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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
15 August 2024 (15.08.2024)

WIPO | PCT

(10) International Publication Number
WO 2024/168092 A2

(51) International Patent Classification:
C12Q 1/6806 (2018.01) *C12Q 1/686* (2018.01)

(21) International Application Number:
PCT/US2024/014893

(22) International Filing Date:
07 February 2024 (07.02.2024)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

63/483,741	07 February 2023 (07.02.2023)	US
63/471,951	08 June 2023 (08.06.2023)	US
63/614,344	22 December 2023 (22.12.2023)	US

(71) Applicant: **PARSE BIOSCIENCES, INC.** [US/US]; 700 Dexter Avenue, Suite 600, Seattle, WA 98109 (US).

(72) Inventors: **ROSENBERG, Alexander, B.**; c/o Parse Biosciences, Inc., 700 Dexter Avenue, Suite 600, Seattle, WA 98109 (US). **ROCO, Charles**; c/o Parse Biosciences, Inc., 700 Dexter Avenue, Suite 600, Seattle, WA 98109 (US). **PAPALEXI, Efthymia**; c/o Parse Biosciences, Inc., 700 Dexter Avenue, Suite 600, Seattle, WA 98109 (US). **PANGALLO, Joseph**; c/o Parse Biosciences, Inc., 700 Dexter Avenue, Suite 600, Seattle, WA 98109 (US). **TRAN, Nam-Vuong, Khai**; c/o Parse Biosciences, Inc., 700 Dexter Avenue, Suite 600, Seattle, WA 98109 (US).

(74) Agent: **BECKER, Daniel M.** et al.; Fenwick & West LLP, 801 California Street, Mountain View, CA 94041 (US).

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

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE,

(54) Title: METHODS AND KITS FOR LABELING CELLULAR MOLECULES FOR MULTIPLEX ANALYSIS

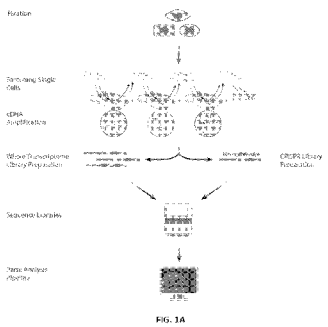


FIG. 1A

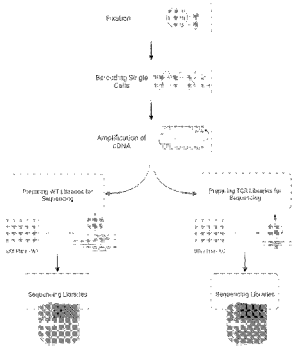


FIG. 1B

(57) Abstract: The present disclosure relates generally to multiplex methods of cell- or nucleus-specific labeling of molecules such as nucleic acids. The present disclosure also relates to kits for multiplex, cell- or nucleus-specific labeling of molecules. In some embodiments, the methods and kits relate to parallel labeling of nucleic acids corresponding to target sequences of interest and to the whole transcriptome.

WO 2024/168092 A2

[Continued on next page]

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:

— *of inventorship (Rule 4.17(iv))*

Published:

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

METHODS AND KITS FOR LABELING CELLULAR MOLECULES FOR MULTIPLEX ANALYSIS

FIELD

[0001] The present disclosure relates generally to methods of uniquely labeling or barcoding molecules within or originating from a nucleus or plurality of nuclei, a cell or plurality of cells, or one or more tissues, organs, or organisms. The present disclosure also relates to kits for uniquely labeling molecules within or originating from a nucleus or plurality of nuclei, a cell or plurality of cells, or one or more tissues, organs, or organisms. In particular, the methods and kits may relate to the labeling of RNAs and/or cDNAs within cells for the preparation of sequencing libraries for multiplex analysis.

CROSS-REFERENCE TO RELATED APPLICATIONS

[0002] The present application claims the benefit of and priority to U.S. Provisional Application No. 63/483,741, filed February 7, 2023, U.S. Provisional Application No. 63/471,951, filed June 8, 2023, and U.S. Provisional Application No. 63/614,344, filed December 22, 2023, each of which is herein incorporated by reference in its entirety.

BACKGROUND

[0003] Next Generation Sequencing (NGS) can be used to identify and/or quantify individual transcripts from a sample of cells. However, such techniques may be complicated to perform on individual cells in large samples. In such methods, RNA transcripts are generally purified from lysed cells (i.e., cells that have been broken apart), followed by conversion of the RNA transcripts into complementary DNA (cDNA) using reverse transcription. The cDNA sequences can then be sequenced using NGS. In such a procedure, all of the cDNA sequences are mixed together before sequencing, such that RNA expression is measured for a whole sample and individual sequences cannot be linked back to an individual cell.

[0004] Methods for uniquely labeling or barcoding transcripts from individual cells can involve the manual separation of individual cells into separate reaction vessels and can require specialized equipment. An alternative approach to sequencing individual transcripts in cells is to use microscopy to identify individual fluorescent bases. However, this technique can be difficult to implement and is limited to sequencing a low number of cells.

[0005] Single-cell sequencing can allow the identification of transcripts from individual cells. However, a limitation of single-cell RNA sequencing is the inability to detect every expressed

gene in a given cell without exhaustive sequencing, which can be cost prohibitive due to the sequencing depth required per cell. This is problematic for applications such as single cell clustered regularly interspaced short palindromic repeats (CRISPR) screens where it is essential to detect the guide RNA (gRNA) in every cell in order to associate a transcriptome-wide perturbation effect. While the development of pooled CRISPR screens with a single-cell RNA sequencing readout have enabled researchers to dissect the effects of genetic perturbations on single-cell transcriptional profiles, the scalability of these assays remains a challenge due to the inability to detect every expressed single guide RNA (sgRNA) transcript without exhaustive sequencing.

[0006] Similarly, understanding the relationship between naturally varying genes in populations of cells and the whole transcriptome in the same cells can provide critical insight into the biological and/or disease-related roles of each variant. For example, analyzing the relationship between T cell receptors (TCRs) and T cell activation during disease pathogenesis and progression can assist, e.g., in the development of next-generation therapeutics with more favorable and sustainable outcomes. However, previous approaches to accomplishing this using, e.g., single-cell sequencing to profile receptor sequences and full transcriptomes, have been limited by a variety of factors, e.g., related to their requirement for microfluidics devices or plate-based protocols, and/or because of their limited sensitivity and throughput (1,000s-10,000s of cells), making these approaches both time consuming and costly.

[0007] Accordingly, there is a need for new methods of single-cell sequencing that allow the specific detection of particular individual targets in every cell without the need for exhaustive sequencing, together with an ability to associate the presence, expression, and/or identity of the individual targets with the whole transcriptome or a subset thereof. The present disclosure addresses this need and provides other advantages as well.

BRIEF DESCRIPTION OF THE DRAWINGS

[0001] FIGs. 1A-1B provide overviews of two embodiments of the present method. In both cases, cells or nuclei are fixed and permeabilized, cDNA is generated within the cells or nuclei by reverse transcription using well-specific barcoded primers, one or more additional barcodes are appended to the cDNA within the cells or nuclei by a split-pool labeling process, the cells or nuclei are lysed, the released barcoded cDNA is isolated and a second cDNA strand is generated by template switching, the cDNA is amplified in a multiplex “preamplification” reaction using both non-specific primers (whole transcriptome, or WT primers) and target-

specific primers, and then the amplified cDNA molecules are used to prepare, in parallel, a Whole Transcriptome sequencing library and a target sequencing library (such as a CRISPR library or a TCR library). Finally, the sequencing libraries are sequenced, and the sequencing reads analyzed. FIG. 1A illustrates an exemplary method for multiplex labeling of nucleic acids in cells in the context of a CRISPR screen. FIG. 1B illustrates an exemplary method for multiplex labeling of nucleic acids in cells in the context of TCR profiling of T cells.

[0008] FIGS. 2A-2C provide an overview of the cDNA products obtained at different steps in one embodiment of the present methods, e.g., using a pair of nested "focal primers" to enrich for one or more target sequences when preparing parallel WT and target-specific sequencing libraries. FIG. 2A shows the products of the combinatorial barcoding of cDNA in cells or nuclei following template switching. Both a generic cDNA molecule representative of the whole transcriptome, and a target cDNA molecule of interest are shown. "BC1," "BC2," "BC3": barcodes added during combinatorial labeling; "TS Primer Sequence": primer (adapter) sequence added during template switching; "Focal Primer 1 Binding Site" and "Focal Primer 2 Binding Site": nested target primer binding sites present on the target cDNA molecule; R2: Illumina Truseq R2 sequence. FIG. 2B shows primer binding sites used during multiplex cDNA amplification with both non-specific whole transcriptome primers and with spiked-in target primers (i.e., the "preamplification" step using WT and target "preamplification primers" as described elsewhere herein). In the image, the whole transcriptome is amplified by PCR using a pair of non-specific WT primers binding to the R2 sequence ("R2 Primer") and the TS primer binding sequence ("PCR Primer"), and the target sequences are enriched by PCR using a pair of primers including one target-specific primer ("Focal Primer 1") and one non-specific primer ("R2 Primer"). FIG. 2C shows the two types of cDNA products resulting from the multiplex amplification (i.e., preamplification) step shown in FIG. 1B: the whole transcriptome (WT) cDNA products, and the target cDNA products ("Gene-specific cDNA from first enrichment). The combined sample containing the two products can then used as input for the subsequent separate preparation of whole transcriptome (WT) sequencing libraries and target (e.g., "focal PCR") sequencing libraries.

[0009] FIGS. 3A-3D provides an overview of *in situ* cell barcoding steps (i.e., combinatorial barcoding, or split-pool labeling) performed in various embodiments of the present methods. FIG. 3A: Round 1 Barcoding. Fixed and permeabilized cells are loaded into a multiple wells (e.g., 48 wells) of a Round 1 plate. RNA is reverse transcribed to generate cDNA using reverse transcription primers comprising a well-specific barcode ("BC1") and, e.g., a poly(dT)

sequence or a random sequence. FIG. 3B: Round 2 Barcoding: The cells containing the cDNA are pooled and loaded into a Round 2 Plate. An adapter (nucleic acid tag) with a well-specific barcode ("BC2") is ligated to the cDNA, e.g., 5' to the first barcode. FIG. 3C: Round 3 Barcoding: The cells are pooled and loaded into a Round 3 Plate. A third barcode is ligated to the cDNA (indicated in red) via another adapter, which also contains an Illumina R2 sequence, and biotin. FIG. 3D: Lysis and Sublibrary Generation: Cells are split into multiple sublibraries (or "samples") (e.g., 8 sublibraries or samples) and lysed.

[0010] FIGS. 4A-4D provide an overview of cDNA capture and amplification steps performed in various embodiments of the present methods. FIG. 4A: cDNA Capture: Following cell lysis, biotinylated cDNA is captured (isolated) in each sublibrary via streptavidin beads. FIG. 4B: cDNA Template Switch: A template switch (TS) reaction adds an adapter to the 3' end of the cDNA. FIG. 4C: WT cDNA Amplification: the whole transcriptome (WT) cDNA is amplified by PCR using primers binding to a Template Switch (TS) sequence and to the Illumina Truseq R2 sequence. FIG. 4D: cDNA amplification with spiked-in target primers. In some embodiments, target specific primers are added, or spiked-in, during the first round of cDNA amplification (i.e., the amplification step shown in FIG. 4C), so as to enrich the presence of the target cDNAs among the whole transcriptome cDNA molecules. FIG. 4C shows a target (hU6-sgRNA-polyA) transcript enriched due to the presence of a spiked-in human U6 specific primer, alongside two non-target cDNAs from the whole transcriptome. The amplified cDNA mixture comprising the WT and target cDNAs can be split and used to separately prepare the Whole Transcriptome and target (e.g., CRISPR) sequencing libraries.

[0011] FIGS. 5A-5C show steps in the preparation of WT sequencing libraries starting from amplified WT cDNA molecules as shown, e.g., in FIG. 4C. FIG. 5A: the cDNA molecules are fragmented, and the ends are repaired and then A-tailed. FIG. 5B: Adapter Ligation: An Illumina Truseq R1 Adapter is ligated to the 5' end of the DNA. FIG. 5C: Round 4 Barcoding: The sequencing library is amplified, adding P5/P7 Adapters and a fourth barcode via the UDI – WT Plate.

[0012] FIGS. 6A-6C show steps in the preparation of target (e.g., CRISPR) sequencing libraries starting from enriched target cDNA molecules amplified with spiked in primers during the "preamplification" PCR step (i.e., the target cDNA molecules shown in FIG. 4D). FIGS. 6A-6B show the products of two additional amplification steps performed after the initial preamplification round of PCR. FIG. 6A: CRISPR PCR: In the first additional round of

amplification, a PCR reaction using a second hU6 specific primer further amplifies and enriches the target (sgRNA) cDNA molecules. This reaction also adds an adaptor (e.g., with an Illumina R1 sequence). FIG. 6B: CRISPR Index PCR: In the second additional round of amplification, the CRISPR Sequencing Library is amplified, adding P5/P7 adaptors and a fourth barcode via the Illumina indexes in the UDI Plate - EC. FIG. 6C: another representation of the products of the two additional rounds of amplification.

[0013] FIGS. 7A-7C show steps in the preparation of target (e.g.,) TCR sequencing libraries. FIGS. 7A-7B show the products of two additional amplification steps performed after the initial preamplification round of PCR. FIG. 7A: TCR Amplification 1: In the first additional round of amplification, a PCR reaction using a TCR specific primer further amplifies and enriches the target (TCR) cDNA molecules, e.g., cDNA molecules from the whole transcriptome that contain V(D)J segments in the CDR3 repertoire of the T cell. This reaction also adds an adaptor (e.g., an adapter with an Illumina Nextera R1 sequence). FIG. 7B: TCR Amplification 2: In the second additional round of amplification, the TCR Sequencing Library is amplified, adding P5/P7 adaptors and a fourth barcode via the Illumina indexes in the UDI Plate - EC. FIG. 7C: another representation of the products of the two additional rounds of amplification.

[0014] FIGS. 8A-8B show validation of enrichment ability of spiked-in target primers for Focal Barcoding protocol. FIG. 8A: Percentage of cells with gene detected in HEK293 or NIH/3T3 cells, in Whole Transcriptome vs. Focal libraries, for GPX4, PRDX2, CHCHD2, KDELR1, Psmd2, GAPDH, RPL5, and Actb genes. Whole Transcriptome libraries were sequenced at 10,000 reads/cell. Focal libraries were sequenced at 250 reads/cell. FIG. 8B: Number of unique transcripts captured in HEK293 or NIH/3T3 cells, in Whole Transcriptome vs. Focal libraries, for GPX4, PRDX2, CHCHD2, KDELR1, Psmd2, GAPDH, RPL5, and Actb genes. Whole Transcriptome libraries were sequenced at 10,000 reads/cell. Focal libraries were sequenced at 250 reads/cell.

[0015] FIGS. 9A-9B illustrate enrichment of spiked-in primers for moderately expressed genes, and improvement in purity following application of a 1 read count threshold filter. FIG. 9A: Two low- to medium-expressing genes were enriched using the herein-disclosed methods in humans (KDELR1) and mice (Psmd2), and the level of enrichment (as measured by virtue of the percentage of cells comprising the gene or reads per cell) was determined in 12k and 62k cell sublibraries. FIG. 9B: Sequencing libraries were prepared according to the herein-described methods using two sets of target genes (Psmd2-KDELR1 or Fn1-RPL5) with 10 ng

or 50 ng of preamplified cDNA introduced into the first round of target sequence specific amplification performed subsequent to the multiplex "preamplification" round of amplification.

[0016] FIG. 10 shows Tapestation image showing the amount of enriched target gene transcripts (RPL5 and Fn1) in samples prepared with (lane C2) or without (lane A2) spiked-in target-specific primers during a first round of multiplex cDNA amplification (i.e., "preamplification").

[0017] FIG. 11 shows the fraction of sequencing reads with valid barcodes that mapped to either of the targeted genes RPL5 or Fn1, in samples prepared with (sublibrary S3) or without (sublibrary S1) spiked-in target-specific primers during a first round of multiplex cDNA amplification (i.e., "preamplification").

[0018] FIG. 12 shows the number of target gene transcripts (i.e., RPL5 or Fn1 transcripts) detected per cell in samples prepared with (sublibrary S3) or without (sublibrary S1) spiked-in target-specific primers during a first round of multiplex cDNA amplification (i.e., "preamplification").

[0019] FIGS. 13A-13B show the fraction of cells that contained an enriched transcript (i.e., an RPL5 or Fn1 transcript), both without any filtering (FIG. 13A) or following a filtering step to only consider transcripts represented by more than 2 reads (FIG. 13B), in samples prepared with (sublibrary S3) or without (sublibrary S1) spiked-in target-specific primers during a first round of multiplex cDNA amplification (i.e., "preamplification").

[0020] FIGS. 14A-14B show high TCR detection in primary T cells showing sensitive clonotype detection. FIG. 14A shows high TCR chain identification rate. Isolated T cells from healthy donor PBMCs were directly profiled (Primary). Alpha, Beta, and Paired detection are represented in percentages. FIG. 14B shows TCR chain assignment across 8 donors. High rate of chain assignments to both TCR alpha and beta. Among T cells with a detected TCR, paired alpha beta chain assignments ranged between 49%-66%.

[0021] FIGS. 15A-15B show comprehensive Immune Repertoire Detection measured by number of unique alpha and beta chain clonotypes across donors. Nearly four hundred thousand unique alpha chain clonotypes and five hundred thousand unique beta chain clonotypes were identified across the 8 donors, with the vast majority being classified as rare clonotypes. The rare clonotypes (darker color, lower shaded) are defined as only being detected in 1 or 2 cells

and the majority of detected clonotypes are rare. FIG. 15A: Unique Alpha Chain. FIG. 15B: Unique Beta Chain.

[0022] FIGS. 16A-16D show increased detection of TCR alpha and beta chains with spiking in of TCR-specific primers during first multiplex cDNA amplification step (i.e., "preamplification" step). FIG. 16A shows percentages of activated T cells with detected alpha chain or with no detected chain, in libraries prepared with or without TCR-specific preamplification primers. FIG. 16B shows percentages of resting T cells with detected alpha chain or with no detected chain, in libraries prepared with or without TCR-specific preamplification primers. FIG. 16C shows percentages of activated T cells with detected beta chain or with no detected chain, in libraries prepared with or without TCR-specific preamplification primers. FIG. 16D shows percentages of resting T cells with detected beta chain or with no detected chain, in libraries prepared with or without TCR-specific preamplification primers. Spike-in +e: TCR-specific preamplification primers added. Spike-in -: TCR-specific preamplification primers not added.

SUMMARY

[0023] The present disclosure provides methods, compositions, kits, and systems for labeling RNA and other molecules within and originating from cells and nuclei. In some embodiments, the present methods relate to single cell methods of labeling, in parallel, target transcripts of interest such as CRISPR gRNAs or T cell receptors, and the transcriptomes of single cells, e.g., whole transcriptomes or subsets of the transcriptome.

[0024] In one aspect, disclosed herein are methods of labeling nucleic acids for multiplex transcriptional analysis in a plurality of cells or nuclei, the method comprising:

(a) providing a plurality of fixed and permeabilized cells or nuclei, each comprising a plurality of RNA molecules, and each comprising one or more target genes or transcripts of interest;

(b) dividing the plurality of cells or nuclei into a first plurality of aliquots, wherein each aliquot comprises more than one cell or nucleus;

(c) generating complementary DNA (cDNA) molecules by reverse transcribing RNA molecules within the cells or nuclei of the first plurality of aliquots, wherein the RNA molecules are reverse transcribed using reverse transcription (RT) primers each comprising: (i) a poly(T) sequence or a random sequence; and (ii) an RT

barcode sequence, wherein the RT barcode sequences present within the RT primers are specific to each aliquot within the first plurality of aliquots;

(d) pooling the cells or nuclei from the first plurality of aliquots;

(e) tagging the cDNA molecules within the pooled cells or nuclei from the first plurality of aliquots with one or more nucleic acid tags, thereby generating tagged cDNA molecules, by performing steps (i) through (iii) one or more times:

(i) dividing the pooled cells or nuclei into an additional plurality of aliquots;

(ii) coupling nucleic acid tags to cDNA molecules within the cells or nuclei of the additional plurality of aliquots, wherein each nucleic acid tag comprises a tag barcode sequence, and wherein the tag barcode sequences present within the nucleic acid tags are specific to each aliquot within the additional plurality of aliquots;

(iii) pooling the cells from the additional plurality of aliquots;

(f) lysing the cells or nuclei to release the tagged cDNA molecules and produce a lysate comprising the released tagged cDNA molecules;

(g) isolating the released tagged cDNA molecules using a binding agent, such that the isolated tagged cDNA molecules are bound to the binding agent;

(h) generating second strands of the isolated tagged cDNA molecules to produce double-stranded tagged cDNA molecules;

(i) amplifying the isolated tagged cDNA molecules in a first multiplex amplification step using a multiplex set of preamplification primers, wherein the multiplex set of preamplification primers comprises:

at least one pair of whole transcriptome (WT) preamplification primers configured to amplify all of the tagged cDNA molecules isolated from the lysate, and

at least one pair of target-specific preamplification primers configured to specifically amplify tagged target cDNA molecules isolated from the lysate,

thereby generating an enriched plurality of amplified tagged cDNA molecules that comprises the whole transcriptome and that is enriched for the one or more target cDNA molecules;

(j) dividing the enriched plurality of amplified cDNA molecules into at least a first and a second portion;

(k) preparing a whole transcriptome (WT) sequencing library using the first portion of the enriched plurality of amplified tagged cDNA molecules; and

(l) preparing a target sequencing library using the second portion of the enriched plurality of amplified tagged cDNA molecules.

[0025] In some embodiments, the preparation of the WT sequencing library comprises fragmenting the first portion of the enriched plurality of amplified tagged cDNA molecules and appending a first adapter comprising a first adapter sequence to the fragment ends, and amplifying the tagged cDNA molecules comprising the first adapter in an index PCR using WT index amplification primers, wherein one or more of the WT index amplification primers comprises a sequence complementary to the first adapter sequence, wherein one or more of the WT index amplification primers comprises an index sequence, and wherein one or more of the WT index amplification primers comprises a next-generation sequencing (NGS) adapter sequence, an NGS primer binding sequence, and/or an NGS flow-cell binding sequence.

[0026] In accordance with any of the embodiments, the preparation of the target sequencing library comprises: amplifying the tagged target cDNA molecules in the second portion of the enriched plurality of amplified tagged cDNA molecules in a second target-specific amplification step, thereby generating a further enriched plurality of tagged target cDNA molecules; wherein the tagged target cDNA molecules are amplified in the second target-specific amplification step using at least one pair of target amplification primers configured to specifically amplify the tagged target cDNA molecules in the second portion; and amplifying the further enriched plurality of tagged target cDNA molecules in an index target PCR using target index amplification primers; wherein one or more of the target index amplification primers comprises an index sequence, and wherein one or more of the target index amplification primers comprises a next-generation sequencing (NGS) adapter sequence, an NGS primer binding sequence, and/or an NGS flow-cell binding sequence.

[0027] In accordance with any of the embodiments, the nucleic acid tags that are coupled to the tagged cDNA molecules during the last of the one or more times that steps (e)(i) to (e)(iii) are repeated comprise one or more elements selected from the group consisting of a random nucleotide sequence to prevent counting of PCR duplicates, a capture agent, and a second adapter sequence.

[0028] In some embodiments, the capture agent comprises biotin.

[0029] In some embodiments, the binding agent comprises streptavidin-coated magnetic beads.

[0030] In accordance with any of the embodiments, the lysing in step (f) is performed in the presence of a protease.

[0031] In some embodiments, the protease is proteinase K.

[0032] In some embodiments, a protease inhibitor is added to the lysate prior to or together with the binding agent.

[0033] In some embodiments, the protease inhibitor is phenylmethanesulfonyl fluoride (PMSF) or 4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride (AEBSF.

[0034] In accordance with any of the embodiments, the second strands are generated in step (h) using a template switching oligo (TSO) comprising a third adapter sequence, such that the template switching introduces the third adapter sequence to the 3'-end of the released cDNA molecules.

[0035] In accordance with any of the embodiments, prior to the lysing in step (f), the plurality of cells or nuclei comprising the tagged cDNA molecules are divided into a plurality of samples.

[0036] In some embodiments, the index sequence in one or more of the WT index amplification primers and/or one or more of the target index amplification primers is sample-specific.

[0037] In accordance with any of the embodiments, at least one target-specific preamplification primer and at least one target amplification primer are complementary to the same sequence within a target cDNA molecule as a target amplification primer.

[0038] In accordance with any of the embodiments, none of the target-specific preamplification primers specifically binds to the same sequence within a target cDNA molecule as a target amplification primer.

[0039] In accordance with any of the embodiments, wherein the at least one pair of target-specific preamplification primers and the at least one pair of WT preamplification primers have at least one preamplification primer in common.

[0040] In some embodiments, the at least one preamplification primer common to both the at least one pair of target-specific preamplification primers and the at least one pair of WT preamplification primers comprises a sequence complementary to the second adapter sequence.

[0041] In accordance with any of the embodiments, the at least one pair of WT preamplification primers comprise a primer comprising a sequence complementary to the third adapter sequence and a primer comprising a sequence complementary to the second adapter sequence.

[0042] In accordance with any of the embodiments, the multiplex set of preamplification primers comprises a single pair of target-specific preamplification primers and/or a single pair of WT preamplification primers.

[0043] In accordance with any of the embodiments, the multiplex set of preamplification primers comprises from 1-10, 10-100, or 100-200 pairs of target-specific preamplification primers.

[0044] In accordance with any of the embodiments, the RT primers and/or the nucleic acid tags are DNA molecules.

[0045] In accordance with any of the embodiments, the method further comprising sequencing the WT sequencing library and the target sequencing library.

[0046] In some embodiments, the WT sequencing library and the target sequencing library are sequenced separately.

[0047] In some embodiments, the WT sequencing library and the target sequencing library are sequenced together.

[0048] In accordance with any of the embodiments, the method further comprising filtering the sequencing reads generated from the sequencing of the WT sequencing library and the target sequencing library to remove transcripts represented by only one read.

[0049] In accordance with any of the embodiments, the method further comprising grouping sequencing reads generated from the sequencing of the WT sequencing library and the target sequencing library according to one or more features selected from the group consisting of RT barcode sequences, tag barcode sequences, series or combinations of tag barcode sequences, and index sequences.

[0050] In some embodiments, the grouped sequencing reads are used to determine the individual cell or nucleus from among the plurality of cells or nuclei from which a given cDNA originated.

[0051] In accordance with any of the embodiments, the method further comprising grouping sequence reads into target cDNA sequences and non-target cDNA sequences.

[0052] In some embodiments, the method further comprising identifying one or more cells or nuclei from among the plurality of cells or nuclei in which a target cDNA is expressed, or in which the target cDNA is expressed at an increased or decreased level relative to a reference value.

[0053] In accordance with any of the embodiments, the method further comprising relating the identity and/or expression level of a target gene or sequence in an individual cell or nucleus to the overall pattern of expression of the whole transcriptome in the same individual cell or nucleus.

[0054] In accordance with any of the embodiments, the method further comprising relating the identity and/or expression level of a target gene or sequence in an individual cell or nucleus to the expression level of one or more individual non-target genes in the same original cell or nucleus.

[0055] In accordance with any of the embodiments, the one or more target cDNAs comprise one or more CRISPR guide RNA (gRNA) sequences.

[0056] In some embodiments, the gRNA sequences are from a gRNA library used in a CRISPR screen.

[0057] In accordance with any of the embodiments, at least one of the target-specific preamplification primers is specific to a Pol III promoter.

[0058] In some embodiments, the Pol III promoter is a U6 promoter.

[0059] In accordance with any of the embodiments, at least one of the target-specific preamplification primers comprises the sequence of SEQ ID NO:11.

[0060] In accordance with any of the embodiments, at least one pair of target-specific preamplification primers includes a primer comprising the sequence of SEQ ID NO:12 or SEQ ID NO:13.

[0061] In accordance with any of the embodiments, at least one of the target amplification primers is specific to a Pol III promoter.

[0062] In some embodiments, the Pol III promoter is a U6 promoter.

[0063] In accordance with any of the embodiments, at least one of the target amplification primers comprises the sequence of SEQ ID NO: 14.

[0064] In accordance with any of the embodiments, at least one of the target amplification primers comprises the sequence of SEQ ID NO: 12.

[0065] In accordance with any of the embodiments, the plurality of cells each comprises an expression construct encoding an RNA-guided nuclease, or inactivated form thereof, that is capable of physically interacting with the guide RNA and being directed to a target locus in the genome by the guide RNA.

[0066] In some embodiments, the RNA-guided nuclease is Cas9 or Cpf1.

[0067] In accordance with any of the embodiments, the target genes or sequences comprise a T cell receptor (TCR) gene or sequence.

[0068] In some embodiments, the at least one pair of target preamplification and/or the at least one pair of target amplification primers comprise a primer specific to a TCR alpha, beta, gamma, or delta chain.

[0069] In some embodiments, the at least one pair of target-specific preamplification primers and/or the at least one pair of target amplification primers comprise a primer specific to a TCR alpha chain and a primer specific to a TCR beta chain.

[0070] In accordance with any of the embodiments, at least one of the target-specific preamplification primers and/or at least one of the target amplification primers is specific to a TCR CDR3 region.

[0071] In accordance with any of the embodiments, the cells or nuclei comprise mammalian cells or nuclei.

[0072] In some embodiments, the cells or nuclei comprise human cells or nuclei.

[0073] In accordance with any of the embodiments, the cells or nuclei comprise mouse cells or nuclei.

[0074] In accordance with any of the embodiments, the cells or nuclei comprise T cells or nuclei derived therefrom.

[0075] In some embodiments, the T cells or nuclei comprise one or more cells or nuclei selected from the group consisting of chimeric antigen receptor (CAR) T cells, activated T cells, primary cells, T cells isolated from a cell line, T cells isolated from a tissue, T cells isolated from a subject, effector T cells, cytotoxic T cells, helper T cells, regulatory T cells, memory T cells, nuclei derived from any of the heretofore listed T cells, and combinations thereof.

[0076] In accordance with any of the embodiments, the first multiplex amplification of step (i) comprises amplifying the tagged cDNA molecules for from 5 to 20 cycles.

[0077] In accordance with any of the embodiments, the first multiplex amplification of step (i) is performed according to the conditions shown in Table 12.

[0078] In accordance with any of the embodiments, one or more of steps (a), (b), (d), (e)(*i*), (e)(*iii*), or (f) are carried out at a temperature of below about 8, 7, 6, 5, 4, 3, 2, 1, 0, -1, -2, -3, or -4 °C, between about -4 to 8, -4 to 0, 0 to 4, 4 to 8, or 0 to 8 °C, or at about 8, 7, 6, 5, 4, 3, 2, 1, 0, -1, -2, -3, or -4 °C.

[0079] In accordance with any of the embodiments, the cells or nuclei were fixed and/or permeabilized at 4 °C or below 4 °C.

[0080] In accordance with any of the embodiments, the nucleic acid tags are coupled to the cDNA molecules in step (e)(ii) by ligation.

[0081] In accordance with any of the embodiments, the RT primers each comprise a 5' overhang comprising a 5' overhang sequence.

[0082] In accordance with any of the embodiments, the nucleic acid tags each comprise a first strand comprising a 3' hybridization sequence and/or a 5' hybridization sequence flanking the 3' end and/or the 5' end of the tag barcode sequence, respectively.

[0083] In some embodiments, the RT primers each comprise a 5' overhang comprising a 5' overhang sequence, and wherein the nucleic acid tags each further comprise a second strand comprising: a first portion complementary to a 5' hybridization sequence of a previously coupled nucleic acid tag or a 5' overhang sequence of an RT primer; and a second portion complementary to the 3' hybridization sequence.

[0084] In accordance with any of the embodiments, the method further comprising: size selecting the enriched plurality of amplified tagged cDNA molecules subsequent to step (h) using solid phase reversible immobilization (SPRI) beads.

[0085] In accordance with any of the embodiments, the method further comprising: size selecting the further enriched plurality of tagged target cDNA molecules using solid phase reversible immobilization (SPRI) beads.

[0086] In some embodiments, the size selection using SPRI beads is single-sided.

[0087] In some embodiments, the size selection using SPRI beads is double-sided.

[0088] In accordance with any of the embodiments, one or more of the pluralities of aliquots or samples are distributed in a multi-well plate.

[0089] In some embodiments, the multi-well plate is a 96-well plate.

[0090] In some embodiments, the nucleic acid tags used in one or more of the additional pluralities of aliquots comprise 96 distinct barcode sequences.

[0091] In some embodiments, each of the 96 distinct barcode sequences is present in only one of the 96 aliquots.

[0092] In accordance with any of the embodiments, at least a subset of the wells of the multi-well plate contain primers comprising well-specific unique dual indexes (UDIs).

[0093] In some embodiments, wherein one or more of the UDIs comprise any of SEQ ID NOS: 13-302.

[0094] In one aspect, disclosed herein are kits for performing the method of any one or more of claims 1-70.

[0095] In one aspect, disclosed herein are kits for preparing multiplex sequencing libraries, the kit comprising at least one set of primers for cDNA labeling and amplification of the whole transcriptome, and at least one set of target-specific preamplification primers for cDNA labeling, amplification and enrichment of one or more target genes of interest.

[0096] In some embodiments, the kit further comprising instructions for preparing the sequencing libraries.

[0097] In accordance with any of the embodiments, the kit comprises target-specific primers comprising SEQ ID NO: 11 or SEQ ID NO:14.

DETAILED DESCRIPTION

Introduction

[0098] The present disclosure relates generally to methods of uniquely labeling or barcoding molecules within a nucleus, a plurality of nuclei, a cell, a plurality of cells, and/or one or more tissues, organs, organisms, or subjects. The present disclosure also relates to kits for uniquely labeling or barcoding molecules within a nucleus, a plurality of nuclei, a cell, a plurality of cells, and/or a tissue, organ or organism. The molecules to be labeled may include, but are not limited to, RNA molecules, cDNA molecules, DNA molecules, proteins, peptides, and/or antigens.

[0099] In various embodiments, the present disclosure provides methods and compositions for creating multiple related sequencing libraries, e.g., transcriptome sequencing libraries for multiplex analyses. For example, some embodiments, single cell whole transcriptome libraries are created that are coupled with gene- or vector-enriched libraries. By creating such libraries, individual genes or transgenes can be robustly detected across cells while limiting the overall sequencing needed, even when the individual genes or transgenes are rare among the cells.

[0100] The present methods and compositions enable the enrichment and highly sensitive detection of individual (or a small number of, e.g., up to 10, 50, 100, 200) target transcripts of interest in parallel with the whole transcriptome (or a subset thereof) in single cells. The target transcripts can correspond to any sequence of interest whose presence or expression level in a cell may be associated with other properties of interest of the cell.

