US 20250257396A1

(19) United States
(12) Patent Application Publication
Belgrader et al.
(10) Pub. No.: US 2025/0257396 A1
(43) Pub. Date: Aug. 14, 2025

(54) **METHODS AND SYSTEMS FOR PROCESSING POLYNUCLEOTIDES**

(71) Applicant: **10x Genomics, Inc.**, Pleasanton, CA (US)

(72) Inventors: **Phillip Belgrader**, Livermore, CA (US); **Zachary Bent**, Pleasanton, CA (US); **Rajiv Bharadwaj**, Pleasanton, CA (US); **Vijay Kumar Sreenivasa Gopalan**, Pleasanton, CA (US); **Josephine Harada**, San Francisco, CA (US); **Christopher Hindson**, Livermore, CA (US); **Mohammad Rahimi Lenji**, Pleasanton, CA (US); **Michael Ybarra Lucero**, South San Francisco, CA (US); **Geoffrey McDermott**, Livermore, CA (US); **Elliott Meer**, Pleasanton, CA (US); **Tarjei Sigurd Mikkelsen**, Dublin, CA (US); **Christopher Joachim O'Keeffe**, Pleasanton, CA (US); **Katherine Pfeiffer**, Berkeley, CA (US); **Andrew D. Price**, Hayward, CA (US); **Paul Ryvkin**, San Jose, CA (US); **Serge Saxonov**, Oakland, CA (US); **John R. Stuelpnagel**, Santa Barbara, CA (US); **Jessica Michele Terry**, Pleasanton, CA (US); **Tobias Daniel Wheeler**, Alameda, CA (US); **Indira Wu**, San Carlos, CA (US); **Solongo Batjargal Ziraldo**, San Ramon, CA (US); **Stephane Claude Boutet**, Burlingame, CA (US); **Sarah Taylor**, Pleasanton, CA (US); **Niranjan Srinivas**, Dublin, CA (US)

(21) Appl. No.: **19/083,328**

(22) Filed: **Mar. 18, 2025**

Related U.S. Application Data

(63) Continuation of application No. 18/786,209, filed on Jul. 26, 2024, which is a continuation of application No. 17/473,760, filed on Sep. 13, 2021, now Pat. No. 12,084,716, which is a continuation of application No. 17/084,481, filed on Oct. 29, 2020, now Pat. No. 11,180,805, which is a continuation of application No. 16/443,638, filed on Jun. 17, 2019, now Pat. No. 10,858,702, which is a continuation of application No. 16/439,675, filed on Jun. 12, 2019, now Pat. No. 10,815,525, which is a continuation-in-part of application No. PCT/US2017/068320, filed on Dec. 22, 2017, which is a continuation-in-part of application No. 15/720,085, filed on Sep. 29, 2017, now Pat. No. 10,011,872, said application No. 16/439,675 is a continuation-in-part of application No. PCT/US2018/064600, filed on Dec. 7, 2018, which is a continuation-in-part of application No. 16/107,685, filed on Aug. 21, 2018, now abandoned, said application No. 16/439,675 is a continuation-in-part of application No. 16/107,685, filed on Aug. 21, 2018, now abandoned, said application No. 16/439,675 is a continuation-in-part of application No. 15/933,299, filed on Mar. 22, 2018, now Pat. No. 10,480,029, which is a continuation of application No. 15/720,085, filed on Sep. 29, 2017, now Pat. No. 10,011,872.

(60) Provisional application No. 62/438,341, filed on Dec. 22, 2016, provisional application No. 62/596,557, filed on Dec. 8, 2017, provisional application No. 62/723,960, filed on Aug. 28, 2018, provisional application No. 62/596,557, filed on Dec. 8, 2017, provisional application No. 62/596,557, filed on Dec. 8, 2017, provisional application No. 62/438,341, filed on Dec. 22, 2016.

Publication Classification

(51) **Int. Cl.**
C12Q 1/6874 (2018.01)
C12Q 1/6806 (2018.01)
C12Q 1/683 (2018.01)

(52) **U.S. Cl.**
CPC ***C12Q 1/6874*** (2013.01); ***C12Q 1/6806*** (2013.01); ***C12Q 1/683*** (2013.01); *C12Q 2525/191* (2013.01); *C12Q 2537/143* (2013.01); *C12Q 2563/179* (2013.01)

(57) **ABSTRACT**

The present disclosure provides compositions, methods, systems, and devices for polynucleotide processing and analyte characterization. Such polynucleotide processing may be useful for a variety of applications, including analyte characterization by polynucleotide sequencing. The compositions, methods, systems, and devices disclosed herein generally describe barcoded oligonucleotides, which can be bound to a bead, such as a gel bead, useful for characterizing one or more analytes including, for example, protein (e.g., cell surface or intracellular proteins), genomic DNA, and RNA (e.g., mRNA or CRISPR guide RNAs). Also described herein, are barcoded labelling agents and oligonucleotide molecules useful for "tagging" analytes for characterization.

Specification includes a Sequence Listing.

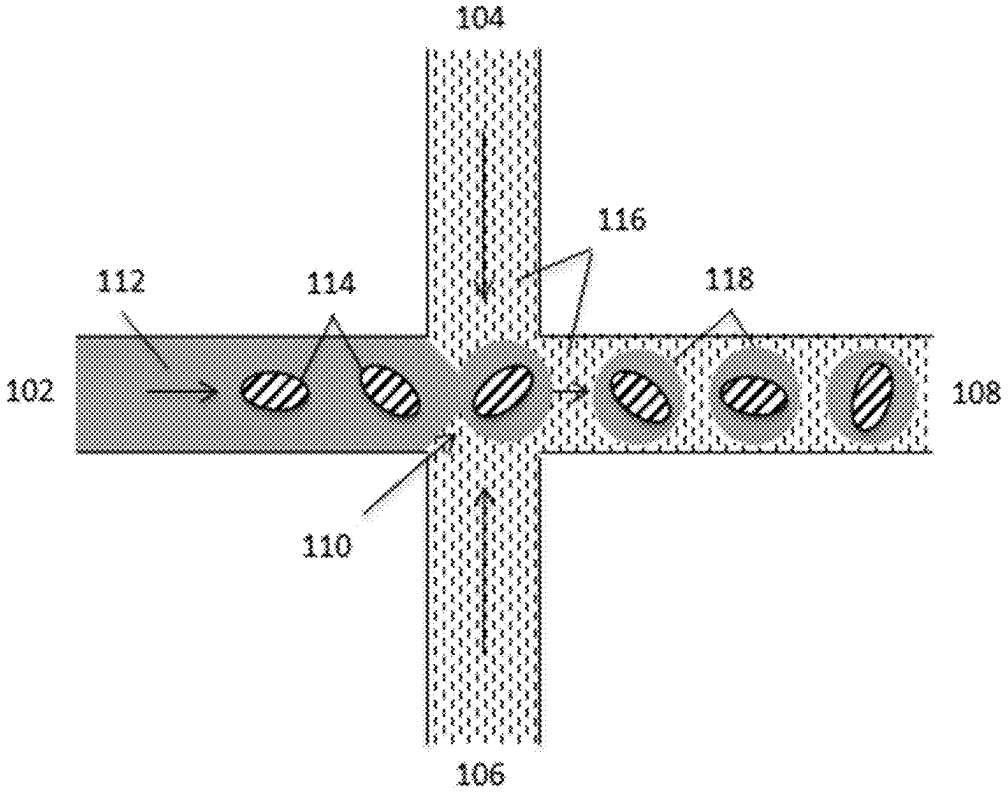


FIG. 1

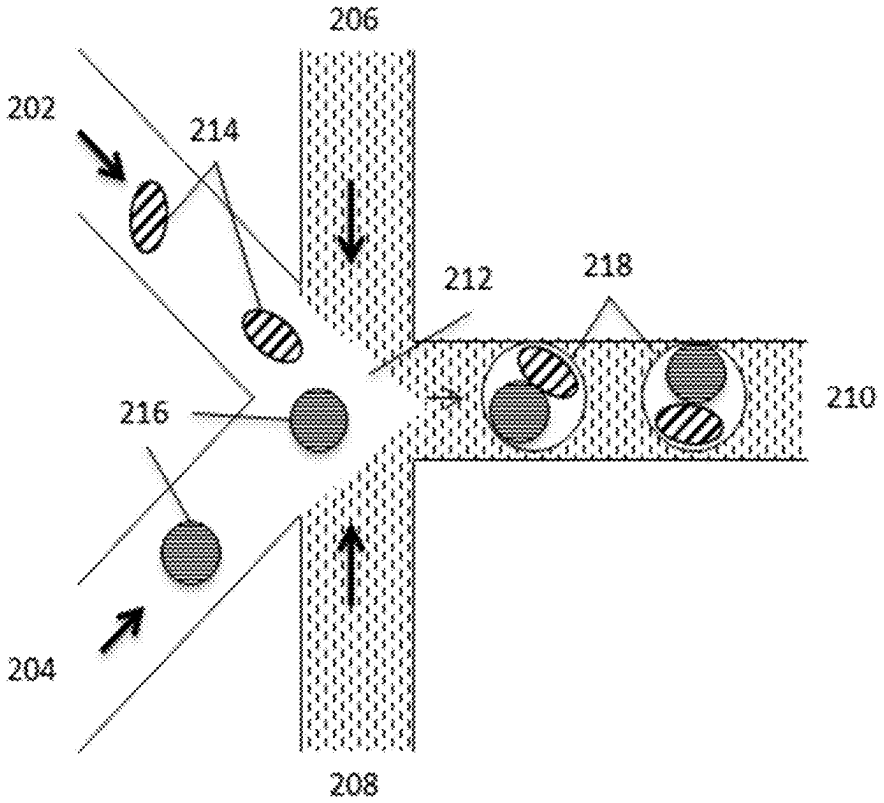


FIG. 2

FIG. 3A

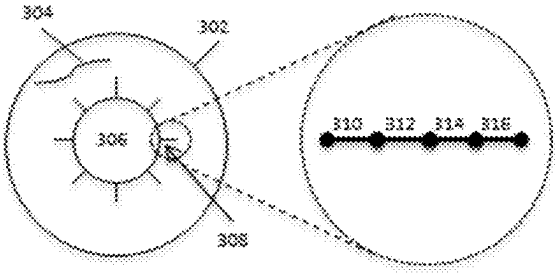


FIG. 3B

FIG. 3C

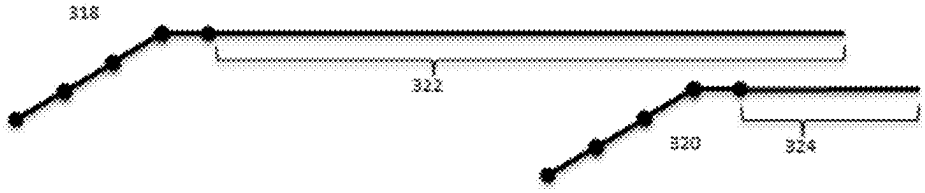


FIG. 3D

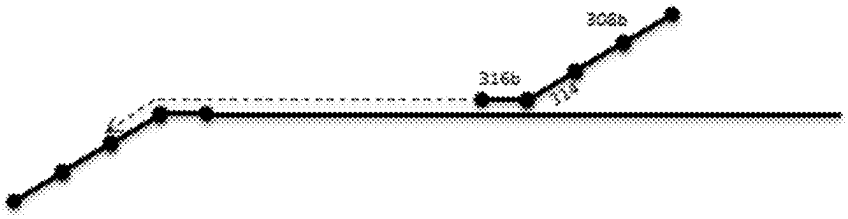


FIG. 3E

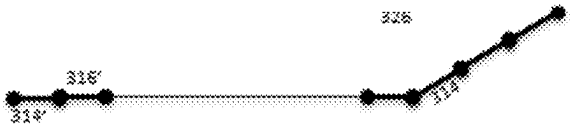
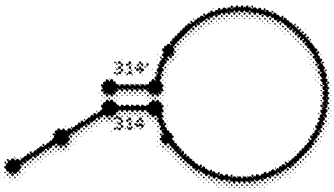


FIG. 3F



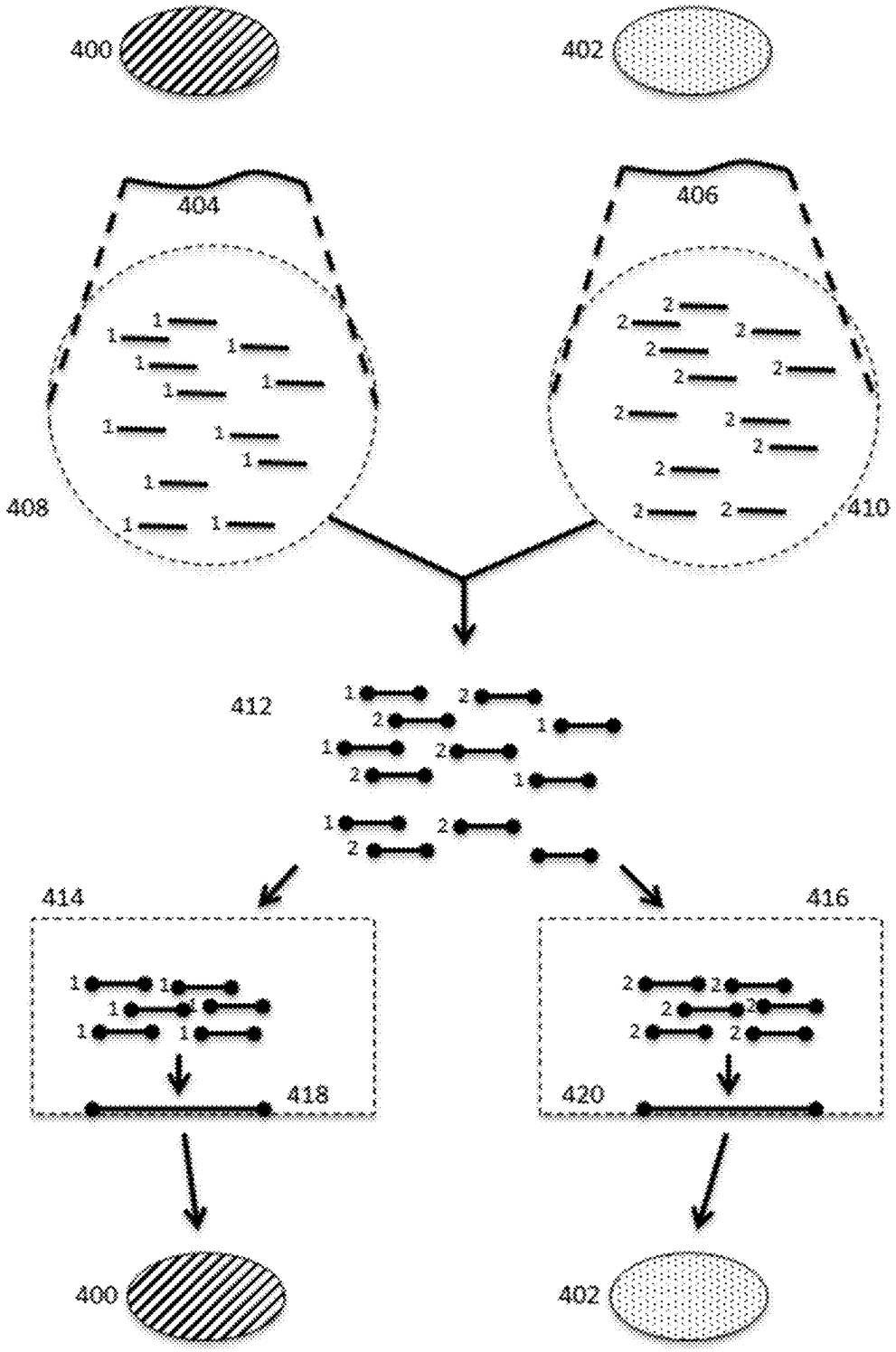


FIG. 4

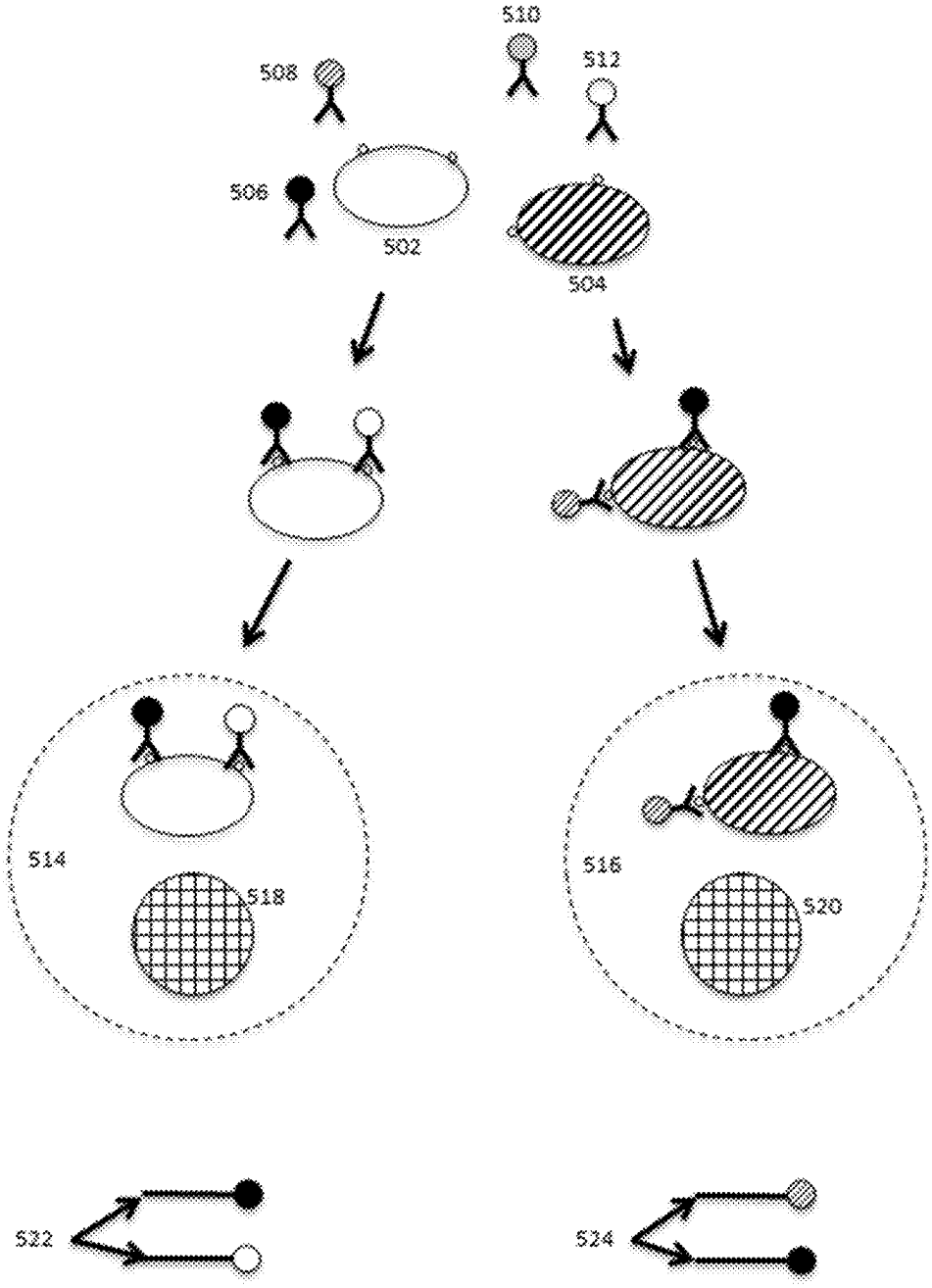


FIG. 5

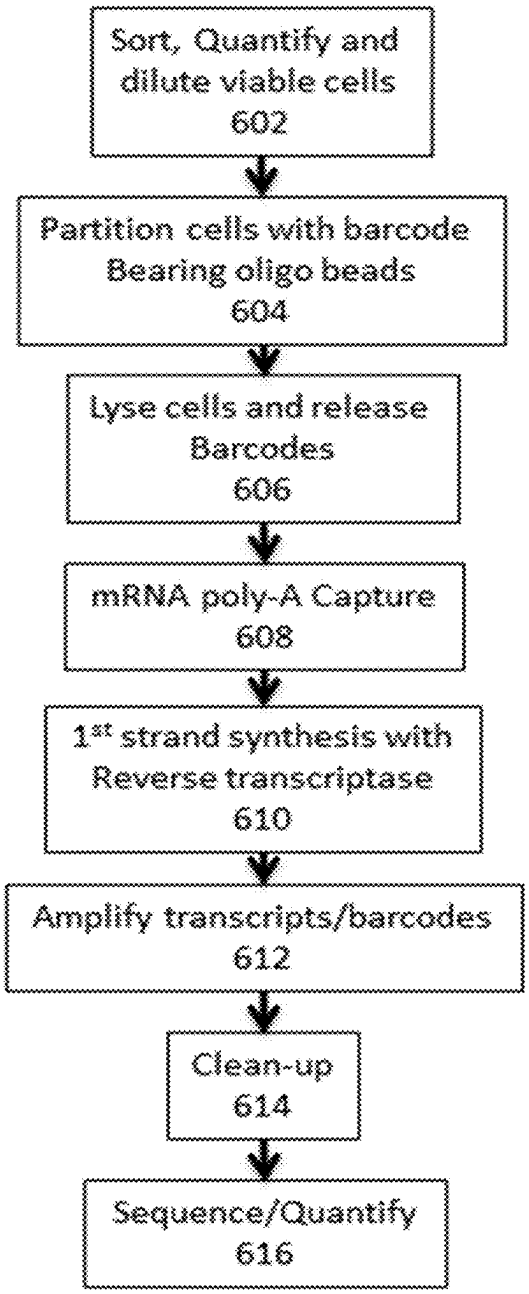

Sort, Quantify and dilute viable cells 602
Partition cells with barcode Bearing oligo beads 604
Lyse cells and release Barcodes 606
mRNA poly-A Capture 608
1st strand synthesis with Reverse transcriptase 610
Amplify transcripts/barcodes 612
Clean-up 614
Sequence/Quantify 616

FIG. 6

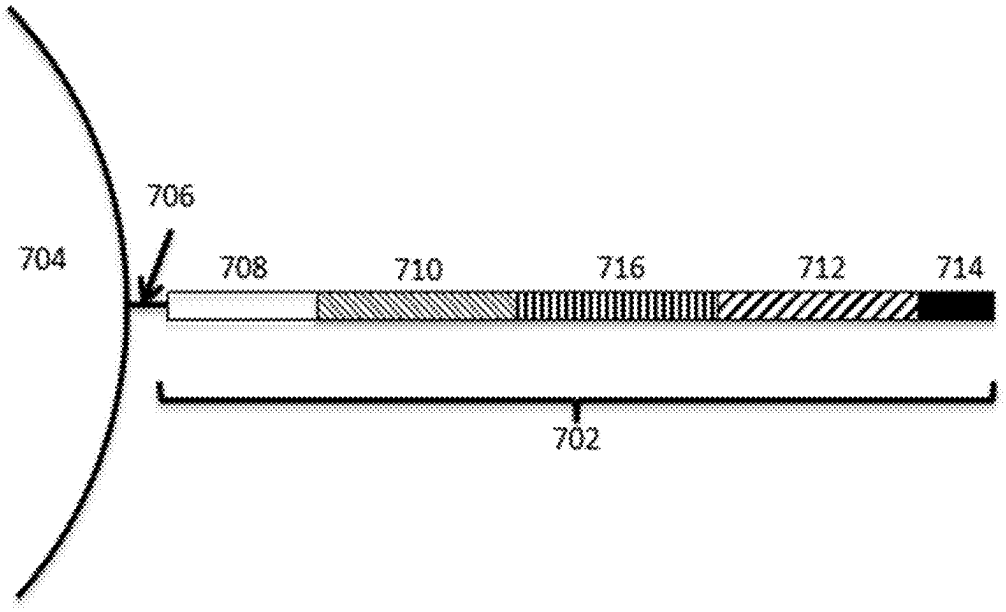
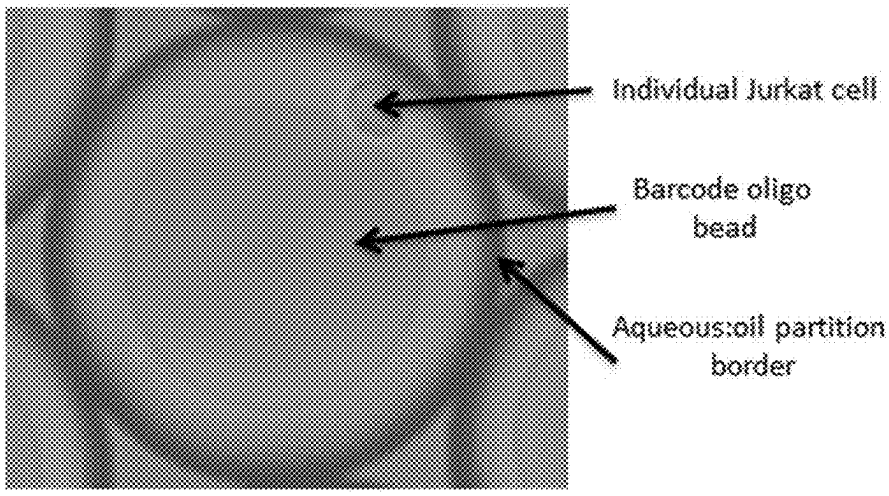


FIG. 7


Individual Jurkat cell
Barcode oligo bead
Aqueous:oil partition border

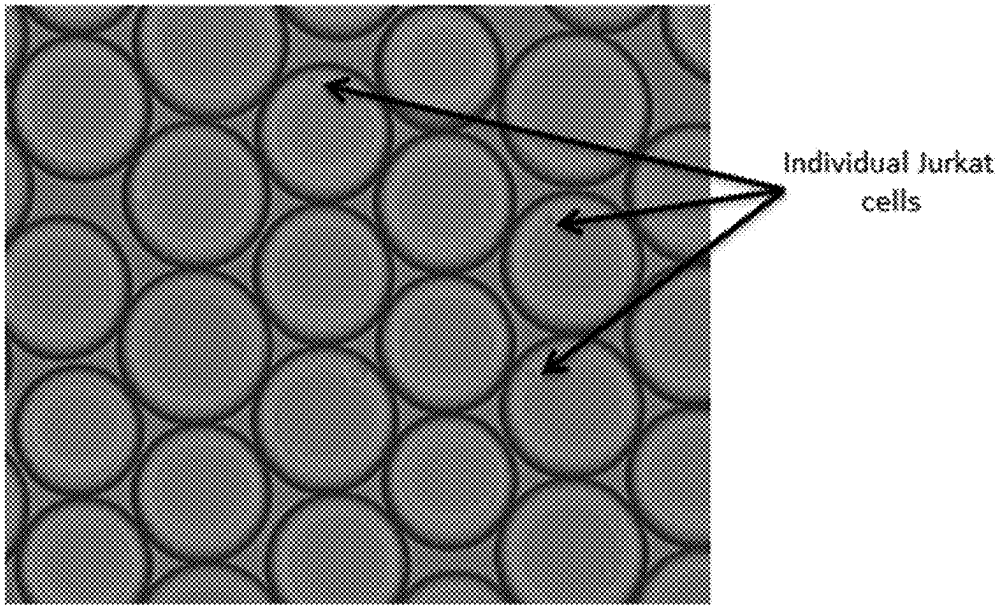

Individual Jurkat cells

FIG. 8

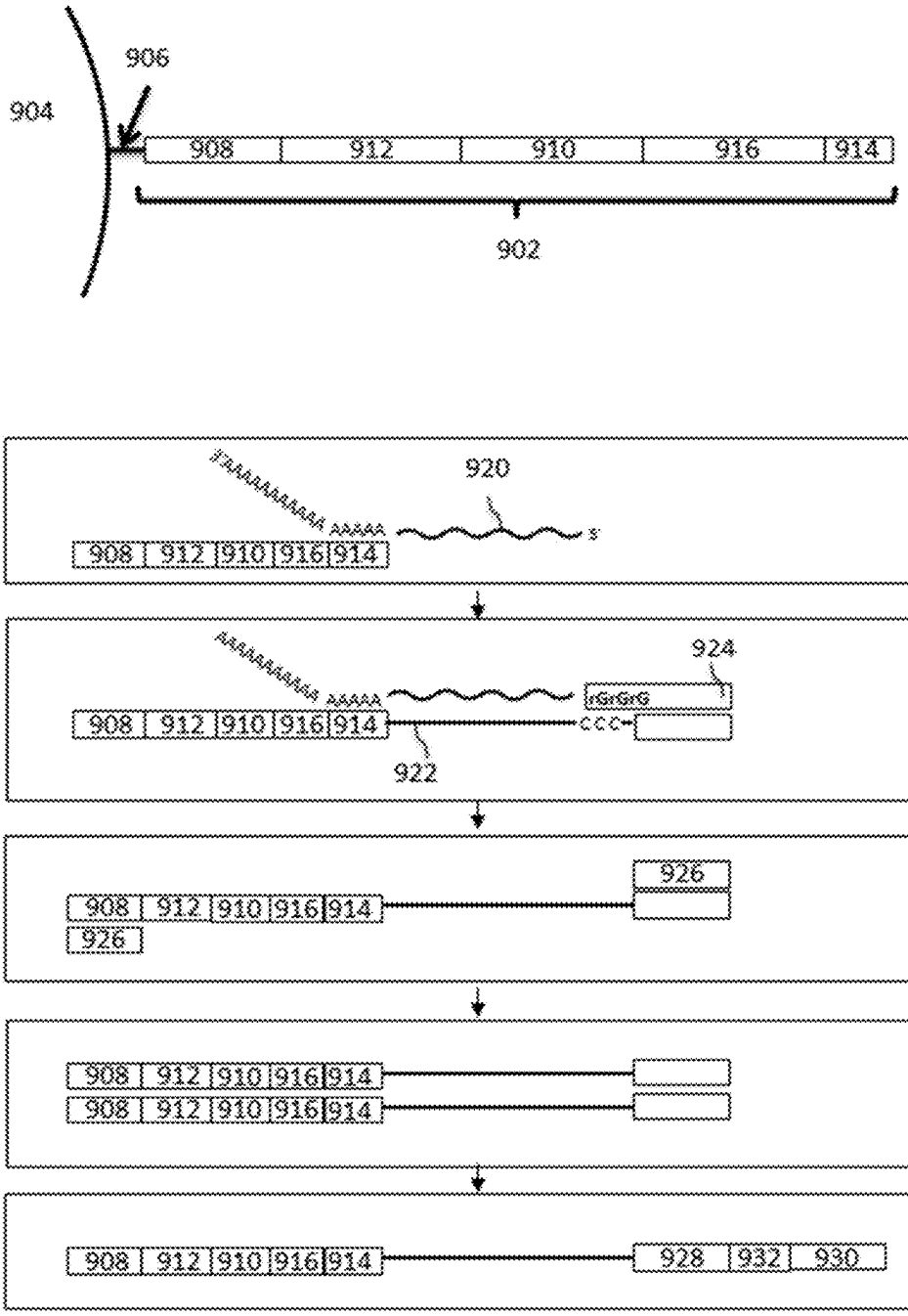

906
904
908
912
910
916
914
902
950
920
908
912
910
916
914
952
924
rGrGrG
908
912
910
916
914
922
954
926
908
912
910
916
914
926
956
908
912
910
916
914
908
912
910
916
914
958
908
912
910
916
914
928
932
930

FIG. 9A

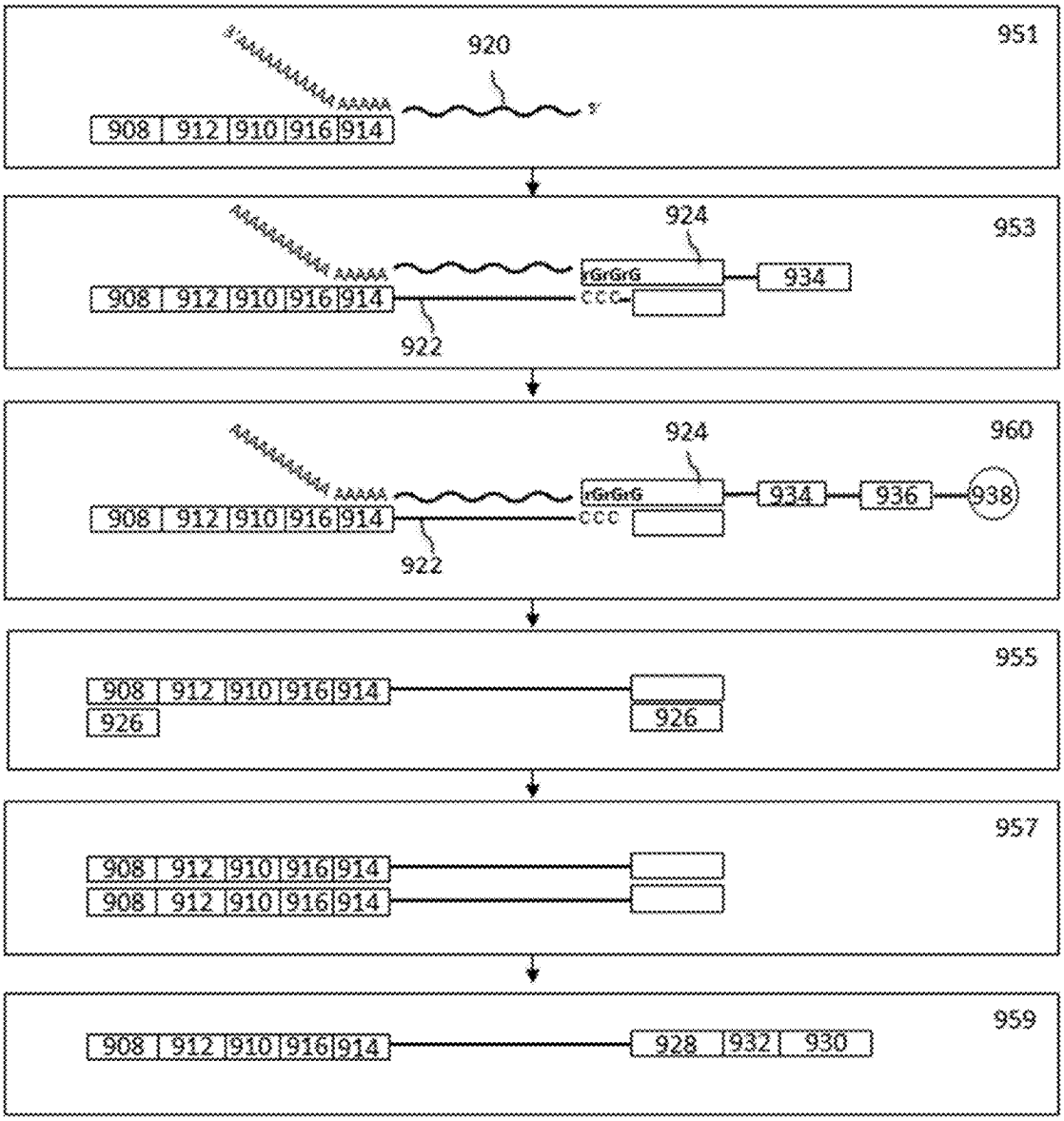

951
920
908
912
910
916
914
953
924
934
922
960
936
938
955
926
957
959
928
932
930

FIG. 9B

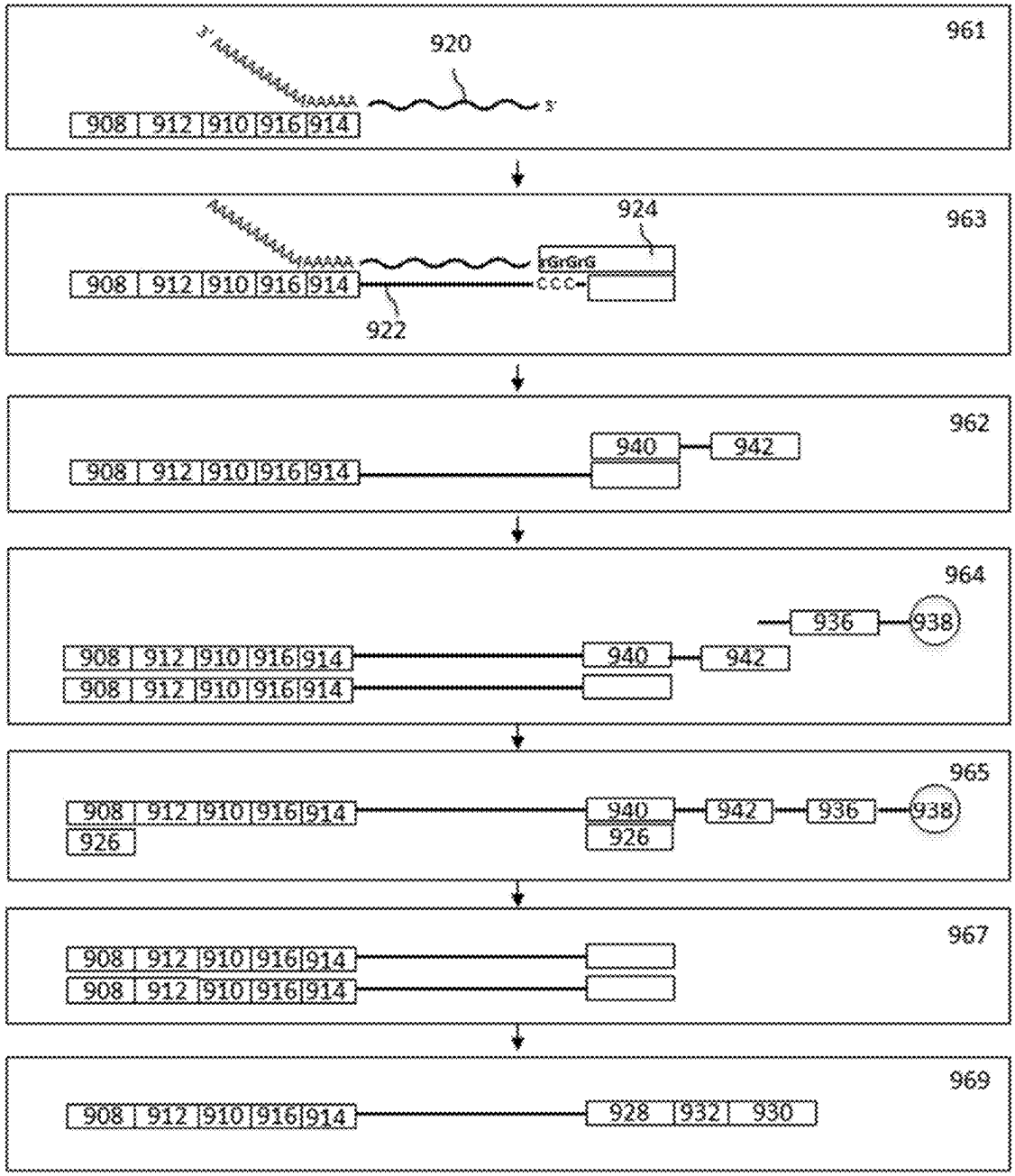

961
920
3'
5'
908
912
910
916
914
963
924
922
962
940
942
964
936
938
965
926
967
969
928
932
930

FIG. 9C

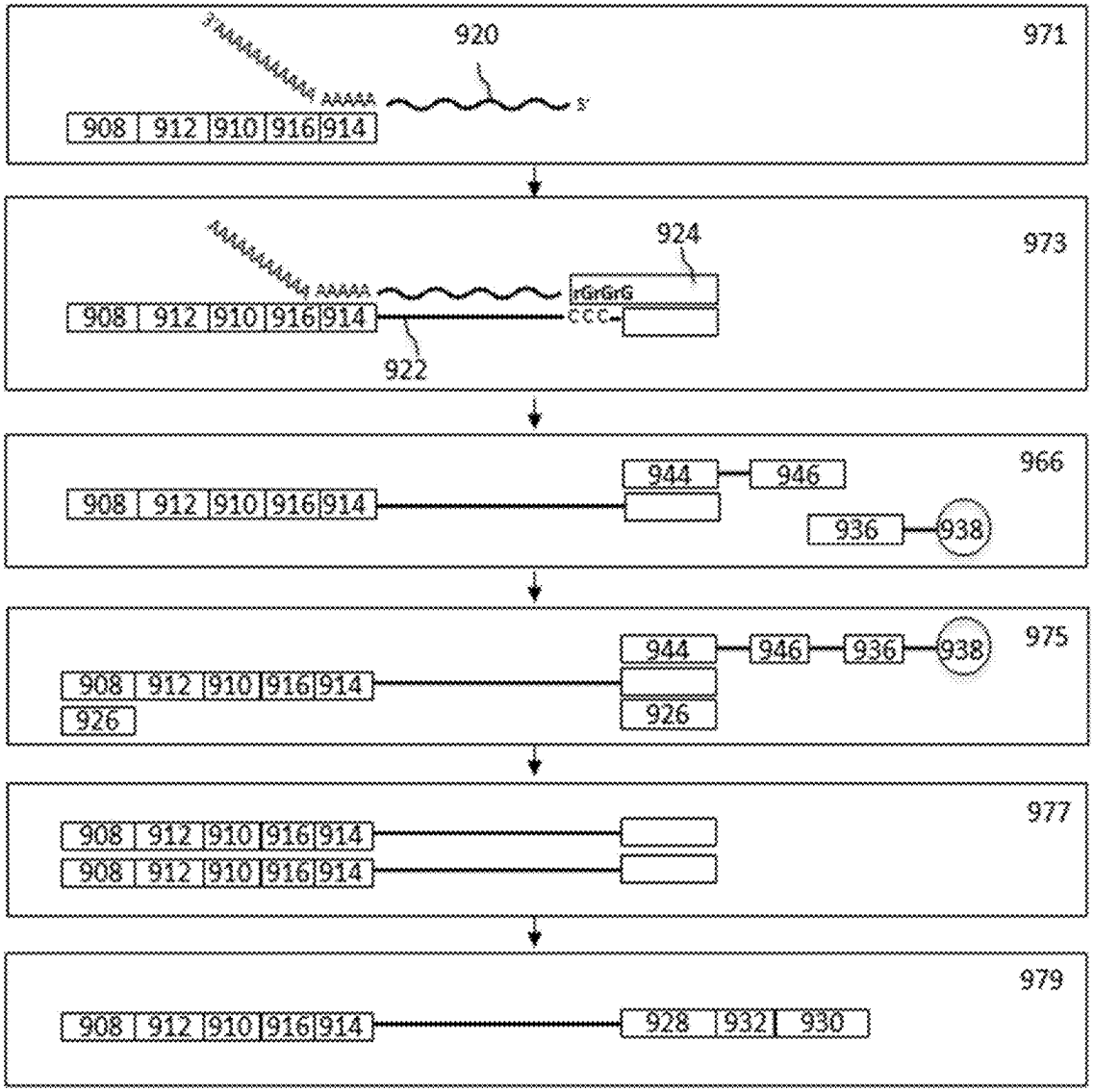

971
920
908
912
910
916
914
973
924
922
966
944
946
936
938
975
926
977
979
928
932
930

FIG. 9D

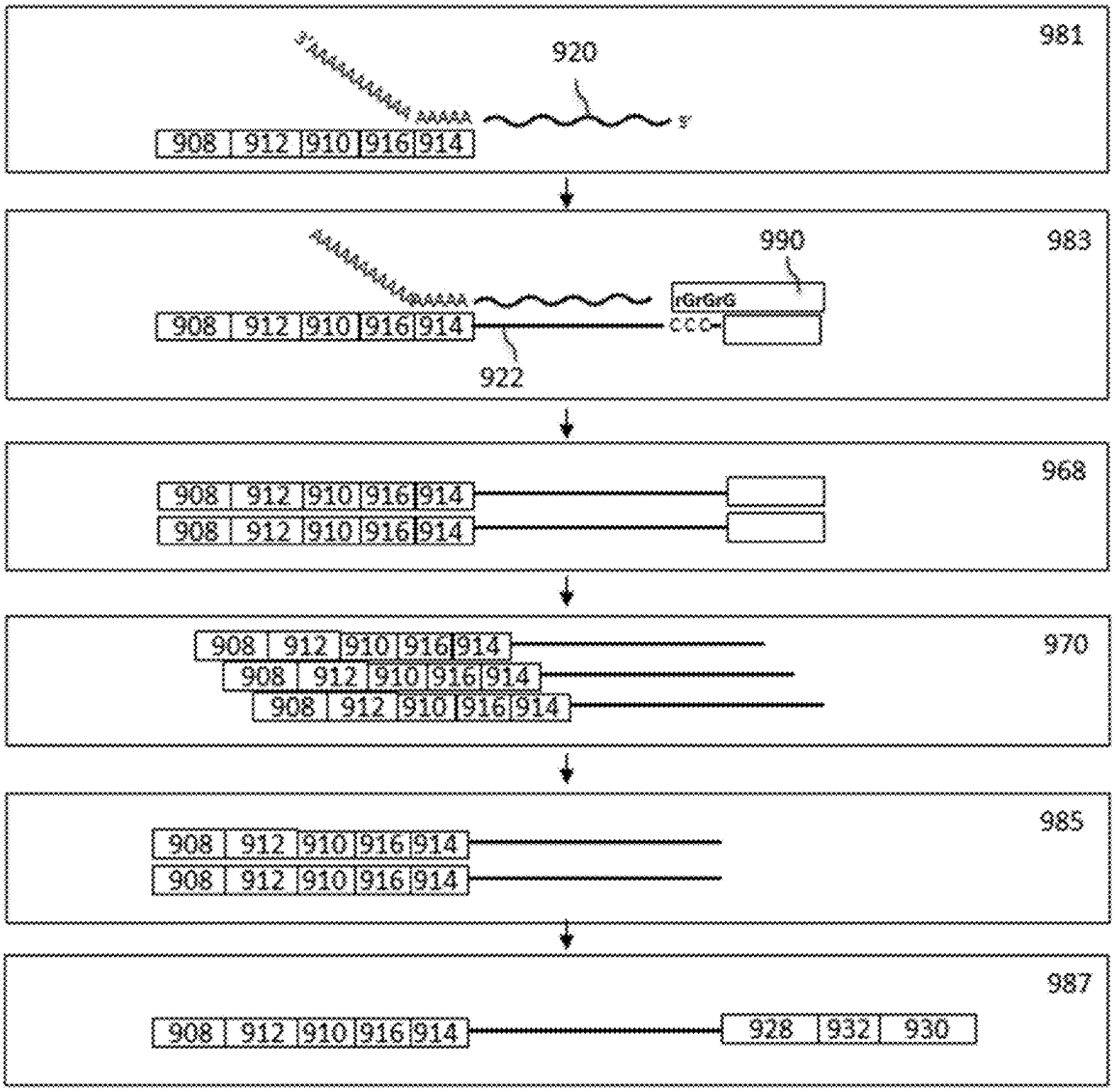

981
920
908
912
910
916
914
983
990
rGrGrG
922
968
970
985
987
928
932
930

FIG. 9E

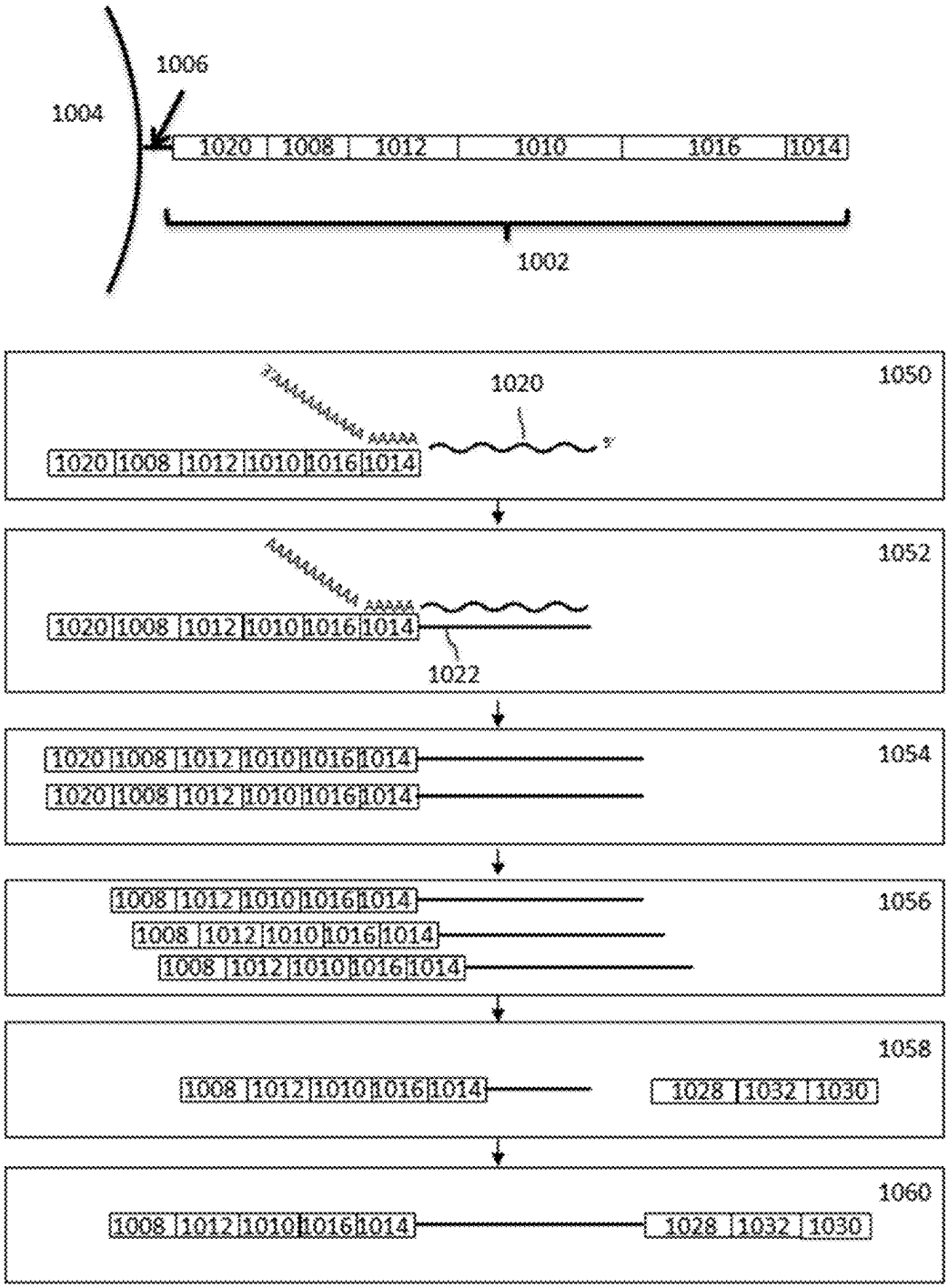
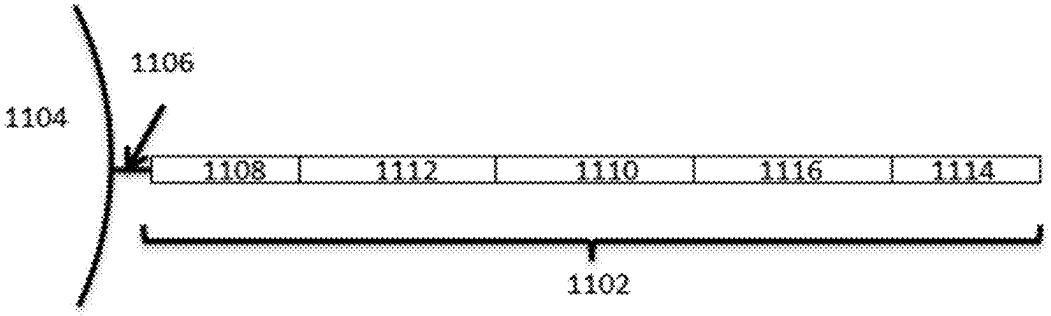

1004
1006
1020
1008
1012
1010
1016
1014
1002
1050
1020
1052
1022
1054
1056
1058
1028
1032
1030
1060

FIG. 10


1106
1104
1108
1112
1110
1116
1114
1102

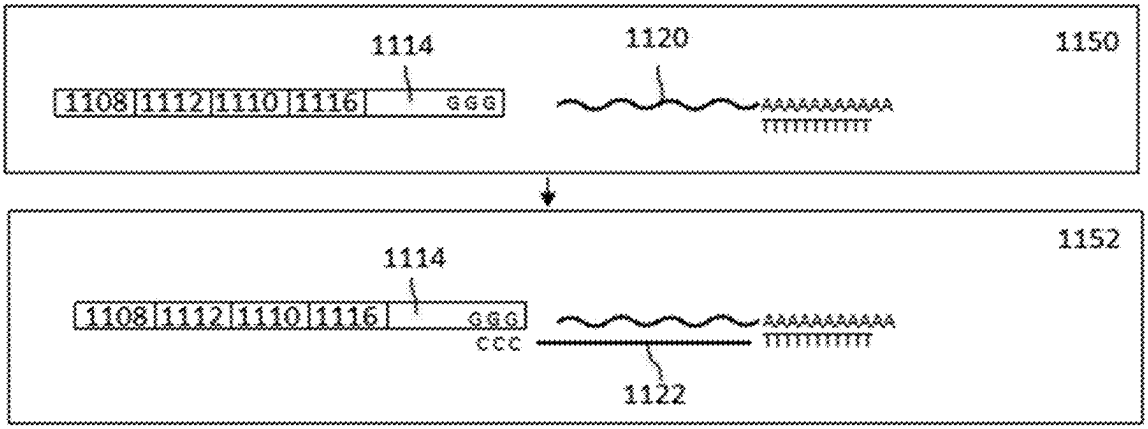

1114
1120
1150
1108
1112
1110
1116
ggg
AAAAAAAAAAA
TTTTTTTTTTT
1152
1114
1108
1112
1110
1116
ggg
ccc
AAAAAAAAAAA
TTTTTTTTTTT
1122

