein target molecule-binding agents that are specific to an identical target molecule share an identical nucleic acid barcode, wherein the exosomes, cells or vesicles are isolated from a patient suffering from a disease or disorder, e.g., cancer, autoimmune disease or inflammatory disease, wherein the plurality of target molecules comprise molecules that are markers indicative of a disease or disorder, e.g., cancer, autoimmune disease or inflammatory disease, efficacy of the therapy, and/or are indicative of undesirable side effects, and wherein expression levels of the target molecules determined in the patient are compared with expression levels of the corresponding target molecules determined in normal controls, thereby diagnosing a disease or disorder, e.g., cancer, autoimmune disease or inflammatory disease, determining the patient's response to a therapy, and/or monitoring for undesired side effects of a therapy.

[0010] In some embodiments, the markers are tested before and after treatment with a therapy.

[0011] In some embodiments, the patient's response to therapy is determined and/or monitored in real-time.

[0012] In some embodiments, the patient is further treated with the same therapy, if the therapy is effective, but associated with little or no undesired side effects; or the patient is treated with a different therapy if the therapy is not effective and/or is associated with undesired side effects.

[0013] In some embodiments, each target molecule-binding agent described herein above further comprises a universal round 1 primer sequence at the 3' end for a first round extension.

[0014] In some embodiments, the methods described herein further comprises:

(b) dividing the exosomes, cells or vesicles into at least two primary aliquots, the at least two primary aliquots comprising at least a first primary aliquot and a second primary aliquot;

(c) adding primary nucleic acid tags to the target molecule-binding agents in the at least two primary aliquots, wherein the primary nucleic acid tags added to the target molecule-binding agents in any one of the at least two primary aliquots are different from the primary nucleic acid tags added to the target molecule-binding agents in any one of the other primary aliquots;

(d) combining the at least two primary aliquots;

(e) dividing the combined primary aliquots into at least two secondary aliquots, the at least two secondary aliquots comprising a first secondary aliquot and a second secondary aliquot;

(f) adding secondary nucleic acid tags to the at least two secondary aliquots, wherein the secondary nucleic acid tags added to the target molecule-binding agents in the first secondary aliquot are different from the secondary nucleic acid tags added to the target molecule-binding agents in the second secondary aliquot; and

(g) repeating steps (d), (e), and (f) with the at least two secondary aliquots a number of times sufficient to generate a unique series of nucleic acid tags for each exosome, cell or vesicle in the sample.

[0015] In some embodiments, the primary nucleic acid tags, the secondary nucleic acid tags and/or subsequent nucleic acid tags described herein above are added by ligation reactions, polymerase extension reactions, and/or chemical syntheses.

[0016] In some embodiments, the nucleic acid tags described herein above are added by polymerase extension reaction,

wherein the nucleic acid barcode bound to each target molecule-binding agent is extended with one of the primary nucleic acid tags by contacting the exosomes, cells or vesicles with a strand displacing polymerase and a first DNA hairpin, wherein the first DNA hairpin comprises:

(i) a first oligonucleotide comprising a sequence complementary to the universal round 1 primer sequence, and

(ii) a second oligonucleotide, wherein the second oligonucleotide comprises a third oligonucleotide comprising a primary nucleic acid tag, and a fourth oligonucleotide comprising a sequence complementary to the primary nucleic acid tag;

wherein the third oligonucleotide is located at the 5' end of the second oligonucleotide, and the fourth oligonucleotide is located at the 3' end of the second oligonucleotide; and

wherein the first oligonucleotide is fused to the 3' end of the fourth oligonucleotide.

[0017] In some embodiments, each primary nucleic acid tag described herein above comprises a unique well-specific first round barcode sequence at the 5' end and a universal round 2 primer sequence at the 3' end.

[0018] In some embodiments, the first DNA hairpin described herein above is disabled at the end of each polymerase extension by removal of the first oligonucleotide from the first DNA hairpin using an exonuclease or by treating the first DNA hairpin with an enzyme that remove a unique base present at the junction of the first oligonucleotide and the fourth oligonucleotide.

[0019] In some embodiments, the method described herein further comprising adding secondary nucleic acid tags by polymerase extension reaction,

wherein the nucleic acid barcode bound to each target molecule-binding agent is further extended with one of the secondary nucleic acid tags by contacting the exosomes, cells or vesicles with a strand displacing polymerase and a second DNA hairpin wherein the second DNA hairpin comprises:

(i) a fifth oligonucleotide comprising a sequence complementarity to the universal round 2 primer sequence, and

(ii) a sixth oligonucleotide, wherein the sixth oligonucleotide further comprises a seventh oligonucleotide comprising a secondary nucleic acid tag, and an eighth oligonucleotide comprising a sequence encoding a sequence complementary to the secondary nucleic acid tag;

wherein the seventh oligonucleotide is located at the 5' end of the sixth oligonucleotide, and the eighth oligonucleotide is located at the 3' end of the sixth oligonucleotide; and

wherein the fifth oligonucleotide is fused to the 3' end of the eighth oligonucleotide.

[0020] In some embodiments, each secondary nucleic acid tag comprises a unique well-specific second round barcode sequence at the 5' end and a universal round 3 primer sequence at the 3' end.

[0021] In some embodiments, the primary nucleic acid tags described herein above are added by ligation reaction in a first round splint-ligation reaction, wherein the primary nucleic acid tags are added to the 3' end of the nucleic acid barcode bound to each target molecule-

binding agent by contacting the exosomes, cells or vesicles with a ligase, a first-round oligonucleotide and a first-round splint sequence,

wherein the first-round oligonucleotide comprises a 5' common region followed by a primary nucleic acid tag terminated by a 3' universal round 2 sequence; and

wherein the first-round splint sequence comprises a region complementary to the universal round 1 sequence at 3' end of the nucleic acid barcode bound to each target molecule-binding agent and a region complementary to the 5' common region of the first-round oligonucleotide.

[0022] In some embodiments, the splint-ligation process described above is terminated at the end of the first-round splint-ligation reaction.

[0023] In some embodiments, secondary nucleic acid tags are added through a second round splint ligation reaction,

wherein the nucleic acid barcode bound to each target molecule-binding agent is further extended with a secondary nucleic acid tag by contacting the exosomes, cells or vesicles with a ligase, a second round oligonucleotide and a second round splint sequence,

wherein the second round oligonucleotide comprises a 5' common region followed by a secondary nucleic acid tag terminated by a 3' universal round 3 sequence or a universal PCR sequence; and

wherein the second round splint sequence comprises a region complimentary to the 3' universal round 2 sequence of the first-round oligonucleotide, and a region complimentary to the 5' common region of the second round oligonucleotide.

[0024] In some embodiments, the methods described herein above further comprise amplifying and sequencing the nucleic acid barcodes, the UMIs and/or the nucleic acid tags. In some embodiments, split-pool sequencing is used to sequence the barcodes, UMIs and/or nucleic acid tags.

[0025] In some embodiments, the target molecules described herein above are proteins, sugar moieties, lipids, and/or polynucleotides.

[0026] In some embodiments, the target molecule-binding agents described herein above comprise an antibody, an antibody fragment, a peptide aptamer, lectins, a phage display system or a yeast display system.

[0027] In some embodiments, the nucleic acid barcode described herein above is a DNA-barcode.

[0028] In some embodiments, the detecting method described herein above is at a single cell

or a single exosome or a single vesicle level.

[0029] In some embodiments, each of the at least two primary aliquots described herein above consists of a single cell, a single exosome, or a single vesicle.

BRIEF DESCRIPTION OF DRAWINGS

[0030] Figures 1A-C. Overview of an exemplary ligation reaction. A barcode is added to the 3' end of an antibody barcode through a splint ligation process.

[0031] Figures 2A-F. Overview of polymerase extension reaction (PER). Barcoding at each "split" step is performed using Polymerase Extension Reaction (PER). Hairpins are designed with a 3' poly dA tail to prevent priming and extension on the 3' end of the hairpin. The barcode sequence in green to be appended to the antibody is well-specific and comprises a sequence that allows for subsequent second round extension or PCR amplification depending on the barcoding capacity required for the experiment. Notably, the hairpin can also possess a modified base in the first paired position (denoted by the star). This allows for enzymatic digestion of the hairpin after the reaction to minimize crosstalk and mis-extension in the subsequent "split" extension/PCR steps. Modifications may include but are not limited to, uracil, inosine, or 5' hairpin phosphorylation (not pictured). (A) Antibodies bound to exosomes are conjugated with oligos that possess an antigen specific barcode and UMI, as well as a 3' terminal landing sequence. They are introduced in solution with a barcoding hairpin, a strand displacing polymerase, appropriate dNTPs, and supplemented with enzyme buffer. The reaction is allowed to proceed isothermally at a temperature appropriate to the sample and the polymerase. (B) The 3' terminal landing sequence anneals on the exposed complementary 3' overhang on the hairpin. (C) The strand displacing polymerase extends the oligo, appending to it the appropriate barcode sequence before encountering a stop signal prompting the reaction to terminate. (D) Branch migration allows the newly extended oligo on the antibody to spontaneous dissociate from the hairpin. (E) The hairpin, newly re-annealed, is recycled to extend other unextended oligos in solution. (F) After the reaction is completed the appropriate enzyme is introduced into the reaction to cleave the landing pad from the hairpin to prevent inappropriate extension of unextended oligos from the current round in subsequent rounds.

[0032] Figures 3A-C. Design of DNA oligonucleotide and barcoding hairpins. (A) Antibody barcodes are conjugated such that each antibody will have an antibody-specific barcode that encode the identity of the corresponding antigen to which the antibody binds as

well as a UMI. All oligos have the same "universal round 1" sequence appended to their 3' end to facilitate extension by PER. (B) The first round hairpin has a 3' poly dA tail to prevent priming and extension on the 3' end of the hairpin. Each hairpin also contains a unique well-specific first round barcode as well as a universal sequence common to all hairpins in the same round for use during the next round of extension. The sequence that gets extended by the strand displacing polymerase is comprised of As, Cs, and Ts. The black circle shown in the diagram represents a C within the hairpin template which would require the addition a G but the reactions are setup devoid of dGTP and as such the polymerase stalls upon the need to add this base and does not extend the oligonucleotide any further. (C) The second round hairpin is similar in structure and function to first except it contains unique sequences which enable its use for extension by only DNA oligonucleotides that were correctly extended during the first round of PER. Similar to the first round, each hairpin that is provided to a well labels all molecules with a unique well-specific second round barcode as well as a universal sequence common to all hairpins in the same round for use during additional rounds of PER as for PCR to generate the final product that is ready for DNA sequencing, with e.g., split-pool sequencing.

[0033] Note that here dashed species represent barcode regions that vary from antibody to antibody or between hairpin barcodes. Solid lines represent sequences that are universal among all oligonucleotides of the same type. Furthermore, the modification from the first and second round extensions (denoted by star and hexagon) must differ to prevent premature degradation of hairpins from enzyme carryover of the first round reaction. Alternatively, a method can be used to disable the enzymes that excise the unconventional base after they have performed their function to enable the same approach to disabling the hairpin to be used for all rounds. While not depicted, other methods of selective hairpin destruction such as, but not limited to, lambda exonuclease based digestion of the hairpin via the addition of 5' phosphate groups to all hairpins, can also be used. (see, e.g., J.Y. Kishi, et al., 2018. Programmable autonomous synthesis of single-stranded DNA. Nature Chemistry, pp. 155-164, the contents of which are hereby incorporated herein by reference).

[0034] Figure 4. Represents an exemplary polymerase extension reaction (PER) comprising two rounds of barcode extension and enzymatic digestion of hairpins following each round, and a third round of barcode PCR.

[0035] Figures 5A-D. Platform for performing single-cell immune profiling: (A) Antibody with a conjugated DNA oligo. Each oligo contains an antibody specific barcode and a UMI. (B) Antibodies bind specifically to their cellular targets. (C) During split-pool sequencing, all

of the oligos associated with a given cell are identically labeled, yet no two cells apply the same tag to their oligos. (D) During analysis, all oligos associated with a given cell are identified via their tag, and the composition and quantity of all bound antibodies is determined by analyzing the BC and UMI, respectively. For simplicity, only antibodies binding to cell surface proteins are shown.

[0036] Figures 6A-B. Overview of a single exosome sequencing (SESeq) approach. (A) A million unique DNA barcoded gel beads are mixed with isolated exosomes along with the required enzymes for reverse transcription using microfluidics. Oil is then used to partition the sample, isolating a single bead with a single exosome. Each exosome is then reverse transcribed and all RNAs from a single exosome are tagged with a common DNA barcode. (B) Upon sequencing the RNAs belonging to each exosome can be determined by using the unique DNA barcode added to each transcript during sample preparation. This allows to know all the RNAs within in each exosome.

[0037] Figure 7. The computational algorithm used in the methods of the disclosure is robust to sparse and noisy data. Box plts show how well one (box) or more cells correlate (Pearson's r) with bulk RNA-seq data. The grey plots show the correlation when using gene expression data. The correlation between two cells is approximately 0.95, while using gene expression about 10-fold more cells are required for a similar correlation.

[0038] Figure 8. Overview of exemplary methods of the disclosure used in clinical diagnostics. A blood sample is first collected from a cancer patient undergoing chemotherapy, exosomes are isolated, SESeq is performed and used to monitor tumor response to therapy and to determine potential deleterious side-effects from chemotherapy.

DETAILED DESCRIPTION

[0039] The present disclosure is directed, at least in part, to methods and systems for accurately quantifying the levels of multiple, e.g., over a hundred or more, target molecules, e.g., proteins, in, or on the surface of, single entities, including single exosomes, single cells or single vesicles. In some embodiments, the disclosure provides methods for high throughput detection of a plurality of target molecules within pooled single entities, including single exosomes, single cells or single vesicles using DNA-barcoded antibodies (DBAs), split-pool sequencing, and/or unique molecule identifiers (UMIs).

[0040] The methods and systems of the disclosure require no specialized equipment to scale to hundreds of protein targets, thereby providing cost-effective, modular and easy-to-use

quantitative single cell protein profiling platform.

[0041] This platform provides methods for use in, e.g., preclinical and clinical settings for diagnostic and therapeutic applications, using DBAs and split-pool sequencing to provide cost effective, modular and simple to use quantitative protein profiling of a single entity, e.g., a single exosome, a single cell, or a single vesicle.

[0042] In some embodiments, the methods and systems of the disclosure can be applied to early stage discovery pipelines to identify pathways of interest (e.g., in a preclinical setting) and in clinical settings for diagnosis, monitoring, or therapy of various diseases or disorders, such as immune diseases and disorders, e.g., cancer, autoimmune disease, and inflammatory disease.

[0043] In some embodiments, the methods and systems of the disclosure can be applied to identify which patients with a disease or disorder, e.g., cancer, autoimmune disease or inflammatory disease, are more or less likely to respond to any of a diverse set of therapeutics, e.g., immune regulators (such as, for example, an anti-PD1 antibody). Furthermore, in some embodiments, the methods and systems of the disclosure can be used to monitor a response to a particular therapeutic.

[0044] Furthermore, the methods of the disclosure can be used for quantifying the abundance and state (e.g. phosphorylation state) of hundreds of proteins at the level of a single entity, e.g., cell, vesicle or exosome, enables comprehensive phenotyping within a variety of disease states, and have translational applications in early stage discovery pipelines or in guiding clinical decision making.