[0101] For example, in some embodiments, the transcripts comprise CRISPR guide RNAs (gRNAs, or single-guide RNAs or sgRNAs), and the present methods can be used to associate the presence and/or expression level of a given gRNA in an individual cell (and therefore any expected effects on the genes targeted by the gRNA), with the whole transcriptome of same cell. Such methods enable, for example, the profiling of up to one million or more cells, while reducing the sequencing needed to detect and assign specific expression events (e.g., the expression of individual gRNAs) to individual cells in a population of cells (e.g., in a pooled single-cell CRISPR screen such as CROP-seq or similar methods). Accordingly, the present disclosure allows for complex phenotypes such as transcriptional profiles to be associated with specific genetic perturbations or expression status.

[0102] The present methods can also be used in methods involving the analysis of T cell receptor (TCR) identity in individual cells, the analyses performed using the presently

described methods can allow the detection or characterization of a disease process (e.g., via the detection and/or tracking of T cell clonotypes) or the development, preparation, or monitoring of cell therapies (e.g., therapies involving chimeric antigen receptor (CAR) T cells).

[0103] The present methods can be used for other applications as well, e.g., any of a number of applications in which it is desirable to associate the presence or expression level of a target with the whole transcriptome in single cells, e.g., for the validation of biomarkers, for the validation of drug targets, for lineage tracing, or for Massively Parallel Reporter Assays (MPRAs).

[0104] It will be readily understood that the embodiments, as generally described herein, are exemplary. The following more detailed description of various embodiments is not intended to limit the scope of the present disclosure, but is merely representative of various embodiments. Moreover, the order of the steps or actions of the methods disclosed herein may be changed by those skilled in the art without departing from the scope of the present disclosure. In other words, unless a specific order of steps or actions is required for proper operation of the embodiment, the order or use of specific steps or actions may be modified.

Definitions

[0105] As will be understood by one of ordinary skill in the art, each embodiment disclosed herein can comprise, consist essentially of, or consist of its particular stated element, step, ingredient, or component. As used herein, the transition term "comprise" or "comprises" means includes, but is not limited to, and allows for the inclusion of unspecified elements, steps, ingredients, or components, even in major amounts. The transitional phrase "consisting of" excludes any element, step, ingredient or component not specified. The transition phrase "consisting essentially of" limits the scope of the embodiment to the specified elements, steps, ingredients or components, and to those that do not materially affect the embodiment.

[0106] Unless otherwise indicated, all numbers expressing quantities of ingredients, properties such as molecular weight, reaction conditions, and so forth used in the specification and claims are to be understood as being modified in all instances by the term "about." Accordingly, unless indicated to the contrary, the numerical parameters set forth in the specification and attached claims are approximations that may vary depending upon the desired properties sought to be obtained by the present disclosure. At the very least, and not as an attempt to limit the application of the doctrine of equivalents to the scope of the claims, each numerical parameter should at least be construed in light of the number of reported significant digits and by applying

ordinary rounding techniques. When further clarity is required, the term "about" has the meaning reasonably ascribed to it by a person skilled in the art when used in conjunction with a stated numerical value or range, i.e., denoting somewhat more or somewhat less than the stated value or range, to within a range of, e.g., ±20% of the stated value; ±19% of the stated value; ±18% of the stated value; ±17% of the stated value; ±16% of the stated value; ±15% of the stated value; ±14% of the stated value; ±13% of the stated value; ±12% of the stated value; ±11% of the stated value; ±10% of the stated value; ±9% of the stated value; ±8% of the stated value; ±7% of the stated value; ±6% of the stated value; ±5% of the stated value; ±4% of the stated value; ±3% of the stated value; ±2% of the stated value; or ±1% of the stated value.

[0107] The terms "a," "an," "the" and similar referents used in the context of describing the disclosure (especially in the context of the following claims) are to be construed to cover both the singular and the plural, unless otherwise indicated herein or clearly contradicted by context. Recitation of ranges of values herein is merely intended to serve as a shorthand method of referring individually to each separate value falling within the range. Unless otherwise indicated herein, each individual value is incorporated into the specification as if it were individually recited herein. All methods described herein can be performed in any suitable order unless otherwise indicated herein or otherwise clearly contradicted by context. The use of any and all examples, or exemplary language (e.g., "such as") provided herein is intended merely to better illuminate the disclosure and does not pose a limitation on the scope of the disclosure otherwise claimed. No language in the specification should be construed as indicating any non-claimed element essential to the practice of the disclosure.

[0108] Notwithstanding that the numerical ranges and parameters setting forth the broad scope of the disclosure are approximations, the numerical values set forth in the specific examples are reported as precisely as possible. Any numerical value, however, inherently contains certain errors necessarily resulting from the standard deviation found in their respective testing measurements.

[0109] The terms "polynucleotide" and "nucleic acid," used interchangeably herein, refer to a polymeric form of nucleotides of any length, either ribonucleotides or deoxynucleotides. Thus, this term includes, but is not limited to, single-, double-, or multi-stranded DNA or RNA, genomic DNA, cDNA, DNA-RNA hybrids, or a polymer comprising purine and pyrimidine bases or other natural, chemically or biochemically modified, non-natural, or derivatized nucleotide bases. The terms "polynucleotide" and "nucleic acid" should be understood to

include, as applicable to the embodiment being described, single-stranded (such as sense or antisense) and double-stranded polynucleotides. Unless specifically limited, the term encompasses nucleic acids containing known analogs or derivatives of natural nucleotides, e.g., molecules that have similar binding properties as the reference nucleic acid. In some embodiments, the nucleic acids can comprise one or more modified nucleotides, e.g., nucleic acids modified at the base moiety, at the sugar moiety, or at the phosphate backbone (e.g., phosphorothioates). In some embodiments, the nucleic acids can comprise one or more moieties to allow or facilitate, e.g., detection, quantification, purification, capture, identification, or selective removal, e.g., biotin, fluorescent labels, etc.

[0110] The term "gene" refers to the segment of DNA involved in producing a polypeptide chain or a non-coding transcript (e.g., mRNA). For coding sequences, it may include regions preceding and following the coding region (leader sequence and/or trailer sequence) as well as intervening sequences (introns) between individual coding segments (exons). A "transgene" refers to a gene that has been introduced into a cell or organism from another source (e.g., from another organism or following synthesis).

[0111] The terms "hybridizable" or "complementary" or "substantially complementary" it is meant that a nucleic acid (e.g. RNA) comprises a sequence of nucleotides that enables it to non-covalently bind, i.e. form Watson-Crick base pairs and/or G/U base pairs, "anneal", or "hybridize," to another nucleic acid in a sequence-specific, antiparallel, manner (i.e., a nucleic acid specifically binds to a complementary nucleic acid) under the appropriate in vitro and/or in vivo conditions of temperature and solution ionic strength. As is known in the art, standard Watson-Crick base-pairing includes: adenine (A) pairing with thymidine (T), adenine (A) pairing with uracil (U), and guanine (G) pairing with cytosine (C) [DNA, RNA]. In addition, it is also known in the art that for hybridization between two RNA molecules (e.g., dsRNA), guanine (G) base pairs with uracil (U). For example, G/U base-pairing is partially responsible for the degeneracy (i.e., redundancy) of the genetic code in the context of tRNA anti-codon base-pairing with codons in mRNA. In the context of this disclosure, a guanine (G) of a protein-binding segment (dsRNA duplex) of a subject DNA-targeting RNA molecule is considered complementary to a uracil (U), and vice versa. As such, when a G/U base-pair can be made at a given nucleotide position a protein-binding segment (dsRNA duplex) of a subject DNA-targeting RNA molecule, the position is not considered to be non-complementary, but is instead considered to be complementary. As used herein, the terms "hybridize" or "complementary"

refer to a first nucleotide sequence capable of forming non-covalently bind (hydrogen bond) with at least a portion of a specified second nucleotide sequence.

[0112] A "promoter" refers to a set of nucleic acid sequences that direct the transcription of a nucleic acid, e.g., an adjacent coding sequence. Promoters can be constitutive or inducible. As used herein, a promoter includes necessary nucleic acid sequences near the start site of transcription, such as, in the case of a polymerase II type promoter, a TATA element. Promoters as used herein can include bacterial promoters or eukaryotic promoters including RNA polymerase II (e.g., EF-1 alpha) and RNA polymerase III (e.g., U6) promoters. A promoter can also include distal enhancer or repressor elements. The promoter can be a heterologous promoter (i.e., not naturally linked to the coding sequence) or homologous (i.e., the promoter that naturally drives the expression of the transcribed sequence).

[0113] The term “binding” or “coupling” is used broadly throughout this disclosure to refer to any form of attaching or coupling two or more components, entities, or objects. For example, two or more components may be bound to each other via chemical bonds, covalent bonds, non-covalent bonds, ionic bonds, hydrogen bonds, electrostatic forces, Watson-Crick hybridization, nucleic acid sequence complementarity, etc.

[0114] An “expression cassette” is a nucleic acid construct, generated recombinantly or synthetically, with a series of specified nucleic acid elements that permit transcription of a particular polynucleotide sequence in a host cell. An expression cassette may be part of a plasmid, viral genome, or nucleic acid fragment. Typically, an expression cassette includes a polynucleotide to be transcribed (e.g., a protein coding sequence or a non-coding RNA such as a guide RNA), operably linked to a promoter. The promoter can be a heterologous promoter, i.e., a promoter not naturally linked to the transcribed sequence.

[0115] A “barcode” or “index” refers to a nucleotide sequence (the “barcode sequence” or “index sequence”) that is used to label an entity such as a cell, plurality of cells, cell populations, cell compartments, nucleic acids, polypeptides, or other molecules, and that varies among or between cells, cell populations, nucleic acids or other molecules, etc. For example, in some embodiments, a barcode is used to label (or tag) cDNA s generated within a given aliquot of cells, e.g., where all of the cDNA s labeled in the aliquot receive the same barcode, or receive a set of barcodes that is specific to the aliquot, i.e., that the specific set of barcodes used in the aliquot is different from the sets used in the other aliquots. Barcodes can be added to polynucleotides (or other molecules) in any of a number of ways. For polynucleotides, for

example, they can be introduced, e.g., in a primer, template, template-switch oligonucleotide (TSO), or other polynucleotide used during a polymerization-based reaction such as reverse transcription, PCR, or other polymerization-based and/or amplification reaction; barcodes can also be added to polynucleotides by hybridization and/or by ligation, e.g., by ligation of an adaptor or other polynucleotide (e.g., a "nucleic acid tag") comprising a barcode. Such adaptors or other barcode-comprising polynucleotides can be appended to a polynucleotide, e.g., by ligation via blunt-end ligation, ligation to compatible restriction ends, ligation to A-tailed or otherwise tailed ends, using a linker strand, etc. The barcode, or adapter or other polynucleotide comprising the barcode, can be single-stranded, double-stranded, partially double-stranded and partially single-stranded (e.g., comprising one or more overhangs at the 3' and/or 5' ends), etc.

[0116] As used herein, when a polynucleotide is said to comprise a "barcode" (or the equivalent term "barcode sequence") it means that the polynucleotide comprises a sequence of nucleotides that can be used to distinguish the polynucleotide comprising the barcode from one or more other polynucleotides, e.g., from polynucleotides originating from another cell, from polynucleotides labeled in a different aliquot or well, or from all other polynucleotides in a sample. In some embodiments, the barcode alone is sufficient to distinguish the polynucleotide from other polynucleotides, whereas in other embodiments the barcode provides information that can contribute to distinguishing the polynucleotide from other polynucleotides, but is not sufficient on its own (e.g., one or more additional sequence elements, or other markers, are also needed to completely distinguish the polynucleotide).

[0117] It will be appreciated that a "barcode" can refer to a single sequence of contiguous nucleotides, or to a combination of individual sequences of contiguous nucleotides. For example, in certain split-pool labeling methods as described in more detail elsewhere herein, multiple rounds of tagging can be performed, e.g., multiple rounds in each of which cells are divided into aliquots, nucleic acid tags (comprising a barcode) are added to molecules (such as cDNA s) in the cells of each aliquot, and the cells are then recombined (or repooled). As such, the tagged cDNA s in the cells of the aliquots can comprise two nucleic acid tags, each comprising a barcode. In some embodiments, the two barcodes present on the same molecule can be referred to herein as a single "barcode," even if there are other sequence elements (such as linker sequences, adapter sequences, primer-binding sequences, etc.) intervening between the two barcodes on the molecule.

[0118] Further, it will be appreciated that when a polynucleotide is said to comprise a "barcode," this can mean that, depending on the context, the specific sequence of the barcode (or "barcode sequence") can vary between the different polynucleotides comprising the barcode, or that the specific barcode sequence is the same between the different polynucleotides. For example, in some embodiments, all of the cDNA s in cells of a given aliquot (or sample) are tagged with nucleic acid tags comprising the same barcode sequence, whereas the cDNA s of cells in other aliquots are tagged with nucleic acid tags comprising other barcode sequences. In some embodiments, however, a polynucleotide comprising a barcode is added to molecules in a given aliquot (or sample), wherein the specific barcode sequence differs between the different polynucleotides used in the aliquot or sample; such polynucleotides and barcodes can be used, for example, to distinguish between the different original molecules in the sample (e.g., to be able to detect errors arising during amplification of the original molecules, such as cDNA s derived from original mRNA molecules).

[0119] In some embodiments of the present disclosure, one or more barcodes is included in primers used for reverse transcription (i.e., RT primers) or amplification (e.g., PCR), or in a nucleic acid tag appended to a polynucleotide, e.g., by ligation, where the cells are divided into two or more (e.g., 2, 4, 8, 12, 16, 24, 32, 48, 96, or more) aliquots or wells prior to the reverse transcription, amplification, or ligation, and the barcode sequences used in the cells are aliquot- or well-specific. When barcode or index sequences are said to be "aliquot-specific" or "well-specific," this means that there is an association between the sequences used and the presence of the different cells within the two or more aliquots or wells, such that the association can be used to derive information about the location of a given cell within the aliquots or wells based upon the specific sequence. For example, in some embodiments each cell within a given aliquot or well has primers or tags with the same barcode sequence, and the barcode sequences are different between each aliquot or well. However, it will be appreciated that even where a less direct relationship exists between the barcode sequences and the aliquots or wells (e.g., where more than one barcode sequences are used within a given aliquot or well, or where more than one aliquot or well share one or more barcode sequences), the barcodes are still considered aliquot-specific or well-specific, so long that some information can be derived from the barcode sequence about the aliquot or well in which a given cell or nucleus was present.

[0120] "Split-pool labeling" or "split-pool barcoding" or "combinatorial labeling" or "combinatorial barcoding" refers to a cell-specific labeling method involving the use of fixed and permeabilized cells or nuclei as containers, wherein a plurality of the cells or nuclei are

first separated into multiple wells (or aliquots), followed by the labeling of RNA or other molecules within each cell or nucleus using a well-specific tag or barcode, followed by the pooling of the cells, and wherein this cycle of separation, tagging, and pooling is repeated one or more times. In this way, at the end of the process each cell or nucleus within the plurality will comprise a combination of tags or barcodes that will reflect the particular combination of wells or aliquots in which it was present throughout the multiple rounds of tagging. As the number of rounds of tagging and/or the number of wells or aliquots used in each round is increased, the number of potential barcode combinations increases correspondingly. As such, for a given number of cells or nuclei in the plurality a suitable experimental design can be prepared that will generate a high likelihood that tagged molecules, e.g., cDNA s, within each cell or nucleus will have the same combination of barcodes that is unique among the overall population of cells or nuclei. Examples of split-pool labeling methods are disclosed, e.g., in US Patent Nos. 10,900,065, 11,634,751, 11,168,355, 11,427,856, 11,555,216, 11,639,519, 11,680,283, 10,633,648, 11,421,221, US Pat. App. Pub. No. US 2021/0388415 A1, in Rosenberg et al., Science 360, 176-182 (2018), Rosenberg et al., BioRxiv (2017), "Scaling single cell transcriptomics through split pool barcoding," doi.org/10.1101/105163, Tran et al. BioRxiv (2022) "High sensitivity single cell RNA sequencing with split pool barcoding," doi.org/10.1101/2022.08.27.505512, the entire disclosures of all of which are herein incorporated by reference (including all supplemental material).

[0121] "CRISPR/Cas9" refers to a class of bacterial systems for defense against foreign nucleic acids. CRISPR-Cas systems are found in a wide range of bacterial and archaeal organisms. CRISPR-Cas systems fall into two classes with six types, I, II, III, IV, V, and VI, with Class 1 including types I, III, and IV CRISPR systems, and Class 2 including types II, V, and VI. Class 2 CRISPR-Cas systems include individual Cas proteins that carry out multiple functions including spacer acquisition, RNA processing from the CRISPR locus, target identification, and cleavage of target nucleic acids, e.g., Cas9 in the case of type II systems and Cas12a (or Cpf1) in the case of type V systems. Any suitable Cas protein or protein assembly can be used in the present methods, i.e., any Cas that can associate with a guide RNA (gRNA), be directed to a target sequence as defined by the gRNA, and, e.g., cleave or otherwise degrade the target sequence, inhibit or activate transcription at the target sequence (in the case of CRISPRi or CRISPRa), etc. An exemplary Cas protein is the Streptococcus pyogenes Cas9 polypeptide (SpyCas9).

[0122] A "guide RNA" (gRNA, or sgRNA) refers to an RNA sequence that comprises a constant, scaffold sequence (tracrRNA) that interacts with the Cas protein (e.g., Cas9), and a variable sequence (crRNA) that defines the target sequence of the nuclease.

[0123] As used herein, the term "tagged cDNA molecules" refers to complementary DNA (cDNA) molecules comprising one or more barcodes (e.g., well-specific barcodes), e.g., cDNA molecules generated within cells or nuclei to which a nucleic acid tag has been appended (e.g., by ligation). The term "tagged target cDNA molecules" refers to tagged cDNA molecules (or transcripts) expressed from a "target gene" or comprising a "target sequence" of interest. Target cDNAs can refer to any transcript whose particular identity, whose presence, or whose level of expression can potentially vary between cells or nuclei, and where it may be of interest to correlate the identity, presence, and/or expression level of the transcript with the whole transcriptome in individual cells or nuclei. Examples of target transcripts are guide RNAs (gRNAs), e.g., as used in the context of a CRISPR screen, or T cell receptors (TCR), where different clonotypes may display associations with, e.g., disease patterns or other physiological or biological features of interest as reflected in the whole transcriptome in single cells.

[0124] As used herein, "preamplification" refers to a round of multiplex PCR performed on the whole transcriptome (WT), using WT primers (or "WT preamplification primers") configured to amplify the whole transcriptome, e.g., primers annealing to adapter sequences (e.g., TSO or R2 sequences) present on all tagged cDNAs generated using the present combinatorial barcoding methods (see, e.g., FIGS. 2A-2C), together with "spiked-in" target-specific or "gene-specific" primers (or target or gene-specific "preamplification primers") configured to amplify target cDNAs, e.g., one or more primers capable of annealing to target cDNA sequences and amplifying target cDNA molecules, either alone or together in combination with a generic primer (e.g., one target specific primer and one R2 primer; see, e.g., FIG. 2B), such that target cDNAs can become enriched during the multiplex amplification relative to non-target cDNAs within the whole transcriptome. When a primer is said to be "target-specific," this can indicate that the primer is complementary to a sequence present in all target cDNAs (e.g., a common sequence such as a U6 primer sequence present in all gRNAs used in a CRISPR library screen) or that it is complementary to a sequence present in one target only (e.g., in a transcript expressed from a specific gene of interest). In all cases, however, a "target-specific" preamplification primer is not configured to bind specifically to non-target cDNAs in the whole transcriptome.

Multiplex analysis methods in single cells or nuclei

[0125] One aspect of the present disclosure relates to methods of labeling nucleic acids, e.g., labeling nucleic acids in a cell- or nucleus-specific manner, such that the nucleic acids within the cells or nuclei each comprise a cell-specific label (or nucleus-specific label when nuclei are used instead of cells). In various embodiments, the methods comprise labeling nucleic acids in one or more cells or nuclei in order to prepare two (or more) sequencing libraries in parallel, with one sequencing library corresponding to the whole transcriptome (or a subset thereof), and one sequencing library corresponding to one or more target sequences (e.g., target genes, target transcripts, or derivatives thereof) of interest. Because both libraries comprise tagged cDNA molecules originating from the same plurality of cells, and because the tagged cDNA molecules in the two libraries collectively share the same cell- or nucleus-specific labels (i.e., each cDNA molecule has a combination of barcode sequences that can be used to identify the individual cell from which the cDNA originated), the present methods can be used to associate the presence or expression level of individual target sequences with the overall pattern of expression of the whole transcriptome in the individual cells.

[0126] The target sequences can be any target of interest, e.g., any target whose particular identity, whose presence, or whose level of expression can vary between cells, and where the identity, presence, and/or expression level has the potential to affect, or be correlated with an effect on, the whole transcriptome or a subset thereof. For example, the target may be a guide RNA (gRNA) used in the context of a CRISPR screen, where the method is used to determine the specific impact of each individual gRNA being evaluated in the screen on the overall expression in the cell. Alternatively, the target may be a potential biomarker or drug target, and the method is used to identify individual markers or transcripts whose alteration has an impact on the overall transcription in the cell, or on the transcription of particular subsets of genes involved, e.g., in a pathway, physiological process, or disease-related phenomenon or property of interest. In some embodiments, the target may be a gene whose identity varies between cells, such as a T cell receptor (TCR), where the present methods allow the association of individual TCR clonotypes with the overall pattern of the whole transcriptome (or a subsets thereof) in the same cells.

[0127] As such, the present disclosure provides a method of labeling nucleic acids for multiplex transcriptional analysis in a plurality of cells or nuclei, the method comprising providing a plurality of fixed and permeabilized cells or nuclei, each comprising a plurality of RNA molecules and one or more target genes or transcripts of interest; dividing the plurality

of cells or nuclei into a first plurality of aliquots, wherein each aliquot comprises more than one cell or nucleus; generating cDNA molecules by reverse transcribing RNA molecules within the cells or nuclei of the first plurality of aliquots using reverse transcription (RT) primers each comprising: (i) a poly(T) sequence or a random sequence; and (ii) an aliquot-specific RT barcode sequence; pooling the cells or nuclei from the first plurality of aliquots; using one or more rounds of split-pool barcoding to cell-specifically (or nucleus-specifically) tag the cDNA molecules with one or more nucleic acid tags; lysing the cells or nuclei to release the tagged cDNA molecules; isolating the released tagged cDNA molecules; generating second strands of the released tagged cDNA molecules; amplifying the released tagged cDNA molecules in a first preamplification step using preamplification primers configured to non-specifically amplify tagged cDNA molecules in the lysate and preamplification primers configured to specifically amplify tagged target cDNA molecules; dividing the plurality of preamplified cDNA molecules into at least a first and a second portion; non-specifically amplifying the preamplified cDNA molecules in the first portion; and specifically amplifying preamplified target cDNA molecules in the second portion.

[0128] It will be appreciated that, as used herein, when a method is said to, e.g., label or tag nucleic acids "in" or "within" a cell or nucleus, this means that at least one of the labeling steps takes place at the interior of the cell or nucleus (e.g., at least a first step or a first set of steps), but does not necessarily mean that all labeling or tagging steps take place at the interior of the cell or nucleus. For example, in some embodiments of the present methods, one or more tagging steps, such as those involving reverse transcription to generate cDNA molecules and the subsequent coupling of one or more nucleic acid tags to the cDNA molecules, may take place at the interior of the cells or nuclei, and one or more subsequent steps, such as template switching and preamplification/amplification steps, may be performed on tagged cDNA molecules isolated from the cells or nuclei following their lysis.

Split-pool barcoding

[0129] The present methods involve labeling cDNA molecules with barcodes in a process called, alternatively, split-pool labeling, tagging, or barcoding, or combinatorial labeling, tagging, or barcoding. In various embodiments, the split-pool barcoding step of the protocol may be repeated a number of times sufficient to generate a unique combination or series of labeling sequences for the cDNAs in each sequencing library such that all (or virtually all, or the great majority) of the cDNAs originating from a given cell (or nucleus) will have the same combination or series of labeling sequences (also referred to as barcode sequences or index

sequences), and that the complexity of the combinations or series of labeling sequences is such that each combination or series is unique, or essentially unique, among all of the cells in the plurality of cells. For example, the labeling could be repeated enough times to generate a sufficient number of distinct combinations or series of labeling sequences that each individual combination or series has, e.g., at least a 95%, 96%, 97%, 98%, 99%, or higher probability of being unique among all of the combinations or series in the cells of the plurality. Stated another way, the split-pool barcoding may be repeated a number of times such that the cDNAs in the first cell may have a first unique series of labeling sequences, the cDNA s in a second cell may have a second unique series of labeling sequences, the cDNAs in a third cell may have a third unique series of labeling sequences, and so on. The methods of the present disclosure may provide for the labeling of cDNA sequences from single cells with unique barcodes, wherein the unique barcodes may identify or aid in identifying the cell from which the cDNA originated. In other words, a portion, a majority, or substantially all of the cDNA from a single cell may have the same barcode, and that barcode may not be repeated in cDNA originating from one or more other cells in a sample (e.g., from a second cell, a third cell, a fourth cell, etc.).

[0130] The barcodes used in the present methods can be added to the cDNA molecules at any of a number of steps, e.g., during reverse transcription, during one or more rounds of ligation-based tagging (e.g., appending a nucleic acid tag to a cDNA molecule), or during amplification (e.g., introduced via a PCR primer). The barcodes introduced at any of the steps can be any length, e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 nucleotides or longer. In some embodiments, the barcodes are at least 8 nucleotides long. In some embodiments, the barcodes are 8 nucleotides long. In some embodiments, the barcodes are greater than 8 nucleotides long. The length of any given barcode used in a given embodiment of the method (e.g., a barcode sequence in a first nucleic acid tag) can be independent of the length of a different barcode used in the embodiment (e.g., a barcode in a second nucleic acid tag). For example, in some embodiments, a RT primer barcode, a barcode in a first nucleic acid tag, a barcode in a second nucleic acid tag, and/or a barcode or index sequence in a PCR primer may all be the same length, whereas in some embodiments they may all be different lengths, while in other embodiments some barcodes may be the same length while some are of different lengths.

[0131] In some embodiments, barcoded (or tagged) cDNA molecules are mixed together and sequenced (e.g., using NGS), such that data can be gathered regarding RNA expression at the level of a single cell. In some embodiments, cDNA molecules are barcoded (or tagged)

together, but then sequenced independently. In either case, so long as the cDNA molecules originated from the same original cells (or nuclei) and were labeled together such that all of the cDNA molecules originating from a given original cell or nucleus comprise the same combination of barcodes, the sequencing data can be combined and analyzed together, even if certain cDNA molecules from the same individual cells are present in different sequencing libraries and/or sequenced independently. Certain embodiments of the methods of the present disclosure may be useful in assessing, analyzing, or studying the transcriptome (i.e., the different RNA species transcribed from the genome of a given cell) of one or more individual cells.

[0132] As discussed above, for any one or more steps of the present methods, including, but not limited to, reverse transcription steps, nucleic acid tag coupling steps, cell lysis steps, template switching steps, amplification steps, or enrichment steps, an aliquot or group of cells (or nuclei, or lysates) can be separated into different reaction vessels or containers. Vessels or containers can also be referred to herein as receptacles, samples, and wells, and the terms vessel, container, receptacle, sample, and well may be used interchangeably herein. In some embodiments, cells or nuclei may be separated into a number of different reaction vessels. For example, the number of reaction vessels may include four 1.5 ml microcentrifuge tubes, a plurality of wells of a 96-well plate, or another suitable number and type of reaction vessels. In some embodiments, the reaction vessels or containers include one or more 96-well plates (or, e.g., 6, 12, 24, 48, 384 well plates).

[0133] For combinatorial barcoding (or split-pool tagging), cells or nuclei can be distributed into a plurality of aliquots and polynucleotides within the cells or nuclei labeled with an aliquot-specific barcode (e.g., by reverse transcription of RNA within the cells or nuclei using primers comprising the barcode, or by appending a nucleic acid tag to polynucleotides within the cells or nuclei wherein the tags comprise aliquot-specific barcodes), the aliquots can then be repooled, washed, and separated again into a new plurality of aliquots, and a further set of barcodes can be added to the polynucleotides. In this way, after repeated rounds of separating, tagging, and repooling, cDNAs or other polynucleotides within each cell or nucleus may be bound to a unique combination or sequence of barcodes, or substantially unique combination or sequence of barcodes. In some embodiments, all (or most, depending, e.g., on the efficiency of the tagging reactions in a given cell or nucleus) of the cDNA molecules or other polynucleotides within any individual cell or nucleus within a plurality of cells or nuclei will comprise the same combination of barcodes (or barcode sequences). In some embodiments, the

combination or sequence of barcodes can be used to identify, or help identify, the individual cell from which a given tagged cDNA molecule originated.

[0134] In some embodiments, in a given barcoding step cells or nuclei within each well or aliquot are tagged with a different barcode, i.e., all of the barcodes (or barcode sequences) used within the well or aliquot are the same, while the barcode sequences are different in each of the wells or aliquots. However, other barcoding strategies are possible as well, e.g., in which more than one barcode sequence is used within a given well or aliquot, or in which one or more barcode sequences are present in multiple wells or aliquots during the barcoding step. In general, any barcoding protocol can be encompassed by the present disclosure so long that during the protocol the labeled molecules (e.g., RNA molecules, or cDNA molecules produced therefrom) within each cell or nucleus acquire a combination of barcodes that reflects the different wells or aliquots in which the cell or nucleus was present.

[0135] The different labeling sequences can be introduced at one or more steps, including during reverse transcription (e.g., wherein each reverse transcription (RT) primer comprises a barcode), during one or more subsequent labeling steps (e.g., ligating, tagmentation, or otherwise coupling a nucleic acid tag comprising a barcode sequence to a cDNA), or during one or more amplification steps (e.g., using one or more primers that include a barcode or index sequence). For example, in some embodiments of the present disclosure, RNA is labeled within cells or nuclei by generating cDNA through reverse transcription (RT) using well-specific barcode-containing primers, and subsequently additional well-specific barcodes are ligated to the cDNA molecules in one or more round of split-pool tagging, and finally yet more barcodes (or indexes) are added to the cDNA molecules during amplification using well- or sample-specific barcoded primers (e.g., unique dual indexes or UDIs). Accordingly, the number of possible barcode combinations can vary by, e.g., increasing or decreasing the number of wells used for reverse transcription, for ligation-based tagging, and/or for indexing during amplification, and/or by changing the number of total barcoding steps, e.g., by varying the number of rounds of split-pool tagging or by omitting barcodes in one or more steps (e.g., by performing RT and/or amplification using non-barcoded primers, or by omitting the ligation tagging steps and/or amplification indexing steps altogether).

[0136] In certain embodiments, steps of the present methods in which a nucleic acid tag is appended or coupled to a cDNA or other polynucleotide within a cell or nucleus may be repeated one or more times, e.g., 1, 2, 3, 4, 5 times, or 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14,

15, 20, 25, 30, 35, 40, 45, 50, 100, or more times. In certain other embodiments, the steps are repeated a sufficient number of times such that the cDNA s of each cell or nucleus would be likely to be bound to a unique barcode (e.g., unique among the plurality of cells or nuclei, or among multiple pluralities of cells or nuclei, e.g., in situations where multiple pluralities may be sequenced together). The number of times may be selected to provide a greater than 50% likelihood, greater than 90% likelihood, greater than 95% likelihood, greater than 99% likelihood, or some other probability that the cDNA s in each cell are bound to a unique barcode.

[0137] The number of total possible barcode combinations in the population will be a function of the number of barcode tagging rounds that are performed, and on the number of different barcodes/aliquots included in each round. The total number of possible barcode combinations can be achieved in any of a number of ways. In some embodiments, the number of total possible barcode combinations is greater than the number of different cells in the population, e.g., such that the probability that a given combination of barcodes is unique among all of the cells of the plurality is, e.g., 95%, 96%, 97%, 98%, 99%, or higher. Accordingly, the present methods could include, e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, or more individual rounds of barcoding.

Cells or nuclei

[0138] The present methods can be used to specifically label molecules, e.g., nucleic acids, in any of a wide variety of cell types. Cells suitable for use in the present methods include, e.g., primary cells, cell lines (e.g., HEK293, HEK293T, HEK293F, NIH3T3, Jurkat cells, or others), cells isolated from an organism, organoid, or a tissue, isolated blood cells, and others. In some embodiments, the cells are healthy cells (e.g., wild-type or control cells). In some embodiments, the cells are disease cells (e.g., cancer cells, infected cells). In some embodiments, the cells are stem cells. In various embodiments, the plurality of cells may be eukaryotic cells, vertebrate cells, mammalian cells, human cells, mouse cells, insect cells, plant cells, fungal cells, yeast cells, or bacterial cells.

[0139] In some embodiments, the cells comprise one or more cell types such as blood cells (e.g., peripheral blood mononuclear cells or PBMCs, immune cells such as T cells, B cells, NK cells), brain cells, liver cells, gut cells, bone marrow cells, pancreatic cells, epithelial cells, endothelial cells, neuronal cells, fibroblast cells, bone cells, muscle cells, skin cells, fat cells, lymphocytes, myeloid cells, macrophages, stem cells, and others. In some embodiments, the

cells are all from a single source, i.e., from a single individual, organism, or tissue. In some embodiments, the cells are from multiple sources, i.e. from multiple individuals, organisms, or tissues. In some embodiments, the cells are autologous cells. In some embodiments, the cells are allogeneic cells. In some embodiments, the cells are all or primarily comprise a single cell type (e.g., from a cell line, or a specific cell type isolated from a primary sample). In some embodiments, the cells comprise a mixture of different cell types. In some embodiments, the cells are adherent cells. In some embodiments, the cells are suspension cells.

[0140] In some embodiments, the cells have been previously frozen. In some embodiments, the cells have been previously fixed and frozen, e.g., the methods are performed using multiple samples that have been fixed and/or frozen at different times. In some embodiments, the cells have been previously fixed, permeabilized, and frozen.

[0141] It will be appreciated that in all of the herein-disclosed embodiments, either cells or nuclei obtained from the cells can be used. Nuclei can be prepared, e.g., using standard methods such as by douncing. In some embodiments, nuclei are prepared from frozen cells, tissue samples, or tissue slices or sections. Nuclei can be prepared, e.g., by placing the frozen sample (cells, tissue, minced tissue sample, slice, section, etc.) into a cooled nuclei isolation (NIM) buffer solution (e.g., NIM1 or NIM2 buffer), then transferred to a dounce and homogenized, e.g., using a pestle (e.g., 10 strokes each with a loose and with a tight pestle). The homogenate can then be filtered (e.g., using a 40 um or 70 um filter) and transferred to, e.g. conical tubes. The tube can then be centrifuged, e.g., 200x or 500x g in a pre-cooled swinging bucket centrifuge for, e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, or more minutes at a low temperature, e.g., 4 °C, or at 1 °C, 2 °C, 3 °C, 4 °C, 5 °C, 6 °C, 7 °C, or 8 °C, or at a temperature less than 8 °C, 7 °C, 6 °C, 5 °C, 4 °C, 3 °C, 2 °C, or 1 °C. In some embodiments, the cells are counted before and/or after centrifugation, e.g., to ensure an appropriate number of cells in each aliquot or tube. The pellets can then be resuspended in an appropriate solution or buffer, e.g., a nuclei buffer containing BSA (e.g., 0.75% BSA), and subsequently fixed and stored, e.g., at -80 °C.