FIG. 11

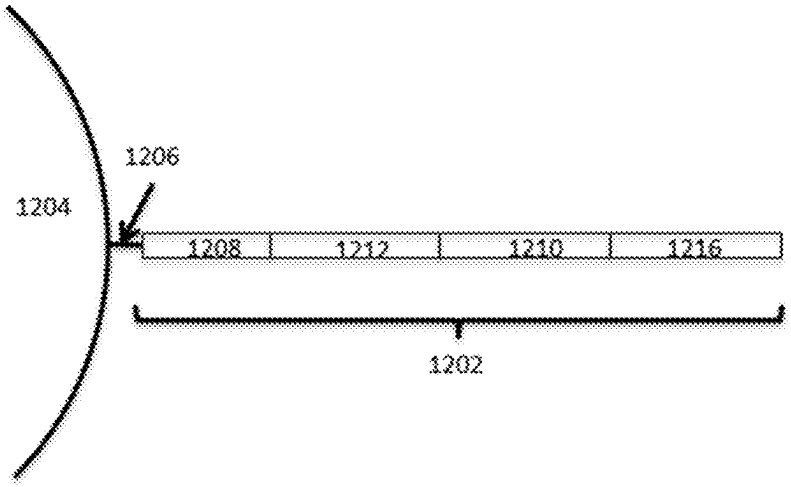


FIG. 12A

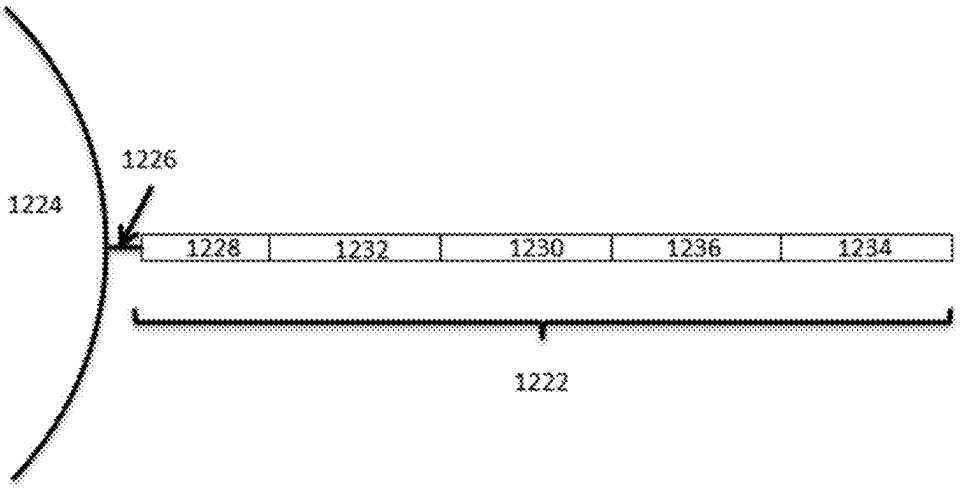


FIG. 12B

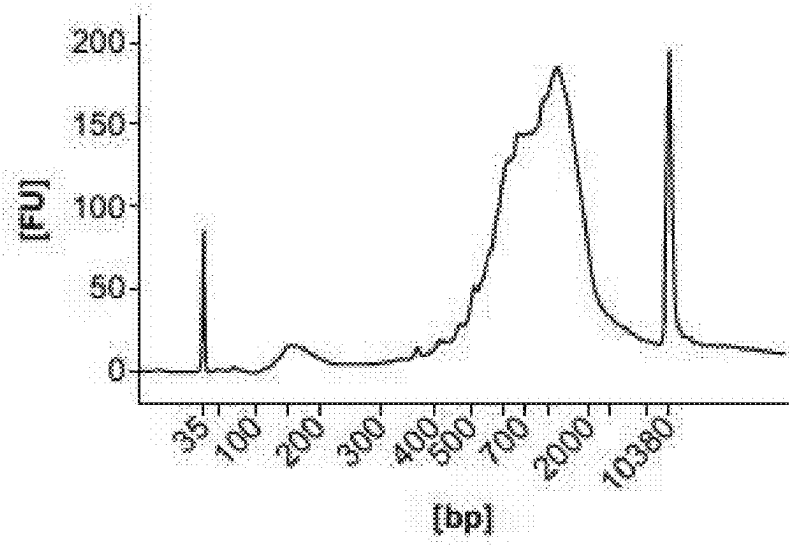


Figure 13A

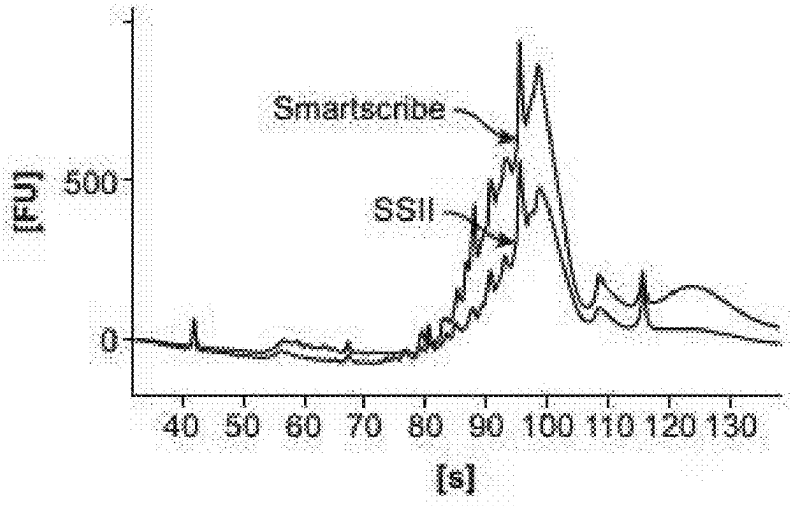


Figure 13B

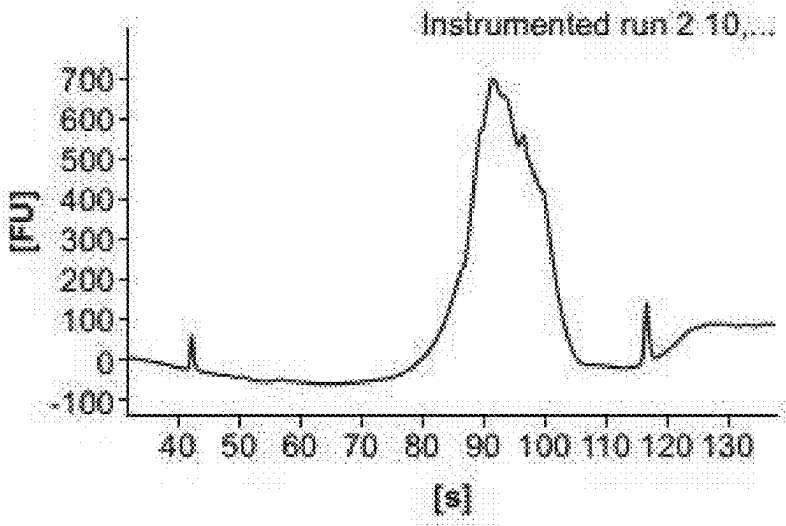


Figure 13C

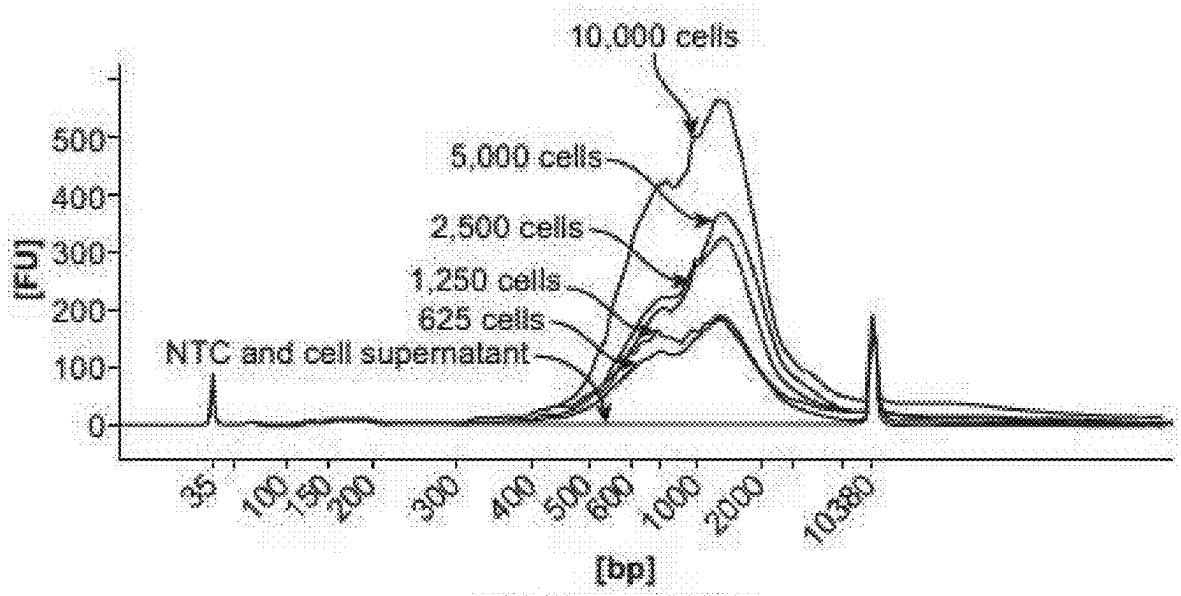


Figure 14A

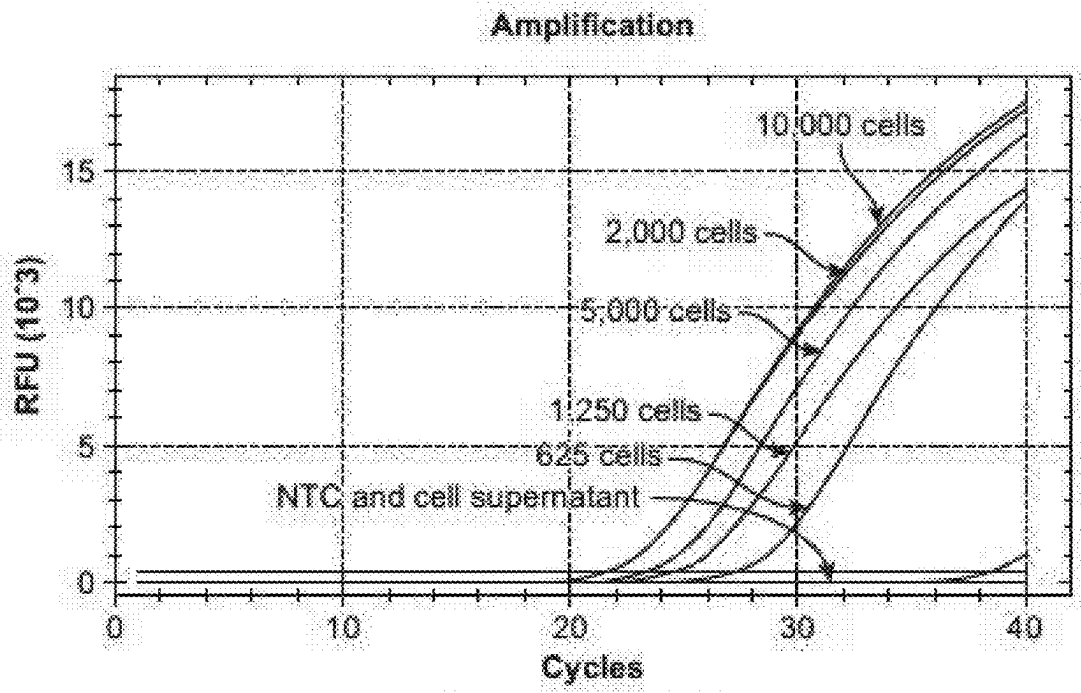


Figure 14B

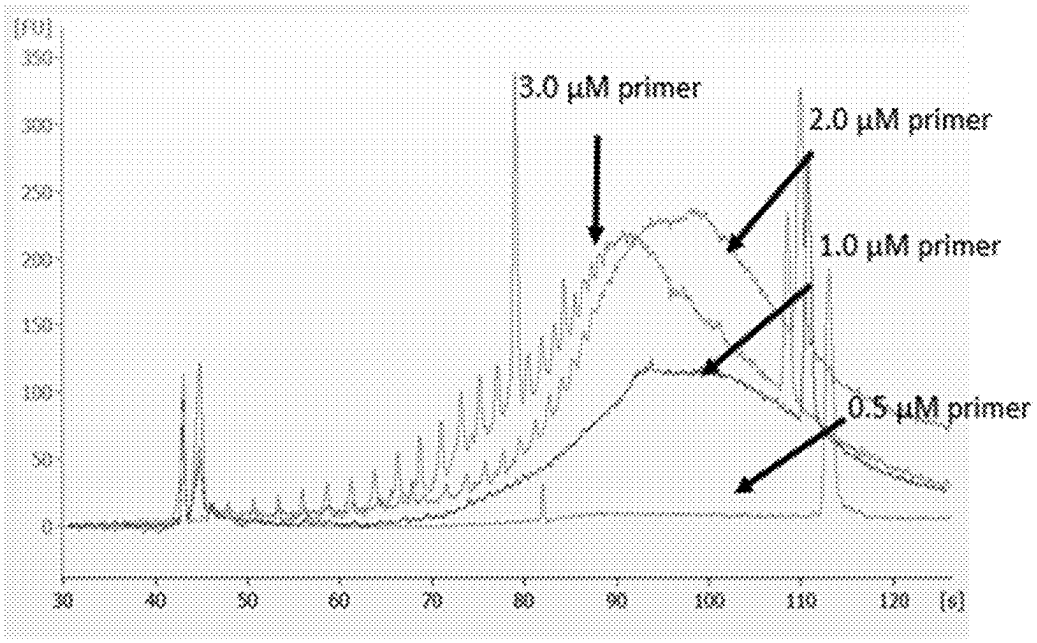

[FU]
350
300
250
200
150
100
50
0
30
40
50
60
70
80
90
100
110
120
[s]
3.0 µM primer
2.0 µM primer
1.0 µM primer
0.5 µM primer

FIG. 15

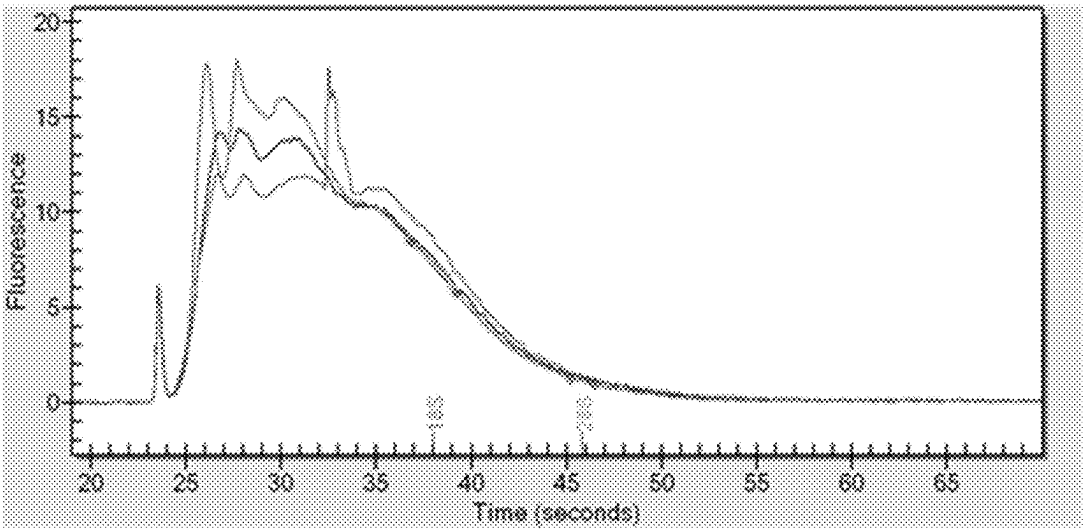

Fluorescence
20
15
10
5
0
20
25
30
35
40
45
50
55
60
65
Time (seconds)

FIG. 16

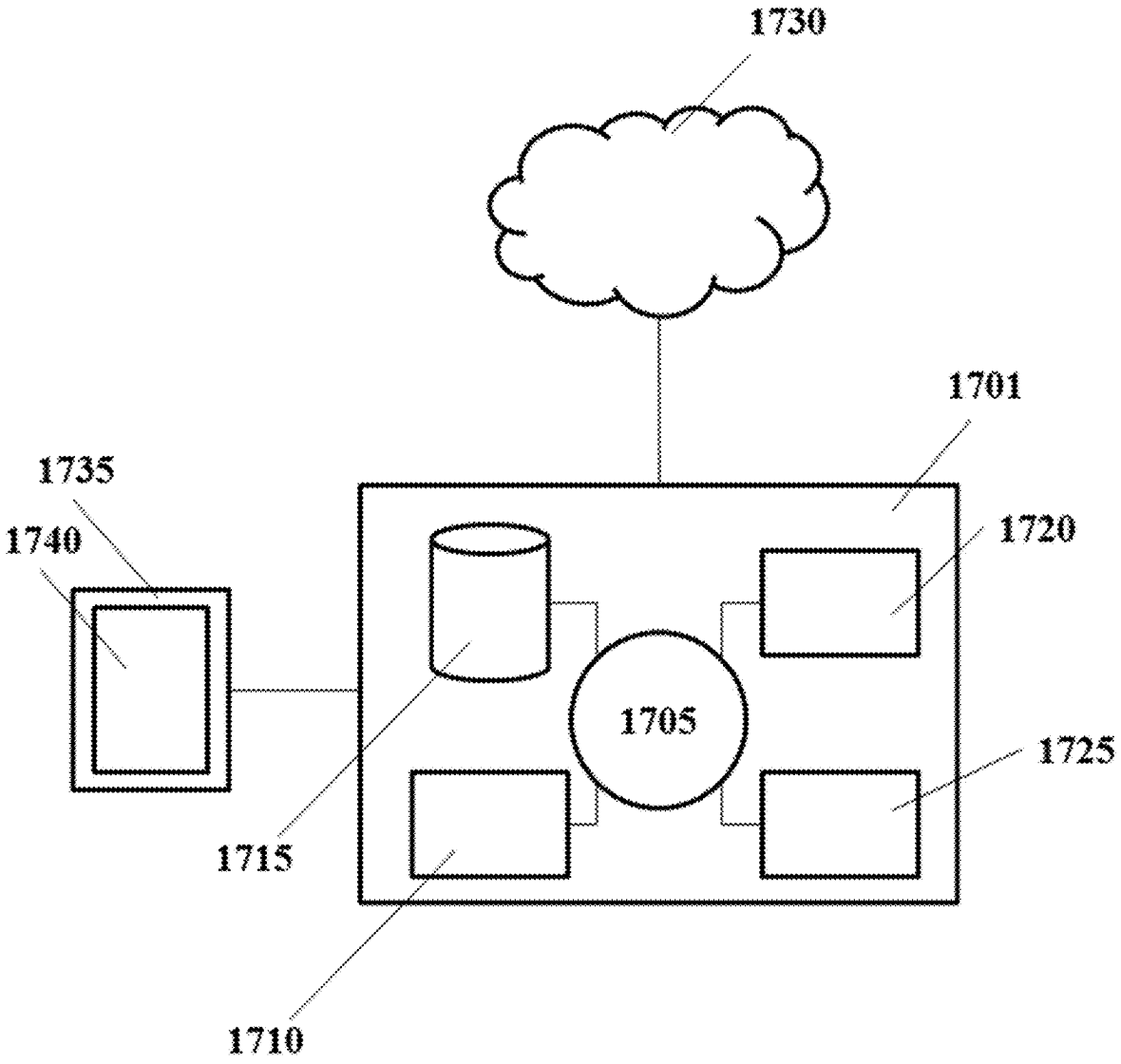


FIG. 17

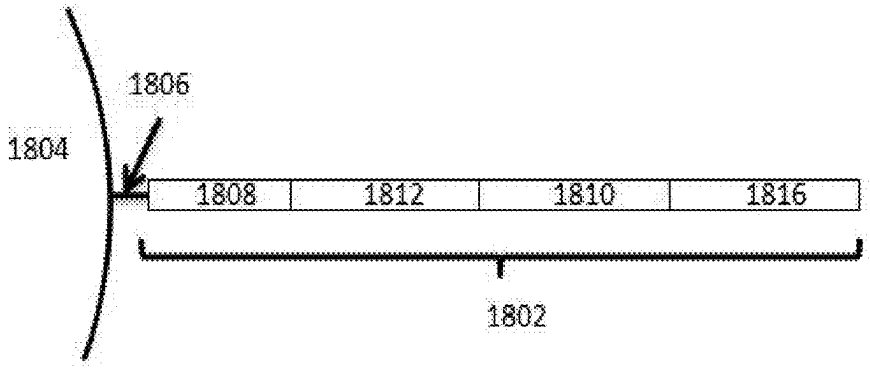

1806
1804
1808
1812
1810
1816
1802

FIG. 18

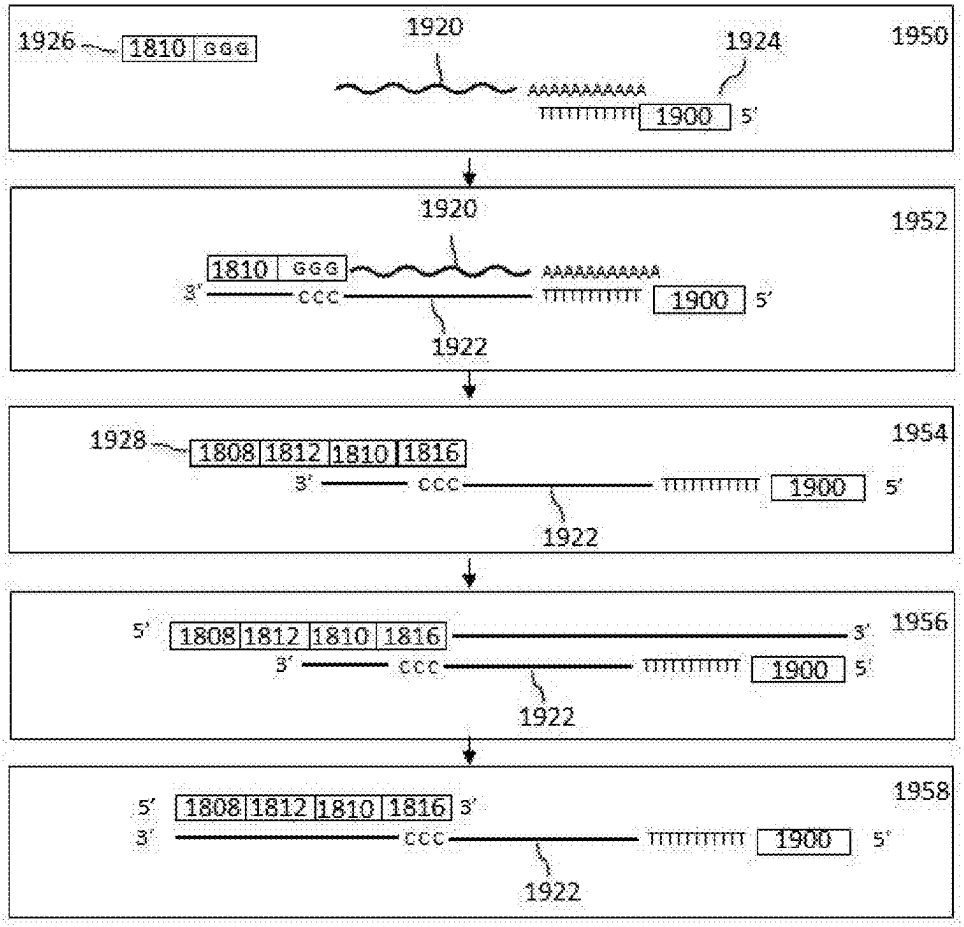

1926
1810
GGG
1920
1924
1950
AAAAAAAAAAA
1900
5'
1952
1920
1810
GGG
AAAAAAAAAAA
3'
CCC
1900
5'
1922
1954
1928
1808
1812
1810
1816
3'
CCC
1900
5'
1922
1956
5'
1808
1812
1810
1816
3'
3'
CCC
1900
5'
1922
1958
5'
1808
1812
1810
1816
3'
3'
CCC
1900
5'
1922

FIG. 19

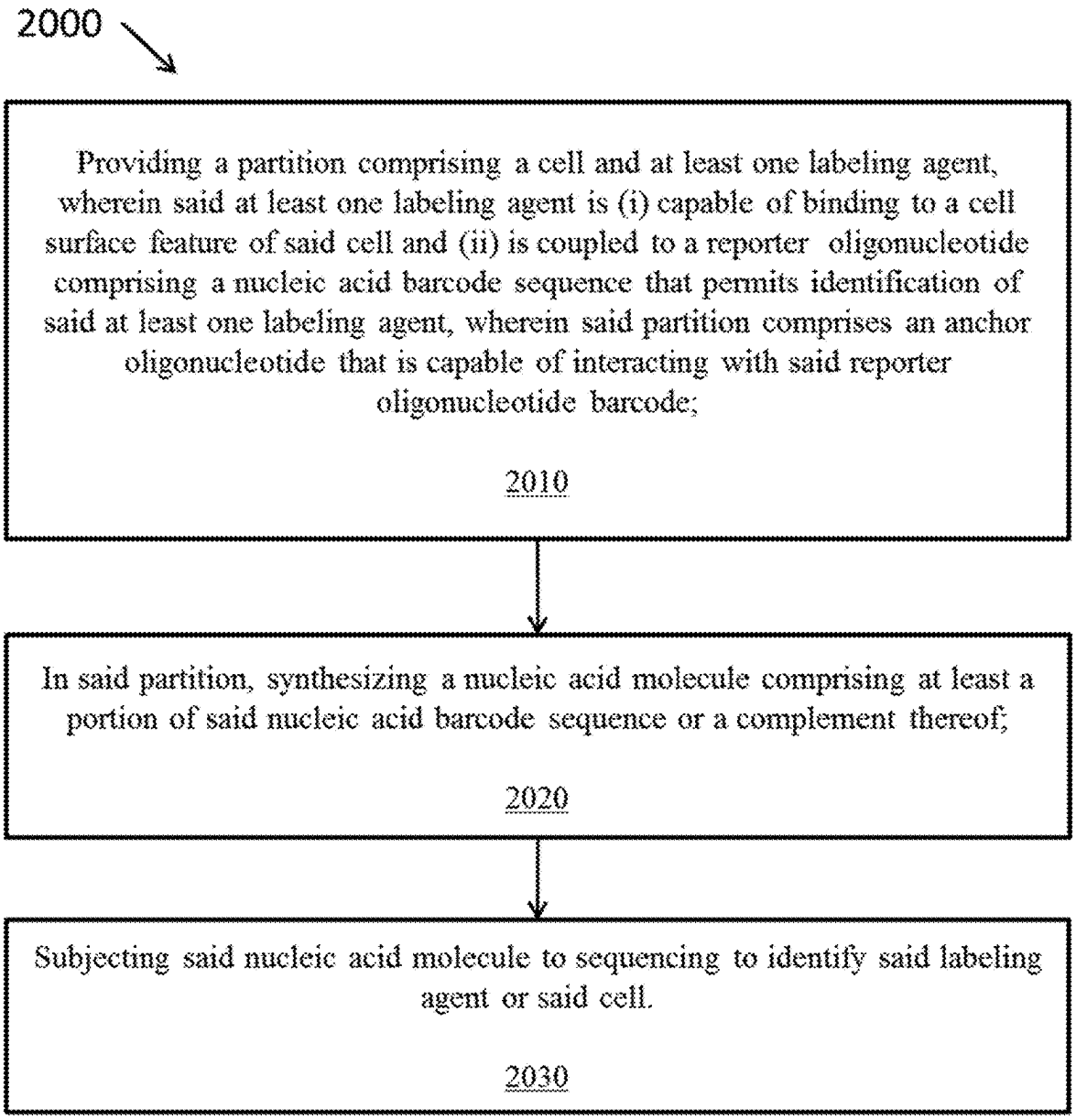

2000
Providing a partition comprising a cell and at least one labeling agent, wherein said at least one labeling agent is (i) capable of binding to a cell surface feature of said cell and (ii) is coupled to a reporter oligonucleotide comprising a nucleic acid barcode sequence that permits identification of said at least one labeling agent, wherein said partition comprises an anchor oligonucleotide that is capable of interacting with said reporter oligonucleotide barcode;
2010
In said partition, synthesizing a nucleic acid molecule comprising at least a portion of said nucleic acid barcode sequence or a complement thereof;
2020
Subjecting said nucleic acid molecule to sequencing to identify said labeling agent or said cell.
2030

FIG. 20

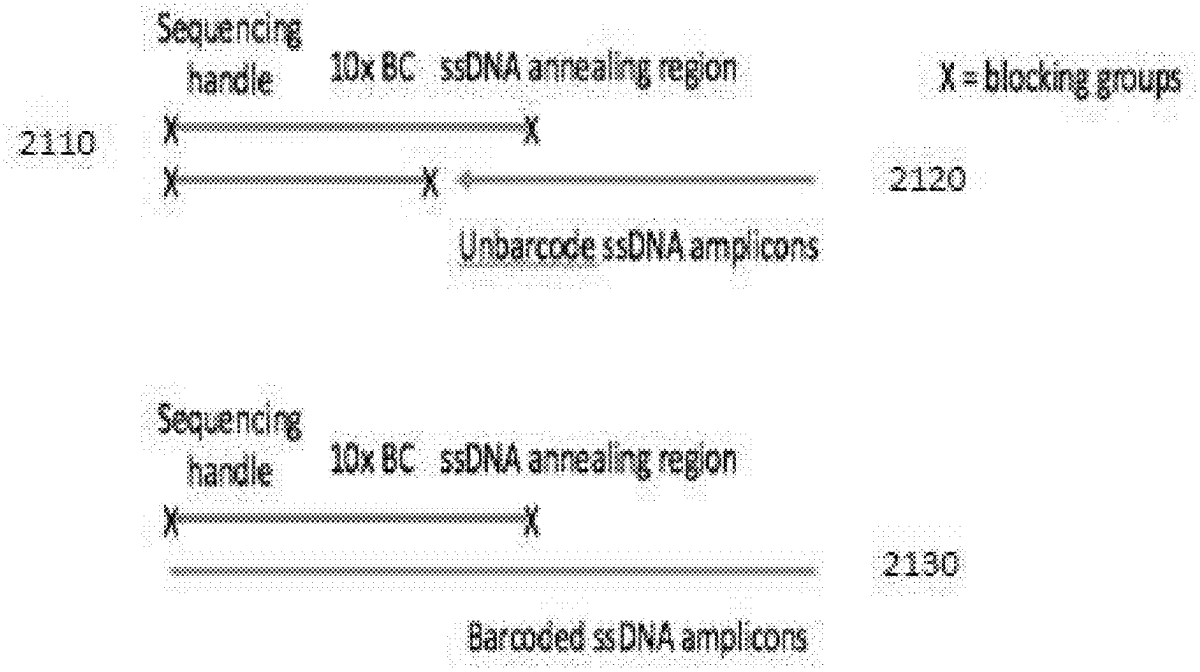

Sequencing handle
10x BC
ssDNA annealing region
X = blocking groups
2110
2120
Unbarcode ssDNA amplicons
Sequencing handle
10x BC
ssDNA annealing region
2130
Barcoded ssDNA amplicons

FIG. 21

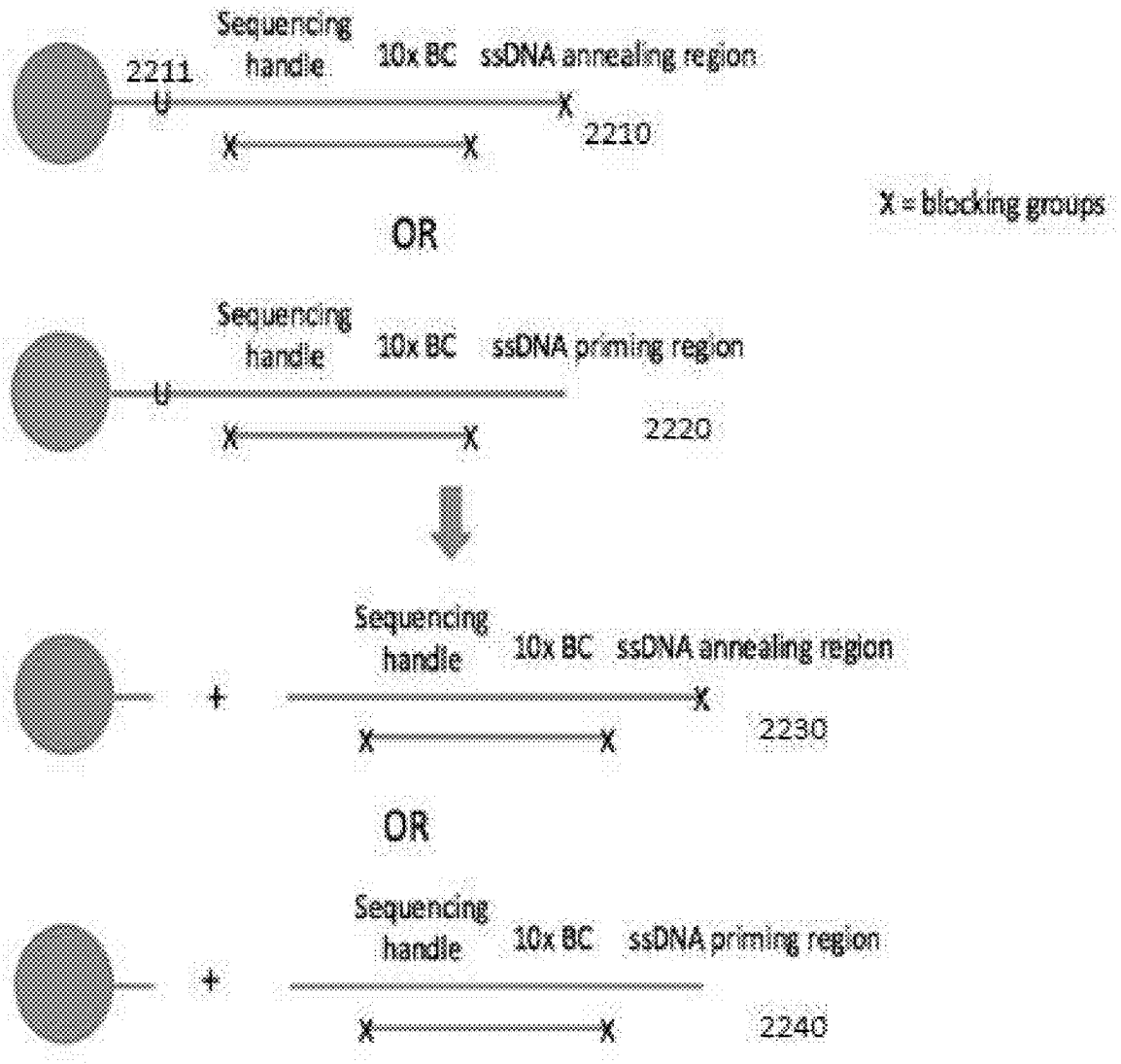

Sequencing handle
10x BC
ssDNA annealing region
2211
2230
X = blocking groups
OR
Sequencing handle
10x BC
ssDNA priming region
2220
Sequencing handle
10x BC
ssDNA annealing region
2230
OR
Sequencing handle
10x BC
ssDNA priming region
2240

FIG. 22

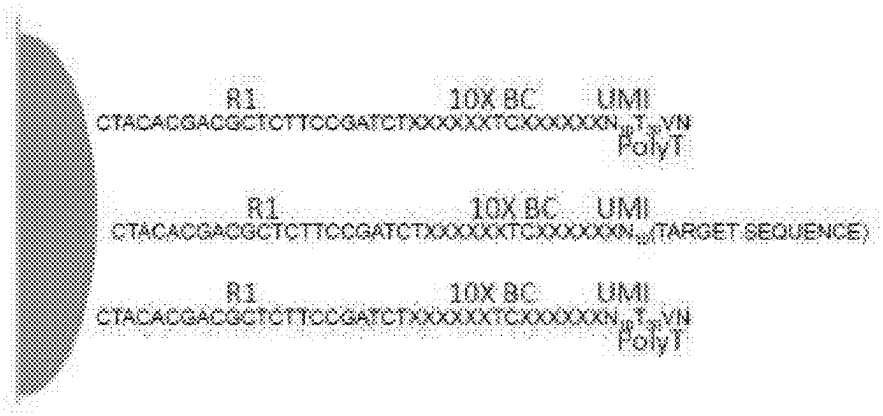


FIG. 23A

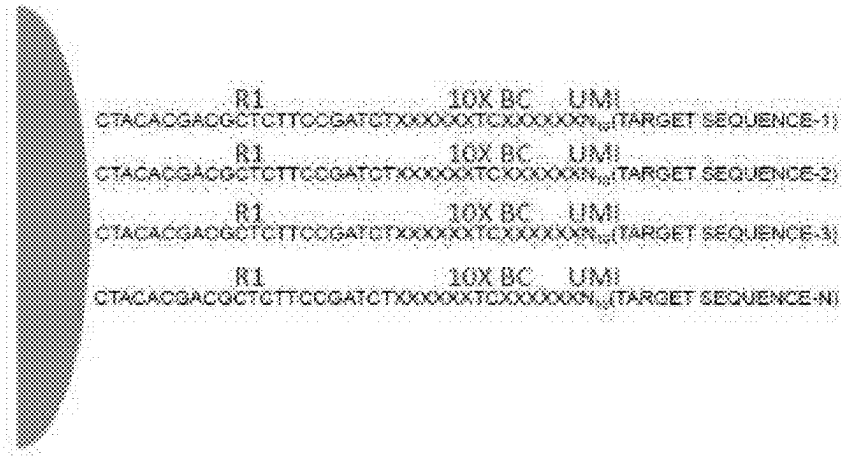


FIG. 23B

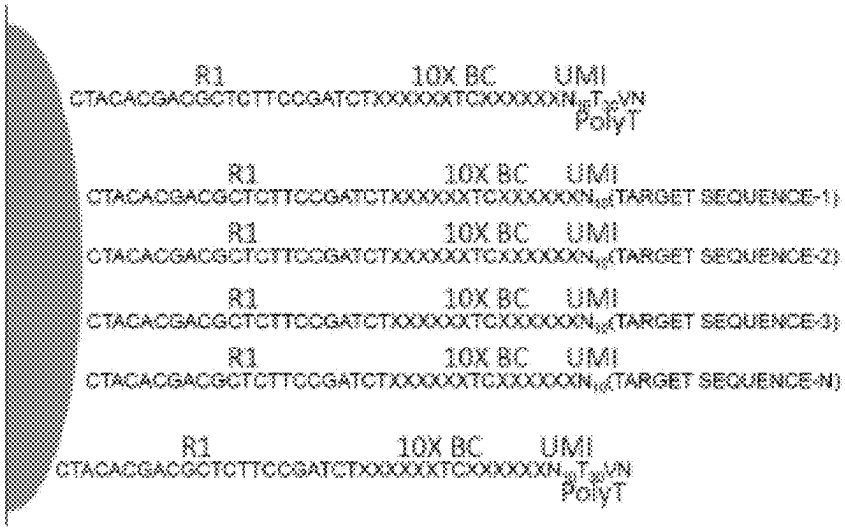


FIG. 23C

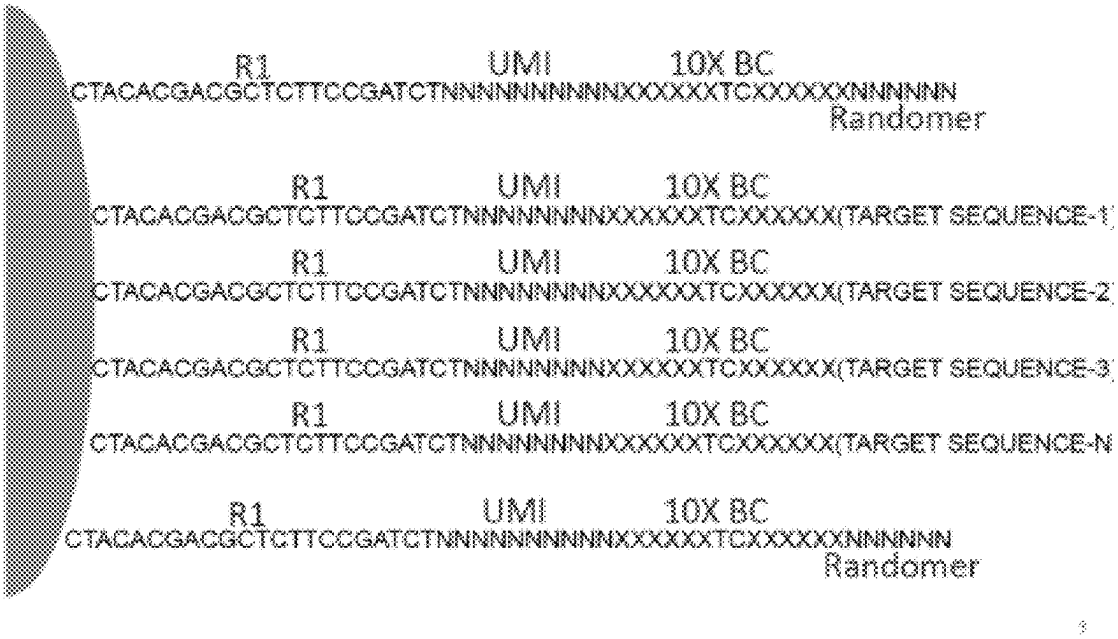

R1
UMI
10X BC
CTACACGACGCTCTTCCGATCTNNNNNNNNNXXXXXXTCXXXXXXNNNNNN
Randomer
R1
UMI
10X BC
CTACACGACGCTCTTCCGATCTNNNNNNNNXXXXXXTCXXXXXX(TARGET SEQUENCE-1)
R1
UMI
10X BC
CTACACGACGCTCTTCCGATCTNNNNNNNNXXXXXXTCXXXXXX(TARGET SEQUENCE-2)
R1
UMI
10X BC
CTACACGACGCTCTTCCGATCTNNNNNNNNXXXXXXTCXXXXXX(TARGET SEQUENCE-3)
R1
UMI
10X BC
CTACACGACGCTCTTCCGATCTNNNNNNNNXXXXXXTCXXXXXX(TARGET SEQUENCE-N)
R1
UMI
10X BC
CTACACGACGCTCTTCCGATCTNNNNNNNNNXXXXXXTCXXXXXXNNNNNN
Randomer

FIG. 24

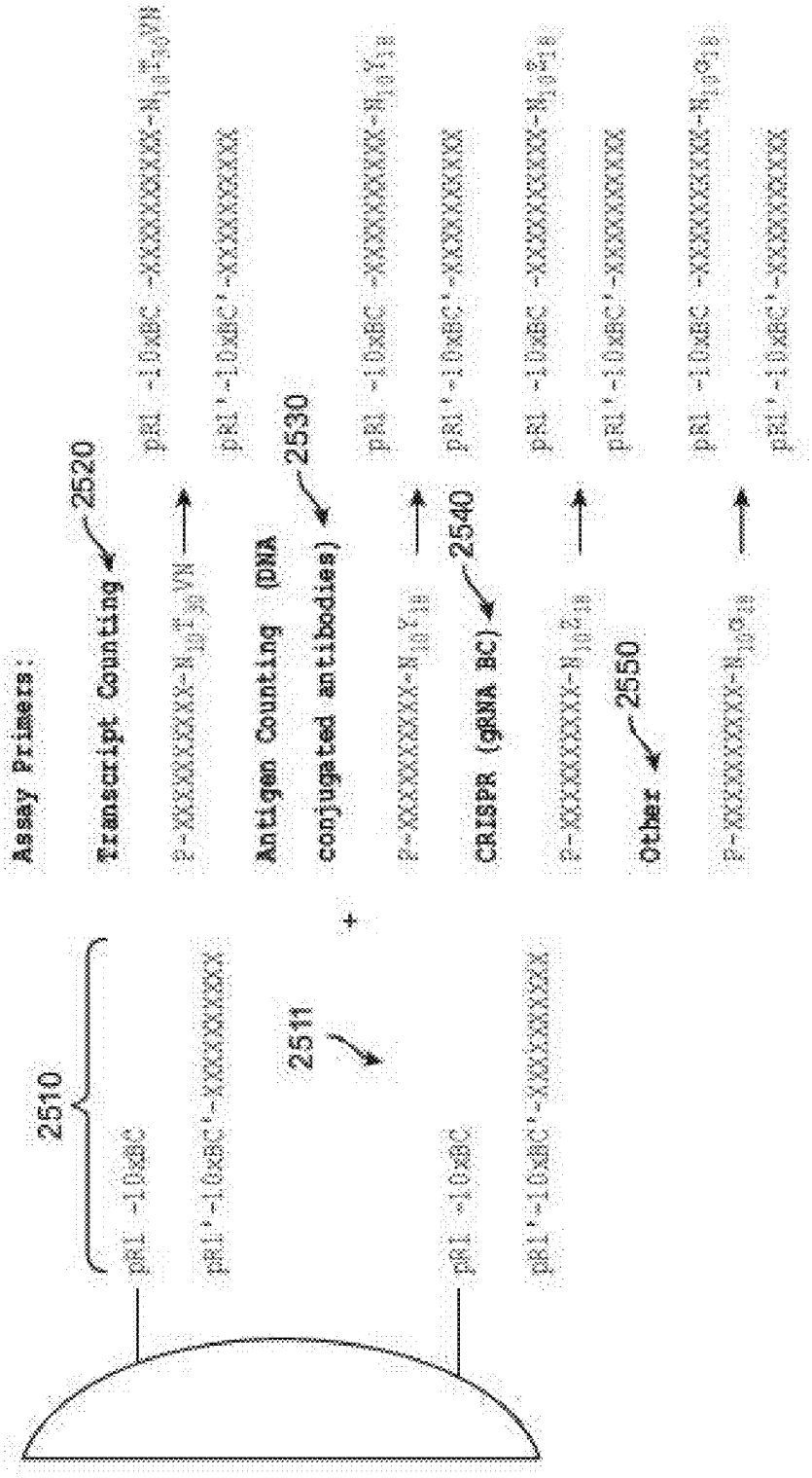

2510
pR1 -10xBC
pR1' -10xBC' -XXXXXXXXXX
2511
+
pR1 -10xBC
pR1' -10xBC' -XXXXXXXXXX
Assay Primers:
Transcript Counting
2520
Antigen Counting (DNA
conjugated antibodies)
2530
CRISPR (gRNA BC)
2540
Other
2550

FIG. 25A

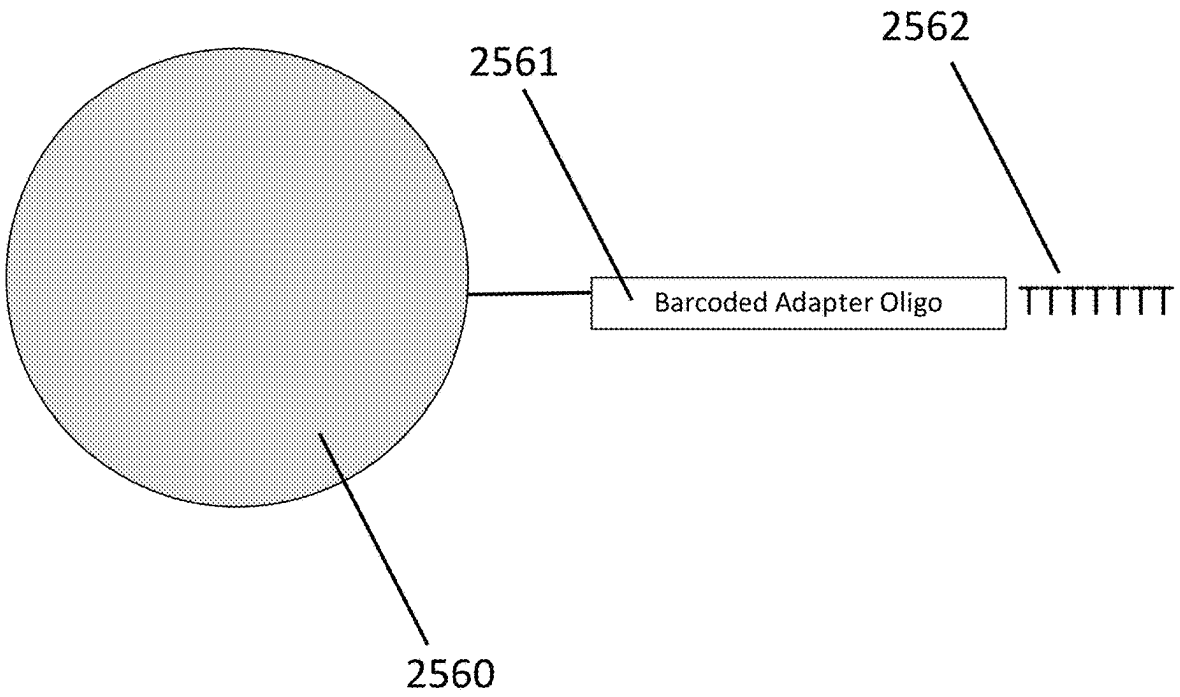


FIG. 25B

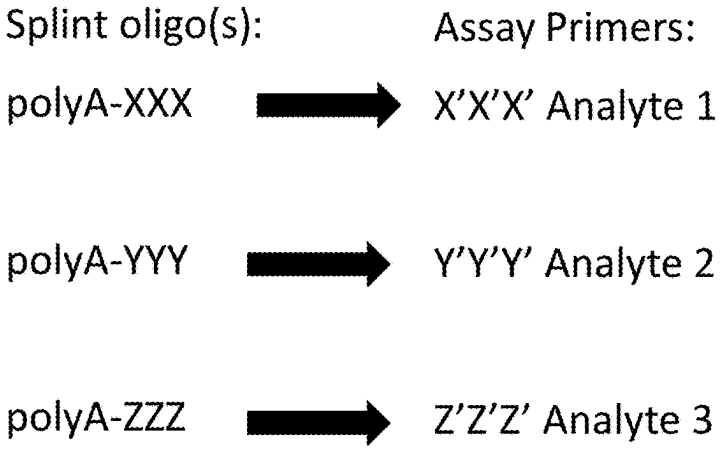
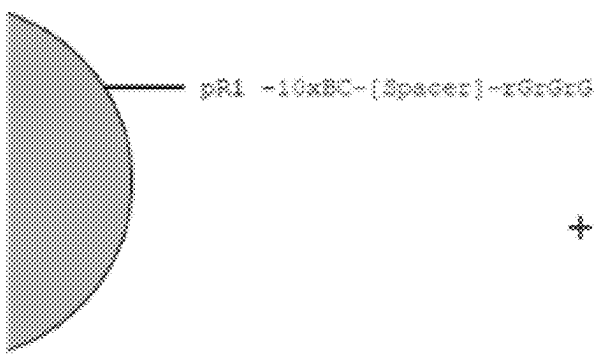


FIG. 25C



Assay Primers (in solution):

Transcript Counting

[PCR primer site]-T_{30}VN

Antigen Counting (DNA conjugated antibodies)

[PCR primer site]-$N_{10}X_{16}$

CRISPR (gRNA BC)

[PCR primer site]-$N_{10}X_{16}$

Other

[PCR primer site]-$N_{10}X_{16}$

FIG. 26

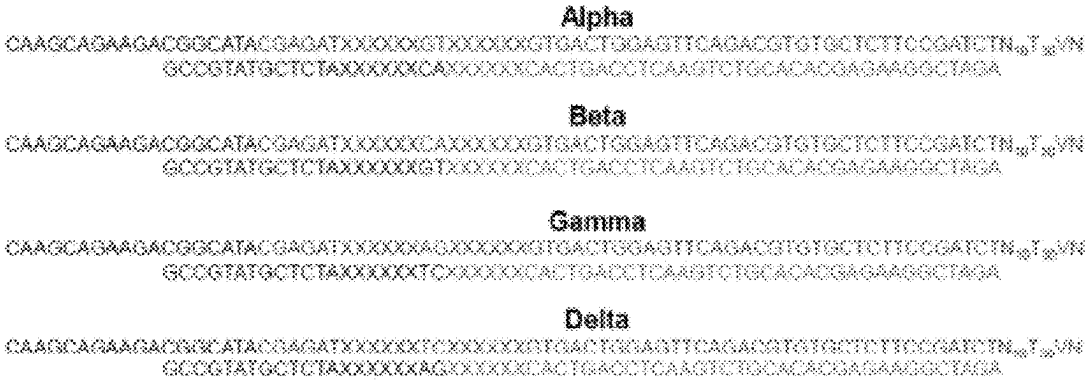

Alpha
CAAGCAGAAGACGGCATACGAGATXXXXXXGTXXXXXXGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTN10T30VN
GCCGTATGCTCTAXXXXXXCAXXXXXXCACTGACCTCAAGTCTGCACACGAGAAGGCTAGA
Beta
CAAGCAGAAGACGGCATACGAGATXXXXXXCAXXXXXXGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTN10T30VN
GCCGTATGCTCTAXXXXXXGTXXXXXXCACTGACCTCAAGTCTGCACACGAGAAGGCTAGA
Gamma
CAAGCAGAAGACGGCATACGAGATXXXXXXAGXXXXXXGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTN10T30VN
GCCGTATGCTCTAXXXXXXTCXXXXXXCACTGACCTCAAGTCTGCACACGAGAAGGCTAGA
Delta
CAAGCAGAAGACGGCATACGAGATXXXXXXTCXXXXXXGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTN10T30VN
GCCGTATGCTCTAXXXXXXAGXXXXXXCACTGACCTCAAGTCTGCACACGAGAAGGCTAGA
P7
10X BC
R2
UMI
CAAGCAGAAGACGGCATACGAGATXXXXXXGTXXXXXXGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTN10T30VN
PolyT

FIG. 27A

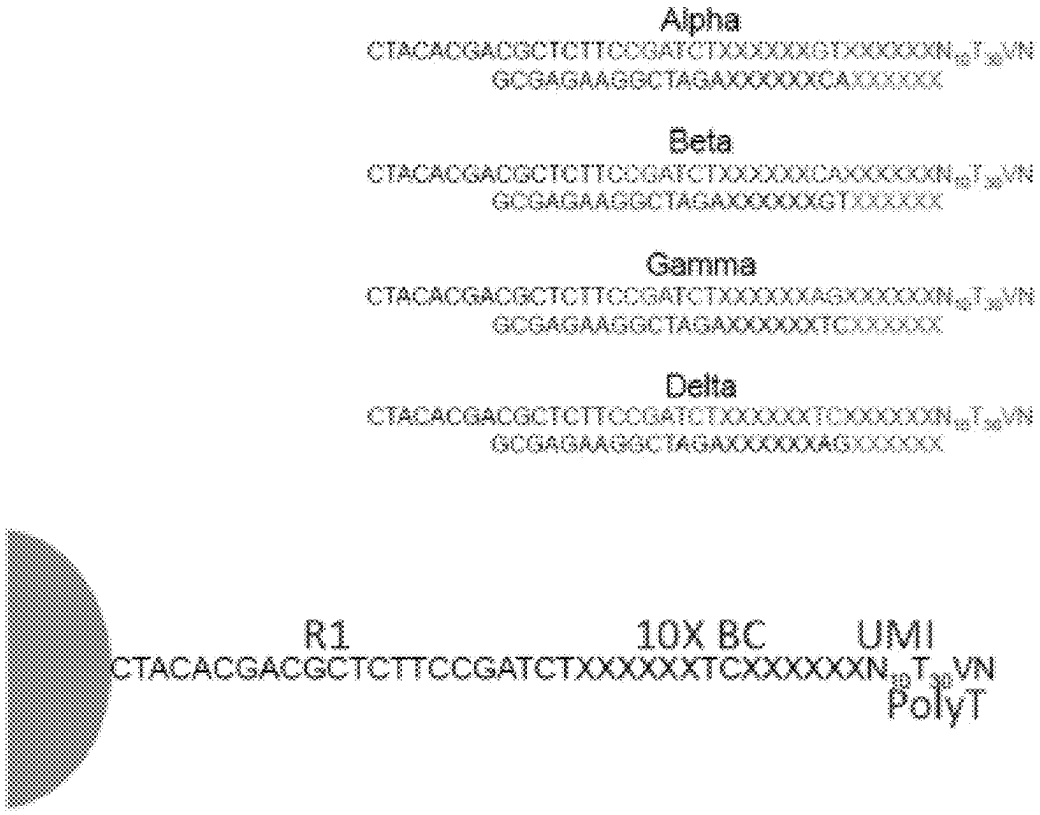

Alpha
CTACACGACGCTCTTCCGATCTXXXXXXGTXXXXXXN18T30VN
GCGAGAAGGCTAGAXXXXXXCAXXXXXX
Beta
CTACACGACGCTCTTCCGATCTXXXXXXCAXXXXXXN18T30VN
GCGAGAAGGCTAGAXXXXXXGTXXXXXX
Gamma
CTACACGACGCTCTTCCGATCTXXXXXXAGXXXXXXN18T30VN
GCGAGAAGGCTAGAXXXXXXTCXXXXXX
Delta
CTACACGACGCTCTTCCGATCTXXXXXXTCXXXXXXN18T30VN
GCGAGAAGGCTAGAXXXXXXAGXXXXXX
R1
10X BC
UMI
CTACACGACGCTCTTCCGATCTXXXXXXTCXXXXXXN18T30VN
PolyT