[0045] The methods of the disclosure are superior to existing technologies which require specialized equipment and lack the ability to scale to hundreds of protein targets (e.g., mass cytometry (CyTOF), immunofluorescence microscopy, flow cytometry).

[0046] Because of its potential to simultaneously monitor multiple tissues, methods of the disclosure are powerful tools for studying off-target drug response studies which may be of interest in, for example, chemotherapies, where both on-target effects at a tumor and off-target effects at healthy tissue can be simultaneously monitored.

[0047] Recent work in bioinformatics, which allows an understanding of genetic interactions and network dynamics of gene signaling, expands the impact of this technology by allowing for educated inferences about the dynamics of intracellular activity by protein signaling manifested on the cell surface.

[0048] Thus, the methods and systems of the disclosure provide a cost-effective, modular, and easy-to-use quantitative single entity (e.g., cell, exosome, or vesicle) protein profiling

platform. These methods utilize DNA-barcoded antibodies and, optionally, unique molecule identifiers (UMIs) in combination with split-pool sequencing to improve information content and to reduce cost. The methods enable multiplexed and highly scalable detection of proteins at the single entity level. These methods and systems do not require in-house specialized equipment or training and can be applied in early stage discovery pipelines to identify pathways of interest and in clinical settings to diagnose and monitor treatment of patients, and to determine and/or predict which patients are more or less likely to respond to specific therapeutics, e.g., immune regulators, such as, for example, anti-PD1 antibodies, for the treatment of cancer, autoimmune disease, or inflammatory disease.

<u>**Definitions**</u>

[0049] As used herein and in the appended claims, the singular forms "**a**", "**an**", and "**the**" include plural reference unless the context clearly dictates otherwise. Thus, for example, reference to a "protein" is a reference to one or more proteins and equivalents thereof known to those skilled in the art, and so forth.

[0050] As used herein, the term "**about**" means plus or minus 20% of the numerical value of the number with which it is being used. Therefore, about 50% means in the range of 40%-60%.

[0051] The transitional term "**comprising**", which is synonymous with "including," "containing," or "characterized by," is inclusive or open-ended and does not exclude additional, unrecited elements or method steps. See MPEP 2111.03.

[0052] As used herein, the term "**target molecule-binding agent**" as used herein refers to any naturally occurring or synthetic biological or chemical molecule which specifically binds an identified target molecule. The binding can be covalent or non-covalent, i.e., conjugated or by any known means taking into account the nature of the target molecule-binding agent and its respective target. In some embodiments, a target molecule-binding agent is a protein, e.g., an antibody, that specifically binds to a target protein. In some embodiments, the target molecule-binding agent comprises an antibody, an antibody fragment, or a peptide aptamer.

[0053] The term "**antibody**", as used herein, broadly refers to any immunoglobulin (Ig) molecule comprised of four polypeptide chains, two heavy (H) chains and two light (L) chains, or any functional fragment, mutant, variant, or derivation thereof, which retains the essential epitope binding features of an Ig molecule. Examples antibody fragments include, but are not limited to, Fab, Fab', F(ab')2, Fv, immunologically functional immunoglobulin

fragments, heavy chain, light chain, and single-chain antibodies. Such mutant, variant, or derivative antibody formats are known in the art. In a full-length antibody, each heavy chain is comprised of a heavy chain variable region (abbreviated herein as HCVR or VH) and a heavy chain constant region. The heavy chain constant region is comprised of three domains, CH1, CH2 and CH3. Each light chain is comprised of a light chain variable region (abbreviated herein as LCVR or VL) and a light chain constant region. The light chain constant region is comprised of one domain, CL. The VH and VL regions can be further subdivided into regions of hypervariability, termed complementarity determining regions (CDR), interspersed with regions that are more conserved, termed framework regions (FR). Each VH and VL is composed of three CDRs and four FRs, arranged from amino-terminus to carboxy-terminus in the following order: FR1, CDR1, FR2, CDR2, FR3, CDR3, FR4. Immunoglobulin molecules can be of any type (*e.g.*, IgG, IgE, IgM, IgD, IgA and IgY), class (*e.g.*, IgG 1, IgG2, IgG 3, IgG4, IgA1 and IgA2) or subclass.

[0054] The term "**barcode**" as used herein, refers to any unique, non-naturally occurring, nucleic acid sequence that may be used to identify the originating source of a molecule. Barcoding may be performed based on any of the compositions or methods disclosed in patent publication WO 2014047561 Al, "Compositions and methods for labeling of agents," incorporated herein in its entirety. Additionally, other barcoding designs and tools have been described (see e.g., Birrell et al., (2001) Proc. Natl Acad. Sci. USA 98, 12608-12613; Giaever, et al., (2002) Nature 418, 387-391; Winzeler et al., (1999) Science 285, 901-906; and Xu et al., (2009) Proc Natl Acad Sci USA. February 17; 106(7):2289-94); Rosenberg et al. (2018) Science, 360; 176-182; Quinodoz, et al. (2018) Cell, 174, 744-757, the contents of each of which are hereby incorporated herein by reference. In some embodiments, the nucleic acid barcode is a DNA barcode. In some embodiments, a barcode, e.g., a DNA barcode, is associated with a protein, e.g., an antibody.

[0055] As used herein, the term "**oligonucleotide**" refers to a nucleic acid such as deoxyribonucleic acid (DNA), ribonucleic acid (RNA), or DNA/RNA hybrids and includes analogs of either DNA or RNA made from nucleotide analogs known in the art (see, *e.g.* U.S. Patent or Patent Application Publications: US 7399845, US 7741457, US 8022193, US 7569686, US 7335765, US 7314923, US 7335765, and US 7816333, US 20110009471, the entire contents of each of which are incorporated herein by reference). Oligonucleotides may be single-stranded (such as sense or antisense oligonucleotides), double-stranded, or partially single-stranded and partially double-stranded.

[0056] As used herein, "**sample**" includes a specimen or culture obtained from any source

which contains a cell, exosome, or vesicle. Biological samples can be obtained from blood (including any blood product, such as whole blood, plasma, serum, or specific types of cells of the blood), a blood fraction containing peripheral blood mononuclear cells, cerebrospinal fluid (CSF), lymph, urine, saliva, semen, sweat, sputum, lacrimal fluid, synovial fluid, cerebrospinal fluid, feces, mucous, vaginal fluid, and spinal fluid. Biological samples also include tissue samples, such as biopsy tissues or pathological tissues that have previously been fixed (*e.g.*, formaline snap frozen, cytological processing, etc.). Methods or obtaining tissue biopsies and body fluids from mammals are well known in the art.

Methods of the Disclosure

[0057] The present disclosure is directed, at least in part, to methods and systems for accurately quantifying the levels of multiple, e.g., over a hundred or more, target molecules, e.g., proteins, in, or on the surface of, single entities, including single exosomes, single cells or single vesicles. In some embodiments, the disclosure provides methods for high throughput detection of a plurality of target molecules in or on pooled single entities, e.g., single exosomes, single cells or single vesicles using DNA-barcoded antibodies (DBAs) and split-pool sequencing.

[0058] Accordingly, in some aspects, the present disclosure provides methods for detecting a plurality of target molecules, e.g., proteins, in, or on the surface of exosomes, cells or vesicles. In some embodiments exosomes, cells or vesicles in a sample are placed in contact with a plurality of target molecule-binding agents, *e.g.,* antibodies, wherein each target molecule-binding agent, *e.g.,* antibody, comprises a nucleic acid barcode and optionally a unique molecular identifier (UMI), and wherein target molecule-binding agents, *e.g.,* antibodies, that are specific to an identical target molecule share an identical nucleic acid barcode.

[0059] In some embodiments, a vesicle a microvesicle, a membrane particle, and/or an apoptotic bleb. In some embodiments, a vesicle is a biological organelles or complexes. In some embodiments, the biological organelle or complex is mitochondria, lysosome, and/or Golgi body. In other embodiments, a vesicle is a synaptosome, a nanoparticle, and/or DNA origami. In some embodiments, the vesicle is are natural or synthetic.

[0060] The cells used in the methods of the disclosure can be of any cell type. In some embodiments, the cells are immune cells. In other embodiments, the cells are cancer cells. In other embodiments, the cells are endothelial cells, stem cells, bone cells, blood cells, muscle

cells, cardiomyocytes, fat cells, skin cells, nerve cells, brain cells, pancreatic cells, sperm cells, egg cells, or any other cell type.

[0061] In some embodiments, each target molecule-binding agent, e.g., antibody, further comprises a universal primer sequence at the 3' end for a first round extension, e.g., a universal round 1 primer sequence. In some embodiments, the universal primer sequence comprises about 9 to 20 nucleotides. In some embodiments, the universal primer sequence is a 10-mer. In some embodiments, the universal primer sequence is a 15-mer.

[0062] In some embodiments, the methods described herein further comprises:

(b) dividing the exosomes, cells or vesicles into at least two primary aliquots, the at least two primary aliquots comprising at least a first primary aliquot and a second primary aliquot;

(c) adding primary nucleic acid tags to the target molecule-binding agents in the at least two primary aliquots, wherein the primary nucleic acid tags added to the target molecule-binding agents in any one of the at least two primary aliquots are different from the primary nucleic acid tags added to the target molecule-binding agents in any one of the other primary aliquots;

(d) combining the at least two primary aliquots;

(e) dividing the combined primary aliquots into at least two secondary aliquots, the at least two secondary aliquots comprising a first secondary aliquot and a second secondary aliquot; and

(f) adding secondary nucleic acid tags to the at least two secondary aliquots, wherein the secondary nucleic acid tags added to the target molecule-binding agents in the first secondary aliquot are different from the secondary nucleic acid tags added to the target molecule-binding agents in the second secondary aliquot.

[0063] In some embodiments, the method further comprises step (g), which comprises repeating steps (d), (e), and (f) with the at least two secondary aliquots a number of times sufficient to generate a unique series of nucleic acid tags for each exosome, cell or vesicle in the sample.

[0064] In some embodiments, the methods described above further comprise amplifying and sequencing the nucleic acid barcodes, the UMIs and the nucleic acid tags. In some embodiments, next generation sequencing is used to sequence the barcodes, UMIs and/or nucleic acid tags. In some embodiments, split-pool sequencing is used to sequence the barcodes, UMIs and/or nucleic acid tags.

[0065] In some embodiments, the target molecules are proteins, sugar moieties, lipids, and/or polynucleotides.

[0066] In some embodiments, the detecting is at a single cell or a single exosome, cell or vesicle level. In some embodiments, each of the at least two primary aliquots consists of a single cell, a single exosome, or a single vesicle. In some embodiments, the at least two primary aliquots is about 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900 or 1000 or more aliquots.

[0067] In some embodiments, the nucleic acid tags are added by ligation reactions, polymerase extension reactions, and/or chemical syntheses.

Ligation Reactions

[0068] In some embodiments, the primary nucleic acid tags and/or the secondary nucleic acid tags and/or subsequent nucleic acid tags are added by ligation reaction(s). Because DNA barcodes are conjugated on their 5' ends to their respective target molecule-binding agents (e.g., antibodies), barcode addition must occur on the 3' end of the antibody-bound DNA barcode. In order to use split-pool sequencing (Rosenberg et al. (2018) Science, 360; 176-182, the contents of which are incorporated herein by reference), the ligation reaction was adapted to accommodate 3' end ligations (see, e.g., **Figs. 1A-C**).

[0069] In such a ligation method, one or more barcodes (*i.e.*, nucleic acid tags) are added to the 3' end of the target molecule-binding agent barcode (e.g., antibody barcode) through at least one splint-ligation reaction.

[0070] As a non-limiting example, **Figs. 1A-C** illustrate an exemplary ligation reaction method of the disclosure using an antibody. In some embodiments, a pool of exosomes, cells or vesicles to which various DNA-conjugated antibodies have bound is split into multiple wells. In some embodiments, a first-round oligonucleotide and a first-round splint sequence are added to each well containing a single exosome, a single cell or a single vesicle, bound to the target molecule-binding agent barcode, and a first-round splint ligation reaction is carried out. In some embodiments, the target molecule-binding agent barcode further comprises a universal round 1 sequence.

[0071] In some embodiments, the first-round oligonucleotide comprises a 5' common region, e.g., a universal round 1 sequence, followed by a unique barcode sequence (primary nucleic acid tag) terminated by a 3' region common to all first round oligonucleotides (e.g., a universal round 2 sequence). In some embodiments, the 5' common region and the 3'

universal round 2 sequence are about 15 bases in length.

[0072] In some embodiments, the first-round splint sequence comprises a region complementary to the universal round 1 sequence at 3' end of the target molecule-binding agent barcode and a region complementary to the 5' end of the first-round oligonucleotide in order to facilitate double stranded DNA ligation (see **Figs. 1A and B**).

[0073] In some embodiments, at the end of the first-round splint ligation reaction, a blocking strand is added to prevent inappropriate subsequent ligations. In some embodiments, the first-round blocking strand sequence is complementary to the first-round splint sequence.

[0074] In some embodiments, following the first-round splint ligation reaction, a second round splint ligation process is carried out. In some embodiments, the exosomes, cells or vesicles are re-pooled and split into second round wells. In some embodiments, a second round oligonucleotide and a second round splint sequence are then added to each well.

[0075] In some embodiments, the second round oligonucleotide comprises a common 5' region (e.g., a universal round 2 sequence), a second round barcode (secondary nucleic acid tag), and a common 3' region (e.g., a universal round 3 sequence or a universal PCR sequence). In some embodiments, the 5' common region and the 3' common region are about 15-bases in length.

[0076] In some embodiments, the second round splint sequence comprises a region complementary to the 3' universal round 2 sequence of the first-round oligonucleotide, as well as a region complementary to the 5' common region of the second round oligonucleotide (see **Fig. 1C**).

[0077] In some embodiments, after ligation, a second round blocking strand that is complementary to the second round splint sequence is added.

[0078] One or more subsequent ligation reactions can be repeated for as many rounds as necessary to generate necessary barcode diversity.

[0079] Before final addition by PCR, the reaction can be terminated by the addition of a buffer containing EDTA. The final round barcode can be added by PCR (see **Fig. 1C**).

Polymerase extension reactions

[0080] In some embodiments, the primary nucleic acid tags and/or the secondary nucleic acid tags and/or subsequent nucleic acid tags are added by polymerase extension reactions (PER). PER is a strand-displacing DNA polymerase-based barcoding approach (generally described in, for example, Kishi et al., *Nature Chemistry*, November 6, 2017, DOI: 10.1038, the

contents of which are hereby incorporated herein by reference). In PER, a DNA oligonucleotide bound to a target of interest, *e.g.*, antibody, is extended by a strand displacing polymerase by providing a DNA hairpin that has complementarity to the 3' end of the oligonucleotide conjugated to targets of interest (**Figs. 2A and B**).

[0081] As a non-limiting example, **Figs. 2-4** illustrate the PER methods of the disclosure using an antibody. However, any target of interest that can have a DNA oligonucleotide associated with it can be used in the methods of the disclosure, including, but not limited to, a cell, organelle, microvesicle, DARPin, fynomer, nanobody, or monobody.