[0142] The present methods allow the labeling and multiplex analysis of cells at a broad range of scales. For example, in some embodiments, the methods are used to label up to 10,000 cells. In some embodiments, the methods are used to label up to 100,000 cells or nuclei. In some embodiments, the methods are used to label up to 1,000,000 cells or nuclei (see, e.g., FIGs. 3A-3D).

[0143] FIGS. 3A-3D provide an overview of *in situ* cell barcoding steps (i.e., combinatorial barcoding, or split-pool labeling) according to embodiments of the present methods. FIG. 3A illustrates Round 1 Barcoding. Fixed and permeabilized cells were loaded into multiple wells of a Round 1 plate. The image shows cells distributed into 48 wells of the Plate, e.g., for labeling up to 100,000 cells. In other embodiments, different numbers of wells are used, e.g., labeling up to 10,000 cells distributed into 8 wells of the plate, or labeling of up to 1,000,000 cells distributed into all 96 wells. In some embodiments, some or all of the different wells (e.g., the 48 wells shown in FIGs. 3A-3C) can correspond to different biological samples or experimental conditions. In other embodiments, the wells contain cells originating from the same samples or experimental conditions. RNA was reverse transcribed via oligo dT and random hexamer primers with a well-specific barcode that are associated with specific samples. FIG. 3B illustrates Round 2 Barcoding. The cells were pooled and loaded into the Round 2 Plate. An adapter with a well-specific barcode was ligated to the first barcode. FIG. 3C illustrates Round 3 Barcoding. The cells were pooled and loaded into the Round 3 Plate. A third barcode was ligated to the cDNA, which also contains an Illumina R2 sequence, and biotin. FIG. 3D illustrates Lysis and Sublibrary Generation. Cells were split into multiple sublibraries (or "samples") and lysed. As with the varying number of wells used for reverse transcription, the number of sublibraries (or samples) can vary according to the number of cells used, e.g., 8 sublibraries with up to 100,000 cells, or 2 sublibraries with up to 10,000 cells, or 16 sublibraries with up to 1,000,000 cells.

[0144] In embodiments of the present methods, the numbers of cells labeled and analyzed are at least 10,000, 20,000, 30,000, 40,000, 50,000, 60,000, 70,000, 80,000, 90,000, 100,000, 200,000, 300,000, 400,000, 500,000, 600,000, 700,000, 800,000, 900,000, 1,000,000, 1,500,000, 2,000,000, 3,000,000, or more.

[0145] In some embodiments, the cells have been genetically modified, e.g., modified using CRISPR to introduce a transgene, delete a gene, locus, or genomic region, or modify the sequence of a gene, locus, or genomic region, e.g., through inducing indels (small insertions or deletions) at the gRNA target site. In some embodiments, the cells have been modified to modulate the expression of one or more endogenous genes, e.g., using CRISPRi, CRISPRa, RNAi, etc. In some embodiments, the modified cells express transgenes, e.g., coding sequences encoding polypeptides with potential activity of interest, e.g., candidate drug targets or biomarkers. Accordingly, in some embodiments, the present methods are used in assays for, e.g., validating biomarkers and/or drug targets. In some embodiments, the modified cells

comprise candidate regulatory sequences and encode reporter genes, e.g., in massively parallel reporter assays (MPRAs). In some embodiments, the present methods are used for lineage tracing.

[0146] In some embodiments, the present methods are used in screens, e.g., CRISPR screens, where the methods can be used both to identify the specific guide RNA (gRNA) expressed in a given cell and assess the effects of the gRNA expression on global transcriptome in the same cell. In exemplary such embodiments, the present methods are used to generate a first sequencing library for a population of cells having undergone CRISPR screening that is enriched for the gRNAs, and a second sequencing library directed to the whole transcriptome (or a subset of transcripts of interest).

[0147] In some embodiments, each of the cells within the population comprises an RNA-guided nuclease (e.g., Cas9 or Cpf1), and the transgenes comprise guide RNAs. For example, in such embodiments the detection of a given guide RNA points to a particular genomic modification as directed by the RNA-guided nuclease at a particular locus specified by the guide RNA. Such modifications could comprise, e.g., deletions, insertions, nucleotide alternations, transcriptional activation (e.g., CRISPRa) or inhibition (e.g., CRISPRi), or other changes. In such embodiments, the RNA-guided nuclease (e.g., Cas9) can be introduced, e.g., by stably integrating a transgene encoding the nuclease into the genome of the cells, by introducing a plasmid or vector (e.g., a viral vector) comprising an expression cassette encoding the nuclease into the cells, by introducing mRNA encoding the nuclease into the cells, or by introducing the nuclease polypeptides into the cells. In some embodiments, the transgenes drive the expression of transcripts that can potentially alter the activity of one or more endogenous genes, e.g., antisense, miRNA, shRNA, or siRNA sequences.

[0148] Expression of the transgenes can be driven using any suitable promoter, e.g., a constitutive or an inducible promoter suitable for use in the cells. In some embodiments, the promoter is a constitutive promoter, e.g. a U6 promoter such as human U6 (hU6), or the EF-1 alpha promoter (EF-1A). In some embodiments, transcription of transgenes comprising small non-coding RNAs (e.g., guide RNA, shRNA, siRNA, miRNA) is driven by a type III RNA polymerase, e.g., U6, 7SK, or H1 promoter. In some embodiments, the transcription of transgenes comprising protein coding sequences is driven by a type III RNA polymerase, e.g., CMV, EF1a, CAG, or PGK promoter.

[0149] In some embodiments, the transgenes have been introduced into the cells using a vector e.g., a viral vector such as a lentiviral vector, adeno associated viral (AAV) vector, or adenoviral vector. In some embodiments, the vector is configured such that a transcript (e.g., gRNA) is expressed under the control of more than one promoter, e.g., an RNA Pol III promoter such as hU6, as well as an RNA Pol II promoter such as EF1a, such that the gRNA is produced in both a non-polyadenylated form (from the Pol III promoter) and a polyadenylated form (expressed from the Pol II promoter). In such embodiments, the non-polyadenylated gRNA can bind to the RNA-guided nuclease (e.g., Cas9) to effect a genomic editing event (e.g., indel) at a target locus defined by the gRNA, and the polyadenylated gRNA can be detected for single-cell RNA sequencing, e.g., using a poly(T) reverse transcription primer according to the present methods. In some embodiments, the gRNA sequence is present in the vector (and/or the genome after integration) within an expression cassette, i.e., a construct comprising an expressed sequence (such as a gRNA), operably linked to a promoter (e.g., hU6). In some embodiments, the expression cassette is present in more than one copy in the genome following genomic integration, e.g., due to duplication of the cassette upon integration. In some embodiments, each vector comprises a gRNA-specific barcode sequence (or a barcode specific to a different, non-gRNA polynucleotide), such that the specific gRNA (or other polynucleotide) expressed from the vector can be determined indirectly by sequencing the barcode.

[0150] In some embodiments, the vector is, or is derived from, a vector disclosed in Datlinger et al. (2017) Nature Methods Vol. 14(3); Jaitin et al. (2016) Cell, 167.7: 1883-1896; Xie et al., (2017) Mol. Cell. 66(2):285-299; Dixit et al. (2016) 167(7):1853-1866; all of which are herein incorporated by reference in their entireties.

[0151] Any number of specific targets can be assessed in the present methods. For example, in some embodiments, a CRISPR screen comprising 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000 or more gRNAs can be analyzed using the present methods. The set of gRNAs (or vectors, e.g,. lentiviral vectors) comprising the gRNA sequences can target any number of endogenous genes. For example, each gene could be targeted by 1, 2, 3, 4, 5, or more different gRNAs used in the screen.

[0152] In some embodiments, the targets (e.g., gRNA, e.g., vectors such as lentiviral vectors comprising a gRNA sequence) are introduced into cells at a multiplicity of infection (m.o.i.)

such that most or all cells in the library will contain at most one gRNA vector. In some embodiments, a higher m.o.i. is used, such that cells (or nuclei) in the library are likely to have more than one gRNA vector. Such multiplex embodiments allow the assessment of the combinatorial contributions of the different gRNAs in cells or nuclei.

Fixation and permeabilization

[0153] In some embodiments, the cells (or nuclei) are fixed prior to barcoding such that the components are immobilized or held in place. In some embodiments, fixation is performed on single cell (or nuclei) suspensions. In some embodiments, the cells (or nuclei) were previously frozen and are thawed before fixation. In some embodiments, the cells (or nuclei) are used directly, e.g., from culture or after isolation from a biological source, without freezing. In some embodiments, the cells (or nuclei) are assessed for quality before fixation. Cells (or nuclei) can be counted (e.g., using a hemocytometer) and/or otherwise assessed in any of a number of standard ways, e.g., by staining with trypan blue, acridine orange, and/or propidium iodide. In some embodiments, the cell (or nuclei) suspensions show 70% or more viability, the suspensions show 5% and/or less aggregation/debris. In some embodiments, suspensions with at least about 100,000 cells or nuclei are used. In some embodiments, adherent cells are first dissociated using TrypLE Express Enzyme (1X), phenol red (Thermo Fisher Scientific).

[0154] In some embodiments, the present methods comprise providing a plurality of fixed and permeabilized cells. In some embodiments, the present methods comprise fixing and permeabilizing the plurality of cells or nuclei prior to, e.g., generating cDNA in the cells or nuclei by reverse transcribing RNA in the cells or nuclei. In some embodiments, the cells or nuclei may be fixed and permeabilized and frozen, e.g., at -80 °C, prior to, e.g., generating cDNA. In some embodiments, the cells or nuclei are fixed and permeabilized and then directly used in the present methods, i.e., without freezing and storing them.

[0155] The plurality of cells (or nuclei) may be fixed using any of a number of suitable reagents or conditions. For example, in some embodiments, the cells (or nuclei) can be fixed in formaldehyde in phosphate buffered saline (PBS) (e.g., in about 1-4% formaldehyde in PBS). In various embodiments, the plurality of cells (or nuclei) may be fixed using methanol (e.g., 100% methanol) at about −20° C. or at about 25° C. In various other embodiments, the plurality of cells (or nuclei) may be fixed using methanol (e.g., 100% methanol), at between about −20° C. and about 25° C. In yet various other embodiments, the plurality of cells (or nuclei) may be fixed using ethanol (e.g., about 70-100% ethanol) at about −20° C. or at room temperature. In

yet various other embodiments, the plurality of cells (or nuclei) may be fixed using ethanol (e.g., about 70-100% ethanol) at between about −20° C. and room temperature. In still various other embodiments, the plurality of cells (or nuclei) may be fixed using acetic acid, for example, at about −20° C. In still various other embodiments, the plurality of cells (or nuclei) may be fixed using acetone, for example, at about −20° C. Other suitable methods of fixing the plurality of cells (or nuclei) are also within the scope of this disclosure.

[0156] In some embodiments, RNases are inactivated or eliminated before, during, and/or after fixation and/or permeabilization using, e.g., RNase decontamination products such as RNaseZap RNAse Decontamination Solution (Thermo Fisher). In some embodiments, BSA (e.g., 5-10%, or 7.5%) is added to the cells or nuclei, e.g., to prevent aggregation.

[0157] In some embodiments, the methods may include fixing and/or permeabilizing the cells or nuclei at a temperature below about 8 °C, below about 7 °C, below about 6 °C, below about 5 °C, below about 4 °C, below about 3 °C, below about 2 °C, below about 1 °C, below about 0 °C, below about -5 °C, below about -10 °C, at about 8 °C, at about 7 °C, at about 6 °C, at about 5 °C, at about 4 °C, at about 3 °C, at about 2 °C, at about 1 °C, or at another suitable temperature. In some embodiments, the cells or nuclei are fixed and/or permeabilized at a temperature of below about 8, 7, 6, 5, 4, 3, 2, 1, 0, -1, -2, -3, or -4 °C, between about -4 to 8, -4 to 0, 0 to 4, 4 to 8, or 0 to 8 °C, or at about 8, 7, 6, 5, 4, 3, 2, 1, 0, -1, -2, -3, or -4 °C. In some embodiments, the cells or nuclei are fixed and/or permeabilized on ice.

[0158] In some embodiments, the cells are adherent cells (i.e., cells that are adhered to a plate, e.g., adherent mammalian cells). In some such embodiments, adherent cells are fixed, permeabilized, and/or undergo reverse transcription, followed by trypsinization to detach the cells from a surface. Alternatively, the adherent cells may be detached prior to the separation and/or tagging steps. In some other embodiments, the adherent cells may be trypsinized prior to the fixing and/or permeabilizing steps.

[0159] Permeabilization of the cells (or nuclei) can be achieved in any of a number of ways. For example, a detergent or surfactant such as TRITON™ X-100 may be added to the plurality of cells (or nuclei), followed by the optional addition of HCl. In some such embodiments, about 0.2% TRITON™ X-100 is added to the plurality of cells (or nuclei), followed by the addition of about 0.1 N HCl. In some embodiments, the plurality of cells (or nuclei) is permeabilized using ethanol (e.g., about 70% ethanol), methanol (e.g., about 100% methanol), Tween 20 (e.g., about 0.2% Tween 20), and/or NP-40 (e.g., about 0.1% NP-40).

Reverse transcription

[0160] In some embodiments, reverse transcription is conducted or performed on the plurality of cells or nuclei. In certain embodiments, reverse transcription may be conducted on a fixed and/or permeabilized plurality of T cells (or nuclei). In some embodiments, variants of M-MuLV reverse transcriptase may be used in the reverse transcription. However, any suitable method of reverse transcription is within the scope of this disclosure.

[0161] In some embodiments, the reverse transcription primers may be configured to reverse transcribe all, or substantially all, RNA in a cell (e.g., a random hexamer with a 5′ overhang). In some other embodiments, the reverse transcription primers may be configured to reverse transcribe RNA having a poly(A) tail (e.g., a poly(dT) primer, such as a dT(15) primer or anchored dT(15) primer, with a 5′ overhang). In some embodiments, the reverse transcription primers are configured to reverse transcribe both all, or substantially all, RNA in a cell, as well as to reverse transcribe polyadenylated RNA. In some embodiments, reverse transcription primers may be included that are configured to reverse transcribe predetermined RNAs (e.g., target RNAs).

[0162] In some embodiments, a portion of a reverse transcription (RT) primer that is configured to bind to RNA and/or initiate reverse transcription may comprise one or more of the following: a random hexamer, random septamer, an octamer, a nonamer, a decamer, or a poly(T) (or polydT) stretch of nucleotides (e.g., comprising 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more consecutive thymine bases). In some embodiments, the poly(T) sequence is anchored, i.e., comprises a base other than T at the 3’ end (e.g., a mixture of anchored primers is used each comprising an A, C, or G at the 3’ end of the poly(T) sequence). In some embodiments, RT primers are used that comprise well-specific barcodes, as described in more detail elsewhere herein.

[0163] In some embodiments, primers comprising a poly(T) sequence are used in the reverse transcription in the absence of primers with random sequences (such as random hexamers). In some embodiments, primers comprising random sequences (e.g., random hexamers) are used in the absence of primers with poly(T) sequences. In some embodiments, the reverse transcription is performed with both primers comprising a random sequence (e.g., random hexamer) and primers comprising a poly(T) sequence (e.g., with 15 consecutive thymidine residues).

[0164] In some embodiments, RT primers are used that are specific for one or more target genes or transcripts. For example, specific primers can be used to enrich for specific transcripts among the population of cells, e.g., genes or sequences (such as gRNAs) or a set of transcripts of interest (e.g., to specifically examine genes relating to a particular physiological or cellular pathway or process). In some embodiments, specific primers can be used to enrich for genes relating to a process connected to the genes targeted by a set of gRNAs being assessed in a CRISPR screen. In some embodiments, target genes of the gRNAs (e.g., genes whose expression is expected to be altered following the expression of individual gRNAs can be enriched, so as to provide a control for the activity of an expressed gRNA). In some embodiments, transcripts corresponding to genes encoding other targets can be used, e.g., to T cell receptors (TCRs), e.g., to one or more of TCR alpha, beta, gamma, or delta chains.

[0165] In some embodiments, each of the RT primers comprises a 5' overhang comprising a 5' overhang sequence located 5' of the poly(T) or the random nucleotide sequence, wherein the 5' overhang sequence is the same in all of the RT primers used in the first plurality of aliquots, and wherein following reverse transcription of the RNA, the 5' overhang sequence is present at the 5' end of each of the cDNA molecules.

[0166] The RT primers can be used at any of a range of suitable concentrations. For example, in some embodiments, the concentration of each RT primer (e.g., a polydT primer, a random hexamer primer, or a target-specific primer) is between about 0.5 μM and about 10 μM. In some embodiments, the concentrations are each between about 1 μM and about 7 μM, between about 1.5 μM and about 4 μM, between about 2 μM and about 3 μM, about 2.5 μM, or another suitable concentration.

[0167] In some embodiments, each of the RT primers comprises a barcode (i.e., a specific barcode sequence) (i.e., an "RT primer barcode"). In some embodiments, the reverse transcription is performed on a population of cells (or nuclei) distributed in a plurality of aliquots or wells, and the RT primer barcodes are aliquot- or well-specific. In some such embodiments, all of the RT primer barcodes used in a given aliquot or well are the same, and a different RT primer barcode is used in each of the aliquots or wells. Stated another way, a first barcode may be added to the cDNA molecules in a first specific container, mixture, reaction, receptacle, sample, well, or vessel (e.g., specific to the given container, mixture, reaction, receptacle, sample, well, or vessel), and a second barcode sequence may be added to the cDNA molecules in a second container, mixture, reaction, receptacle, sample, well, or

vessel (and the same for a third, fourth, etc. barcode sequence and third, fourth, etc. container, mixture, reaction, receptacle, sample, well, or vessel). For example, in some embodiments, 48 sets of different well-specific RT primers are used (e.g., in a 48-well plate). Accordingly, if there are 48 samples (e.g., cells, tissues, etc.), each sample can get a unique well-specific barcode. However, if there are only four samples, each sample could have 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or more different sets of well-specific RT primers. A user can know which barcodes correspond to each sample, so the user can recover sample identities. Other numbers of specific barcodes (or well-specific RT primers) are also within the scope of this disclosure. Such a configuration may allow or provide for the multiplexing of the method. In general, any distribution of barcode sequences that can provide some information about the aliquot or well in which a given cell was present can be used. For example, even if more than one barcode sequence is used in one or more aliquots or wells, or if a given barcode sequence is used in more than one aliquot or well, the methods are encompassed by the present disclosure.

[0168] The barcode sequences present within the RT primers can have any of a range of lengths, e.g., 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 nucleotides or longer. In some embodiments, the RT primer barcode sequences are 8 nucleotides in length. By varying 8 nucleotides, there are 65,536 possible unique sequences. In some embodiments, the RT primer barcodes comprise more than 8 nucleotides. In some other embodiments, the RT primers comprise fewer than 8 nucleotides.

[0169] The plurality of cells can be divided prior to reverse transcription into any of a number of aliquots or wells, and using any of a number of suitable reaction vessels or containers. For example, the plurality of cells or nuclei can be distributed into individual tubes or containers, or into a plurality of wells in a multi-well plate. Any multi-well plate can be used, e.g., 4, 6, 8, 12, 24, 48, 96, 384, or 1536 well plates. In some embodiments, the plurality of cells or nuclei are distributed into one or more 96-well plates. In some embodiments, all 96 wells of a 96-well plate are used (i.e., the plurality of cells or nuclei is divided into 96 aliquots). In some embodiments, a fraction of the wells on the plate are used, e.g., the cells or nuclei are distributed into, e.g., 8, 16, 24, 32, 40, 48, 56, 64, 72, 80, 88, 12, 24, 36, 48, 60, 72, or 84 wells of a 96-well plate. In some embodiments, up to 12 wells are used for up to 10,000 cells or nuclei, or up to 48 wells are used for up to 100,000 cells or nuclei, or up to 96 wells are used for up to 1,000,000 cells or nuclei.

[0170] Suitable methods of reverse transcribing RNA within cells for use in the herein-described methods are described, e.g., in US Patent Nos. 10,900,065, 11,634,751, 11,168,355, 11,427,856, 11,555,216, 11,639,519, 11,680,283, 10,633,648, 11,421,221, US Pat. App. Pub. No. US 2021/0388415 A1, in Rosenberg et al., Science 360, 176-182 (2018), Rosenberg et al., BioRxiv (2017), "Scaling single cell transcriptomics through split pool barcoding," doi.org/10.1101/105163, Tran et al. BioRxiv (2022) "High sensitivity single cell RNA sequencing with split pool barcoding," doi.org/10.1101/2022.08.27.505512, the entire disclosures of all of which are herein incorporated by reference (including all supplemental material).

Labeling with nucleic acid tags

[0171] As described above, in some embodiments of the present disclosure a plurality of cells or nuclei is subjected to combinatorial barcoding, such that molecules (e.g., cDNAs synthesized in cells by reverse transcription) are labeled with barcodes that when viewed in combination provide a cell- or nucleus-specific label for the labeled molecules. In such labeling steps, a plurality of cells is split into multiple aliquots or wells, labeled with well-specific barcodes (e.g., during reverse transcription using barcoded primers and/or by ligating barcoded nucleic acid tags), and repooled. This cycle of dividing the plurality of cells, well-specific labeling, and repooling can be repeated any number of times, with each round adding more tags to the cDNAs and thereby creating a set of nucleic acid tags that together can act as, e.g., a cell-specific (or nucleus-specific) barcode. As more and more rounds are added, the number of paths that a cell can take increases and consequently the number of possible barcodes that can be created also increases. Given enough rounds and divisions, the number of possible barcodes will be much higher than the number of cells, resulting in a high likelihood that each cell (or nucleus) in the population has a unique barcode. For example, if the division took place in a 96-well plate, after 4 divisions there would be 964=84,934,656 possible barcodes. In some embodiments of the present methods, enough rounds of labeling is performed such that the number of possible barcodes is 50%, 60%, 70%, 80%, 90%, 100%, 110%, 120%, 130%, 140%, 150%, 200%, 300%, 400%, 500%, 600%, 700%, 800%, 900%, 1000%, or 1x, 2x, 3x, 4x, 5x, 6x, 7x, 8x, 9x, 10x, 20x, 30x, 40x, 50x, 60x, 70x, 80x, 90x, 100x, or more of the number of cells (or nuclei) in the plurality of cells (or nuclei). In some embodiments, enough rounds of labeling are performed such that the likelihood that the tagged molecules within a given cell or nucleus have a unique barcode combination among the plurality of cells is about 95%, 96%,

97%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or greater.

[0172] In each round of tagging, the plurality of cells or nuclei can be divided into any of a number of aliquots or wells, and using any of a number of suitable reaction vessels or containers. For example, the plurality of cells or nuclei can be distributed into individual tubes or containers, or into a plurality of wells in a multi-well plate. Any multi-well plate can be used, e.g., 4, 6, 8, 12, 24, 48, 96, 384, or 1536 well plates. In some embodiments, the plurality of cells or nuclei are distributed into one or more 96-well plates. In some embodiments, all 96 wells of a 96-well plate are used (i.e., the plurality of cells or nuclei is divided into 96 aliquots). In some embodiments, a fraction of the wells on the plate are used, e.g., the cells or nuclei are distributed into, e.g., 8, 16, 24, 32, 40, 48, 56, 64, 72, 80, 88, 12, 24, 36, 48, 60, 72, or 84 wells of a 96-well plate. In some embodiments, 96 distinct barcode sequences are present among the nucleic acid tags used in the plurality of aliquots. In some embodiments, the plurality of aliquots comprises 96 aliquots distributed in a 96-well plate. In some embodiments, each of the 96 distinct barcode sequences are present in only one of the 96 aliquots.

[0173] FIGS. 3A-3D illustrate an exemplary embodiment of labelling or tagging nucleic tags. As shown, during reverse transcription of the mRNA molecules (FIG. 3A), Poly T and random hexamer primers anneal to mRNA within single cells. Each primer contains a barcode and optionally a 5' overhang comprising a 5' overhang sequence. Reverse transcriptase extends cDNA to form a cDNA/mRNA hybrid comprising the barcode (BC1), for example, a well-specific barcode. The cells may be pooled and split and redistributed into individual wells. In subsequent steps, nucleic acid tags can be appended to the cDNA/mRNA hybrid via the 5' overhang. The nucleic acid tags can contain a second barcode (e.g., BC2) and optionally a second DNA linker. The cells can be pooled and split and redistributed into individual wells. A second nucleic acid tag can be ligated to the growing cDNA. The adaptor contains a third barcode, an Illumina adaptor sequence (e.g., R2) or a compatible equivalent thereof, and a biotin molecule.

[0174] Suitable methods of combinatorial barcoding are described, e.g., in US Patent Nos. 10,900,065, 11,634,751, 11,168,355, 11,427,856, 11,555,216, 11,639,519, 11,680,283, 10,633,648, 11,421,221, US Pat. App. Pub. No. US 2021/0388415 A1, in Rosenberg et al., Science 360, 176-182 (2018), Rosenberg et al., BioRxiv (2017), "Scaling single cell transcriptomics through split pool barcoding," doi.org/10.1101/105163, Tran et al. BioRxiv

(2022) "High sensitivity single cell RNA sequencing with split pool barcoding," doi.org/10.1101/2022.08.27.505512, the entire disclosures of all of which are herein incorporated by reference (including all supplemental material).

[0175] In some embodiments, combinatorial barcoding is accomplished by coupling the cDNA molecules generated in each cell during reverse transcription (or DNA fragments or adapters appended to other molecules as described elsewhere herein) with a nucleic acid tag, wherein each nucleic acid tag comprises a barcode sequence, e.g., a well-specific barcode sequence or tag barcode. In some embodiments, coupling the cDNA molecule with the nucleic tag comprises ligating the nucleic tag to the cDNA molecule. In some embodiments, each nucleic acid tag comprises a first strand comprising the barcode sequence, and further comprises a 3' and/or 5' region located 3' and/or 5' of the barcode. In some embodiments, the first strand comprises a 3′ hybridization sequence extending from a 3′ end of a labeling (i.e., barcode) sequence and/or a 5′ hybridization sequence extending from a 5′ end of the labeling (i.e., barcode) sequence. Each nucleic acid tag may also comprise a second (linker) strand including an overhang sequence. The overhang sequence may include (i) a first portion complementary to a 5′ hybridization sequence of a different nucleic acid tag (e.g., a nucleic acid tag appended in a previous round of tagging) or to a 5′ overhang sequence of an RT primer. The second (linker) strand may also comprise a sequence complementary to the 3′ hybridization sequence of the first strand. In some embodiments, the first and second strands are preannealed before being added to the wells or aliquots containing the cells or nuclei.

[0176] In some embodiments, the nucleic acid tags comprise i) a tag barcode sequence, and ii) a 3' hybridization sequence located 3' of the barcode sequence and/or a 5' hybridization sequence located 5' of the barcode sequence, wherein multiple distinct tag barcode sequences are present among the nucleic acid tags used in the second plurality of aliquots, and wherein the tag barcode sequences present in each individual aliquot of the second plurality of aliquots are specific to the individual aliquot.

[0177] In some embodiments, the 3' end of the nucleic acid tag is present within the 3' hybridization sequence, and the 3' end of the nucleic acid tag is brought into proximity of the 5' end of the cDNA molecule by being preannealed to a linker nucleic acid strand that is complementary to the 3' hybridization sequence of the nucleic acid tag and to the 5' overhang sequence of the RT primer, or to the 3' hybridization sequence of the nucleic acid tag and to the 5' hybridization sequence of a previously coupled nucleic acid tag.

[0178] The barcode sequences present within the nucleic acid tags can be any of range of lengths, e.g., 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 nucleotides or longer. In some embodiments, the barcode sequences are 8 nucleotides in length. By varying 8 nucleotides, there are 65,536 possible unique sequences. In some embodiments, more than 8 nucleotides are used. In some other embodiments, fewer than 8 nucleotides are used. In some embodiments, the first strand of the nucleic acid tag is preannealed to the second (linker) strand. In some embodiments, the linker strand includes sequence complementary to part of the RT primer or to a 5' region in a previously coupled nucleic acid tag (i.e., appended to the cDNA in a previous round of tagging), thereby allowing it to hybridize and bring the 3′ end of the barcodes into close proximity to the 5′ end of the reverse transcription primer or the previously added tag. In some embodiments, the phosphate of the reverse transcription primer or previous tag can is ligated to the 3′ end of the first-round barcodes by, e.g., T4 DNA ligase. A 5' region of the nucleic acid tag (e.g., domain s2) can then provide an accessible binding domain for a linker oligo to be used in another round of barcoding. The nucleic acid tags (barcode oligos) can include a 5′ phosphate that can allow ligation to the 3′ end of another oligo by T4 DNA ligase. In some embodiments, the nucleic acid tags are ligated to the cDNA s (or adapter molecules) during each round of labeling. In some embodiments, the methods of labeling nucleic acids in the first cell may comprise ligating at least two of the nucleic acid tags that are bound to the cDNA s. In some embodiments, the nucleic acid tags are hybridized to the cDNA s (or adapters) during each round, and ligation is performed for all of the hybridized tags at a later stage, e.g., following cell lysis, i.e., ligation may be conducted before or after the lysing and/or the cDNA purification steps. Ligation can comprise covalently linking the 5′ phosphate sequences on the nucleic acid tags to the 3′ end of an adjacent strand or nucleic acid tag such that individual tags are formed into a continuous, or substantially continuous, barcode sequence that is bound to the 3′ end of the cDNA sequence. In various embodiments, a double-stranded DNA or RNA ligase may be used with an additional linker strand that is configured to hold a nucleic acid tag together with an adjacent nucleic acid in a "nicked" double-stranded conformation. The double-stranded DNA or RNA ligase can then be used to seal the "nick." In various other embodiments, a single-stranded DNA or RNA ligase may be used without an additional linker. In certain embodiments, the ligation may be performed within the plurality of cells. FIGS. 6 and 7 illustrate the ligation of a plurality of nucleic acid tags to form a substantially continuous label or barcode.

[0179] In some embodiments, following the ligation of the nucleic acid tags during each round, one or more unbound nucleic acid tags are removed (e.g., by washing the plurality of cells). For example, the methods may comprise removing a portion, a majority, or substantially all of the unbound nucleic acid tags. Unbound nucleic acid tags may be removed such that further rounds of the disclosed methods are not contaminated with one or more unbound nucleic acid tags from a previous round of a given method. In some embodiments, unbound nucleic acid tags may be removed via centrifugation. For example, the plurality of cells can be centrifuged such that a pellet of cells is formed at the bottom of a centrifuge tube. The supernatant (i.e., liquid containing the unbound nucleic acid tags) can be removed from the centrifuged cells. The cells may then be resuspended in a buffer (e.g., a fresh buffer that is free or substantially free of unbound nucleic acid tags). In another example, the plurality of cells may be coupled or linked to magnetic beads that are coated with an antibody that is configured to bind the cell or nuclear membrane. The plurality of cells can then be pelleted using a magnet to draw them to one side of the reaction vessel. In some other embodiments, the plurality of cells may be placed in a cell strainer (e.g., a PLURISTRAINER® cell strainer) and washed with a wash buffer. For example, the plurality of cells may remain in the cell strainer while the wash buffer passes through the cell strainer. Wash buffer may include, e.g., a surfactant, a detergent, and/or about 5-60% formamide.

[0180] In some embodiments, the ligation can be stopped during each round of combinatorial labeling by adding an excess of oligo that is complementary to all or part of the linker (second) strand used during the same round of labeling. To stop each barcode ligation, oligo strands that are fully complementary to the linker oligos can be added. These oligos can bind the linker strands attached to unligated barcodes and displace the unligated barcodes through a strand displacement reaction. The unligated barcodes can then be completely single-stranded. As T4 DNA ligase, for example, is unable to ligate single-stranded DNA to other single-stranded DNA , the ligation reaction will stop progressing. In some embodiments, to ensure that all linker oligos are bound by the complementary (i.e., stop) oligos, a molar excess of the stop oligos (relative to the linker oligos) is added. In some embodiments, stop ligation strands are diluted into 10X Ligase Buffer and water, e.g., 264 µl stop ligation strand, 300 µl 10X T4 DNA Ligase Buffer, and 636 µl nuclease-free water.

[0181] In some embodiments, the nucleic acid tag (e.g., a final nucleic acid tag appended during the last of multiple rounds of tagging) may comprise a capture agent such as, but not limited to, biotin, e.g., a 5′ biotin. A cDNA labeled with a 5′ biotin-comprising nucleic acid tag

may allow or permit the attachment or coupling of the cDNA to a streptavidin-coated magnetic bead, e.g., C1 beads. In some other embodiments, a plurality of beads may be coated with a capture strand (i.e., a nucleic acid sequence) that is configured to hybridize to a final sequence overhang of a barcode. In yet some other embodiments, cDNA may be purified or isolated by use of a commercially available kit (e.g., an RNEASY™ kit).

[0182] In some embodiments, one or more nucleic acid tags may comprise additional elements (in addition to biotin or another capture agent) such as a nucleotide sequence (e.g., comprising random and/or degenerate nucleotides) allowing the detection of PCR duplicates, primer binding sequences, adapter sequences for next-generation sequencing (NGS) (e.g., Illumina adapter sequences), and others. The random nucleotide sequences allow the computational removal of PCR duplicates, since these duplicates will have the same random sequence. In this way, each original transcript will only be counted once, even if multiple PCR duplicates are sequenced. Such sequences can contain any number of random successive nucleotides, e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10 nucleotides or more, and can be used either alone or in conjunction with other sequence elements (such as the other barcode sequences described herein) to allow the identification of the PCR duplicates. The number of different random sequences per cell indicates how many unique RNA molecules can be detected per cell, which is directly related to how efficiently the molecules can be barcoded and processed to enable detection by next generation sequencing.

Lysis

[0183] In some embodiments, the methods include lysing (i.e., breaking down the cell or nuclear structure) the plurality of cells (or nuclei) to release the tagged cDNA molecules from the plurality of cells or nuclei following combinatorial barcoding, thereby forming a lysate comprising the released tagged cDNA molecules. In some embodiments, the cells or nuclei comprising the tagged cDNAs (or other molecules) are divided into one or more samples or sublibraries, and the lysis is performed separately on each sample or sublibrary. In some embodiments, as described below each sample or sublibrary can be tagged with one or more index or barcode sequences during a subsequent step of the herein-described methods (e.g., using unique dual indices, or UDIs, as described elsewhere herein).