FIG. 27B

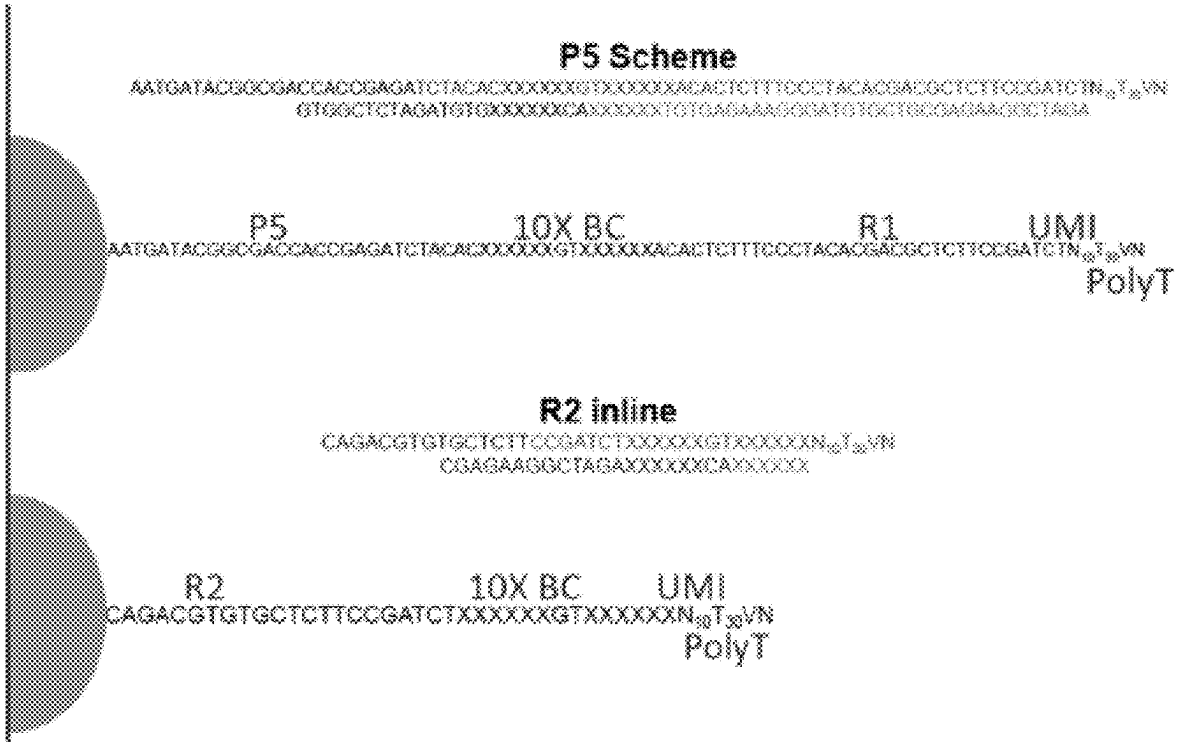


FIG. 27C

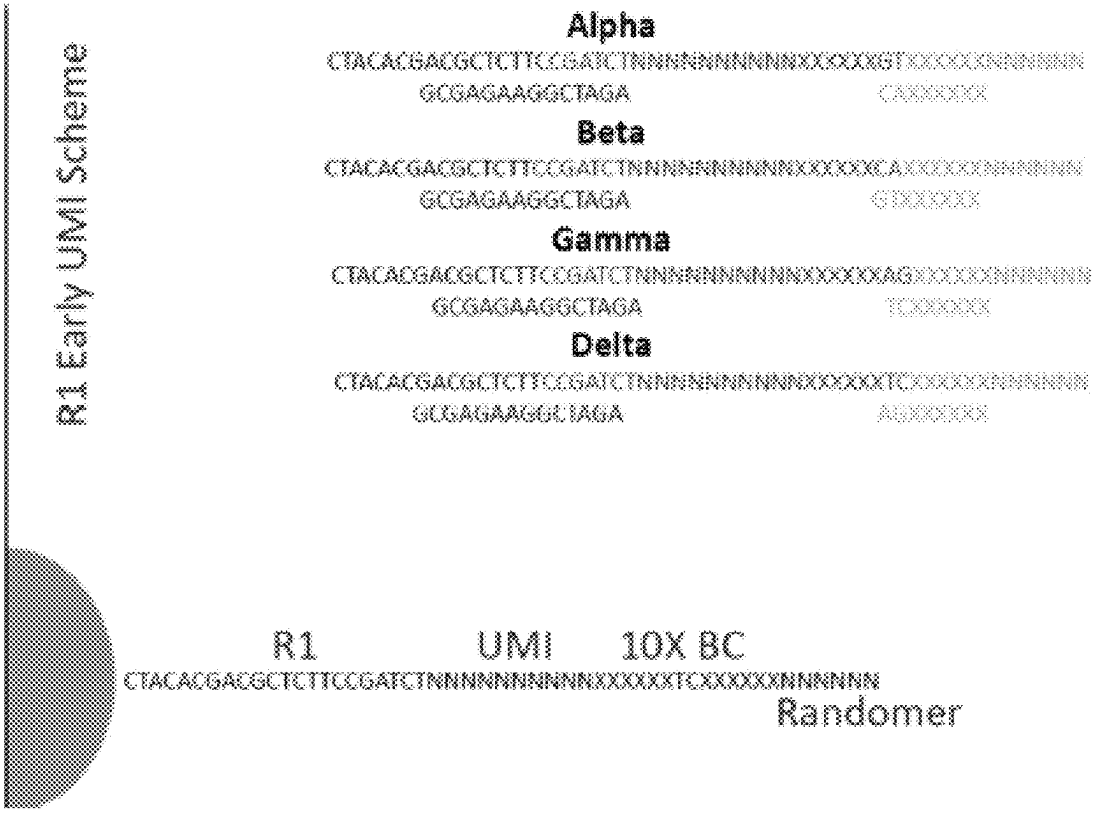

R1 Early UMI Scheme
Alpha
CTACACGACGCTCTTCCGATCTNNNNNNNNNNXXXXXGTXXXXXXNNNNNN
GCGAGAAGGCTAGA
CXXXXXX
Beta
CTACACGACGCTCTTCCGATCTNNNNNNNNNNXXXXXXCAXXXXXXNNNNNN
GCGAGAAGGCTAGA
GTXXXXXX
Gamma
CTACACGACGCTCTTCCGATCTNNNNNNNNNNXXXXXXAGXXXXXXNNNNNN
GCGAGAAGGCTAGA
TCXXXXXX
Delta
CTACACGACGCTCTTCCGATCTNNNNNNNNNNXXXXXXTCXXXXXXNNNNNN
GCGAGAAGGCTAGA
AGXXXXXX
R1
UMI
10X BC
CTACACGACGCTCTTCCGATCTNNNNNNNNNNXXXXXXTCXXXXXXNNNNNN
Randomer

FIG. 27D

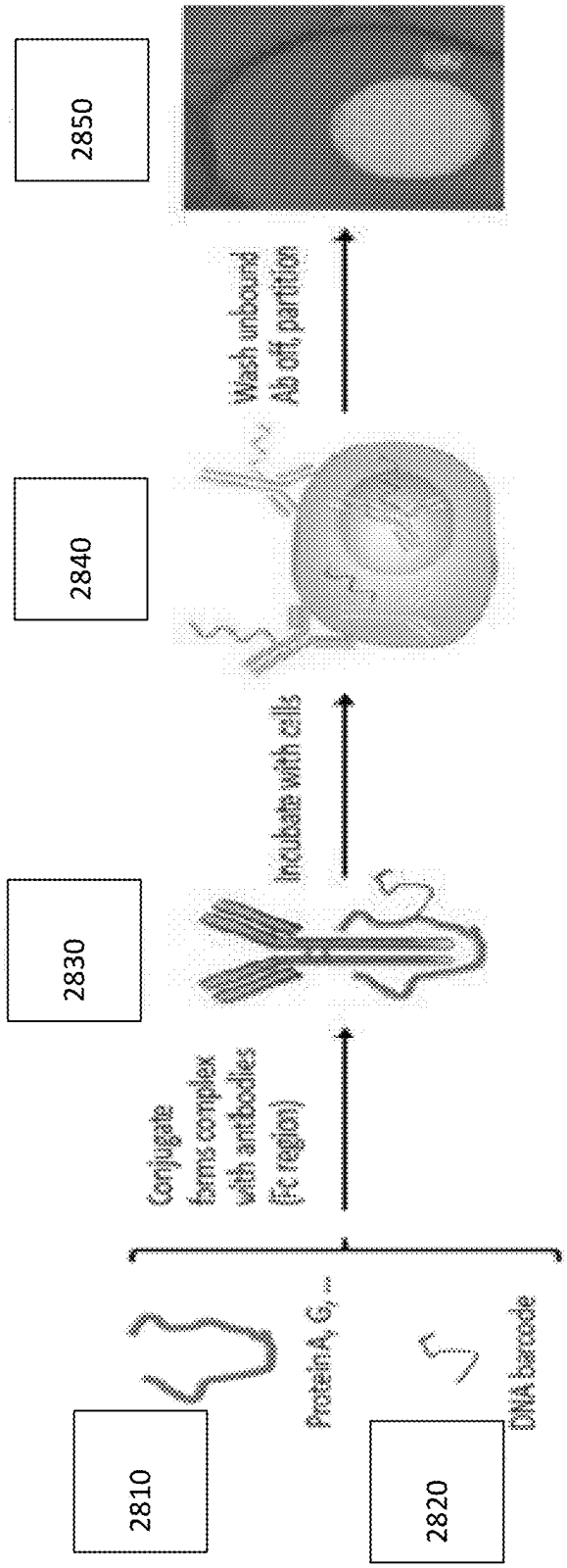

2810
2820
Protein A, G, ...
DNA barcode
Conjugate forms complex with antibodies (Fc region)
2830
Incubate with cells
2840
Wash unbound Ab off, partition
2850

Figure 28

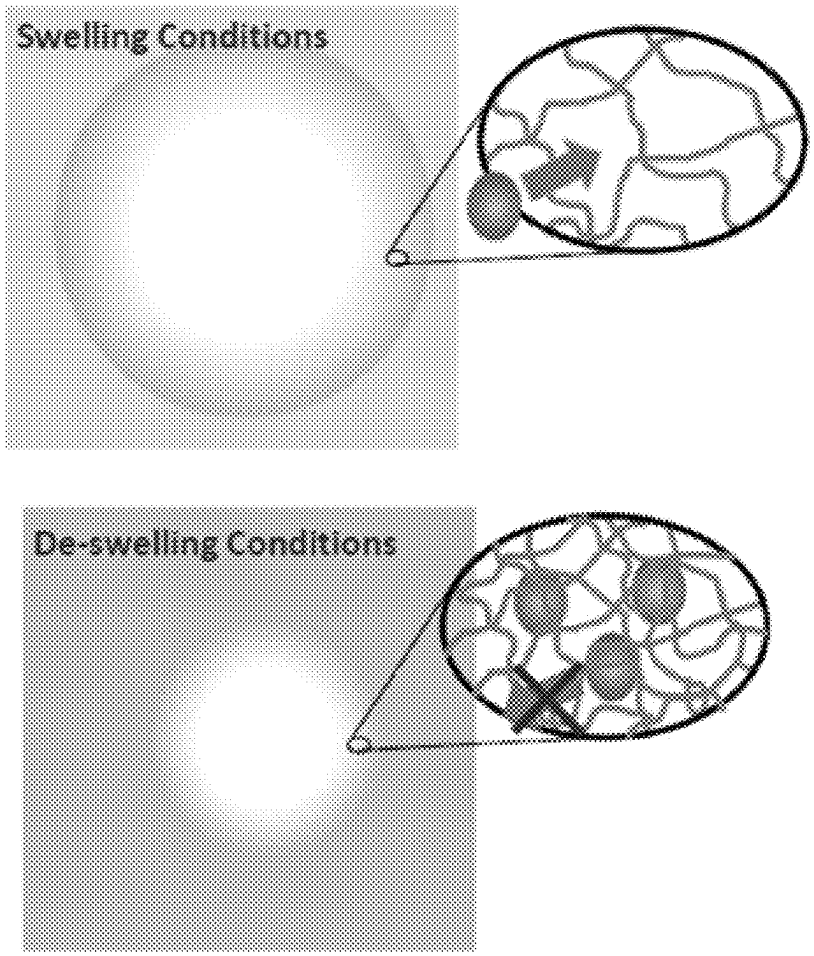


FIG. 29

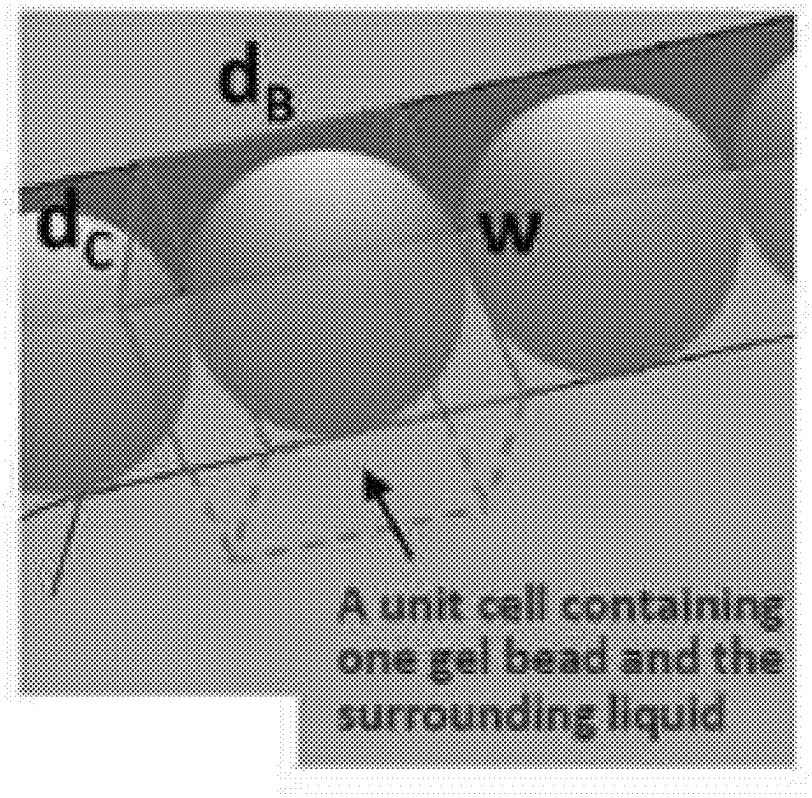

d_B
d_C
w
A unit cell containing one gel bead and the surrounding liquid

FIG. 30

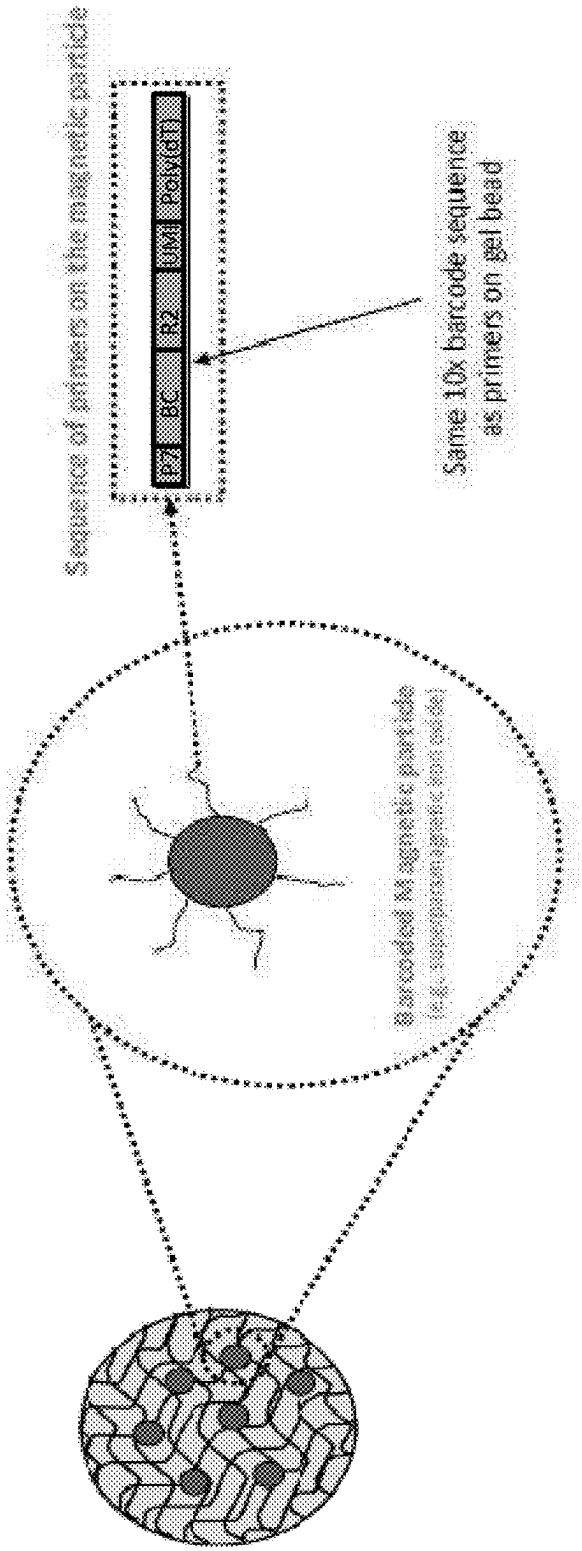

Sequence of primers on the magnetic particle
P7
BC
R2
UMI
PolydT
Same 10x barcode sequence as primers on gel bead
Barcoded Magnetic particle
(e.g. oligonucleotide coated)

FIG. 31

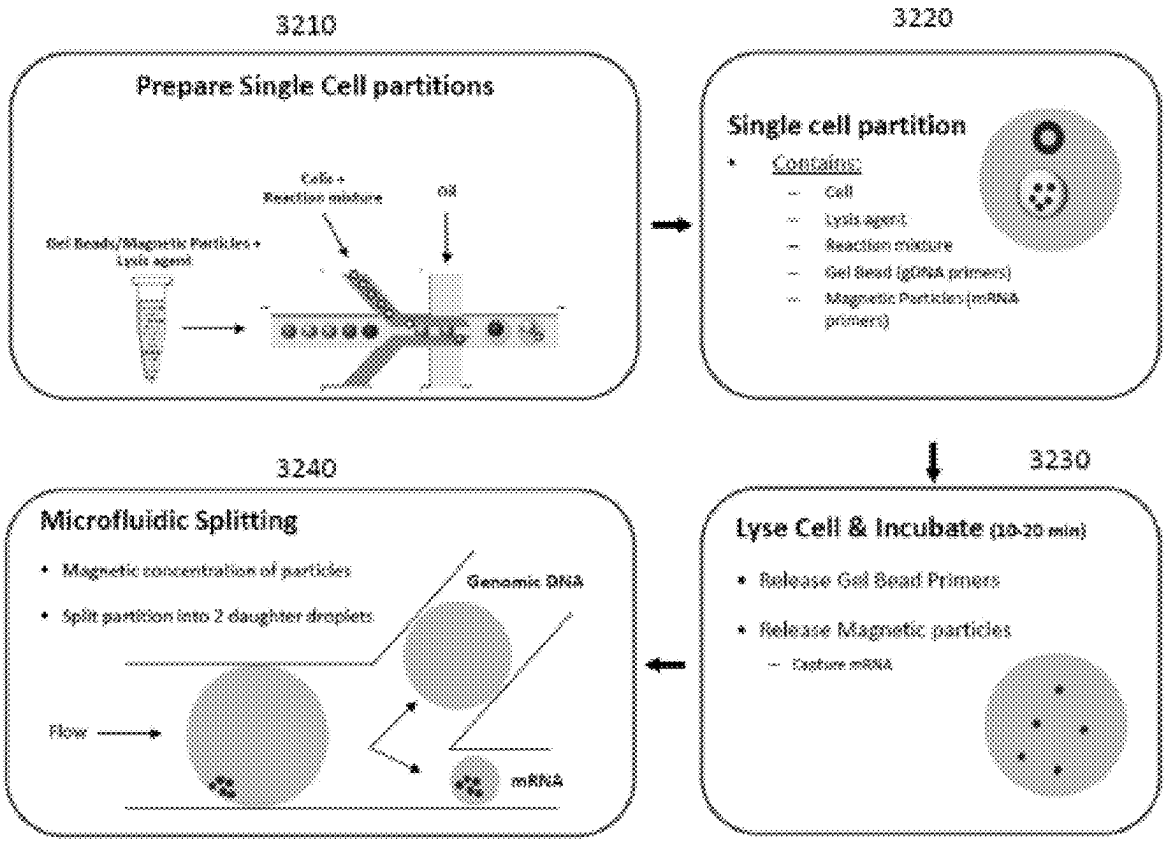

3210
Prepare Single Cell partitions
Cells + Reaction mixture
Oil
Gel Beads/Magnetic Particles + Lysis agent
3220
Single cell partition
Contains:
Cell
Lysis agent
Reaction mixture
Gel Bead (gDNA primers)
Magnetic Particles (mRNA primers)
3230
Lyse Cell & Incubate (10-20 min)
Release Gel Bead Primers
Release Magnetic particles
Capture mRNA
3240
Microfluidic Splitting
Magnetic concentration of particles
Split partition into 2 daughter droplets
Genomic DNA
Flow
mRNA

FIG. 32

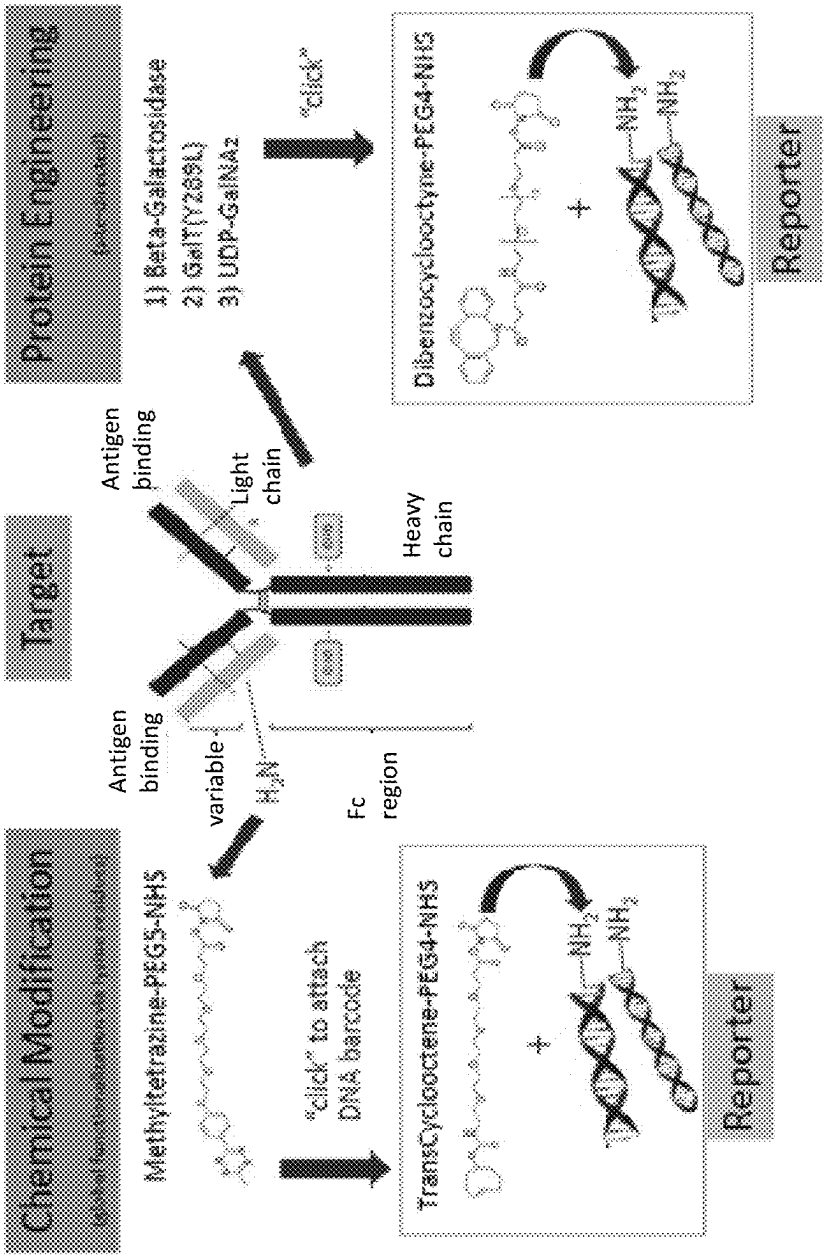

Chemical Modification
Protein Engineering
Target
Methyltetrazine-PEG5-NHS
"click" to attach DNA barcode
TransCyclooctene-PEG4-NHS
Reporter
Antigen binding
variable
H_2N
Fc region
Antigen binding
Light chain
Heavy chain
1) Beta-Galactosidase
2) GalT(Y289L)
3) UDP-GalNAz
"click"
Dibenzocyclooctyne-PEG4-NHS
Reporter

FIG. 33

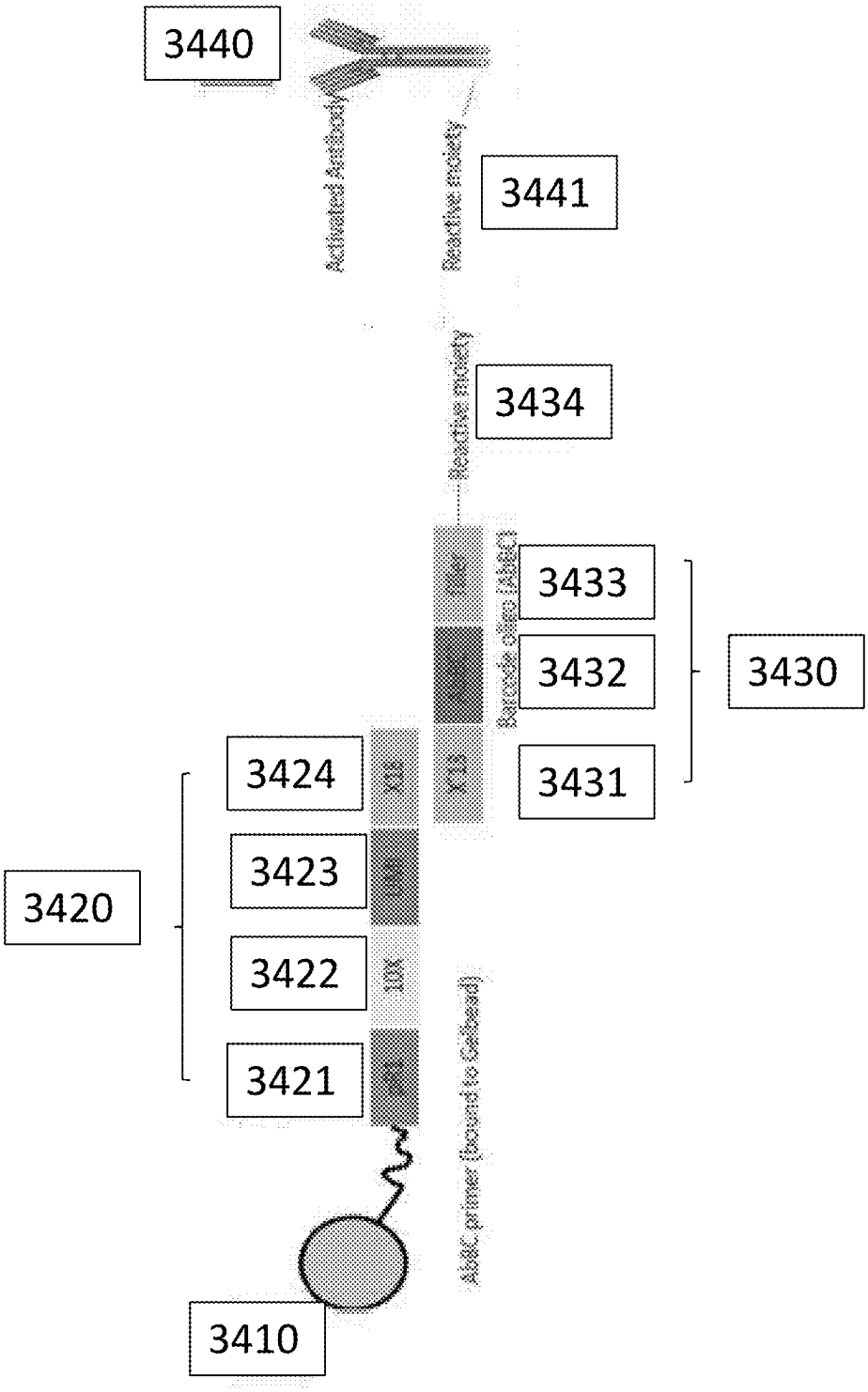

3440
Activated Antibody
Reactive moiety
3441
Reactive moiety
3434
Barcode oligo (ssDNA)
3433
3432
3431
3430
3424
3423
3422
3421
3420
AbMC primer (bound to Gelbead)
3410

FIG. 34

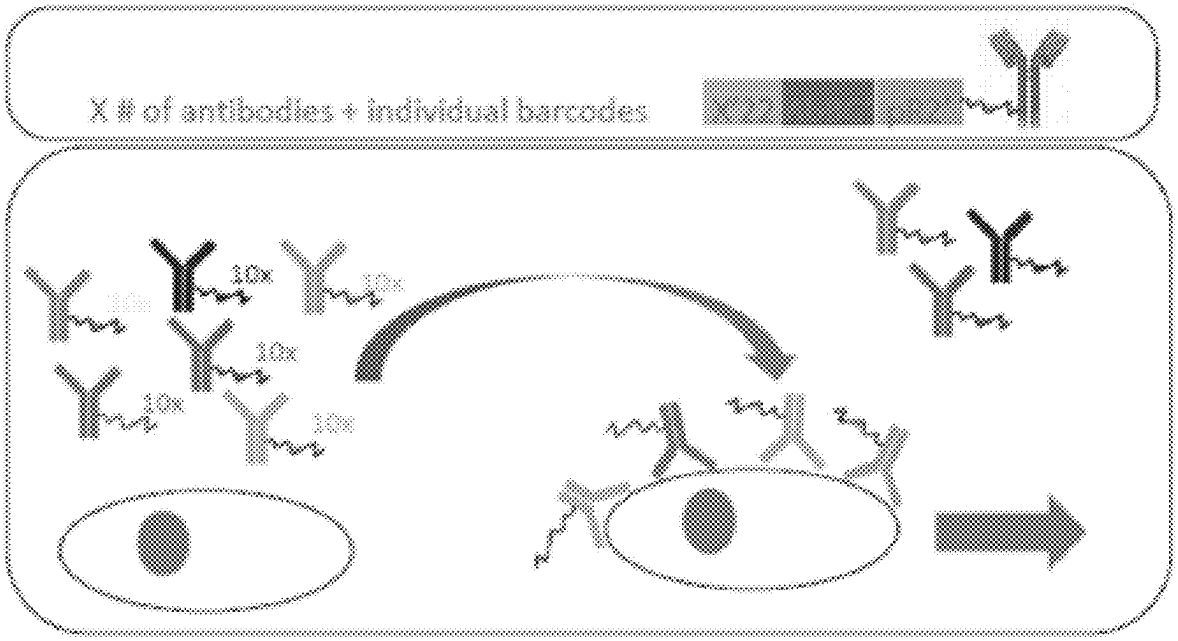


FIG. 35A

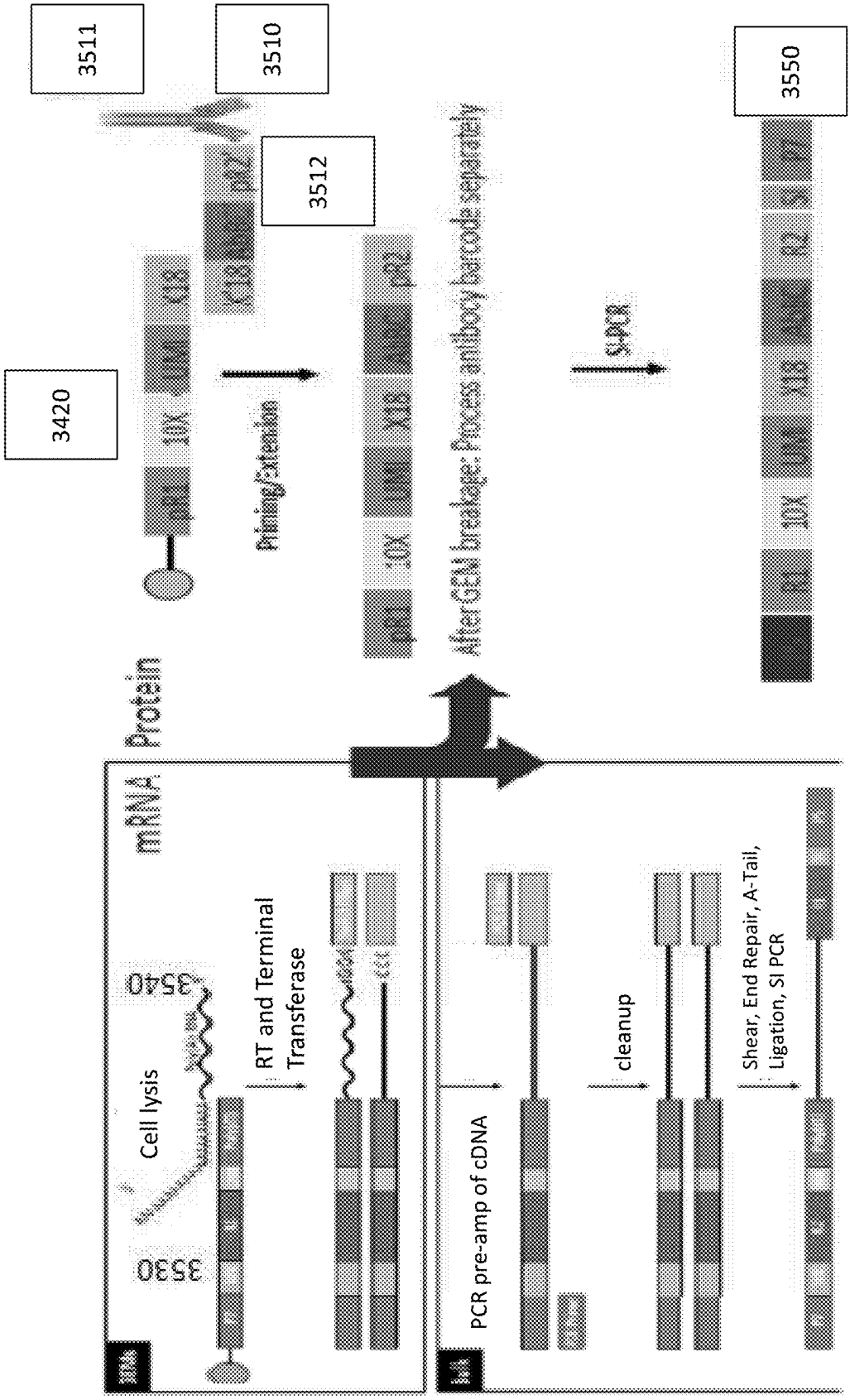

3511
3510
3512
3420
3550
Protein
mRNA
3540
3530
Cell lysis
RT and Terminal Transferase
Priming/Extension
After GEM breakage: Process antibody barcode separately
SI-PCR
PCR pre-amp of cDNA
cleanup
Shear, End Repair, A-Tail, Ligation, SI PCR

FIG. 35B

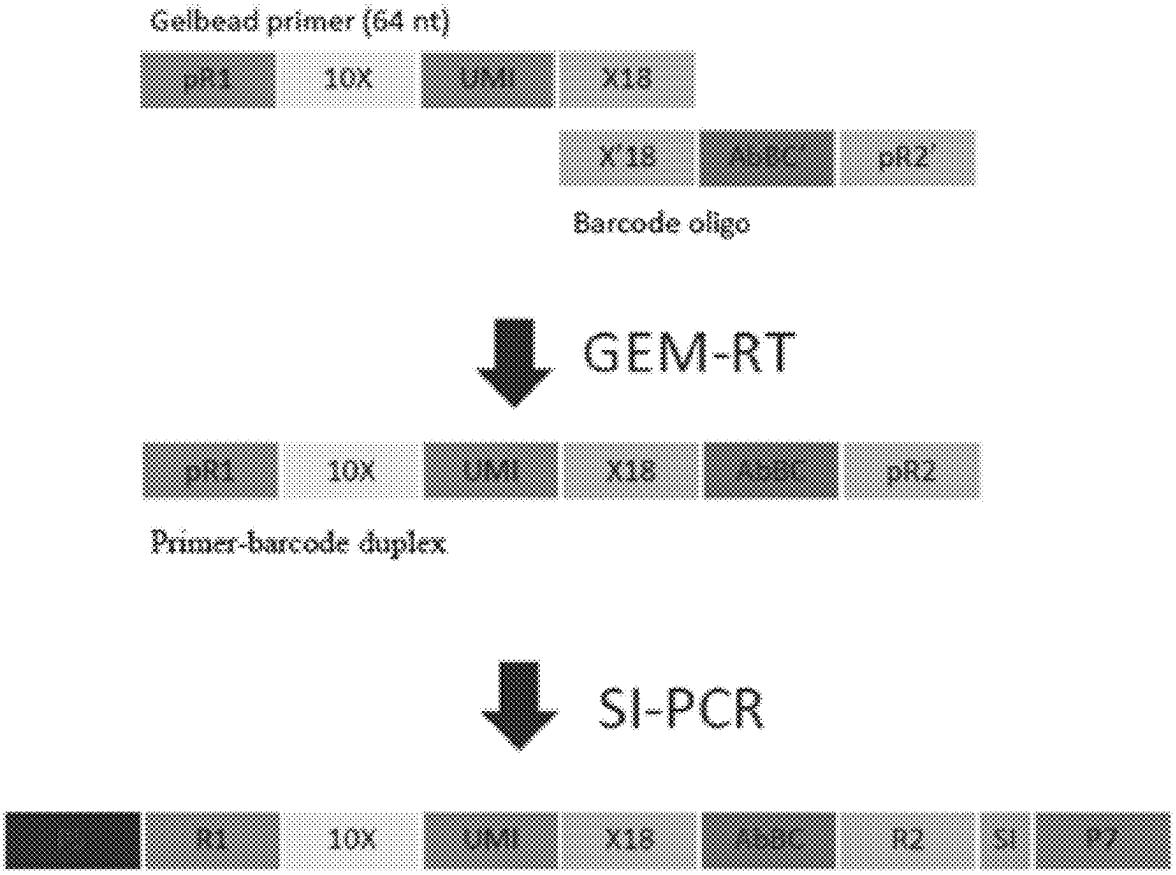

Gelbead primer (64 nt)
pR1
10X
UMI
X16
X'16
pR2'
Barcode oligo
GEM-RT
pR1
10X
UMI
X16
pR2
Primer-barcode duplex
SI-PCR
R1
10X
UMI
X16
R2
SI
R2

FIG. 35C

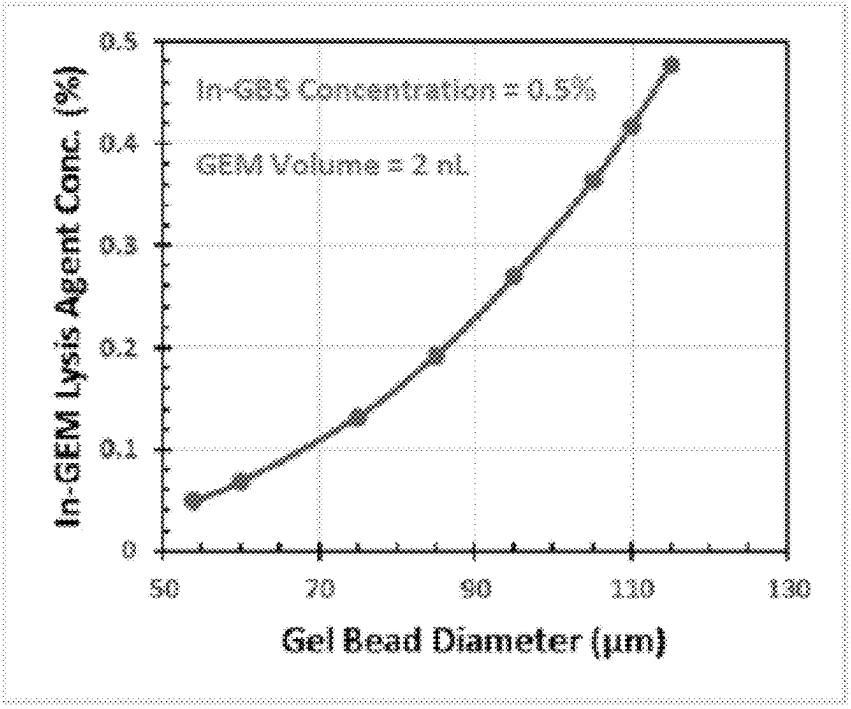


FIG. 36A

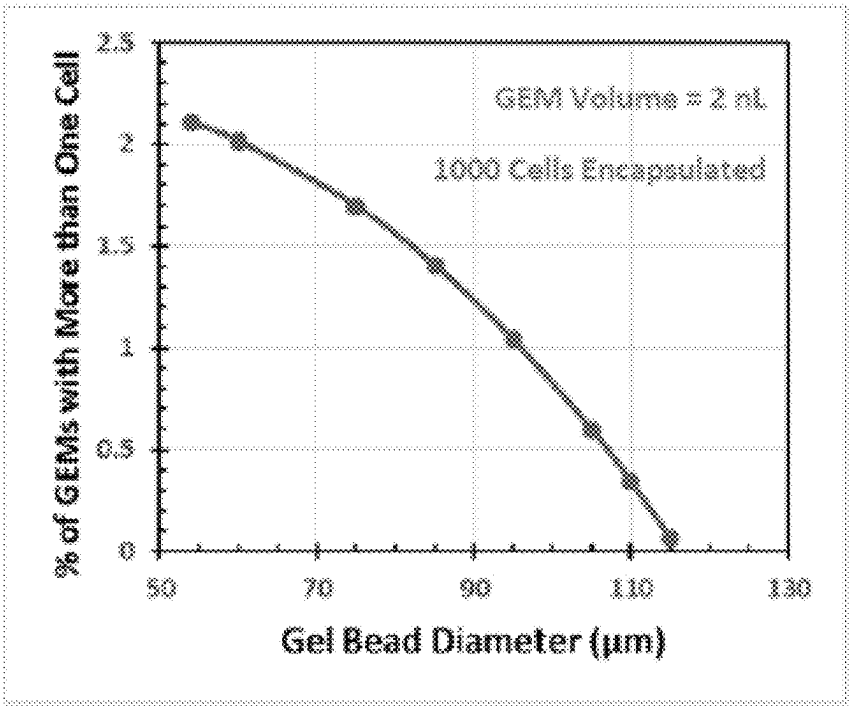


FIG. 36B

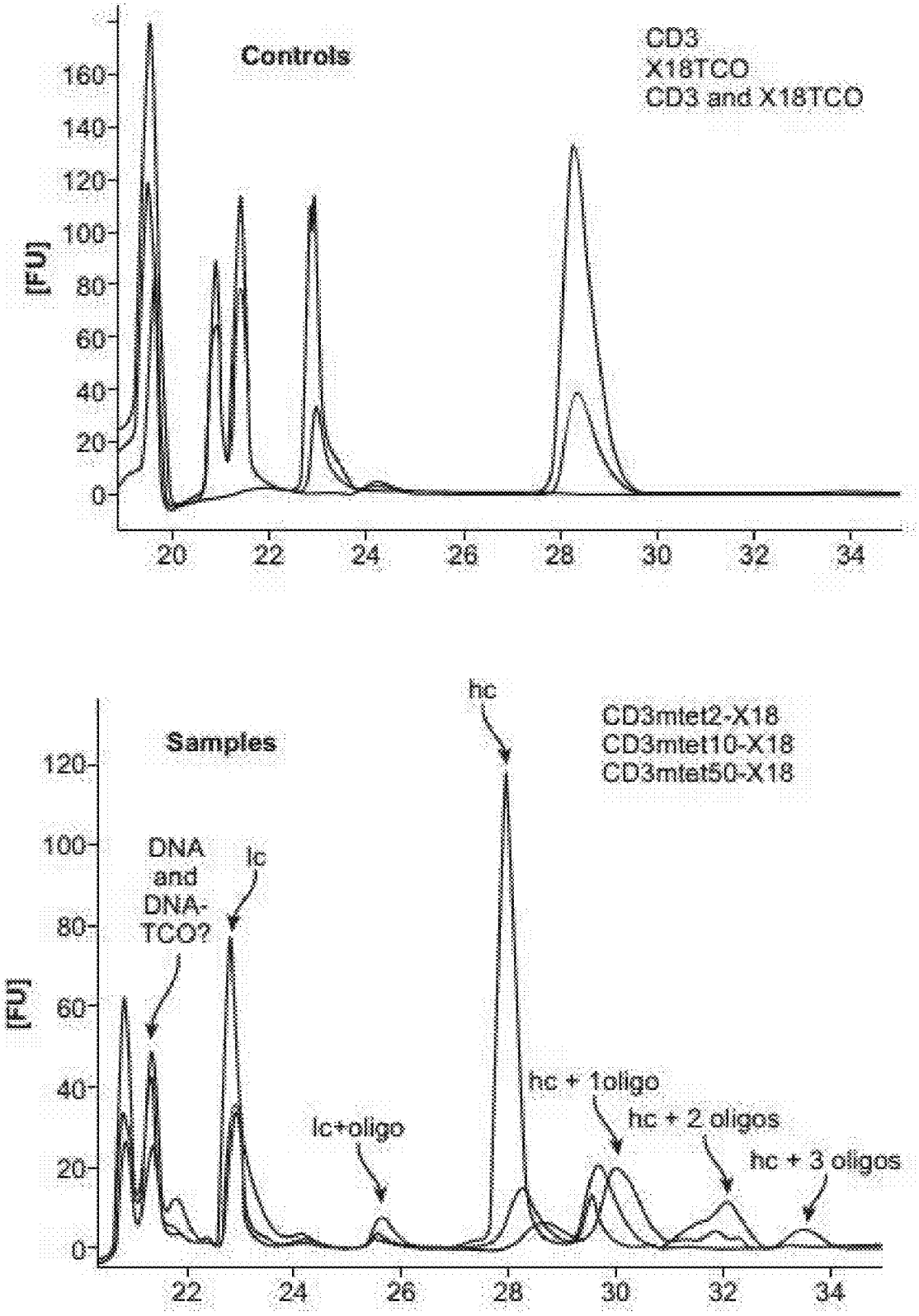

Controls
CD3
X18TCO
CD3 and X18TCO
[FU]
160
140
120
100
80
60
40
20
0
20
22
24
26
28
30
32
34
Samples
hc
CD3mtet2-X18
CD3mtet10-X18
CD3mtet50-X18
DNA
and
DNA-
TCO?
lc
lc+oligo
hc + 1oligo
hc + 2 oligos
hc + 3 oligos
[FU]
120
100
80
60
40
20
0
22
24
26
28
30
32
34

Figure 37

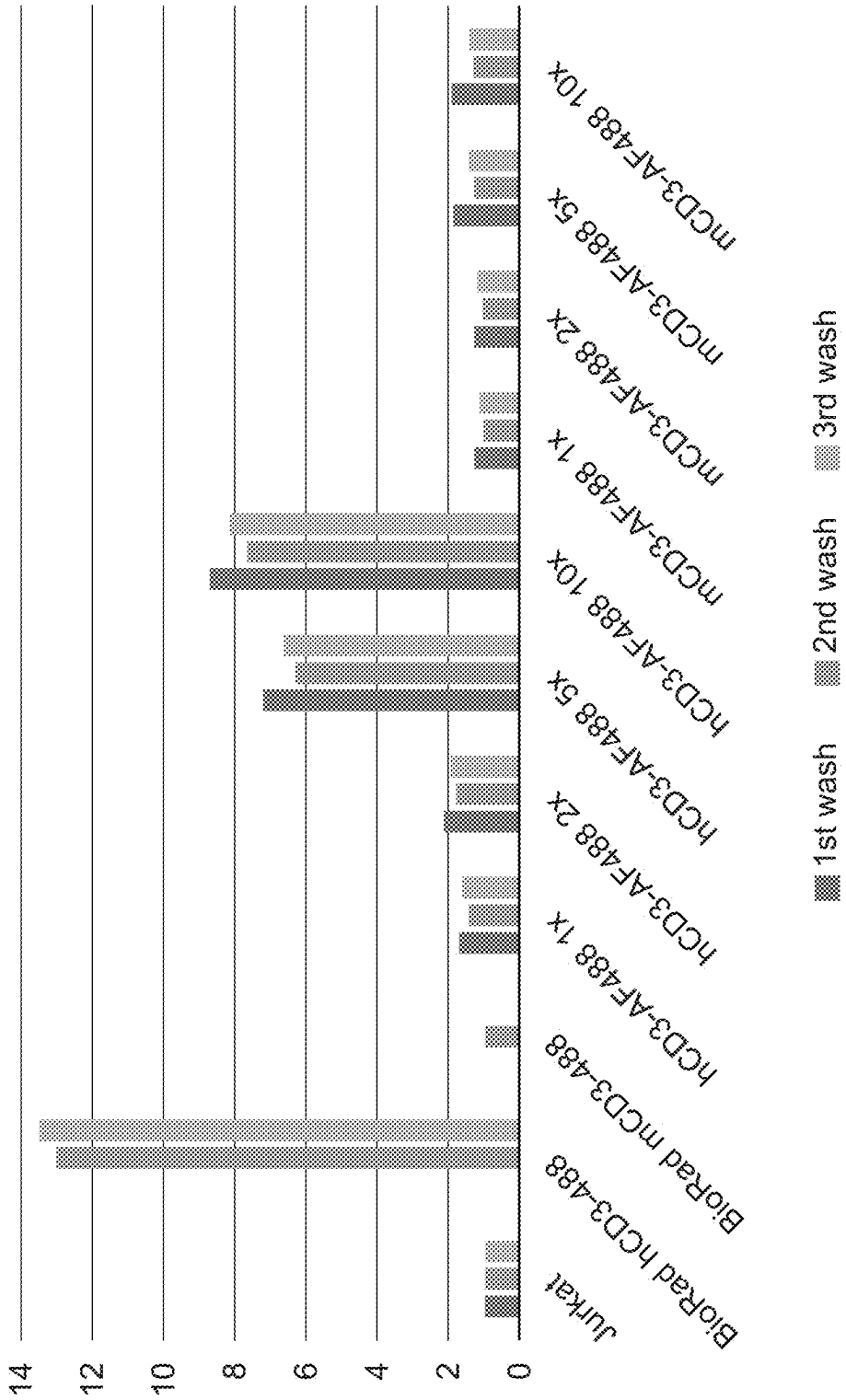

14
12
10
8
6
4
2
0
Jurkat
BioRad hCD3-488
BioRad mCD3-488
hCD3-AF488 1x
hCD3-AF488 2x
hCD3-AF488 5x
hCD3-AF488 10x
mCD3-AF488 1x
mCD3-AF488 2x
mCD3-AF488 5x
mCD3-AF488 10x
1st wash
2nd wash
3rd wash

FIG. 38

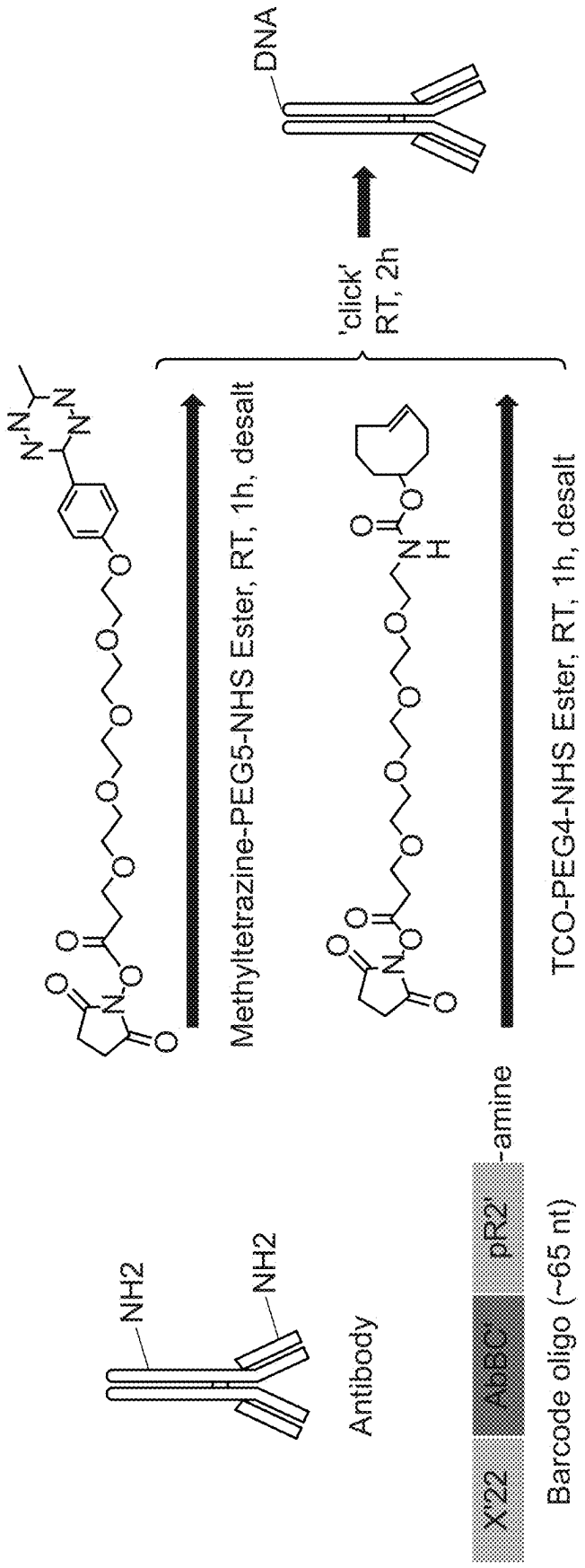

DNA
'click'
RT, 2h
Methyltetrazine-PEG5-NHS Ester, RT, 1h, desalt
TCO-PEG4-NHS Ester, RT, 1h, desalt
NH2
NH2
Antibody
X'22
AnBC
pR2'
-amine
Barcode oligo (~65 nt)

FIG. 39A

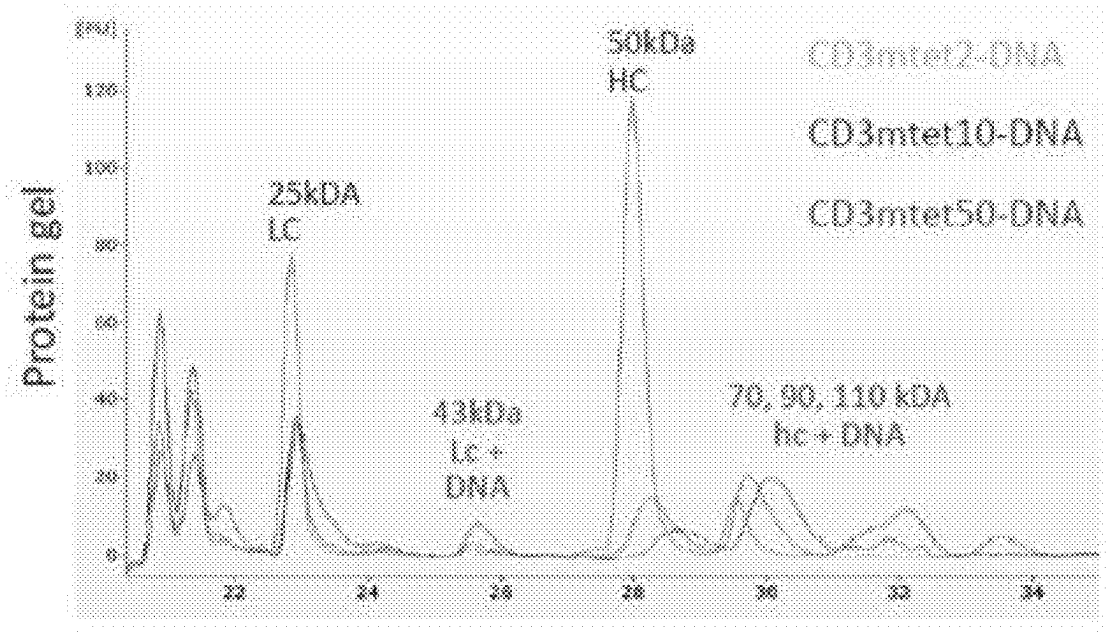

Protein gel
25kDA
LC
43kDa
Lc +
DNA
50kDa
HC
70, 90, 110 kDA
hc + DNA
CD3mtet2-DNA
CD3mtet10-DNA
CD3mtet50-DNA

FIG. 39B

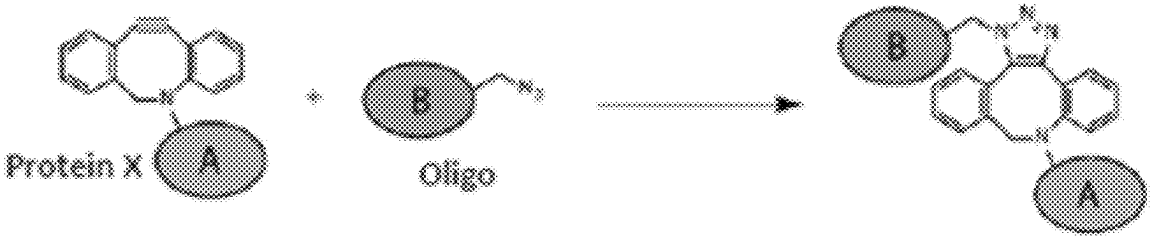


FIG. 40

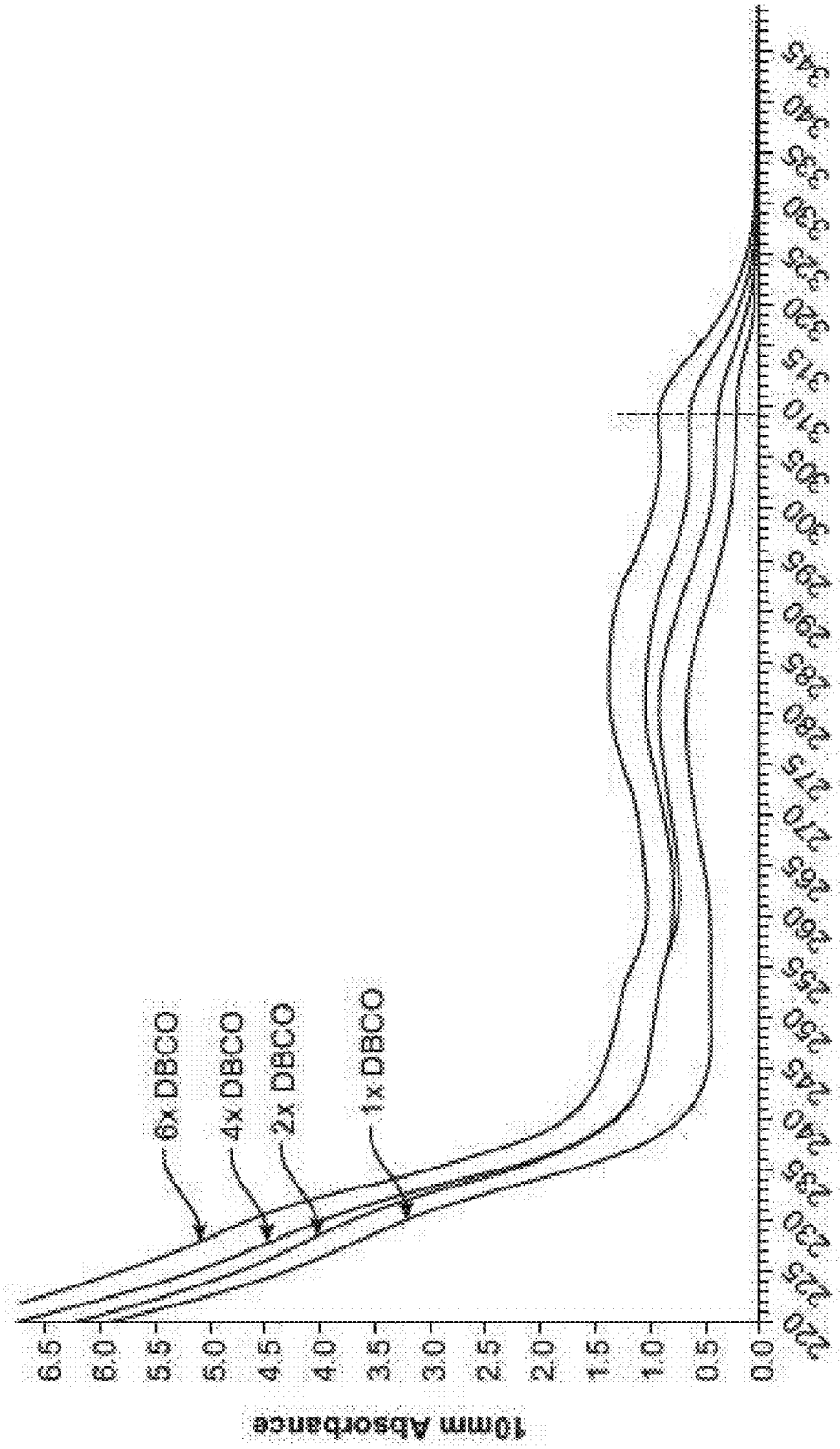

6x DBCO
4x DBCO
2x DBCO
1x DBCO
10mm Absorbance
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
220
225
230
235
240
245
250
255
260
265
270
275
280
285
290
295
300
305
310
315
320
325
330
335
340
345