[0082] In some embodiments, during the PER reaction, the oligonucleotide associated with the target of interest, e.g., the antibody, is extended, and during this extension it obtains a unique "first round barcode" or "primary nucleic acid" (**Fig. 2C**). At the reaction temperatures used to perform PER, the complex between the hairpin and the DNA oligonucleotide is susceptible to branch migration, a process in which the displaced hairpin flap competes and causes the extended oligonucleotide to disassociate from the hairpin, thus freeing the hairpin so that it can be used in subsequent rounds of PER (**Fig. 2E**). As the outcome of PER is the addition of a unique sequence to the DNA oligonucleotide attached to a target of interest (**Fig. 2D**), this principle can be used as a method of barcoding, as an alternative to a ligation-based approach.

[0083] In some embodiments, to enable PER for use in for antibody barcoding, a pool of exosomes, cells or vesicles to which various DNA-conjugated antibodies have bound is split into multiple wells. Each oligonucleotide in the same well is given a first round hairpin that appends a unique "well-specific first round barcode" sequence along with a common "universal round 2 landing sequence" to all the oligonucleotides in the well. After this first round of labeling, the hairpins within each of the wells are disabled using, for example, an exonuclease or by treating the hairpins with enzymes that will remove a unique base present within each of the hairpin separating the landing pad sequence from the rest of the hairpin (**Fig. 2F**). By disabling the first round hairpins, inappropriate hairpin priming between wells when they are pooled in subsequent rounds is prevented.

[0084] Following the first round of barcoding and hairpin destruction, all wells are pooled together, mixed, and then aliquoted into a fresh set of wells. The same process of hairpin-based extension is then repeated using a set of second round hairpins, with the process repeated as many times as desired to give the amount of barcode diversity needed for the user's purposes.

[0085] After the final round of PER, the universal landing sequence that is added can be used

as a PCR binding site to add further library diversity and also prepare the library for next generation sequencing.

[0086] Accordingly, in some embodiments, the present disclosure provides methods wherein the nucleic acid barcode bound to each target molecule-binding agent, e.g., antibody, is extended with one of the primary nucleic acid tags by contacting the exosomes, cells or vesicles with a strand displacing polymerase and a first DNA hairpin.

[0087] In some embodiments, the first DNA hairpin comprises

(i) a first oligonucleotide comprising a sequence complementary to a universal round 1 primer sequence, and

(ii) a second oligonucleotide, wherein the second oligonucleotide comprises a third oligonucleotide comprising a sequence comprising a primary nucleic acid tag, and a fourth oligonucleotide comprising a sequence complementary to the primary nucleic acid tag;

wherein the third oligonucleotide is located at the 5' end of the second oligonucleotide, and the fourth oligonucleotide is located at the 3' end of the second oligonucleotide; and

wherein the first oligonucleotide is fused to the 3' end of the fourth oligonucleotide.

[0088] In some embodiments, each primary nucleic acid tag comprises a unique well-specific first round barcode sequence at the 5' end and a universal round 2 primer sequence at the 3' end.

[0089] In some embodiments, the second oligonucleotide further comprises a connecting sequence between the third and fourth oligonucleotides. In some embodiments, the connecting sequence comprises the nucleic acid sequence GGGCCTTTTGGCCC (SEQ ID NO: 1).

[0090] In some embodiments, the first DNA hairpin further comprises a polyA sequence at the 3' end. In some embodiments, the polyA nucleic acid sequence is AAAAAAAA.

[0091] An exemplary hairpin DNA sequence that can be used in the methods of the disclosure is as follows:

ATACTTATTCTT<u>ACTAAATTCA</u>GGGCCTTTTGGCCCTGAATTTAGTAAGAA TA AGTA/ideoxyU/TTGCTAGGACAAAAAAAA (SEQ ID NO: 2),

wherein ATACTTATTCTT (SEQ ID NO: 3), as a third oligonucleotide as mentioned above, is the well index,

TTGCTAGGAC (SEQ ID NO: 4) is the landing pad sequence for the current round, which is complementary to the universal round 1 primer,

ACTAAATTCA (SEQ ID NO: 5) is the common landing pad sequence for the next round,

GGGCCTTTTGGCCC (SEQ ID NO: 1) serves as the connecting sequence, and

TGAATTTAGTAAGAATAAGTA/ideoxyU/ (SEQ ID NO: 6) is complementary to the landing pad sequence for the next round and the well index.

The poly-A tail is used to stop priming and extension in the wrong direction. (See, for example, Kishi et al., and **Fig. 2B**).

[0092] In some embodiments, the first DNA hairpin is disabled at the end of each polymerase extension. In some embodiments, the hairpin can be disabled using, for example, an exonuclease or by treating the first DNA hairpin with enzymes that will remove a unique base present at the junction of the first oligonucleotide and the fourth oligonucleotide. In some embodiments, the exonuclease is, for example, but not limited to, lambda exonuclease or exonuclease VIII.

[0093] In some embodiments, the unique base is ideoxyU. In some embodiments, an ideoxyU is used to replace T at the 3' end of the fourth oligonucleotide.

[0094] In some embodiments, the secondary nucleic acid tags are added by polymerase extension reaction, wherein the nucleic acid barcode bound to each target molecule-binding agent is further extended with one of the secondary nucleic acid tags by a strand displacing polymerase by providing a second DNA hairpin that comprises: (i) a fifth oligonucleotide, having sequence complementarity to the universal round 2 primer sequence, and (ii) a sixth oligonucleotide, wherein the sixth oligonucleotide further comprises a seventh oligonucleotide comprising a sequence encoding a secondary nucleic acid tag, and an eighth oligonucleotide comprising a sequence encoding a sequence complementary to the seventh oligonucleotide; wherein the seventh oligonucleotide is located at the 5' end of the sixth oligonucleotide, and the eighth oligonucleotide is located at the 3' end of the sixth oligonucleotide; and wherein the fifth oligonucleotide is fused to the 3' end of the eighth oligonucleotide.

[0095] In some embodiments, each secondary nucleic acid tag comprises a unique well-specific second round barcode sequence at the 5' end and a universal round 3 primer sequence at the 3' end. In some embodiments, the sixth oligonucleotide further comprises a connecting sequence between the seventh and eighth oligonucleotides. In some embodiments, the connecting sequence is GGGCCTTTTGGCCC (SEQ ID NO: 1).

[0096] In some embodiments, the second DNA hairpin further comprises a polyA sequence at the 3'-end. In some embodiments, the polyA sequence is AAAAAAAA.

[0097] In some embodiments, the second DNA hairpin is disabled at the end of the second extension of the nucleic acid barcode. In some embodiments, the hairpin can be disabled using, for example, an exonuclease or by treating the second DNA hairpin with enzymes that will remove a unique base present at the junction of the fifth oligonucleotide and the eighth oligonucleotide. In some embodiments, the exonuclease is lambda exonuclease or exonuclease VIII. In some embodiments, the unique base is ideoxyU. In some embodiments, an ideoxyU is used to replace T at 3' end of the eighth oligonucleotide.

[0098] In some embodiments, a nucleic acid tag described herein above comprises a unique barcode and a common 3' region, e.g., a universal round 1 primer sequence, a universal round 2 primer sequence, a universal round 3 primer sequence or a universal subsequent round primer sequence. The nucleic acid tag is exemplified by the third oligonucleotide and the seventh oligonucleotide described herein above.

[0099] In some embodiments, different aliquots described herein above in each round are separately placed in different wells.

[00100] In some embodiments, the methods of the disclosure further comprise: step (g), which comprises repeating steps (d), (e), and (f), as set forth above, with the at least two secondary aliquots. In some embodiments, step (g) is repeated any number of times sufficient to generate a unique series of nucleic acid tags for each exosome, cell or vesicle in the sample. In some embodiments, step (g) is repeated any number of times sufficient to generate multiple unique nucleic acid tag for each exosome, cell or vesicle in a sample.

[00101] In some embodiments, the number of times is about 20 to 100 times the number of the exosomes, cells or vesicles present in the sample.

DNA-barcoded antibodies (DBAs)

[00102] In some embodiments, the present methods use DNA barcoded antibodies (DBAs) against target proteins and UMIs, in combination with split-pool sequencing, to develop a novel single entity, e.g., vesicle, cell or exosome, phenotyping/proteomics platform, which can be applied to vesicle, cell or exosome profiling within patients. In some embodiments, the present method requires use of at least a target molecule-binding agent, *e.g.,* antibody, with a conjugated DNA oligo. Each oligo contains an antibody specific barcode and a UMI. Each antibody binds specifically to its cellular target. During split-pool sequencing, all of the oligos associated with a given cell are identically labeled, yet no two cells apply the same tag to their oligos. During analysis, all oligos associated with a given cell are identified via their tag, and the composition and quantity of all bound antibodies is

determined by analyzing the BC and UMI, respectively. See **Figs. 5A-D**. For simplicity, only antibodies binding to cell surface proteins are shown in **Figs. 5A-D**.

[00103] DBAs can comprise a library of >50 DBAs against extracellular and intracellular proteins. The target proteins comprise markers of immune cell identity or function (**Fig. 5A**). The specificities of the DBAs are determined by applying them to cell lines with previously defined markers (e.g., EBV immortalized B-cells, Sup-T1 cells) (**Fig. 5B**). The antibody staining, library preparation, DNA-sequencing and analysis pipelines such that antibodies that are directed against antigens present in a particular cell line show a 100-fold increase in abundance versus antibodies whose antigens are not expressed. Complex DBA libraries have been used for determination of protein contents within bulk cellular samples (e.g., fine needle aspirates from patients). The present method combines the validated DBAs with a split-pool sequencing method to specifically tag all the antibodies bound to a single cell with a common identifier. DNA-barcodes that have been specifically labeled with a unique cell tag are then PCR- amplified, sequenced, and analyzed, to determine the composition of antibodies bound to each cell (**Figs. 5C and D**). As compared to other single cell, vesicle or exosome sequencing strategies, the use of a split-pool labeling approach removes the need for in-house specialized equipment while also decreasing the costs and greatly expanding the number of cells which can be analyzed within a single experiment. Split-pool sequencing is used for identifying short DNA oligos conjugated to antibodies.

[00104] A catalogue of well characterized antibodies against hundreds of human targets exists. These antibodies are able to bind to their corresponding antigen with exquisite sensitivity and specificity. By adding DNA-barcodes onto previously developed antibody reagents, the methods of the disclosure take advantage of the decades of effort and validation that have gone into creating high quality antibody reagents (Edfors, F., et al., Nature Communications. 2018; 8: 4130, the contents of which are hereby incorporated herein by reference). Adding DNA-barcodes to antibodies does not affect their target binding efficiency or specificity (Ullal, A.V., et al., Science Translational Medicine. 2014; 6(219): 219ra9, the contents of which are hereby incorporated herein by reference). By selectively conjugating defined DNA-barcodes to specific antibodies, an antibody binding to its antigen on a single cell can be detected by sequencing for the presence of the DNA-barcode attached to the antibody. In addition, UMIs have been shown to further improve the ability to precisely quantify the prevalence of a given element (e.g., a DNA-barcode) within a sample. This is possible because while all of the antibodies against a single antigen will share the same

DNA-barcode, each antibody molecule will also have an additional sequence (the UMI) which is unique to an individual antibody molecule and thus enable a more precise quantitation of the number of antibodies bound to each cell. The combination of DNA-barcodes with UMIs drastically improves the information content of multiplex screens (Michlits, G., et al., Nature Methods. 2017; 14:1191-1197, the contents of which are hereby incorporated herein by reference).

[00105] DBAs are a versatile tool for sensitive detection of proteins in vitro. They are made by conjugating short DNA oligos – either single- or double-stranded – to conventional antibodies used for immunohistochemistry. Traditional methods of conjugation used the biotin-streptavidin system but covalent conjugation can also be used. DBAs can be used in experiments for multiplexed protein analysis (M. Morishita, et al., *Journal of Pharmaceutical Sciences,* 2017; A. V. Ullal, et al., *Science Translational Medicine,* 2014; M. Stoeckius, et al., *Nature Methods,* 2017; and V. M. Peterson, et al., *Nature Biotechnology,* 2017, 35(10):936-939, the contents of which are hereby incorporated herein by reference), immuno-PCR (T. Sano, et al., *Scientific Reports,* 2016, the contents of which are hereby incorporated herein by reference), and FISH imaging (S. S. Agasti, et al., *Chemical Science,* 2017, the contents of which are hereby incorporated herein by reference). Due to the small size of exosomes, and consequently the low amounts of protein in a single exosome, sensitivity in detection is a critical consideration for the methods of the disclosure. DBAs fulfill this requirement because they allow antibodies to be counted by amplifying their signal using PCR and detected via next generation sequencing (NGS). DBAs also provide higher multiplexing capacity because they are not limited by overlapping emission/absorption spectra, which limit fluorescence based methods, allowing to probe dozens of proteins in parallel.

[00106] To produce DBAs, thiolated single-stranded oligos are used and conjugated to amines on lysine residues in the antibody via $SM(PEG)_2$ crosslinkers (S. S. Agasti, et al., *Chemical Science,* 2017, the contents of which are hereby incorporated herein by reference). Optimization will ensure an average of 1-2 oligos are conjugated to each antibody to minimize potential steric interference from the oligo on antigen binding and to ensure interpretable correlation between sequenced oligos and antibodies bound. To show that binding capacity of the antibody is unhindered by any oligo conjugation, flow cytometry results on cells incubated with oligo-coupled antibodies as primary antibodies are compared those on cells incubated with control primary antibodies. Specifically, the primary antibodies are first incubated with the cells expressing the target protein, further incubated with a secondary fluorophore-conjugated antibody against the primary antibodies, and quantified via

flow cytometry.

[00107] Additionally, to ensure consistent stoichiometric conjugation between different antibodies, a control PEG molecule can be used to be detected on an SDS-PAGE gel to visualize the average number of conjugants (J. A. G. L. van Buggenum, et al., *Scientific Reports,* 2016, the contents of which are hereby incorporated herein by reference). The DBAs are incubated with protein A coated beads and bind to protein A coated beads with high affinity and specificity. Antibody-coated beads are then taken through the split-pool protocol. In some embodiments, the linkers between the oligo and the antibody are uncleavable. Alternatively, in other embodiments, the linkers between the oligo and the antibody are cleavable.

Split-Pool Sequencing

[00108] The methods of the disclosure can be scaled to any number of protein targets, and take advantage of several disparate technologies, i.e., DNA-barcoded antibodies, unique molecular identifiers, and split-pool sequencing.

[00109] Split-pool sequencing is a method of sequencing which involves serially ligating small tags onto DNA molecules which enables users to label all DNA molecules associated with a single entity, e.g., cell, vesicle or exosome, with a unique barcode (see, for example, Rosenberg, A., et al., Science. 2018; 360:176-182, the contents of which are hereby incorporated herein by reference).

[00110] Split-pool sequencing dramatically reduces the cost of single-cell sequencing (Rosenberg AB, et al., Science. 2018 Apr 13; 360(6385): pp.176-182, the contents of which are hereby incorporated herein by reference). The combination of DNA-barcodes with UMIs drastically improves the information content of multiplex screens. The methods of the disclosure describe a cost-effective, modular, and easy-to-use quantitative single entity protein profiling platform. For example, the present methods use DBAs and UMIs in combination with split-pool sequencing to improve information content and to reduce cost. The methods of the disclosure enable multiplexed and highly scalable detection of proteins at the single-entity level, but do not require in-house specialized equipment or specialized training.