[0184] The total number of sublibraries prepared can depend on various factors, including the number of cells in the plurality of cells or nuclei. For example, in some embodiments, the plurality of cells or nuclei comprises up to 10,000 cells or nuclei, and two sublibraries are

prepared. In some embodiments, the plurality of cells or nuclei comprises up to 100,000 cells or nuclei, and 8 sublibraries are prepared. In some embodiments, the plurality of cells or nuclei comprises up to 1,000,000 cells or nuclei, and 16 sublibraries are prepared. Other numbers of cells or nuclei, including less than 10,000 and greater than 1,000,000, or various numbers between 10,000 and 1,000,000, can be used, and a skilled artisan will be able to determine a suitable number of sublibraries.

[0185] Further, different numbers of cells can be added to each sublibrary as desired. Sublibraries with small cell numbers (e.g., 200-500) will be easier to sequence to saturation, and can serve, e.g., as a good quality control (QC) measure before sequencing additional sublibraries with much larger cell numbers. In some embodiments, the number of cells in each library can be, for example, 200, 500, 1,000, 2,000, 3,000, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, or more.

[0186] Once prepared, each sublibrary can be immediately processed, e.g., to prepare a sequencing library, or stored, e.g., at -80 °C. Further, while all of the sublibraries can be processed together, each sublibrary can be sequenced separately.

[0187] In some embodiments, the plurality of cells is lysed in a lysis solution (e.g., a solution comprising Tris-HCl, EDTA, NaCl, and SDS, e.g., 10 mM Tris-HCl (pH 7.9), 50 mM EDTA (pH 7.9), 0.2 M NaCl, 2.2% SDS) comprising an RNase inhibitor (e.g., 0.5 mg/ml ANTI-RNase, AMBION®) and a protease such as a serine protease, e.g., Proteinase K (e.g., 1000 mg/ml proteinase K (AMBION®)). The skilled artisan will be able to identify suitable conditions for cell (or nuclear) lysis. In some embodiments, lysis is performed at about 55 °C for about 3 hours with shaking (e.g., vigorous shaking). In some other embodiments, the plurality of cells is lysed using ultrasonication and/or by being passed through an 18-25 gauge syringe needle at least once. In yet some other embodiments, the plurality of cells is lysed by being heated to about 70-90 °C. For example, the plurality of cells may be lysed by being heated to about 70-90 °C for about one or more hours.

cDNA isolation

[0188] Following cell or nuclear lysis, the cDNA molecules may be isolated from the lysed cells or nuclei. In some embodiments, RNase H (or another RNase) may be added to the cDNA to remove RNA. The methods may further comprise ligating at least two of the nucleic acid tags that are bound to the released cDNA s (i.e., in embodiments in which the tags were not ligated during each round of split-pool tagging). In some such embodiments, the methods may

comprise ligating at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more nucleic acid tags that are bound to the cDNA molecules.

[0189] In some embodiments, the tagged cDNA molecules are isolated from the lysis solution with a purification or clean-up step, e.g. an SPRI bead cleanup, before binding the desired nucleic acids (i.e., cDNAs containing 5′ biotin) to streptavidin beads. In some embodiments, a protease inhibitor is added to lysates and then streptavidin beads are directly added (i.e., skipping the first SPRI isolation of nucleic acids). The protease inhibitor may include phenylmethanesulfonyl fluoride (PMSF), 4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride (AEBSF), a combination thereof, and/or another suitable protease inhibitor.

[0190] In some embodiments, the cDNA molecules are isolated using Streptavidin beads, e.g., C1 beads. For example, 20 µl of resuspended DYNABEADS® MYONE™ Streptavidin C1 beads (for each aliquot of cells) can be added to a 1.7 ml microcentrifuge tube (EPPENDORF®). The beads can be washed, e.g., 3 times, with, e.g., 1× phosphate buffered saline Tween 20 (PBST) and resuspended in PBST (e.g., 20 µl PBST). In some embodiments, 900 µl PBST is added to the cell aliquot and 20 µl of washed C1 beads are added to the aliquot of lysed cells. The samples can, e.g., be placed on a gentle roller for 15 minutes at room temperature and then washed, e.g., 3 times with 800 µl PBST using a magnetic tube rack (EPPENDORF®). The beads can then be resuspended in, e.g., PBS such as 100 µl PBS.

[0191] In some embodiments, a microcentrifuge tube (EPPENDORF®) comprising a sample can be placed against a magnetic tube rack (EPPENDORF®) for, e.g., 2 minutes and then the liquid can be aspirated. In some embodiments, the beads can be resuspended in, e.g., an RNase solution (3 µl RNase Mix (ROCHE™), 1 µl RNase H (NEW ENGLAND BIOLABS®), 5 µl RNase H 10× Buffer (NEW ENGLAND BIOLABS®), and 41 µl nuclease-free water). The sample can be incubated under suitable conditions, e.g., at 37 °C for 1 hour, and then removed from the conditions and placed against a magnetic tube rack (EPPENDORF®) for, e.g., 2 minutes. The sample can be washed with, e.g., 750 µl of nuclease-free water+0.01% Tween 20 (H2O-T), without resuspending the beads and keeping the tube disposed against the magnetic tube rack. The liquid can then be aspirated. The sample can be washed with 750 µl H2O-T without resuspending the beads and while keeping the tube disposed against the magnetic tube rack. Next, the liquid can be aspirated while keeping the tube disposed against the magnetic tube rack. The tube can then be removed from the magnetic tube rack and the sample can be resuspended in 40 µl of nuclease-free water.

[0192] FIGS. 4A-4B illustrates an exemplary embodiment of isolating cDNA molecules. After cell lysis, the biotinylated cDNA/mRNA hybrid binds to a streptavidin binder bead (FIG. 4A). Molecules having biotin are collected and molecules lacking biotin are removed. Next, a template switch reaction is performed (FIG. 4B) using a template switching oligonucleotide (TSO) to append a template switching (TS) adapter comprising, e.g., a primer binding site (e.g., a template switching primer "TS primer") to the 3' end of the cDNA molecule for cDNA amplification. During cDNA amplification, two products can be created: (i) whole transcriptome and (ii) target cDNA molecules (see, e.g., FIGS. 4C-4D).

Second strand synthesis

[0193] In some embodiments, to facilitate subsequent amplification, a common adapter sequence (or NGS adapter) is added to the 3′-end of the released cDNA molecules following isolation of the released cDNA (i.e., cDNA /mRNA duplex). In some embodiments, the common (or NGS) adapter sequence is the same, or substantially the same, for each of the cDNA molecules (i.e., within a given experiment). The addition of the common adapter may be conducted or performed in a solution including up to about 10% w/v of PEG, wherein the molecular weight of the PEG is between about 7,000 g/mol and 9,000 g/mol. In some embodiments, to prevent concatemers of the adapter oligo, dideoxycytidine (ddC) can be included at the 3′ end of the adapter oligo. In some embodiments, adapters are used with a phosphate at the 5′ end and ddC at the 3′ end. Several enzymes are capable of ligating single-stranded oligo to the 3′ end of single-stranded DNA, e.g., T4 RNA ligase 1 (NEW ENGLAND BIOLABS®) or thermostable 5′ AppDNA /RNA Ligase (NEW ENGLAND BIOLABS®).

[0194] In some embodiments, the adapter sequence is added to the 3′-end of the released cDNA molecules by template switching (see Picelli, S, et al. Nature Methods 10, 1096-1098 (2013)). For example, template switching can be performed on the cDNA molecules, i.e., the cDNA/RNA duplexes attached to streptavidin beads, using a template switching oligonucleotide (TSO) or template switching primer (TS) comprising the adapter sequence. In some embodiments, up to 10% w/v PEG (molecular weight 7000-9000) is used in the template switch reaction.

[0195] In some embodiments of the present methods, the adapter sequence introduced by the TSO is used for the amplification of tagged cDNA molecules, e.g., as illustrated in FIGs. 4A-4D.

[0196] FIGS. 4A-4D provide an overview of cDNA capture and amplification (multiplex preamplification and/or target amplification) steps according to certain embodiments of the present methods, e.g., in the context of CRISPR screens such as CROP-seq or similar methods. FIG. 4A shows cDNA Capture. Biotinylated cDNA was captured via streptavidin beads. FIG. 4B shows cDNA Template Switch. A template switch (TS) reaction adds an adapter to the 3' end of the cDNA. FIG. 4C shows cDNA Amplification. The cDNA was amplified with template switch adaptor and Illumina Truseq R2 specific primers. In some embodiments, the first round of amplification (or multiplex amplification) shown here, also referred to as "preamplification" in the context of the present disclosure, additional primers specific for one or more target cDNAs (e.g., expressed from target genes or comprising target sequences) were added, or "spiked-in," at this step, so as to enrich the presence of the target cDNAs in the mixture. For example, FIG. 4C shows hU6-sgRNA-polyA transcripts which had been enriched due to the addition of a human U6 specific primer. Amplified cDNA was split to create Whole Transcriptome and CRISPR sequencing libraries, as shown in FIGS. 5 and 6 and as described elsewhere herein. FIG. 4D presents another view of the products of the preamplification step, showing a preamplified cDNA representative of the whole transcriptome, and a preamplified target cDNA, in this case comprising a gRNA sequence.

[0197] FIGS. 5A-5C provide an overview of the preparation of whole transcriptome (WT) sequencing libraries according to certain embodiments of the present methods, e.g., in the context of CRISPR screening, TCR profiling of T cells, or other multiplex applications. FIG. 5A shows cDNA transcripts were fragmented, ends were repaired, and A-tailed. FIG. 5B shows Adapter Ligation. As an example, an Illumina Truseq R1 Adapter was ligated to the 5' end of the DNA. FIG. 5C shows Round 4 Barcoding. The sequencing library was amplified, adding P5/P7 Adapters and a fourth barcode via the UDI – WT Plate.

[0198] FIGS. 6A-6C provide an overview of the preparation of CRISPR sequencing libraries according to certain embodiments of the present methods. FIGS. 6A-6B show two additional amplification steps after the initial preamplification and subsequent separation of the enriched preamplified tagged cDNAs. FIG. 6A shows CRISPR PCR. In the first additional round of amplification, a PCR reaction enriched the sgRNA containing cDNA with a second hU6 specific primer. This reaction could also add an adaptor, e.g., Illumina adaptor or any compatible equivalents thereof. FIG. 6B shows CRISPR Index PCR. In the second additional round of amplification, the CRISPR Sequencing Library was amplified, adding P5/P7 adaptors

and a fourth barcode via the Illumina indexes in the UDI Plate - EC. FIG. 6Cshows another view of the products of the two additional rounds of amplification.

Multiplex cDNA amplification (or "preamplification")

[0199] Following the isolation of the tagged cDNA molecules and second strand synthesis, the cDNA molecules are amplified in a first round of "multiplex" amplification, in which PCR is performed using both non-specific (i.e., non-target specific) WT primers (e.g., TSO and R2 primers) configured to amplify all tagged cDNAs in the transcriptome, and target-specific primers configured to amplify target sequences only. Because this initial round of amplification is a first enrichment step (due to the spiked-in target primers) taking place before a subsequent, additional round of target-specific amplification (i.e., a target-specific "enrichment" round of PCR performed exclusively with target-specific primers during the preparation of the target sequencing library), this round is referred to herein as a "preamplification" step. Similarly, the primers used in this round are referred to herein as "preamplification primers." This preamplification step is a multiplex PCR in which both the whole transcriptome and the specific target sequences to be enriched (e.g., gRNA or TCR sequences) are amplified simultaneously, but with the target sequences being amplified preferentially (due to the presence of the spiked-in target specific primers or gene specific primers) such that they become enriched within the overall transcriptome.

[0200] In some embodiments, preamplification comprises amplifying cDNA molecules using at least one pair of primers (i.e., whole transcriptome (WT) preamplification primers) configured to broadly amplify tagged cDNA molecules in the mixture but which do not specifically anneal to target cDNAs, as well as at least one pair of "spiked-in" target-specific primers ("target preamplification primers") configured to specifically amplify one or more target sequences. In some embodiments, the at least one pair of WT primers can comprise one primer complementary to an adapter sequence introduced by the TSO, and one primer complementary to an adapter sequence (e.g., R2 sequence) introduced by the last nucleic acid tag appended to the cDNA during split-pool labeling. In some embodiments, a pair of spiked-in target preamplification primers comprises two target-specific primers, such that neither of the primers is configured to amplify non-target cDNAs within the whole transcriptome. In some embodiments, a pair of spiked-in preamplification primers comprises one target-specific primer, and one non-target-specific primer (e.g., a WT primer such as an RT primer), such that PCR using the pair of primers will only amplify (i.e., exponentially amplify) target cDNAs.

[0201] In some embodiments, one or more WT preamplification primers used in the preamplification round comprise the sequence of SEQ ID NO:12 or SEQ ID NO:13, or a sequence comprising 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity to SEQ ID NO:12 or SEQ ID NO:13, or a sequence comprising not more than 1, 2, 3, 4, or 5 mismatches relative to SEQ ID NO:12 or SEQ ID NO:13.

[0202] In some embodiments, one or more target-specific primers used in the preamplification round (e.g., for CRISPR screens) comprise the sequence of any one of SEQ ID NO:11, or a sequence comprising 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity to SEQ ID NO:11, or a sequence comprising not more than 1, 2, 3, 4, or 5 mismatches relative to SEQ ID NO:11. In some embodiments, one or more pairs of target preamplification primers used in the preamplification round comprise the sequence of SEQ ID NO:12, or a sequence comprising 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity to SEQ ID NO:12, or a sequence comprising not more than 1, 2, 3, 4, or 5 mismatches relative to SEQ ID NO:12. In some embodiments, one or more target-specific primers comprises any of the sequences shown as SEQ ID NOS:1-10, or a sequence comprising 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity to any of SEQ ID NOS:1-10, or a sequence comprising not more than 1, 2, 3, 4, or 5 mismatches relative to any of SEQ ID NOS:1-10.

[0203] The primers can be used at a range of concentrations. In some embodiments, the primers are each added at from 1 to 10 μM, e.g., at 1.2 μM, 2.4 μM, 4.8 μM, 7.2 μM, 9.6 μM, or 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more μM, or from 100 nM to 1 μM, e.g., 100, 200, 300, 400, 500, 600, 700, 800, 900, or more nM for each target preamplification primer. In some embodiments, the ratio (e.g., molar ratio) of target-specific primers to WT primers used is 1:1000, 1:900, 1:800, 1:700, 1:600, 1:500, 1:400, 1:300, 1:200, 1:100, 1:90, 1:80, 1:70, 1:60, 1:50, 1:40, 1:30, 1:20, 1:10, 1:9, 1:8, 1:7, 1:6, 1:5, 1:4, 1:3, 1:2, 1:1, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, 10:1, 20:1, 30:1, 40:1, 50:1, 60:1, 70:1, 80:1, 90:1, 100:1, 200:1, 300:1, 400:1, 500:1, 600:1, 700:1, 800:1, 900:1, or 1000:1.

[0204] It will be appreciated that primers said to "non-specifically" amplify the tagged cDNAs (e.g., WT preamplification primers) in the mixture means that they are configured broadly amplify all tagged cDNAs generated using the present methods (e.g., all cDNAs comprising a TSO adapter sequence and an R2 adapter sequence). It does not necessarily mean, however, that the primers will have identical affinity for each of the tagged cDNAs in the mixture, or that they will amplify all tagged cDNAs in the mixture at an identical rate. Instead, primers

said to "non-specifically" amplify the tagged cDNAs simply means that the primers are not designed to specifically anneal to any particular cDNAs in the mixture, e.g., are not designed to be complementary to a target sequence, such that the primers can generally be expected to amplify the overall collection of tagged cDNA molecules in the mixture. It will also be appreciated that a pair of primers said to "specifically" amplify a target sequence can comprise two primers with complementarity to the target sequence, or one primer with complementarity to the target sequence and one primer capable of annealing to all tagged cDNAs in the mixture (including non-target sequences). For example, the primer capable of annealing to all of the tagged cDNAs could comprise complementarity to an R2 adapter sequence present the last nucleic acid tag coupled to all tagged cDNAs during combinatorial barcoding.

[0205] For example, the at least one pair of target preamplification primers may comprise a primer with complementarity to one or more target sequence or sequences (e.g. to a specific gRNA in a CRISPR screen, or to a particular TCR variable region sequence), or to a sequence common to a plurality of different target sequences (such as a U6 promoter sequence found upstream of Pol III generated transcripts in the case of CRISPR screens such as CROP-seq screens, or to a TCR sequence, e.g., a variable region sequence common to multiple TCR genes in the population. As a result of the use of these non-specific (i.e., WT preamplification primers) and target-specific primers (i.e., target preamplification primers), the cDNA amplification will give rise to two products: (1) the whole transcriptome, and (2) an enriched pool of tagged cDNAs corresponding to the specific target or targets (e.g., gRNA or TCR sequences) amplified by the target preamplification primers, meaning that the resulting mixture of cDNAs will be an enriched cDNA amplification product that comprises cDNAs for the whole transcriptome that is enriched for transcripts corresponding to the specific target transcripts of interest (e.g., gRNA TCR sequences).

[0206] In some embodiments, the preamplification step is performed by placing tubes comprising the (magnetic) beads with bound tagged cDNA molecules against a magnetic rack, removing and discarding the clear supernatant, and resuspending the beads in bind buffer, then removing and discarding the supernatant, then removing the tubes from the magnetic rack and resuspending the beads in amplification reaction solution (e.g., a solution comprising an amplification master buffer, WT preamplification primers for amplifying the whole transcriptome, and/or target specific preamplification primers (e.g., TCR-specific primers). The resuspended beads can then be stored on ice (or a suitable temperature, e.g., at about 0, 1, 2, 3, 4, 5, 6, 7, or 8 °C). The tubes can then be placed in a thermocycler and subjected to suitable

conditions for PCR amplification of the cDNAs. Following amplification, the tubes can be removed and stored, e.g., at 4 °C.

[0207] Following amplification of the cDNA molecules, the PCR products can be cleaned up, e.g., by the addition of solid phase reversible immobilization (SPRI) beads. SPRI beads are used to remove polynucleotides of less than about 200 base pairs, less than about 175 base pairs, or less than about 150 base pairs (see DeAngelis, M M, et al. Nucleic Acids Research (1995) 23(22):4742). In some embodiments, SPRI beads are used to remove polynucleotides of less than about 200 base pairs. The ratio of SPRI bead solution to amplified cDNA molecule solution may be between about 0.9:1 and about 0.7:1, between about 0.875:1 and about 0.775:1, between about 0.85:1 and about 0.75:1, between about 0.825:1 and about 0.725:1, about 0.8:1, or another suitable ratio. In some embodiments, the SPRI clean-up is single-sided. In some embodiments, the SPRI clean-up is double-sided.

[0208] Furthermore, the SPRI bead solution may include between about 1 M and 4 M NaCl, between about 2 M and 3 M NaCl, between about 2.25 M and 2.75 M NaCl, about 2.5 M NaCl, or another suitable amount of NaCl. The SPRI bead solution may also include between about 15% w/v and 25% w/v polyethylene glycol (PEG), wherein the molecular weight of the PEG is between about 7,000 g/mol and 9,000 g/mol (PEG 8000). In various embodiments, the SPRI bead solution may include between about 17% w/v and 23% w/v PEG 8000, between about 18% w/v and 22% w/v PEG 8000, between about 19% w/v and 21% w/v PEG 8000, about 20% w/v PEG 8000, or another suitable % w/v PEG 8000.

[0209] Specifically, 20 µl of the RNase-treated beads can be added to a single PCR tube. 80 µl of ligase mix (5 µl T4 RNA Ligase 1 (NEW ENGLAND BIOLABS®), 10 µl 10X T4 RNA ligase buffer, 5 µl BC_0047 oligo at 50 µM, 50 µl 50% PEG 8000, and 10 µl 10 mM ATP) can be added to the 20 µl of beads in the PCR tube. 50 µl of the ligase mixed with the beads can be transferred into a new PCR tube to prevent too many beads from settling to the bottom of a single tube and the sample can be incubated at 25 °C for 16 hours.

Preparation of Sequencing libraries

[0210] Following the first (preamplification) phase of cDNA amplification, the combined products of preamplification (i.e., mixture of whole transcriptome cDNA molecules and enriched target cDNA molecules) are used as inputs to separately prepare two sequencing libraries are prepared for each cDNA sublibrary: a Whole Transcriptome (WT) Library (WT Sequencing Library) and a target-specific library (Target Sequencing Library). In embodiments

in which, e.g., 8 sublibraries have been prepared prior to lysis, this will result in a total of 16 sequencing libraries from the previous 8 sublibraries made. In some embodiments, during preparation of the WT and target libraries, an additional (e.g., fourth) sublibrary-specific barcode is added to the cDNA molecules. In some embodiments, this additional (e.g., fourth) barcode is an Illumina Unique Dual Index (UDI).

Whole Transcriptome (WT) Sequencing Libraries

[0211] In some embodiments, to prepare WT sequencing libraries, the preamplified WT cDNA molecules are fragmented, an adapter comprising, inter alia, a primer binding sequence is appended to the fragmented ends, and an additional amplification reaction is performed to introduce one or more index sequences (e.g., unique dual indexes, or UDIs) and sequencing primer binding sites for NGS sequencing (see, e.g., FIGS. 5A-5C). The WT cDNAs can be fragmented, e.g., using a fragmentation enzyme and fragmentation buffer. In some embodiments, the preamplified cDNA molecules are fragmented by incubating with the fragmentation enzyme and buffer at 32 °C for 10 minutes and are then held at 65 °C for, e.g., 30 minutes. In some embodiments, following fragmentation of the DNA, the fragment ends are repaired and A-tailed, and the adapter is ligated to the ends. For example, in some embodiments, an Illumina Truseq R1 Adapter is ligated to the 5' end of the DNA. In some embodiments, the ligation of the adapter to the ends of the fragments of the preamplified cDNA can be preceded and/or followed by an SPRI clean-up step, e.g., using Ampure XP or KAPA Pure Beads.

[0212] In some embodiments, the cleaned-up molecules are then subjected to an additional round of amplification, e.g., adding P5/P7 adapter sequences. A fourth barcode can also be added. In some embodiments, the additional barcodes correspond to unique dual indexes (UDI), e.g., with different well-specific index primers used for each sublibrary. In some embodiments, the indexing round of amplification be preceded and/or followed by an additional SPRI clean up step (e.g., using Ampure XP or KAPA Pure Beads).

[0213] FIGS. 5A-5C illustrate an exemplary embodiment of preparation of whole transcriptome libraries for sequencing. Sublibrary cDNA is fragmented to a size compatible with a suitable sequencing platform (FIG. 5A), e.g., Illumina sequencing or other compatible sequencing platforms. A second adaptor (e.g., an R1 adapter) is ligated to the fragmented end of the cDNA (FIG. 5B). Lastly, a final PCR (FIG. 5C) amplifies the fragmented cDNA and appends the fourth DNA barcode, UDIs, and P5 and P7 adaptors.

Target Sequencing Libraries

[0214] In parallel to the preparation of the WT sequencing library, a target sequencing library is prepared. In some embodiments, the preparation of the target library comprises one or more additional rounds of PCR amplification, e.g., using the product of the first PCR amplification as a starting material for amplification. This second round of "enrichment" PCR uses target-specific primer pairs to specifically amplify target cDNA molecules, in the absence of WT primer pairs. In some embodiments, a first additional round of amplification is performed using at least one target-specific amplification primer that is different than the target-specific preamplification primer(s) used in the previous, preamplification step. For example, in some embodiments, distinct, nested primers are used in the preamplification and amplification steps, e.g., as shown in FIGS. 2A-2C (i.e., Focal Primer 1 and Focal Primer 2). In other embodiments, the same primers are used in the preamplification and amplification steps, or, alternatively, different primers are used that nevertheless bind to the same binding site on the target cDNA (i.e., the primers bind to the same sites on the target cDNA, but may comprise distinct additional elements outside of their hybridization region).

[0215] In some embodiments, the first additional round of amplification is followed by a second additional round of amplification. For example, the first round of amplification can be performed to further enrich the target sequences and also add, e.g., an adapter sequence such as an Illumina adapter or other compatible adapters. Subsequently, the second round of amplification can further enrich the target sequences as well as introduce additional elements such as flow cell binding sites (e.g., P5 and P7) and index sequences (e.g., unique dual indexes, or UDIs). UDIs as used herein can be Illumina UDIs or any compatible UDIs, (e.g., Zymo-Seq SwitchFree™ 3'mRNA Library Kits (Zymo-Seq RiboFree® Total RNA Library Kits (Cat. No. R3000/R3003). In some embodiments, the UDIs used are those shown in Table 16 (i.e., in a UDI – WT plate for WT index PCR in the preparation of WT sequencing libraries) and Table 20 (i.e., in a UDI – EC plate, for target specific index PCR in the preparation of target sequencing libraries.

[0216] FIGS. 2A-2C provide an overview of one embodiment of the present methods, e.g., using a pair of nested "focal primers" to enrich for one or more target sequences when preparing multiple sequencing libraries for multiplex single-cell analysis purposes. FIG. 2A shows the products of the combinatorial barcoding of cDNA in cells or nuclei, e.g., using split-pool labeling as described herein. Shown are both a generic cDNA molecule representative of the entire genome, as well as a cDNA molecule representative of a subset of the whole

transcriptome comprising a target cDNA of interest. Two nested primer binding sites are indicated on the target cDNA, corresponding to one embodiment of the present disclosure. FIG. 2B shows the primer binding sites during the "preamplification" step according to one embodiment of the present disclosure. In the image, the cDNA representative of the whole transcriptome was amplified using a primer (the "R2 primer") specific to an adapter sequence introduced, e.g., by a nucleic acid tag during the last round of combinatorial barcoding, and a second primer (the "PCR primer") specific to a second adapter sequence introduced, e.g., by the template switching oligonucleotide (TSO). In parallel, the target cDNA was amplified by the same R2 primer and by a second primer (the "Focal Primer") that binds specifically to the target cDNA. In this way, the target cDNAs were enriched within the overall mixture of preamplification products. FIG. 2C shows the two types of cDNA products present in the same sample following preamplification: the whole transcriptome (WT) cDNA products representative of the whole transcriptome in the cell or nucleus, and the enriched target (or "gene-specific") cDNA molecules. The sample containing the two products was then used as input for the subsequent separate preparation of the whole transcriptome (WT) sequencing libraries and target (e.g., "focal PCR") sequencing libraries.

[0217] In some embodiments, the first additional round of amplification is followed by a second additional round of amplification. For example, the first round of amplification can be performed to further enrich the target sequences and also add, e.g., an adapter sequence such as an Illumina adapter or other compatible adapters. Subsequently, the second round of amplification can further enrich the target sequences as well as introduce additional elements such as flow cell binding sites (e.g., P5 and P7) and index sequences (e.g., unique dual indexes, or UDIs). UDIs as used herein can be Illumina UDIs or any compatible UDIs, (e.g., Zymo-Seq SwitchFreeTM 3'mRNA Library Kits, Cat. No. R3008/R3009; Zymo-Seq RiboFree® Total RNA Library Kits, Cat. No. R3000/R3003).

[0218] One exemplary method of preparing a target-specific sequencing library is shown in FIGS. 6A-6C, where the first additional round of amplification further enriches the sgRNA-containing cDNA using a second hU6 specific primer, and also adds an Illumina adaptor, and the second additional round adds P5/P7 adaptors and a fourth barcode via the Illumina indexes in the UDI Plate - EC. Another exemplary method is shown in FIGS. 7A-7C, where the first additional round amplifies a subset of the cDNA from the whole transcriptome that contains V(D)J segments in the CDR3 repertoire of the T cell, and also adds an Illumina adaptor, and the second, final round amplifies the TCR Amplification product from the previous round and

also appends UDIs from the UDI - EC Plate as well as the P5 and P7 adaptors. In some embodiments, the indexing round of amplification can be preceded and/or followed by an additional SPRI clean up step (e.g., using Ampure XP or KAPA Pure Beads).

[0219] FIGS. 7A-7C provide an overview of the preparation of TCR sequencing libraries according to certain embodiments of the present methods. FIGS. 7A-7B show two additional amplification steps after the initial preamplification and subsequent separation of the enriched preamplified tagged cDNAs. FIG. 7A illustrates TCR Amplification 1. A PCR reaction amplified a subset of the cDNA from the whole transcriptome that contains V(D)J segments in the CDR3 repertoire of the T cell. This reaction also added an adaptor, e.g., Illumina adaptor or any compactible equivalents thereof. FIG. 7B illustrates TCR Amplification 2. A final PCR amplifies the TCR Amplification 1 product and appends the fourth DNA barcodes, UDIs from the UDI - EC Plate, as well as the P5 and P7 adaptors. FIG. 7C shows another view of the products of the two additional rounds of amplification.

[0220] In some embodiments, e.g., in the context of a CRISPR screen such as a CROP-seq screen, one or more primers are used in the preamplification and/or amplification steps that are specific to a Pol III promoter, e.g., a U6 promoter such as a human U6 promoter. In some such embodiments, one or more target primers used in the preamplification round comprise the sequence of SEQ ID NO:1, or a sequence comprising 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity to SEQ ID NO:1, or a sequence comprising not more than 1, 2, 3, 4, or 5 mismatches relative to SEQ ID NO:1. In some embodiments, one or more target amplification primers used in the amplification round, e.g., the first round of amplification following the preamplification step, comprise the sequence of SEQ ID NO:4, or a sequence comprising 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity to SEQ ID NO:4, or a sequence comprising not more than 1, 2, 3, 4, or 5 mismatches relative to SEQ ID NO:4. In some embodiments, one or more pairs of target amplification primers used in the amplification round, e.g., the first round of amplification following the preamplification step, comprise the sequence of SEQ ID NO:2, or a sequence comprising 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity to SEQ ID NO:2, or a sequence comprising not more than 1, 2, 3, 4, or 5 mismatches relative to SEQ ID NO:2.

[0221] In some embodiments, e.g., for the preparation of target cDNA libraries, or for the preparation of specific sets of transcripts within the overall transcriptome, specific cDNAs are enriched by hybridization-based methods, e.g., by gene capture using a panel specific to the

cDNAs of interest. For example, in some embodiments, e.g., where it is desired to evaluate the effects of different gene modulations caused by different gRNAs in a CRISPR screen, or the effects of different drug target candidates, on a particular cellular, physiological, or disease-related process, such methods can be used to restrict the analysis to a subset of the whole transcriptome related to the process in question, rather than on the whole transcriptome itself. Similarly, in some embodiments, such hybridization-based methods (e.g., gene capture) can be used in conjunction with the herein-disclosed amplification steps to further enrich for particular targets.

[0222] FIGs. 7A-7C illustrate an exemplary embodiment of preparing TCR libraries for sequencing. During first TCR amplification, a PCR reaction amplifies a subset of the cDNA from the whole transcriptome that contains V(D)J segments in the CDR3 repertoire of the T cell. This reaction also adds an Illumina adaptor. During the second TCR amplification, a final PCR amplifies the amplification product from the first amplification and appends the fourth DNA barcodes, UDIs, and P5 and P7 adaptors.

[0223] FIGs. 6A-6C illustrate an exemplary embodiment of preparing CRISPR libraries for sequencing. A PCR reaction enriches the sgRNA containing cDNA. This reaction also adds an Illumina adaptor. A final PCR amplifies the CRISPR PCR product and appends the fourth DNA barcodes as UDI, and the P5 and P7 adaptors.

Sequencing and analysis

[0224] In some embodiments, sequencing reads from both libraries (i.e., target and WT libraries) are grouped by cell barcodes (e.g., RT primer barcodes, nucleic acid tag barcodes, UDI barcodes, and combinations thereof). Each barcode combination should correspond to the cDNA from a single cell. In some embodiments, only reads with valid barcodes are retained. For the WT library, the sequencing reads with each barcode combination can be aligned to a reference genome, e.g., to a reference human genome. Multiple reads with the same random identifier sequence are counted as a single read. In some embodiments, reads with random identifier sequences with two or less mismatches are assumed to be generated by sequencing errors and are counted as a single read. In some embodiments, transcripts represented by one read only are filtered out and not included in subsequent analysis of the libraries.

[0225] Sequence reads from each library (i.e., the whole transcriptome library and the CRISPR or other target library) comprising the same barcode sequence combinations are then associated to correlate the expression of, e.g., a given gRNA with the whole transcriptome in the same

cell or nucleus. In some embodiments, specific other targets can be assessed as well, e.g., the genes specifically targeted by the different gRNAs, e.g., as a control to confirm an expected effect of a given gRNA on the expression of its target. Such controls can provide further support for the association between specific genetic perturbations or expression events and complex phenotypes such as altered transcriptional profiles.

Kits

[0226] Another aspect of the disclosure relates to kits for preparing and sequencing enriched sequencing libraries, e.g., sequencing libraries for performing whole transcriptome sequencing together with enriched sequencing of a target transcript or transcripts of interest according to the herein-disclosed methods. In some embodiments, the kit may comprise at least one reverse transcription (RT) primer, e.g., an RT primer as disclosed herein comprising an RT barcode, a sequence such as a poly(dT) sequence, random sequence, or a target sequence, and/or a 5′ overhang sequence. The kit may also comprise a plurality of nucleic acid tags, e.g., nucleic acid tags with well-specific barcodes. Each first nucleic acid tag may comprise a first strand. The first strand may include a barcode sequence, flanked by a 3’ and/or 5’ sequence located 3’ or 5’ of the barcode. In some embodiments, the first strand comprises a 3′ hybridization sequence extending from a 3′ end of a first labeling sequence and/or a 5′ hybridization sequence extending from a 5′ end of the first labeling sequence. Each first nucleic acid tag may further comprise a second strand.

[0227] The kit may also comprise one or more linker strands as described herein, and/or one or more stop oligos according to the present disclosure. In some embodiments, the nucleic acid tags are pre-complexed with their corresponding linker strands in the kit.

[0228] In some embodiments, the kit may comprise one or more sets of nucleic acid tags, e.g., sets of nucleic acid tags configured to be used in a given round of ligation-based tagging. For example, each tag in a given set of nucleic acid tags may comprise the same 5’ and/or 3’ hybridization sequence, with the 5’ and/or 3’ hybridization sequences differing from the 5’ and/or 3’ hybridization sequences used in other sets of tags. Each set of tags may also comprise a plurality of distinct barcode sequences, e.g., 96 different barcode sequences for distinctively labeling cDNAs present in cells or nuclei within each well of a 96-well plate. The different sets of nucleic acid tags may also differ with respect to the presence or absence of additional elements such as capture agent (e.g., biotin), a random sequence, and/or an adapter sequence such as an NGS adapter sequence.

[0229] In various embodiments, the kit may further comprise at least one of a reverse transcriptase, a fixation agent, a permeabilization agent, a ligation agent, and/or a lysis agent.