Figure 41

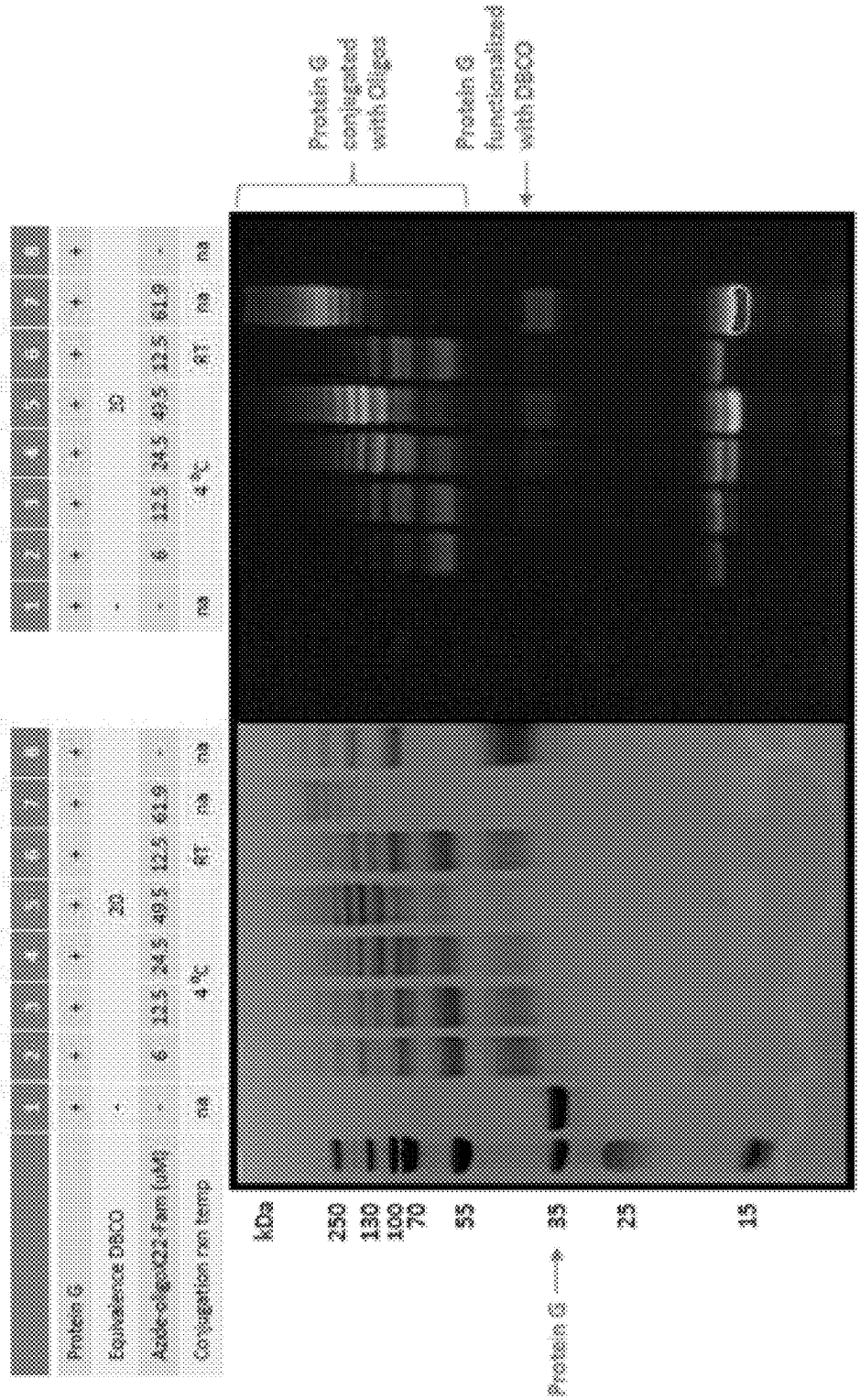

Protein G
Equivalence DBCO
Azide-oligo(dC)-Fam (uM)
Conjugation run temp
4°C
RT
kDa
250
130
100
70
55
35
25
15
Protein G
Protein G conjugated with Oligos
Protein G functionalized with DBCO

FIG. 42

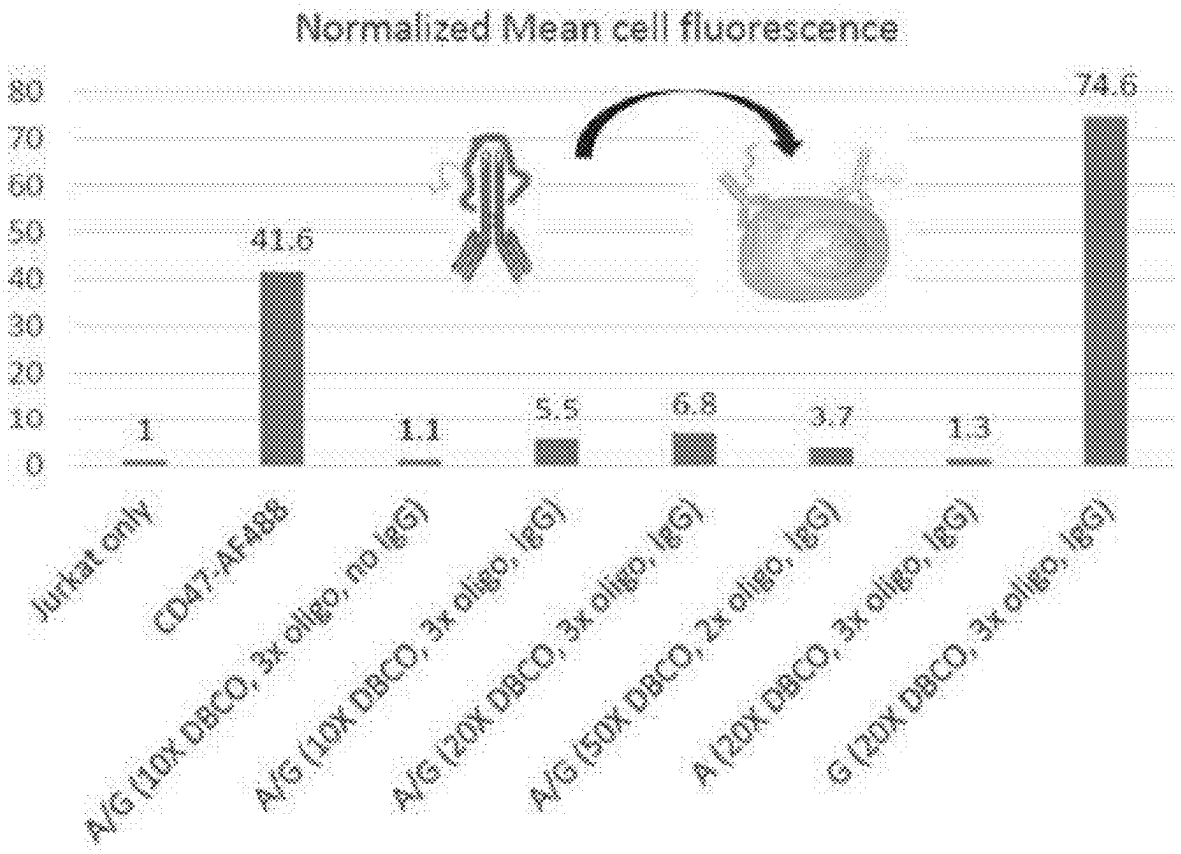

Normalized Mean cell fluorescence
80
70
60
50
40
30
20
10
0
1
41.6
1.1
5.5
6.8
3.7
1.3
74.6
Jurkat only
CD47-AF488
A/G (10X DBCO, 3x oligo, no IgG)
A/G (10X DBCO, 3x oligo, IgG)
A/G (20X DBCO, 3x oligo, IgG)
A/G (50X DBCO, 2x oligo, IgG)
A (20X DBCO, 3x oligo, IgG)
G (20X DBCO, 3x oligo, IgG)

FIG. 43

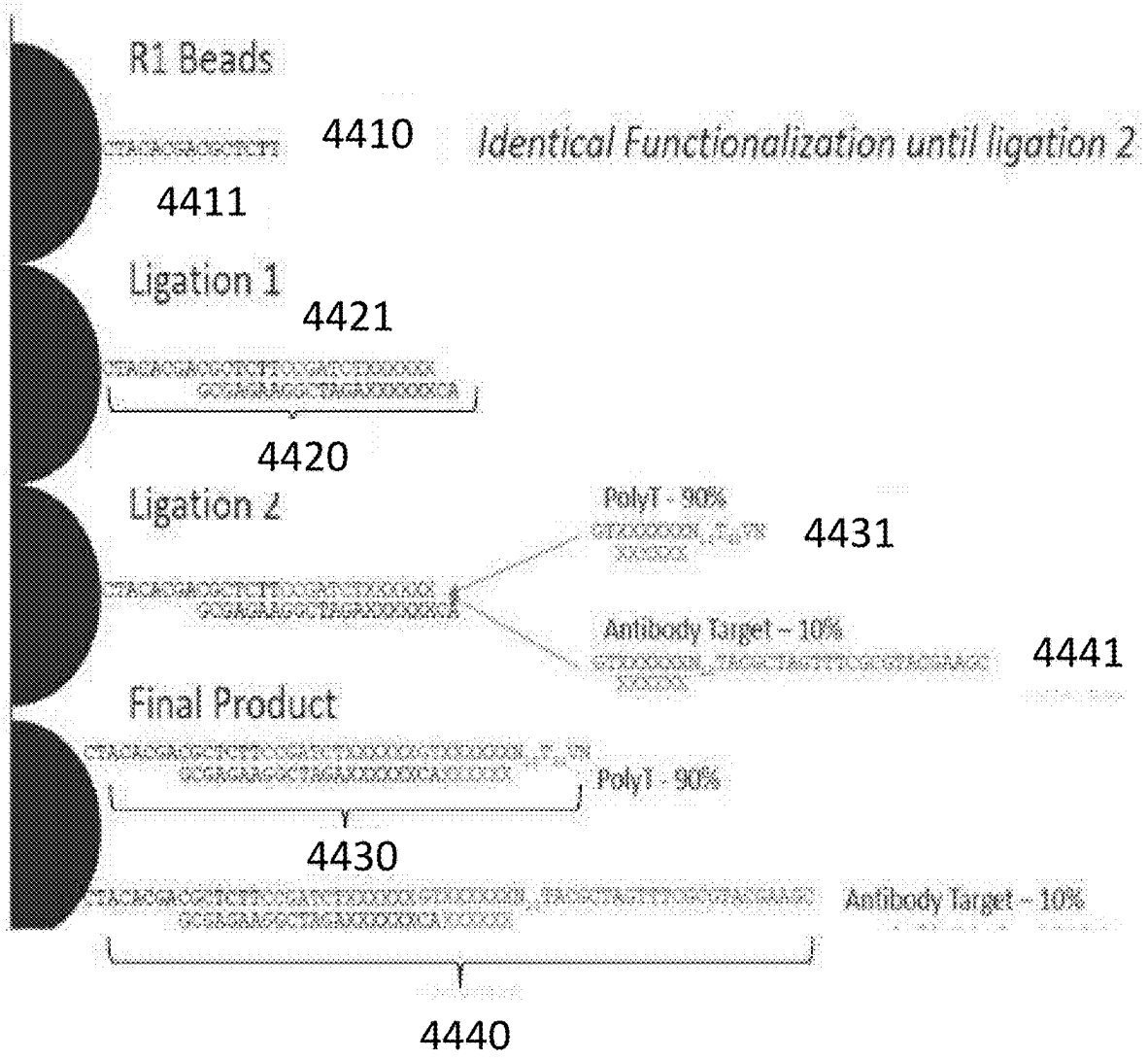

R1 Beads
4410
Identical Functionalization until ligation 2
4411
Ligation 1
4421
4420
Ligation 2
PolyT - 90%
4431
Antibody Target ~ 10%
4441
Final Product
PolyT - 90%
4430
Antibody Target ~ 10%
4440

FIG. 44

FIG. 45A

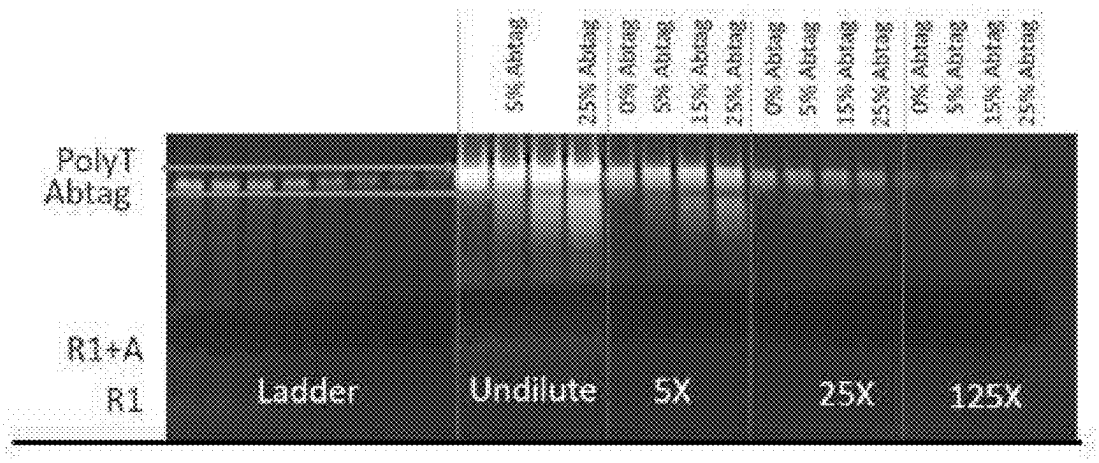
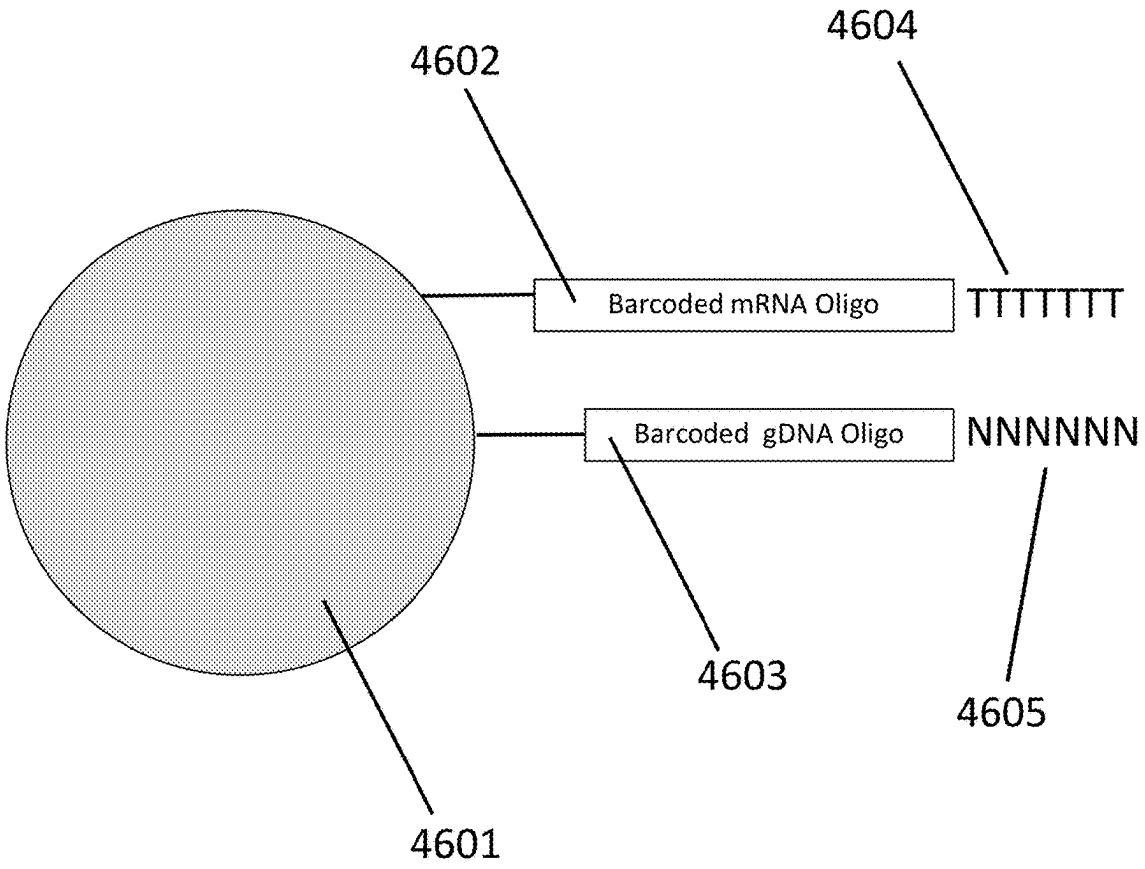


FIG. 45B


4604
4602
Barcoded mRNA Oligo
TTTTTTT
Barcoded gDNA Oligo
NNNNNN
4603
4605
4601

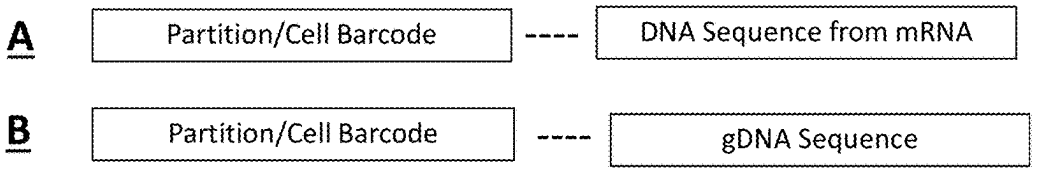

A
Partition/Cell Barcode
DNA Sequence from mRNA
B
Partition/Cell Barcode
gDNA Sequence

FIG. 46A

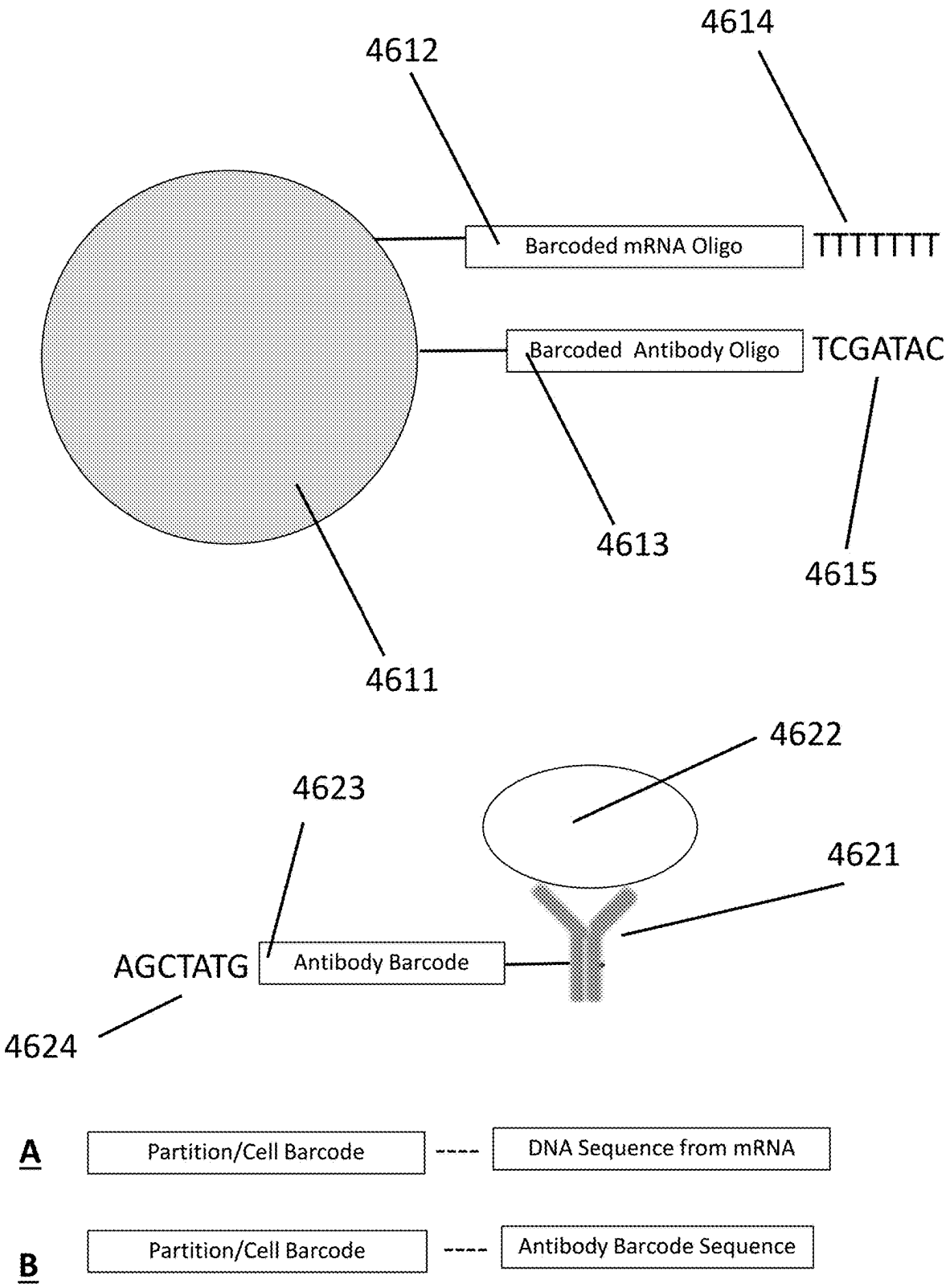
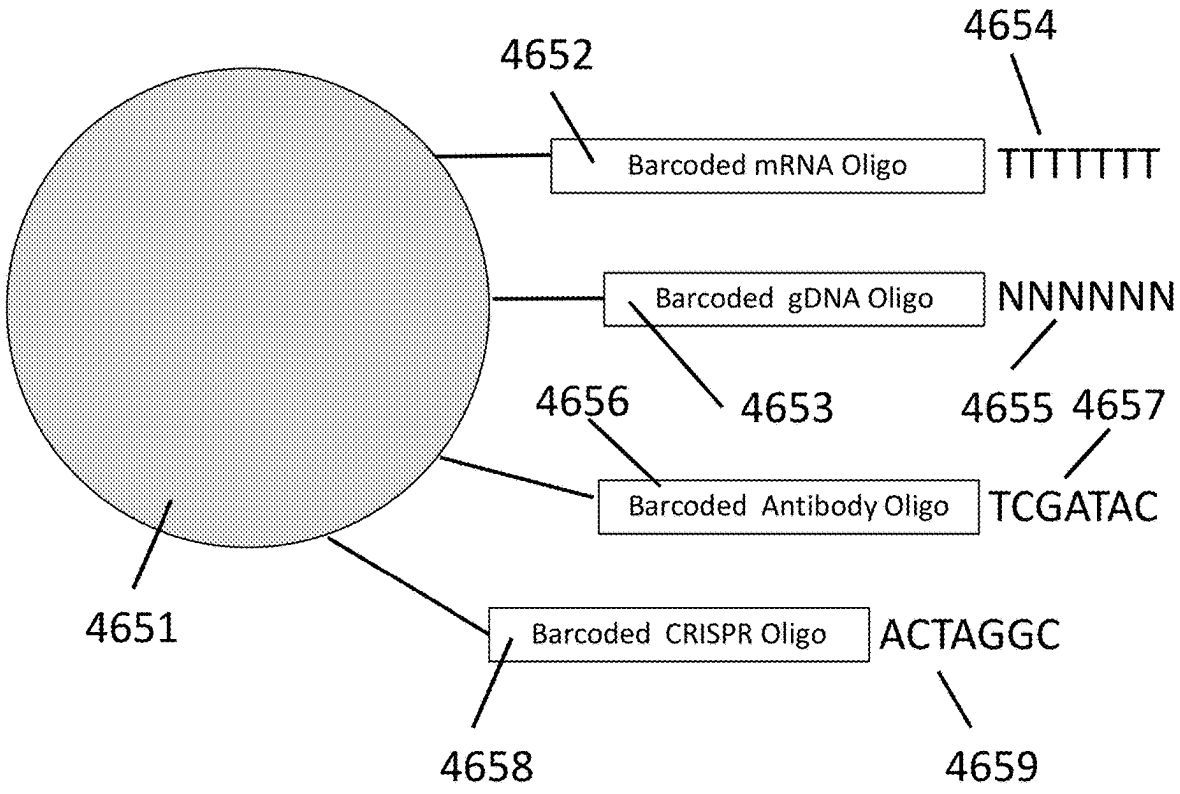

4614
4612
Barcoded mRNA Oligo
TTTTTTT
Barcoded Antibody Oligo
TCGATAC
4613
4615
4611
4622
4623
4621
AGCTATG
Antibody Barcode
4624
A
Partition/Cell Barcode
DNA Sequence from mRNA
B
Partition/Cell Barcode
Antibody Barcode Sequence

FIG. 46B


4652
4654
Barcoded mRNA Oligo
TTTTTTT
Barcoded gDNA Oligo
NNNNNN
4656
4653
4655
4657
Barcoded Antibody Oligo
TCGATAC
4651
Barcoded CRISPR Oligo
ACTAGGC
4658
4659

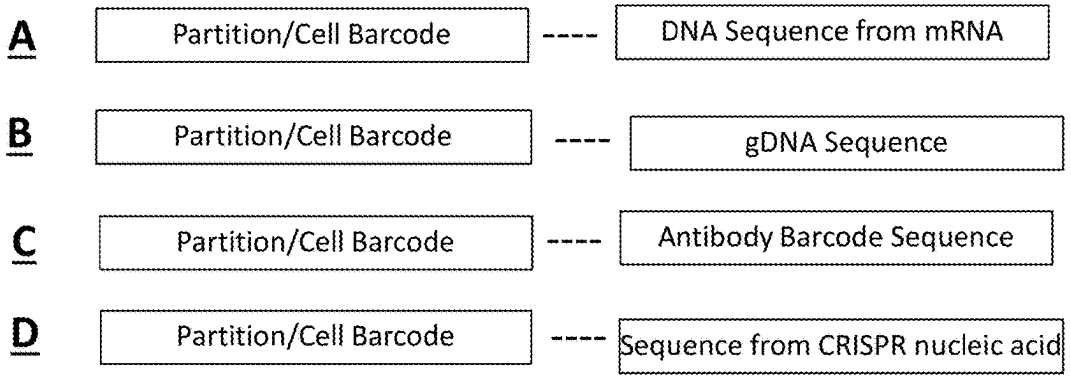

A
Partition/Cell Barcode
DNA Sequence from mRNA
B
Partition/Cell Barcode
gDNA Sequence
C
Partition/Cell Barcode
Antibody Barcode Sequence
D
Partition/Cell Barcode
Sequence from CRISPR nucleic acid

FIG. 46C

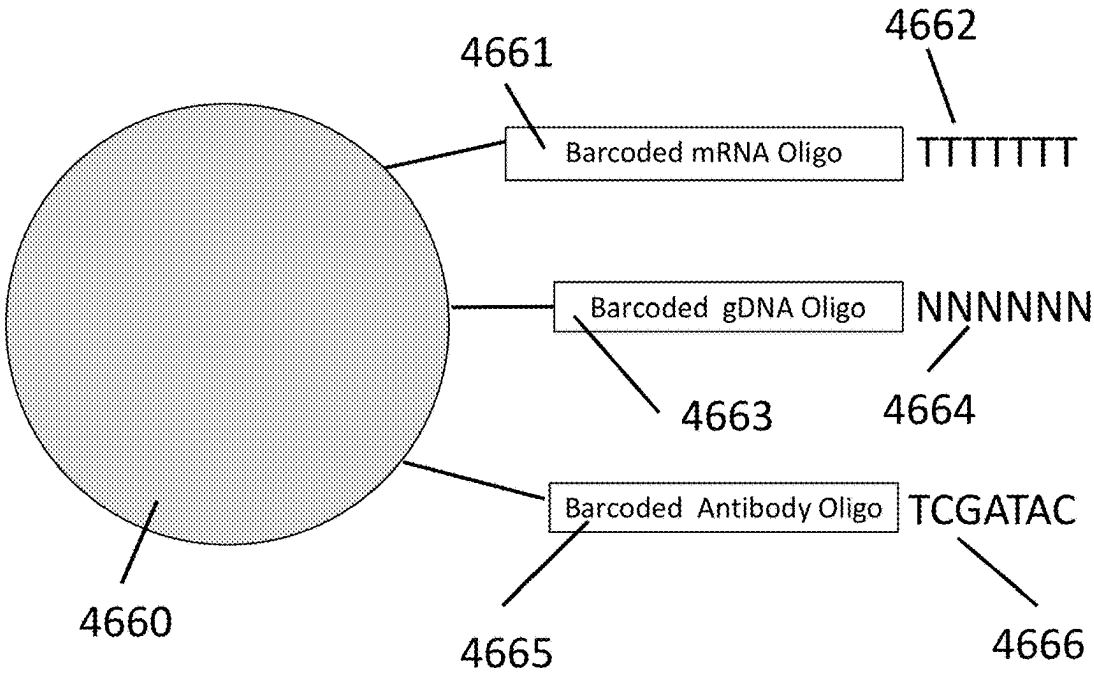


FIG. 46D

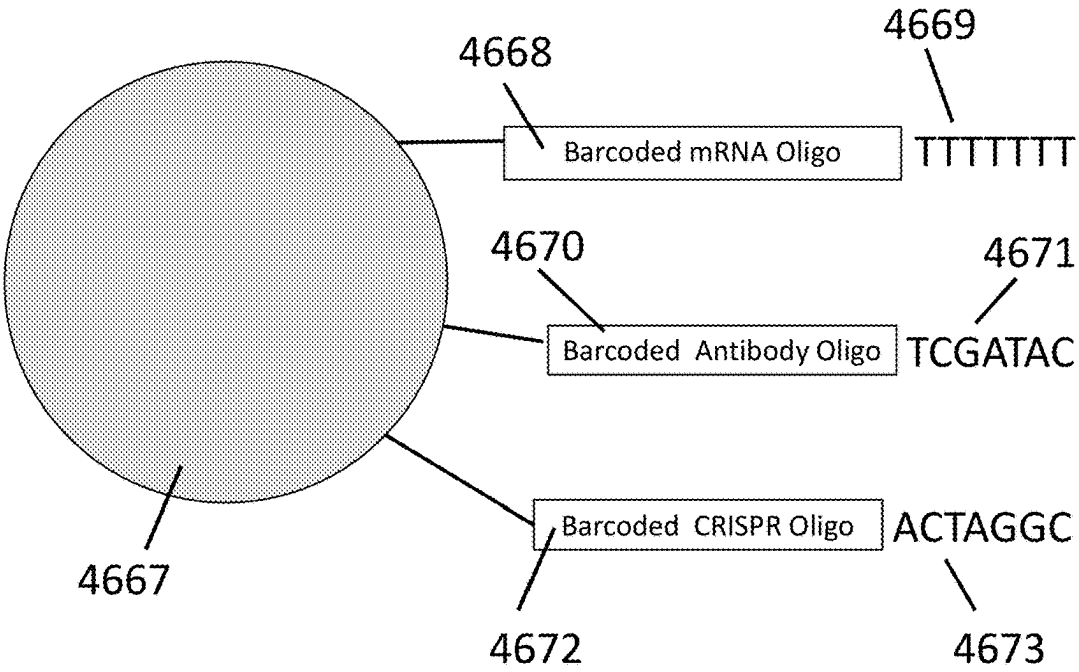


FIG. 46E

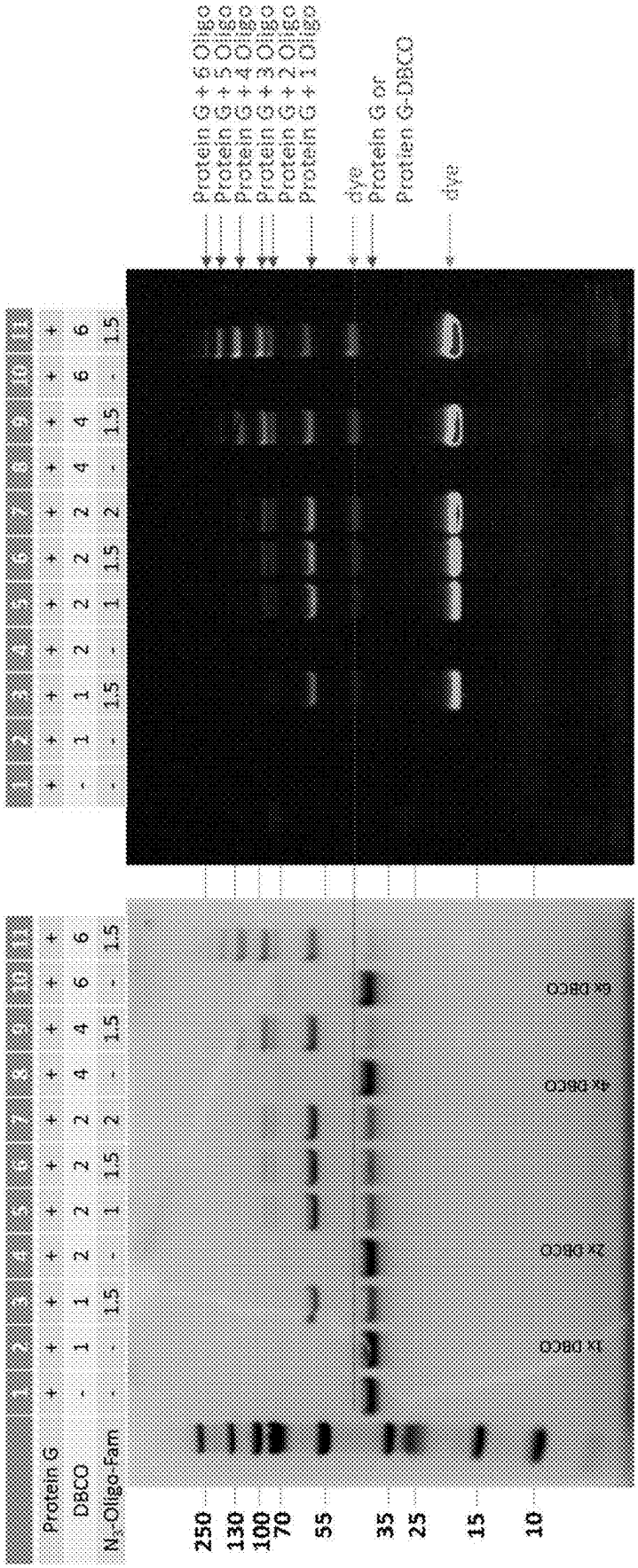
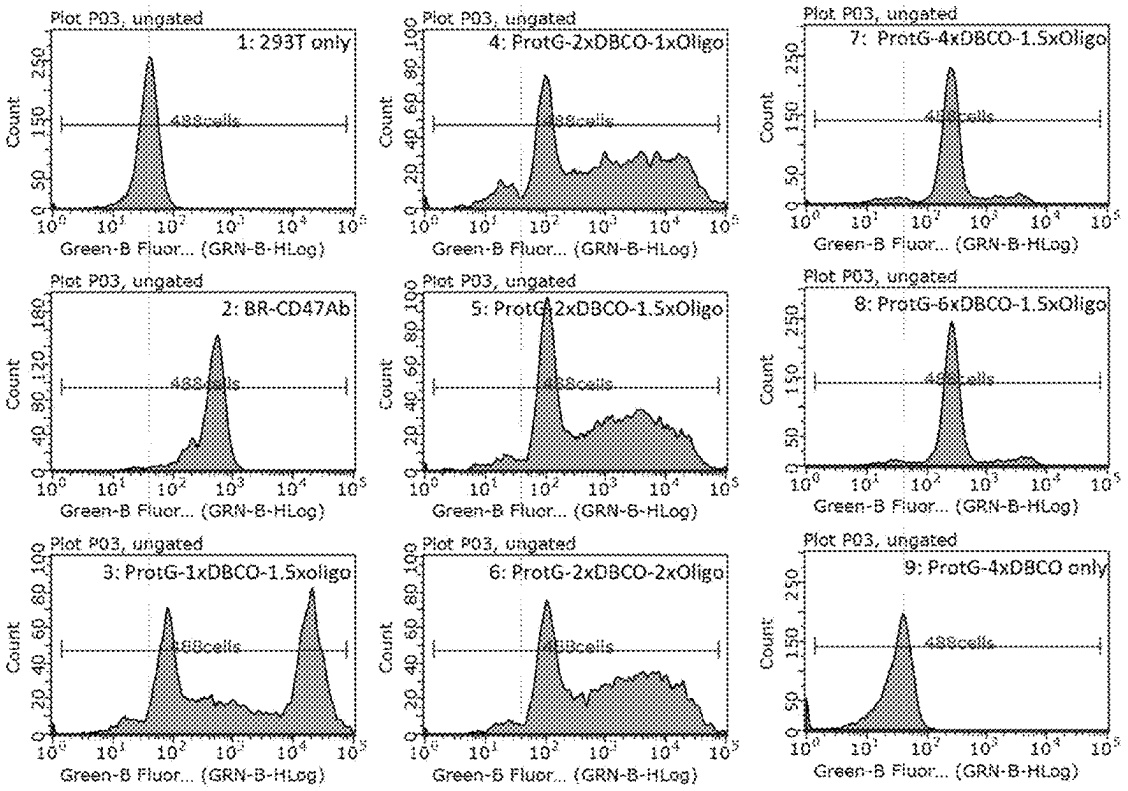

Protein G
DBCO
N3-Oligo-Fam
1x DBCO
2x DBCO
4x DBCO
6x DBCO
250
130
100
70
55
35
25
15
10
Protein G + 6 Oligo
Protein G + 5 Oligo
Protein G + 4 Oligo
Protein G + 3 Oligo
Protein G + 2 Oligo
Protein G + 1 Oligo
dye
Protein G or
Protien G-DBCO
dye

FIG. 47A

	Relative Band % (Coomassie gel)						
			2X DBCO				
MW (kDa)	Protein G	1x DBCO 1.5x oligo	1x oligo	1.5x oligo	2x oligo	4x DBCO 1.5x oligo	6x DBCO 1.5x oligo
155.6							7.0
117.6						4.6	20.1
85.0			2.8	4.6	6.9	20.3	21.2
70.0			1.6	2.9	4.0	7.7	5.4
58.2		32.9	46.6	45.8	44.4	44.4	37.2
39.6	100	67.1	49.0	46.7	44.7	23.0	9.1

FIG. 47B


Plot P03, ungated
1: 293T only
4: ProtG-2xDBCO-1xOligo
7: ProtG-4xDBCO-1.5xOligo
2: BR-CD47Ab
5: ProtG-2xDBCO-1.5xOligo
8: ProtG-6xDBCO-1.5xOligo
3: ProtG-1xDBCO-1.5xoligo
6: ProtG-2xDBCO-2xOligo
9: ProtG-4xDBCO only
Count
Green-B Fluor... (GRN-B-HLog)

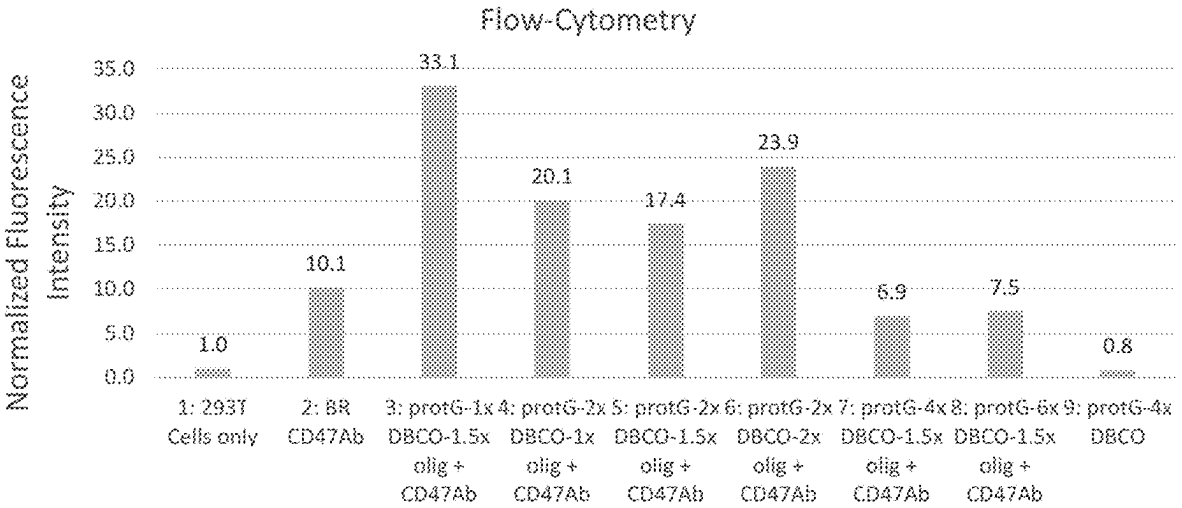
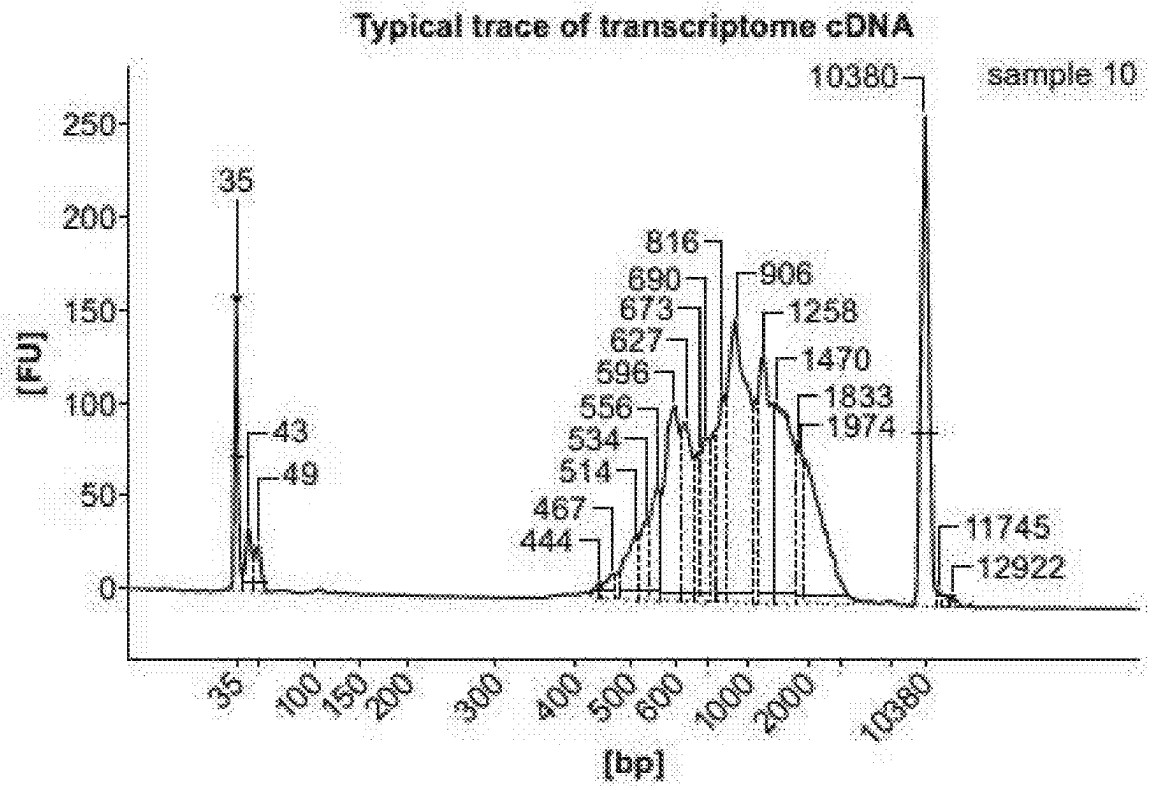

Flow-Cytometry
Normalized Fluorescence Intensity
35.0
30.0
25.0
20.0
15.0
10.0
5.0
0.0
1.0
10.1
33.1
20.1
17.4
23.9
6.9
7.5
0.8
1: 293T Cells only
2: BR CD47Ab
3: protG-1x DBCO-1.5x olig + CD47Ab
4: protG-2x DBCO-1x olig + CD47Ab
5: protG-2x DBCO-1.5x olig + CD47Ab
6: protG-2x DBCO-2x olig + CD47Ab
7: protG-4x DBCO-1.5x olig + CD47Ab
8: protG-6x DBCO-1.5x olig + CD47Ab
9: protG-4x DBCO

FIG. 48


Typical trace of transcriptome cDNA
sample 10
[FU]
250
200
150
100
50
0
35
43
49
467
444
514
534
556
596
627
673
690
816
906
1258
1470
1833
1974
10380
11745
12922
35
100
150
200
300
400
500
600
1000
2000
10380
[bp]

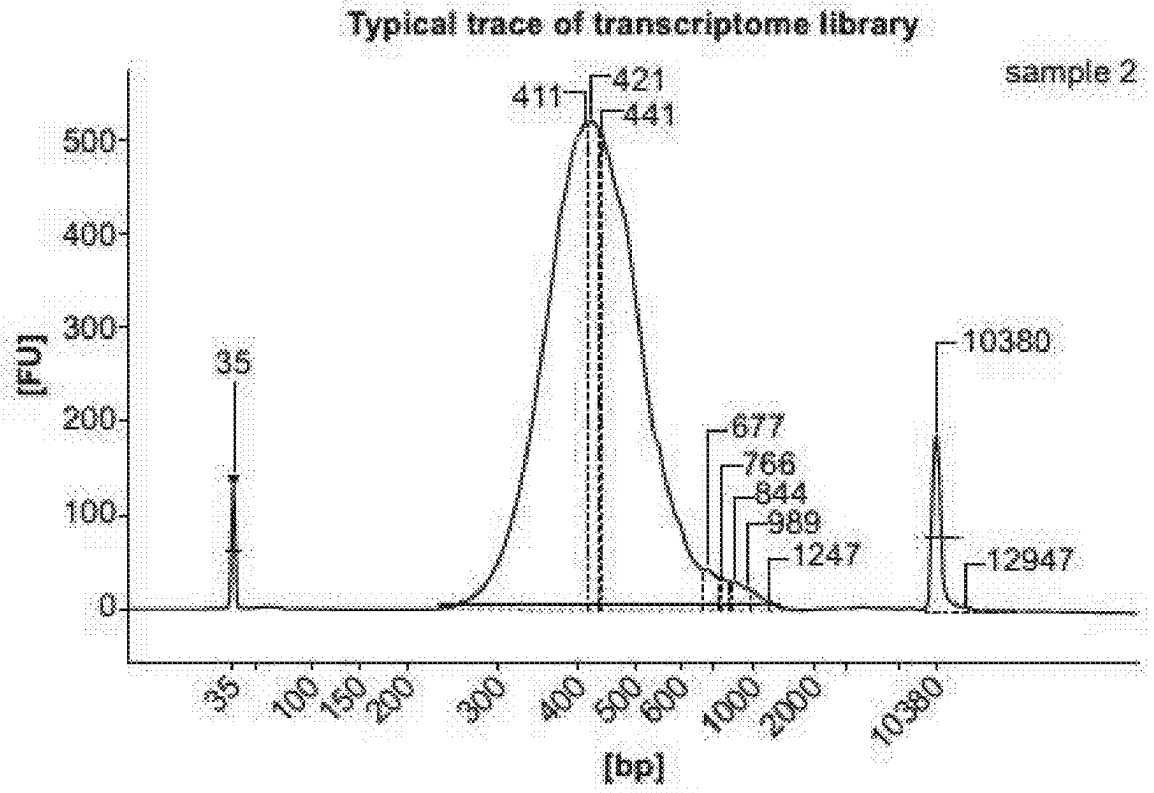

Typical trace of transcriptome library
sample 2
[FU]
500
400
300
200
100
0
35
411
421
441
677
766
844
989
1247
10380
12947
35
100
150
200
300
400
500
600
1000
2000
10380
[bp]

Figure 49A

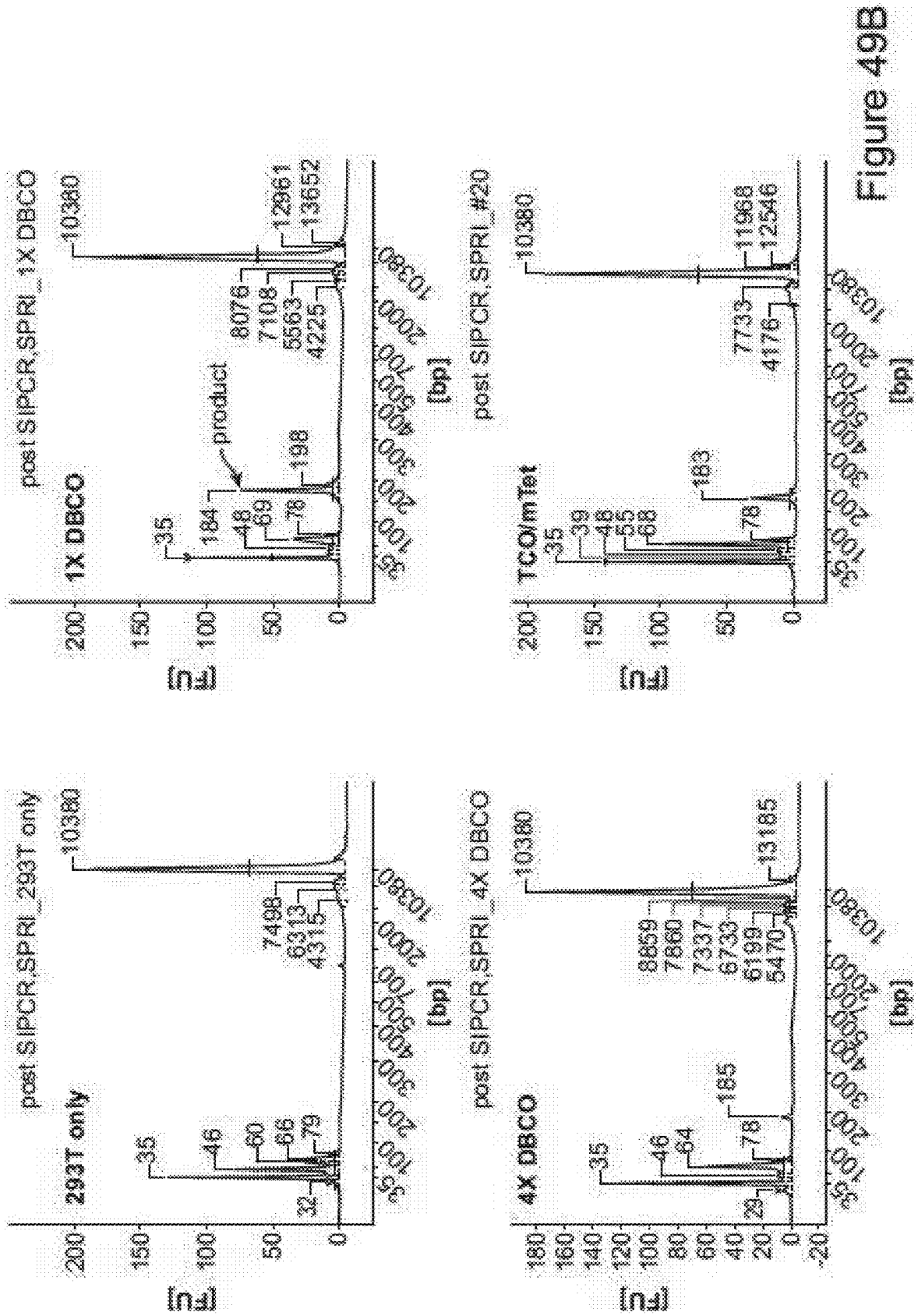

post SIPCR_SPRI_1X DBCO
1X DBCO
product
[FU]
[bp]
post SIPCR_SPRI_#20
TCO/mTet
post SIPCR_SPRI_293T only
293T only
post SIPCR_SPRI_4X DBCO
4X DBCO

Figure 49B

Description	Correted Area	Product Peak Size (bp)
293T Cells only	11.0	-
protG-1x DBCO-1.5x oligo + CD47Ab	178.4	184
protG-2x DBCO-2x oligo + CD47Ab	120.5	184
protG-4x DBCO-1.5x oligo + CD47Ab	17.6	185
protG-6x DBCO-1.5x oligo + CD47Ab	32.4	184
protG-TCO/mtet-10x oligo + CD47Ab	45.6	183
protG-oligo	5.0	-

Figure 49B (Cont.)