[00111] The present disclosure also relates, at least in part, to a method of directly quantifying the abundance and state (e.g., phosphorylation state) of over a hundred proteins on tens of thousands of individual cells, exosomes, or vesicles, through the creation of a

single-entity profiling platform which functions independently of specialized equipment. The methods of the disclosure enable comprehensive cellular phenotyping within a variety of diseases, and have innumerable translational applications in early stage discovery pipelines or in helping guide clinical decision making.

[00112] Other potential uses exist in the area of cellular therapies to better characterize the composition of modified cellular products or to delineate the properties of those cells that are most tumor invasive/effective.

[00113] Split-pool sequencing can be used to analyze over a hundred thousand cells, exosomes, or vesicles within a single experiment at a fraction of the cost of other single-entity sequencing methods which all require specialized equipment and proprietary reagents (e.g., Fluidigm C1, 1CellBio inDrop, 10x Genomics Chromium).

[00114] Computer program products can analyze data generated by any sequencing instrument. Non-limiting examples of sequencers include: a) DNA sequencers produced by Illumina™, for example, HiSeg™, HiScanSQ™, Genome Analyzer GAIIX™, and MiSeg™ models; b) DNA sequencers produced by Life Technologies™, for example, DNA sequencers under the AB Applied Biosystems™ and/or Ion Torrent™ brands; c) DNA sequencers manufactured by Beckman Coulter™; d) DNA sequencers manufactured by 454 Life Sciences™; and e) DNA sequencers manufactured by Pacific Biosciences™.

[00115] While some exemplary methods for sequencing are provided herein, these are exemplary and not meant to limit the scope of the present disclosure. Additional suitable methods for sequencing will be apparent to those of skill in the art based on the present disclosure in view of the knowledge in the art. Additional methods of sequencing are described in US Application No. 20150066385; Quail et al., 2012, BMC Genomics. 13 (1): 34; Liu et al., 2012, Journal of Biomedicine and Biotechnology. 2012: 1–11; each of which is incorporated herein by reference in its entirety.

Exosomes

[00116] In some embodiments, exosomes are analyzed using the methods of the disclosure. By "exosome" is meant any cell-derived, extracellular vesicle composed of a membrane enclosing an internal space, wherein the vesicle is generated from a cell by fusion of the late endosome with the plasma membrane or by direct plasma membrane budding, and wherein the vesicle has a longest dimension, such as a longest cross-sectional dimension, such as a cross-sectional diameter, ranging for example, from 10 nm to 150 nm, such as 20

nm to 150 nm, such as 20 nm to 130 nm, such as 20 nm to 120 nm, such as 20 to 100 nm, such as 40 to 130 nm, such as 30 to 150 nm, such as 40 to 150 nm, or from 30 nm to 200 nm, such as 30 to 100 nm , such as 30 nm to 150 nm, such as 40 nm to 120 nm, such as 40 to 150 nm, such as 40 to 200 nm, such as 50 to 150 nm, such as 50 to 200 nm, such as 50 to 100 nm, or from 10 to 400 nm, such as 10 to 250 nm, such as 50 to 250 nm, such as 100 to 250 nm, such as 200 to 250 nm, such as 10 to 300 nm, such as 50 to 400 nm, such as 100 to 400 nm, such as 200 to 400 nm, each range inclusive. As used herein, "inclusive" refers to a provided range including each of the listed numbers. Unless noted otherwise herein, all provided ranges are inclusive.

[00117] An exosome is typically created intracellularly when a segment of the cell membrane spontaneously invaginates and is ultimately exocytosed. As used herein, exosomes can also include any shed membrane bound particle that is derived from either the plasma membrane or an internal membrane. Exosomes can also include cell-derived structures bounded by a lipid bilayer membrane arising from both herniated evagination (blebbing) separation and sealing of portions of the plasma membrane or from the export of any intracellular membrane-bounded vesicular structure containing various membrane-associated proteins, including surface-bound molecules derived from the host circulation that bind selectively to the exosomal proteins together with molecules contained in the exosome lumen, including but not limited to mRNAs, microRNAs or intracellular proteins. Blebs and blebbing are further described in Charras et al, Nature Reviews Molecular and Cell Biology, Vol. 9, No. 11, p. 730-736 (2008). Exosomes can also include membrane fragments.

[00118] In some embodiments, size exclusion chromatography, such as gel permeation columns, centrifugation or density gradient centrifugation, and filtration methods can be used to isolate exosomes from a sample prior to use in the methods of the disclosure. Differential centrifugation, anion exchange and/or gel permeation chromatography, sucrose density gradients, organelle electrophoresis, magnetic activated cell sorting (MACS), or with a nanomembrane ultrafiltration concentrator can also be used to isolate exosomes. In a preferred embodiments, the purification method is a series of ultracentrifugation steps (Kowal, E. J. K., Ter-Ovanesyan, D., Regev, A. & Church, G. M. Extracellular Vesicle Isolation and Analysis by Western Blotting. *Methods Mol. Biol. Clifton NJ 2017,* **1660:** 143-152). Known methods for isolation of exosomes also include those disclosed in, for example, Monguio-Tortajada, et al., Cellular and Molecular Life Sciences (2019) 76:2369-2382, the contents of which are hereby incorporated by reference.

[00119] Highly abundant proteins, such as albumin and immunoglobulins, may hinder

isolation of exosomes from a biological sample. Therefore, exosomes can be isolated by a system that utilizes multiple antibodies that are specific to the most abundant proteins found in blood. Such a system can remove up to several proteins at once, thus unveiling the lower abundance species such as cell-of-origin specific exosomes. Other known methods for exosome isolation include high abundant protein removal methods as described in Chromy et al. J. Proteome Res 2004; 3: 1120-1127. In another embodiment, the isolation of exosomes from a sample may also be enhanced by removing serum proteins using glycopeptide capture as described in Zhang et al, Mol Cell Proteomics 2005; 4: 144-155.

[00120] A detailed understanding of a patient's disease and response to treatment on a molecular level can help elucidate mechanisms of actions and inform downstream treatment follow-ups. The present disclosure includes methods of diagnosing disease, monitoring disease progression and monitoring response to treatment comprising non-invasive and unbiased tissue health surveillance by using already existing particles found in bodily fluids such as, for example, extracellular vesicles, including exosomes. To provide real-time, quantitative monitoring of cellular response through examining the patient's proteome, the present method is based on a new platform for single entity, e.g., cell, exosome or vesicle, proteomic analysis. The present disclosure relates, at least in part, to a method for split-pool sequencing coupled with DBAs in order to obtain a tissue specific surface proteome. The tissue specific surface proteome is an indicator of cellular homeostasis and can be used to monitor disease progression and treatment in a quantitative and sensitive fashion.

[00121] Exosomes are shed from all cells tested to date into bodily fluids which can be used to track changes in the proteome of all tissues within an organism at any given point in time. Specifically, exosomes are small extracellular vesicles secreted by many cell types (M. Li, et al., *Philosophical transactions of the Royal Society of London. Series B, Biological sciences,* 2014; and A. Vlassov, et al., *Biochim. Biophys. Acta,* 2012, the contents of which are hereby incorporated herein by reference) and found in all bodily fluids including urine (M. Li, et al., *Philosophical transactions of the Royal Society of London. Series B, Biological sciences,* 2014, the contents of which are hereby incorporated herein by reference), blood (M. Li, et al., *Philosophical transactions of the Royal Society of London. Series B, Biological sciences,* 2014, the contents of which are hereby incorporated herein by reference), cerebrospinal fluid (D. Chiasserini, et al., *Journal of Proteomics,* 2014, the contents of which are hereby incorporated herein by reference), and amniotic fluid (A. Asea, et al., *Journal of Reproductive Immunology,* 2008, the contents of which are hereby incorporated herein by reference). They are characterized by several highly expressed surface transmembrane

markers including CD9, CD81, and CD63 (M. Simons, et al., *Current Opinion in Cell Biology,* 2009, the contents of which are hereby incorporated herein by reference). Exosome biogenesis follows the endosomal pathway. Endosomes are intracellular compartments that have an inverted membrane composition from that of the cellular plasma membrane. Molecules from the plasma membrane are sorted in the endosome for recycling or degradation in the lysosome mediated by an ESCRT-dependent pathway. In order for the release of exosomes to occur, the endosomal compartments must bud into the lumen of the endosome turning it into a multivesicular body (MVB). If the MVB fuses with the lysosome, the contents are degraded. However, if the MVB fuses with the plasma membrane, the vesicles are then released to the extracellular space. The secreted vesicles are known as exosomes. Due to their means of biogenesis it is not surprising that several studies have shown that exosome surfaces are rich in membrane proteins reflective of the cell.

[00122] Exosomes have been shown to contain cellular components such lipids, proteins, RNA, and DNA. Exosomes are important for cell signaling; intact mRNA transported from one cell via exosome to another has been shown to be able to be used in protein translation (H. Valadi, et al., *Nature Cell Biology,* 2007, the contents of which are hereby incorporated herein by reference). Network based gene analysis has shown that these mRNA transcripts are enriched for such functions as cellular development, protein synthesis, RNA post-transcriptional modification. Exosomes also may play an important role in immune response. Exosomes secreted by tumor cells can be captured by dendritic cells and used to present antigens to activate the immune system (O. G. de Jong, et al., *Journal of Extracellular Vesicles,* 2012; Q. Lian, et al., *Cell Research,* 2017; and F. M. Barros, et al., *Frontiers in Immunology,* 2018, the contents of each of which are hereby incorporated herein by reference). Concurrently, other small molecules, like PD-L1 which inhibit T-cell response, have been shown to be encapsulated in exosomes which dampen immune response (F. M. Barros, et al., *Frontiers in Immunology,* 2018; and Y. Yang, et al., *Cell Research,* 2018, the contents of each of which are hereby incorporated herein by reference). In addition, studies have also shown that exosomes serve as a method of secretion for harmful cytoplasmic DNAs (A. Takahashi, et al., *Nature Communications,* 2016, the contents of which are hereby incorporated herein by reference).

[00123] Because of the ubiquitous property that nearly all cells shed exosomes into extracellular spaces, they present a versatile tool for interrogating a diverse panel of tissues simultaneously. Taking advantage of the fact that the plasma membrane proteins of exosomes are representative of the membrane proteins of the parent cell, monitoring proteome

alterations of exosomes are a convenient, non-invasive way to monitor cell surface proteome alterations of tissues. Exosomes also have the ability to cross the blood-brain barrier, carrying with them information about brain tissue health, a property now being explored for diagnosing brain diseases from exosome analysis (K. M. Kanninen, et al., *Biochimica et Biophysica Acta - Molecular Basis of Disease,* 2016, the contents of which are hereby incorporated herein by reference).

[00124] Exosome diagnostics in the clinical setting have largely relied on identifying specific markers for disease from a bulk population of purified particles. The most obvious limitation with this strategy is its lack of sensitivity. A more recent technology has demonstrated that exosomes and other secreted vesicles have the potential to be powerful tools in monitoring cancer therapy response (H. Shao, et al., *Nature Medicine,* 2012, the contents of which are hereby incorporated herein by reference). Expanding upon this result, exosomes are probed for their tissue specific proteome in order to diagnose and monitor the progression of disease. In some embodiments, the present disclosure provides methods for determining the proteomic profile of individual exosomes are obtained by combining single entity sequencing technologies and DBAs.

[00125] The natural exosome biogenesis incorporates cellular material, especially membrane proteins from the parent cell, and is a reflection of the biological status of the cell of origin (H. Shao, et al., *Nature Medicine,* 2012; K. L. Schey, et al., *Methods,* 2015; H. Im, et al., *Nature Biotechnology,* 2013; O. G. de Jong, et al., *Journal of Extracellular Vesicles,* 2012; and B. Gyorgy, et al., *Cellular and Molecular Life Sciences,* 2011. F., the contents of each of which are hereby incorporated herein by reference). In some embodiments, the methods of the invention provide for sequencing the antibodies bound to single exosomes to detect changes in the exosome surface proteome in order to quantitatively monitor changes in cell state from multiple organs in parallel. However, because exosomes are 100 times smaller in diameter than the average mammalian cell, it is anticipated that making definitive conclusions about cell state from the sparse data obtained from a single exosomes may be challenging in some instances. Thus, in some embodiments, pooling exosomes with the same tissue-specific markers, allows for generation of a collective surface proteome for the tissue of interest. The methods of the present disclosure are a more time-sensitive diagnostic tool that are superior to current serum-based diagnostic standards. Furthermore, the methods of the disclosure can be used for studying any pathology of interest.

Diagnostic and Therapeutic Methods

[00126] In some aspects, the methods and systems of the disclosure are applied to discovery pipelines to identify pathways of interest (e.g., in a preclinical setting) and in a clinical setting for diagnostic, monitoring, and/or therapeutic applications for a variety of diseases and disorders. Thus, in some embodiments, the present methods of the disclosure are used to analyze samples obtained from patients with disease (e.g., cancer, organ transplant, and autoimmune disease). In some embodiments, the samples are compared to control samples, e.g., those obtained from healthy subjects or other control samples, to diagnose disease or monitor response to therapy.

[00127] In some aspects, the present disclosure is directed to methods for diagnosing a disease or disorder, e.g., cancer, autoimmune disease or inflammatory disease, determining a patient's response to a therapy, and/or monitoring the patient for undesired side effects to the therapy, the method comprising: (a) contacting exosomes, cells or vesicles with a plurality of target molecule-binding agents, wherein each target molecule-binding agent comprises a nucleic acid barcode and optionally a unique molecular identifier (UMI), wherein target molecule-binding agents that are specific to an identical target molecule share an identical nucleic acid barcode, wherein the exosomes, cells or vesicles are isolated from blood of a patient suffering from a disease or disorder, e.g., cancer, autoimmune disease or inflammatory disease, before or after the patient has been treated with a therapy, wherein the plurality of target molecules comprise molecules that are markers indicative of a disease or disorder, e.g., cancer, autoimmune disease or inflammatory disease, efficacy of the therapy, and/or are indicative of undesirable side effects, and wherein expression levels of the target molecules determined in the patient before or after the therapy are compared with expression levels of the corresponding target molecules determined in normal controls, thereby diagnosing a disease or disorder, e.g., cancer, autoimmune disease or inflammatory disease, determining the patient's response to a therapy, and/or monitoring for undesired side effects of a therapy.

[00128] In some embodiments, the patient's response to therapy is determined and/or monitored in real-time.

[00129] In some embodiments, the patient is further treated with the same therapy, if the therapy is effective, but associated with little or no undesired side effects; or the patient is treated with a different therapy if the therapy is not effective and/or is associated with undesired side effects.