[0230] In some embodiments, the kit comprises primers for amplifying cDNA molecules according to the present disclosure. In some embodiments, the kit comprises primers for amplifying transcripts being enriched according to the present disclosure, e.g., a U6 primer for amplifying gRNAs according to the present methods. In some embodiments, the kit comprises a second primer for amplifying transcripts being enriched, e.g., for a second round of amplification as described herein. In some embodiments, the kit comprises one or more universal primers for amplifying the whole transcriptome, and/or for amplifying target transcripts when combined with a target transcript specific primer (e.g., a U6 primer). In some embodiments, the kit comprises one or more primers comprising any of the nucleotide sequence shown as SEQ ID NOS:1-302.

[0231] In some embodiments, the kit comprises one or more reaction vessels or containers for the any one or more of the herein-described compositions or methods. For example, in some embodiments the kit comprises one or more multi-well plates such as 96-well plates. In some embodiments, the kit comprises one or more multi-well plates pre-loaded with barcoded RT primers, with nucleic acid tags, or with indexed primers (e.g. UDI primers) according to the present disclosure.

EXAMPLES

[0232] The following examples are illustrative of disclosed methods and compositions. In light of this disclosure, those of skill in the art will recognize that variations of these examples and other examples of the disclosed methods and compositions would be possible without undue experimentation.

Example 1. Validation of enrichment ability for focal barcoding.

[0233] Whole transcriptome and focal libraries were prepared for HEK293 and NIH/3T3 cells. FIG. 8A shows the percentage of cells with gene detected in HEK293 or NIH/3T3 cells in Whole Transcriptome vs. Focal libraries. Whole Transcriptome libraries were sequenced at 10,000 reads/cell. Focal libraries were sequenced at 250 reads/cell. FIG. 8B shows number of unique transcripts captured in HEK293 or NIH/3T3 cells in Whole Transcriptome vs. Focal libraries. Whole Transcriptome libraries were sequenced at 10,000 reads/cell. Focal libraries were sequenced at 250 reads/cell.

Example 2. Enrichment of low to medium expressing genes in 12k and 62k cell human and mouse sublibraries.

[0234] An enrichment protocol was performed comprising a spike-in preamplification step as described herein on two low- to medium-expressing genes in humans (KDELR1) and mice (Psmd2), and the level of enrichment (as measured by virtue of the percentage of cells comprising the gene) in 12k and 62k cell sublibraries. The results are shown in FIGS. 9A-9B. FIGS. 9A-9B illustrate enrichment of moderately expressed genes in human , and improvement in purity following application of a 1 read count threshold filter. FIG. 9A shows two low- to medium-expressing genes were enriched using the herein-disclosed methods in humans (KDELR1) and mice (Psmd2), and the level of enrichment (as measured by virtue of the percentage of cells comprising the gene or reads per cell) was determined in 12k and 62k cell sublibraries.

[0235] As shown in FIG. 9A, whereas in the absence of enrichment the target genes were only present in a fraction of the cells, e.g., in less than 5% of the cells for KDELR1 in the 62k human sublibrary, and at most less than 40% of the cells with the Psmd2 gene in the mouse 12k sublibrary, with enrichment the percentage was about 70%, and typically exceeded 90%.

[0236] FIG. 9B shows sequencing libraries were prepared according to the herein-described methods using two sets of target genes (Psmd2-KDELR1 or Fn1-RPL5) with 10 ng or 50 ng of preamplified cDNA introduced into the first round of target sequence specific amplification performed subsequent to the multiplex "preamplification" round of amplification. The number of amplification cycles performed in this round of amplification was varied from 13-21 cycles. The number of cells with 100% purity was determined, and a filter was applied (read count threshold = 1, 2, 5, or 10) or not applied (read count threshold = 0) to remove transcripts represented by only 1 read, by 2 or fewer reads, by 5 or fewer reads, or by 10 or fewer reads. A substantial increase in purity was observed when a 1-read filter was applied relative to the purity in the absence of a filter. In this experiment, the purity did not increase substantially with more stringent filters (i.e., filters requiring higher numbers of reads per transcript).

Example 3. Filtration of sequencing reads.

[0237] Sequencing libraries were prepared according to the herein-described methods, using either Psmd2 with KDELR1 or Fn1 with RPL5 as target genes, and varying the amounts of preamplified cDNA (e.g., 10 ng or 50 ng) introduced into the first additional round of

amplification performed during the preparation of the target-specific sequencing library (e.g., the first round of amplification subsequent to the multiplex preamplification step, analogous to the "CRISPR PCR" shown in FIG. 6C). The number of amplification cycles performed in this round of amplification was also varied (e.g., from 13-21 cycles).

[0238] Following sequencing, the sequencing reads were filtered to remove transcripts represented by only 1 read, by 2 reads, by 5 reads, or by 10 reads. As shown in FIG. 9B, a substantial increase in purity was observed when a 1-read filter was applied relative to the purity in the absence of a filter. In this experiment, the purity did not increase substantially with more stringent filters (i.e., filters requiring higher numbers of reads per transcript).

Example 4. Spiking in target primers during wt cdna amplification improves the yield and enrichment of target cdna molecules in a combinatorial barcoding-based multiplex scrna-seq protocol

[0239] Cells were fixed and permeabilized, and cDNA generated by reverse transcribing (RT) RNA within the cells using RT primers with well-specific barcodes. The cDNA molecules were tagged within the cells by appending nucleic acid tags (comprising barcodes) to the cDNA via ligation, with the nucleic acid tags appended in the final round also comprising biotin and an adapter sequence. The cells were lysed, and the tagged cDNA molecules isolated from the lysate using streptavidin-coated magnetic beads.

[0240] Second strand synthesis was carried out using a template switching oligonucleotide (TSO) comprising an adapter sequence, and cDNA was amplified with generic WT primers (i.e., primers binding to the TSO adapter sequence and to a TruSeq R2 sequence (see, e.g., FIG. 1A)), or with generic WT primers with gene-specific primers spiked in. Primers specific to RPL5 and to Fn1 genes were used. The target primers used are shown in Table 1, and PCR cycling times and conditions are shown in Table 2.

Table 1. RPL5 and FN1 spike-in primer sequences

primer	Length (nt)	Tm (50 mM NaCl)	Sequence	gene	SEQ ID NO
BC_0452	21	56.5	GTGGAACCGTCCCAAAATGTC	RPL5	SEQ ID NO: 1
BC_0456	21	58.3	GTTCAAACACCCCTCCCACAG	Fn1	SEQ ID NO: 2

Table 1. RPL5 and FN1 spike-in primer sequences					
primer	**Length (nt)**	**Tm (50 mM NaCl)**	**Sequence**	**gene**	**SEQ ID NO**
BC_0460	54 (21)	70.4 (56.5)	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**GTGGAACCGTCCCAAAATGTC**	NexteraR1/RPL5	SEQ ID NO: 3
BC_0463	54 (21)	71.0 (58.3)	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**GTTCAAACACCCCTCCCACAG**	NexteraR1/Fn1	SEQ ID NO: 4
BC_0468	84	72.0	AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCCTTGCTCAGAAGAAGGATCG	P5/NexteraR1/RPL5	SEQ ID NO: 5
BC_0469	85	72.0	AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCACAGGACTCACTTTGTCCCAAC	P5/NexteraR1/Fn1	SEQ ID NO: 6
BC_0467	48	70.8	AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCAGATG	P5/Partial NexteraR1	SEQ ID NO: 7
BC_0468	84	72.0	AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCCTTGCTCAGAAGAAGGATCG	P5/NexteraR1/RPL5	SEQ ID NO: 8
BC_0469	85	72.0	AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCACAGGACTCACTTTGTCCCAAC	P5/NexteraR1/Fn1	SEQ ID NO: 9

Table 1. RPL5 and FN1 spike-in primer sequences					
primer	**Length (nt)**	**Tm (50 mM NaCl)**	**Sequence**	**gene**	**SEQ ID NO**
BC_0467	48	70.8	AATGATACGGCGACCAC CGAGATCTACACTCGTCG GCAGCGTCAGATG	P5/Partial NexteraR 1	SEQ ID NO: 10

Table 2. cDNA amplification (with spike-in) PCR conditions			
Step	**Time**	**Temperature**	**Cycles**
1	3 min	95 °C	1
2	20 sec	98 °C	8 (steps 2-4 repeated 8 times)
3	45 sec	65 °C	8
4	3 min	72 °C	8
5	20 sec	98 °C	N (steps 5-7 repeated N times: N= 7 cycles for 100 ng samples, or 12 cycles for 10 ng samples)
6	20 sec	69 °C	N
7	1 min	72 °C	N

[0241] Following the initial round of cDNA amplification (with or without the spiking in of gene-specific primers), the amplified cDNA was used in an additional round of PCR performed using gene-specific primers in order to enrich the target genes. The enriched cDNA product was analyzed by Tapestation (FIG. 10). Lane A2 contains enriched cDNA from a sample in which no gene-specific primers were spiked in during the initial round of cDNA amplification, and lane C2 contains cDNA from a sample in which primers were spiked-in. The results show strong enrichment in the sample with spiked-in primers (lane C2) relative to the sample with no spike-in (lane A2).

[0242] Sequencing reads obtained from libraries prepared from the different samples were analyzed to determine the fraction of reads with valid barcodes that mapped to either of the targeted genes (RPL5 or Fn1) (FIG. 11). Sublibrary S3, which was prepared from a sample in which target-specific primers were spiked in during cDNA amplification, showed a significantly higher fraction of mapped reads than sublibrary S1, which was prepared from a sample without spiked-in primers.

[0243] The sequencing reads were analyzed to determine the number of target gene transcripts (i.e., RPL5 or Fn1 transcripts) detected per cell (FIG. 12). Sublibrary S3, which was prepared from a sample in which target-specific primers were spiked in during cDNA amplification, showed a much higher number of RPL5 or Fn1 transcripts detected per cell than sublibrary S1, which was prepared from a sample without spiked-in primers.

[0244] The sequencing reads were analyzed to determine the fraction of cells that contained an enriched transcript (i.e., an RPL5 or Fn1 transcript), both without any filtering (FIG. 13A) or following a filtering step to only consider transcripts represented by more than 2 reads (FIG. 13B). The fraction of cells with a target gene transcript was higher in Sublibrary S3 (with spiked-in primers) than in S1 (no spiked-in primers) with or without filtering.

Example 5. Using combinatorial barcoding to simultaneously profile the transcriptome and immune repertoire of 1 million T cells.

[0245] Split-pool combinatorial barcoding technology was used to simultaneously characterize T cell receptors (TCRs) alongside the full transcriptomes of up to 1 million T cells. Using this approach, we were able to recover paired TCR sequences in >50% of individual T cells together with their corresponding transcriptomes. We identified hundreds of thousands of unique alpha and beta clonotypes across 8 donors. Finally, while most are rare in frequency, we identified some hyper-expanded clonotypes that are donor- and cell-type specific.

[0246] Isolated T cells from healthy donor PBMCs were directly profiled, and showed high levels of TCR chain detection, both for all cells (FIG. 13A) and for each of the 8 donors (FIG. 13B). Further, a number of unique alpha and beta chain clonotypes were detected across donors (FIGS. 14A-14B), including nearly four hundred thousand unique alpha chain clonotypes and five hundred thousand unique beta chain clonotypes identified across the 8 donors, with the vast majority being classified as rare clonotypes.

[0247] The results are shown in FIGs. 13A-13B and 14A-14B. As shown in FIGS. 13A-13B, high TCR detection in primary T cells showing sensitive clonotype detection. FIG. 13A shows high TCR chain identification rate. Isolated T cells from healthy donor PBMCs were directly profiled (Primary). Alpha, Beta, and Paired detection were represented in percentages. FIG. 13B shows TCR chain assignment across 8 donors. High rate of chain assignments to both TCR alpha and beta was observed. Among T cells with a detected TCR, paired alpha beta chain assignments ranged between 49%-66%.

[0248] FIGS. 14A-14B show comprehensive Immune Repertoire Detection measured by number of unique alpha and beta chain clonotypes across donors. Nearly four hundred thousand (~400,000) unique alpha chain clonotypes and five hundred thousand (~500,000) unique beta chain clonotypes were identified across the 8 donors. The rare clonotypes (darker color, lower shaded) are defined as only being detected in 1 or 2 cells and the majority of detected clonotypes are rare. FIG. 14A: Unique Alpha Chain. FIG. 14B: Unique Beta Chain.

Example 6. Contribution of spike-in of TCR CDR3 primers in preamplification step.

[0249] To investigate the contribution of spiked-in TCR-specific primers during the preamplification step (i.e., during the initial round of multiplex cDNA amplification prior to the separate preparation of WT sequencing libraries and TCR sequencing libraries), WT and TCR sequencing libraries were prepared with or without the inclusion of TCR-specific preamplification primers (i.e., with or without the "spiking-in" of TCR-specific primers during preamplification).

[0250] Two sets of sequencing libraries were prepared in parallel from activated and resting T cells according to the present methods. In one set of sequencing libraries, no TCR-specific primers were included during the preamplification step (i.e., only non-specific primers capable of amplifying the whole transcriptome were included). For example, these are generic primers (e.g., specific to adapter sequences) introduced to all cDNA molecules via the template switching oligonucleotide (TSO) or the last-added nucleic acid tag. In the second set of sequencing libraries, primers specific to CDR3 sequences within the variable regions of both TCR alpha and beta chains were included (or "spiked-in") to the amplification reaction.

[0251] The results, shown in FIGS. 16A-16D, indicate that the spiking in of the TCR-specific preamplification primers (spike-in +ve) led to an increase in the detection of both alpha and

beta chains in both resting and activated T cells compared sequencing library prepared without TCR-specific preamplification primers added (spike-in -ve).

[0252] FIGS. 16A-16D show increased detection of TCR alpha and beta chains with spiking in of TCR-specific primers during first multiplex cDNA amplification step (i.e., "preamplification" step). Whole transcriptome (WT) and TCR-specific sequencing libraries were prepared from activated and resting T cells with or without spiked-in primers, and the percentages of cells with detected alpha (FIGs. 16A and 16B) and/or beta (FIGs. 62C and 16D) chains were determined. FIG. 16A shows percentages of activated T cells with detected alpha chain or with no detected chain, in libraries prepared with or without TCR-specific preamplification primers. FIG. 16B shows percentages of resting T cells with detected alpha chain or with no detected chain, in libraries prepared with or without TCR-specific preamplification primers. FIG. 16C shows percentages of activated T cells with detected beta chain or with no detected chain, in libraries prepared with or without TCR-specific preamplification primers. FIG. 16D shows percentages of resting T cells with detected beta chain or with no detected chain, in libraries prepared with or without TCR-specific preamplification primers.

Example 7. Exemplary primers for use in embodiments of the present methods.

[0253] This example provides exemplary multiplex preamplification primers for use in embodiments of the present methods, such as CRISPR, show in Tables 3 and 4.

Table 3. Multiplex CRISPR-specific preamplification primers

Name	Sequence	Description	SEQ ID NO.
BC_0747	GGGCCTATTTCCCATG ATTCCTTC	CRISPR specific preamplification primer	SEQ ID NO:11
BC_0363	GTGACTGGAGTTCAG ACGTGTGCTCTTCCGA TCT	Generic (non-specific) cDNA preamplification/amplificat ion primer	SEQ ID NO:12
BC_0108	AAGCAGTGGTATCAA CGCAGAGT	Generic (non-specific) cDNA preamplification/amplificat ion primer	SEQ ID NO:13

Table 4. CRISPR specific library amplification primers			
Name	Sequence	Description	SEQ ID
BC_0748	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGCTTTATATATCTTGTGGAAAGGACGA	CRISPR specific amplification primer	SEQ ID NO:14
BC_0363	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	Generic (non-specific) cDNA preamplification/amplification primer	SEQ ID NO: 12

Example 8. Exemplary protocols for the multiplex labeling of the whole transcriptome and target sequences of interest

Workflow

[0254] CRISPR Detect enables analysis of single guide RNAs (sgRNAs) in studies using CROP-seq or similar methods. Compatible methods generate a polyadenylated transcript containing the sgRNA sequence downstream of a human U6 promoter. When CRISPR Detect is combined with Evercode WT Mini v2, paired sgRNA detection and whole transcriptome expression can be analyzed in up to 10,000 cells across up to 12 different biological samples or experimental conditions. Evercode Cell Fixation kits convert the cells into individualized reaction compartments, thus avoiding the requirement for dedicated microfluids hardware. Through three rounds of barcoding, the transcriptome of each fixed cell is uniquely labeled. In each round, pooled cells are randomly distributed into different wells, and transcripts are labeled with well-specific barcodes. Barcoded transcripts are amplified during cDNA amplification, and sgRNA containing polyadenylated transcripts are enriched with human U6 specific primers. Amplified cDNA is split to create Whole Transcriptome and CRISPR sequencing libraries. After sequencing, the Parse Biosciences Analysis Pipeline assigns reads that share the same four barcode combination to a single cell and associates sgRNAs to the appropriate cell.

in situ cell barcoding

[0255] FIGS. 3A-3D provide an overview of *in situ* cell barcoding steps (i.e., combinatorial barcoding, or split-pool labeling) according to embodiments of the present methods. FIGS. 4A-4D provide an overview of cDNA capture and amplification steps for CRISPR screens such as CROP-seq or similar methods.

[0256] Set up and Sample Counting Before barcoding, cells are thawed and counted. Cool a centrifuge with swinging bucket rotors to 4°C. Set a heat block to 37°C. Fill a bucket with ice. Prepare a hemocytometer, flow cytometer, or other cell counting device. Place the Round 1 Plate into a thermocycler and thaw. Thaw the previously fixed cell samples in a heat block set to 37°C until all ice crystals dissolve. Thoroughly mix each sample by pipetting and store on ice. While minimizing time on ice, count the number of cells in your sample with a hemocytometer or alternative cell counting device. Record the cell count. Dilute each sample to appropriate extent and store on ice. Do not discard the Dilution Buffer.

Barcoding Round 1

[0257] Samples are loaded into Round 1 Plate. An *in situ* reverse transcription reaction adds well-specific barcodes that also serve as sample barcodes. Cells are then pooled, centrifuged, and resuspended. To add round 1 barcodes: Gently remove the Round 1 Plate from the thermocycler, return to the green rack, and centrifuge for 1 minute at 100 x g at 4°C. Remove the Round 1 Plate from the centrifuge. Remove the plate seal and store on ice. Without reusing tips, add 14 µL (or other appropriate volume) of each diluted sample to the appropriate wells of Round 1 Plate. Mix immediately after dispensing each sample by pipetting 3x and keeping the plate on ice. Remove the Round 1 Plate from ice. Add a new PCR plate seal. Place the Round 1 Plate into a thermocycler and run the program shown in Table 5. Upon completion, proceed immediately to the next step.

Table 5. Barcoding Round 1

Run time	Lid Temperature	Sample Volume	
40 min	70 °C	40 uL	
Step	Time	Temperature	Cycles
1	10 min	50 °C	1
2	12 s	8 °C	3
3	45 s	15 °C	3

Table 5. Barcoding Round 1			
4	45 s	20 °C	3
5	30 s	30 °C	3
6	2 min	42 °C	3
7	3 min	50 °C	3
8	5 min	50 °C	1
9	Hold	4 °C	Hold

[0258] Remove the Round 1 Plate from the thermocycler, return to the green rack, and store on ice. Remove the Round 2 Plate from the blue rack, place it into a thermocycler, and thaw. Proceed immediately to the next step while the (thaw) program is still running. While secured in the green rack on a flat surface, remove the plate seal from the Round 1 Plate and store the uncovered plate on ice. With both on ice, transfer all the liquid in the Round 1 Plate to a 15 mL tube. Add 2.4 µL of Spin Additive to the 15 mL tube with pooled cells. Mix by gently inverting the tube just once. Centrifuge the 15 mL tube in a swinging bucket rotor for 5-10 minutes at 200-500 x g at 4°C. Immediately move to the next step after centrifugation. Remove the supernatant until about ~40 µL of liquid remains above the pellet. Fully but gently resuspend the pellet in 1 mL of Resuspension Buffer. 5. Add an additional 1 mL of Resuspension Buffer for a total addition of 2 mL. Store on ice. Proceed immediately to Round 2 Barcoding.

Barcoding Round 2

[0259] The pooled cells are added to the Ligation Master Mix, which is loaded into the Round 2 Plate. An *in situ* ligation reaction adds a well-specific barcode to the 3' end of the cDNA. The ligation reaction is quenched with Round 2 Stop Mix, and the cells are pooled and strained. On ice, add 20 µL of Round 2 Ligation Enzyme directly into the Ligation Mix tube to create the Ligation Master Mix. Using a P1000 pipette, add 2 mL of sample in Resuspension Buffer into the Ligation Master Mix. Mix thoroughly by pipetting 10x with a P1000 set to 1000 µL. Store on ice. Remove the Round 2 Plate from the thermocycler, return to the blue rack, and centrifuge for 1 minute at 100 x g at 4°C. Keep at room temperature. While secured in the blue rack on a flat surface, remove the plate seal from the Round 2 Plate. Transfer the entire volume in the Ligation Master Mix tube to a basin with a P1000. At room temperature and without reusing tips, transfer 40 µL from the basin to each well in the Round 2 Plate. While secured in

the blue rack on a flat surface, add a new plate seal to the Round 2 Plate. Place the Round 2 Plate into a thermocycler and run the program shown in Table 6. Upon completion, proceed immediately to the next step.

Table 6. Barcoding Round 2		
Run time	Lid Temperature	Sample Volume
30 min	50 °C	50 uL
Step	Time	Temperature
1	30 min	37 °C
2	Hold	4 °C

[0260] Briefly vortex the Round 2 Stop Mix. Transfer the entire volume of this tube to a new basin with a P1000. Remove the Round 2 Plate from the thermocycler and return to the blue rack. While secured in the blue rack on a flat surface, remove the plate seal from the Round 2 Plate. At room temperature and without reusing tips, transfer 10 µL of the Round 2 Stop Mix from the basin to each well in the Round 2 Plate with a multichannel P20. After each transfer, mix by pipetting exactly 3x. While secured in the blue rack on a flat surface, add a new plate seal to the Round 2 Plate. Place the Round 2 Plate into a thermocycler and run the program shown in Table 7. Upon completion, proceed immediately to the next step.

Table 7. Round 2 Stop.		
Run time	Lid Temperature	Sample Volume
30 min	50 °C	60 uL
Step	Time	Temperature
1	30 min	37 °C
2	Hold	4 °C

[0261] Transfer the Round 2 Plate from the thermocycler to the blue rack and store it at room temperature. Place the Round 3 Plate into a thermocycler and thaw. Proceed immediately to the next step while the (thaw) program is still running. While secured in the blue rack on a flat surface, remove the plate seal from Round 2 Plate. At room temperature, transfer all the liquid in the Round 2 Plate into a new basin. Pipette the sample through a 40 µm strainer into a new basin with a P1000. Before each transfer, gently mix the cells in the basin by pipetting 2x and tilt the basin to recover as much liquid as possible.

Barcoding Round 3

[0262] Round 3 Ligation Enzyme is added to the pooled cells, which are then loaded into the Round 3 Plate. A second in situ ligation reaction adds a third well-specific barcode, the Illumina Truseq R2 sequence, and a biotin. The sample is then pooled and strained. To add round 3 barcodes: Add 20 µL of Round 3 Ligation Enzyme to the basin containing the strained sample. Mix by gently pipetting 20x with a P1000 set to 1000 µL. Remove the Round 3 Plate from the thermocycler, return to the orange rack, and centrifuge for 1 minute at 100 x g at 4°C. While secured in the orange rack on a flat surface, remove the plate seal from the Round 3 Plate. At room temperature and without reusing tips, transfer 50 µL from the basin to each well in the Round 3 Plate. While secured in the orange rack on a flat surface, add a new plate seal to the Round 3 Plate. Place the Round 3 Plate into a thermocycler and run the program shown in Table 8.

Table 8. Barcoding Round 3		
Run time	Lid Temperature	Sample Volume
30 min	50 °C	60 uL
Step	Time	Temperature
1	30 min	37 °C
2	Hold	4 °C

[0263] Briefly vortex the Round 3 Stop Mix. Transfer the entire volume to a new basin with a P1000. Remove the Round 3 Plate from the thermocycler and return it to the orange rack. While secured in the orange rack on a flat surface, remove the plate seal from the Round 3 Plate. At room temperature and without reusing tips, transfer 20 µL of the Round 3 Stop Mix from the basin to each well in the Round 3 Plate with a multichannel P20. After each transfer, mix by pipetting exactly 3x. Without incubation, proceed immediately to the next step. At room temperature, transfer all the liquid in the Round 3 Plate into a new basin. Pipette the sample through a 40 µm strainer into a new 15 mL tube with a P1000. Before each transfer, gently mix the cells in the basin by pipetting 2x and tilt the basin to recover as much liquid as possible. Proceed immediately to Lysis and Sublibrary Generation.

Lysis and Sublibrary Generation

[0264] The cell pool is centrifuged, washed, and resuspended in Dilution Buffer. The cells are counted and divided into sublibraries. These sublibraries are lysed and stored at -80°C. To generate and lyse sublibraries: Add 70 µL of Spin Additive to the 15 mL tube with the sample. Gently invert once to mix. Centrifuge the 15 mL tube in a swinging bucket rotor for 5-10 minutes at 200-500 x g at 4°C. Immediately move to the next step after centrifugation. Remove the supernatant until about ~40 µL of liquid remains above the pellet. Use a P1000 for the first 6 mL and then a P200 for the remaining volume. Fully but gently resuspend the pellet in 1 mL of Pre-Lyse Wash Buffer. Add an additional 3 mL of Pre-Lyse Wash Buffer for a total addition of 4 mL. Centrifuge the 15 mL tube in a swinging bucket rotor for 5-10 minutes at 200-500 x g at 4°C. Immediately move to the next step after centrifugation. Remove the supernatant until about ~40 µL of liquid remains above the pellet. Use a P1000 for the first 3 mL and then a P200 for the remaining volume. Fully but gently resuspend the pellet with the remaining supernatant in the 15mL tube. Measure the volume of the resuspended cells with a P200 pipette. If less than 60 µL are measured, add Dilution Buffer to a total of 60 µL. If more than 60 µL are measured, proceed without adding Dilution Buffer. Store on ice.

[0265] While minimizing time on ice, count the number of cells in your sample with a hemocytometer or alternative cell counting device. Record the count. Decide how to divide cells across the sublibraries. Do not overload a sublibrary. The recommended loading is up to 5,000 cells/sublibrary, with a maximum of 10,000 cells/sublibrary. Adding additional cells may result in an increased multiplet rates. Ensure the cells are in suspension by pipetting 5x with a P200 set to 45 µL before each transfer. Add the appropriate volume of sample to 2 separate 0.2 mL tubes. Add the appropriate volume of Dilution Buffer to the 0.2 mL tubes for a total volume of 25 µL. Prepare the Lysis Master Mix in a new 1.5 mL tube. Add 30 µL of Lysis Master Mix to each 0.2 mL tube with diluted cells. Store at room temperature. Vortex the 0.2 mL tube(s) for 10 seconds. Briefly centrifuge. Place the tubes into a thermocycler and run the program shown in Table 9.

Table 9. Cell Lysis		
Run time	Lid Temperature	Sample Volume
60 min	80 °C	55 uL
Step	Time	Temperature
1	60 min	65 °C

Table 9. Cell Lysis		
2	Hold	4 °C

[0266] Freeze the lysate(s) at -80°C. Safe stopping point: Sublibrary lysates are stable for up to 6 months at -80°C.

cDNA Capture and Amplification – Binder Bead Preparation

[0267] During the Binder Bead Preparation step, Streptavidin-coated Binder Beads are washed. To prepare for cDNA capture and amplification: Fill an ice bucket. Prepare 400 µL of 85% ethanol per lysate with nuclease-free water. Equilibrate 80 µL of SPRI beads per lysate to room temperature. Vortex with an adaptor for 96 well plates. Vortex Binder Beads until fully mixed. Add an appropriate volume of Binder Beads to a new 1.5 mL tube, depending on the number of lysates being processed. Place the tube on the magnetic rack for 1.5 mL tubes until the solution clears (~2 minutes). Remove and discard the supernatant. Remove the tube from the magnetic rack and fully resuspend the bead pellet in the appropriate volume of Bead Wash Buffer. Place the tube on the magnetic rack for 1.5 mL tubes until the solution clears (~2 minutes). Remove and discard the supernatant. Repeat wash steps twice for a total of 3 washes. Remove the tube from the magnetic rack. Fully resuspend the pellet in the appropriate volume of Bind Buffer A. Store the 1.5 mL tube at room temperature and proceed immediately to cDNA capture.

cDNA Capture

[0268] The barcoded cDNA is captured with streptavidin-coated magnetic Binder Beads and washed to remove cellular debris. To capture the cDNA: Remove the desired tube(s) of lysate from -80°C. Incubate the tube(s) in a heat block or thermocycler at 37°C for 5 minutes. Briefly centrifuge and store at room temperature. Briefly centrifuge Lysis Neutralizer and gently mix by pipetting 2x with a P20 set to 15 µL. Add 2.5 µL of Lysis Neutralizer to each tube of lysate. Mix 5x with a P200 pipette set to 40 µL. Briefly centrifuge. Incubate for 10 minutes at room temperature. Fully mix the Binder Beads by pipetting 3x. Add 50 µL of Binder Beads to each tube of lysate. Fully mix by pipetting 5x with a P200 set to 90 µL. Place the tube(s) into a 96 well PCR tube rack, press to secure, and ensure the caps are secured tightly. Place the lid on the rack. Place the rack onto a vortex mixer with a plate adaptor. Push to secure. Vortex on 20% power (~800-1000 RPM) for 60 minutes at room temperature. Remove the tube(s) from

the vortex mixer with a plate adaptor. Briefly vortex the tube(s) on a standard vortex adaptor. Briefly centrifuge without letting beads collect at the bottom of the tube(s).

[0269] Place the tube(s) on the high magnet position of the Parse Biosciences Magnetic Rack so the magnet is closer to the top of the 0.2 mL tubes. Incubate until the solution clears (~2 minutes. While still on the magnetic rack, remove and discard the supernatant. Remove the tube(s) from the magnetic rack. Fully resuspend each bead pellet with 125 µL Bind Buffer B. Incubate for 1 minute at room temperature. Return the tube(s) to the high position of the magnetic rack. Incubate until the solution clears (~2 minutes). While still on the magnetic rack, remove and discard the supernatant. Repeat wash steps once for a total of 2 washes with Bind Buffer B. Remove the tube(s) from the magnetic rack. Fully resuspend each bead pellet with 125 µL Bead Storage Buffer. Incubate for 1 minute at room temperature. Proceed immediately to cDNA template switch.

cDNA Template Switch

[0270] After an additional wash, a template switching reaction is added to the captured cDNA. This template switching reaction adds a 5' adaptor to the cDNA. To template switch: Prepare the Template Switch Master Mix in a new 1.5 mL tube. Mix by pipetting 10x and store on ice. Place each tube of captured cDNA on the high position of the magnetic rack. Incubate until the solution clears (~2 minutes). While still on the magnetic rack, remove and discard the supernatant. While still on the magnetic rack, add 125 µL of Bind Buffer C to each tube. While still on the magnetic rack, remove and discard the Bind Buffer C. Remove the tube(s) from the magnetic rack. Fully resuspend each bead pellet with 100 µL of the Template Switch Master Mix. Briefly centrifuge without letting beads collect at the bottom of the tube(s). Incubate for 30 minutes at room temperature. Fully resuspend each bead pellet by mixing up and down 5x with a P200 set to 75 µL. Place the tube(s) into a thermocycler and run the program shown in Table 10.

Table 1. cDNA template switch.		
Run time	Lid Temperature	Sample Volume
90 min	70 °C	100 uL
Step	Time	Temperature
1	90 min	42 °C
2	Hold	4 °C

[0271] If not immediately continuing to cDNA amplification, place the tube(s) on the high position of the magnetic rack. Incubate until the solution clears (~2 minutes). While still on the magnetic rack, remove and discard the supernatant. Remove the tube(s) from the magnetic rack. Fully resuspend each bead pellet with 125 µL Bead Storage Buffer. Safe stopping point: Captured cDNA can be stored at 4°C for up to 18 hours. Do not freeze.

cDNA Amplification ("multiplex cDNA amplification," or "preamplification")

[0272] The captured cDNA is washed and amplified with TS- and WT preamplification primers (e.g., Illumina Truseq R2- specific primers, or compatible equivalent thereof). The target cDNA is enriched using target pairs of primers (target specific preamplification primers). For example, sgRNA transcripts are enriched with Human U6 primers. To amplify the cDNA: Prepare the Amplification Reaction Solution Master Mix in a new 1.5 mL tube as follows. Mix by pipetting 10x and store on ice. Place each tube on the high position of the magnetic rack. Incubate until the solution clears (~2 minutes). While still on the magnetic rack, remove and discard the supernatant. While still on the magnetic rack, add 125 µL of Bind Buffer C to each tube. Incubate for 1 minute at room temperature. While still on the magnetic rack, remove and discard the Bind Buffer C. Remove tube(s) from the magnetic rack. Fully resuspend each bead pellet with 100 µL of the Amplification Reaction Master Mix. Store on ice. Determine the number of PCR cycles required for cDNA amplification based on the recommendations in Table 11. Although these recommendations are appropriate for many cell types, the number of cycles may need to be optimized for your sample type.

Table 11. Number of PCR cycles in multiplex amplification or preamplification step

Cells in sublibrary	High RNA content such as cell lines	Low RNA content such as PBMCs
200-1,000	12	14
1,000-2,000	10	12
2,000-6,000	8	10
6,000-10,000	6	8

[0273] Place the tube(s) into a thermocycler and run the program shown in Table 12.

Table 12. cDNA amplification conditions ("preamplification" phase, for multiplex WT amplification with spiked in primers for target cDNA enrichment			
Run time	Lid Temperature	Sample Volume	
60-80 min	105 °C	100 uL	
Step	Time	Temperature	Cycles
1	3 min	95 °C	1
2	20 sec	98 °C	5
3	45 sec	65 °C	5
4	3 min	72 °C	5
5	20 sec	98 °C	Variable (see Table 11)
6	20 sec	67 °C	Variable (see Table 11)
7	3 min	72 °C	Variable (see Table 11)
8	5 min	72 °C	1
9	Hold	4 °C	1

Post-Amplification Purification

[0274] Amplified cDNA is purified with a 0.8x SPRI bead cleanup. To purify the cDNA: Gather 400 µL of freshly prepared 85% ethanol per tube of amplified cDNA with nuclease-free water. Gather room temperature SPRI beads (80 µL per tube of amplified cDNA). Place each tube of amplified ("preamplified") cDNA on the high position of the magnetic rack. Incubate until the solution clears (~2 minutes). While still on the magnetic rack, transfer 90 µL of the supernatant containing the cDNA into a new 0.2 mL tube(s). Store at room temperature. Vortex the SPRI beads until fully mixed. Add 72 µL of SPRI beads to each tube with amplified cDNA. Vortex the tube(s) for 5 seconds. Briefly centrifuge. Incubate for 5 minutes at room temperature. Place the tube(s) on the high position of the magnetic rack. Incubate until the solution clears (~2 minutes). While still on the magnetic rack, remove and discard the supernatant. While still on the magnetic rack, add 180 µL of 85% ethanol to each tube. Incubate for 1 minute at room temperature. While still on the magnetic rack, remove and discard the supernatant. Repeat steps 10-12 once for a total of 2 washes.