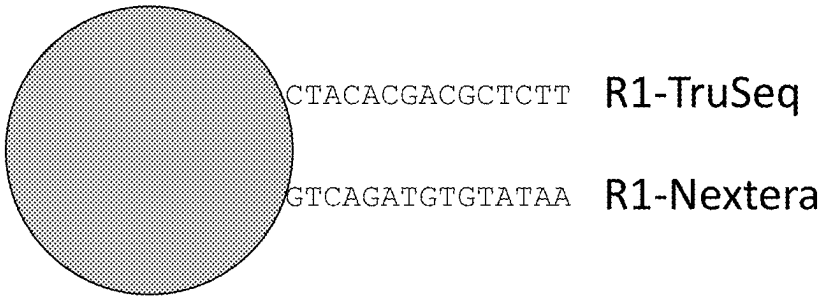


FIG. 50A

Alpha

CTACACGACGCTCTTCCGATCTXXXXXXXGTXXXXXXXN$_{10}$T$_{30}$VN
GCGAGAAGGCTAGAXXXXXXXCAXXXXXXX

Beta

CTACACGACGCTCTTCCGATCTXXXXXXXCAXXXXXXXN$_{10}$T$_{30}$VN
GCGAGAAGGCTAGAXXXXXXXGTXXXXXXX

Gamma

CTACACGACGCTCTTCCGATCTXXXXXXXAGXXXXXXXN$_{10}$T$_{30}$VN
GCGAGAAGGCTAGAXXXXXXXTCXXXXXXX

Delta

CTACACGACGCTCTTCCGATCTXXXXXXXTCXXXXXXXN$_{10}$T$_{30}$VN
GCGAGAAGGCTAGAXXXXXXXAGXXXXXXX

FIG. 50B

Alpha

GTCAGATGTGTATAAGAGACAGXXXXXXXGTXXXXXXXN$_{10}$GCTTCGTACGCGAAACTAGCGT
CACATATTCTCTGTCXXXXXXXCAXXXXXXX

Beta

GTCAGATGTGTATAAGAGACAGXXXXXXXGTXXXXXXXN$_{10}$GCTTCGTACGCGAAACTAGCGT
CACATATTCTCTGTCXXXXXXXCAXXXXXXX

Gamma

GTCAGATGTGTATAAGAGACAGXXXXXXXGTXXXXXXXN$_{10}$GCTTCGTACGCGAAACTAGCGT
CACATATTCTCTGTCXXXXXXXCAXXXXXXX

Delta

GTCAGATGTGTATAAGAGACAGXXXXXXXGTXXXXXXXN$_{10}$GCTTCGTACGCGAAACTAGCGT
CACATATTCTCTGTCXXXXXXXCAXXXXXXX

FIG. 50C

5101

5102

5′CTACACGACGCTCTTCCGATCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTTCTTATATrGrGrG
3′ AAAGAATATA C C CTAGACTGACGTGGAACCTGGCGATTTCAAC-**FAM**

FIG. 51A

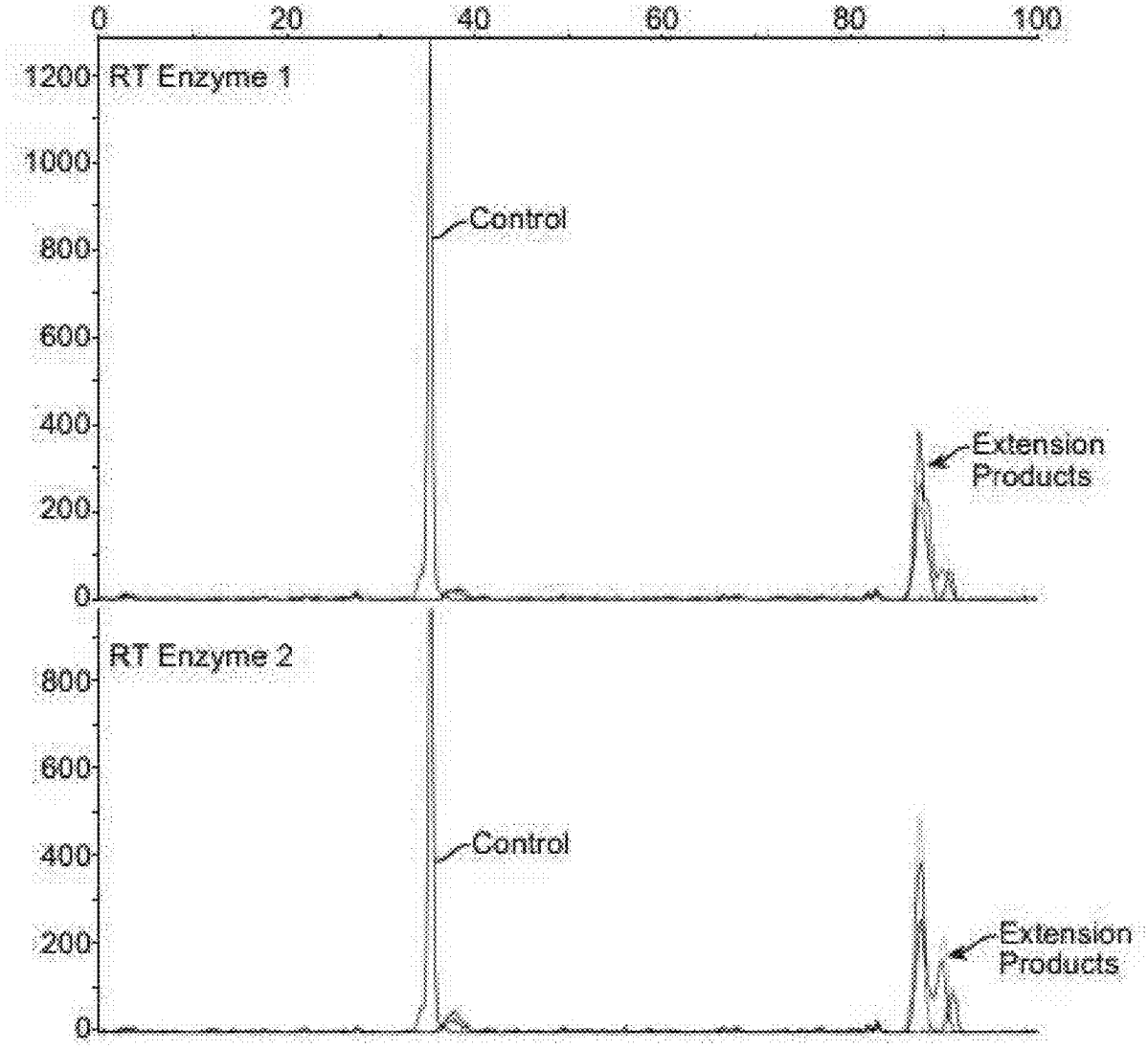

0
20
40
60
80
100
1200
1000
800
600
400
200
0
RT Enzyme 1
Control
Extension Products
800
600
400
200
0
RT Enzyme 2
Control
Extension Products

Figure 51B

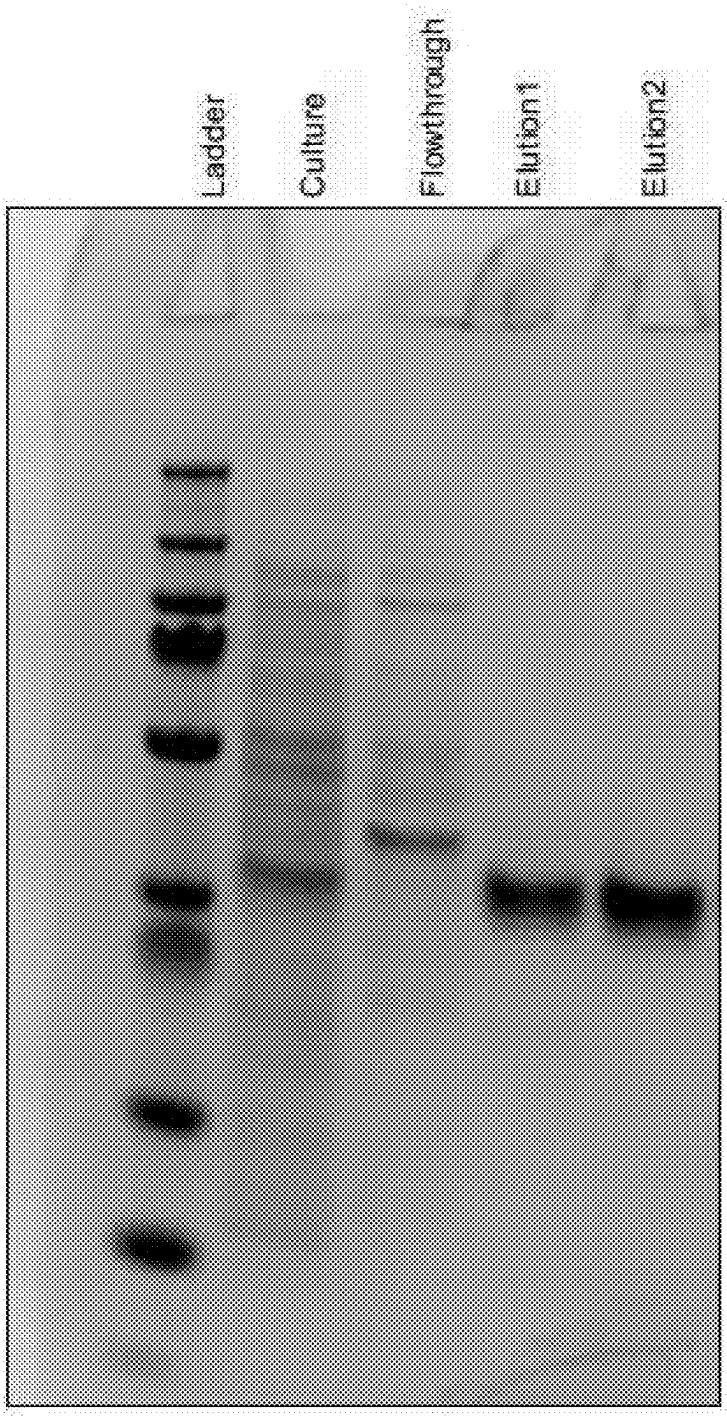


Figure 52A

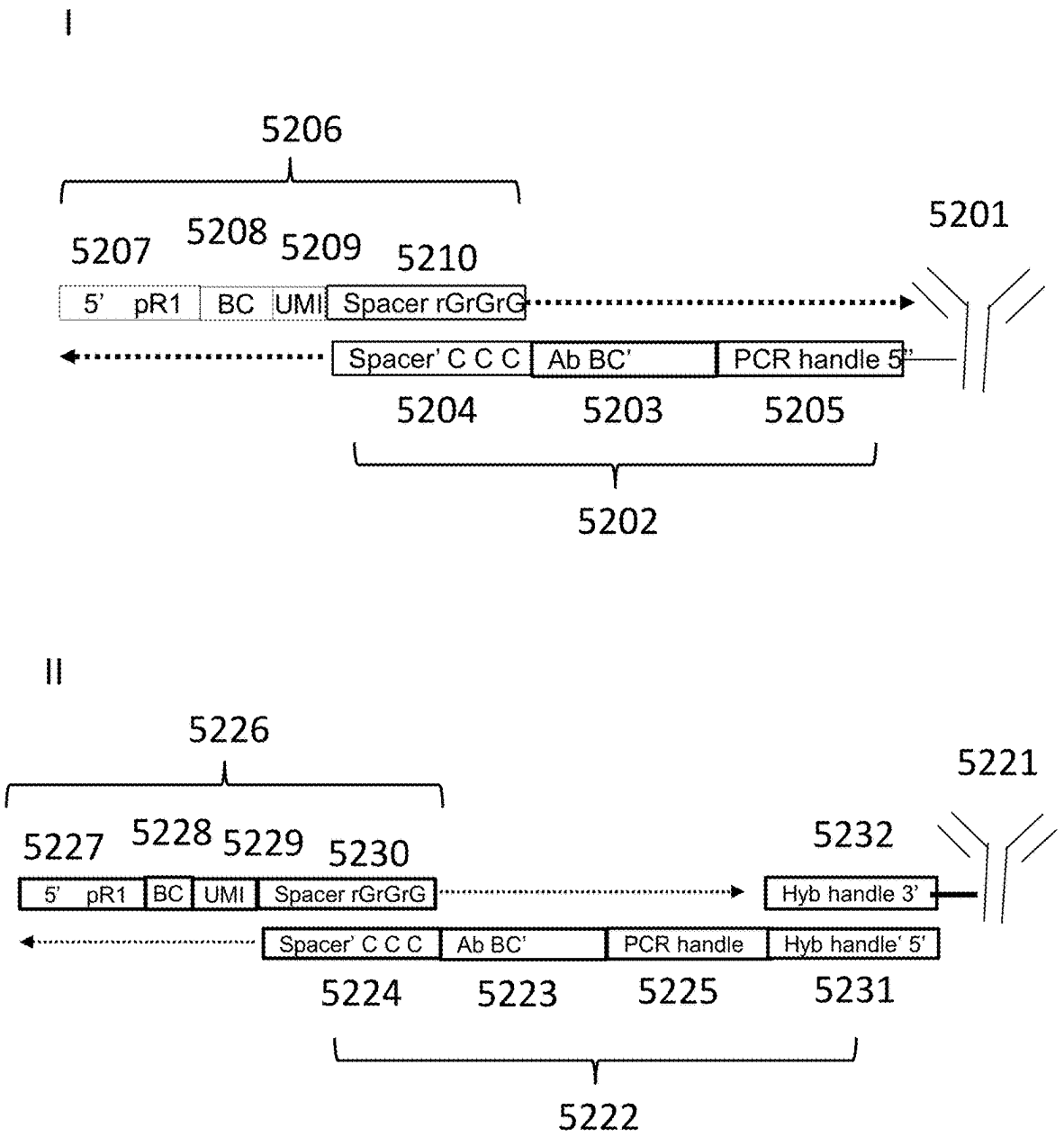

I
5206
5207
5208
5209
5210
5201
5' pR1
BC
UMI
Spacer rGrGrG
Spacer' C C C
Ab BC'
PCR handle 5'
5204
5203
5205
5202
II
5226
5227
5228
5229
5230
5221
5232
5' pR1
BC
UMI
Spacer rGrGrG
Hyb handle 3'
Spacer' C C C
Ab BC'
PCR handle
Hyb handle' 5'
5224
5223
5225
5231
5222

FIG. 52B

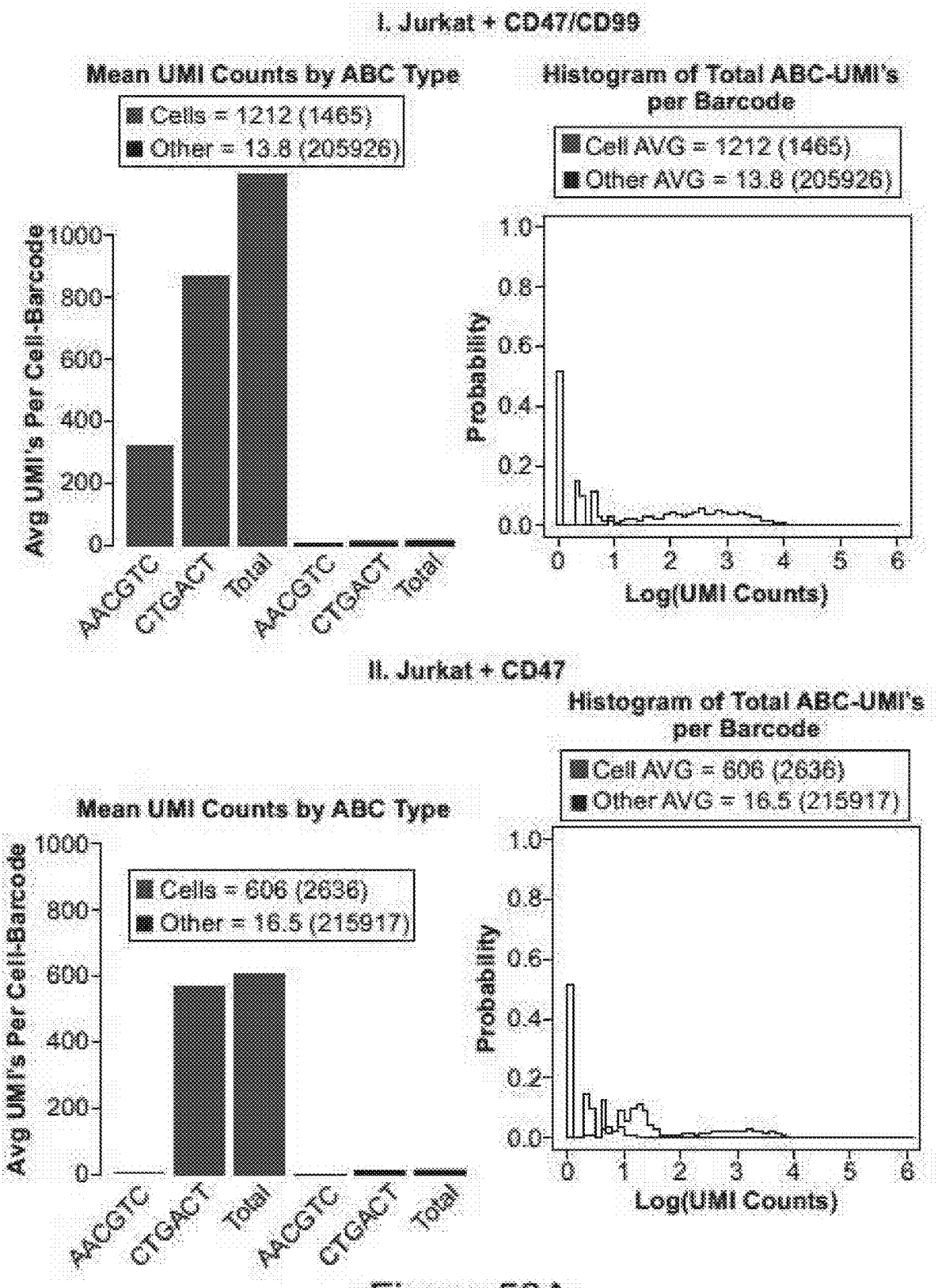

I. Jurkat + CD47/CD99
Mean UMI Counts by ABC Type
Cells = 1212 (1465)
Other = 13.8 (205926)
Avg UMI's Per Cell-Barcode
1000
800
600
400
200
0
AACGTC
CTGACT
Total
AACGTC
CTGACT
Total
Histogram of Total ABC-UMI's per Barcode
Cell AVG = 1212 (1465)
Other AVG = 13.8 (205926)
Probability
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
6
Log(UMI Counts)
II. Jurkat + CD47
Mean UMI Counts by ABC Type
Cells = 606 (2636)
Other = 16.5 (215917)
Avg UMI's Per Cell-Barcode
1000
800
600
400
200
0
AACGTC
CTGACT
Total
AACGTC
CTGACT
Total
Histogram of Total ABC-UMI's per Barcode
Cell AVG = 606 (2636)
Other AVG = 16.5 (215917)
Probability
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
6
Log(UMI Counts)

Figure 53A

Sample	Sample ID	Avg UMIs Per Cell	Avg UMIs Per Other	BC3 per cell	BC4 per cell	Total # of cell BCs detected in AbBC expt	# cells, SC5	# cells, AbBC topN (~#cells, SC5)	AbBC topN cells ∩ SC5 cells	AbBC called cells	AbBC cells ∩ SC5 cells	SC5 cells ∩ AbBC topN / SC5 cells	BCAb cells ∩ SC5' cells / AbBC cells
Unbiased Jurkat + BC4	33600	498	13.9	64647	635932	128481	1419	1418	378	1362	370	0.27	0.27
Unbiased Jurkat + BC3/4	33601	577	17.1	247587	404815	126514	1141	1141	368	992	316	0.32	0.32
Unbiased Jurkat + BC4	33602	996	9.14	1474578	6119	144263	1505	1504	862	1193	762	0.57	0.64
Unbiased Jurkat spike BC3/4	33603	38.8	14.1	1163	71316	206337	1898	1980	103	62631	1570	0.05	0.03

FIG. 53B

Sample	description	Mean Reads per Cell	Valid Barcodes	Reads Mapped Confidently to Transcriptome	Fraction Reads in Cells	Fraction reads usable	Reads Mapped Confidently to Exonic Regions	Reads Mapped Confidently to Intronic Regions	Reads Mapped Confidently to Intergenic Regions	Reads Mapped Antisense to Gene	Fraction reads unmapped	cDNA PCR Duplication (20k raw reads per cell)	Median genes per cell (20k raw reads per cell)	Median UMI counts per cell (20k raw reads per cell)
37358	**cDNA control**	217,543	83.60%	56.00%	95.70%	26.20%	64.50%	12.70%	8.90%	9.00%	3.30%	46.90%	1,103	2,764
37772	Unbiased-5: CG-CD3 GEM 53/85C D	31,899	70.20%	32.90%	95.10%	15.20%	41.60%	9.00%	16.60%	12.80%	4.70%	43.30%	730	1,690
37773	Unbiased-6: CG-CD3 GEM 53/85C D	29,602	69.20%	29.90%	94.40%	13.70%	38.90%	9.30%	16.80%	14.00%	4.70%	42.80%	684	1,562
37776	Unbiased-21: SC-CD3 GEM 53/85C D	36,549	76.00%	37.10%	95.40%	17.20%	47.40%	13.20%	13.50%	13.90%	5.10%	44.60%	812	1,935
37777	Unbiased-22: SC-CD3 GEM 53/85C D	44,884	76.40%	39.20%	96.10%	18.30%	48.90%	12.60%	13.10%	13.00%	5.10%	43.20%	855	2,077
37780	Unbiased-7: CG-CD3 GEM 53/85C St	33,828	76.20%	38.30%	94.40%	17.60%	47.80%	12.00%	14.50%	12.80%	4.50%	48.30%	781	1,796
37781	Unbiased-8: CG-CD3 GEM 53/85C St	33,085	71.90%	33.90%	94.60%	15.60%	43.90%	11.20%	15.30%	14.30%	4.60%	46.00%	750	1,666
37784	Unbiased-23: SC-CD3 GEM 53/85C S	40,137	77.70%	40.20%	95.30%	18.60%	50.40%	14.90%	12.90%	12.90%	4.60%	49.60%	803	1,880
37785	Unbiased-24: SC-CD3 GEM 53/85C S	40,931	80.30%	39.10%	95.00%	18.10%	50.30%	16.20%	13.40%	14.00%	4.90%	52.30%	747	1,714
37770	Unbiased-1: CG-CD3 GEM 53C Dyna	29,136	91.40%	64.50%	94.90%	29.80%	71.10%	12.00%	7.90%	6.20%	6.50%	68.70%	673	1,824
37771	Unbiased-2: CG-CD3 GEM 53C Dyna	28,533	91.30%	65.40%	95.30%	30.30%	71.60%	11.90%	7.40%	5.80%	6.40%	64.80%	751	2,057
37774	Unbiased-17: SC-CD3 GEM 53C Dyn	42,069	91.30%	62.40%	95.10%	28.80%	69.40%	15.30%	7.30%	6.70%	6.70%	64.70%	752	2,009
37775	Unbiased-18: SC-CD3 GEM 53C Dyna	38,234	91.90%	63.00%	95.30%	29.10%	69.90%	15.30%	7.10%	6.60%	6.60%	65.30%	751	2,000
37778	Unbiased-[illegible] GEM [illegible]	32,773	92.60%	67.70%	94.90%	31.30%	73.60%	11.70%	6.90%	5.70%	5.90%	64.60%	848	2,133
37779	Unbiased-4: CG-CD3 GEM 53C Strep	30,309	92.80%	67.40%	94.40%	30.90%	73.40%	12.30%	6.70%	5.90%	6.10%	63.90%	855	2,147
37782	Unbiased-19: SC-CD3 GEM 53C Stre	45,118	92.00%	61.40%	95.20%	28.30%	68.40%	18.60%	6.70%	6.70%	6.30%	68.00%	757	1,744
37783	Unbiased-20: SC-CD3 GEM 53C Stre	37,621	92.00%	61.60%	94.50%	28.20%	68.60%	18.20%	6.50%	6.80%	6.80%	67.90%	741	1,702

Heat kill

No heat kill

FIG. 54

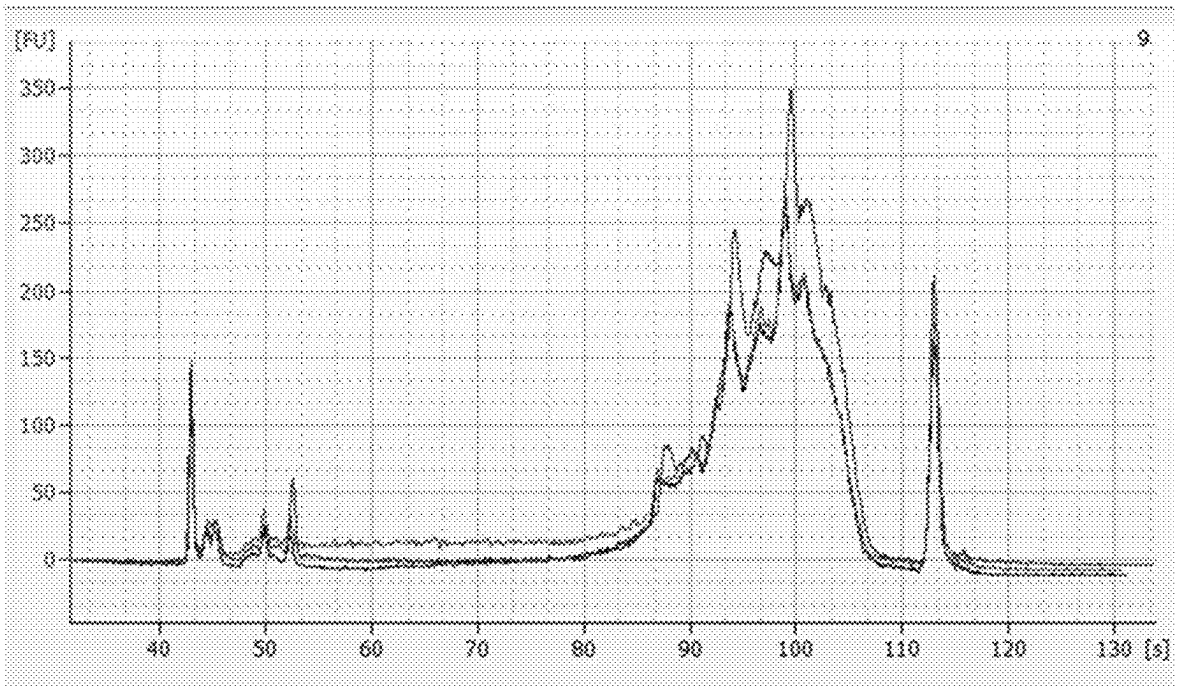


FIG. 55

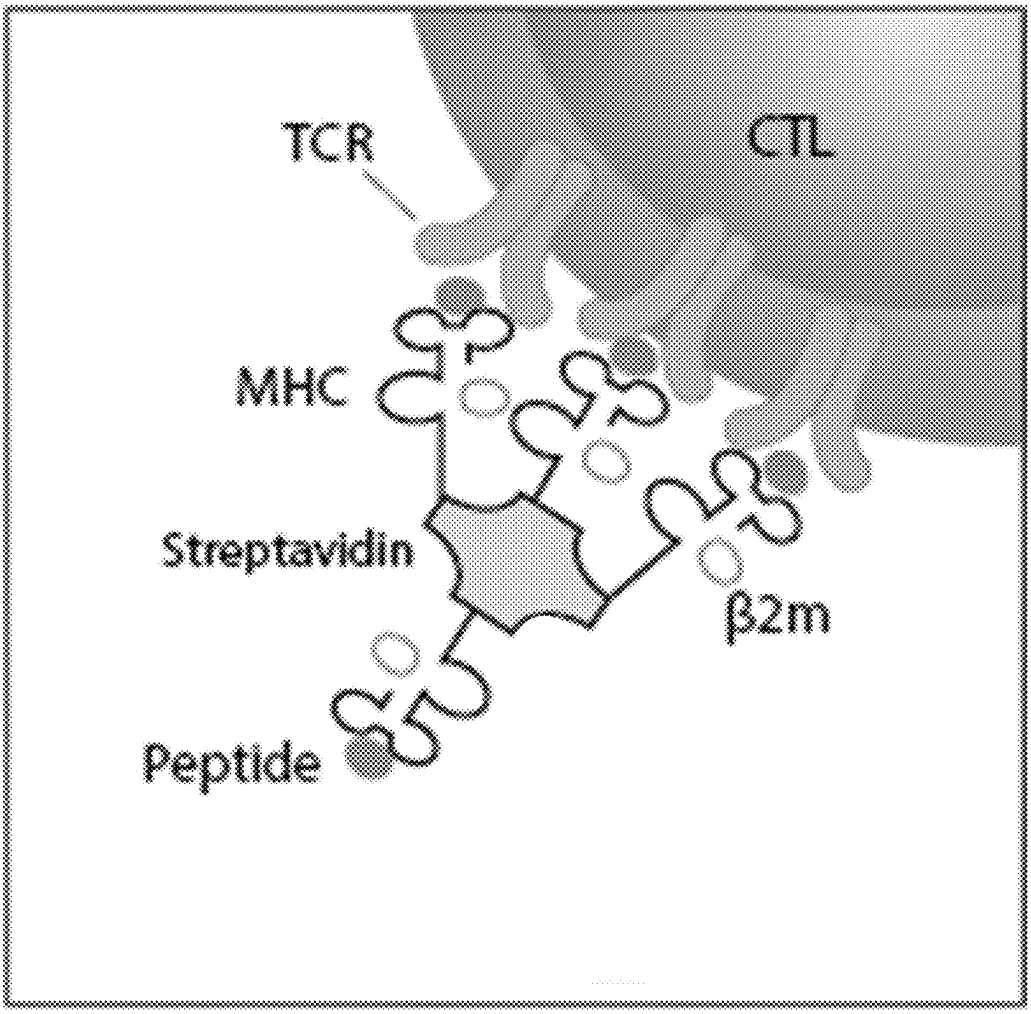


FIG. 56A

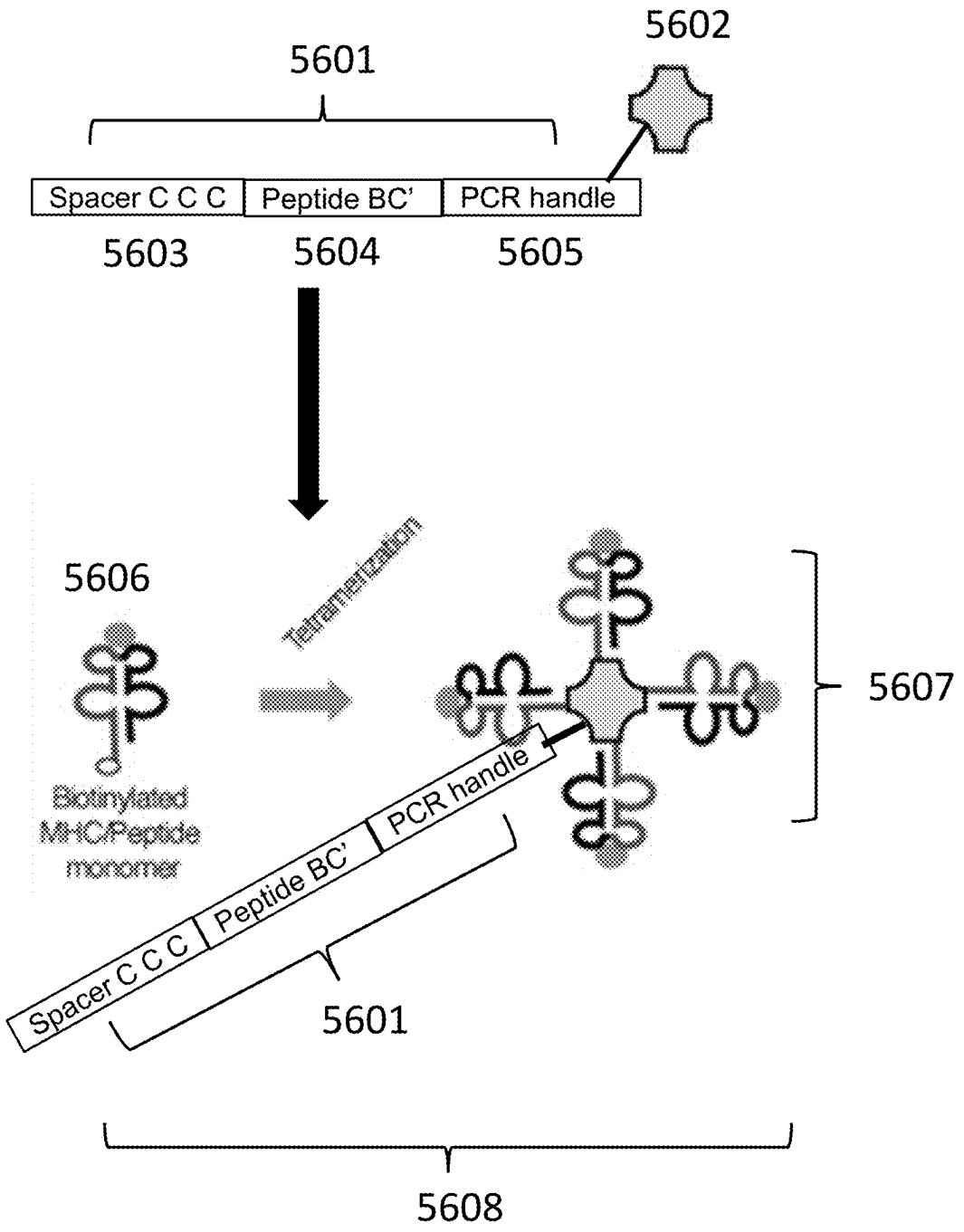

5601
5602
Spacer C C C
Peptide BC'
PCR handle
5603
5604
5605
5606
Tetramerization
Biotinylated MHC/Peptide monomer
5607
PCR handle
Peptide BC'
Spacer C C C
5601
5608

FIG. 56B

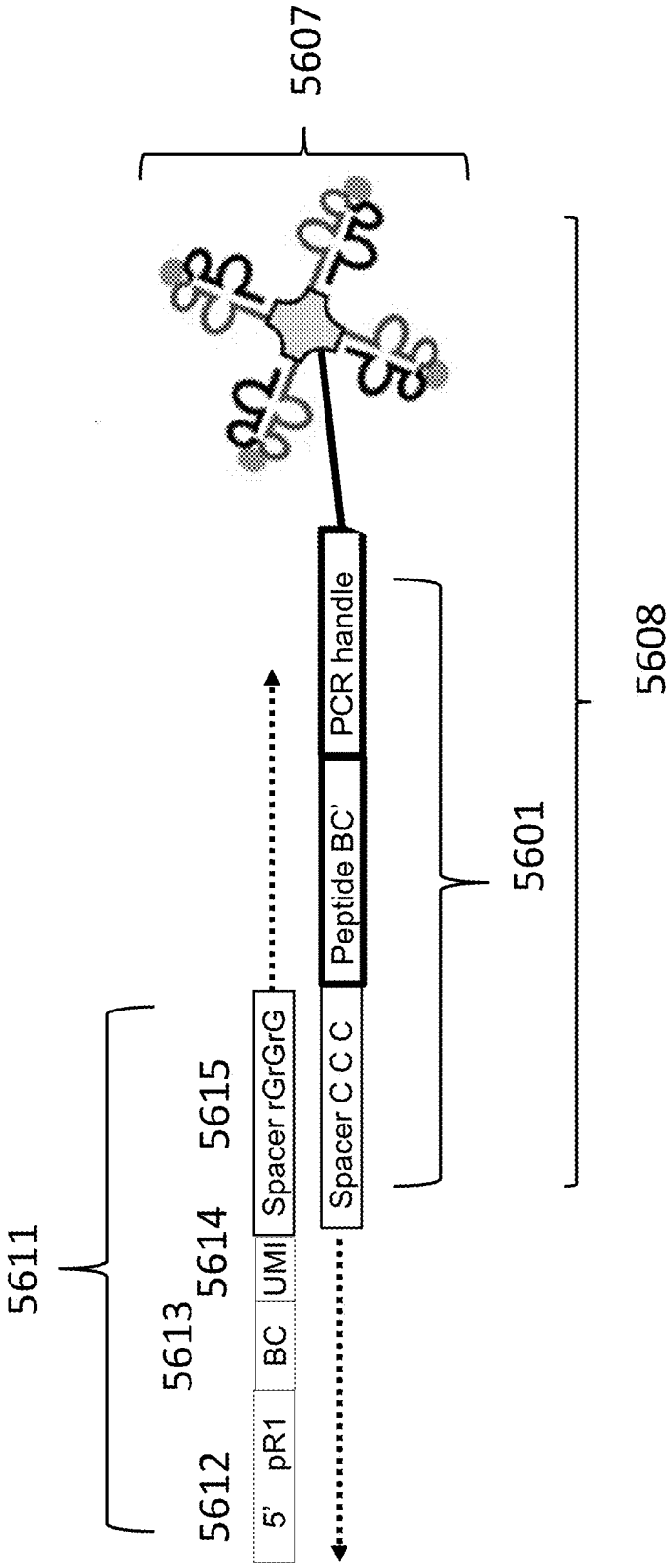

5607
5611
5612
5613
5614
5615
5601
5608
5' pR1
BC
UMI
Spacer rGrGrG
Spacer C C C
Peptide BC'
PCR handle

FIG. 56C

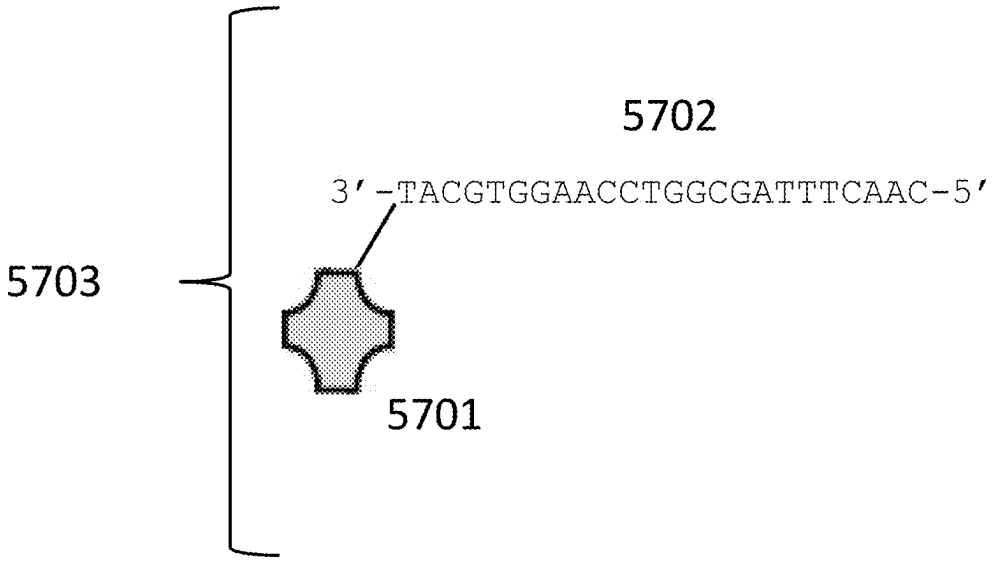


FIG. 57A

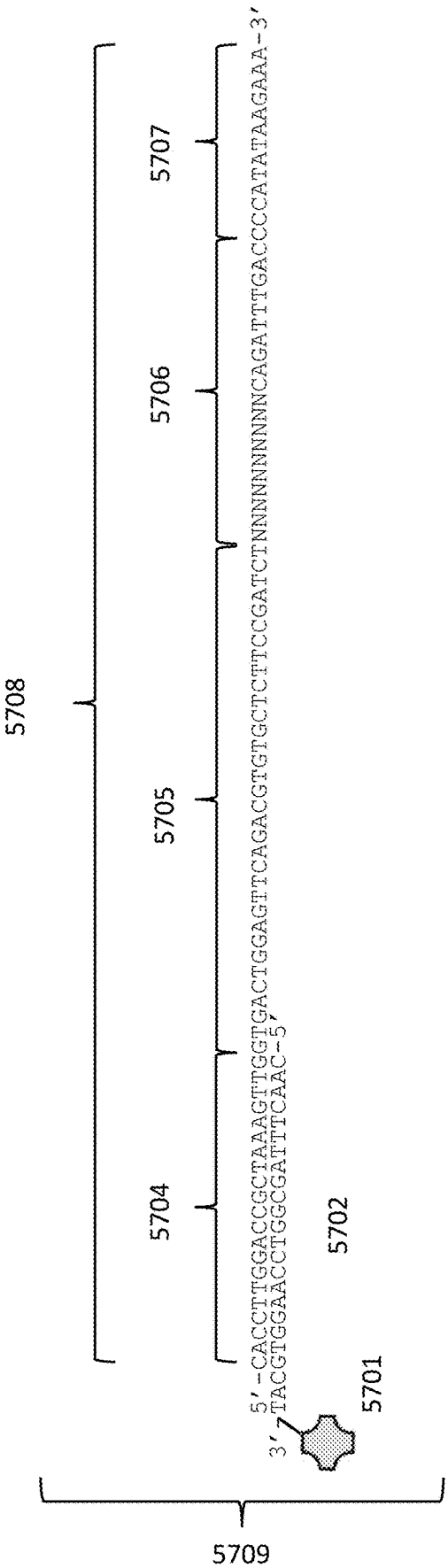

5'-CACCTTGGACCGCTAAAGTTGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNCAGATTTGACCCCATATAAGAAA-3'
3'-TACGTGGAACCTGGCGATTTCAAC-5'
5708
5704
5705
5706
5707
5701
5702
5709

FIG. 57B

FIG. 58A

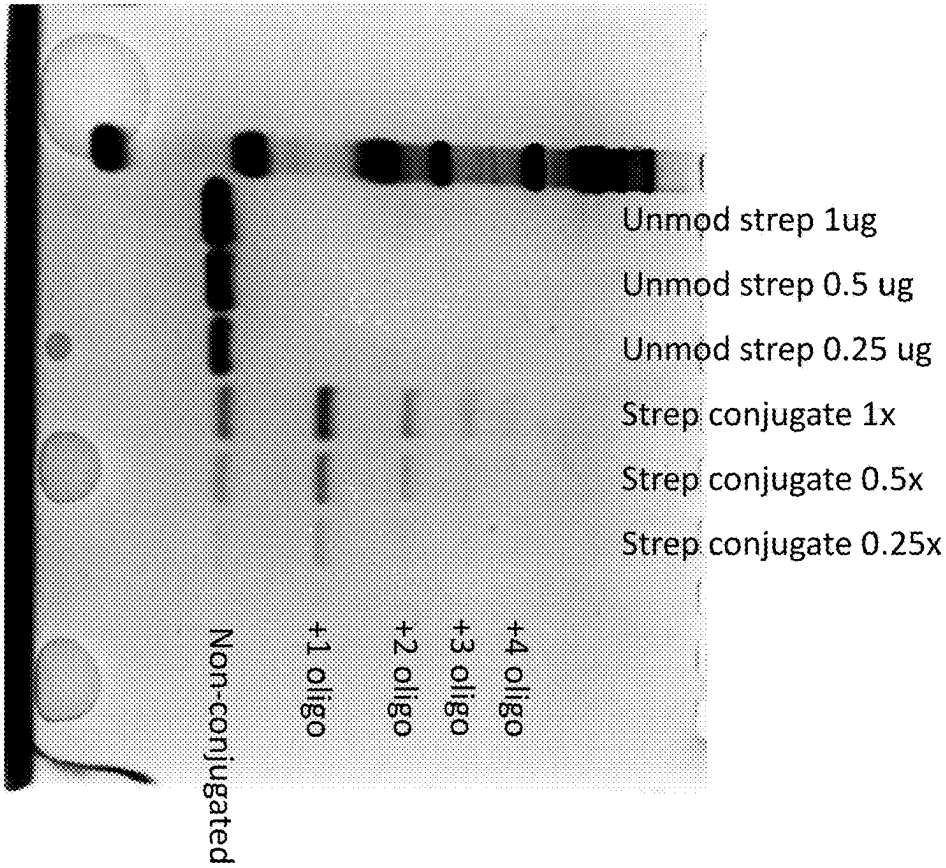


FIG. 58B

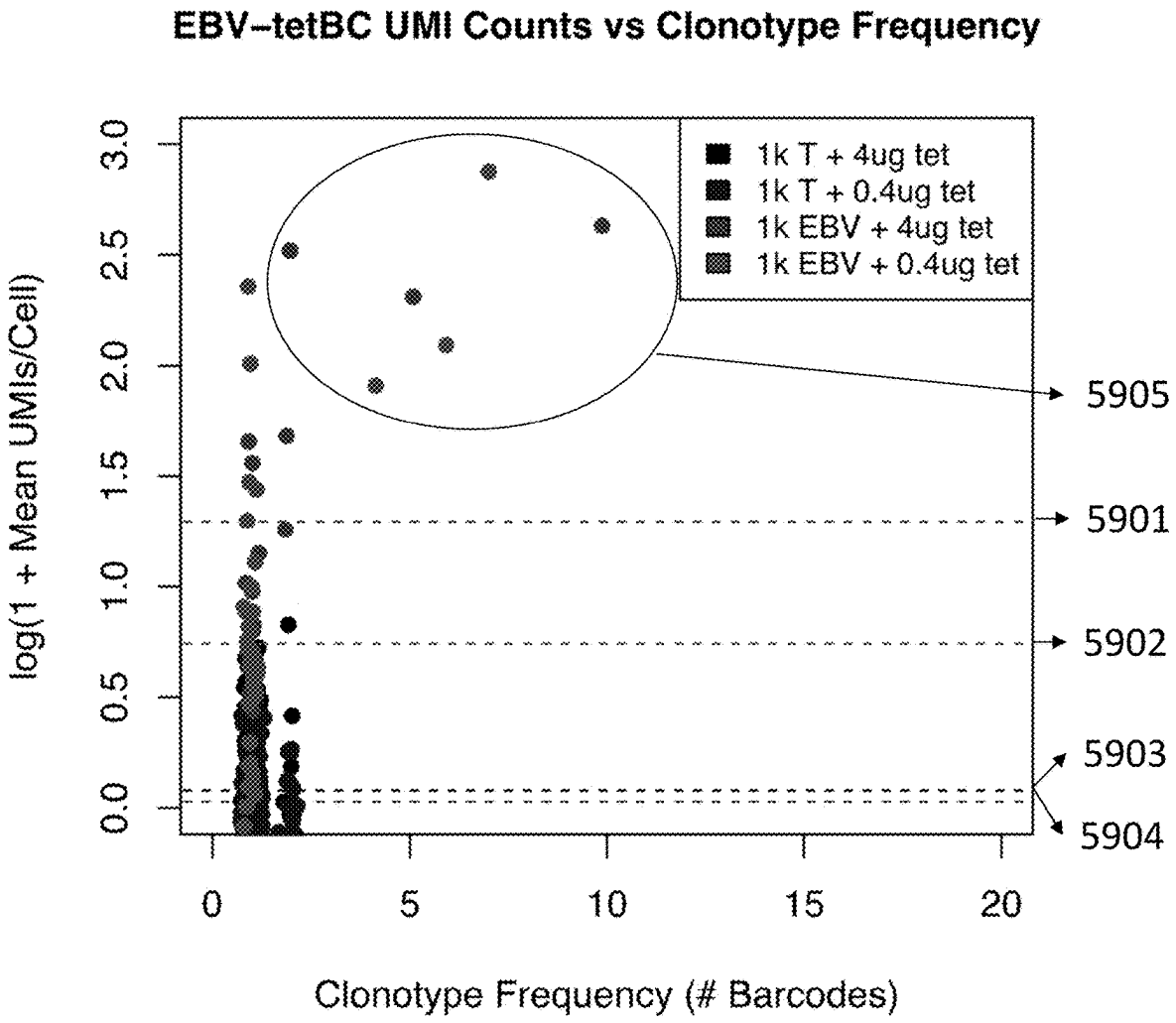

EBV–tetBC UMI Counts vs Clonotype Frequency
1k T + 4ug tet
1k T + 0.4ug tet
1k EBV + 4ug tet
1k EBV + 0.4ug tet
log(1 + Mean UMIs/Cell)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
5
10
15
20
Clonotype Frequency (# Barcodes)
5905
5901
5902
5903
5904

FIG. 59

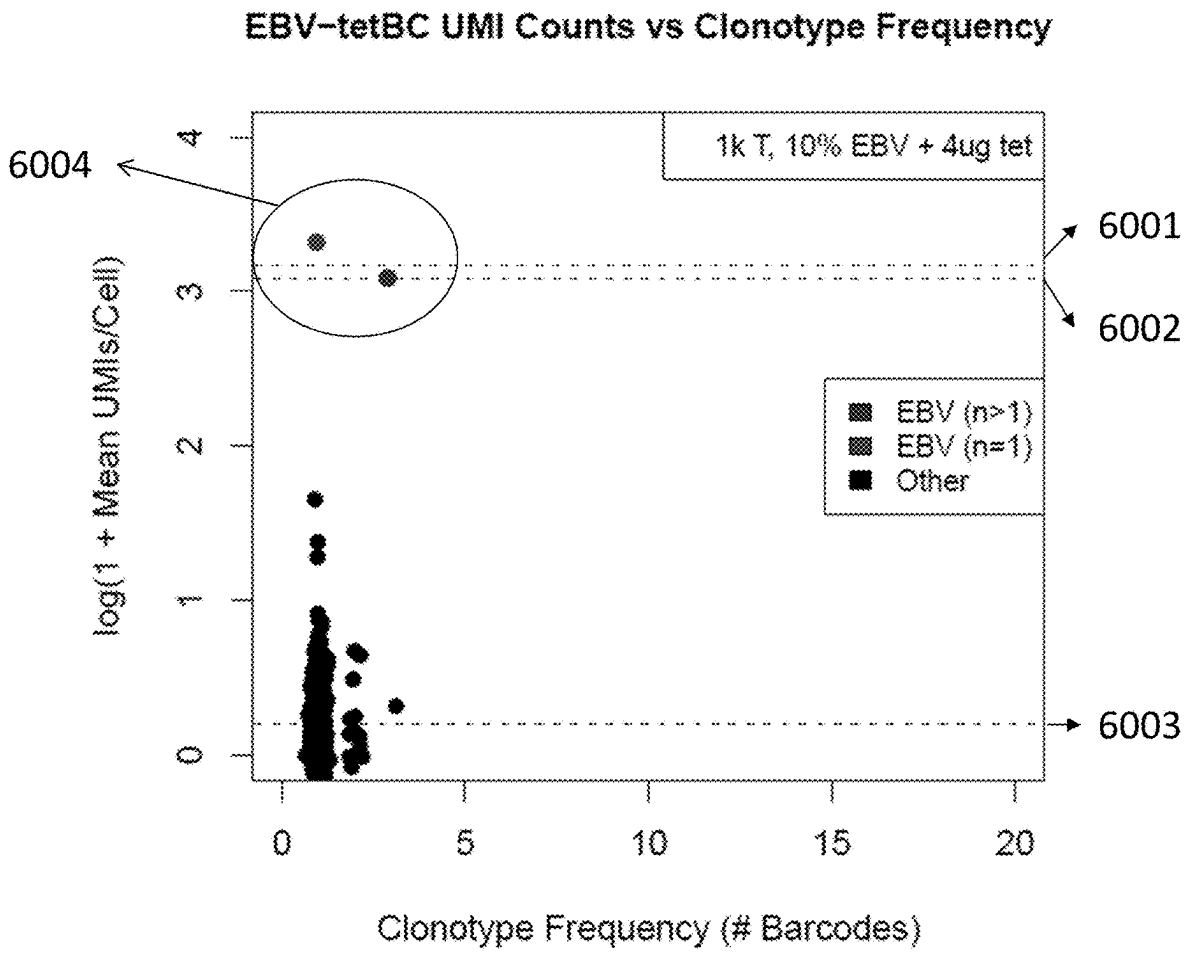

EBV−tetBC UMI Counts vs Clonotype Frequency
1k T, 10% EBV + 4ug tet
6004
6001
6002
6003
EBV (n>1)
EBV (n=1)
Other
log(1 + Mean UMIs/Cell)
0
1
2
3
4
0
5
10
15
20
Clonotype Frequency (# Barcodes)

FIG. 60

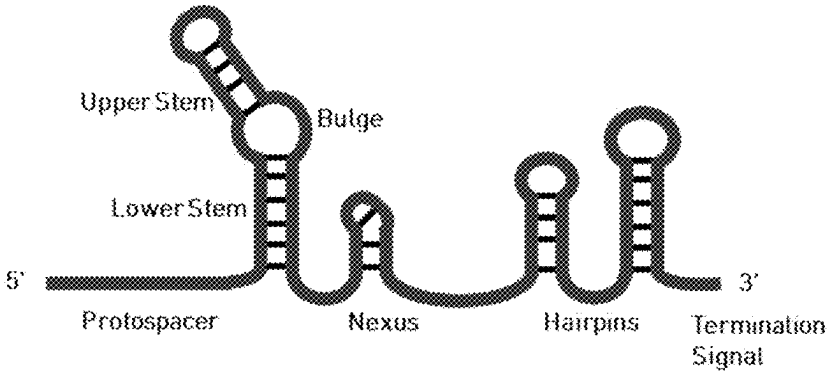


FIG. 61A

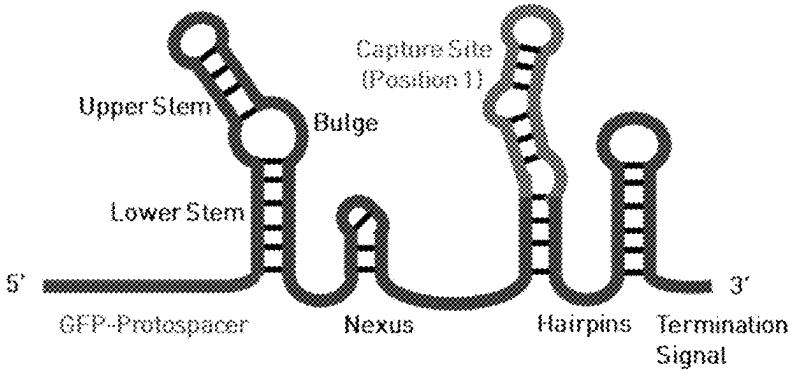


FIG. 61B

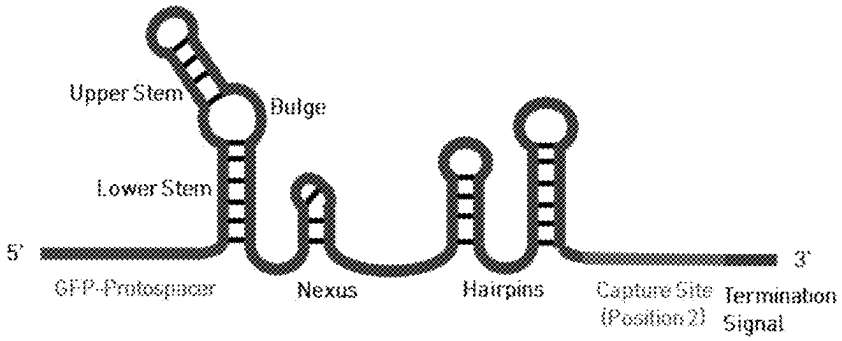


FIG. 61C

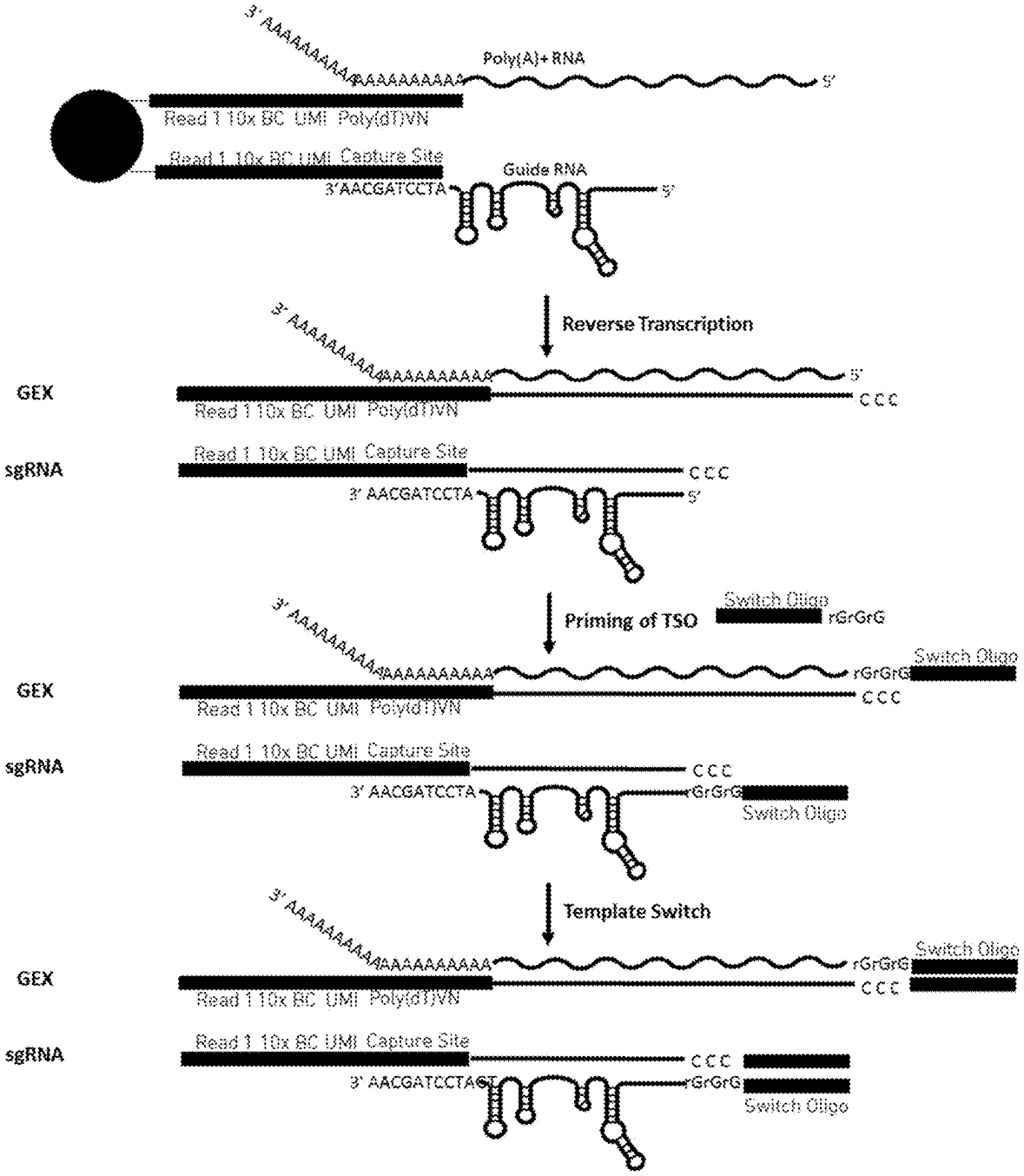

3' AAAAAAAAAAAAAAAAAAAA
Poly(A)+ RNA
5'
Read 1 10x BC UMI Poly(dT)VN
Read 1 10x BC UMI Capture Site
Guide RNA
3'AACGATCCTA
5'
Reverse Transcription
GEX
CCC
sgRNA
Priming of TSO
Switch Oligo
rGrGrG
Template Switch