Real-Time Cancer Diagnostics

[00130] Chemotherapy is a pillar of the clinical oncology arsenal. Each year over half a million Americans are treated with chemotherapy, a collection of toxic small molecules and proteins designed to selectively destroy tumor cells (DeVita V. T. Chu, EA history of cancer chemotherapy. which are hereby incorporated herein by reference). For all patients, the choice of therapy is largely based on historical experience and tumor molecular profile (Santos, F. N., de Castria, T. B., Cruz, M. R. S. & Riera, R. *Cochrane Database Syst. Rev.* CD010463, 2015; and Ahmadzada, T. *et al.* J. *Clin. Med. 2018,* **7**, the contents of each of which are hereby incorporated herein by reference). Given the sizable genetic, expression and cellular heterogeneity between patients with the same tumor classification, it is not surprising that patients treated with the identical chemotherapeutic regimen can have vastly different responses. Despite recent advances in precision medicine, patients still regularly fail to respond to treatment, and nearly all patients undergoing chemotherapy suffer from the undesirable "side-effects" ranging from extreme nausea and diarrhea to more severe complications such as heart failure or life-threatening bacterial infections (Pearce, A. et *al.*, *PloS One* 12, e0184360; Thatcher, N. First- and second-line treatment of advanced metastatic non-small-cell lung cancer: a global view. *BMC Proc. 2008,* 2 Suppl 2: S3; Cardoso, F. et al. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol. 2002,* 13: 197-207; and Catalano, V. et al. *Br. J. Cancer 2008,* 99: 1402-1407). In fact, it is not uncommon for a course of chemotherapy to show little to no effect on tumor growth, yet during the treatment, the patient is still forced to suffer from undesired side-effects (Temel, J. S. *et al. N. Engl. J. Med. 2010,* 363: 733-742; and Harrington, S. E. & Smith, T. J. *JAMA 2008,* 299: 2667-2678, the contents of each of which are hereby incorporated herein by reference). Given the sizable cost of therapy and the life-threatening complications that can ensue, it is of critical importance to maximize the benefits that patients receive from chemotherapy (Pritzker, K. P. H. *Expert Rev. Mol. Diagn. 2015,* 15: 971-974, the contents of which are hereby incorporated herein by reference).

[00131] The present disclosure describes a diagnostic system that reports a patient's response to therapy in real-time, while also monitoring for undesired side-effects. The present disclosure describes a method of SESeq that can be applied to real-time monitoring of patients' response to therapy.

[00132] Exosomes, as small extracellular vesicles containing RNA, are secreted into the circulation by all cells, including tumor cells (Momen-Heravi, F., Getting, S. J. & Moschos, S. A. *Pharmacol. Ther.* 2018, the contents of which are hereby incorporated herein by reference). Blood contains a mixture of extracellular vesicles from various tissues, so in

order to gain insight into tumor and normal tissue response to chemotherapy, extracellular vesicles need to be sequenced individually. In some embodiments, single entity sequencing is used to selectively analyze the contents of individual extracellular vesicles isolated from blood in a cancer patient. The present disclosure describes a non-invasive technology to allow clinicians to dynamically tailor patient therapy to tumor response while also enabling them to catch undesired off-target effects before they become clinically evident.

[00133] Since extracellular vesicles are secreted by all cells into the circulation and contain RNA from their cell of origin, it is hypothesized that individually sequencing the contents of each extracellular vesicle will provide insights into the cellular processes occurring within both the tumor and normal tissues. The present disclosure includes methods and analysis pipelines for performing single entity sequencing, e.g., SESeq. In some embodiments, single entity sequencing, e.g., SESeq is used to determine the response of tumor and normal cells to chemotherapy within simulated patient samples. The present method is the a real-time method for individually analyzing the contents of extracellular vesicles and used this data to infer the cellular state of tumor and normal cells, thereby transforming cancer care by monitoring efficacy of chemotherapy against tumor while helping better detect and manage undesired side-effects.

[00134] The present disclosure is directed, at least in part, to a scalable method with high levels of sensitivity and scalability and requires combining split-pool sequencing and DBA technologies for quantifying the abundance of dozens of proteins on individual extracellular vesicles. Specifically, by coating single extracellular vesicles with a library of DBAs followed by sequencing all antibodies bound to each extracellular vesicle, a set of specific proteins in each extracellular vesicle can be identified. This information can be amalgamated to understand which tissue the extracellular vesicle came from and the state of that tissue at the time the extracellular vesicle was released. The platform's versatility lends itself to be a good tool for not only diagnostic purposes but also for downstream treatment monitoring or even off-target drug effect discovery. A single extracellular vesicle proteome sequencing platform allows a physician to gain dynamic biological insight into disease progression in a quantitative manner that surpasses current standards of disease monitoring.

[00135] In order to attain single extracellular vesicle resolution in sequencing, a split-pool barcoding is used as a sequencing approach. Split-pool barcoding relies on randomly distributing a mixture of cells each containing a series of in situ transcribed cDNA into one of 96 wells. Each well contains a unique DNA primer that is ligated onto all of the cDNA molecules associated with each cell. After the ligation, all wells are pooled and the process is

repeated several more times ultimately tagging all DNA molecules in a given cell with a unique barcode combination.

[00136] All publications, patents and patent applications cited in this specification are herein incorporated by reference in their entirety for all purposes as if each individual publication, patent, or patent application were specifically and individually indicated to be incorporated by reference for all purposes. The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the inventors described herein are not entitled to antedate such disclosure by virtue of prior disclosure or for any other reason.

[00137] The following Examples are set forth as being representative of the present disclosure. These examples are not to be construed as limiting the scope of the disclosure as these and other equivalent embodiments will be apparent in view of the present disclosure, figures, tables and accompanying claims.

EXAMPLES

Example 1: Sequencing the Surface Proteome of Single Cells Using DBAs

[00138] This example describes methods of sequencing proteins on cells. However, it is applicable to exosomes and vesicles as well.

[00139] Two cell types (human liver-HEPG2 and human glioblastoma (GBM)-LNZ308) and a panel of 6-8 antibodies specific to each of the two cell types are chosen. The panel of the antibodies are used to demonstrate that a single cell protein quantification method using split-pool sequencing. Liver tissue can be chosen through an analysis of the Human Protein Atlas database (M. Uhlen, et al., *Science,* 2015). Liver tissue is selected because it exhibits the highest number of tissue specific membrane proteins and because it plays roles in drug metabolism and toxin filtration. A human glioblastoma cell line is chosen because bulk glioblastoma exosomes have previously been shown to allow real time monitoring of GBM therapy (H. Shao, et al., *Nature Medicine,* 2012).

[00140] Antibodies to these proteins will be procured and validated for specificity by CRISPR knockout or knockdown of their targeted antigen and finally conjugated to oligos for single cell antibody protein profiling sequencing. Two cell types in a mixed tissue experiment are co-cultured and incubated with DBAs for labeling prior to taking them through the split-pool pipeline. In the two-tissue experiment, clean single cells are obtained through split-pool without particle aggregation. It will be confirmed whether the cell type of each of the single

cells is profiled with high accuracy. Specifically, it will be confirmed whether the single cell protein profiles show correlation of protein expression levels between cells of the same lines, providing an indication of the variance to be expected in these experiments.

[00141] As long-term passaging of immortalized cell lines tend to be quite homogeneous as compared to the native tissue. Within-cell comparisons will allow testing of the dynamic range of the methods of the disclosure. In order to test that the results derived from the aforementioned split-pool sequencing method are accurate and representative of single cell populations, the integrity and accuracy of the results is validated with a lower throughput single cell sequencing method based on the Chromium 10X.

Example 2: Sequencing the Proteome of Single Exosomes Using DBAs

[00142] Experiments on exosomes present additional challenges to whole cell experiments. Because exosomes are much smaller, the protein information is much more sparse, it is expected that the number individual exosomes to be processed will be much larger than in traditional single cells experiments-on the scale of tens to hundreds of thousands. Two exosome populations are mixed and are incubated with DBAs for labeling prior to being taken through the split-pool pipeline. It will be confirmed that the tissue origin of each of the exosome is profiled with high accuracy. Specifically, it well be confirmed if the surface proteome profile of individual exosomes correlates well with the surface proteomes of individual cells generated from whole cell profiling.

[00143] Studies have shown high degrees of correlation between cell surface proteome and bulk exosome surface proteome in GBM, colorectal, and B-cell leukemia cell lines (H. Shao, et al., *Nature Medicine,* 2012; and L. Belov, et al., *Journal of Extracellular Vesicles,* 2016). It is predicted that the methods of the disclosure will enable pooling of single exosomes identified from each cell type and reconstruction of two tissue specific proteomes that have good surface proteome correlations with their parent cell line. From single exosome data generated from split-pool, the antigen density on the surface of exosomes will be determined and the sensitivity detecting protein level changes for downstream applications will be determined. If it is confirmed that the specificity of tissue origin of exosomes can be readily called, a layer of complexity will be added by titrating a third exosomal population, which is obtained from neither glioblastoma nor liver.

[00144] The method developed in this Example will be used to establish a library of DBAs for single exosomes isolated from different tissues of origin and optimize an experimental pipeline for sequencing DBAs from single exosomes and correctly calling their

tissues of origin.

Example 3: Acetaminophen Toxicity Has a Detectable Phenotype in Cell Culture and Can Be Similarly Observed at The Bulk Exosome Level

[00145] HepG2 cells are cultured and treated with 10μM to 10mM of acetaminophen for 24-48 hours (I. Manov, et al., *Basic & Clinical Pharmacology & Toxicology,* 2004; and R. H. Pierce, et al., *Biochemical Pharmacology,* 2002). A bulk RNA sequencing analysis is done on both the drug-treated and untreated cell lines as well as exosomes harvested from the media of the drug-treated and untreated cells. In addition, mass spectrometry is performed on drug-treated and untreated HepG2 cells as well as their respective exosome populations to confirm similarity in their protein expression levels under drug stress. It will be confirmed whether these protein changes match already known results in literature a comprehensive list of hits that may be specific to exosomes will be generated.

[00146] It is predicted that some the differentially expressed genes after drug treatment will match hits found in the literature because HepG2 cells are one of the most widely used cell lines to study acetaminophen toxicity (J.-M. Prot, et al., "*Toxicology and Applied Pharmacology,* 2012; and R. M. Rodrigues, et al., "*Data in Brief,* 2016). Though many of the existing datasets that study acetominophen toxicity are from microarray experiments, it is possible to confirm these results with bulk RNA sequencing experiments. These results will also be confirmed by mass spectrometry. It is expected that the hits from mass spectrometry and those identified in RNA sequencing will show correlated modulation in response to drug toxicity. If, in fact, it is found that proteomic profiling results do not correlate with RNA sequencing, protein data generated from the mass spectroscopy will be used as markers of acetaminophen toxicity. From these pure, bulk samples, it is possible ascertain the signal to expect from exosome samples obtained from downstream animal experiments.

Example 4: Performing Single Exosome Sequencing on a Complex Fluid to Confirm the Sensitivity Limits

[00147] It will be shown that single exosomes of liver or brain origin can be detected. Blood is extracted from mice. Exosomes from blood will be isolated using the ultracentrifugation method and subsequently incubated with prepared DBAs. Samples will be taken through the split-pool sequencing pipeline and analysis.

[00148] It is predicted that the single exosome population detected in the mixed complex bodily fluid will have generally similar expression patterns to that of the pure

population isolated from cell culture supernatant. However, it is expected that immortalized cell lines do not exactly recapitulate the state of cells in vivo, therefore a first round of complex fluid analysis is performed with a spiked in population of cell culture supernatant exosomes that have been previously analyzed to confirm robustness of the assay. Because many of the antibodies to be used in identifying the tissue origin of the exosome will be vetted for specificity, there should be minimal binding to off-target tissues. Nevertheless, it is inevitable that there will be noise from non-specific antibody binding, ligation, or PCR amplification. It is predicted that signals coming from biological and technical artifacts will be relatively low and have a discernible profile compared to the true signal.

[00149] Based on this experiment, it is possible to find a pattern for the noise generated in the sequencing of complex fluids and apply some filtering rules to discriminate between signal and noise for downstream experiments. The complexity of the exosome sample can be reduced by enriching for particular exosomal fractions via antibody-based pre-enrichment before applying DBAs to the samples.

Example 5: Single Exosome Antibody Sequencing Is a Sensitive Readout of Pathology in an Animal Model of Disease

[00150] An animal model of acetaminophen toxicity is used to study the time dependent change of liver exosome proteomic profiles. For this study, a cohort of animals are administered acetaminophen and terminal bleeds are performed at 0, 6 and 24 hours post treatment to extract exosomes. In parallel, AST and ALT levels are assayed in the blood, along with examination of liver histology and performance of RNA-seq on liver tissue. Histology and RNA-seq provide independent methods for tracking liver function and health over the course of the experiments. These data will allow assessment of how well methods of the disclosure compares to traditional tests, while also enabling testing of whether the present method shows additional performance characteristics not seen with blood based tests of AST/ALT.

[00151] It is predicted that the exosome proteomic profiles will show improved sensitivity to up-regulation of protein markers that respond to liver toxicity. It is also predicted that sensitivity in post-treatment recovery will be improved. Whereas enzyme levels may be slow to decrease in the circulating bloodstream-ALT/AST have half-lives ranging from 14-72 hours (E. G. Giannini, et al., *CMAJ,* 2005) - exosomes biogenesis may occur on a faster timescale (M. Morishita, et al., *Journal of Pharmaceutical Sciences,* 2017) - exosomes have been shown to be absorbed with a half-life of 2-4 minutes-and allow to track

organ function in near real time as compared to traditional enzyme tests.

Example 6: Performing Single Exosome Sequencing (SESeq)

[00152] Exosomes are harvested from two well-characterized and easily-manipulatable cell lines, human embryonic kidney 293T cells and mouse 3T3 fibroblasts (Todaro, G. J. & Green, H. *J. Cell Biol. 1963,* 17: 299-313; and Graham, F. L, Smiley, J., Russell, W. C. & Nairn, R. J. *Gen. Virol. 1977,* 36: 59-74). Pure populations of exosomes are isolated by a series of ultracentrifugation steps and are characterized for purity using both nanoparticle tracking analysis and scanning electron microscopy (Kowal, E. J. K., Ter-Ovanesyan, D., Regev, A. & Church, G. M. *Methods Mol. Biol. Clifton NJ 2017,* 1660: 143-152; and Rupert, D. L. M., Claudio, V., Lasser, C. & Bally, M. Biochim. Biophys. Acta, 2017, 1861:3164-3179). Upon obtaining pure exosome populations from human embryonic kidney 293T cells and mouse 3T3 fibroblasts, equal quantities of two populations of exosomes are mixed together. To detect the proteins present on each exosome, the mixture of exosomes are incubated with a panel of DBAs, which can be used to quantify the level of hundreds of proteins of interest using conventional single cell sequencing pipelines (M Stoeckius, *et al., Nat. Methods 2017,* **14**: 865-868). This exosome mixture, taken as an input into the Columbia Genome Center's Chromium controller (10x Genomics), and each exosome is individually encapsulated and reverse transcribed. All RNAs from the same exosome are tagged with a common DNA barcode (**Fig. 6A-B**). The Chromium-generated libraries will then be run on the Illumina NextSeq 500 or NovaSeq sequencing platform. Because human and mouse genomes have large differences in character (neuronal-like versus fibroblast and organism origin, i.e., human versus mouse), it can be readily determined if the method of performing SESeq is successful, as the system should report only human or only mouse-specific RNA-sequences and antibody barcodes associated with each individual exosome.