[0275] Remove any residual ethanol with a P20. While still on the magnetic rack, air dry the SPRI beads (~2 minutes). Remove the tube(s) from the magnetic rack. Fully resuspend each bead pellet with 20 µL of nuclease-free water. Incubate for 10 minutes at 37°C in a thermocycler. Place the tube(s) on the low magnet position of the magnetic rack, so the magnet is closer to the bottom of the 0.2 mL tubes. Incubate until the solution clears (~2 minutes). While still on the magnetic rack, transfer 20 µL of the supernatant containing the purified cDNA into new 0.2 mL tube(s). Store on ice. Safe stopping point: Amplified cDNA can be stored at 4°C for up to 48 hours or at -20°C for up to 3 months.

cDNA Quantification

[0276] The concentration and size distribution of the cDNA are measured with fluorescent dyes and capillary electrophoresis. The cDNA is then stored at 4°C for up to 48 hours or at -20°C for up to 3 months. To quantify the cDNA: Measure the concentration of each tube of purified cDNA with, e.g., the Qubit dsDNA HS (High Sensitivity) Assay Kit according to the manufacturer's instructions. Record the concentration(s). Assess the size distribution of each tube of purified cDNA with a High Sensitivity DNA Kit on the Agilent Bioanalyzer System or High Sensitivity D5000 ScreenTape and Reagents on the Agilent TapeStation System according to the manufacturer's instructions. Samples may need to be diluted to be within the manufacturer's recommended concentration range. Typically, between a 1:3 to 1:10 dilution is appropriate. Purified cDNA can be stored at 4°C for up to 48 hours or at -20°C for up to 3 months. Otherwise, proceed immediately to WT Sequencing Library Preparation.

WT Sequencing Library Preparation – Fragmentation, End Repair, and A-tailing

[0277] Barcoded and amplified cDNA is fragmented, end repaired and A-tailed in a single reaction. To prepare for sequencing library preparation: Prepare 1.2 mL 85% ethanol per sublibrary with nuclease-free water. Equilibrate approximately 180 µL of SPRI beads per sublibrary to room temperature. Fill an ice bucket. Take out the magnetic rack for 0.2 mL PCR tubes. Obtain recorded cDNA concentrations. Vortex the tube(s) of cDNA for 5 seconds. Briefly centrifuge. Prepare diluted cDNA in new 0.2 mL tube(s) as follows and store on ice. Bring the total volume up to 35 µL. Store any remaining sublibrary cDNA at -20°C. Vortex the Fragmentation Buffer for 5 seconds. Briefly centrifuge. Prepare the Fragmentation Master Mix in a new 1.5 mL tube. Mix by pipetting 10x and store on ice. Add 15 µL of Fragmentation Master Mix to each tube of diluted cDNA. Mix by pipetting 10x with a P200 multichannel

pipette set to 40 µL. Briefly centrifuge. Place the tube(s) into a cooled thermocycler and perform the program shown in Table 13.

Table 13. Sublibrary Fragmentation, End Repair, and A-Tailing		
Run time	Lid Temperature	Sample Volume
40 min	70 °C	50 uL
Step	Time	Temperature
1	Hold	4 °C
2	10 min	32 °C
3	30 min	65 °C
4	Hold	4 °C

[0278] As soon as the program reaches 4°C (step 4 of the thermal cycling program), store the tube(s) on ice and proceed immediately to Post-A-tailing Size Selection.

Post-A-tailing Size Selection

[0279] Fragmented and A-tailed DNA is size selected with a double sided SPRI cleanup. To size select the DNA: Gather freshly prepared 85% ethanol. Gather room temperature SPRI beads (~50 µL per sublibrary). Vortex the SPRI beads until fully mixed. Add 30 µL of SPRI beads to each tube of Atailed DNA. Vortex the tube(s) for 5 seconds. Briefly centrifuge. Incubate for 5 minutes at room temperature. Place the tube(s) on the high position of the magnetic rack for 0.2 mL tubes. Incubate until the solution clears (~2 minutes). While still on the magnetic rack, transfer 75 µL of the supernatant containing the Atailed DNA into new 0.2 mL tube(s). Discard the tube(s) with bead pellet(s). Add 10 µL of SPRI beads to each tube. Vortex the tube(s) for 5 seconds. Briefly centrifuge. Incubate for 5 minutes at room temperature. Place the tube(s) on the high position of the magnetic rack for 0.2 mL tubes. Incubate until the solution clears (~3 minutes) before proceeding. While still on the magnetic rack, remove and discard the supernatant. While still on the magnetic rack, add 180 µL of 85% ethanol to each tube. Incubate for 1 minute at room temperature. While still on the magnetic rack, remove and discard the supernatant. Repeat steps 13-15 once for a total of 2 washes. Remove any residual ethanol with a P20. While still on the magnetic rack, air dry the SPRI beads (~30 seconds). Remove the tube(s) from the magnetic rack. Fully resuspend each bead pellet with 50 µL of nuclease-free water. Incubate for 5 minutes at room temperature. Place the tube(s) on the high position of the magnetic rack. Incubate until the solution clears (~2

minutes). While still on the magnetic rack, transfer 50 µL of the supernatant into new 0.2 mL tube(s). Safe stopping point:The size-selected A-tailed DNA can be stored at 4°C for up to 18 hours or at -20°C for up to 2 weeks.

Adaptor Ligation

[0280] Adaptors with an Illumina Truseq R2 sequence are ligated to the 5' end of the fragmented DNA. To ligate adaptors: Prepare the Adaptor Ligation Master Mix in a new 1.5 mL tube. Mix by pipetting 10x and store on ice. Add 50 µL of Adaptor Ligation Master Mix to each tube of purified A-tailed DNA. Mix by pipetting 10x with a P200 multichannel pipette set to 80 µL. Briefly centrifuge. Place the tube(s) into a thermocycler and run the program shown in Table 14. As soon as the program reaches 4°C, store the adaptor ligated DNA on ice and proceed immediately to Post-ligation purification.

Table 14. Sublibrary Adaptor Ligation		
Run time	Lid Temperature	Sample Volume
15 min	30 °C	100 uL
Step	Time	Temperature
1	15 min	20 °C
2	Hold	4 °C

Post-Ligation Purification

[0281] Adaptor ligated DNA is purified with a 0.8x SPRI bead cleanup. To purify the ligated DNA: Gather freshly prepared 85% ethanol. Gather room temperature SPRI beads (~90 µL per sublibrary). Vortex the SPRI beads until fully mixed. Add 80 µL of SPRI beads to each tube of adaptor ligated DNA. Vortex the tube(s) for 5 seconds. Briefly centrifuge. Incubate for 5 minutes at room temperature. Place the tube(s) on the high position of the magnetic rack for 0.2 mL tubes. Incubate until the solution clears (~2 minutes). While still on the magnetic rack, remove and discard the supernatant. While still on the magnetic rack, add 180 µL of 85% ethanol to each tube. Incubate for 1 minute at room temperature. While still on the magnetic rack, remove and discard the supernatant. Repeat steps 8-10 once for a total of 2 washes. Remove any residual ethanol with a P20. While still on the magnetic rack, air dry the SPRI beads (~2 minutes). Remove the tube(s) from the magnetic rack. Fully resuspend each bead pellet with 23 µL of nuclease-free water. Incubate for 5 minutes at room temperature. Place the

tube(s) on the low magnet position of the magnetic rack so the magnet is closer to the bottom of the 0.2 mL tubes. Incubate until the solution clears (~2 minutes). While still on the magnetic rack, transfer exactly 21 µL of the supernatant containing the purified adaptor ligated DNA into new 0.2 mL tube(s). Store on ice. Proceed immediately to Barcoding Round 4.

Barcoding Round 4

[0282] Purified adaptor ligated DNA is PCR amplified with Illumina Truseq R1 and R2 primers. This indexing PCR generates sequencing libraries and adds i5/i7 UDIs that act as a fourth cell barcode. To add round 4 barcodes: Centrifuge the UDI Plate - WT at 100 x g for 1 minute. Wipe the surface of the plate with 70% ethanol and allow to dry. Orient the UDI Plate - WT with the notch on the bottom left as pictured below. For each sublibrary being processed, choose one unused well of the UDI Plate - WT and record the well position and number for each sublibrary. With a multichannel P20, pierce the seal of the chosen wells in the UDI Plate - WT. With a multichannel P20 and new tips, add 4 µL from a different well of the UDI Plate - WT to each tube of purified adaptor ligated DNA. If any unused wells remain in the UDI Plate - WT, store the plate at -20°C. Add 25 µL of Index PCR Mix to each tube. Mix by pipetting 10x with a P200 multichannel pipette set to 25 µL. Briefly centrifuge. Determine the number of PCR cycles required for indexing PCR based on the concentration of cDNA added to the fragmentation reaction as previously recorded, based on the recommendations shown in Table 15. The UDI sequences in the UDI Plate – WT are shown in Table 16.

Table 15. Number of PCR cycles

cDNA input (ng)	PCR cycles
10-24	13
25-49	12
50-99	11
100-299	10
300-999	8
1,000 or more	7

Table 16. UDI WT Plate sequences							
Index name	**well_position**	**i7_index**	**SEQ ID NO.:**	**i5_index_reverse_complement**	**SEQ ID NO.:**	**i5_index**	**SEQ ID NO.:**
UDI_Plate_WT_1	A1	CAGATCAC	15	ATGTGAAG	63	CTTCACAT	111
UDI_Plate_WT_2	B1	ACTGATAG	16	GTCCAACC	64	GGTTGGAC	112
UDI_Plate_WT_3	C1	GATCAGTC	17	AGAGTCAA	65	TTGACTCT	113
UDI_Plate_WT_4	D1	CTTGTAAT	18	AGTTGGCT	66	AGCCAACT	114
UDI_Plate_WT_5	E1	AGTCAAGA	19	ATAAGGCG	67	CGCCTTAT	115
UDI_Plate_WT_6	F1	CCGTCCTA	20	CCGTACAG	68	CTGTACGG	116
UDI_Plate_WT_7	G1	GTAGAGTA	21	CATTCATG	69	CATGAATG	117
UDI_Plate_WT_8	H1	GTCCGCCT	22	AGATACGG	70	CCGTATCT	118
UDI_Plate_WT_9	A2	GTGAAACT	23	TACAGACT	71	AGTCTGTA	119
UDI_Plate_WT_10	B2	TCATTCCT	24	AATGCCTG	72	CAGGCATT	120
UDI_Plate_WT_11	C2	GGTAGCAT	25	TGCTTGCC	73	GGCAAGCA	121
UDI_Plate_WT_12	D2	ACTTGATC	26	TTTGGGTG	74	CACCCAAA	122
UDI_Plate_WT_13	E2	ATGAGCAT	27	GAATCTGA	75	TCAGATTC	123
UDI_Plate_WT_14	F2	GCGCTATC	28	CGACTGGA	76	TCCAGTCG	124
UDI_Plate_WT_15	G2	TGACCAGT	29	ACATTGGC	77	GCCAATGT	125
UDI_Plate_WT_16	H2	TATAATCA	30	ACCACTGT	78	ACAGTGGT	126
UDI_Plate_WT_17	A3	CAAAAGTC	31	CGGTTGTT	79	AACAACCG	127
UDI_Plate_WT_18	B3	CGATGTCA	32	CATGAGGA	80	TCCTCATG	128
UDI_Plate_WT_19	C3	CTCAGAGT	33	TGGAGAGT	81	ACTCTCCA	129
UDI_Plate_WT_20	D3	TAATCGAC	34	TGACTTCG	82	CGAAGTCA	130
UDI_Plate_WT_21	E3	CATTTTCT	35	GGAAGGAT	83	ATCCTTCC	131
UDI_Plate_WT_22	F3	CTATACTC	36	TGTTCGAG	84	CTCGAACA	132
UDI_Plate_WT_23	G3	CACTCACA	37	AAGGCTGA	85	TCAGCCTT	133

Table 16. UDI WT Plate sequences							
Index name	**well_position**	**i7_index**	**SEQ ID NO.:**	**i5_index_reverse_complement**	**SEQ ID NO.:**	**i5_index**	**SEQ ID NO.:**
UDI_Plate_WT_24	H3	CTCGAACA	38	CTCGAGTG	86	CACTCGAG	134
UDI_Plate_WT_25	A4	CTCTATCG	39	ATCGGTGG	87	CCACCGAT	135
UDI_Plate_WT_26	B4	TCCTCATG	40	AGGTCTTG	88	CAAGACCT	136
UDI_Plate_WT_27	C4	AACAACCG	41	AGGAAGCG	89	CGCTTCCT	137
UDI_Plate_WT_28	D4	GCCAATGT	42	ACATGTGT	90	ACACATGT	138
UDI_Plate_WT_29	E4	TGGTTGTT	43	ATACAGTT	91	AACTGTAT	139
UDI_Plate_WT_30	F4	TCTGCTGT	44	ATCGCCTT	92	AAGGCGAT	140
UDI_Plate_WT_31	G4	TTGGAGGT	45	TTCGACGC	93	GCGTCGAA	141
UDI_Plate_WT_32	H4	TCGAGCGT	46	TGTCGTTC	94	GAACGACA	142
UDI_Plate_WT_33	A5	TGCGATCT	47	TCCATAGC	95	GCTATGGA	143
UDI_Plate_WT_34	B5	TTCCTGCT	48	TAAGTGTC	96	GACACTTA	144
UDI_Plate_WT_35	C5	TTCCATTG	49	CTGGCATA	97	TATGCCAG	145
UDI_Plate_WT_36	D5	TAACGCTG	50	CTGAGCCA	98	TGGCTCAG	146
UDI_Plate_WT_37	E5	TTGGTATG	51	CTCAATGA	99	TCATTGAG	147
UDI_Plate_WT_38	F5	TGAACTGG	52	CGCATACA	100	TGTATGCG	148
UDI_Plate_WT_39	G5	TCCAGTCG	53	CCGAAGTA	101	TACTTCGG	149
UDI_Plate_WT_40	H5	TGTATGCG	54	CCAGTTCA	102	TGAACTGG	150
UDI_Plate_WT_41	A6	TGGCTCAG	55	CAGCGTTA	103	TAACGCTG	151
UDI_Plate_WT_42	B6	TATGCCAG	56	CAATGGAA	104	TTCCATTG	152
UDI_Plate_WT_43	C6	GGTTGGAC	57	ATCCTGTA	105	TACAGGAT	153
UDI_Plate_WT_44	D6	GACACTTA	58	AGCAGGAA	106	TTCCTGCT	154
UDI_Plate_WT_45	E6	GAACGACA	59	ACGCTCGA	107	TCGAGCGT	155
UDI_Plate_WT_46	F6	AAGGCGAT	60	ACAGCAGA	108	TCTGCTGT	156

Table 16. UDI WT Plate sequences							
Index name	**well_position**	**i7_index**	**SEQ ID NO.:**	**i5_index_reverse_complement**	**SEQ ID NO.:**	**i5_index**	**SEQ ID NO.:**
UDI_Plate_WT_47	G6	ATGCTTGA	61	ACAAGCTA	109	TAGCTTGT	157
UDI_Plate_WT_48	H6	AGTATCTG	62	CATCAAGT	110	ACTTGATG	158

[0283] Place the tube(s) into a thermocycler and run the program shown in Table 17.

Table 17. Indexing PCR			
Run time	Lid Temperature	Sample Volume	
~30 min	105 °C	50 uL	
Step	Time	Temperature	Cycles
1	3 min	95 °C	1
2	20 sec	98 °C	Variable (see Table 15)
3	20 sec	67 °C	Variable (see Table 15)
4	1 min	72 °C	Variable (see Table 15)
5	5 min	72 °C	1
6	Hold	4 °C	1

[0284] Safe stopping point: Sequencing libraries can be stored at 4°C for up to 18 hours.

Post-Round 4 Barcoding Size Selection

[0285] The sequencing libraries are size selected with a double sided SPRI cleanup. To size select the sequencing libraries: Gather freshly prepared 85% ethanol. Gather room temperature SPRI beads (~50 µL per sublibrary). Vortex the SPRI beads until fully mixed. Add 30 µL of SPRI beads to each sequencing library tube. Vortex the tube(s) for 5 seconds. Briefly centrifuge. Incubate for 5 minutes at room temperature. Place the tube(s) on the high position of the magnetic rack for 0.2 mL tubes. Incubate until the solution clears (~2 minutes). While still on the magnetic rack, transfer 75 µL of the supernatant containing the DNA into new 0.2 mL tube(s). Discard the tube(s) with bead pellet(s). Add 10 µL of SPRI beads to each tube. Vortex the tube(s) for 5 seconds. Briefly centrifuge. Incubate for 5 minutes at room temperature. Place the tube(s) on the high position of the magnetic rack for 0.2 mL tubes.

Incubate until the solution clears (~3 minutes). While still on the magnetic rack, remove and discard the supernatant. While still on the magnetic rack, add 180 µL of 85% ethanol to each tube. Incubate for 1 minute at room temperature. While still on the magnetic rack, remove and discard the supernatant. Repeat wash steps once for a total of 2 washes. Remove any residual ethanol with a P20. While still on the magnetic rack, air dry the SPRI beads (~30 seconds). Remove the tube(s) from the magnetic rack. Fully resuspend each bead pellet with 20 µL of nuclease-free water. Incubate for 5 minutes at room temperature. Place the tube(s) on the low position of the magnetic rack. Incubate until the solution clears (~2 minutes). While still on the magnetic rack, transfer all elutant into new 0.2 mL tube(s). Store on ice. Safe stopping point: Sequencing libraries can be stored at -20°C for up to 3 months.

Sequencing Library Quantification

[0286] The concentration and size distribution of the sequencing libraries are measured with fluorescent dyes and capillary electrophoresis. Libraries are then stored at 4°C for up to 48 hours or at -20°C for up to 3 months. To quantify the sequencing libraries: Measure the concentration of each purified sequencing library with, e.g., a Qubit dsDNA HS (High Sensitivity) Assay Kit according to the manufacturer's instructions. Assess the size distribution of each purified sequencing library with a High Sensitivity DNA Kit on the Agilent Bioanalyzer System or High Sensitivity D1000 ScreenTape and Reagents on the Agilent TapeStation System according to the manufacturer's instructions. Samples may need to be diluted to be within the manufacturer's recommended concentration range. Typically, between a 1:3 to 1:10 dilution is appropriate. Safe stopping point: Sequencing libraries can be stored at -20°C for up to 3 months.

Target Sequencing Library Preparation – Target Amplification 1 (e.g., CRISPR PCR)

[0287] A PCR reaction amplifies a subset of the cDNA enriched for target transcripts, e.g., sgRNA transcripts. To amplify the target (e.g., sgRNA transcripts): Prepare 1.2 mL 85% ethanol per sublibrary with nuclease-free water. Equilibrate approximately 90 µL of SPRI beads per sublibrary to room temperature. Fill an ice bucket. Take out a magnetic rack for 0.2 mL PCR tubes. Obtain amplified cDNA from preamplification step. Vortex the tube(s) of amplified cDNA for 2-3 seconds. Briefly centrifuge. Prepare diluted cDNA in a new 0.2 mL tube(s) and store on ice. Bring the total volume up to 21 µL using nuclease-free water. Store any remaining cDNA at -20°C. Prepare a suitable target PCR Reaction Solution. Mix by pipetting 10x and store on ice. PCR reaction solution to each sublibrary (e.g., 29 µL of the

CRISPR PCR Reaction Solution). Mix sublibraries 10x with a P200 pipette set to 40 µL. Briefly centrifuge. Place the tube(s) into a thermocycler and run the program shown in Table 18.

Table 18. Target cDNA amplification round 1 (round 1 post-preamplification)			
Run time	Lid Temperature	Sample Volume	
~50 min	105 °C	50 uL	
Step	Time	Temperature	Cycles
1	3 min	95 °C	1
2	20 sec	98 °C	5
3	20 sec	65 °C	5
4	1 min	72 °C	5
5	20 sec	98 °C	13
6	20 sec	72 °C	13
7	1 min	72 °C	13
8	5 min	72 °C	1
9	Hold	4 °C	1

[0288] Safe stopping point: Amplified cDNA can be stored at 4°C for up to 18 hours.

Post target PCR (Target Amplification 1) Size Selection

[0289] Target (e.g., gRNA, or sgRNA) enriched cDNA sublibraries are size selected with a double sided SPRI cleanup. To size select the sublibraries: Gather freshly prepared 85% ethanol. Gather room temperature SPRI beads (~45 µL per sublibrary). Vortex the SPRI beads until fully mixed. Add 30 µL of SPRI beads to each tube of target (e.g., CRISPR/gRNA) enriched cDNA. Vortex the tube(s) for 2-3 seconds. Briefly centrifuge. Incubate for 5 minutes at room temperature. Place the tube(s) on the high position of the magnetic rack for 0.2 mL tubes. Incubate until the solution clears (~2 minutes). While still on the magnetic rack, transfer 75 µL of the supernatant containing the target enriched cDNA into new 0.2 mL tube(s). Discard the tube(s) with bead pellet(s). Add 10 µL of SPRI beads to each tube. Vortex the tube(s) for 2-3 seconds. Briefly centrifuge. Incubate for 5 minutes at room temperature. Place the tube(s) on the high position of the magnetic rack for 0.2 mL tubes. Incubate until the solution clears (~2 minutes). While still on the magnetic rack, remove and discard the supernatant. While still on the magnetic rack, add 180 µL of 85% ethanol to each tube. Incubate for 1 minute at room temperature.

[0290] While still on the magnetic rack, remove and discard the supernatant. Repeat wash steps once for a total of 2 washes. Remove any residual ethanol with a P20. While still on the magnetic rack, air dry the SPRI beads (~30 seconds). Remove the tube(s) from the magnetic rack. Fully resuspend each bead pellet with 20 µL of nuclease-free water. Incubate for 5 minutes at room temperature. Place the tube(s) on the low position of the magnetic rack. Incubate until the solution clears (~2 minutes). While still on the magnetic rack, transfer 20 µL of the supernatant into new 0.2 mL tube(s). Discard the tube(s) with bead pellet(s). Measure the concentration of each tube of size-selected target (e.g., CRISPR/gRNA) enriched cDNA with the Qubit dsDNA HS (High Sensitivity) Assay Kit according to the manufacturer's instructions. Safe stopping point: The size-selected target enriched cDNA can be stored at 4°C for up to 2 days or -20°C for up to 3 months.

Target Amplification 2 (e.g., CRISPR Index PCR)

[0291] A final PCR (i.e., second post-preamplification round of PCR) adds i5/i7 UDIs that act as a fourth cell barcode. To prepare the final target (e.g., CRISPR/gRNA) sequencing libraries: Obtain amplified and purified target cDNA. Vortex the tube(s) for 2-3 seconds. Briefly centrifuge. Prepare diluted target cDNA in new 0.2 mL tube(s) as follows and store on ice. Bring the total volume up to 21 µL using nuclease-free water. Store any remaining product at -20°C. Centrifuge the UDI Plate - EC at 100 x g for 1 minute. Wipe the surface of the plate with 70% ethanol and allow to dry. Orient the UDI Plate - EC with the notch on the bottom left. For each sublibrary being processed, choose one unused well of the UDI Plate - EC and record the well position and number for each sublibrary. With a multichannel P20, pierce the seal of the chosen wells of the UDI Plate - EC. With a multichannel P20 and new tips, add 4 µL from a different well of the UDI Plate - EC to each tube of diluted CRISPR cDNA. If any unused wells remain in the UDI Plate - EC, store the plate at -20°C. Add 25 µL of CRISPR Amplification Master Mix to each tube. Mix by pipetting 10x with a P200 multichannel pipette set to 40 µL. Briefly centrifuge. Place the tube(s) into a thermocycler and run the program shown in Table 19. The UDI sequences in the UDI Plate – EC are shown in Table 20.

Table 19. Sublibrary Index Amplification			
Run time	Lid Temperature	Sample Volume	
~30 min	105 °C	50 uL	
Step	Time	Temperature	Cycles
1	3 min	95 °C	1

Table 19. Sublibrary Index Amplification			
2	20 sec	98 °C	9
3	20 sec	67 °C	9
4	1 min	72 °C	9
5	5 min	72 °C	1
6	Hold	4 °C	1

[0292] Safe stopping point: target (e.g., CRISPR) sequencing libraries can be stored at 4°C for up to 18 hours.

Table 20. UDI EC Plate sequences

Index name	**well_ position**	**i7_index**	**SEQ ID No.:**	**i5_index_reverse_ complement**	**SEQ ID No.:**	**i5_index**	**SEQ ID No.:**
UDI_Plate_EC_1	A1	AACAGATC	159	TGCCATGA	207	TCATGGCA	255
UDI_Plate_EC_2	B1	CCTGTCTT	160	CTAACGAT	208	ATCGTTAG	256
UDI_Plate_EC_3	C1	ATATCGAG	161	TTCCGGAA	209	TTCCGGAA	257
UDI_Plate_EC_4	D1	TAGTCCGC	162	TCTCCATA	210	TATGGAGA	258
UDI_Plate_EC_5	E1	TGCTGTTA	163	AGTGACGT	211	ACGTCACT	259
UDI_Plate_EC_6	F1	TAGCGAAT	164	CTCTGGCA	212	TGCCAGAG	260
UDI_Plate_EC_7	G1	AGGACCGT	165	AGAAGATG	213	CATCTTCT	261
UDI_Plate_EC_8	H1	TTATCAGG	166	TCTAACAG	214	CTGTTAGA	262
UDI_Plate_EC_9	A2	CCTCGGAA	167	TACGTAGA	215	TCTACGTA	263
UDI_Plate_EC_10	B2	TTGATCGA	168	TAAGCTCT	216	AGAGCTTA	264
UDI_Plate_EC_11	C2	TCAACCTC	169	TTGGTCGA	217	TCGACCAA	265
UDI_Plate_EC_12	D2	CGGAATAA	170	CGGTAATA	218	TATTACCG	266
UDI_Plate_EC_13	E2	TATGAGAC	171	CACAGTGG	219	CCACTGTG	267
UDI_Plate_EC_14	F2	CCTACCAT	172	AGACGCGA	220	TCGCGTCT	268
UDI_Plate_EC_15	G2	CATAGGCC	173	AGCAATGG	221	CCATTGCT	269

Table 20. UDI EC Plate sequences

Index name	well_ position	i7_index	SEQ ID No.:	i5_index_reverse_ complement	SEQ ID No.:	i5_index	SEQ ID No.:
UDI_Plate_EC_16	H2	CACGATAA	174	CCTCTTGA	222	TCAAGAGG	270
UDI_Plate_EC_17	A3	TTCTGGCG	175	TTGTCTAG	223	CTAGACAA	271
UDI_Plate_EC_18	B3	TCGTAACT	176	TGTTGCTA	224	TAGCAACA	272
UDI_Plate_EC_19	C3	TCCGTGGT	177	TACTGTCA	225	TGACAGTA	273
UDI_Plate_EC_20	D3	AAGACAAC	178	AAGCCATA	226	TATGGCTT	274
UDI_Plate_EC_21	E3	ATTGCTTC	179	CGGCACTT	227	AAGTGCCG	275
UDI_Plate_EC_22	F3	AGAGAAGT	180	TCATGGAT	228	ATCCATGA	276
UDI_Plate_EC_23	G3	TACGTCTT	181	TAATCAGG	229	CCTGATTA	277
UDI_Plate_EC_24	H3	TCCACGTT	182	CATCTGAG	230	CTCAGATG	278
UDI_Plate_EC_25	A4	CAATGAGT	183	CAGAATCG	231	CGATTCTG	279
UDI_Plate_EC_26	B4	CGTCAGTT	184	TTGATGCA	232	TGCATCAA	280
UDI_Plate_EC_27	C4	TTCACTCC	185	ACGTATGG	233	CCATACGT	281
UDI_Plate_EC_28	D4	CTACCTGA	186	TGCGGATA	234	TATCCGCA	282
UDI_Plate_EC_29	E4	AGGTGATT	187	TAACAGCA	235	TGCTGTTA	283
UDI_Plate_EC_30	F4	TCTCACAT	188	AATGTTCG	236	CGAACATT	284
UDI_Plate_EC_31	G4	TTAGTGAG	189	TTACGTGT	237	ACACGTAA	285
UDI_Plate_EC_32	H4	TGTTCACT	190	ATCTACTG	238	CAGTAGAT	286
UDI_Plate_EC_33	A5	CGAAGCCT	191	ATACCTCT	239	AGAGGTAT	287
UDI_Plate_EC_34	B5	TGGCCGTA	192	TACATCTG	240	CAGATGTA	288
UDI_Plate_EC_35	C5	CGCTTCAC	193	AAGGAGCA	241	TGCTCCTT	289
UDI_Plate_EC_36	D5	CGGAGAAC	194	CTCCTAGA	242	TCTAGGAG	290
UDI_Plate_EC_37	E5	AAGTTCAG	195	TCAGCCTA	243	TAGGCTGA	291
UDI_Plate_EC_38	F5	AAGCCTTC	196	TTCGCTCA	244	TGAGCGAA	292

Table 20. UDI EC Plate sequences							
Index name	**well_ position**	**i7_index**	**SEQ ID No.:**	**i5_index_reverse_ complement**	**SEQ ID No.:**	**i5_index**	**SEQ ID No.:**
UDI_Plate_EC_39	G5	ACACCTCA	197	CGGATTAA	245	TTAATCCG	293
UDI_Plate_EC_40	H5	TTCTCCTA	198	ATTAGAGG	246	CCTCTAAT	294
UDI_Plate_EC_41	A6	TGACACGC	199	CGTGTGAA	247	TTCACACG	295
UDI_Plate_EC_42	B6	AAGCGCCT	200	TGTCACGG	248	CCGTGACA	296
UDI_Plate_EC_43	C6	TGATAACC	201	CGTAATCT	249	AGATTACG	297
UDI_Plate_EC_44	D6	AAGAGTGT	202	CCTTCTGG	250	CCAGAAGG	298
UDI_Plate_EC_45	E6	TGCTACCG	203	AATCGCTA	251	TAGCGATT	299
UDI_Plate_EC_46	F6	CCATCGTC	204	TGGACCAA	252	TTGGTCCA	300
UDI_Plate_EC_47	G6	CCACGAGA	205	TGTGCACT	253	AGTGCACA	301
UDI_Plate_EC_48	H6	ACGGTCAT	206	CTTATGGA	254	TCCATAAG	302

Post-Target Amplification 2 (e.g., CRISPR Index PCR) Size Selection

[0293] Target (e.g., CRISPR) sequencing libraries are size selected with a double sided SPRI cleanup. To size select the sublibraries: Gather freshly prepared 85% ethanol. Gather room temperature SPRI beads (~45 µL per sublibrary). Vortex the SPRI beads until fully mixed. Add 30 µL of SPRI beads to each tube of target sequencing library. Vortex the tube(s) for 2-3 seconds. Briefly centrifuge. Incubate for 5 minutes at room temperature. Place the tube(s) on the high position of the magnetic rack for 0.2 mL tubes. Incubate until the solution clears (~2 minutes). While still on the magnetic rack, transfer 75 µL of the supernatant containing the target sequencing library into new 0.2 mL tube(s). Discard the tube(s) with bead pellet(s). Add 10 µL of SPRI beads to each tube. Vortex the tube(s) for 2-3 seconds. Briefly centrifuge. Incubate for 5 minutes at room temperature. Place the tube(s) on the high position of the magnetic rack for 0.2 mL tubes. Incubate until the solution clears (~2 minutes). While still on the magnetic rack, remove and discard the supernatant. While still on the magnetic rack, add 180 µL of 85% ethanol to each tube. Incubate for 1 minute at room temperature.

[0294] While still on the magnetic rack, remove and discard the supernatant. Repeat steps 13-15 once for a total of 2 washes. Remove any residual ethanol with a P20. While still on the magnetic rack, air dry the SPRI beads (~30 seconds). Remove the tube(s) from the magnetic rack. Fully resuspend each bead pellet with 20 µL of nuclease-free water. Incubate for 5 minutes at room temperature. Place the tube(s) on the low position of the magnetic rack. Incubate until the solution clears (~2 minutes). While still on the magnetic rack, transfer 20 µL of the supernatant into new 0.2 mL tube(s). Discard the tube(s) with bead pellet(s). Product is ready for quantification.

Target (e.g., CRISPR) Sequencing Library Quantification

[0295] The concentration and size distribution of target sequencing libraries are measured with fluorescent dyes and capillary electrophoresis. Libraries are then stored at -20°C for up to 3 months. To quantify the target sequencing libraries: Measure the concentration of each purified target sequencing library with, e.g., the Qubit dsDNA HS (High Sensitivity) Assay Kit according to the manufacturer's instructions. Assess the size distribution of each purified sequencing library with a High Sensitivity DNA Kit on the Agilent Bioanalyzer System or High Sensitivity D5000 ScreenTape and Reagents on the Agilent TapeStation System according to the manufacturer's instructions. Samples may need to be diluted to be within the manufacturer's recommended concentration range. See below for an expected Tapestation trace when amplifying target sequencing libraries. There should be an expected peak around 530 bp. Typically, between a 1:3 to 1:10 dilution is appropriate. Safe stopping point: Sequencing libraries can be stored at -20°C for up to 3 months.

Sequencing

[0296] Sequencing libraries should be diluted and denatured according to the instruction for the relevant sequencing instrument. Target sequencing libraries can be sequenced together or separately from Whole Transcriptome libraries. We recommend sequencing target (e.g., CRISPR) libraries at 2,000-5,000 reads per cell, so the ratio between the libraries may need to be adjusted if they are sequenced together. If sequenced alone or with target libraries, we recommend using a 5% PhiX spike for Whole Transcriptome libraries. For target libraries sequenced alone, we suggest using a minimum of 10% PhiX due to their limited complexity. Recommended cycles: Read 1: 66; i7 Index (Index 1): 8; i5 Index (Index 2): 8; Read 2: 86.

EQUIVALENTS AND INCORPORATION BY REFERENCE

[0297] All references cited herein are incorporated by reference to the same extent as if each individual publication, database entry (e.g., Genbank sequences or GeneID entries), patent application, or patent, was specifically and individually indicated incorporated by reference in its entirety, for all purposes. This statement of incorporation by reference is intended by Applicants, pursuant to 37 C.F.R. §1.57(b)(1), to relate to each and every individual publication, database entry (e.g., Genbank sequences or GeneID entries), patent application, or patent, each of which is clearly identified in compliance with 37 C.F.R. §1.57(b)(2), even if such citation is not immediately adjacent to a dedicated statement of incorporation by reference. The inclusion of dedicated statements of incorporation by reference, if any, within the specification does not in any way weaken this general statement of incorporation by reference. Citation of the references herein is not intended as an admission that the reference is pertinent prior art, nor does it constitute any admission as to the contents or date of these publications or documents.