FIG. 61D

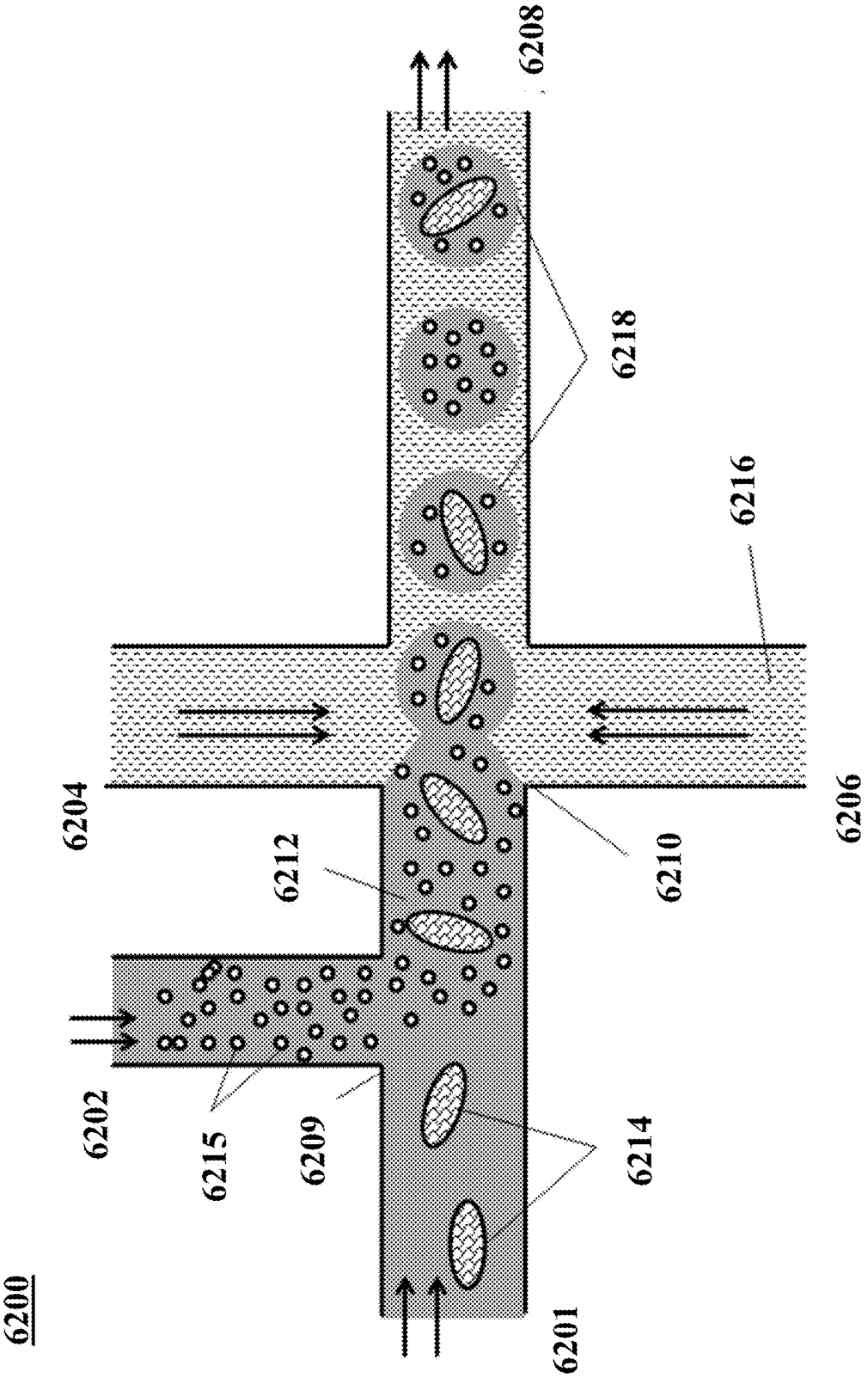

6200
6202
6204
6206
6208
6209
6210
6212
6214
6215
6216
6218
6201

FIG. 62

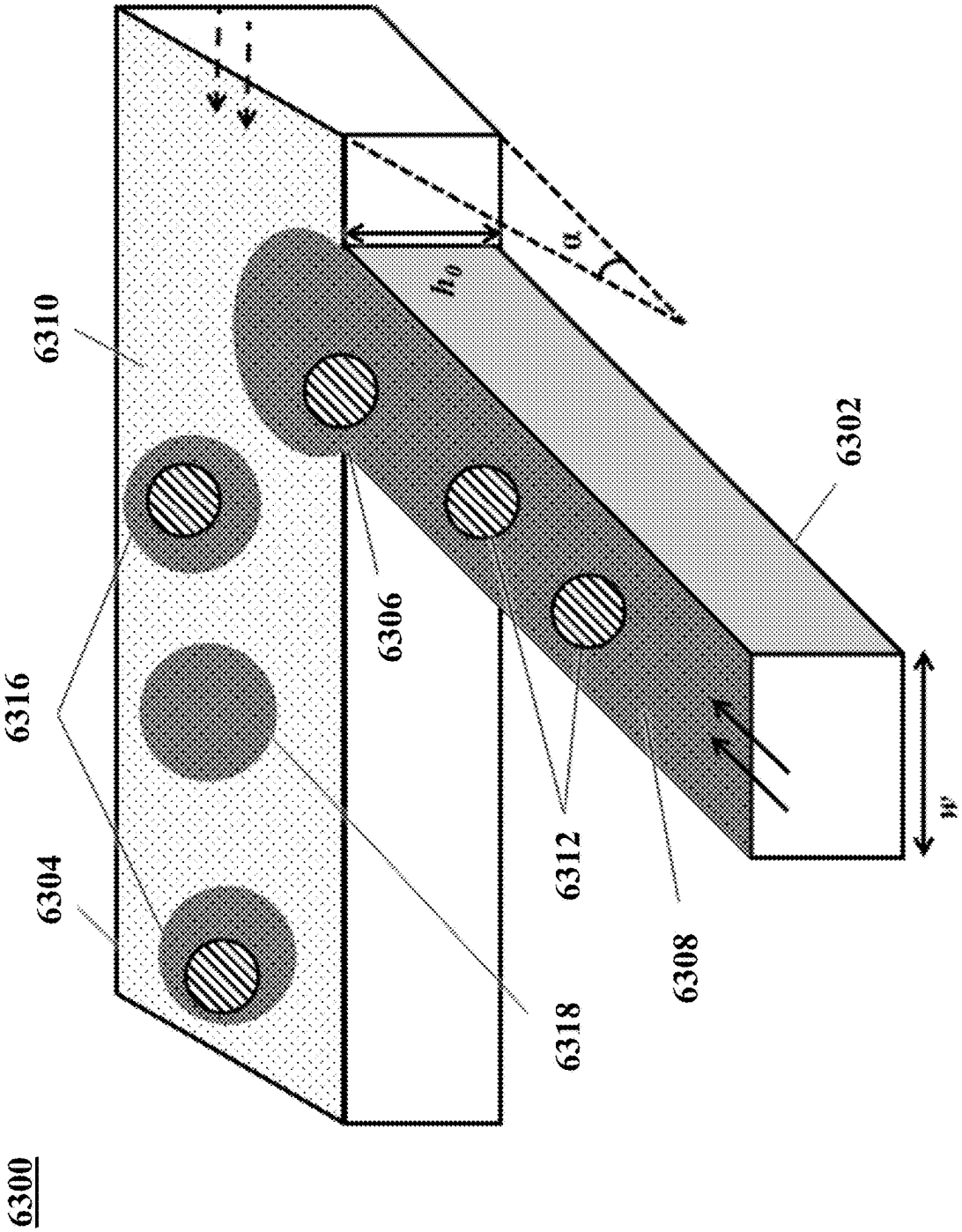


FIG. 63

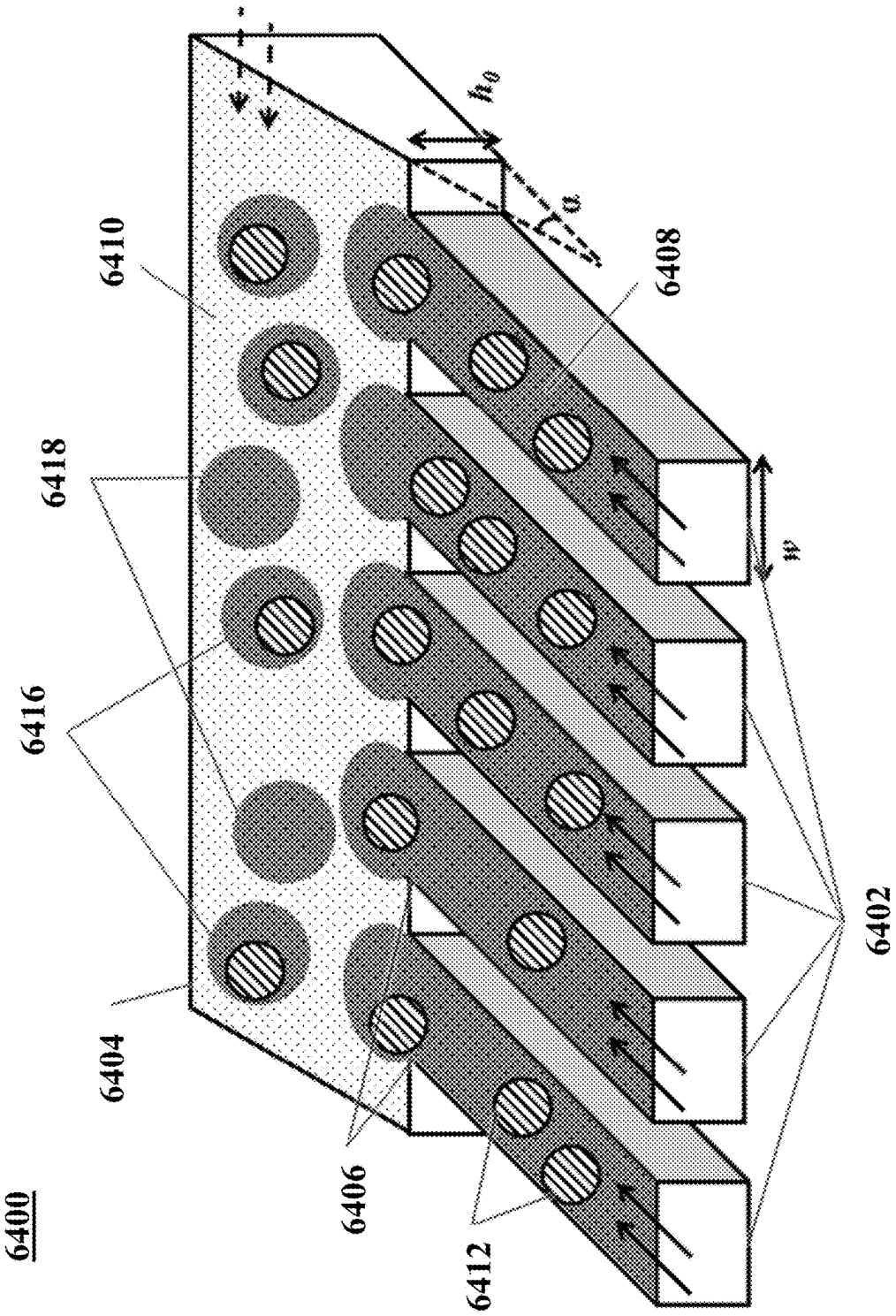


FIG. 64

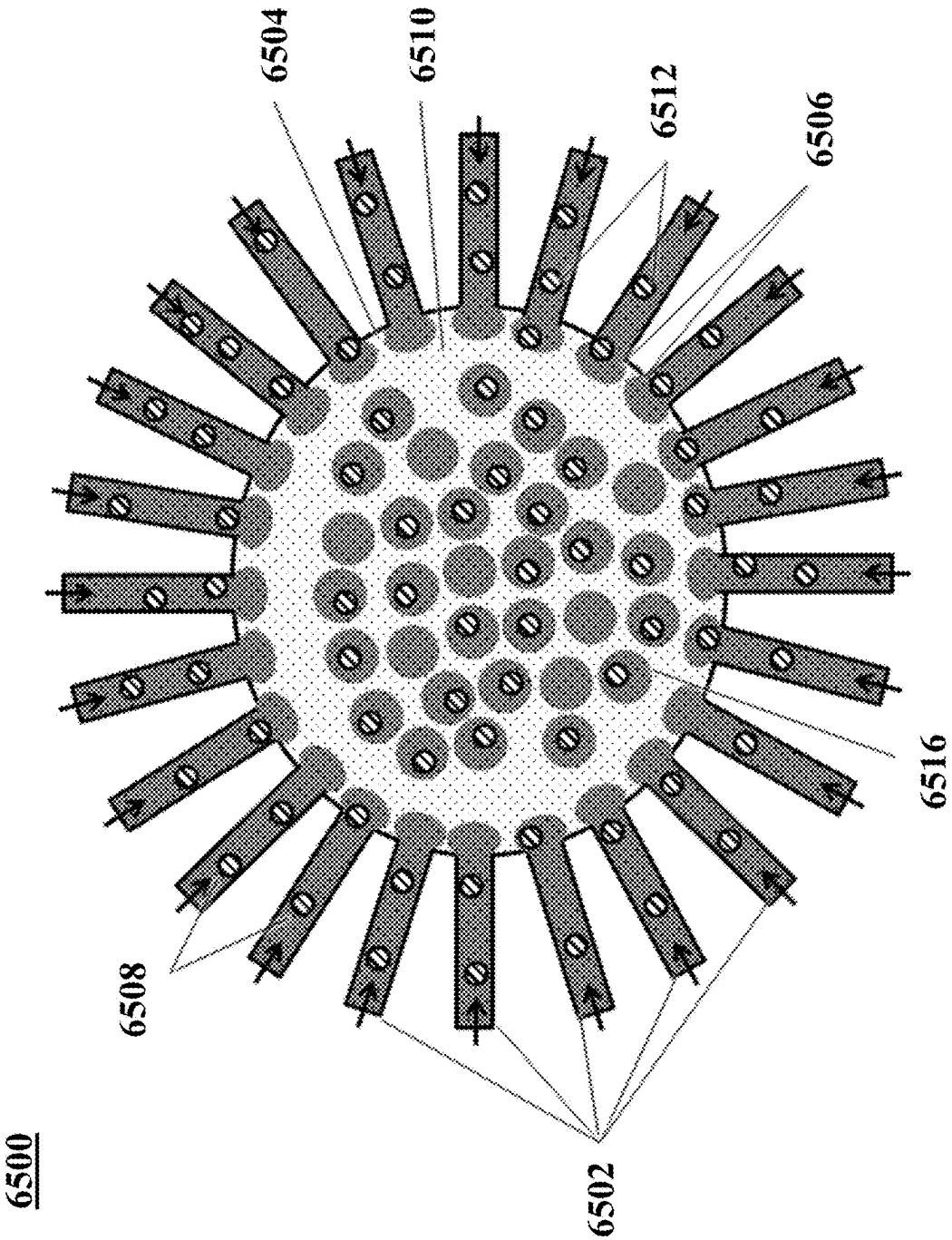


FIG. 65

1	2	3
4	5	6
7	8	9

FIG. 66A

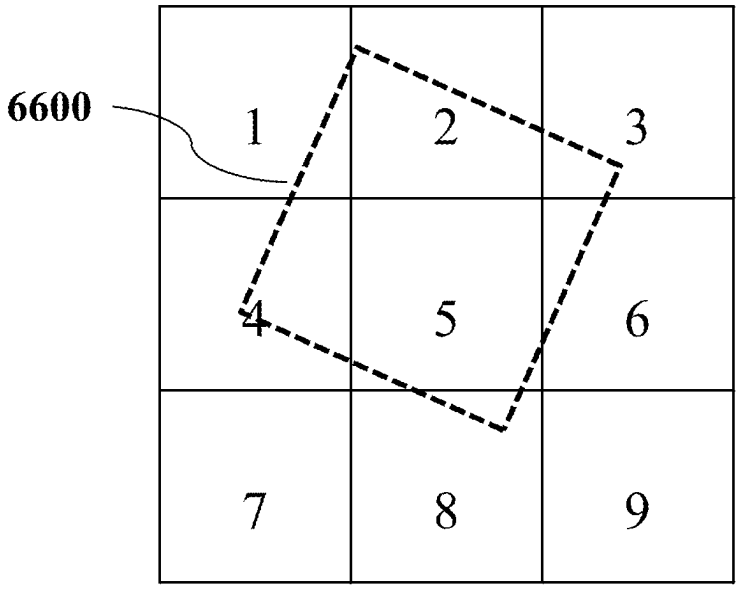


FIG. 66B

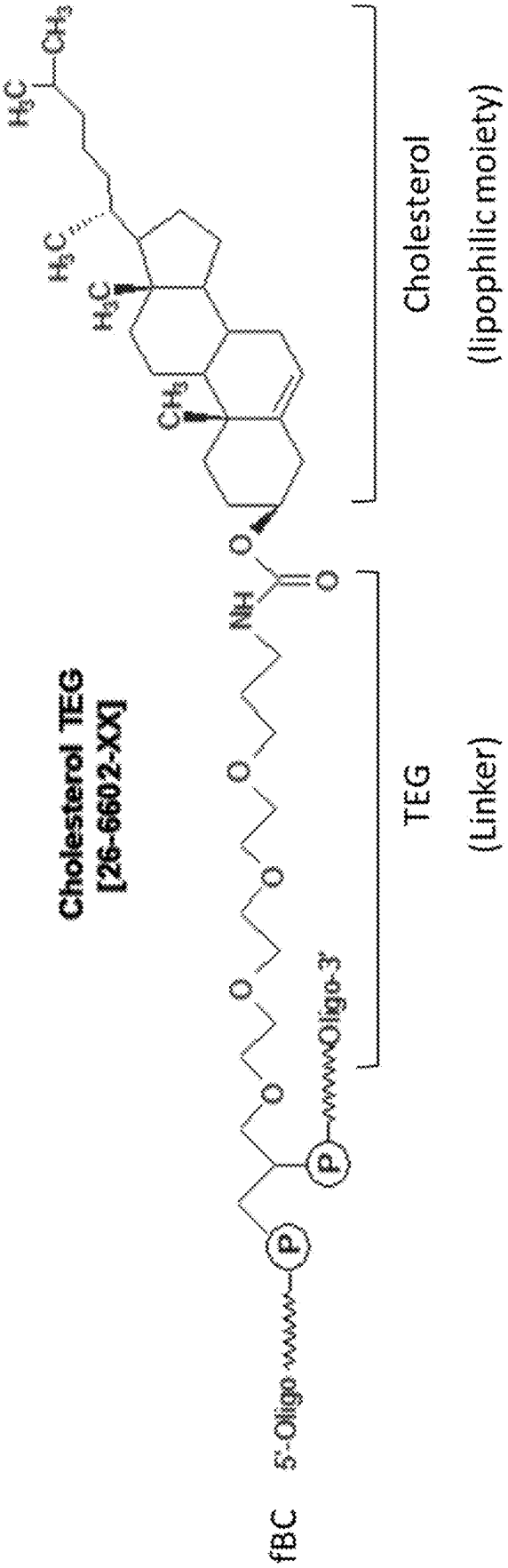

Cholesterol TEG
[26-6602-XX]
fBC
5'-Oligo
Oligo-3'
TEG
(Linker)
Cholesterol
(lipophilic moiety)

FIG. 67

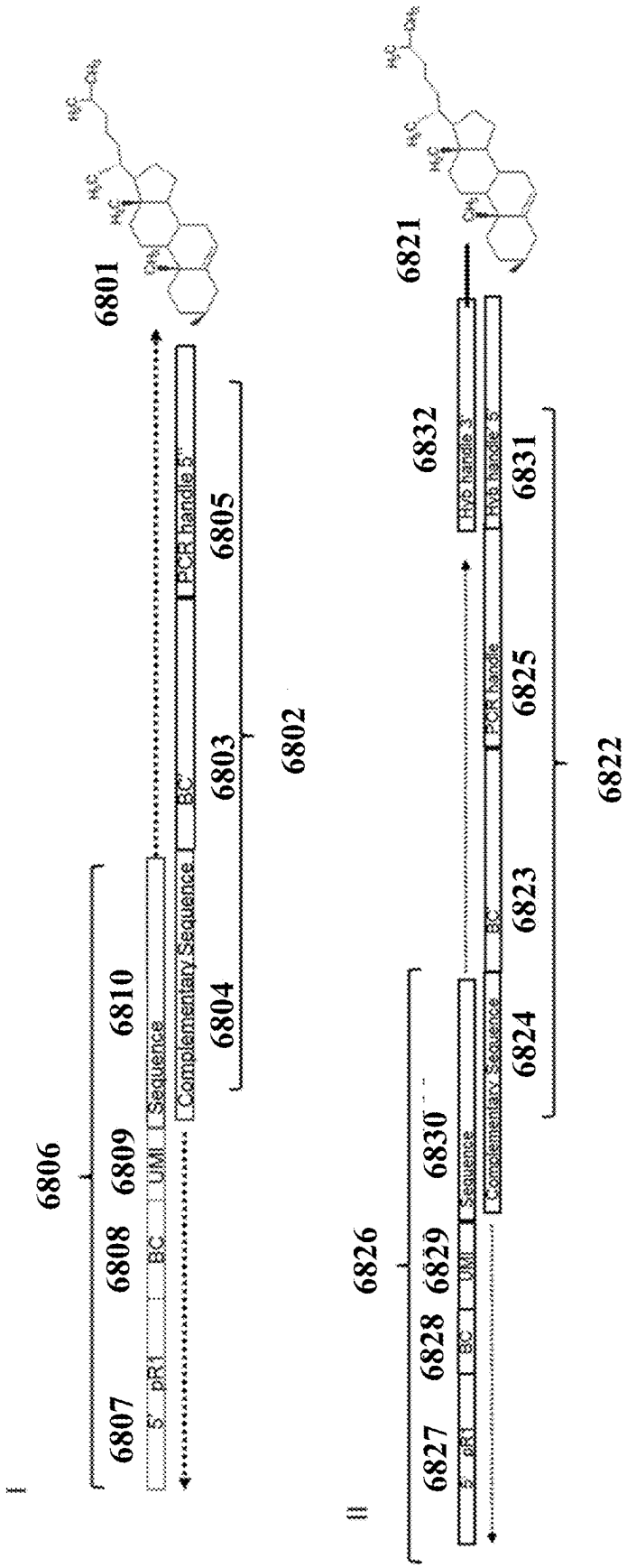

I
6806
6807
6808
6809
6810
5'
pRT
BC
UMI
Sequence
Complementary Sequence
BC
PCR handle 5'
6801
6802
6803
6804
6805
II
6826
6827
6828
6829
6830
6821
6822
6823
6824
6825
6831
6832
PCR handle

FIG. 68

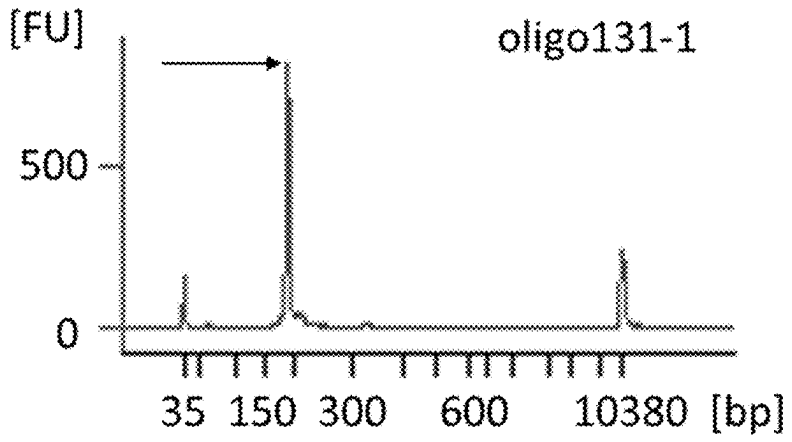


FIG. 69A

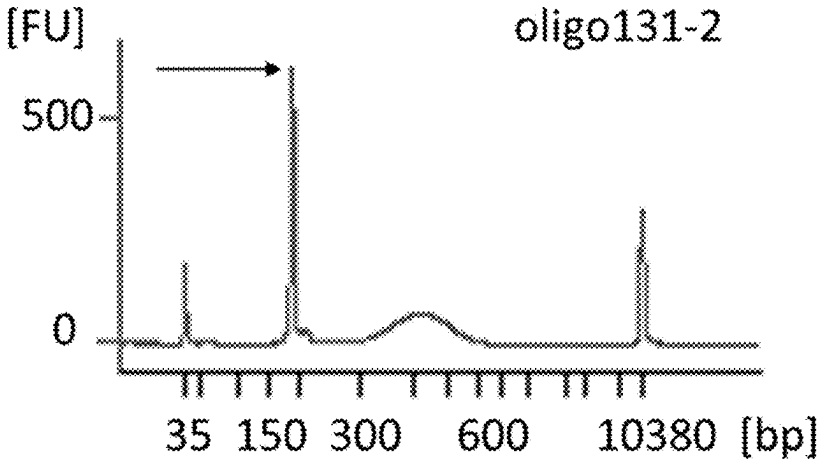


FIG. 69B

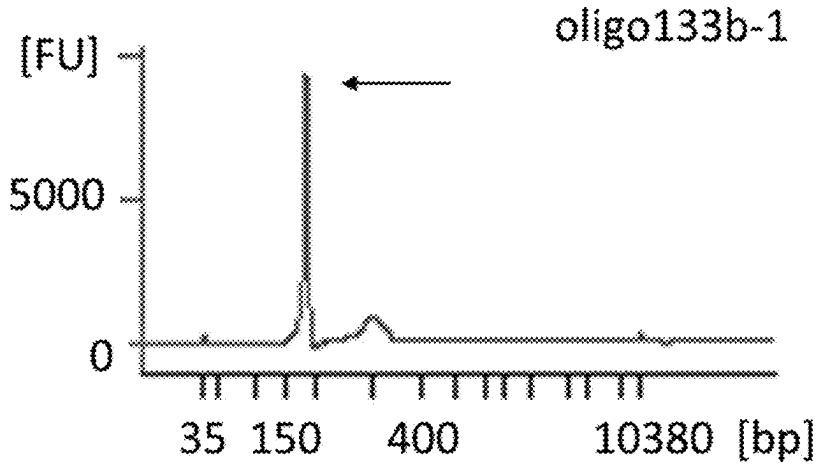


FIG. 69C

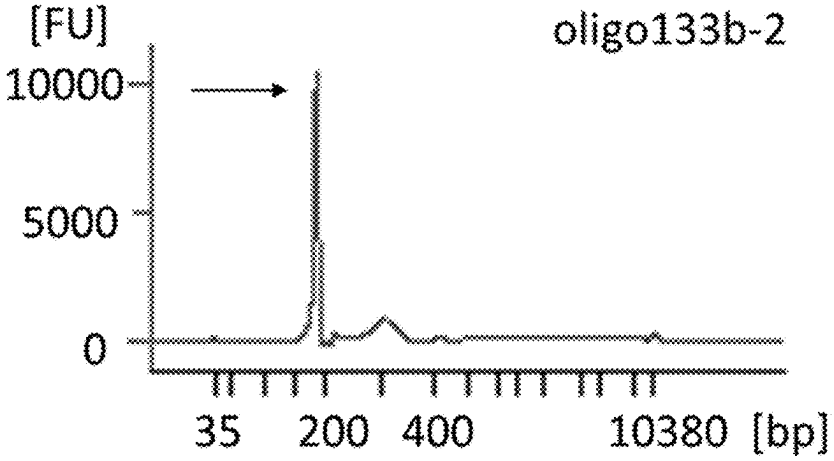


FIG. 69D

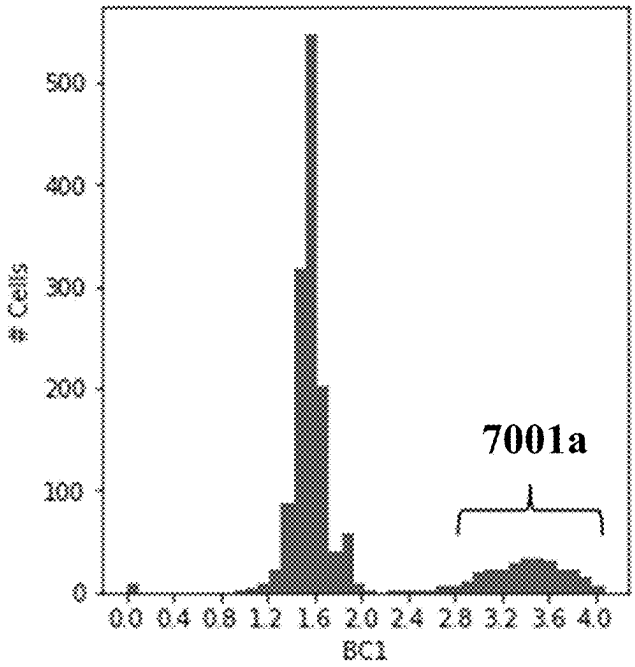


FIG. 70A

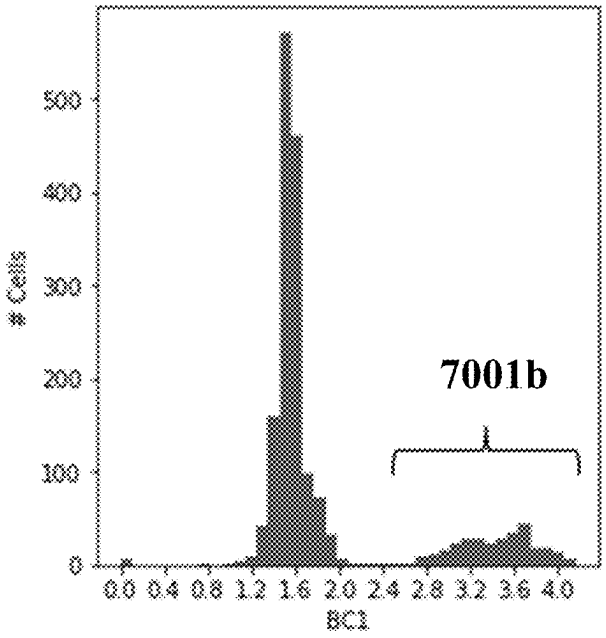


FIG. 70B

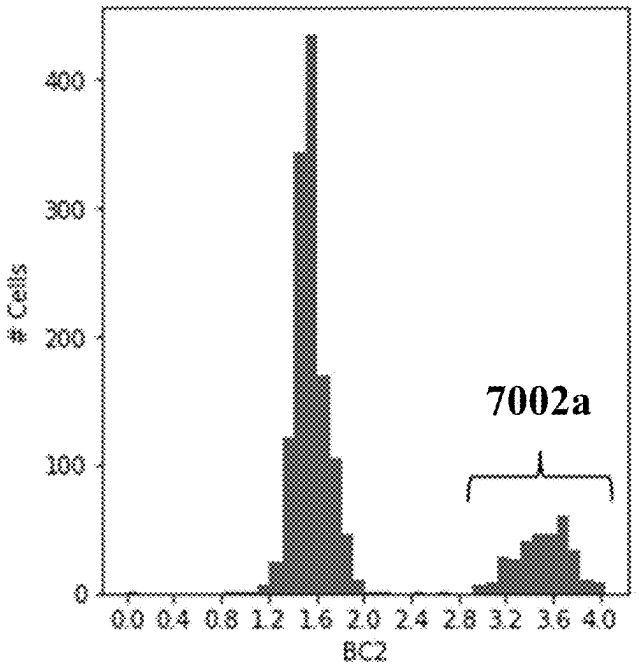


FIG. 70C

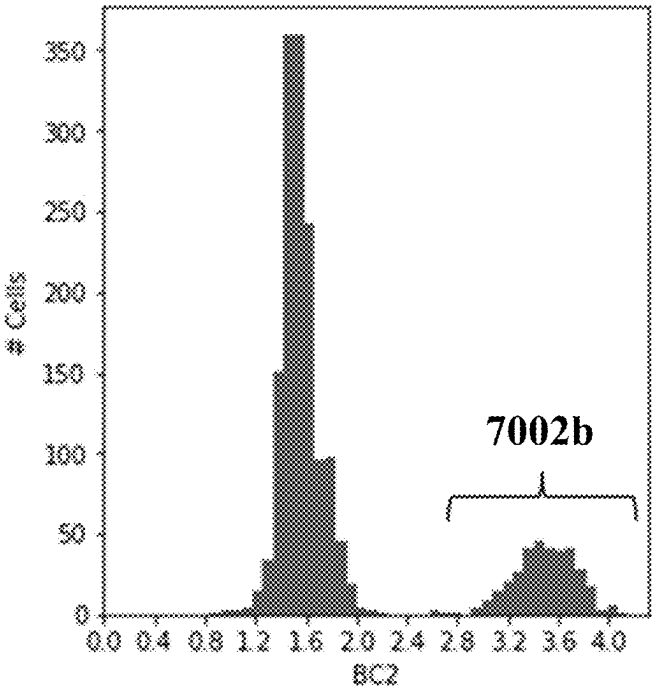


FIG. 70D

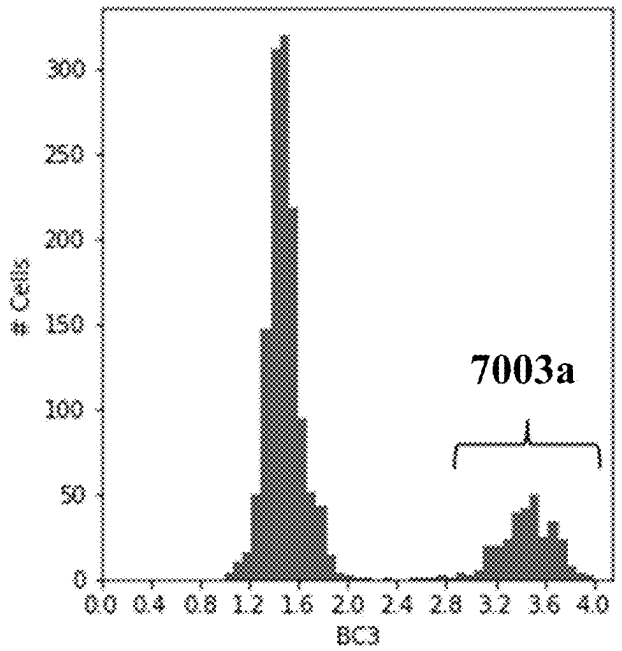


FIG. 70E

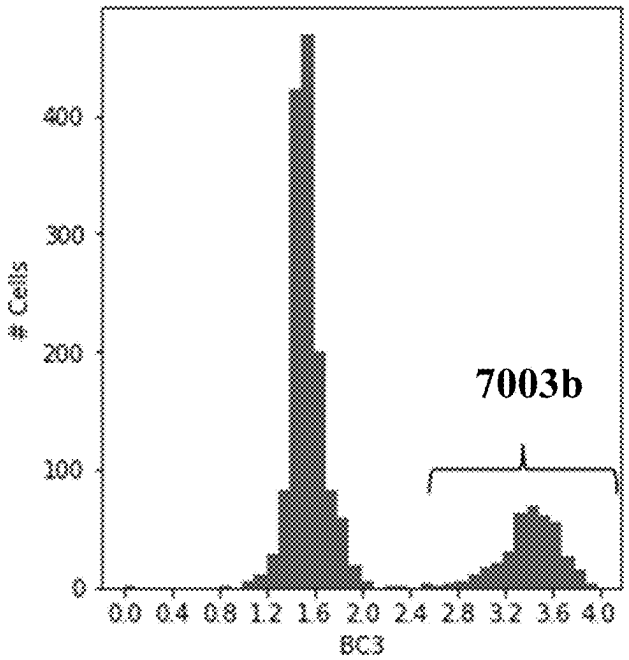


FIG. 70F

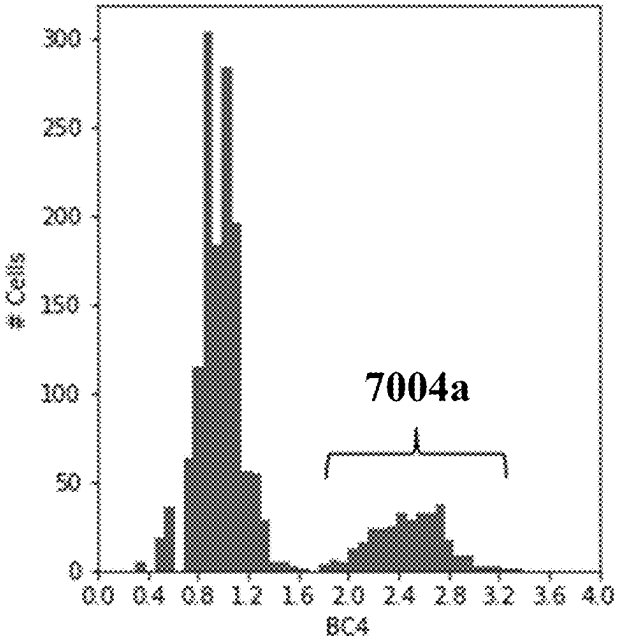


FIG. 70G

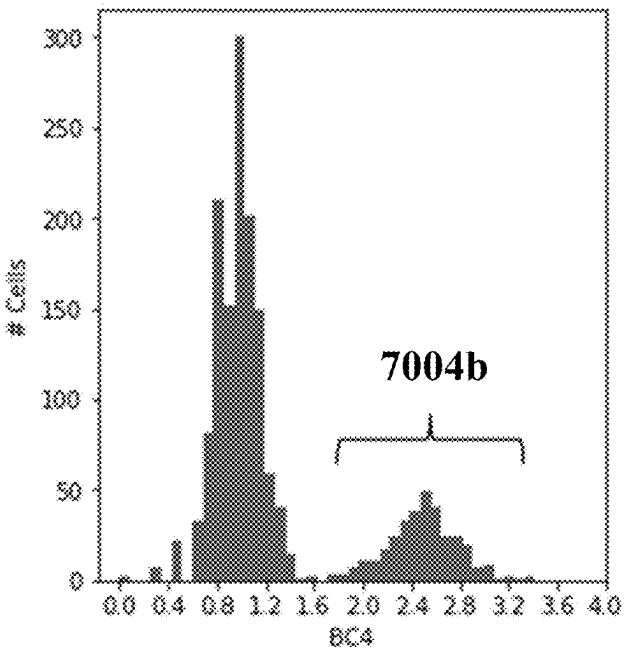


FIG. 70H

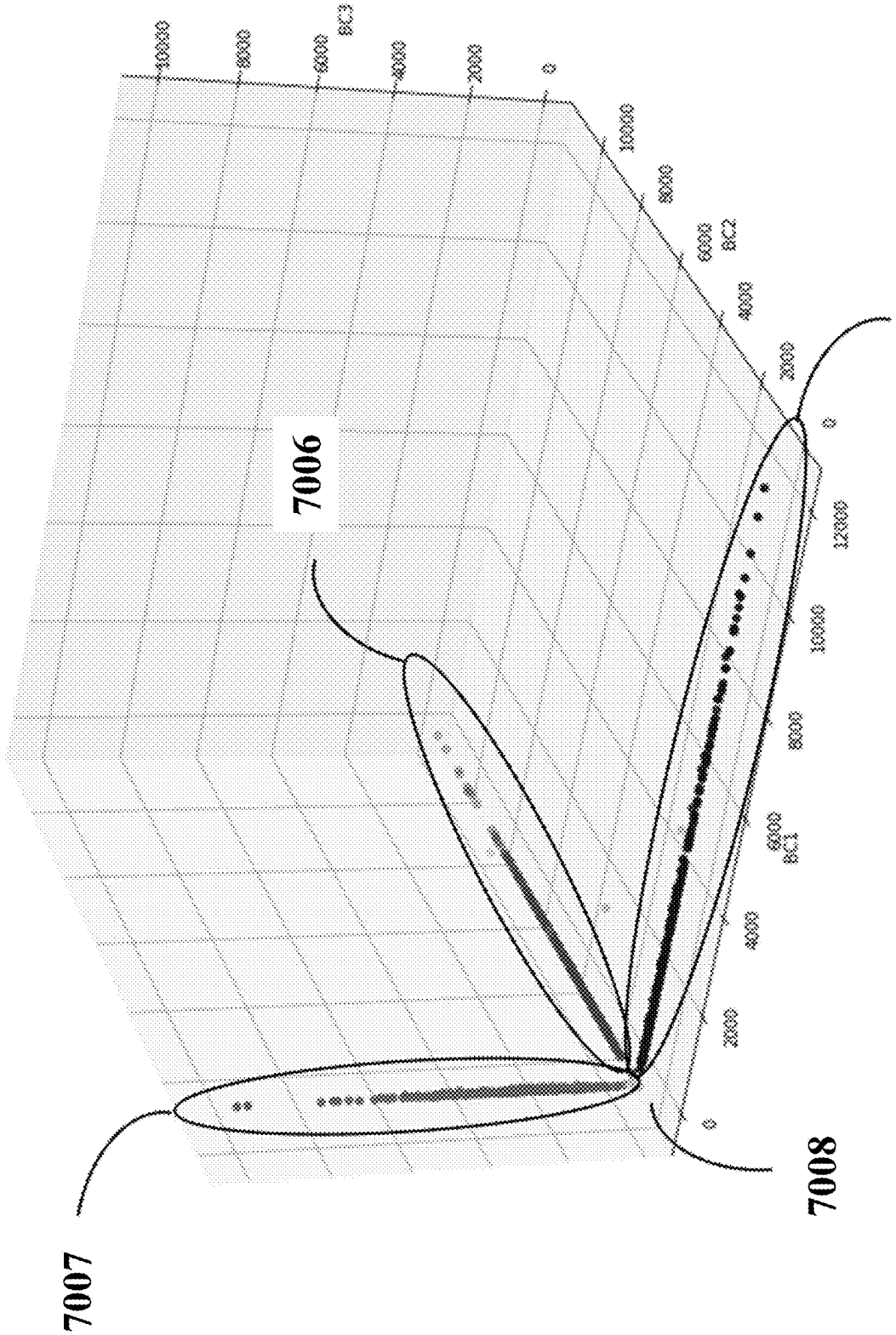

7005
7006
7007
7008
BC1
BC2
BC3
0
2000
4000
6000
8000
10000
12000

FIG. 70I

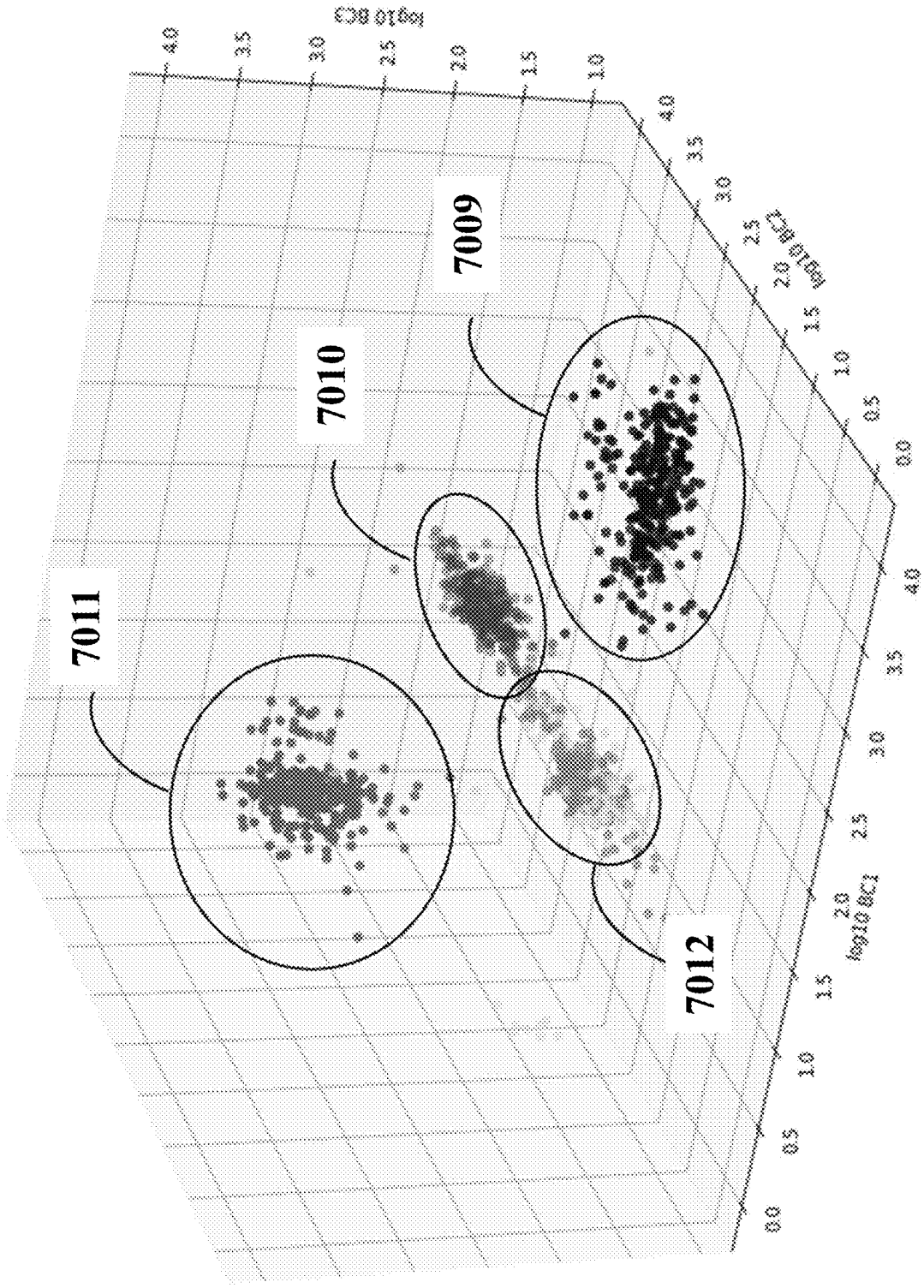

7011
7010
7009
7012
log10 BC1
log10 BC2
log10 BC3
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0

FIG. 70J

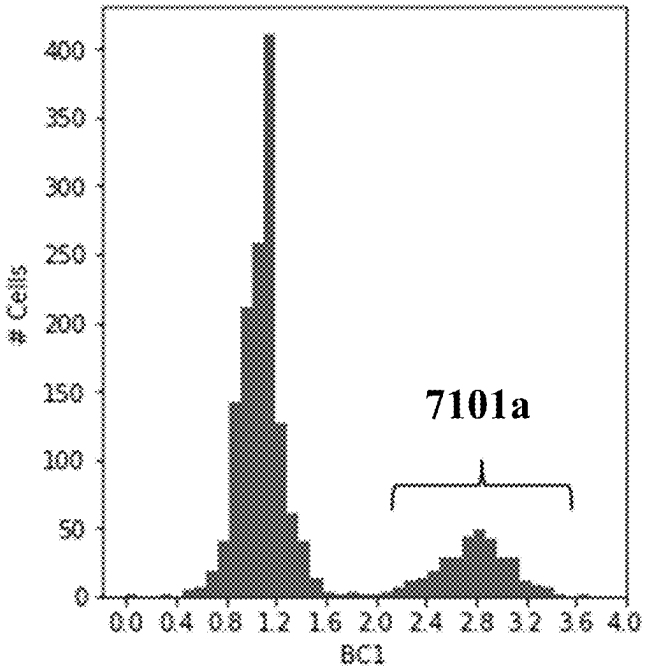


FIG. 71A

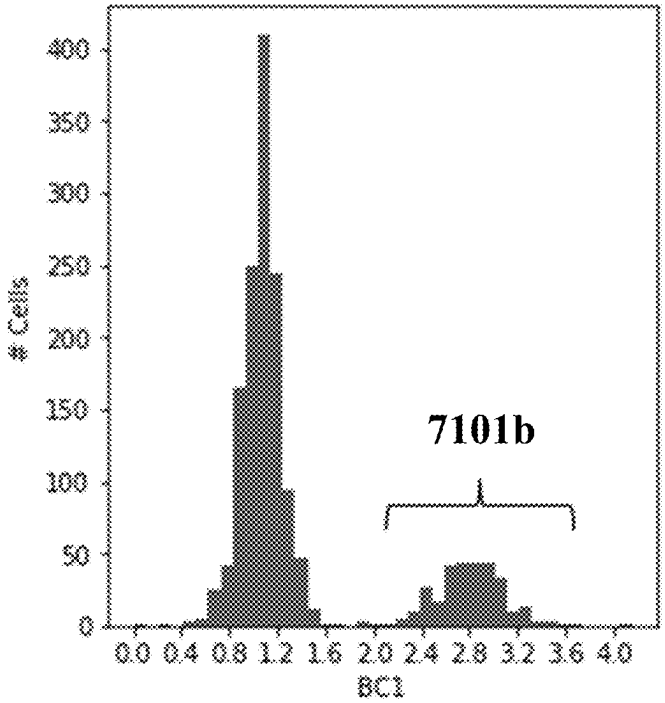


FIG. 71B

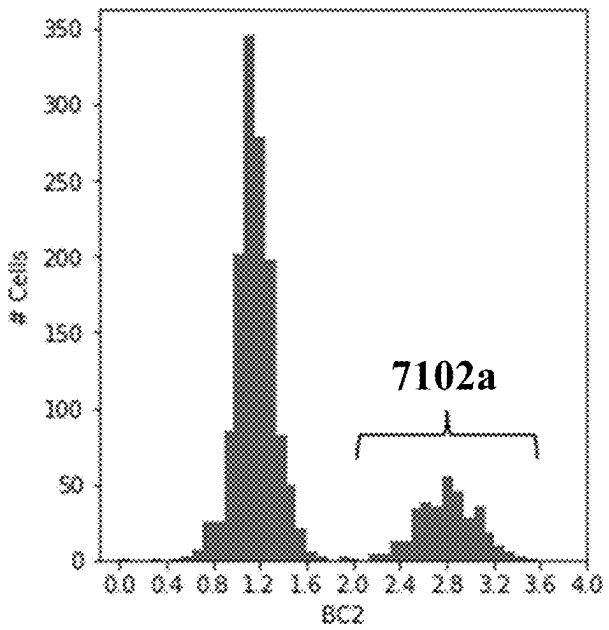


FIG. 71C

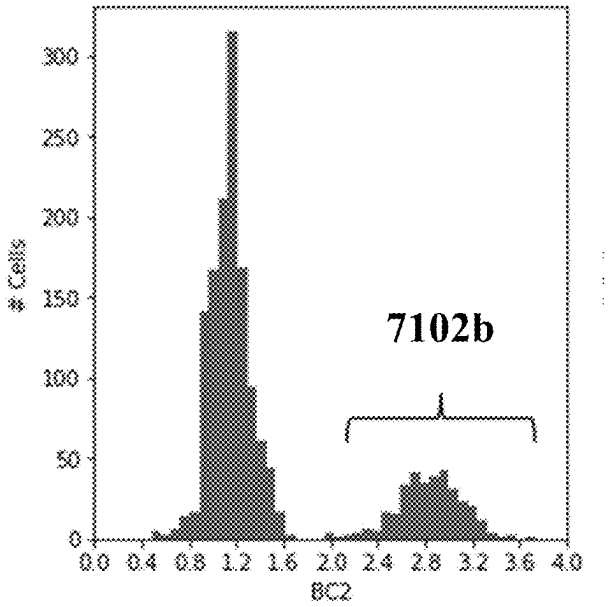


FIG. 71D

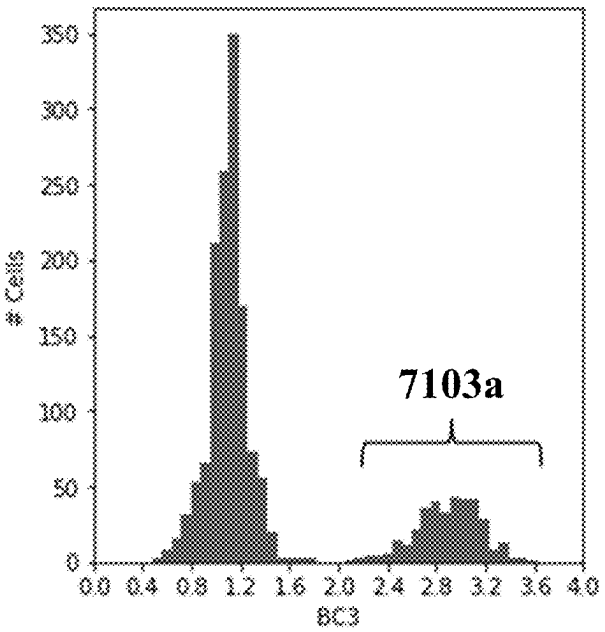


FIG. 71E

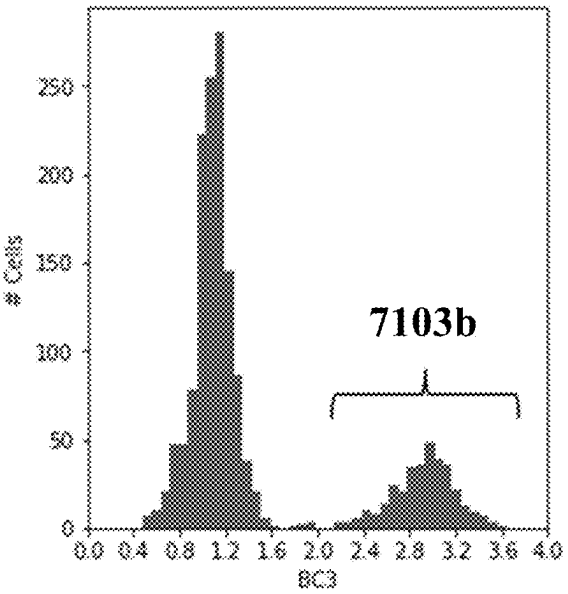


FIG. 71F

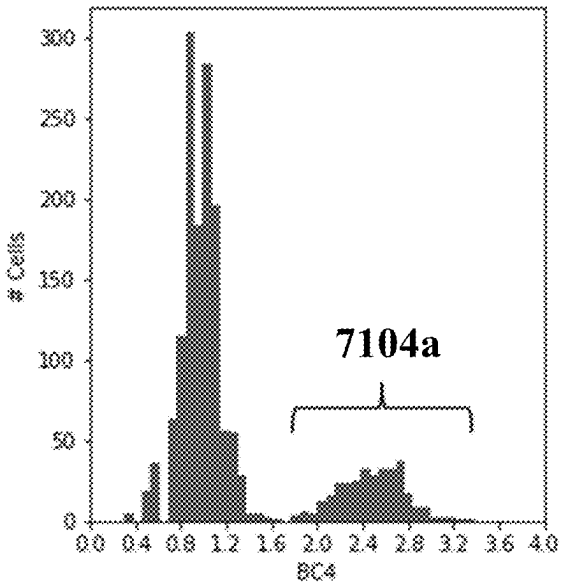


FIG. 71G

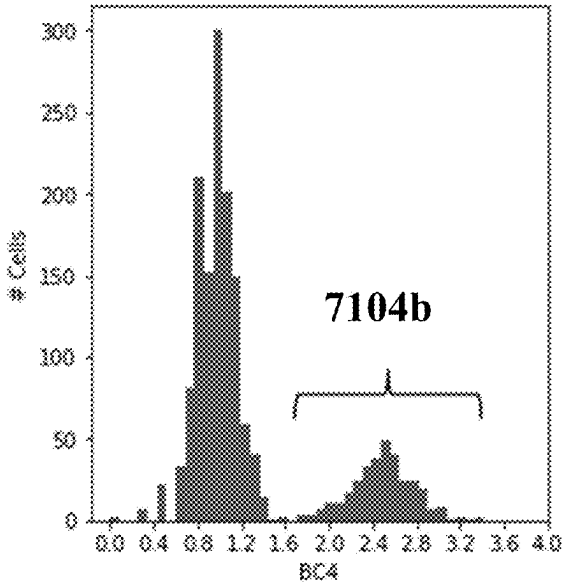


FIG. 71H

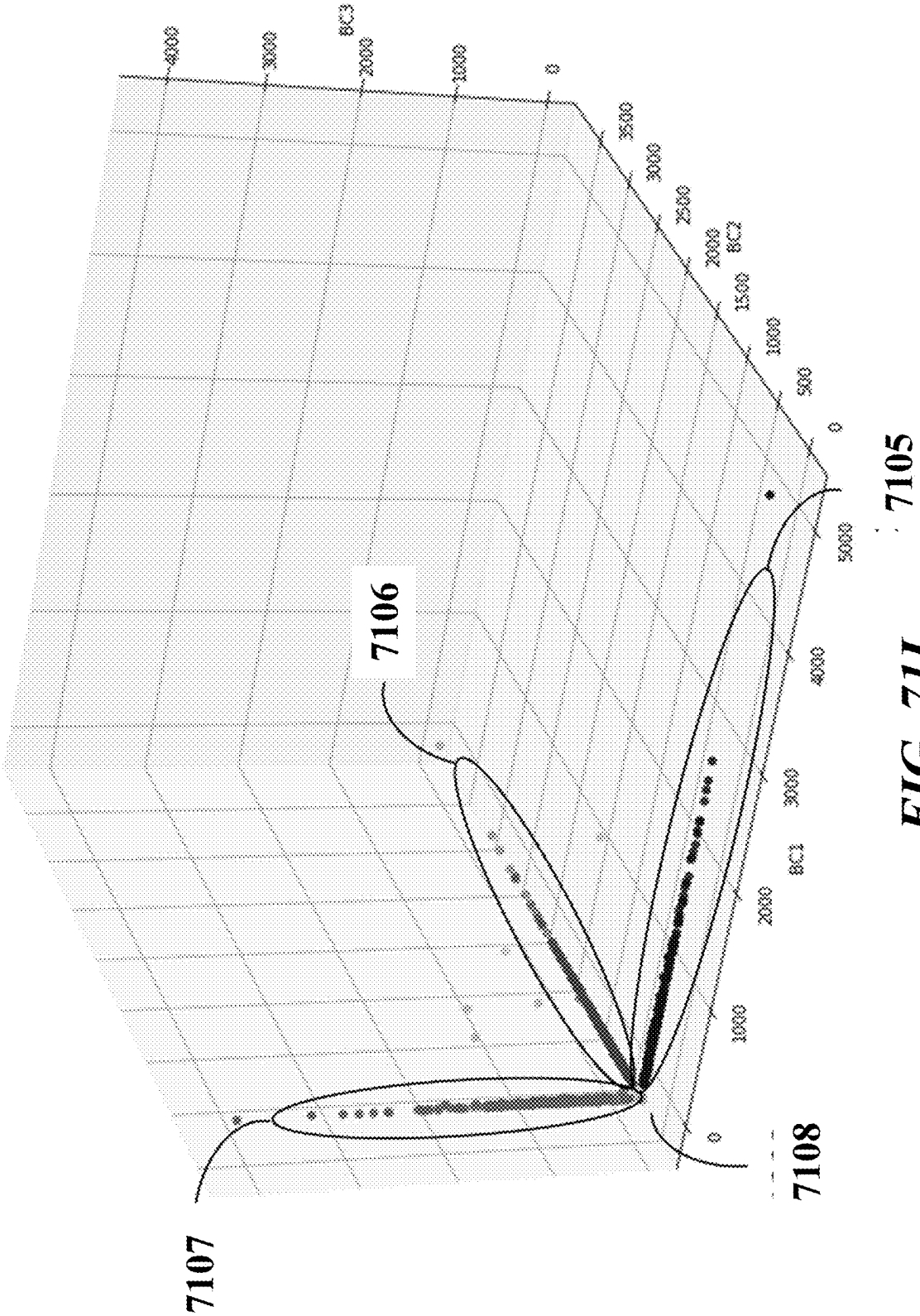

BC3
4000
3500
3000
2500
2000
1500
1000
500
0
BC2
BC1
7105
7106
7107
7108

FIG. 71I

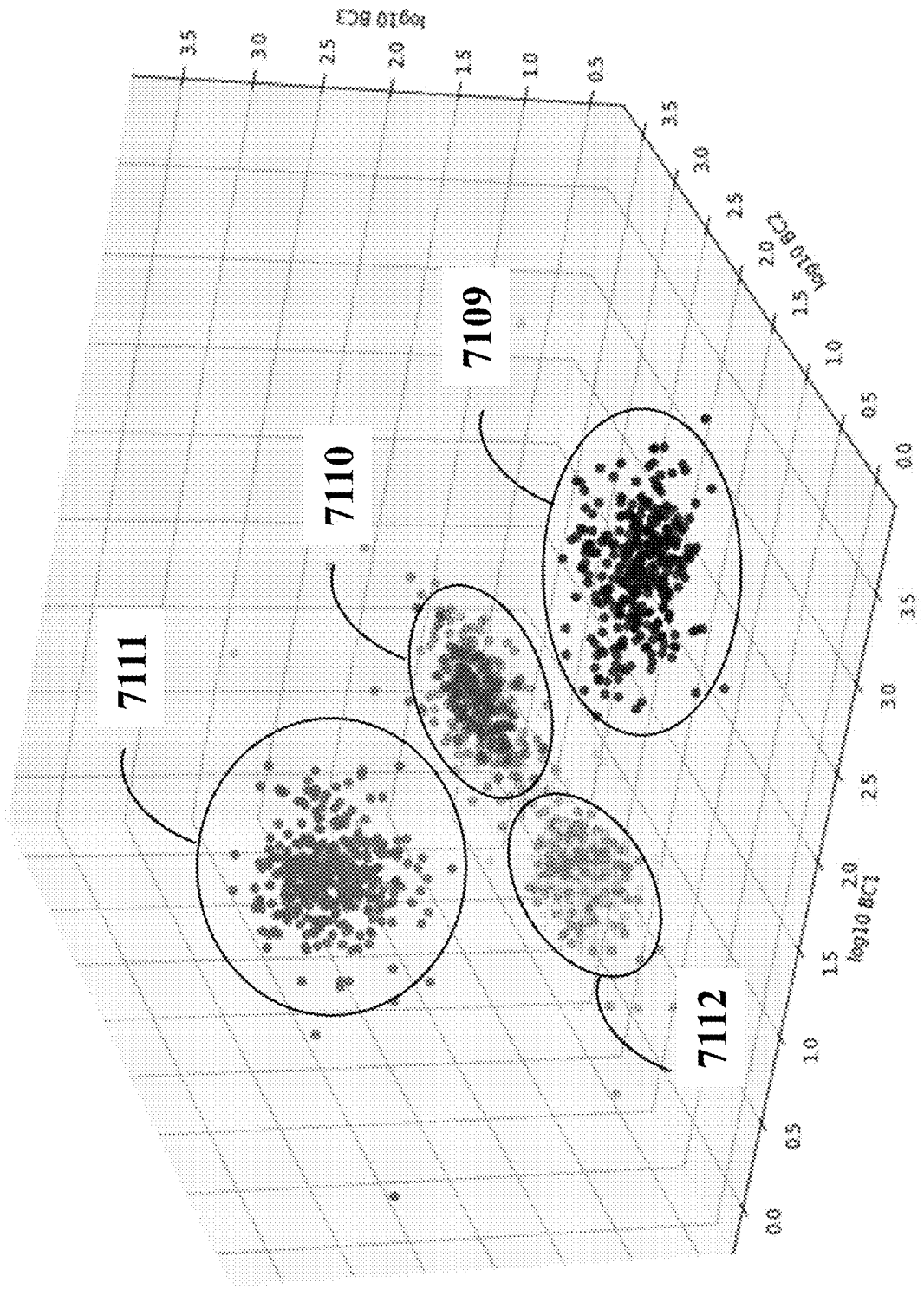

7111
7110
7109
7112
log10 BC3
log10 BC2
log10 BC1
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5