[00153] Once SESeq protocol and analysis pipeline using mixtures of mouse and human exosomes are established, it will be applied to distinguishing between exosomes secreted from different types of human tissue. Specifically, exosomes from 8 immortalized human cell lines derived from various tissues (e.g., HUVEC for vascular endothelium, Ker-CT for skin, RPTEC for kidney tubular epithelium) will be isolated and then SESeq performed on the exosomes from each cell line individually and on exosome pools containing exosomes from 2, 4 or 8 lines. Sequencing data from the exosome pools will then be clustered in the absence of any knowledge of the number of input samples *(de novo* clustering) or clustered using the reference SESeq data obtained when the exosomes from

each cell line were sequenced individually (reference-based clustering). Finally, each cell line chosen will also be sequenced using conventional bulk RNA sequencing to better understand how well SESeq recapitulates the transcript diversity within each cell line.

[00154] Methods for exosome isolation and manipulation, next-generation sequencing, and clinical laboratory diagnostics required for successfully developing the SESeq platform are described in, for example, Alvarez, M. J. *et al. Nat. Genet. 2016,* 48: 838-847; Ding, H. *et al. Nat. Common. 2018,* 9: 1471; Guo, X. *et al. Nat. Biotechnol. 2018,* 36: 540-546; Bester, A. C. *et al. Cell 2018,* 173: 649-664.e20; DiCarlo, J. E., et al., Safeguarding CRISPR-Cas9 gene drives in yeast. *Nat. Biotechnol. 2015,* 33: 1250-1255; Chavez, A. *et al.*, *Nat. Methods 2015,* 12: 326-328; and Yeo, N. C. *et al.*, *Nat. Methods 2018,* 15: 611-616.

[00155] If upon running the SESeq pipeline, it is found that there are too few transcripts per exosome to be able to identify the cell of origin for the majority of exosomes, this can be overcome by 1) adapting single cell clustering algorithms, such as the metaVIPER, that are specifically designed to deal with issues of data sparsity such as gene dropout which violate many of the assumptions made by orthodox statistical tools; 2) modifying the exosome isolation procedure to capture various sub-fractions by size or density to determine if a particular fraction is most suitable for SESeq analysis (exosomes are a fraction of a heterogeneous group of extracellular vesicles) (Collino, F. et al. *Stem Cell Rev. 2017,* 13: 226-243; 3) adapting other methods such as SPLiT-seq which are able to capture >10x more events than methods using commercial devices at a similar costs (Li, M. *et al. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 2014,* 369; Alvarez, M. J. *et al., Nat. Genet. 2016,* 48: 838-847; Ding, H. *et al. Nat Commun 2018,* 9: 1471; Collino, F. *et al. Stem Cell Rev 2017,* 13: 226-243; and Rosenberg, A. B. *et al. Science 2018,* 360: 176-182).

Example 7: Determining the Response of Tumor and Normal Cells to Chemotherapy within Simulated Patient Samples

[00156] To establish methods of using SESeq to infer cellular state from both tumor and normal cells, human A549 non-small cell lung cancer (NSCLC) cells and RPTEC kidney cells are treated with paclitaxel or cisplatin. Exosomes will be collected from untreated and treated cells 4, 12, 24 and 48 hours after addition of drug, and SESeq will be performed on each sample. In addition, to better understand the relationship between contents of RNA and protein in exosomes and cellular state, conventional RNA sequencing and unbiased proteomics will be also performed on the A549 and RPTEC cells at all collection times.

[00157] Paclitaxel and cisplatin have been chosen for these studies because they are

commonly used in first line lung cancer treatment regimens, show varied efficacy within patients, have known nephrotoxic side-effects, and have diverse mechanisms of action (Malhotra, V. & Perry, M. C. *Cancer Biol. Ther. 2003,* 2: 1-3; Lameire, N. *Clin. Kidney J. 2014,* 7: 11-22; Perazella, M. A. *Clin. J. Am. Soc. Nephrol. CJASN7, 2012,* 1713-1721; and Ettinger, D. S. et al. Non-Small Cell Lung Cancer, Version 5.2017, NCCN Clinical Practice Guidelines in Oncology. *J. Natl. Compr. Cancer Netw. JNCCN 2017,* 15: 504-535).

[00158] Using the data obtained from SESeq, the cellular state of the A549 or RPTEC cells at each of the time points (e.g., cell cycle arrested or activating DNA damage response) will be inferred, and correlation with bulk cell analysis will be determined. Finally, to test the utility of SESeq in a more clinically-relevant setting, exosomes from the blood of several healthy human donors are isolated and spiked in varying amounts of A549 or RPTEC derived exosomes from paclitaxel-treated, cisplatin-treated or non-treated cells. Each sample will undergo SESeq, and a researcher blind to the composition of each sample will determine 1) if A549 and/or RPTEC cell-derived exosomes were added to the sample and 2) if the A549 and/or RPTEC cells were treated with drug.

[00159] It is predicted that it is possible to assign individual exosomes to a particular tissue by sequencing all RNAs along with determining the abundance of upwards of a hundred different proteins through the use of DBAs within the exosomes (Li, Y., Zhang, Y., Qiu, F. & Qiu, Z. *Electrophoresis 2011,* **32:** 1976-1983; and Lin, J. et al. *Scientific World Journal 2015,* **2015:** 657086). Once exosomes have been assigned, the data is pooled from all exosomes that have been assigned to a given tissue. Markers of cell state, such as transcription of stress response gene, or increased presence of markers of apoptosis, such as phosphatidylserine on the surface of the exosomes, are identified.

Example 8: Determining the Response of Tumor in a Mouse Model to Chemotherapy

[00160] Mouse models of NSCLC (Kwon, M. & Berns, A. *Mol. Oncol. 2013, 7:* 165-177) are prepared by injecting human A549 cells to mice. Once tumors have formed within the animals, the animals will be treated with a variety of drugs, some of which have previously been shown to cause tumor regression and others that are known to not affect tumor growth. At the start of the experiment and at various time points after initiating treatment, exosomes will be isolated from the blood of animals to determine the efficacy of drug treatment on the tumor. In addition, as all tissues within the animal are experiencing the effects of chemotherapy, insights gained from SESeq are analyzed to determine correlation with potential complications of treatment, such as cardiotoxicity, nephrotoxicity or gut

epithelial dysfunction.

Example 9: Determining the Response of Tumor in a NSCLC Patient to Chemotherapy

[00161] A study will be performed using SESeq on samples from NSCLC patients being treated with chemotherapy to determine whether insights gained from SESeq correlate with clinical outcome (e.g., patients predicted by SESeq to have tumors responding to treatment show decrease in tumor mass or patients predicted to be suffering from cardiomyocyte stress show increase in circulating cardiac troponin-l). To date, there are no universal solutions for rapidly determining a patient's response to treatment. Commonly used methods such as secreted tumor antigen are not available for the majority of tumors and also suffers from issues of specificity (Adhyam, M. & Gupta, A. K. A Review on the Clinical Utility of PSA in Cancer Prostate. *Indian J. Surg. Oncol. 2012,* **3:** 120-129). On the opposite end of the spectrum, serial imaging can be used to track nearly any tumor over time, yet it suffers from low assay sensitivity and at times requires weeks to months between imaging to confidently determine if there has been a change in a patient's tumor burden (Weber, W. A. *J. Nucl. Med. Off. Publ. Soc. Nucl. Med.2009,* 50 Suppl 1: 1S-10S).

[00162] SESeq can be used as a diagnostic for following a patient's response to therapy (see **Fig. 8**). Finally, by obtaining transcriptional information on not just tumor cells but all cells that secrete exosomes, SESeq can enable clinicians to better determine patients at high risk of deleterious drug side-effects, allowing patients to switch to an alternative chemotherapeutic regimen before suffering from irreversible toxicity.

We claim:

1. A method for detecting a plurality of target molecules in, or on the surface of, exosomes, cells or vesicles in a sample, the method comprising:

   (a) contacting the exosomes, cells or vesicles with a plurality of target molecule-binding agents, wherein each target molecule-binding agent comprises a nucleic acid barcode and optionally a unique molecular identifier (UMI), and wherein target molecule-binding agents that are specific to an identical target molecule share an identical nucleic acid barcode.

2. A method for diagnosing a disease or disorder, determining a patient's response to a therapy, and/or monitoring the patient for undesired side effects to the therapy, the method comprising:

(a) contacting exosomes, cells or vesicles with a plurality of target molecule-binding agents,

wherein each target molecule-binding agent comprises a nucleic acid barcode and optionally a unique molecular identifier (UMI),

wherein target molecule-binding agents that are specific to an identical target molecule share an identical nucleic acid barcode,

wherein the exosomes, cells or vesicles are isolated from a patient suffering from a disease or disorder,

wherein the plurality of target molecules comprise molecules that are markers indicative of the disease or disorder, efficacy of the therapy, and/or are indicative of undesirable side effects of the therapy, and

wherein expression levels of the target molecules determined in the patient are compared with expression levels of the corresponding target molecules determined in normal controls, and

thereby diagnosing a disease or disorder, e.g., cancer, autoimmune disease or inflammatory disease, determining the patient's response to a therapy, and/or monitoring for undesired side effects of a therapy.

3. The method of claim 2, wherein the patient is further treated with the same therapy, if the therapy is effective, but associated with little or no undesired side effects; or wherein the

patient is further treated with a different therapy if the therapy is not effective and/or is associated with undesired side effects.

4. The method of any one of claims 1-3, wherein each target molecule-binding agent further comprises a universal round 1 primer sequence at the 3' end for a first round extension.

5. The method of any one of claims 1-4, further comprising:

(b) dividing the exosomes, cells or vesicles, to which the target molecule-binding agents are bound, into at least two primary aliquots, the at least two primary aliquots comprising a first primary aliquot and a second primary aliquot;

(c) adding primary nucleic acid tags to the target molecule-binding agents in the at least two primary aliquots, wherein the primary nucleic acid tags added to the target molecule-binding agents in any one of the at least two primary aliquots are different from the primary nucleic acid tags added to the target molecule-binding agents in any one of the other primary aliquots;

(d) combining the at least two primary aliquots;

(e) dividing the combined primary aliquots into at least two secondary aliquots, the at least two secondary aliquots comprising a first secondary aliquot and a second secondary aliquot;

(f) adding secondary nucleic acid tags to the at least two secondary aliquots, wherein the secondary nucleic acid tags added to the target molecule-binding agents in any one of the at least two secondary aliquots are different from the secondary nucleic acid tags added to the target molecule-binding agents in any one of the other secondary aliquots; and

(g) repeating steps (d), (e), and (f) with the at least two secondary aliquots a number of times sufficient to generate a unique series of nucleic acid tags for each exosome, cell or vesicle in the sample.

6. The method of claim 5, wherein the primary nucleic acid tags, the secondary nucleic acid tags and/or subsequent nucleic acid tags are added by ligation reactions, polymerase extension reactions, and/or chemical syntheses.

7. The method of claim 6, wherein the nucleic acid tags are added by polymerase extension reaction,

wherein the nucleic acid barcode bound to each target molecule-binding agent is extended with one of the primary nucleic acid tags by contacting the exosomes, cells or vesicles with a strand displacing polymerase and a first DNA hairpin, wherein the first DNA hairpin comprises

(i) a first oligonucleotide comprising a sequence complementary to the universal round 1 primer sequence, and

(ii) a second oligonucleotide, wherein the second oligonucleotide comprises a third oligonucleotide comprising a primary nucleic acid tag, and a fourth oligonucleotide comprising a sequence complementary to the primary nucleic acid tag;

wherein the third oligonucleotide is located at the 5' end of the second oligonucleotide, and the fourth oligonucleotide is located at the 3' end of the second oligonucleotide; and

wherein the first oligonucleotide is fused to the 3' end of the fourth oligonucleotide.

8. The method of claim 7, wherein each primary nucleic acid tag comprises a unique well-specific first round barcode sequence at the 5' end and a universal round 2 primer sequence at the 3' end.

9. The method of claim 7 or 8, wherein the first DNA hairpin is disabled at the end of each polymerase extension by removal of the first oligonucleotide from the first DNA hairpin using an exonuclease or by treating the first DNA hairpin with an enzyme that remove a unique base present at the junction of the first oligonucleotide and the fourth oligonucleotide.

10. The method of claim 9, further comprising adding secondary nucleic acid tags by polymerase extension reaction,

wherein the nucleic acid barcode bound to each target molecule-binding agent is further extended with one of the secondary nucleic acid tags by contacting the exosomes, cells or vesicles with a strand displacing polymerase and a second DNA hairpin wherein the second DNA hairpin comprises:

(i) a fifth oligonucleotide comprising a sequence complementarity to the universal round 2 primer sequence, and

(ii) a sixth oligonucleotide, wherein the sixth oligonucleotide further comprises a seventh oligonucleotide comprising a secondary nucleic acid tag, and an eighth oligonucleotide comprising a sequence encoding a sequence complementary to the secondary nucleic acid tag;

wherein the seventh oligonucleotide is located at the 5' end of the sixth oligonucleotide, and the eighth oligonucleotide is located at the 3' end of the sixth oligonucleotide; and

wherein the fifth oligonucleotide is fused to the 3' end of the eighth oligonucleotide.

11. The method of claim 10, wherein each secondary nucleic acid tag comprises a unique well-specific second round barcode sequence at the 5' end and a universal round 3 primer sequence at the 3' end.

12. The method of claim 6, wherein the primary nucleic acid tags are added by ligation reaction in a first round splint-ligation reaction, wherein the primary nucleic acid tags are added to the 3' end of the nucleic acid barcode bound to each target molecule-binding agent by contacting the exosomes, cells or vesicles with a ligase, a first-round oligonucleotide and a first-round splint sequence,

wherein the first-round oligonucleotide comprises a 5' common region followed by a primary nucleic acid tag terminated by a 3' universal round 2 sequence; and

wherein the first-round splint sequence comprises a region complementary to the universal round 1 sequence at 3' end of the nucleic acid barcode bound to each target molecule-binding agent and a region complementary to the 5' common region of the first-round oligonucleotide.

13. The method of claim 12, wherein the splint-ligation process is terminated at the end of the first-round splint-ligation reaction.

14. The method of claims 12 or 13, wherein secondary nucleic acid tags are added through a second round splint ligation reaction,

wherein the nucleic acid barcode bound to each target molecule-binding agent is further extended with a secondary nucleic acid tag by contacting the exosomes, cells or vesicles with a ligase, a second round oligonucleotide and a second round splint sequence,

wherein the second round oligonucleotide comprises a 5' common region followed by a secondary nucleic acid tag terminated by a 3' universal round 3 sequence or a universal PCR sequence; and

wherein the second round splint sequence comprises a region complimentary to the 3' universal round 2 sequence of the first-round oligonucleotide, and a region complimentary to the 5' common region of the second round oligonucleotide.

15. The method of any one of claims 5-14, further comprising amplifying and sequencing the nucleic acid barcodes, the UMIs and the nucleic acid tags.

16. The method of any one of claims 1-15, wherein the target molecules are proteins, sugar moieties, lipids, and/or polynucleotides.

17. The method of any one of claims 1-16, wherein the target molecule-binding agents comprise an antibody, an antibody fragment, a peptide aptamer, lectins, a phage display system or a yeast display system.

18. The method of any one of claims 1-17, wherein the nucleic acid barcode is a DNA-barcode.

19. The method of any one of claims 1-18, wherein the detecting is at a single cell or a single exosome or a single vesicle level.

20. The method of any one of claim 19, wherein each of the at least two primary aliquots consists of a single cell, a single exosome, or a single vesicle.