[0298] While the invention has been particularly shown and described with reference to a preferred embodiment and various alternate embodiments, it is understood by persons skilled in the relevant art that various changes in form and details can be made therein without departing from the spirit and scope of the invention.

CLAIMS

1. A method of labeling nucleic acids for multiplex transcriptional analysis in a plurality of cells or nuclei, the method comprising:

(a) providing a plurality of fixed and permeabilized cells or nuclei, each comprising a plurality of RNA molecules, and each comprising one or more target genes or transcripts of interest;

(b) dividing the plurality of cells or nuclei into a first plurality of aliquots, wherein each aliquot comprises more than one cell or nucleus;

(c) generating complementary DNA (cDNA) molecules by reverse transcribing RNA molecules within the cells or nuclei of the first plurality of aliquots, wherein the RNA molecules are reverse transcribed using reverse transcription (RT) primers each comprising: (i) a poly(T) sequence or a random sequence; and (ii) an RT barcode sequence, wherein the RT barcode sequences present within the RT primers are specific to each aliquot within the first plurality of aliquots;

(d) pooling the cells or nuclei from the first plurality of aliquots;

(e) tagging the cDNA molecules within the pooled cells or nuclei from the first plurality of aliquots with one or more nucleic acid tags, thereby generating tagged cDNA molecules, by performing steps (i) through (iii) one or more times:

(i) dividing the pooled cells or nuclei into an additional plurality of aliquots;

(ii) coupling nucleic acid tags to cDNA molecules within the cells or nuclei of the additional plurality of aliquots, wherein each nucleic acid tag comprises a tag barcode sequence, and wherein the tag barcode sequences present within the nucleic acid tags are specific to each aliquot within the additional plurality of aliquots;

(iii) pooling the cells from the additional plurality of aliquots;

(f) lysing the cells or nuclei to release the tagged cDNA molecules and produce a lysate comprising the released tagged cDNA molecules;

(g) isolating the released tagged cDNA molecules using a binding agent, such that the isolated tagged cDNA molecules are bound to the binding agent;

(h) generating second strands of the isolated tagged cDNA molecules to produce double-stranded tagged cDNA molecules;

(i) amplifying the isolated tagged cDNA molecules in a first multiplex amplification step using a multiplex set of preamplification primers, wherein the multiplex set of preamplification primers comprises:

at least one pair of whole transcriptome (WT) preamplification primers configured to amplify all of the tagged cDNA molecules isolated from the lysate, and

at least one pair of target-specific preamplification primers configured to specifically amplify tagged target cDNA molecules isolated from the lysate,

thereby generating an enriched plurality of amplified tagged cDNA molecules that comprises the whole transcriptome and that is enriched for the one or more target cDNA molecules;

(j) dividing the enriched plurality of amplified cDNA molecules into at least a first and a second portion;

(k) preparing a whole transcriptome (WT) sequencing library using the first portion of the enriched plurality of amplified tagged cDNA molecules; and

(l) preparing a target sequencing library using the second portion of the enriched plurality of amplified tagged cDNA molecules.

2. The method of claim 1,

wherein the preparation of the WT sequencing library comprises fragmenting the first portion of the enriched plurality of amplified tagged cDNA molecules and appending a first adapter comprising a first adapter sequence to the fragment ends, and

amplifying the tagged cDNA molecules comprising the first adapter in an index PCR using WT index amplification primers,

wherein one or more of the WT index amplification primers comprises a sequence complementary to the first adapter sequence,

wherein one or more of the WT index amplification primers comprises an index sequence, and

wherein one or more of the WT index amplification primers comprises a next-generation sequencing (NGS) adapter sequence, an NGS primer binding sequence, and/or an NGS flow-cell binding sequence.

3. The method of claim 1 or claim 2, wherein the preparation of the target sequencing library comprises:

amplifying the tagged target cDNA molecules in the second portion of the enriched plurality of amplified tagged cDNA molecules in a second target-specific amplification step, thereby generating a further enriched plurality of tagged target cDNA molecules;

wherein the tagged target cDNA molecules are amplified in the second target-specific amplification step using at least one pair of target amplification primers configured to specifically amplify the tagged target cDNA molecules in the second portion; and

amplifying the further enriched plurality of tagged target cDNA molecules in an index target PCR using target index amplification primers;

wherein one or more of the target index amplification primers comprises an index sequence, and

wherein one or more of the target index amplification primers comprises a next-generation sequencing (NGS) adapter sequence, an NGS primer binding sequence, and/or an NGS flow-cell binding sequence.

4. The method of any one of claims 1-3, wherein the nucleic acid tags that are coupled to the tagged cDNA molecules during the last of the one or more times that steps (e)(i) to (e)(iii) are repeated comprise one or more elements selected from the group consisting of a random nucleotide sequence to prevent counting of PCR duplicates, a capture agent, and a second adapter sequence.

5. The method of claim 4, wherein the capture agent comprises biotin.

6. The method of claim 5, wherein the binding agent comprises streptavidin-coated magnetic beads.

7. The method of any one of claims 1-6, wherein the lysing in step (f) is performed in the presence of a protease.

8. The method of claim 7, wherein the protease is proteinase K.

9. The method of claim 7 or claim 8, wherein a protease inhibitor is added to the lysate prior to or together with the binding agent.

10. The method of claim 9, wherein the protease inhibitor is phenylmethanesulfonyl fluoride (PMSF) or 4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride (AEBSF).

11. The method of any one of claims 1-10, wherein the second strands are generated in step (h) using a template switching oligo (TSO) comprising a third adapter sequence, such that the template switching introduces the third adapter sequence to the 3'-end of the released cDNA molecules.

12. The method of any one of claims 1-11, wherein prior to the lysing in step (f), the plurality of cells or nuclei comprising the tagged cDNA molecules are divided into a plurality of samples.

13. The method of claim 12, wherein the index sequence in one or more of the WT index amplification primers and/or one or more of the target index amplification primers is sample-specific.

14. The method of any one of claims 1-13, wherein at least one target-specific preamplification primer and at least one target amplification primer are complementary to the same sequence within a target cDNA molecule as a target amplification primer.

15. The method of any one of claims 1-14, wherein none of the target-specific preamplification primers specifically binds to the same sequence within a target cDNA molecule as a target amplification primer.

16. The method of any one of claims 1-15, wherein the at least one pair of target-specific preamplification primers and the at least one pair of WT preamplification primers have at least one preamplification primer in common.

17. The method of claim 16, wherein the at least one preamplification primer common to both the at least one pair of target-specific preamplification primers and the at least one pair of WT preamplification primers comprises a sequence complementary to the second adapter sequence.

18. The method of any one of claims 1-17, wherein the at least one pair of WT preamplification primers comprise a primer comprising a sequence complementary to the third adapter sequence and a primer comprising a sequence complementary to the second adapter sequence.

19. The method of any one of claims 1-18, wherein the multiplex set of preamplification primers comprises a single pair of target-specific preamplification primers and/or a single pair of WT preamplification primers.

20. The method of any one of claims 1-18, wherein the multiplex set of preamplification primers comprises from 1-10, 10-100, or 100-200 pairs of target-specific preamplification primers.

21. The method of any one of claims 1-20, wherein the RT primers and/or the nucleic acid tags are DNA molecules.

22. The method of any one of claims 1-21, further comprising sequencing the WT sequencing library and the target sequencing library.

23. The method of claim 22, wherein the WT sequencing library and the target sequencing library are sequenced separately.

24. The method of claim 22, wherein the WT sequencing library and the target sequencing library are sequenced together.

25. The method of any one of claims 22-24, further comprising filtering the sequencing reads generated from the sequencing of the WT sequencing library and the target sequencing library to remove transcripts represented by only one read.

26. The method of any one of claims 22-25, further comprising grouping sequencing reads generated from the sequencing of the WT sequencing library and the target sequencing library according to one or more features selected from the group consisting of RT barcode sequences, tag barcode sequences, series or combinations of tag barcode sequences, and index sequences.

27. The method of claim 26, wherein the grouped sequencing reads are used to determine the individual cell or nucleus from among the plurality of cells or nuclei from which a given cDNA originated.

28. The method of any one of claims 22-26, further comprising grouping sequence reads into target cDNA sequences and non-target cDNA sequences.

29. The method of claim 28, further comprising identifying one or more cells or nuclei from among the plurality of cells or nuclei in which a target cDNA is expressed, or in which the target cDNA is expressed at an increased or decreased level relative to a reference value.

30. The method of claim 28 or claim 29, further comprising relating the identity and/or expression level of a target gene or sequence in an individual cell or nucleus to the overall pattern of expression of the whole transcriptome in the same individual cell or nucleus.

31. The method of any one of claims 28-30, further comprising relating the identity and/or expression level of a target gene or sequence in an individual cell or nucleus to the expression level of one or more individual non-target genes in the same original cell or nucleus.

32. The method of any one of claims 1-31, wherein the one or more target cDNAs comprise one or more CRISPR guide RNA (gRNA) sequences.

33. The method of claim 32, wherein the gRNA sequences are from a gRNA library used in a CRISPR screen.

34. The method of claim 32 or 33, wherein at least one of the target-specific preamplification primers is specific to a Pol III promoter.

35. The method of claim 34, wherein the Pol III promoter is a U6 promoter.

36. The method of any one of claims 32-35, wherein at least one of the target-specific preamplification primers comprises the sequence of SEQ ID NO:11.

37. The method of any one of claims 1-36, wherein at least one pair of target-specific preamplification primers includes a primer comprising the sequence of SEQ ID NO:12 or SEQ ID NO:13.

38. The method of any one of claims 32-37, wherein at least one of the target amplification primers is specific to a Pol III promoter.

39. The method of claim 38, wherein the Pol III promoter is a U6 promoter.

40. The method of any one of claims 32-39, wherein at least one of the target amplification primers comprises the sequence of SEQ ID NO: 14.

41. The method of any one of claims 1-40, wherein at least one of the target amplification primers comprises the sequence of SEQ ID NO: 12.

42. The method of any one of claims 32-40, wherein the plurality of cells each comprises an expression construct encoding an RNA-guided nuclease, or inactivated form thereof, that is capable of physically interacting with the guide RNA and being directed to a target locus in the genome by the guide RNA.

43. The method of claim 42, wherein the RNA-guided nuclease is Cas9 or Cpf1.

44. The method of any one of claims 1-31, wherein the target genes or sequences comprise a T cell receptor (TCR) gene or sequence.

45. The method of claim 44, wherein the at least one pair of target preamplification and/or the at least one pair of target amplification primers comprise a primer specific to a TCR alpha, beta, gamma, or delta chain.

46. The method of claim 45, wherein the at least one pair of target-specific preamplification primers and/or the at least one pair of target amplification primers comprise a primer specific to a TCR alpha chain and a primer specific to a TCR beta chain.

47. The method of any one of claims 44-46, wherein at least one of the target-specific preamplification primers and/or at least one of the target amplification primers is specific to a TCR CDR3 region.

48. The method of any one of claims 1-47, wherein the cells or nuclei comprise mammalian cells or nuclei.

49. The method of claim 48, wherein the cells or nuclei comprise human cells or nuclei.

50. The method of claim 48 or claim 49, wherein the cells or nuclei comprise mouse cells or nuclei.

51. The method of any one of claims 48-50, wherein the cells or nuclei comprise T cells or nuclei derived therefrom.

52. The method of claim 51, wherein the T cells or nuclei comprise one or more cells or nuclei selected from the group consisting of chimeric antigen receptor (CAR) T cells, activated T cells, primary cells, T cells isolated from a cell line, T cells isolated from a tissue, T cells isolated from a subject, effector T cells, cytotoxic T cells, helper T cells, regulatory T cells, memory T cells, nuclei derived from any of the heretofore listed T cells, and combinations thereof.

53. The method of any one of claims 1-52, wherein the first multiplex amplification of step (i) comprises amplifying the tagged cDNA molecules for from 5 to 20 cycles.

54. The method of any one of claims 1-53, wherein the first multiplex amplification of step (i) is performed according to the conditions shown in Table 12.

55. The method of any one of claims 1-54, wherein one or more of steps (a), (b), (d), (e)(*i*), (e)(*iii*), or (f) are carried out at a temperature of below about 8, 7, 6, 5, 4, 3, 2, 1, 0, -1, -2, -3, or -4 °C, between about -4 to 8, -4 to 0, 0 to 4, 4 to 8, or 0 to 8 °C, or at about 8, 7, 6, 5, 4, 3, 2, 1, 0, -1, -2, -3, or -4 °C.

56. The method of any one of claims 1-55, wherein the cells or nuclei were fixed and/or permeabilized at 4 °C or below 4 °C.

57. The method of any one of claims 1-56, wherein the nucleic acid tags are coupled to the cDNA molecules in step (e)(ii) by ligation.

58. The method of any one of claims 1-57, wherein the RT primers each comprise a 5' overhang comprising a 5' overhang sequence.

59. The method of any one of claims 1-58, wherein the nucleic acid tags each comprise a first strand comprising a 3' hybridization sequence and/or a 5' hybridization sequence flanking the 3' end and/or the 5' end of the tag barcode sequence, respectively.

60. The method of claim 59, wherein the RT primers each comprise a 5' overhang comprising a 5' overhang sequence, and wherein the nucleic acid tags each further comprise a second strand comprising:

a first portion complementary to a 5' hybridization sequence of a previously coupled nucleic acid tag or a 5' overhang sequence of an RT primer; and

a second portion complementary to the 3' hybridization sequence.

61. The method of any one of claims 1-60, further comprising:

size selecting the enriched plurality of amplified tagged cDNA molecules subsequent to step (h) using solid phase reversible immobilization (SPRI) beads.

62. The method of any one of claims 3-61, further comprising:

size selecting the further enriched plurality of tagged target cDNA molecules using solid phase reversible immobilization (SPRI) beads.

63. The method of claim 61 or 62, wherein the size selection using SPRI beads is single-sided.

64. The method of claim 61 or 62, wherein the size selection using SPRI beads is double-sided.

65. The method of any one of claims 1-64, wherein one or more of the pluralities of aliquots or samples are distributed in a multi-well plate.

66. The method of claim 65, wherein the multi-well plate is a 96-well plate.

67. The method of claim 66, wherein the nucleic acid tags used in one or more of the additional pluralities of aliquots comprise 96 distinct barcode sequences.

68. The method of claim 67, wherein each of the 96 distinct barcode sequences is present in only one of the 96 aliquots.

69. The method of any one of claims 65-68, wherein at least a subset of the wells of the multi-well plate contains primers comprising well-specific unique dual indexes (UDIs).

70. The method of claim 69, wherein one or more of the UDIs comprise any of SEQ ID NOS: 13-302.

71. A kit for performing the method of any one or more of claims 1-70.

72. A kit for preparing multiplex sequencing libraries, the kit comprising at least one set of primers for cDNA labeling and amplification of the whole transcriptome, and at least one set of target-specific preamplification primers for cDNA labeling, amplification and enrichment of one or more target genes of interest.

73. The kit of claim 72, further comprising instructions for preparing the sequencing libraries.

74. The kit of claim 72 or 72, wherein the target-specific primers comprise SEQ ID NO: 11 or SEQ ID NO:14.

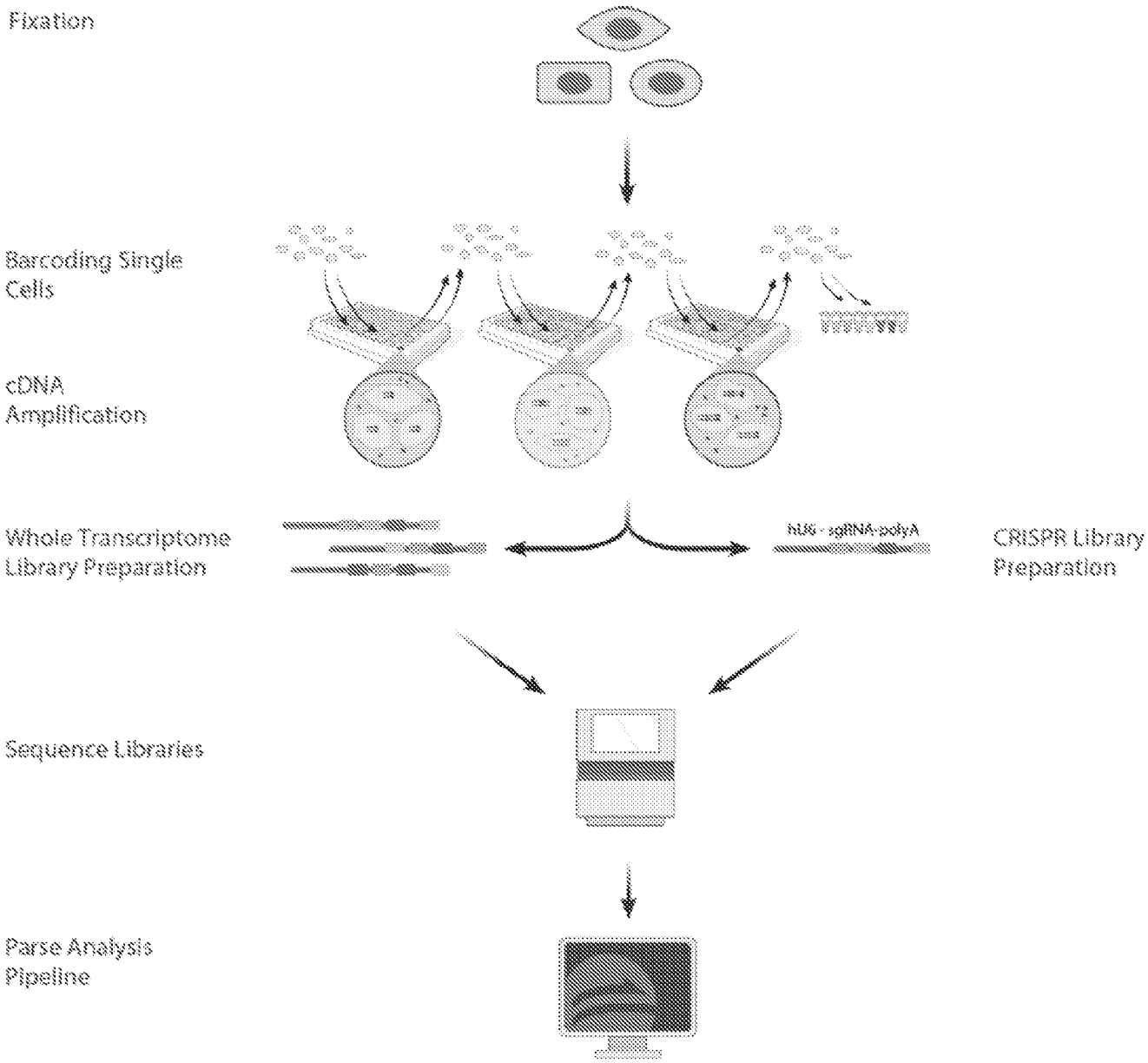

Fixation
Barcoding Single Cells
cDNA Amplification
Whole Transcriptome Library Preparation
hU6 - sgRNA-polyA
CRISPR Library Preparation
Sequence Libraries
Parse Analysis Pipeline

FIG. 1A

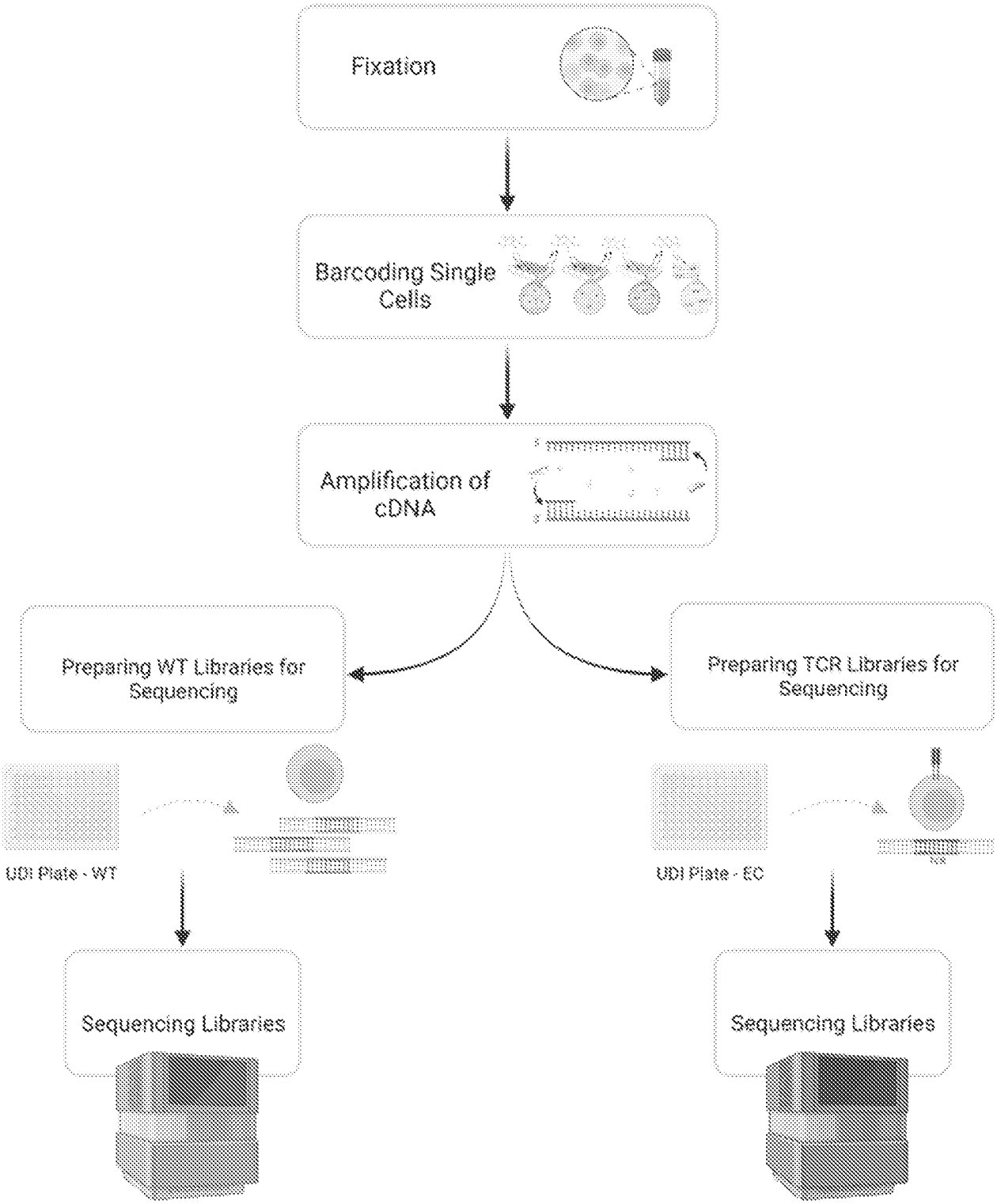

Fixation
Barcoding Single Cells
Amplification of cDNA
Preparing WT Libraries for Sequencing
Preparing TCR Libraries for Sequencing
UDI Plate - WT
UDI Plate - EC
Sequencing Libraries
Sequencing Libraries

FIG. 1B

FIG. 2A

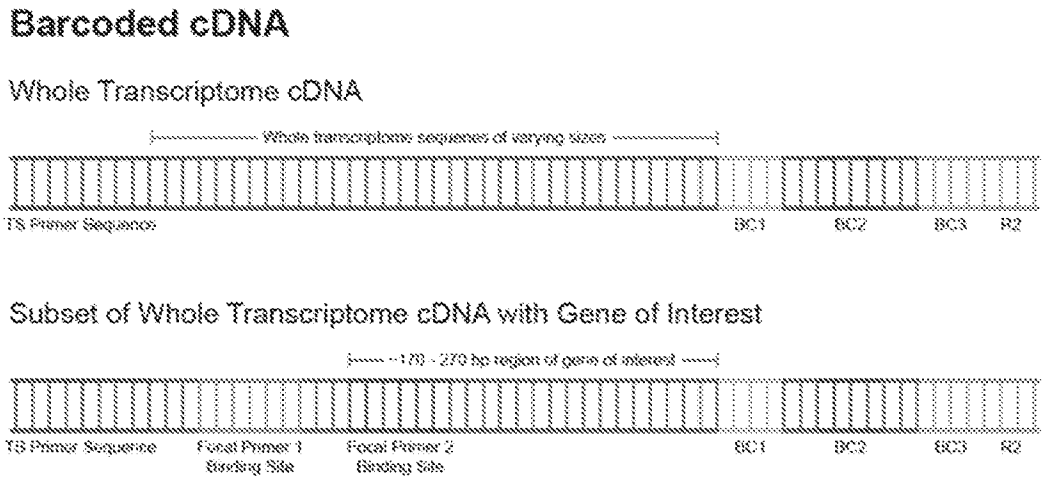


FIG. 2B

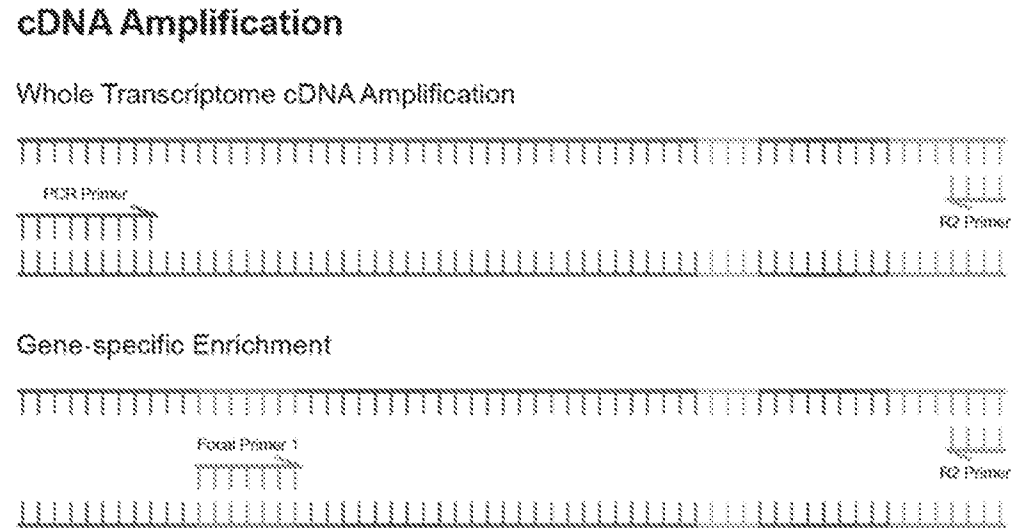


FIG. 2C

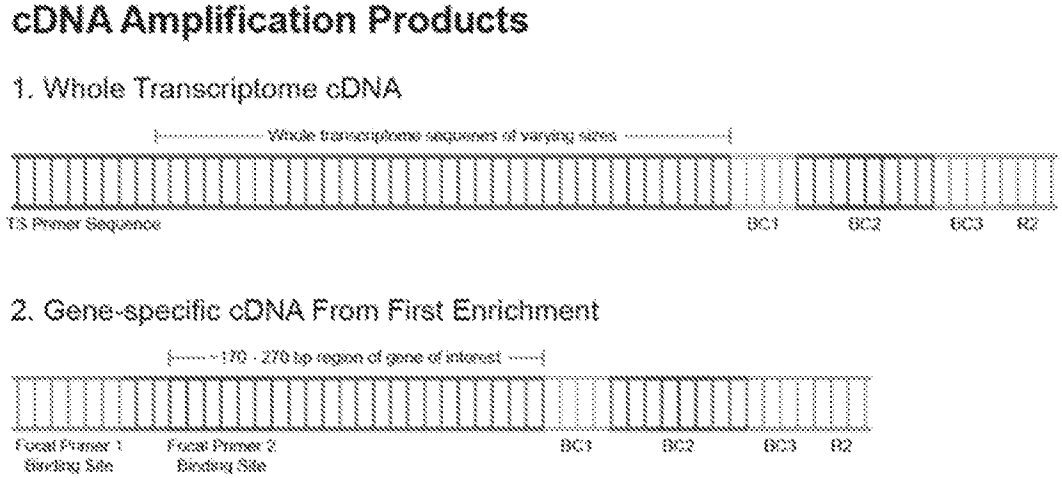


FIG. 3A

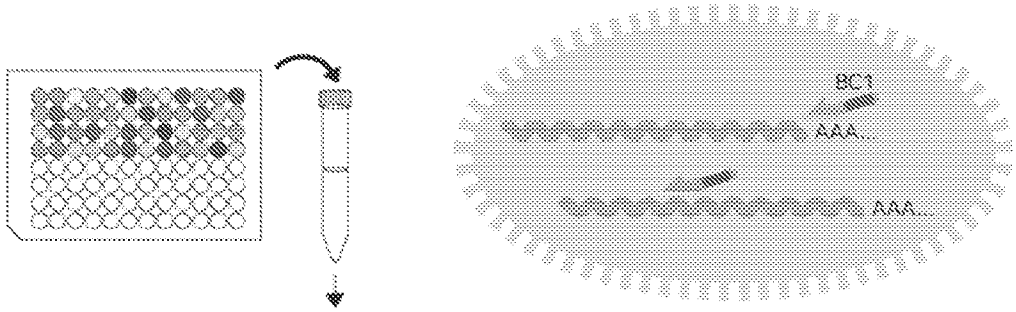


FIG. 3B

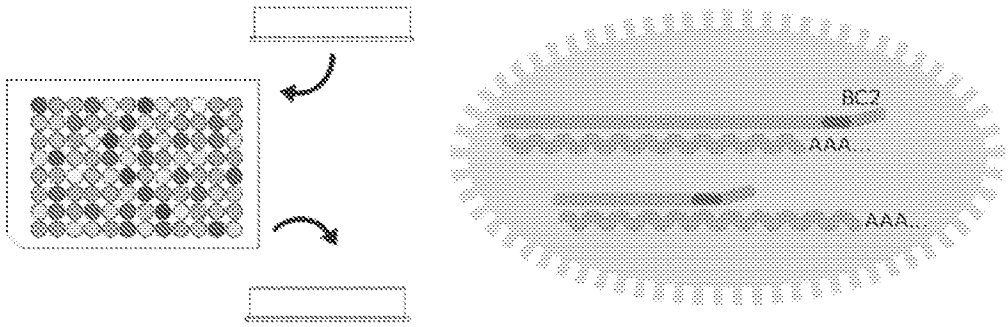


FIG. 3C

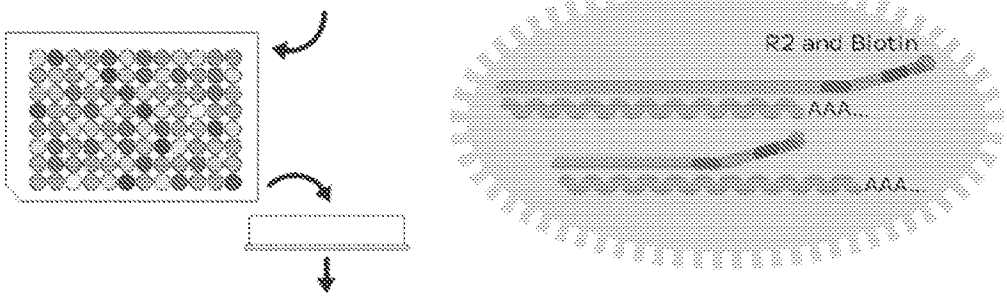


FIG. 3D

FIG. 4A

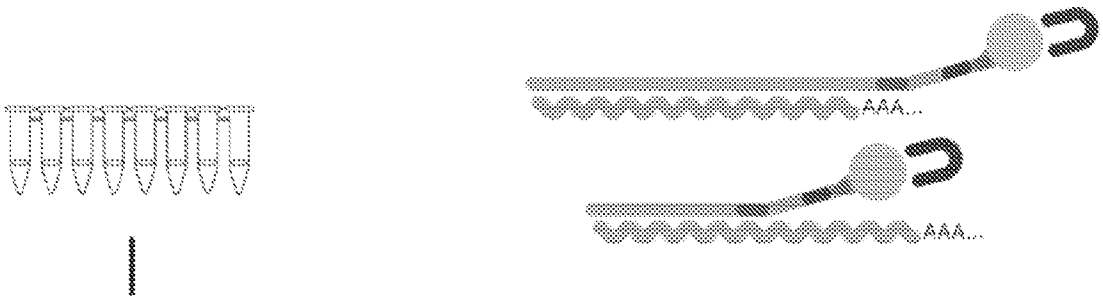


FIG. 4B

FIG. 4C

FIG. 4D

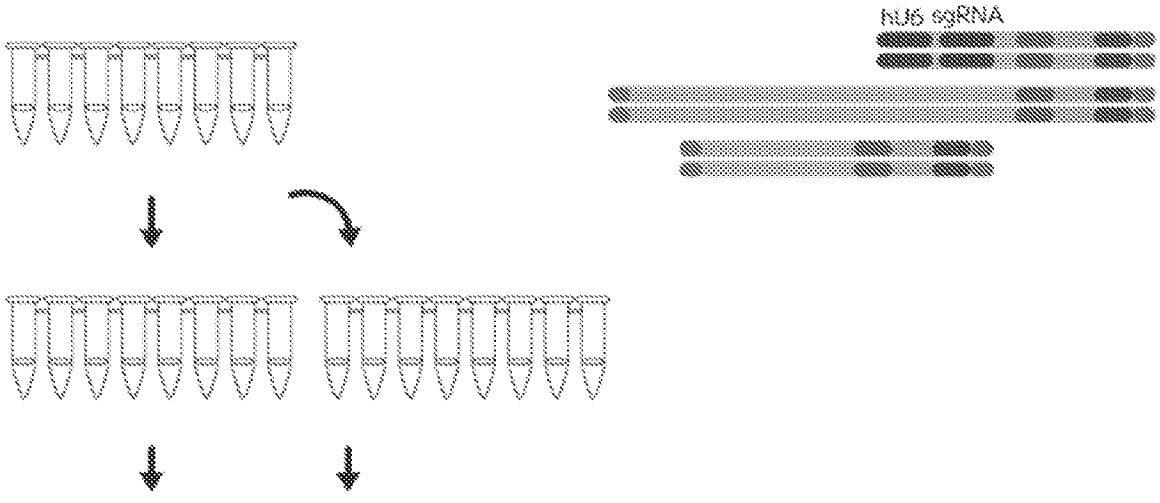
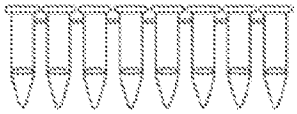


FIG. 5A



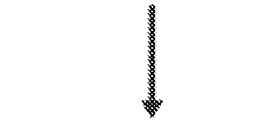
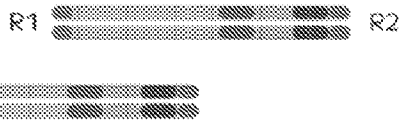