FIG. 71J

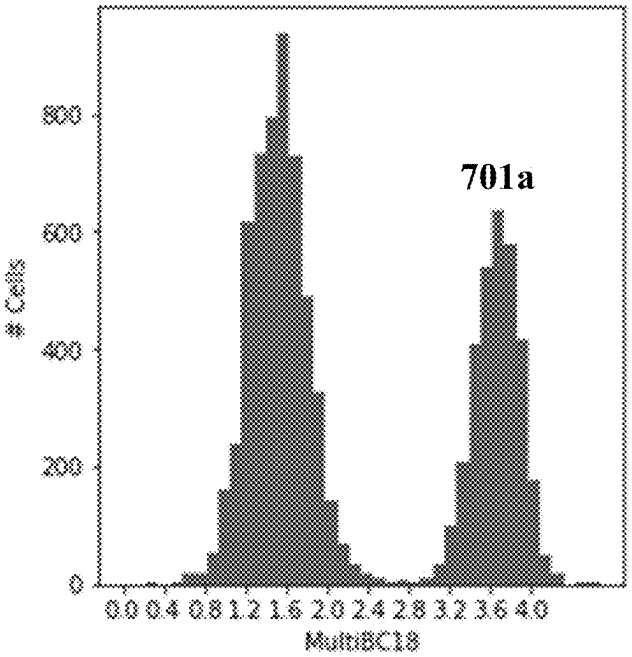


FIG. 72A

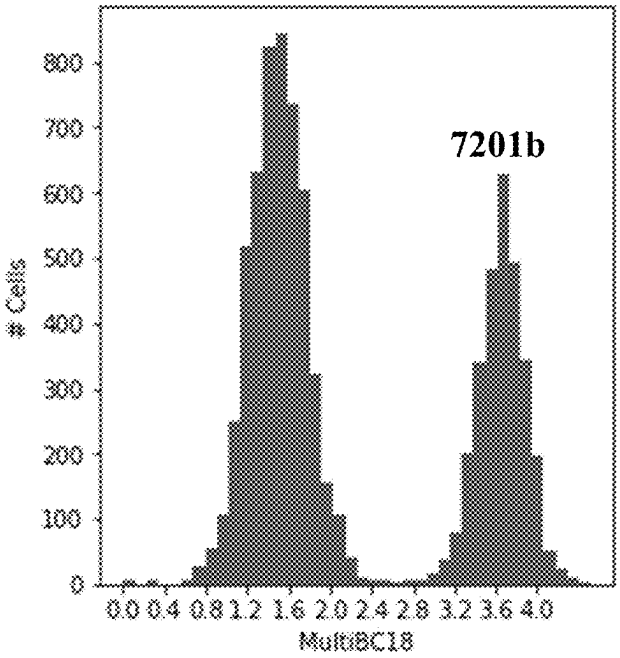


FIG. 72B

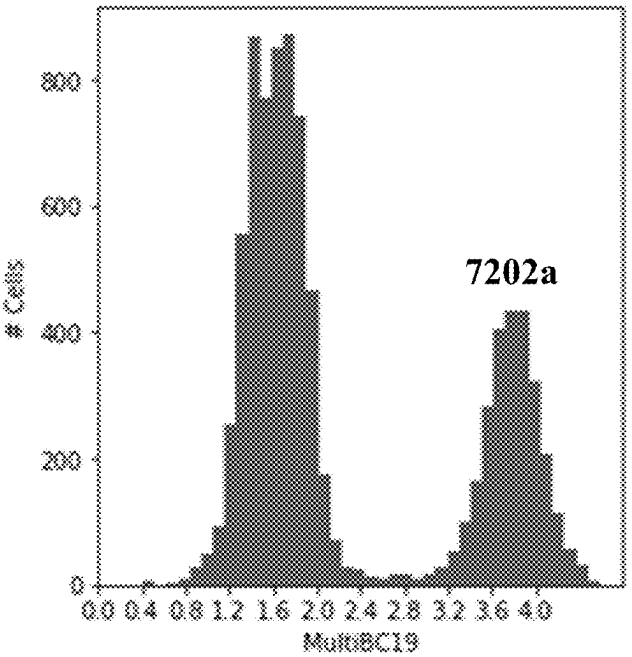


FIG. 72C

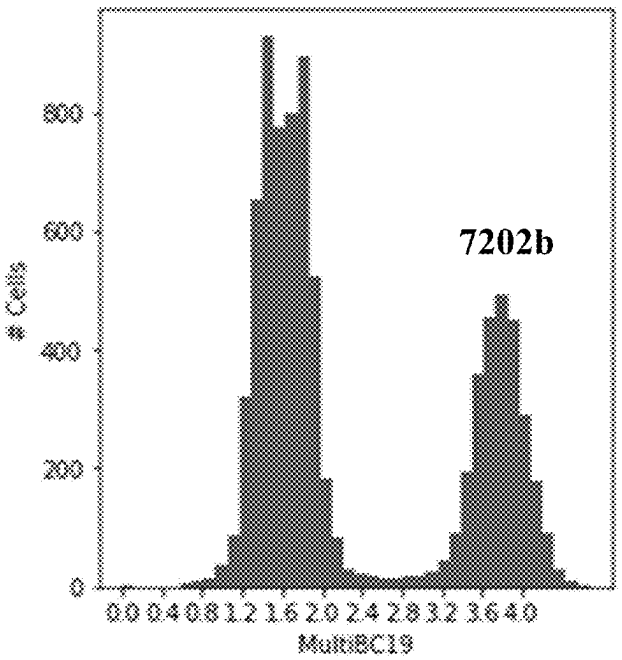


FIG. 72D

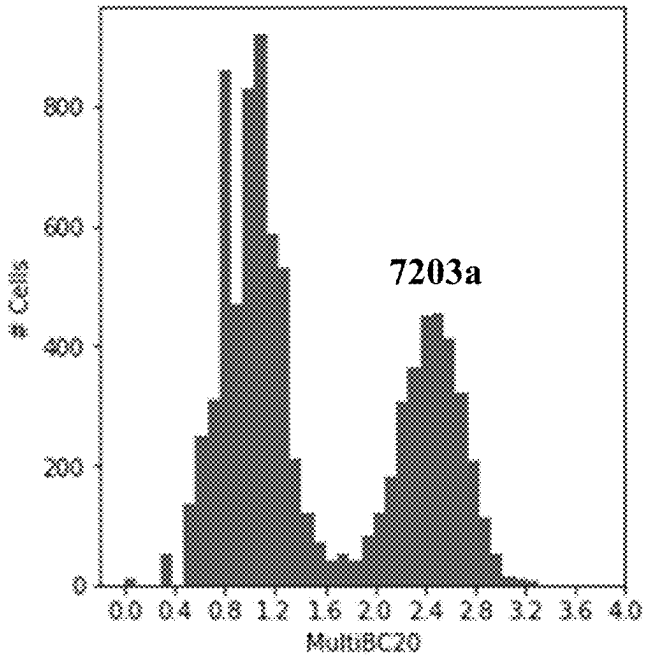


FIG. 72E

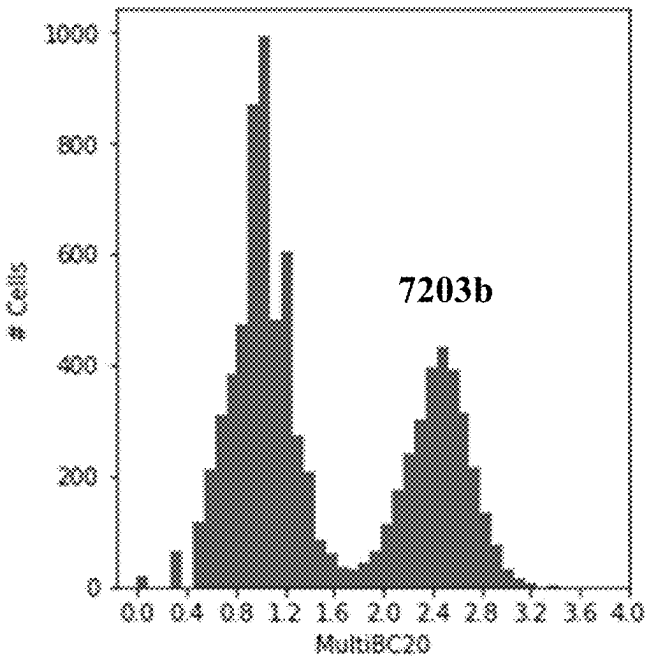


FIG. 72F

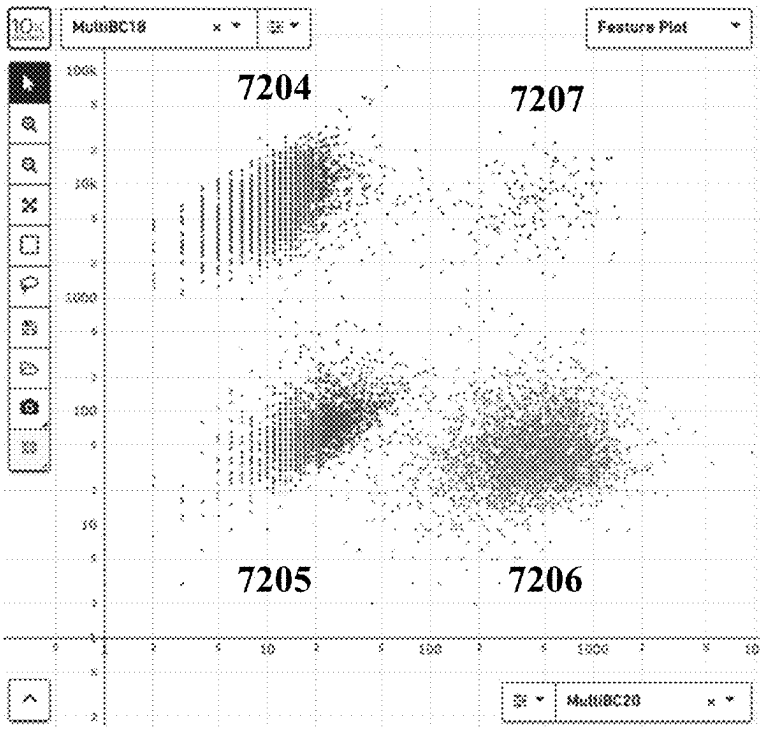


FIG. 72G

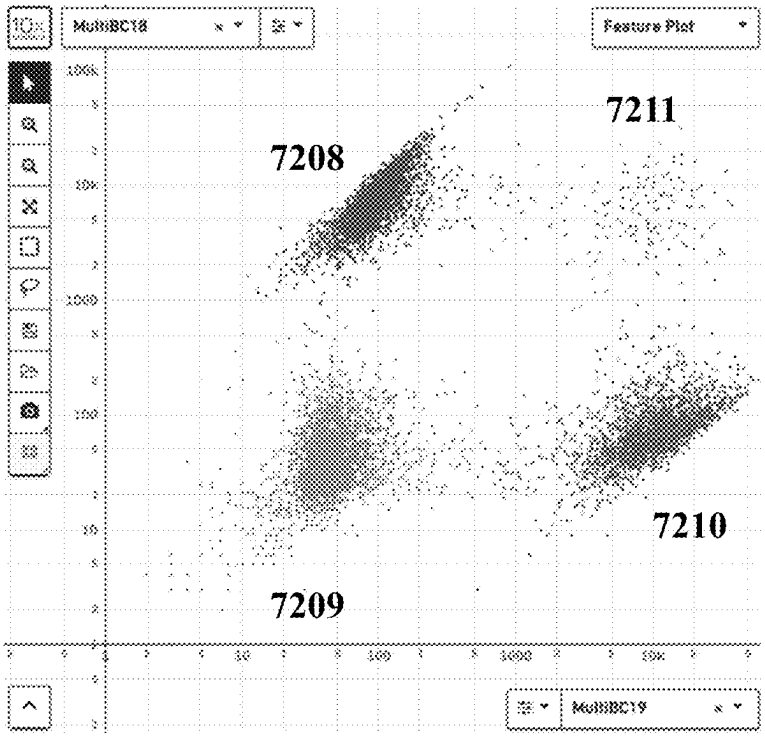


FIG. 72H

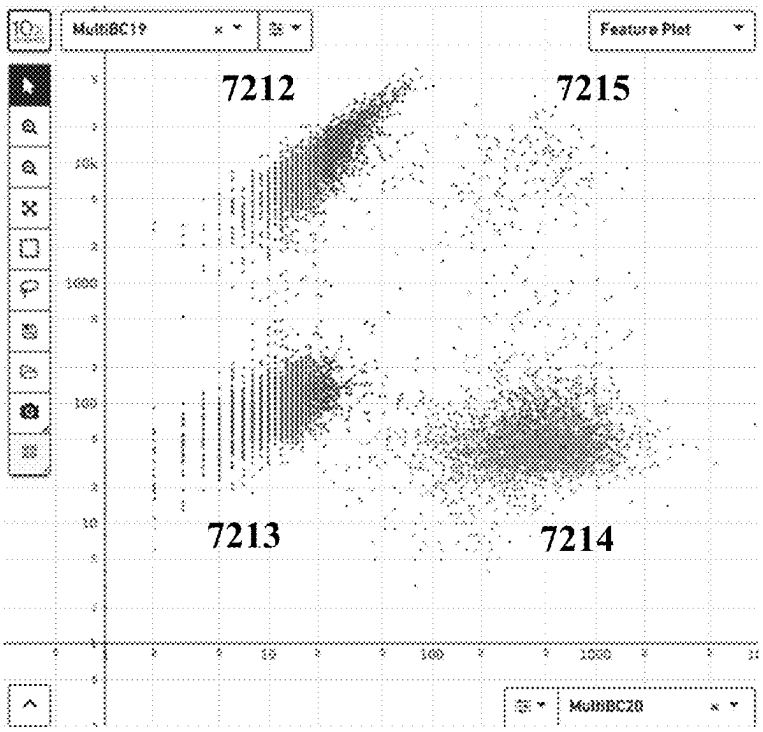


FIG. 72I

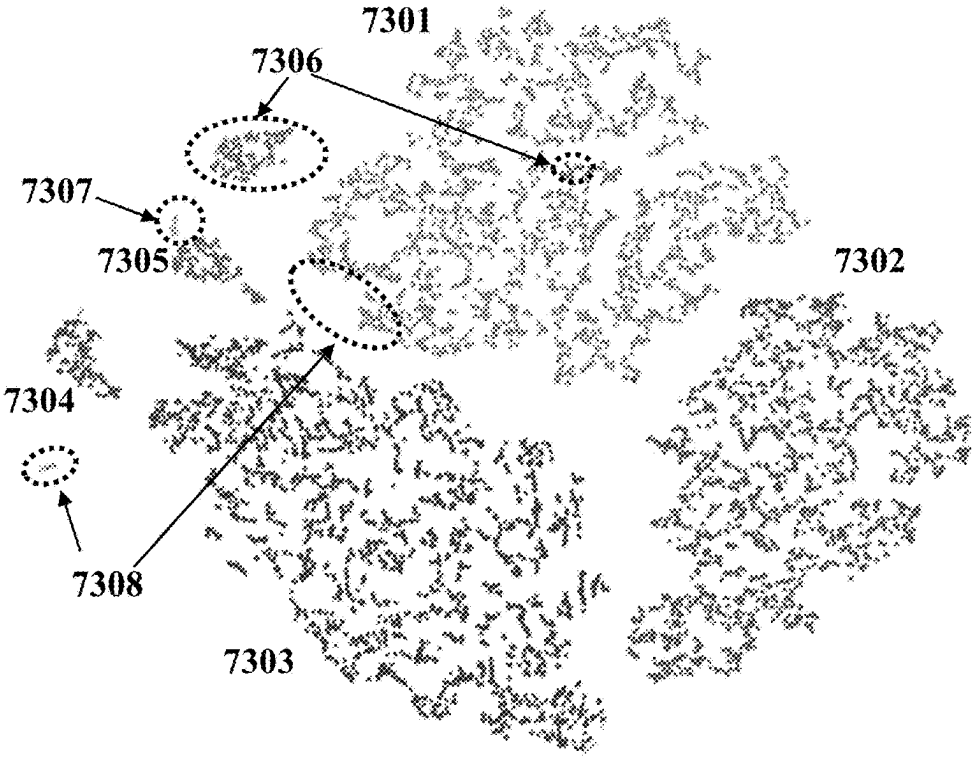


FIG. 73A

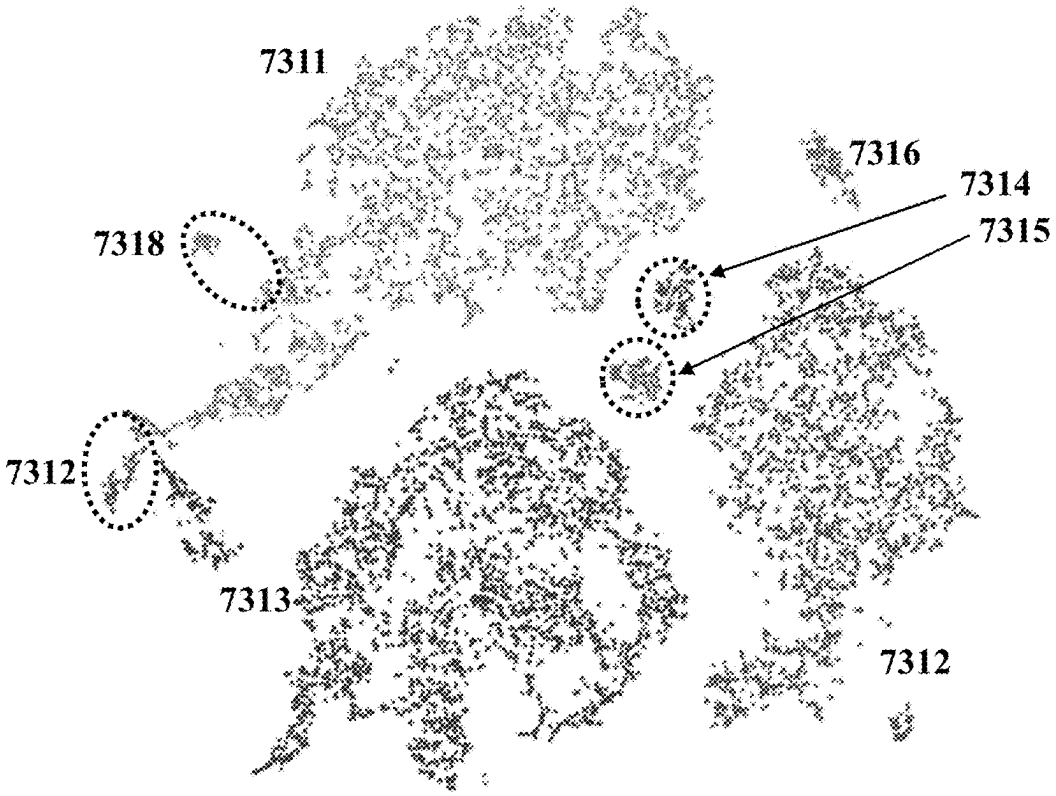


FIG. 73B

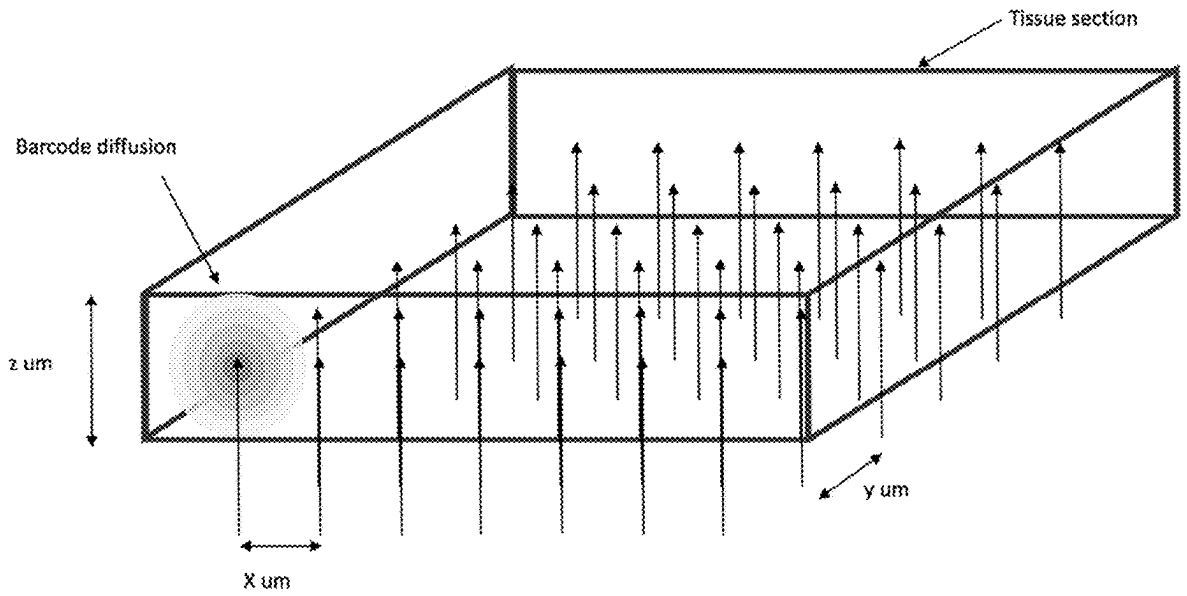


FIG. 74

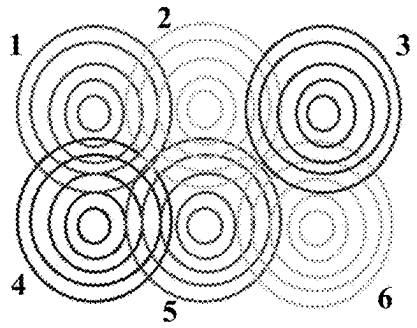


FIG. 75

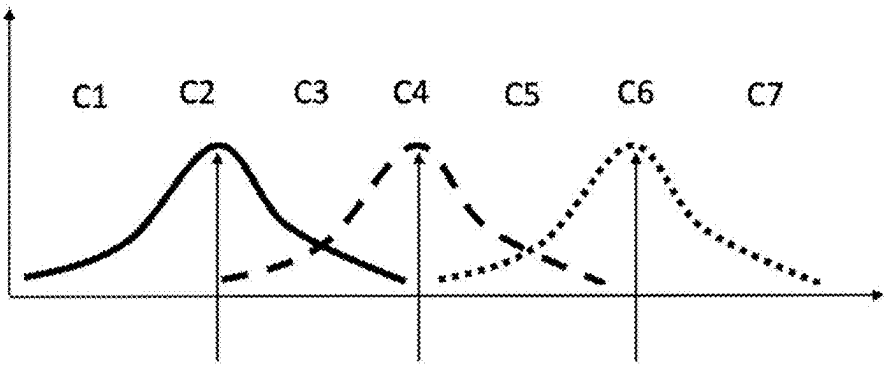


FIG. 76

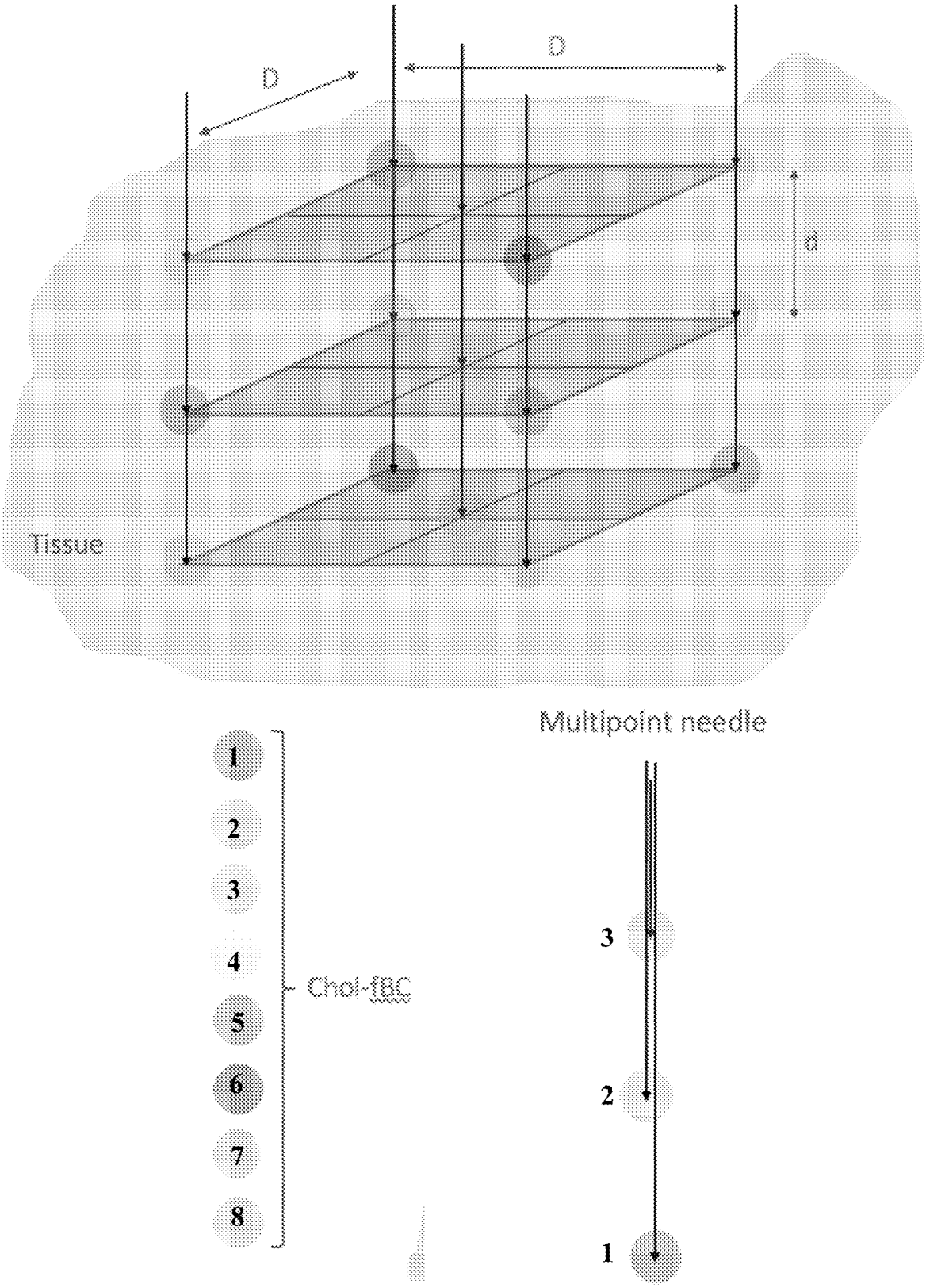


FIG. 77

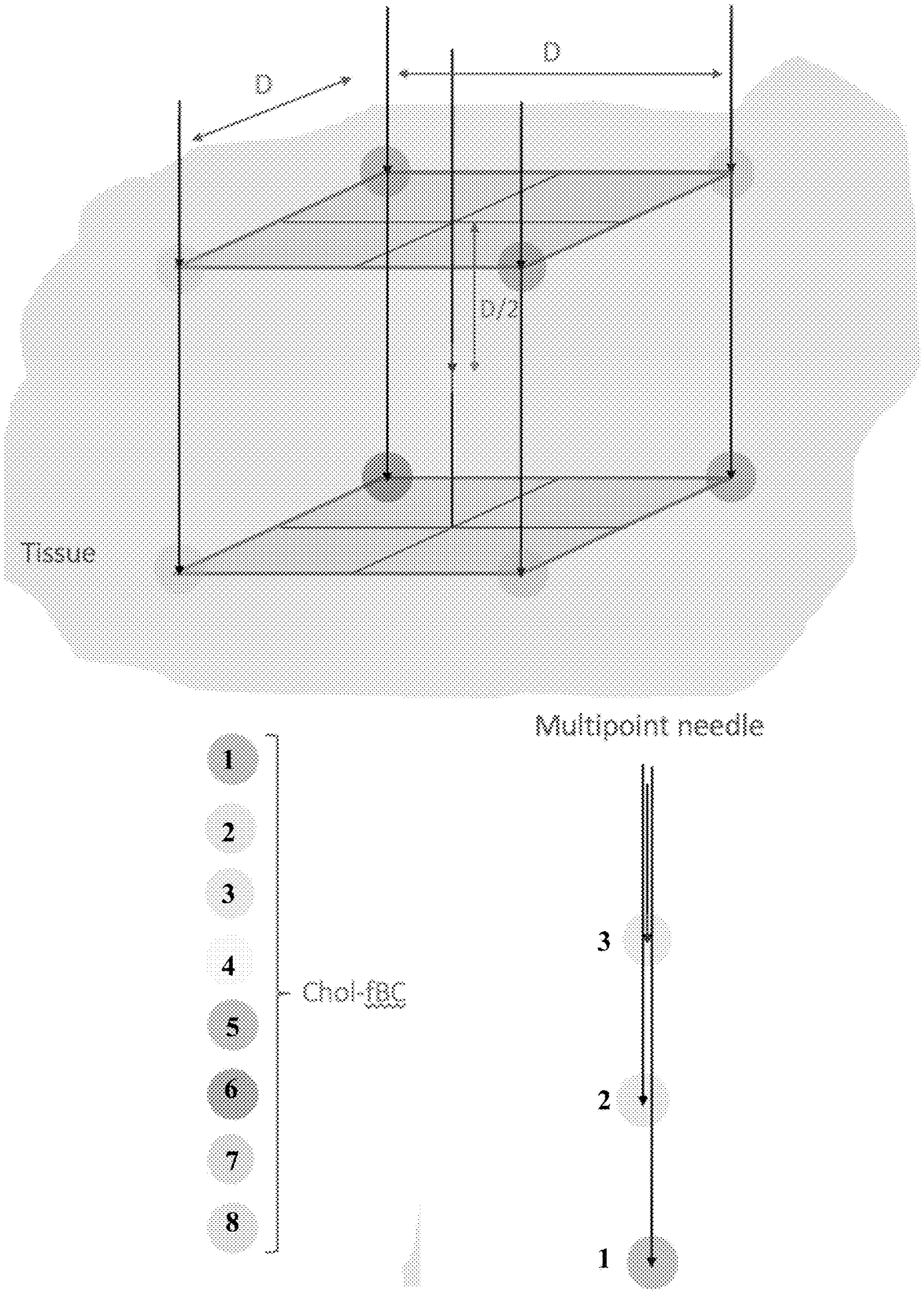


FIG. 78

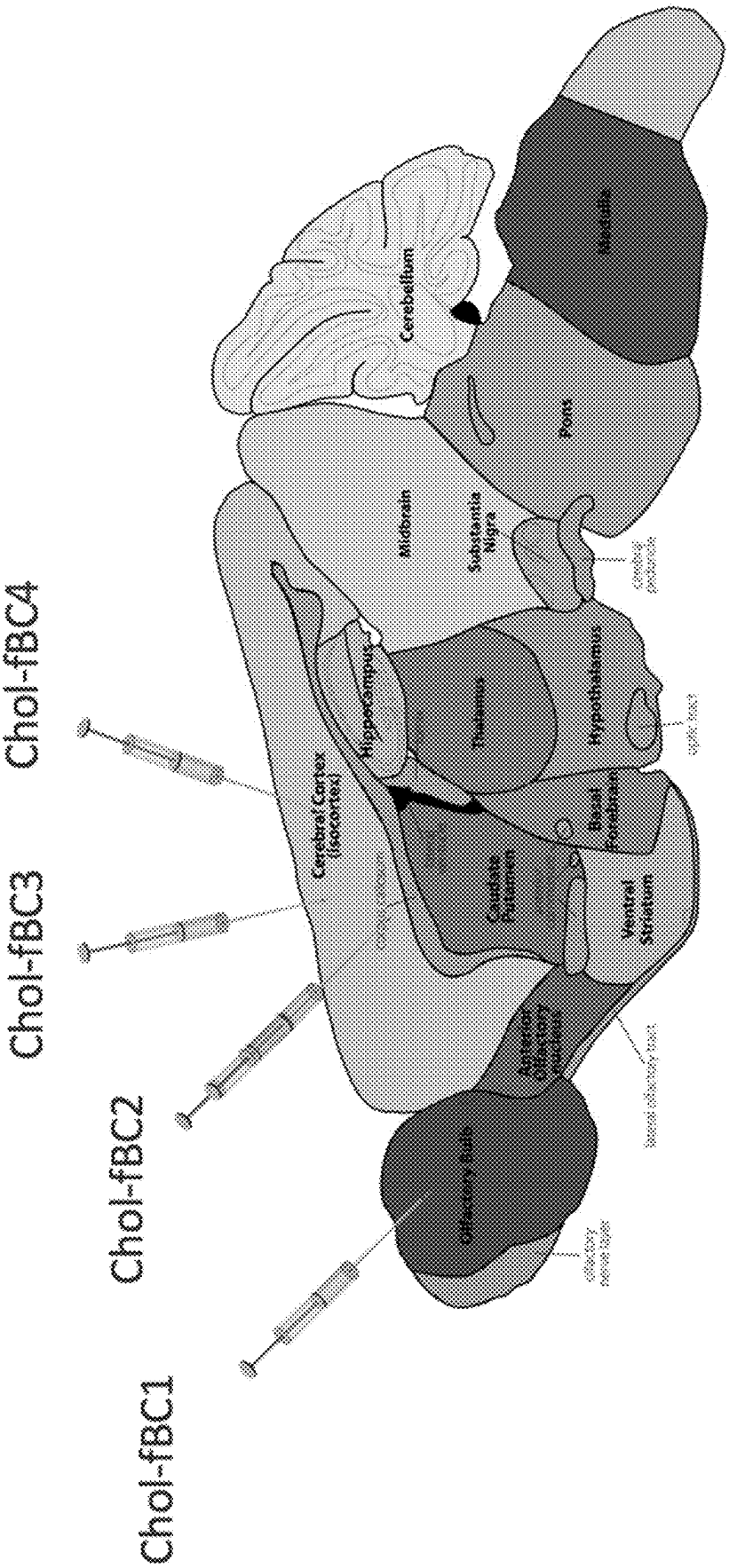

Chol-fBC1
Chol-fBC2
Chol-fBC3
Chol-fBC4
Cerebellum
Medulla
Pons
Midbrain
Substantia Nigra
Hippocampus
Thalamus
Hypothalamus
Cerebral Cortex (isocortex)
Caudate Putamen
Basal Forebrain
Ventral Striatum
Anterior Olfactory Nucleus
Olfactory Bulb

FIG. 79A

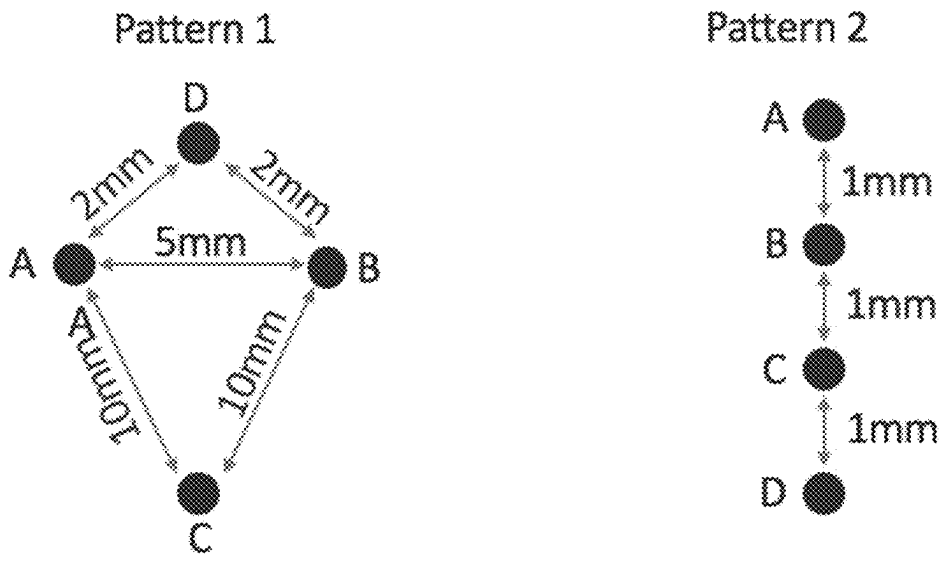

Pattern 1
D
2mm
2mm
A
5mm
B
10mm
10mm
C
Pattern 2
A
1mm
B
1mm
C
1mm
D

FIG. 79B

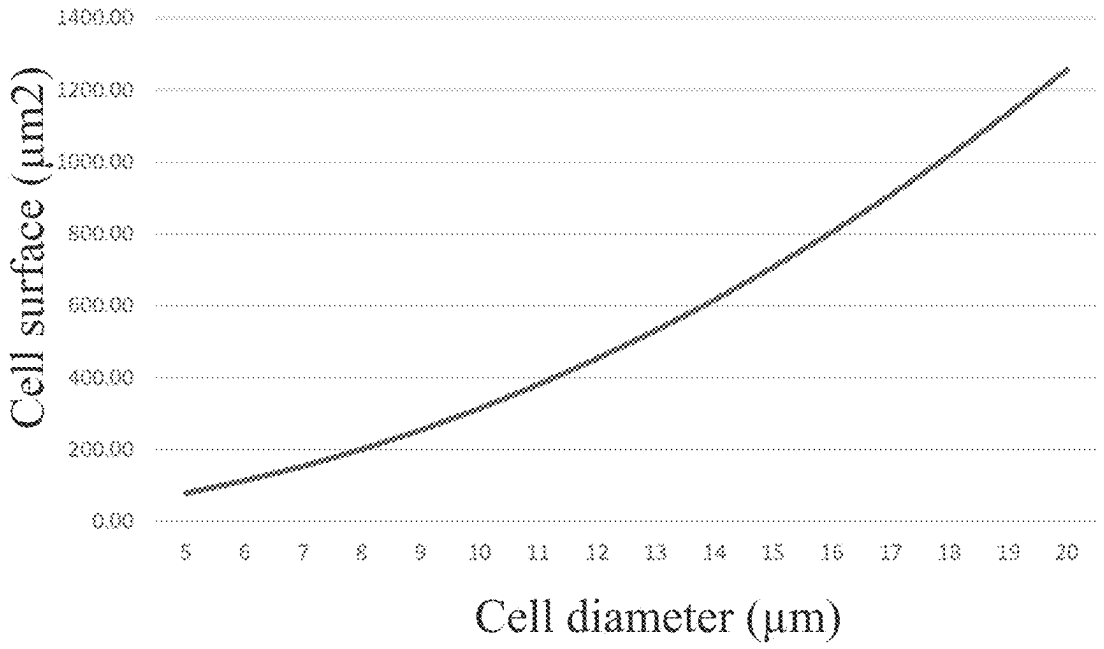

1400.00
1200.00
1000.00
800.00
600.00
400.00
200.00
0.00
Cell surface (µm2)
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
Cell diameter (µm)

FIG. 80

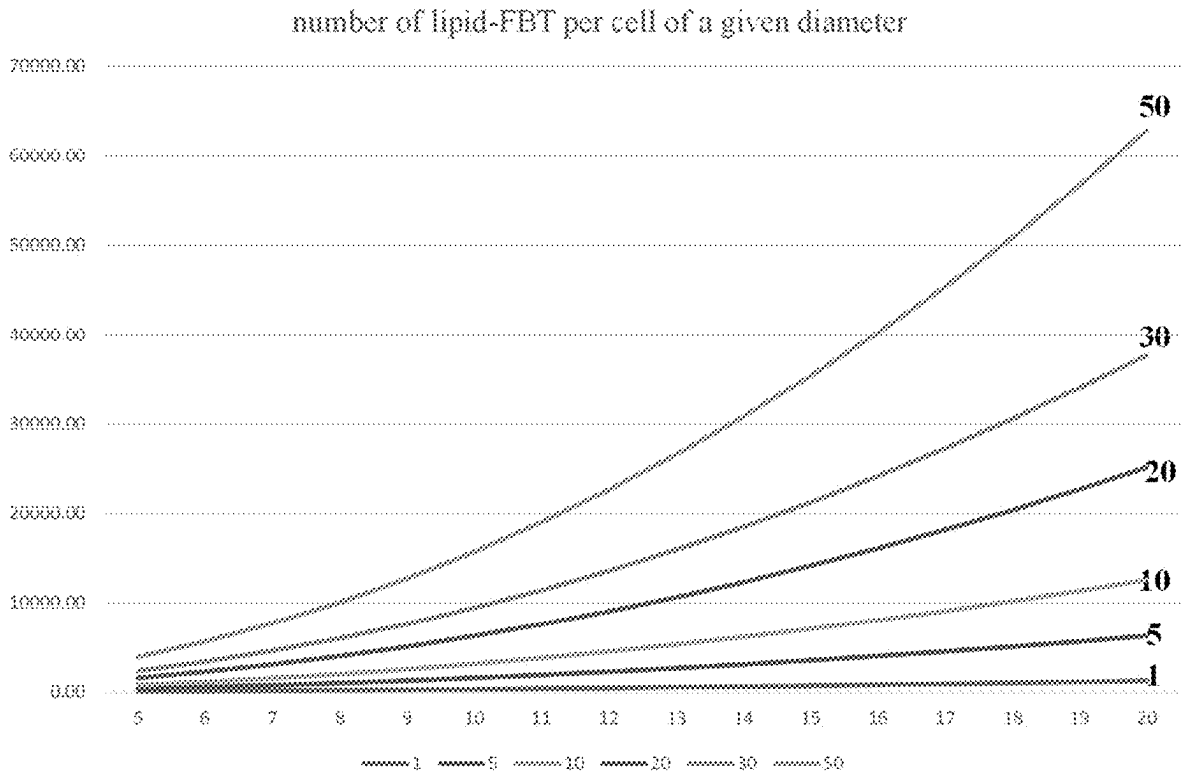

number of lipid-FBT per cell of a given diameter
70000.00
60000.00
50000.00
40000.00
30000.00
20000.00
10000.00
0.00
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
50
30
20
10
5
1
1
5
10
20
30
50

FIG. 81

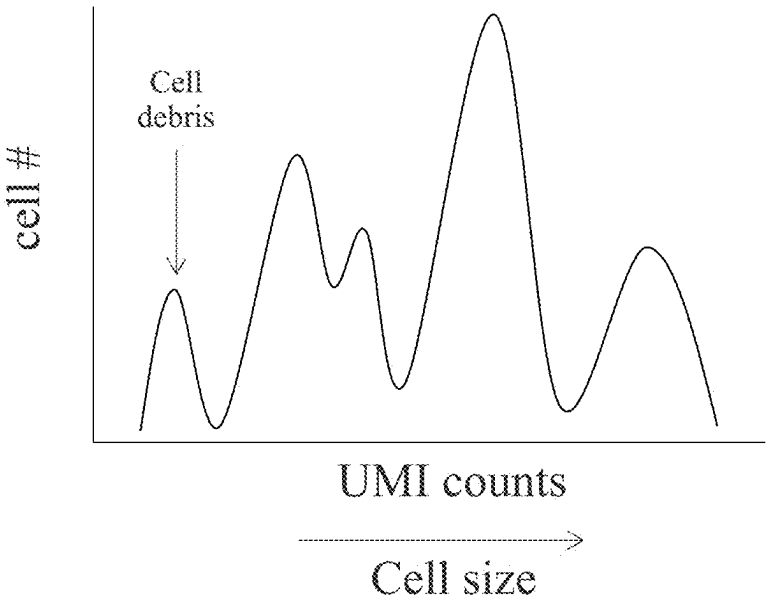


FIG. 82

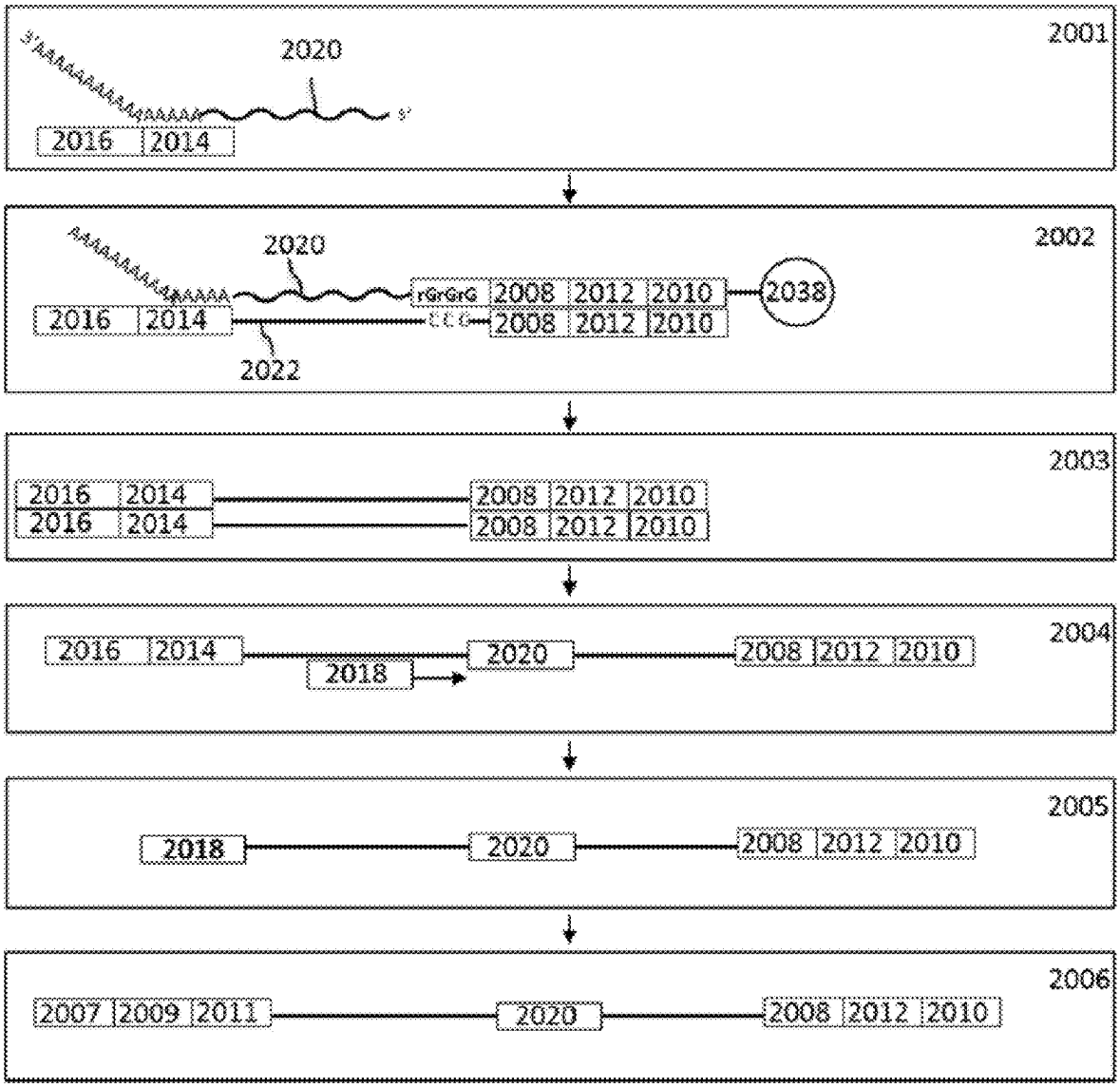

2001
2020
2016
2014
2002
2020
rGrGrG
2008
2012
2010
2038
2016
2014
2008
2012
2010
2022
2003
2016
2014
2008
2012
2010
2016
2014
2008
2012
2010
2004
2016
2014
2018
2020
2008
2012
2010
2005
2018
2020
2008
2012
2010
2006
2007
2009
2011
2020
2008
2012
2010

FIG. 83

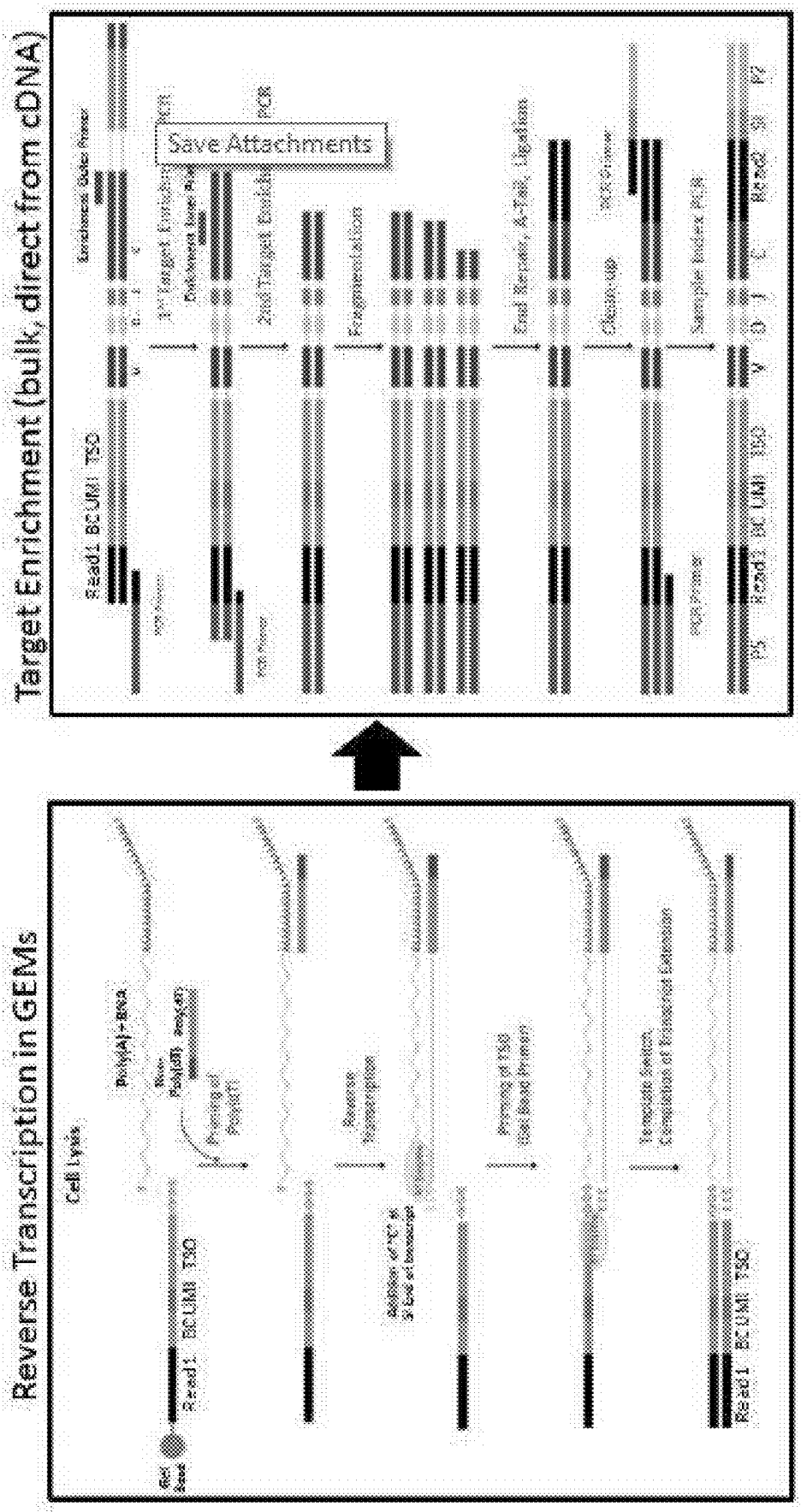

Target Enrichment (bulk, direct from cDNA)
Reverse Transcription in GEMs
Save Attachments
Cell Lysis
Poly(A)+ RNA
Read1
BC
UMI
TSO
Enrichment Gene Primer
3' Target Enrichment PCR
Enhanced Gene Pool
2nd Target Enrichment PCR
Fragmentation
End Repair, A-Tail, Ligation
Cleanup
Sample Index PCR
P5
Read1
BC
UMI
TSO
V
D
J
C
Read2
P7
Reverse Transcription
Primer of TSO (at Read 2 Primer)
Template Switch, Completion of Transcript Extension