Figure 1

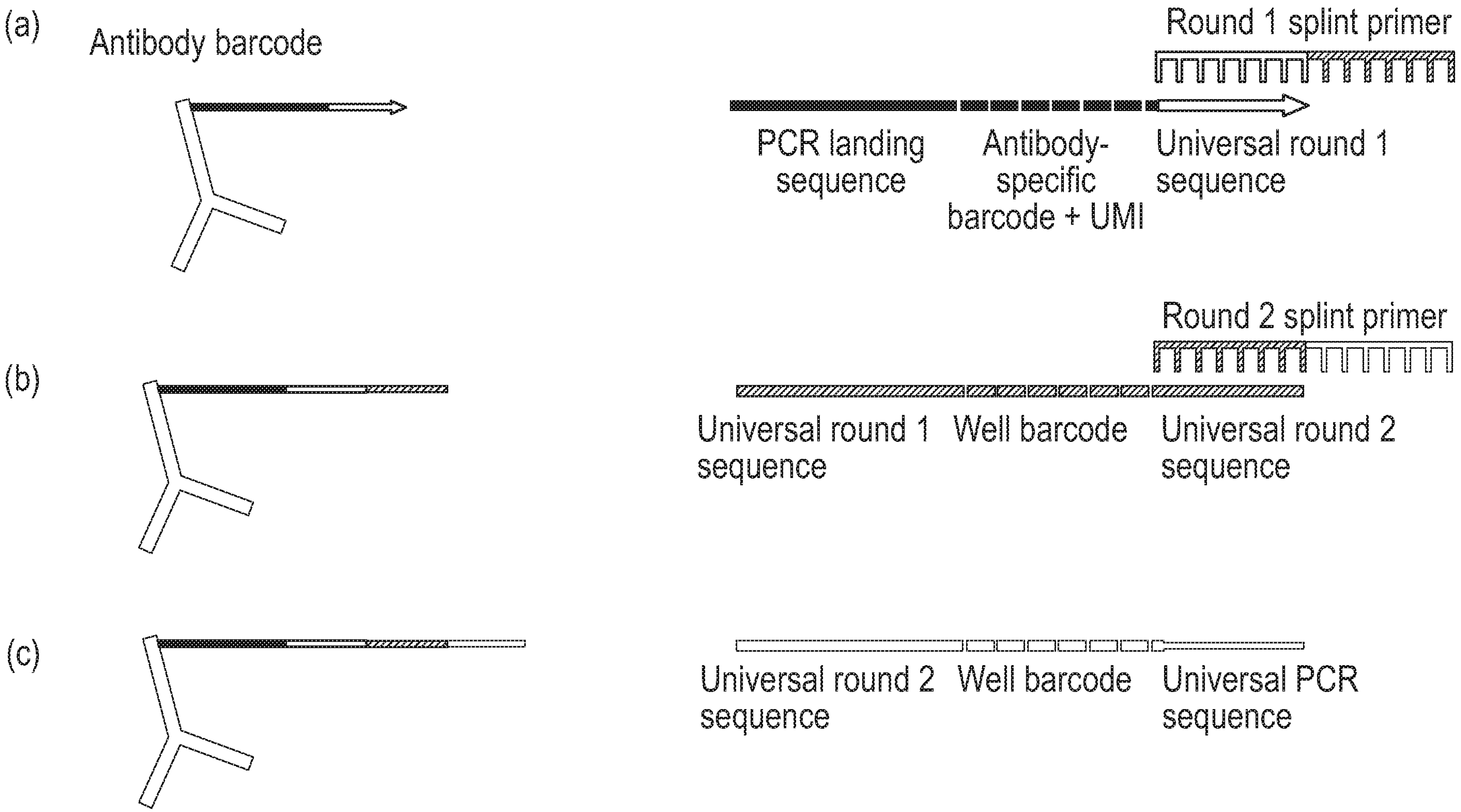
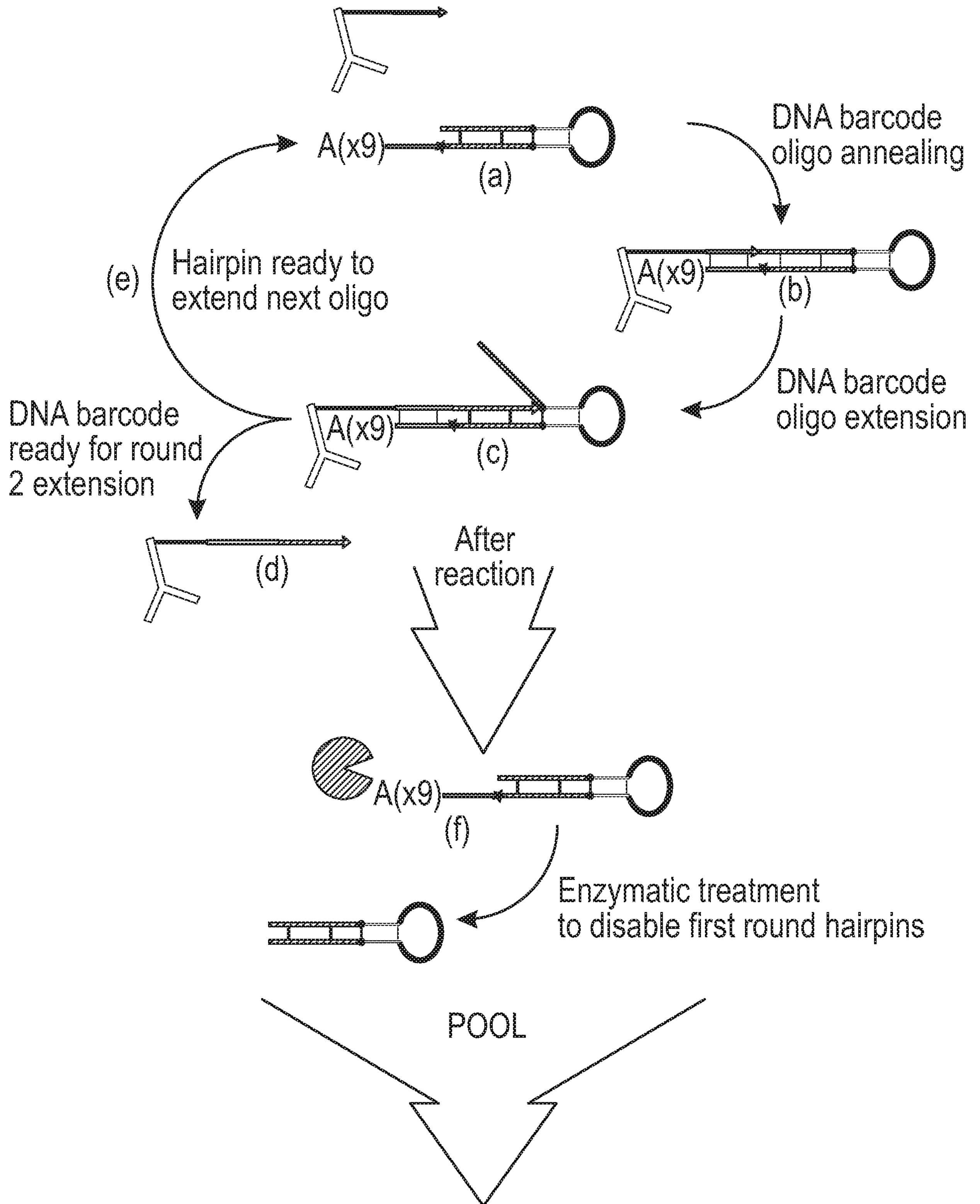


Figure 2



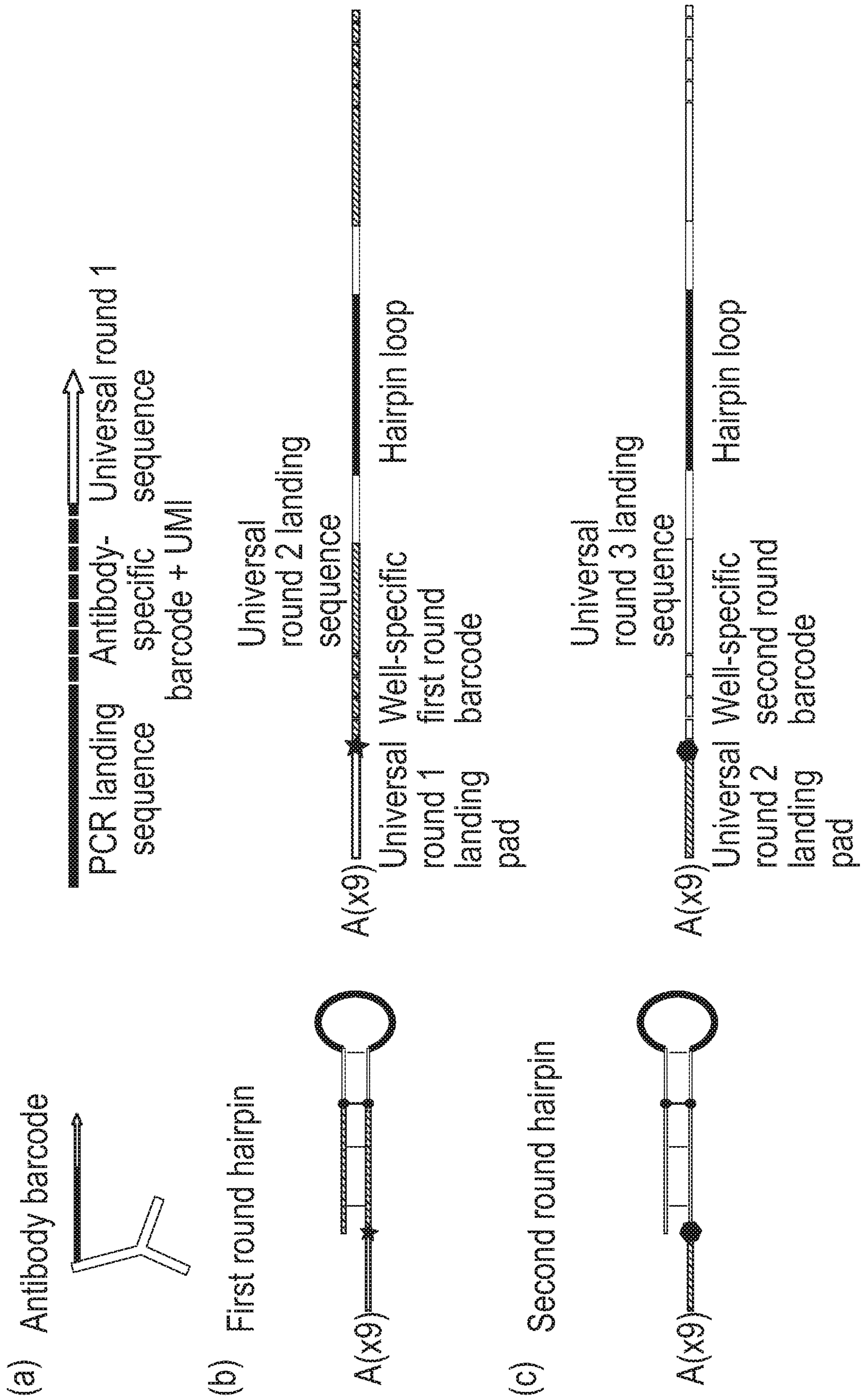
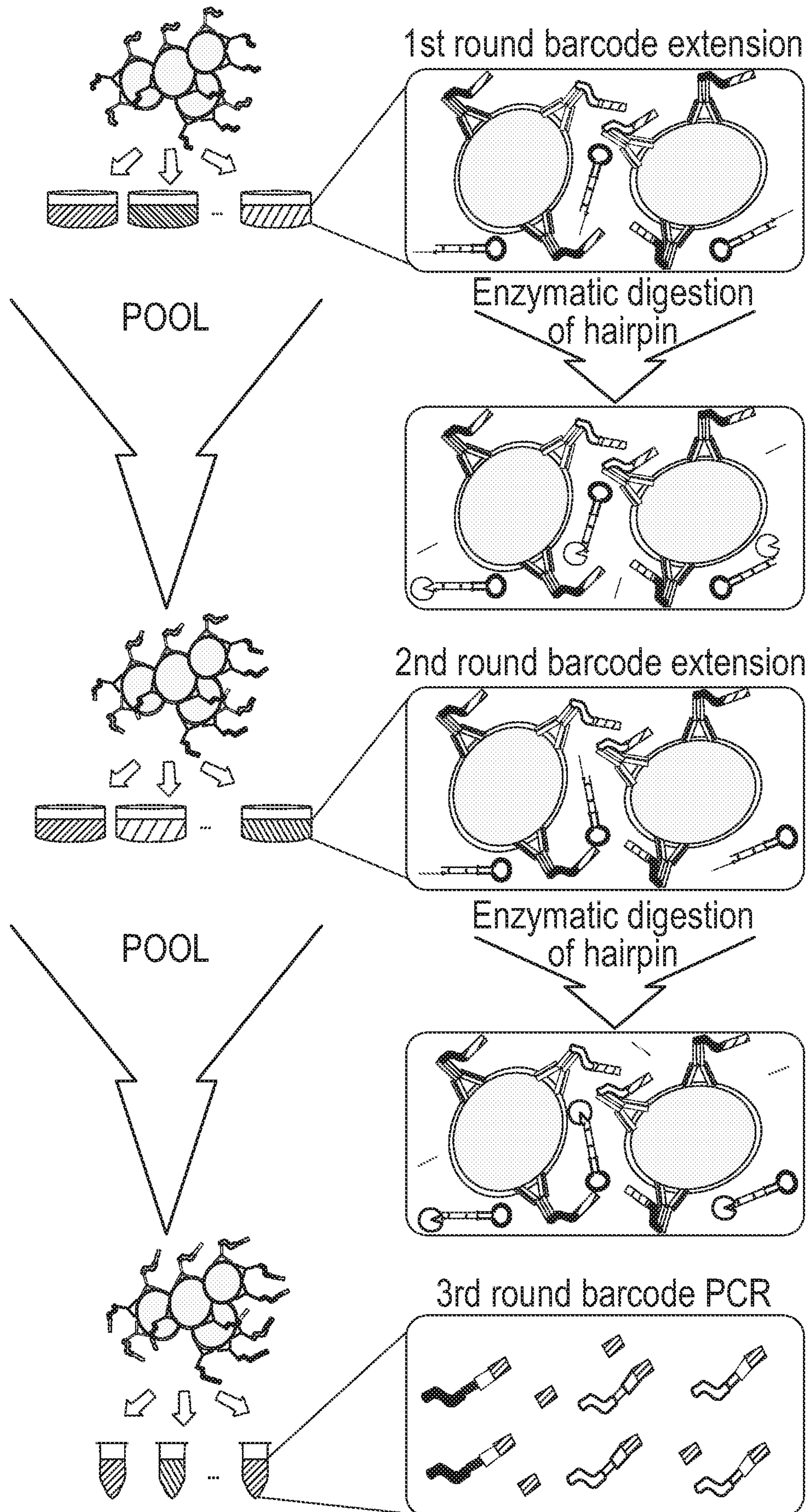


Figure 3
(a) Antibody barcode
PCR landing sequence
Antibody-specific barcode + UMI
Universal round 1 sequence
(b) First round hairpin
A(x9)
Universal round 1 landing pad
Well-specific first round barcode
Universal round 2 landing sequence
Hairpin loop
(c) Second round hairpin
A(x9)
Universal round 2 landing pad
Well-specific second round barcode
Universal round 3 landing sequence
Hairpin loop
First round modification for pre-pooling enzymatic destruction
Second round modification for pre-pooling enzymatic destruction

Figure 4



*Note that leftover products of the enzymatic digestion steps have been left out for clarity.