FIG. 5B



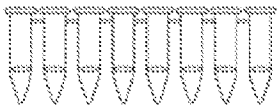


FIG. 5C

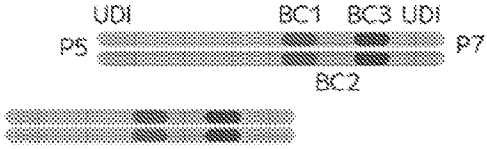


FIG. 6A

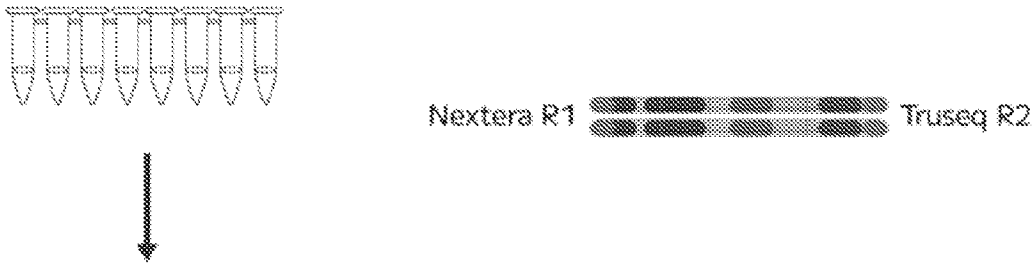
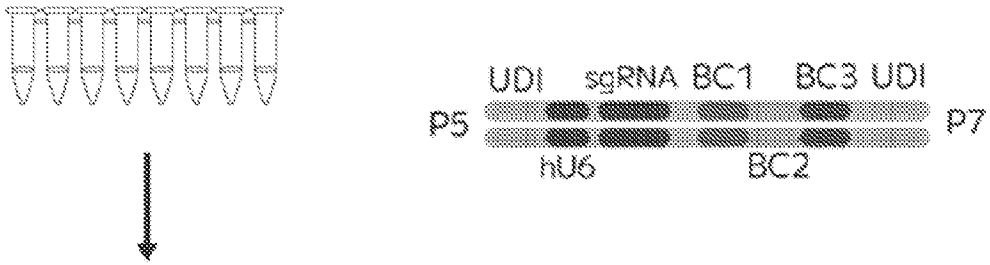


FIG. 6B



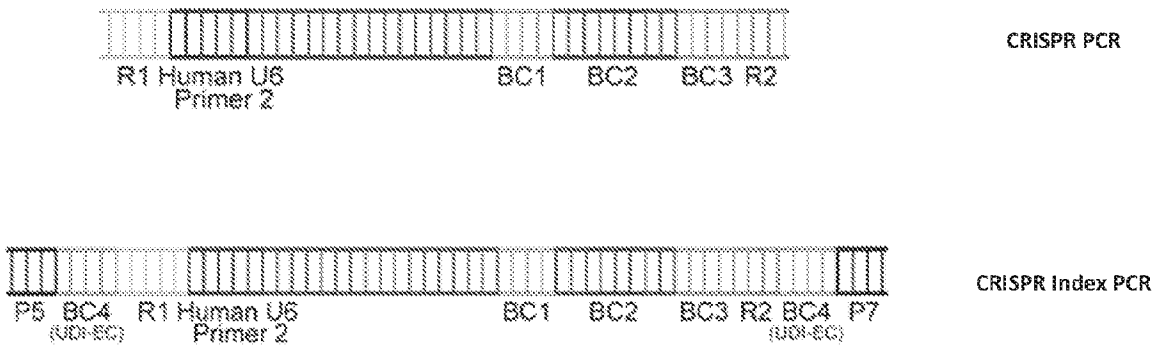


FIG. 6C

FIG. 7A

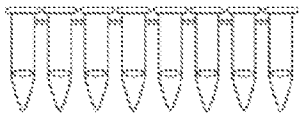
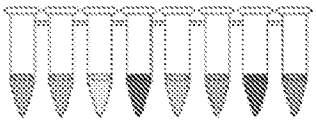
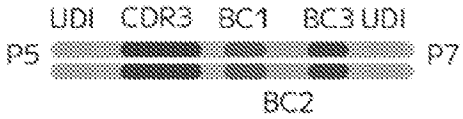


FIG. 7B





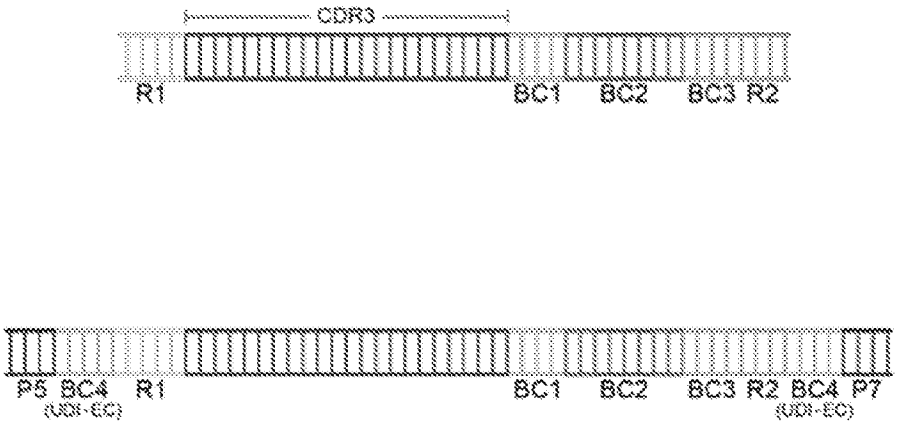


FIG. 7C

FIG. 8A

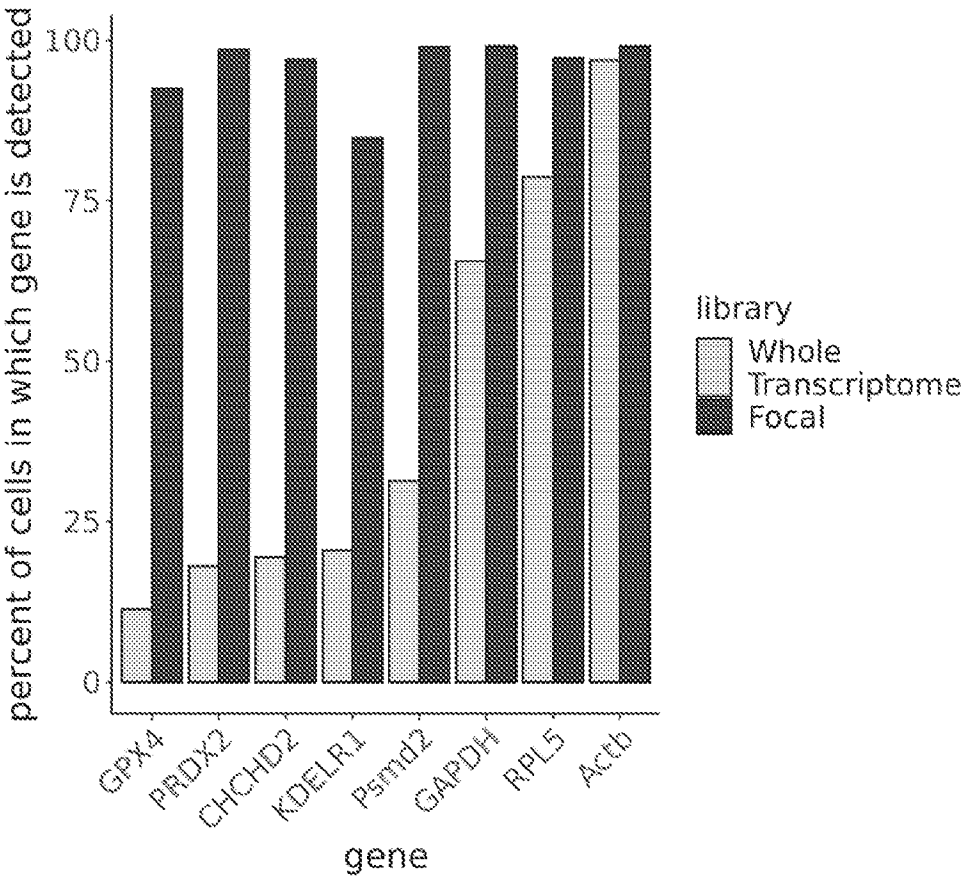

percent of cells in which gene is detected
100
75
50
25
0
GPX4
PRDX2
CHCHD2
KDELR1
Psmd2
GAPDH
RPL5
Actb
gene
library
Whole
Transcriptome
Focal

FIG. 8B

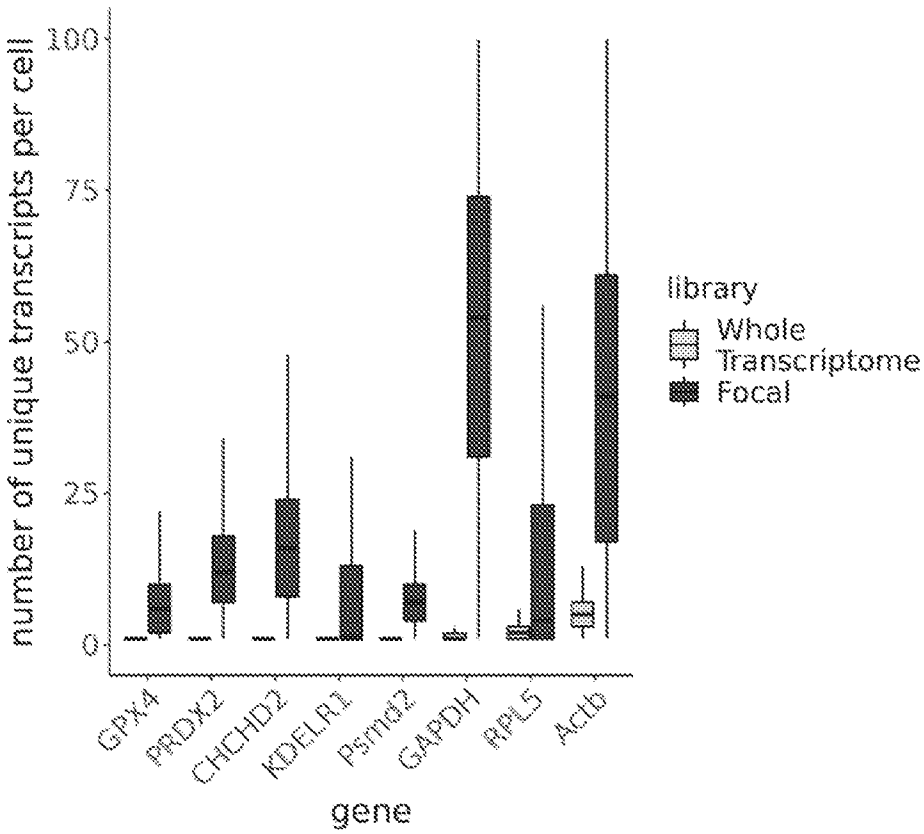
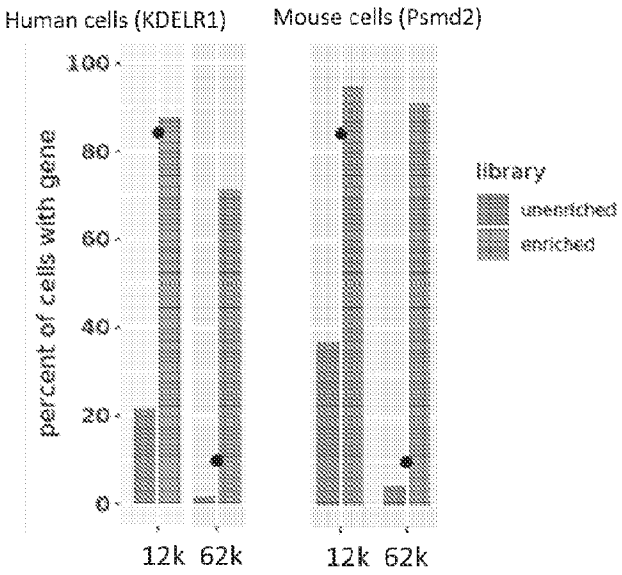

number of unique transcripts per cell
100
75
50
25
0
GPX4
PRDX2
CHCHD2
KDELR1
Psmd2
GAPDH
RPL5
Actb
gene
library
Whole
Transcriptome
Focal

FIG. 9A


Human cells (KDELR1)
Mouse cells (Psmd2)
percent of cells with gene
100
80
60
40
20
0
12k 62k
12k 62k
library
unenriched
enriched

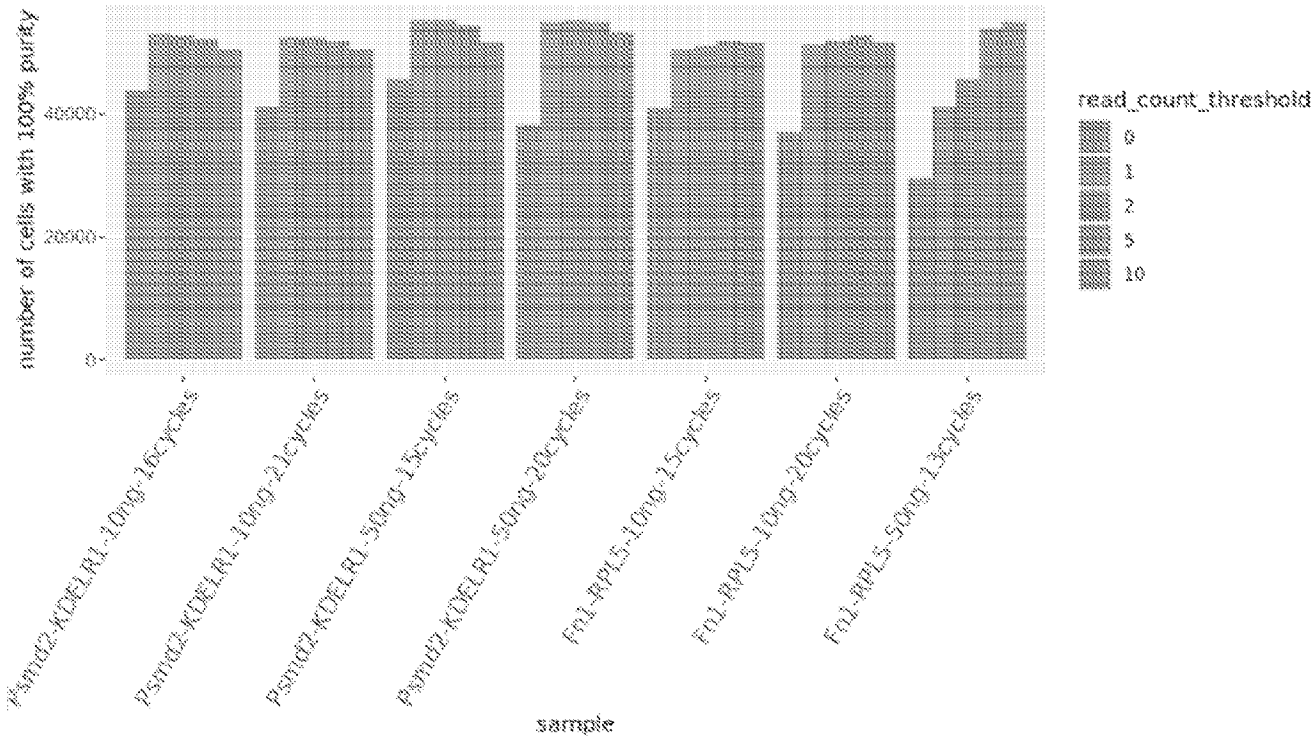

number of cells with 100% purity
40000
20000
0
Psmd2-KDELR1-10ng-16cycles
Psmd2-KDELR1-10ng-21cycles
Psmd2-KDELR1-5ng-15cycles
Psmd2-KDELR1-5ng-20cycles
Fn1-RPL5-10ng-15cycles
Fn1-RPL5-10ng-20cycles
Fn1-RPL5-5ng-13cycles
sample
read_count_threshold
0
1
2
5
10

FIG. 9B

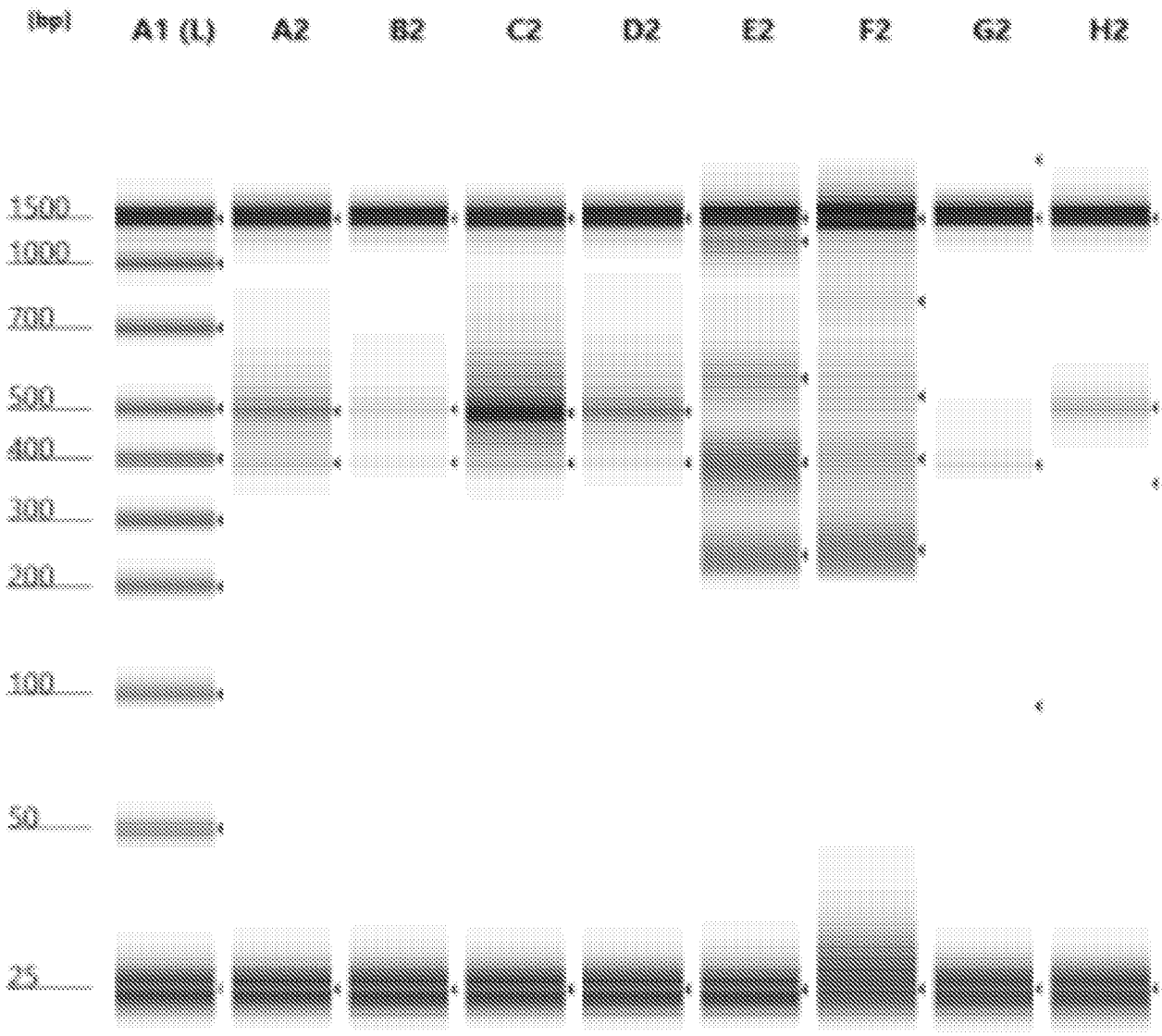

(bp)
A1 (L)
A2
B2
C2
D2
E2
F2
G2
H2
1500
1000
700
500
400
300
200
100
50
25

FIG. 10

sublibrary	fraction_valid_barcode_reads_mapping
s1_C-1947_10ng_BC_0468_0469	0.80255874
s2_C-1947_100ng_BC_0468_0469	0.79076673
s3_C-1948_10ng_BC_0468_0469	0.84253029
s4_C-1948_100ng_BC_0468_0469	0.83863621
s5_C-1949_10ng_BC_0467	0.06949885
s6_C-1949_100ng_BC_0467	0.09053601
s7_C-1948_100ng_BC_0468	0.63802261
s8_C-1948_100ng_BC_0469	0.82712926

FIG. 11

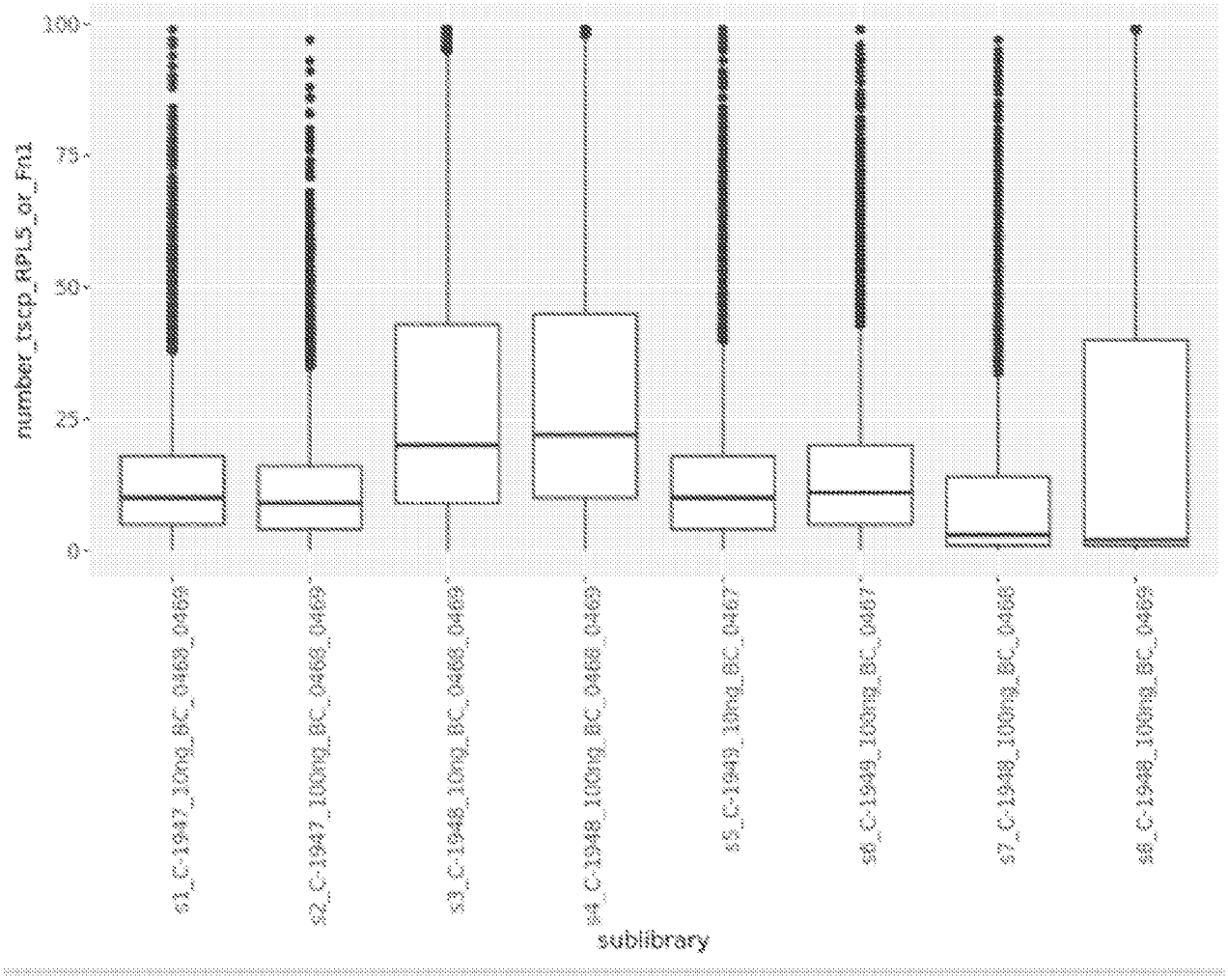

number_tscp_APLS_or_Fnl
100
75
50
25
0
s1_C-1947_32nq_BC_0469_0468
s2_C-1947_100nq_BC_0469_0469
s3_C-1948_10nq_BC_0468_0469
s4_C-1948_100nq_BC_0469_0469
s5_C-1949_10nq_BC_0467
s6_C-1949_100nq_BC_0467
s7_C-1949_100nq_BC_0468
s8_C-1949_100nq_BC_0469
sublibrary

FIG. 12

sublibrary	n_cells_with_enriched_product	possible_number_of_cells	fraction_cells_with_enriched_product
s1_C-1947_10ng_8C_0468_0469	13087	13510	0.9672095
s2_C-1947_100ng_8C_0468_0469	13145	13510	0.9729830
s3_C-1948_10ng_8C_0468_0469	13880	13985	0.9924920
s4_C-1948_100ng_8C_0468_0469	13922	13985	0.9954952
s5_C-1949_10ng_8C_0467	13584	13998	0.9704243
s6_C-1949_100ng_8C_0467	13683	13998	0.9774968
s7_C-1948_100ng_8C_0468	12852	13985	0.9189846
s8_C-1948_100ng_8C_0469	11195	13985	0.8005005

FIG. 13A

sublibrary	n_cells_with_enriched_product	possible_number_of_cells	fraction_cells_with_enriched_product
s1_C-1947_10ng_8C_0468_0469	12934	13510	0.9573648
s2_C-1947_100ng_8C_0468_0469	12869	13510	0.9525537
s3_C-1948_10ng_8C_0468_0469	13754	13985	0.9834823
s4_C-1948_100ng_8C_0468_0469	13673	13985	0.9776904
s5_C-1949_10ng_8C_0467	10897	13998	0.7784684
s6_C-1949_100ng_8C_0467	9323	13998	0.6660237
s7_C-1948_100ng_8C_0468	10326	13985	0.7383625
s8_C-1948_100ng_8C_0469	7988	13985	0.5711834

FIG. 13B

FIG. 14A

FIG. 14B

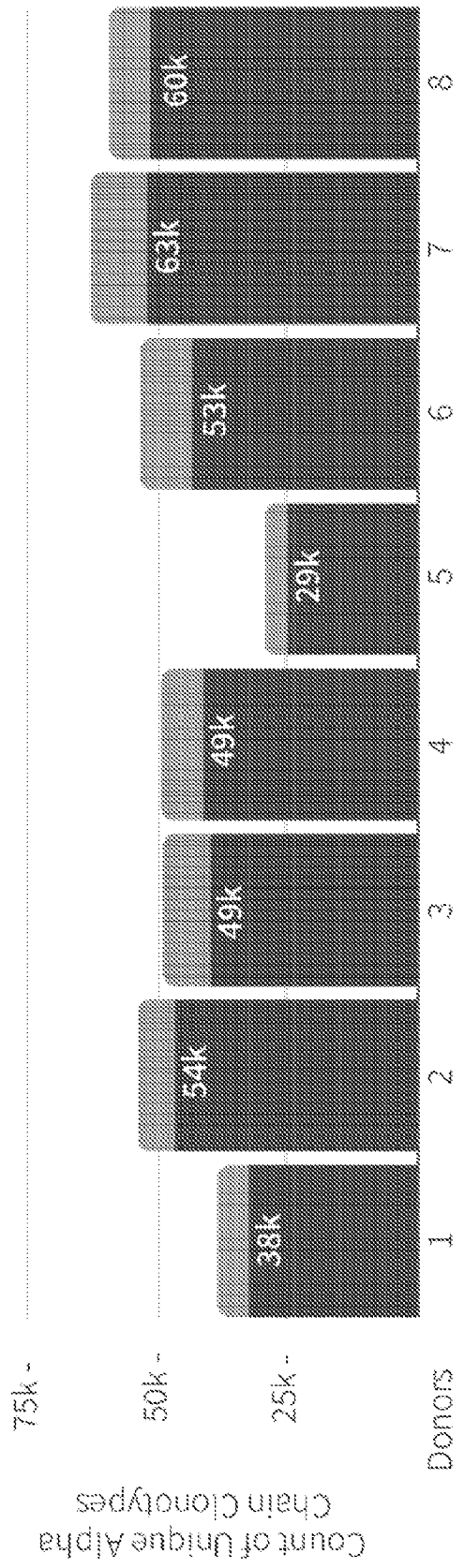


FIG. 15A

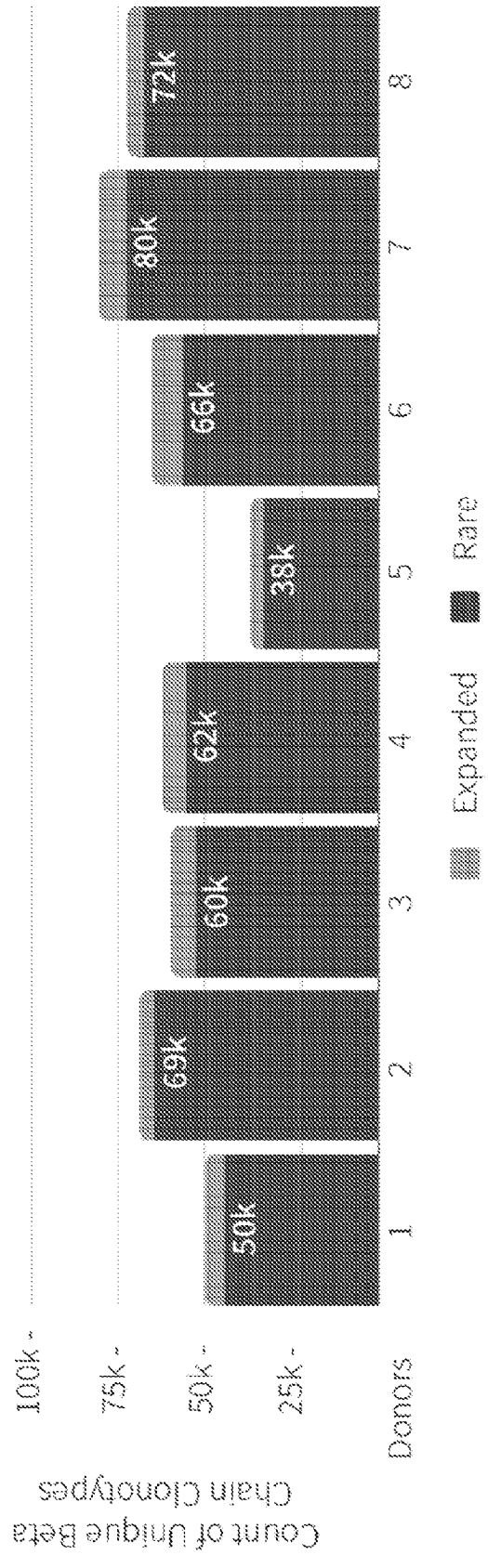


FIG. 15B

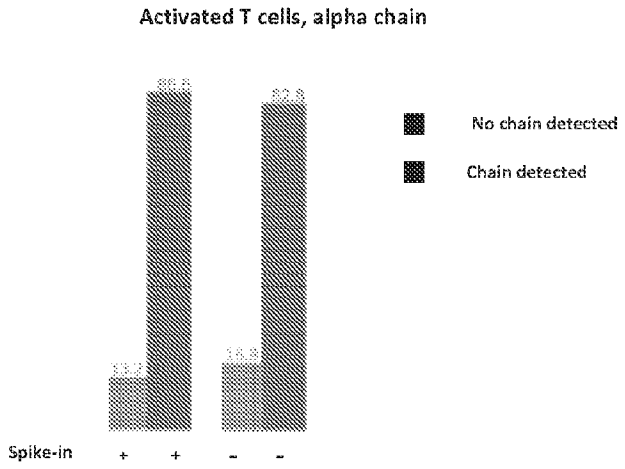


FIG. 16A

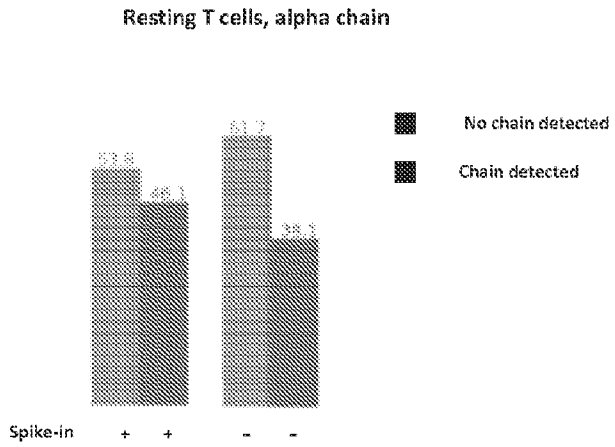


FIG. 16B

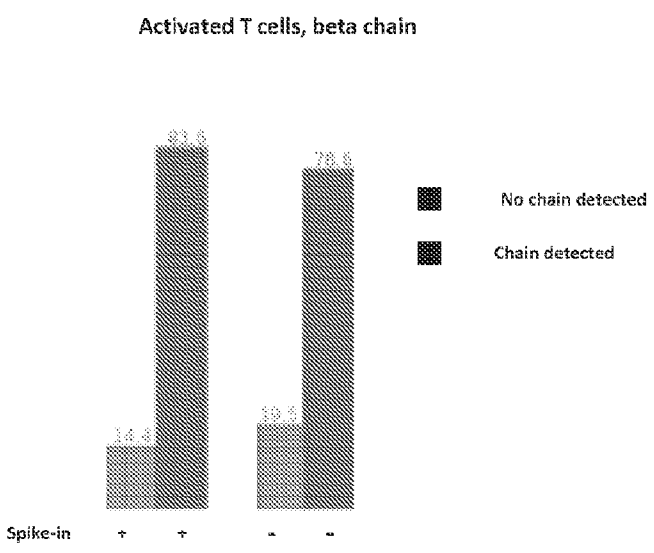


FIG. 16C

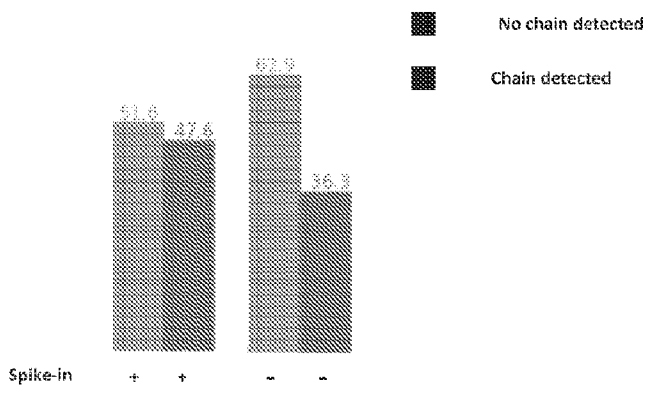


FIG. 16D