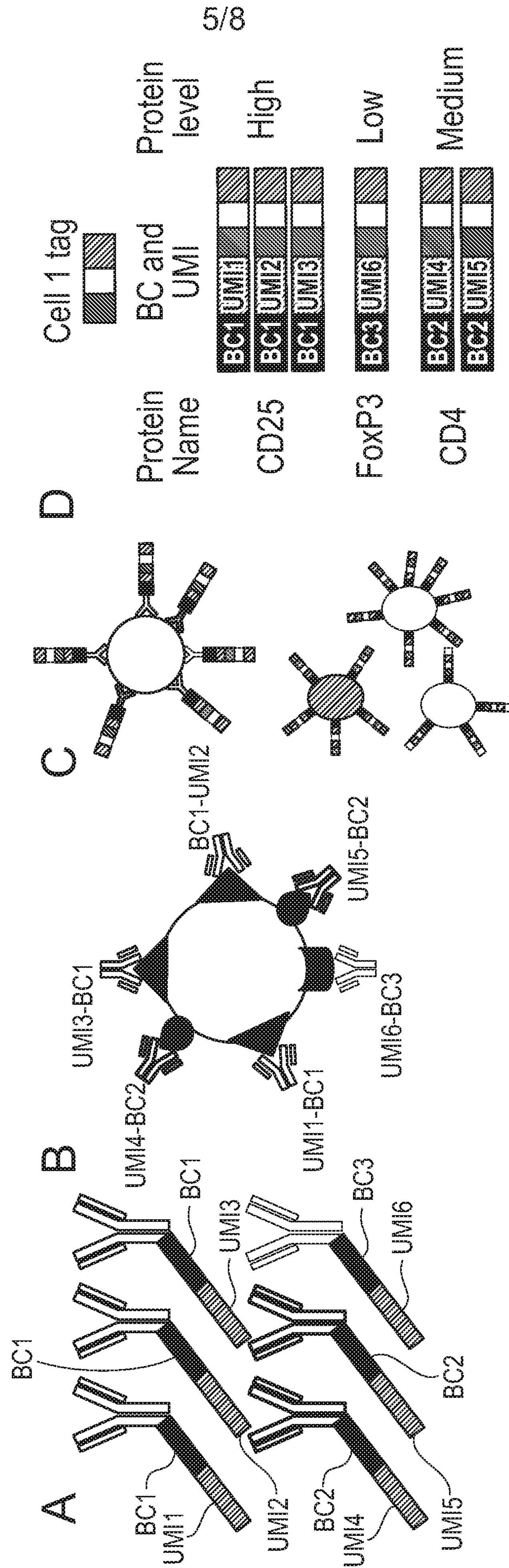


Figure 5

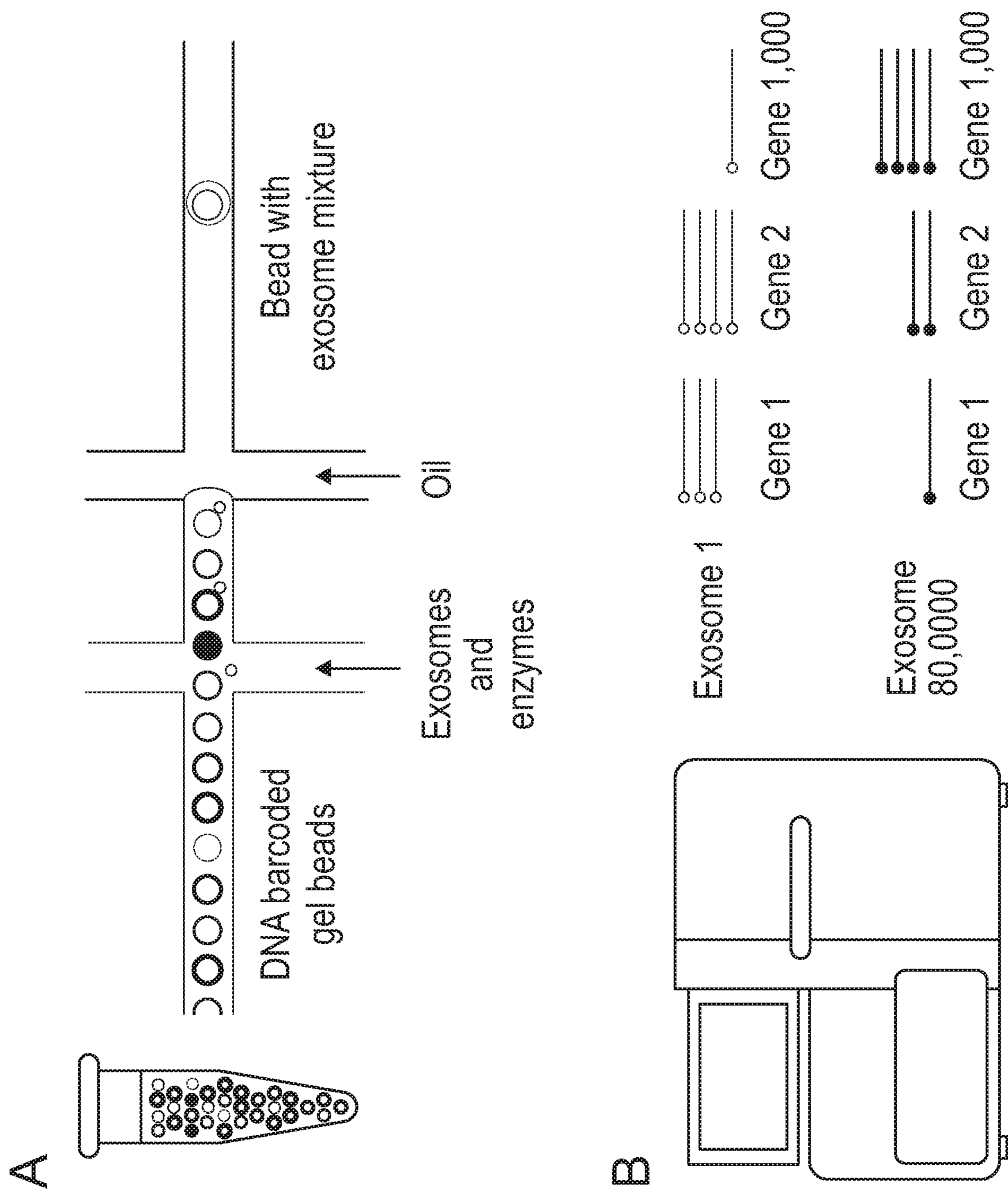


Figure 6

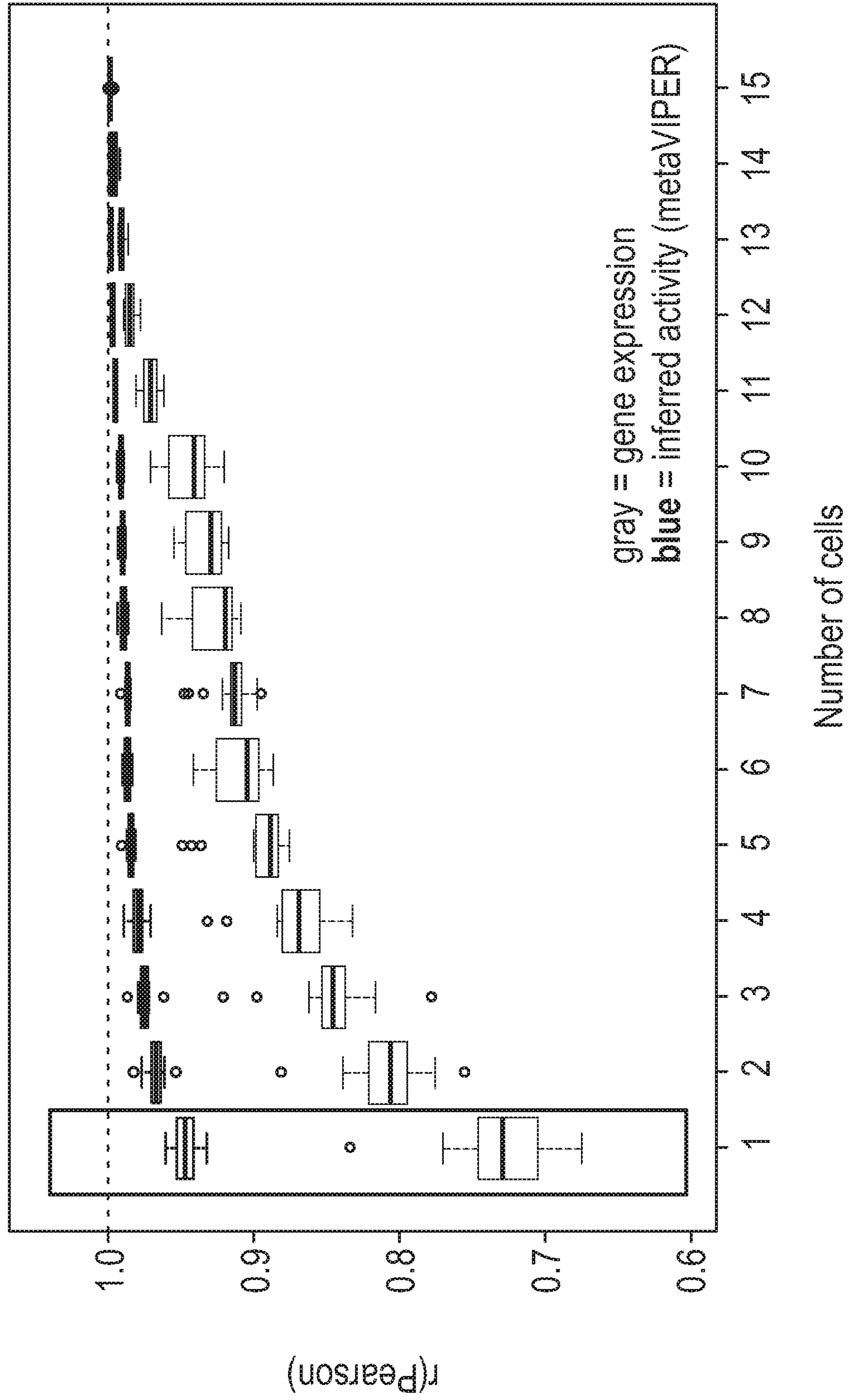


Figure 7
r(Pearson)
1.0
0.9
0.8
0.7
0.6
gray = gene expression
blue = inferred activity (metaVIPER)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
Number of cells

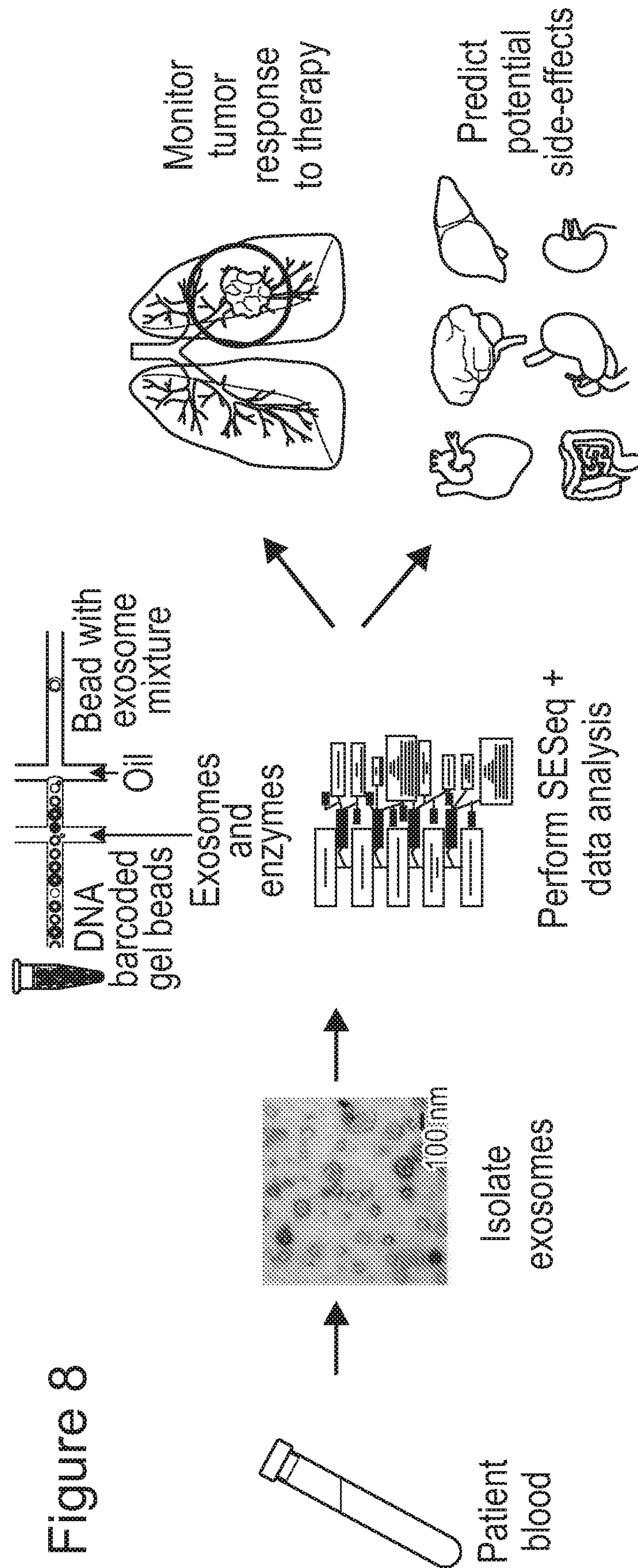


Figure 8

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2020/017327

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - G01N 33/532; G01N 33/566 (2020.01)
CPC - B01J 2219/00547; G01N 33/532; G01N 33/566; G01N 2458/10 (2020.02)

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC - 506/4; 506/9; 506/14 (keyword delimited)

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X -- Y	US 2018/0267036 A1 (CELLULAR RESEARCH, INC.) 20 September 2018 (20.09.2018) entire document	1, 4 ----- 2, 3
Y	US 2014/0287422 A1 (UNIVERSITY OF TOLEDO) 25 September 2014 (25.09.2014) entire document	2, 3
Y	US 2017/0299598 A1 (ONCOMED PHARMACEUTICALS, INC.) 19 October 2017 (19.10.2017) entire document	3
A	WO 2018/226293 A1 (BECTON, DICKINSON AND COMPANY) 13 December 2018 (13.12.2018) entire document	1-4
A	WO 2017/100251 A1 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE) 15 June 2017 (15.06.2017) entire document	1-4
A	US 2018/0251825 A1 (NEW YORK GENOME CENTER INC.) 06 September 2018 (06.09.2018) entire document	1-4

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

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

Date of the actual completion of the international search	Date of mailing of the international search report
30 March 2020	23 APR 2020

Name and mailing address of the ISA/US	Authorized officer
Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, VA 22313-1450 Facsimile No. 571-273-8300	Blaine R. Copenheaver PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2020/017327

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

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

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 5-20
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

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

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

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.