FIG. 84A

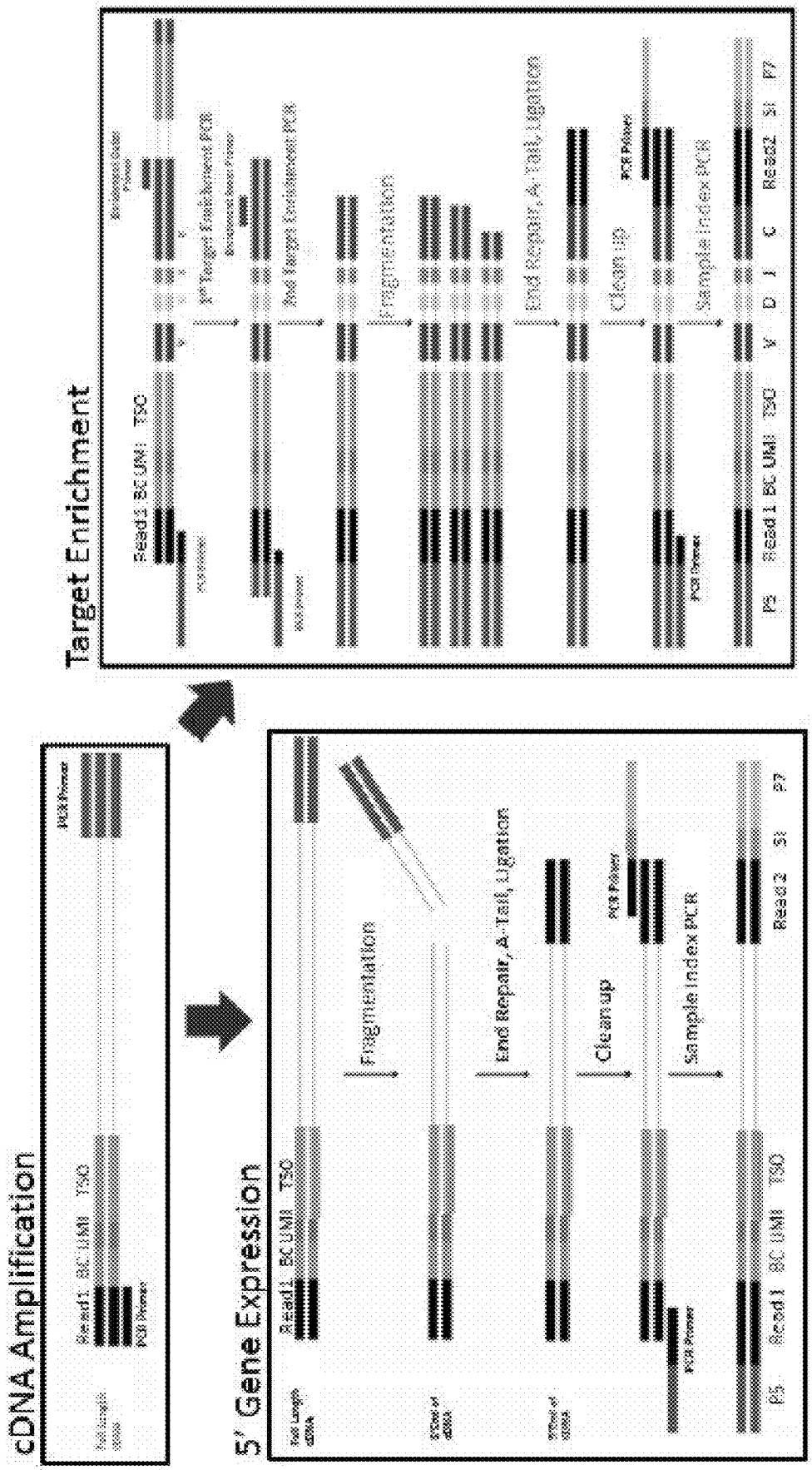

cDNA Amplification
Read1 BC UMI TSO
PCR Primer
5' Gene Expression
Fragmentation
End Repair, A-Tail, Ligation
Clean up
Sample Index PCR
P5 Read1 BC UMI TSO
Read2 SI P7
Target Enrichment
1st Target Enrichment PCR
2nd Target Enrichment PCR

FIG. 84B

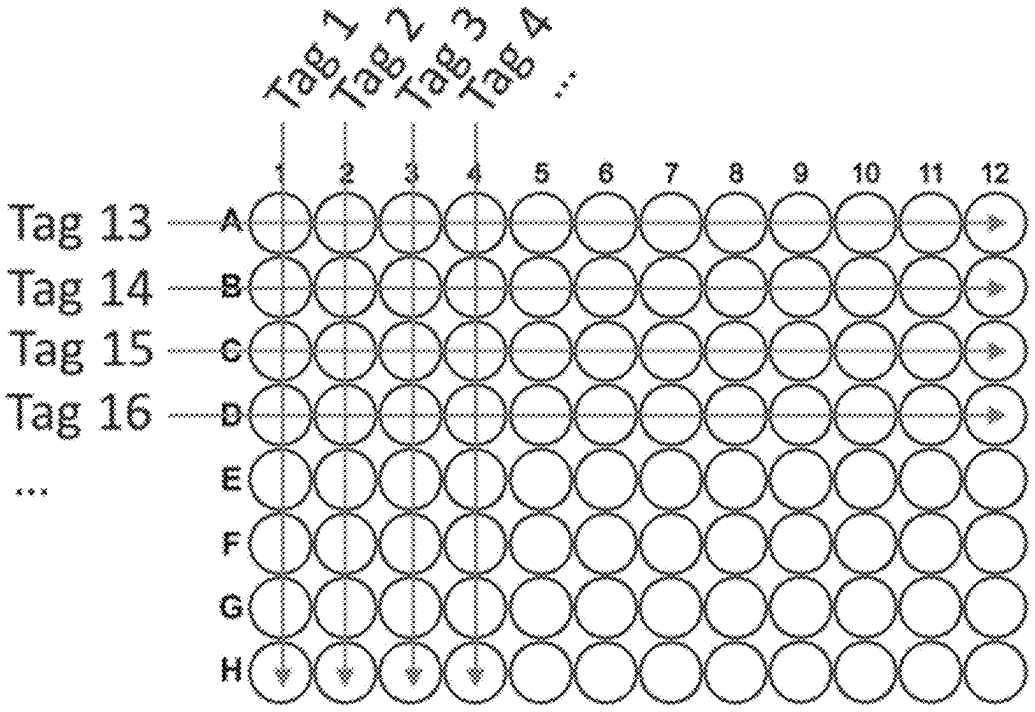

Tag 1
Tag 2
Tag 3
Tag 4
...
1
2
3
4
5
6
7
8
9
10
11
12
Tag 13
Tag 14
Tag 15
Tag 16
...
A
B
C
D
E
F
G
H

FIG. 85

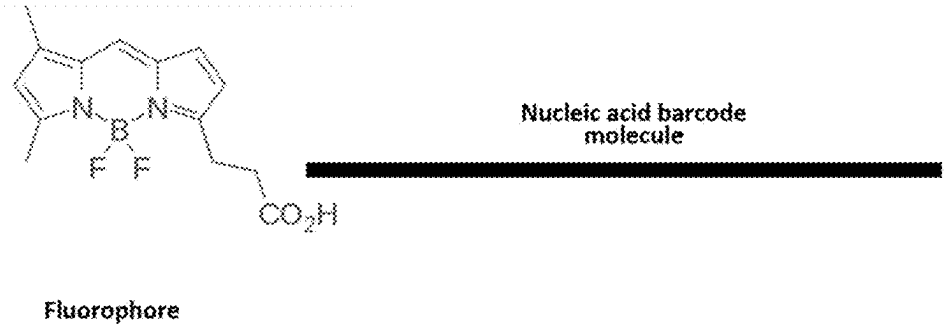


FIG. 86

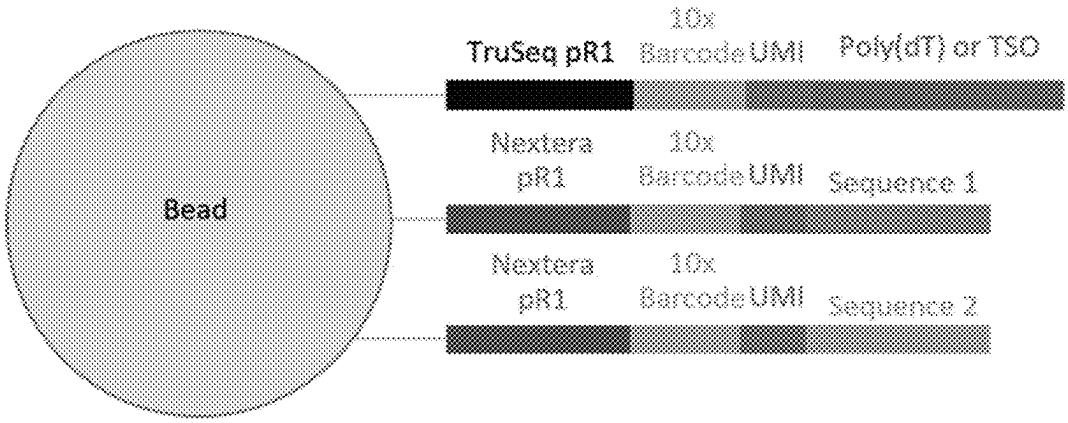

Bead
TruSeq pR1
10x Barcode UMI
Poly(dT) or TSO
Nextera pR1
10x Barcode UMI
Sequence 1
Nextera pR1
10x Barcode UMI
Sequence 2

FIG. 87

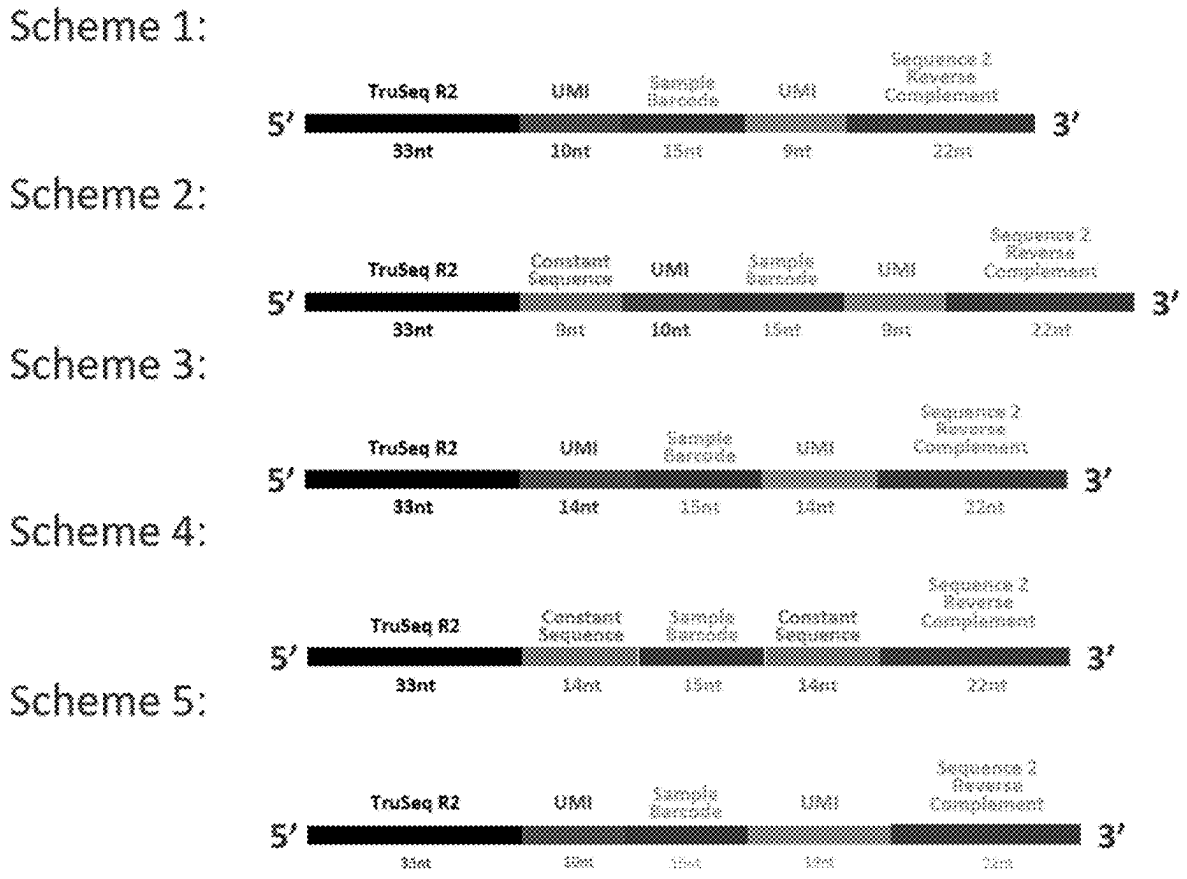

Scheme 1:
TruSeq R2
UMI
Sample Barcode
UMI
Sequence 2 Reverse Complement
5′
3′
33nt
10nt
15nt
9nt
22nt
Scheme 2:
TruSeq R2
Constant Sequence
UMI
Sample Barcode
UMI
Sequence 2 Reverse Complement
5′
3′
33nt
9nt
10nt
15nt
9nt
22nt
Scheme 3:
TruSeq R2
UMI
Sample Barcode
UMI
Sequence 2 Reverse Complement
5′
3′
33nt
14nt
15nt
14nt
22nt
Scheme 4:
TruSeq R2
Constant Sequence
Sample Barcode
Constant Sequence
Sequence 2 Reverse Complement
5′
3′
33nt
14nt
15nt
14nt
22nt
Scheme 5:
TruSeq R2
UMI
Sample Barcode
UMI
Sequence 2 Reverse Complement
5′
3′

FIG. 88

METHODS AND SYSTEMS FOR PROCESSING POLYNUCLEOTIDES

CROSS-REFERENCE

[0001] This application is a continuation of U.S. application Ser. No. 18/786,209, filed Jul. 26, 2024, which is a continuation of U.S. application Ser. No. 17/473,760, filed Sep. 13, 2021, now U.S. Pat. No. 12,084,716, which is a continuation of U.S. patent application Ser. No. 17/084,481, filed Oct. 29, 2020, now U.S. Pat. No. 11,180,805, which is a continuation of U.S. patent application Ser. No. 16/443,638, filed Jun. 17, 2019, now U.S. Pat. No. 10,858,702, which is a continuation of U.S. patent application Ser. No. 16/439,675, filed Jun. 12, 2019, now U.S. Pat. No. 10,815,525, which is a continuation-in-part of U.S. patent application Ser. No. 15/933,299, filed Mar. 22, 2018, now U.S. Pat. No. 10,480,029, which is a continuation of U.S. patent application Ser. No. 15/720,085, filed on Sep. 29, 2017, now U.S. Pat. No. 10,011,872 on Jul. 3, 2018, which claims priority to U.S. Provisional Patent Application No. 62/438,341, filed on Dec. 22, 2016. U.S. patent application Ser. No. 16/439,675 is also a continuation-in-part of International Patent Application No. PCT/US2017/068320, filed Dec. 22, 2017, which claims priority to U.S. Provisional Patent Application No. 62/438,341, filed on Dec. 22, 2016, and is also a continuation application of U.S. patent application Ser. No. 15/720,085, filed on Sep. 29, 2017, which claims priority to U.S. Provisional Patent Application No. 62/438,341, filed on Dec. 22, 2016; U.S. patent application Ser. No. 16/439,675 is also a continuation-in-part of International Patent Application No. PCT/US2018/064600, filed Dec. 7, 2018, which application claims the benefit of U.S. Provisional Applications Nos. 62/596,557, filed Dec. 8, 2017, and 62/723,960, filed Aug. 28, 2018, and is also a continuation application of U.S. Non-Provisional application Ser. No. 16/107,685, filed Aug. 21, 2018, now abandoned, which claims priority to U.S. Provisional Application No. 62/596,557, filed Dec. 8, 2017; U.S. patent application Ser. No. 16/439,675 is also a continuation-in-part of U.S. Non-Provisional application Ser. No. 16/107,685, filed Aug. 21, 2018, now abandoned, which claims priority to U.S. Provisional Application No. 62/596,557, filed Dec. 8, 2017. Each of the above-referenced applications is herein incorporated by reference in its entirety for all purposes.

SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing which has been submitted electronically in XML format and is hereby incorporated by reference in its entirety. Said XML copy, created on Nov. 15, 2022, is named 43487742313_SL.xml and is 59,162 bytes in size.

BACKGROUND

[0003] Significant advances in analyzing and characterizing biological and biochemical materials and systems have led to unprecedented advances in understanding the mechanisms of life, health, disease and treatment. Among these advances, technologies that target and characterize the genomic make up of biological systems have yielded some of the most groundbreaking results, including advances in the use and exploitation of genetic amplification technologies, and nucleic acid sequencing technologies.

[0004] Nucleic acid sequencing can be used to obtain information in a wide variety of biomedical contexts, including diagnostics, prognostics, biotechnology, and forensic biology. Sequencing may involve basic methods including Maxam-Gilbert sequencing and chain-termination methods, or de novo sequencing methods including shotgun sequencing and bridge PCR, or next-generation methods including polony sequencing, 454 pyrosequencing, Illumina sequencing, SOLiD sequencing, Ion Torrent semiconductor sequencing, HeliScope single molecule sequencing, SMRT® sequencing, and others. Nucleic acid sequencing technologies, including next-generation DNA sequencing, have been useful for genomic and proteomic analysis of cell populations.

SUMMARY

[0005] Recognized herein is the need for methods, compositions and systems for analyzing genomic and proteomic information from individual cells or a small population of cells. Such cells include, but are not limited to, cancer cells, fetal cells, and immune cells involved in immune responses. Provided herein are methods, compositions and systems for analyzing individual cells or a small population of cells, including the analysis and attribution of nucleic acids and proteins from and to these individual cells or cell populations.

[0006] In an aspect, the present disclosure provides a method of characterizing a cell. The method comprises (a) providing a partition comprising a cell and at least one labelling agent, wherein the at least one labelling agent is (i) capable of binding to a cell surface feature of the cell and (ii) is coupled to a reporter oligonucleotide comprising a nucleic acid barcode sequence that permits identification of the at least one labelling agent, wherein the partition comprises an anchor oligonucleotide that is capable of interacting with the reporter oligonucleotide barcode; (b) in the partition, synthesizing a nucleic acid molecule comprising at least a portion of the nucleic acid barcode sequence or a complement thereof, and (c) subjecting the nucleic acid molecule to sequencing to identify the labelling agent or the cell.

[0007] In some embodiments, in (a), the at least one labelling agent is bound to the cell surface feature. In some embodiments, prior to (a), the at least one labelling agent is subjected to conditions suitable for binding the at least one labelling agent to the cell surface feature. In some embodiments, subjecting the at least one labelling agent to the conditions suitable for binding the at least one labelling agent to the cell surface feature is performed when the cell and the at least one labelling agent are free from the partition. In some embodiments, prior to (a), the at least one labelling agent is coupled to the reporter oligonucleotide.

[0008] In some embodiments, in (b), the reporter oligonucleotide is subjected to a primer extension reaction that generates the nucleic acid molecule. In some embodiments, the primer extension reaction comprises subjecting the reporter oligonucleotide to conditions suitable to hybridize the anchor oligonucleotide to the reporter oligonucleotide and extend the anchor oligonucleotide using the reporter oligonucleotide as a template.

[0009] In some embodiments, in (b), the anchor oligonucleotide is coupled to a bead. In some embodiments, in (b), the anchor oligonucleotide is coupled to a bead and the method further comprises releasing the anchor oligonucleotide from the bead prior to the synthesizing. In some

embodiments, the bead is a gel bead. In some embodiments, the releasing comprises subjecting the bead to a stimulus that degrades the bead. In some embodiments, the stimulus is a chemical stimulus. In some embodiments, the bead comprises at least about 1,000 copies of the anchor oligonucleotide. In some embodiments, the bead comprises at least about 10,000 copies of the anchor oligonucleotide. In some embodiments, the bead comprises at least about 100,000 copies of the anchor oligonucleotide.

[0010] In some embodiments, prior to (c), the nucleic acid molecule is released from the partition. In some embodiments, (c) comprises identifying the at least one labelling agent. In some embodiments, (c) comprises identifying the cell surface feature from identifying the at least one labelling agent. In some embodiments, (c) comprises determining an abundance of the given cell surface feature on the cell. In some embodiments, (c) comprises identifying the cell. In some embodiments, (c) comprises identifying the at least one labelling agent and the cell.

[0011] In some embodiments, the reporter oligonucleotide comprises a unique molecular identification (UMI) sequence. In some embodiments, the UMI sequence permits identification of the cell. In some embodiments, (c) comprises determining a sequence of the UMI sequence and identifying the cell.

[0012] In some embodiments, the partition is a droplet in an emulsion. In some embodiments, the at least one labelling agent is selected from the group comprising of an antibody, or an epitope binding fragment thereof, a cell surface receptor binding molecule, a receptor ligand, a small molecule, a bi-specific antibody, a bi-specific T-cell engager, a T-cell receptor engager, a B-cell receptor engager, a probody, an aptamer, a monobody, an affimer, a darpin, a protein scaffold, an antigen, an antigen presenting particle and a major histocompatibility complex (MHC). In some embodiments, the cell surface feature is selected from the group comprising of a receptor, an antigen, a surface protein, a transmembrane protein, a cluster of differentiation protein, a protein channel, a protein pump, a carrier protein, a phospholipid, a glycoprotein, a glycolipid, a cell-cell interaction protein complex, an antigen-presenting complex, a major histocompatibility complex, an engineered T-cell receptor, a T-cell receptor, a B-cell receptor, a chimeric antigen receptor, a gap junction, and an adherens junction. In some embodiments, the partition comprises only one cell.

[0013] In some embodiments, the cell is bound to at least one of the at least one labelling agent. In some embodiments, the at least one of the at least one labelling agent comprises at least two of the same labelling agent. In some embodiments, the at least one of the at least one labelling agent comprises at least two different labelling agents. In some embodiments, the cell is bound to at least about 5 different labelling agents. In some embodiments, the cell is bound to at least about 10 different labelling agents. In some embodiments, the cell is bound to at least about 50 different labelling agents. In some embodiments, the cell is bound to at least about 100 different labelling agents. In some embodiments, the (c) comprises determining an identity of at least a subset of the different labelling agents.

[0014] In some embodiments, the method further comprises (i) liberating nucleic acid from the cell and (ii) subjecting the nucleic acid or a derivative thereof to sequencing. In some embodiments, the nucleic acid is liberated from the cell into the partition.

[0015] In an aspect, the present disclosure provides a system for characterizing a cell. The system comprises an electronic display screen comprising a user interface that displays a graphical element that is accessible by a user to execute a protocol to characterize the cell; and a computer processor coupled to the electronic display screen and programmed to execute the protocol upon selection of the graphical element by the user, which protocol comprises: (a) providing a partition comprising a cell and at least one labelling agent, wherein the at least one labelling agent is (i) capable of binding to a cell surface feature of the cell and (ii) is coupled to a reporter oligonucleotide comprising a nucleic acid barcode sequence that permits identification of the at least one labelling agent, wherein the partition comprises an anchor oligonucleotide that is capable of interacting with the reporter oligonucleotide barcode; (b) in the partition, synthesizing a nucleic acid molecule comprising at least a portion of the nucleic acid barcode sequence or a complement thereof, and (c) subjecting the nucleic acid molecule to sequencing to identify the labelling agent or the cell.

[0016] In some embodiments, in protocol (a), the at least one labelling agent is bound to the cell surface feature. In some embodiments, prior to protocol (a), the at least one labelling agent is subjected to conditions suitable for binding the at least one labelling agent to the cell surface feature. In some embodiments, subjecting the at least one labelling agent to the conditions suitable for binding the at least one labelling agent to the cell surface feature is performed when the cell and the at least one labelling agent are free from the partition. In some embodiments, prior to protocol (a), the at least one labelling agent is coupled to the reporter oligonucleotide.

[0017] In some embodiments, in protocol (b), the reporter oligonucleotide is subjected to a primer extension reaction that generates the nucleic acid molecule. In some embodiments, the primer extension reaction comprises subjecting the reporter oligonucleotide to conditions suitable to hybridize the anchor oligonucleotide to the reporter oligonucleotide and extend the anchor oligonucleotide using the reporter oligonucleotide as a template.

[0018] In some embodiments, in protocol (b), the anchor oligonucleotide is coupled to a bead. In some embodiments, in (b), the anchor oligonucleotide is coupled to a bead and the method further comprises releasing the anchor oligonucleotide from the bead prior to the synthesizing. In some embodiments, the bead is a gel bead. In some embodiments, the releasing comprises subjecting the bead to a stimulus that degrades the bead. In some embodiments, the stimulus is a chemical stimulus. In some embodiments, the bead comprises at least about 1,000 copies of the anchor oligonucleotide. In some embodiments, the bead comprises at least about 10,000 copies of the anchor oligonucleotide. In some embodiments, the bead comprises at least about 100,000 copies of the anchor oligonucleotide.

[0019] In some embodiments, prior to protocol (c), the nucleic acid molecule is released from the partition. In some embodiments, protocol (c) comprises identifying the at least one labelling agent. In some embodiments, protocol (c) comprises identifying the cell surface feature from identifying the at least one labelling agent. In some embodiments, protocol (c) comprises determining an abundance of the given cell surface feature on the cell. In some embodiments,

protocol (c) comprises identifying the cell. In some embodiments, protocol (c) comprises identifying the at least one labelling agent and the cell.

[0020] In some embodiments, the reporter oligonucleotide comprises a unique molecular identification (UMI) sequence. In some embodiments, the UMI sequence permits identification of the cell. In some embodiments, protocol (c) comprises determining a sequence of the UMI sequence and identifying the cell.

[0021] In some embodiments, the partition is a droplet in an emulsion. In some embodiments, the at least one labelling agent is selected from the group comprising of an antibody, or an epitope binding fragment thereof, a cell surface receptor binding molecule, a receptor ligand, a small molecule, a bi-specific antibody, a bi-specific T-cell engager, a T-cell receptor engager, a B-cell receptor engager, a probody, an aptamer, a monobody, an affimer, a darpin, a protein scaffold, an antigen, an antigen presenting particle and a major histocompatibility complex (MHC). In some embodiments, the cell surface feature is selected from the group comprising of a receptor, an antigen, a surface protein, a transmembrane protein, a cluster of differentiation protein, a protein channel, a protein pump, a carrier protein, a phospholipid, a glycoprotein, a glycolipid, a cell-cell interaction protein complex, an antigen-presenting complex, a major histocompatibility complex, an engineered T-cell receptor, a T-cell receptor, a B-cell receptor, a chimeric antigen receptor, a gap junction, and an adherens junction. In some embodiments, the partition comprises only one cell.

[0022] In some embodiments, the cell is bound to at least one of the at least one labelling agent. In some embodiments, the at least one of the at least one labelling agent comprises at least two of the same labelling agent. In some embodiments, the at least one of the at least one labelling agent comprises at least two different labelling agents. In some embodiments, the cell is bound to at least about 5 different labelling agents. In some embodiments, the cell is bound to at least about 10 different labelling agents. In some embodiments, the cell is bound to at least about 50 different labelling agents. In some embodiments, the cell is bound to at least about 100 different labelling agents. In some embodiments, protocol (c) comprises determining an identity of at least a subset of the different labelling agents.

[0023] In some embodiments, protocol comprises (i) liberating nucleic acid from the cell and (ii) subjecting the nucleic acid or a derivative thereof to sequencing. In some embodiments, the nucleic acid is liberated from the cell into the partition.

[0024] An additional aspect of the disclosure provides a method for analyte characterization. The method includes: (a) providing a plurality of partitions, where a given partition of the plurality of partitions comprises a plurality of barcode molecules and a plurality of analytes. In some cases, the plurality of barcode molecules comprises at least 1,000 barcode molecules. In addition, (i) a first individual barcode molecule of the plurality of barcode molecules can comprise a first nucleic acid barcode sequence that is capable of coupling to a first analyte of the plurality of analytes, and (ii) a second individual barcode molecule of the plurality of barcoded molecules can comprise a second nucleic acid barcode sequence that is capable of coupling to a second analyte of the plurality of analytes where the first analyte and the second analyte are different types of analytes. The method also includes (b) in the given partition, (i) synthesizing a first nucleic acid molecule comprising at least a portion of the first nucleic acid barcode sequence or complement thereof, and (ii) synthesizing a second nucleic acid molecule comprising at least a portion of the second nucleic acid barcode sequence or complement thereof, and (c) removing the first nucleic acid molecule and the second nucleic acid molecule from the given partition.

[0025] In some embodiments, the method further comprises subjecting the first nucleic acid molecule and the second nucleic acid molecule, or a derivative of the first nucleic acid molecule and/or the second nucleic acid molecule, to sequencing to characterize the first analyte and/or the second analyte. In some embodiments, the method further comprises repeating (a)-(c) based on a characterization of the first analyte or the second analyte from the sequencing. In some embodiments, the method further comprises selecting the first analyte or the second analyte based on a characterization of the first analyte or the second analyte obtained from the sequencing or a subsequent sequencing upon repeating (a)-(c).

[0026] In some embodiments, (b) further comprises: (1) synthesizing the first nucleic acid molecule comprising at least a portion of the first nucleic acid barcode sequence or complement thereof, and (2) synthesizing the second nucleic acid molecule comprising at least a portion of the second nucleic acid barcode sequence or complement thereof.

[0027] In some embodiments, the first analyte is a nucleic acid molecule, such as genomic deoxyribonucleic acid (gDNA) or messenger RNA (mRNA).

[0028] In some embodiments, the first analyte is a labelling agent capable of coupling to a cell surface feature of a cell. In some embodiments, the first individual barcode molecule or the second individual barcode molecule is capable of coupling to the labelling agent via a third nucleic acid molecule coupled to the labelling agent. In some embodiments, the cell surface feature is a receptor, an antigen, or a protein. In some embodiments, the labelling agent is an antibody, an antibody fragment or a major histocompatibility complex (MHC). In some embodiments, the given partition comprises the cell or one or more components of the cell. In some embodiments, the given partition comprises a single cell. In some embodiments, the first nucleic acid molecule or the second nucleic molecule comprises a third barcode sequence. In some embodiments, the third barcode sequence is derived from a third nucleic acid molecule. In some embodiments, the third nucleic acid molecule is coupled to a labelling agent capable of binding to a cell surface feature of a cell.

[0029] In some embodiments, the first analyte and second analyte are different types of nucleic acid molecules. In some embodiments, the first analyte is a ribonucleic acid molecule and the second analyte is a deoxyribonucleic acid molecule. In some embodiments, (i) the first individual barcode molecule comprises a first priming sequence capable of hybridizing to the first analyte; or (ii) the second individual barcode molecule comprises a second priming sequence capable of hybridizing to the second analyte. In some embodiments, the first barcode molecule or the second barcode molecule comprises a unique molecular identification (UMI) sequence.

[0030] In some embodiments, the first analyte is a nucleic acid molecule and the second analyte is a labelling agent capable of coupling to a cell surface feature. In some embodiments, the first analyte is a messenger ribonucleic

acid molecule. In some embodiments, (i) the first individual barcode molecule comprises a first priming sequence capable of hybridizing to the first analyte; or (ii) the second individual barcode molecule comprises a second priming sequence capable of hybridizing to the labelling agent. In some embodiments, the labelling agent is an antibody, or an epitope binding fragment thereof, or a major histocompatibility complex (MHC). In some embodiments, the cell surface feature is selected from the group consisting of a receptor, an antigen, or a protein.

[0031] In some embodiments, the first analyte comprises a nucleic acid molecule with a nucleic acid sequence or complement thereof that encodes at least a portion of a V(D)J sequence of an immune cell receptor. In some embodiments, the nucleic acid molecule is a messenger ribonucleic acid. In some embodiments, the nucleic acid molecule is complementary DNA (cDNA) derived from reverse transcription of an mRNA encoding the at least a portion of the V(D)J sequence.

[0032] In some embodiments, the first analyte comprises a nucleic acid molecule with a nucleic acid sequence that is capable of functioning as a component of a gene editing reaction. In some embodiments, the gene editing reaction comprises clustered regularly interspaced short palindromic repeats (CRISPR)-based gene editing.

[0033] In some embodiments, at least one of the first individual barcode molecule and the second individual barcode molecule is coupled to a bead, such as a gel bead. The bead can be degradable. In some embodiments, the method further comprises, after (a), releasing the first individual barcode molecule or the second individual barcode from the bead. In some embodiments, the given partition further comprises an agent capable of releasing the first individual barcode molecule or the second individual barcode from the bead.

[0034] In some embodiments, the given partition selected is a droplet among a plurality of droplets or a well among a plurality of wells. In some embodiments, the first nucleic acid barcode sequence and the second nucleic barcode sequence are identical. In some embodiments, the method further comprises performing one or more reactions subsequent to removing the first nucleic acid molecule and the second nucleic acid molecule from the given partition.

[0035] Another aspect of the disclosure provides a composition for characterizing a plurality of analytes. The composition comprises a partition comprising a plurality of barcode molecules and the plurality of analytes. The plurality of barcode molecules can comprise at least 1,000 barcode molecules. In addition, (i) a first individual barcode molecule of the plurality of barcode molecules can comprise a first nucleic acid barcode sequence that is capable of coupling to a first analyte of the plurality of analytes; and (ii) a second individual barcode molecule of the plurality of barcoded molecules can comprise a second nucleic acid barcode sequence that is capable of coupling to a second analyte of the plurality of analytes, where the first analyte and the second analyte are different types of analytes.

[0036] In some embodiments, the first analyte is a nucleic acid molecule, such as genomic deoxyribonucleic acid (gDNA) or is messenger RNA (mRNA).

[0037] In some embodiments, the first analyte is a labelling agent capable of coupling to a cell surface feature of a cell. In some embodiments, the first individual barcode molecule or the second individual barcode molecule is capable of coupling to the labelling agent via a third nucleic acid molecule coupled to the labelling agent. In some embodiments, the cell surface feature is a receptor, an antigen, or a protein. In some embodiments, the labelling agent is an antibody, or an epitope binding fragment thereof, or a major histocompatibility complex (MHC). In some embodiments, the partition comprises the cell or one or more components of the cell. In some embodiments, the partition comprises a single cell. In some embodiments, the first nucleic acid molecule or the second nucleic molecule comprises a third barcode sequence. In some embodiments, the third barcode sequence is derived from a third nucleic acid molecule. In some embodiments, the third nucleic acid molecule is coupled to a labelling agent capable of binding to a cell surface feature of a cell.

[0038] In some embodiments, the first analyte and second analyte are different types nucleic acid molecules. In some embodiments, the first analyte is a ribonucleic acid molecule and the second analyte is a deoxyribonucleic acid molecule. In some embodiments, (i) the first individual barcode molecule comprises a first priming sequence capable of hybridizing to the first analyte; or (ii) the second individual barcode molecule comprises a second priming sequence capable of hybridizing to the second analyte. In some embodiments, the first barcode molecule or the second barcode molecule comprises a unique molecular identification (UMI) sequence.

[0039] In some embodiments, the first analyte is a nucleic acid molecule and the second analyte is a labelling agent capable of coupling to a cell surface feature. In some embodiments, the first analyte is a ribonucleic acid molecule. In some embodiments, (i) the first individual barcode molecule comprises a first priming sequence capable of hybridizing to the first analyte; or (ii) the second individual barcode molecule comprises a second priming sequence capable of hybridizing to a third nucleic acid molecule coupled to the labelling agent. In some embodiments, the labelling agent is an antibody, or an epitope binding fragment thereof, or a major histocompatibility complex (MHC). In some embodiments, the cell surface feature is a receptor, an antigen, or a protein.

[0040] In some embodiments, the first analyte comprises a nucleic acid molecule with a nucleic acid sequence or complement thereof that encodes at least a portion of a V(D)J sequence of an immune cell receptor. In some embodiments, the nucleic acid sequence is a ribonucleic acid molecule. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence that is complementary DNA (cDNA) derived from reverse transcription of an mRNA encoding the at least a portion of the V(D)J sequence.

[0041] In some embodiments, the first analyte comprises a nucleic acid molecule with a nucleic acid sequence that is capable of functioning as a component of a gene editing reaction. In some embodiments, the gene editing reaction comprises clustered regularly interspaced short palindromic repeats (CRISPR)-based gene editing. In some embodiments, at least one of the first individual barcode molecule and the second individual barcode molecule is coupled to a bead, such as a gel bead. The bead may be degradable. In some embodiments, the given partition further comprises an agent capable of releasing the first individual barcode molecule or the second individual barcode from the bead. In some embodiments, the given partition is a droplet among a

plurality of droplets or a well among a plurality of wells. In some embodiments, the first nucleic acid barcode sequence and the second nucleic barcode sequence are identical.

[0042] An additional aspect of the disclosure provides a system for characterizing a plurality of analytes. The system comprises a partitioning unit for providing a partition comprising a plurality of barcode molecules and the plurality of analytes, where: (i) a first individual barcode molecule of the plurality of barcode molecules comprises a first nucleic acid barcode sequence and is capable of coupling to a first analyte of the plurality of analytes; and (ii) a second individual barcode molecule of the plurality of barcode molecules comprises a second nucleic acid barcode sequence and is capable of coupling to a second analyte of the plurality of analytes, where the first analyte and the second analyte are different types of analytes. The system also includes a controller coupled to the partitioning unit, where the controller is programmed to (i) direct the partitioning unit to provide the partition; (ii) subject the partition to conditions that are sufficient to: (1) synthesize a first nucleic acid molecule comprising at least a portion of the first nucleic acid barcode sequence or complement thereof, and (2) synthesize a second nucleic acid molecule comprising at least a portion of the second nucleic acid barcode sequence or complement thereof, where sequencing of the first nucleic acid molecule and the second nucleic acid molecule, or derivatives thereof, characterizes the first analyte or the second analyte.

[0043] In some embodiments, the partitioning unit comprises a plurality of channels. In some embodiments, the partitioning unit further comprises at least one channel junction, where the at least one channel junction is configured to provide the partition. In some embodiments, the system also includes (i) a first channel fluidically connected to the at least one channel junction and configured to provide a first fluid to the at least one channel junction; (ii) and a second channel fluidically connected to the at least one channel junction and configured to provide a second fluid, immiscible with the first fluid, to the at least one channel junction.

[0044] In some embodiments, the first analyte is a nucleic acid molecule, such as genomic deoxyribonucleic acid (gDNA) or messenger RNA (mRNA).

[0045] In some embodiments, the first analyte is a labelling agent capable of coupling to a cell surface feature of a cell. In some embodiments, the first individual barcode molecule or the second individual barcode molecule is capable of coupling to the labelling agent via a third nucleic acid molecule coupled to the labelling agent. In some embodiments, the cell surface feature is a receptor, an antigen, or a protein. In some embodiments, the labelling agent is an antibody, or an epitope binding fragment thereof, or a major histocompatibility complex (MHC). In some embodiments, the partition comprises the cell or one or more components of the cell. In some embodiments, the partition comprises a single cell. In some embodiments, the first nucleic acid molecule or the second nucleic molecule comprises a third barcode sequence. In some embodiments, the third barcode sequence is derived from a third nucleic acid molecule. In some embodiments, the third nucleic acid molecule is coupled to a labelling agent capable of binding to a cell surface feature of a cell.

[0046] In some embodiments, the first analyte and second analyte are different types nucleic acid molecules. In some embodiments, the first analyte is a ribonucleic acid molecule and the second analyte is a deoxyribonucleic acid molecule. In some embodiments, (i) the first individual barcode molecule comprises a first priming sequence capable of hybridizing to the first analyte; or (ii) the second individual barcode molecule comprises a second priming sequence capable of hybridizing to the second analyte. In some embodiments, the first barcode molecule or the second barcode molecule comprises a unique molecular identification (UMI) sequence.

[0047] In some embodiments, the first analyte is a nucleic acid molecule and the second analyte is a labelling agent capable of coupling to a cell surface feature. In some embodiments, the first analyte is a ribonucleic acid molecule. In some embodiments, (i) the first individual barcode molecule comprises a first priming sequence capable of hybridizing to the first analyte; or (ii) the second individual barcode molecule comprises a second priming sequence capable of hybridizing to a third nucleic acid molecule coupled to the labelling agent. In some embodiments, the labelling agent is an antibody, or an epitope binding fragment thereof, or a major histocompatibility complex (MHC). In some embodiments, the cell surface feature is a receptor, an antigen, or a protein.

[0048] In some embodiments, the first analyte comprises a nucleic acid molecule with a nucleic acid sequence or complement thereof that encodes at least a portion of a V(D)J sequence of an immune cell receptor. In some embodiments, the nucleic acid sequence is a messenger ribonucleic acid molecule. In some embodiments, the nucleic acid molecule is complementary DNA (cDNA) derived from reverse transcription of an mRNA encoding the at least a portion of the V(D)J sequence.

[0049] In some embodiments, the first analyte comprises a nucleic acid molecule with a nucleic acid sequence that is capable of functioning as a component of a gene editing reaction. In some embodiments, the gene editing reaction comprises clustered regularly interspaced short palindromic repeats (CRISPR)-based gene editing.

[0050] In some embodiments, at least one of the first individual barcode molecule and the second individual barcode molecule is coupled to a bead, such as a gel bead. The bead may be degradable. In some embodiments, the partition further comprises an agent capable of releasing the first individual barcode molecule or the second individual barcode from the bead. In some embodiments, the partition is a droplet among a plurality of droplets or a well among a plurality of wells. In some embodiments, the nucleic acid barcode sequence and the second nucleic barcode sequence are identical. In some embodiments, the partition comprises at least 1,000 barcode molecules.

[0051] In an aspect, the present disclosure provides a method for analyzing cellular occupancy of partitions, comprising: (a) labelling a plurality of cells with a plurality of cell nucleic acid barcode sequences to generate a plurality of labelled cells, wherein each of the plurality of labelled cells comprises a different cell nucleic acid barcode sequence; (b) generating a plurality of partitions comprising the plurality of labelled cells and a plurality of partition nucleic acid barcode sequences, wherein each of the plurality of partitions comprises a different partition nucleic barcode sequence, and wherein at least a fraction of the plurality of partitions comprises more than one labelled cell of the plurality of labelled cells; and (c) identifying at least two

labelled cells of the plurality of labelled cells as originating from a same partition using (i) cell nucleic acid barcode sequences from the plurality of cell nucleic acid barcode sequences or complements thereof and (ii) partition nucleic acid barcode sequences of the plurality of partition nucleic acid barcode sequences or complements thereof.

[0052] In some embodiments, a given cell nucleic acid barcode sequence of the plurality of cell nucleic acid barcode sequences identifies a sample from which an associated cell of the plurality of labelled cells originates.

[0053] In some embodiments, the method further comprises, after (b), synthesizing a plurality of barcoded nucleic acid products from the plurality of labelled cells, wherein a given barcoded nucleic acid product of the plurality of barcoded nucleic acid products comprises (iii) a cell identification sequence comprising a given cell nucleic acid barcode sequence of the plurality of cell nucleic acid barcode sequences or a complement of the given cell nucleic acid barcode sequence; and (iv) a partition identification sequence comprising a given partition nucleic acid barcode sequence of the plurality of partition nucleic acid barcode sequences or a complement of the given partition nucleic acid barcode sequence. In some embodiments, (v) a plurality of partition nucleic acid barcode molecules comprises the plurality of partition nucleic acid barcode sequences, each of the plurality of partition nucleic acid barcode molecules comprising a single partition nucleic acid barcode sequence of the plurality of partition nucleic acid barcode sequences, and (vi) a plurality of cell nucleic acid barcode molecules comprises the plurality of cell nucleic acid barcode sequences, each of the plurality of cell nucleic acid barcode molecules comprising a single cell nucleic acid barcode sequence of the plurality of cell nucleic acid barcode sequences. In some embodiments, a given partition nucleic acid barcode molecule of the plurality of partition nucleic acid barcode molecules comprises a priming sequence that is capable of hybridizing to a sequence of a given cell nucleic acid barcode molecule of the plurality of cell nucleic acid barcode molecules. In some embodiments, a given partition nucleic acid barcode molecule of the plurality of partition nucleic acid barcode molecules comprises a priming sequence that is capable of hybridizing to a sequence of each of the plurality of cell nucleic acid barcode molecules.

[0054] In some embodiments, the plurality of barcoded nucleic acid products is synthesized via one or more primer extension reactions. In some embodiments, the plurality of barcoded nucleic acid products is synthesized via one or more ligation reactions. In some embodiments, the plurality of barcoded nucleic acid products is synthesized via one or more nucleic acid amplification reactions. In some embodiments, the method further comprises sequencing the plurality of barcoded nucleic acid products or derivatives thereof to yield a plurality of sequencing reads. In some embodiments, the method further comprises associating each of the plurality of sequencing reads with an individual labelled cell of the plurality of labelled cells via its respective cell identification sequence, and associating each of the plurality of sequencing reads with an individual partition of the plurality of partitions via its respective partition identification sequence.

[0055] In some embodiments, the method further comprises, in (b), partitioning the plurality of labelled cells with a plurality of beads, wherein each of the plurality of beads comprises a partition nucleic acid barcode sequence of the plurality of partition nucleic acid barcode sequences. In some embodiments, each of the plurality of partitions comprises a single bead of the plurality of beads. In some embodiments, each of the plurality of beads comprises a plurality of partition nucleic acid barcode molecules, wherein each of the partition nucleic acid barcode molecules comprises a single partition nucleic acid barcode sequence of the plurality of partition nucleic acid barcode sequences. In some embodiments, each of the plurality of partition nucleic acid barcode sequences is releasably coupled to its respective bead of the plurality of beads. In some embodiments, the method further comprises, after (b), releasing partition nucleic acid barcode sequences from each of the plurality of beads. In some embodiments, the method further comprises degrading each of the plurality of beads to release the partition nucleic acid barcode sequences from each of the plurality of beads. In some embodiments, each of the plurality of partitions comprises an agent that is capable of degrading each of the plurality of beads. In some embodiments, the plurality of beads is a plurality of gel beads.

[0056] In some embodiments, the plurality of partitions is a plurality of droplets.

[0057] In some embodiments, the plurality of partitions is a plurality of wells.

[0058] In some embodiments, in (a), the plurality of cells is labelled with the plurality of cell nucleic acid barcode sequences by binding cell binding moieties, each coupled to a given cell nucleic acid barcode sequence of the plurality of cell nucleic acid barcode sequences, to each of the plurality of cells. In some embodiments, the cell binding moieties are antibodies, cell surface receptor binding molecules, receptor ligands, small molecules, pro-bodies, aptamers, monobodies, affimers, darpins or protein scaffolds. In some embodiments, the cell binding moieties are antibodies. In some embodiments, the cell binding moieties bind to a protein of cells of the plurality of cells. In some embodiments, the cell binding moieties bind to a cell surface species of cells of the plurality of cells. In some embodiments, the cell binding moieties bind to a species common to each of the plurality of cells.

[0059] In some embodiments, in (a), the plurality of cells is labelled with the plurality of cell nucleic acid barcode sequences by delivering nucleic acid barcode molecules each comprising an individual cell nucleic acid barcode sequence of the plurality of cell nucleic acid barcode sequences to each of the plurality of cells with the aid of a cell-penetrating peptide.

[0060] In some embodiments, in (a), the plurality of cells is labelled with the plurality of cell nucleic acid barcode sequences with the aid of liposomes, nanoparticles, electroporation, or mechanical force.

[0061] In another aspect, the present disclosure provides a method for analyzing cellular occupancy of a partition, comprising: (a) labelling a first cell with a first cell nucleic acid barcode sequence and a second cell with a second cell nucleic acid barcode sequence to generate a first labelled cell and a second labelled cell, wherein the first cell nucleic acid barcode sequence has a different sequence than the second cell nucleic acid barcode sequence; (b) generating a partition comprising the first labelled cell and the second labelled cell, wherein the partition further comprises a partition nucleic acid barcode sequence; and (c) generating (i) a first barcoded nucleic acid molecule comprising the first cell nucleic acid barcode sequence or a complement thereof and the partition

nucleic acid barcode sequence or a complement thereof and (ii) a second barcoded nucleic acid molecule comprising the second cell nucleic acid barcode sequence or a complement thereof and a partition nucleic acid barcode sequence or a complement thereof; and (d) identifying the first labelled cell and the second labelled cell as originating from the partition based on the first barcoded nucleic acid molecule and the second barcoded nucleic acid molecule having the same partition nucleic acid barcode sequence or a complement thereof.

[0062] In an aspect, the present disclosure provides a method for analyzing cellular occupancy of partitions, comprising: (a) providing a plurality of cell nucleic acid barcode molecules comprises a plurality of cell nucleic acid barcode sequences, each cell nucleic acid barcode molecule of the plurality of cell nucleic acid barcode molecules comprising (i) a single cell nucleic acid barcode sequence of the plurality of cell nucleic acid barcode sequences and (ii) a lipophilic moiety; (b) labeling a plurality of cells with the plurality of cell nucleic acid barcode sequences to generate a plurality of labeled cells, wherein each labeled cell of the plurality of labeled cells comprises a different cell nucleic acid barcode sequence of the plurality of cell nucleic acid barcode sequences; (c) generating a plurality of partitions comprising the plurality of labeled cells and a plurality of partition nucleic acid barcode sequences, wherein each partition of the plurality of partitions comprises a different partition nucleic barcode sequence of the plurality of partition nucleic acid barcode sequences, and wherein at least a fraction of the plurality of partitions comprises more than one labeled cell of the plurality of labeled cells; and (d) identifying at least two labeled cells of the plurality of labeled cells as originating from a same partition using (i) cell nucleic acid barcode sequences of the plurality of cell nucleic acid barcode sequences, or complements thereof, and (ii) partition nucleic acid barcode sequences of the plurality of partition nucleic acid barcode sequences, or complements thereof.

[0063] In some embodiments, a given cell nucleic acid barcode sequence of the plurality of cell nucleic acid barcode sequences identifies a sample from which an associated cell of the plurality of labeled cells originates. In some embodiments, the sample is derived from a biological fluid.

[0064] In some embodiments, the method further comprises after (c), synthesizing a plurality of barcoded nucleic acid products from the plurality of labeled cells, wherein a given barcoded nucleic acid product of the plurality of barcoded nucleic acid products comprises (i) a cell identification sequence comprising a given cell nucleic acid barcode sequence of the plurality of cell nucleic acid barcode sequences, or a complement of the given cell nucleic acid barcode sequence; and (ii) a partition identification sequence comprising a given partition nucleic acid barcode sequence of the plurality of partition nucleic acid barcode sequences, or a complement of the given partition nucleic acid barcode sequence. In some embodiments, a plurality of partition nucleic acid barcode molecules comprises the plurality of partition nucleic acid barcode sequences, each partition nucleic acid barcode molecule of the plurality of partition nucleic acid barcode molecules comprising a single partition nucleic acid barcode sequence of the plurality of partition nucleic acid barcode sequences. In some embodiments, a given partition nucleic acid barcode molecule of the plurality of partition nucleic acid barcode molecules comprises a priming sequence that is capable of hybridizing to a sequence of a given cell nucleic acid barcode molecule of the plurality of cell nucleic acid barcode molecules. In some embodiments, the priming sequence is a targeted priming sequence or a random N-mer sequence. In some embodiments, each cell nucleic acid barcode molecule of the plurality of cell nucleic acid barcode molecules comprises the sequence. In some embodiments, the plurality of barcoded nucleic acid products is synthesized via one or more primer extension reactions, ligation reactions, or one or more nucleic acid amplification reactions. In some embodiments, the method further comprises sequencing the plurality of barcoded nucleic acid products or derivatives thereof to yield a plurality of sequencing reads. In some embodiments, the method further comprises associating each sequencing read of the plurality of sequencing reads with a labeled cell of the plurality of labeled cells via its respective cell identification sequence, and associating each sequencing read of the plurality of sequencing reads with a partition of the plurality of partitions via its respective partition identification sequence.

[0065] In some embodiments, the method further comprises in (c), partitioning the plurality of labeled cells with a plurality of beads, wherein each bead of the plurality of beads comprises a partition nucleic acid barcode sequence of the plurality of partition nucleic acid barcode sequences. In some embodiments, each partition of the plurality of partitions comprises a single bead of the plurality of beads. In some embodiments, each bead of the plurality of beads comprises a plurality of partition nucleic acid barcode molecules, wherein each partition nucleic acid barcode molecule of the plurality of partition nucleic acid barcode molecules comprises a single partition nucleic acid barcode sequence of the plurality of partition nucleic acid barcode sequences. In some embodiments, each partition nucleic acid barcode sequence of the plurality of partition nucleic acid barcode sequences is releasably coupled to its respective bead of the plurality of beads. In some embodiments, each partition nucleic acid barcode sequence of the plurality of partition nucleic acid barcode sequences is releasable from its respective bead of the plurality of beads upon application of a stimulus. In some embodiments, the method further comprises after (c), releasing partition nucleic acid barcode sequences of the plurality of partition nucleic acid barcode sequences from each bead of the plurality of beads. In some embodiments, the method further comprises degrading each bead of the plurality of beads to release the partition nucleic acid barcode sequences from each bead of the plurality of beads. In some embodiments, each partition of the plurality of partitions comprises an agent that is capable of degrading each bead of the plurality of beads. In some embodiments, the plurality of beads is a plurality of gel beads.

[0066] In some embodiments, the plurality of partitions is a plurality of droplets or a plurality of wells.

[0067] In some embodiments, in (b), the plurality of cells is labeled with the plurality of cell nucleic acid barcode sequences by binding cell binding moieties, each coupled to a given cell nucleic acid barcode sequence of the plurality of cell nucleic acid barcode sequences, to each cell of the plurality of cells. In some embodiments, the cell binding moieties are antibodies, cell surface receptor binding mol-

ecules, receptor ligands, small molecules, pro-bodies, aptamers, monobodies, affimers, darpins, or protein scaffolds.

[0068] In some embodiments, in (b), the plurality of cells is labeled with the plurality of cell nucleic acid barcode sequences by delivering nucleic acid barcode molecules each comprising an individual cell nucleic acid barcode sequence of the plurality of cell nucleic acid barcode sequences to each cell of the plurality of cells with the aid of a cell-penetrating peptide.

[0069] In some embodiments, in (b), the plurality of cells is labeled with the plurality of cell nucleic acid barcode sequences with the aid of liposomes, nanoparticles, electroporation, or mechanical force.

[0070] In some embodiments, the lipophilic moiety of each nucleic acid barcode molecule of the plurality of cell nucleic acid barcode molecules is a cholesterol.

[0071] In some embodiments, the lipophilic moiety is linked to the plurality of cell nucleic acid barcode molecules via a linker.

[0072] In some embodiments, each cell of the plurality of cells comprises a plurality of nucleic acid molecules. In some embodiments, the labeled cells are lysed or permeabilized to provide access to the plurality of nucleic acid molecules. In some embodiments, a plurality of partition nucleic acid barcode molecules comprises the plurality of partition nucleic acid barcode sequences, each partition nucleic acid barcode molecule of the plurality of partition nucleic acid barcode molecules comprising a single partition nucleic acid barcode sequence of the plurality of partition nucleic acid barcode sequences and a priming sequence that is capable of hybridizing to a sequence of at least a subset of the plurality of nucleic acid molecules.

[0073] Additional aspects and advantages of the present disclosure will become readily apparent to those skilled in this art from the following detailed description, wherein only illustrative embodiments of the present disclosure are shown and described. As will be realized, the present disclosure is capable of other and different embodiments, and its several details are capable of modifications in various obvious respects, all without departing from the disclosure. Accordingly, the drawings and description are to be regarded as illustrative in nature, and not as restrictive.

INCORPORATION BY REFERENCE

[0074] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[0075] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings (also "Figure" and "FIG." herein), of which:

[0076] FIG. **1** schematically illustrates a microfluidic channel structure for partitioning individual or small groups of cells;

[0077] FIG. **2** schematically illustrates a microfluidic channel structure for co-partitioning cells and microcapsules (e.g., beads) comprising additional reagents;

[0078] FIGS. **3**A-**3**F schematically illustrate an example process for amplification and barcoding of cell's nucleic acids;

[0079] FIG. **4** provides a schematic illustration of use of barcoding of cell's nucleic acids in attributing sequence data to individual cells or groups of cells for use in their characterization;

[0080] FIG. **5** provides a schematic illustration of cells associated with labeled cell-binding ligands;

[0081] FIG. **6** provides a schematic illustration of an example workflow for performing RNA analysis using the methods described herein;

[0082] FIG. **7** provides a schematic illustration of an example barcoded oligonucleotide structure for use in analysis of ribonucleic (RNA) using the methods described herein;

[0083] FIG. **8** provides an image of individual cells co-partitioned along with individual barcode bearing beads;

[0084] FIGS. **9**A-**9**E provide schematic illustrations of example barcoded oligonucleotide structures for use in analysis of RNA and example operations for performing RNA analysis ("AAAAAAAAAAAAAAAA" disclosed as SEQ ID NO: 1);

[0085] FIG. **10** provides a schematic illustration of example barcoded oligonucleotide structure for use in example analysis of RNA and use of a sequence for in vitro transcription ("AAAAAAAAAAAAAAAA" disclosed as SEQ ID NO: 1);

[0086] FIG. **11** provides a schematic illustration of an example barcoded oligonucleotide structure for use in analysis of RNA and example operations for performing RNA analysis (SEQ ID NOS 2-3 and 2-3, respectively, in order of appearance);

[0087] FIGS. **12**A-**12**B provide schematic illustrations of example barcoded oligonucleotide structure for use in analysis of RNA;

[0088] FIGS. **13**A-**13**C provide illustrations of example yields from template switch reverse transcription and PCR in partitions;

[0089] FIGS. **14**A-**14**B provide illustrations of example yields from reverse transcription and cDNA amplification in partitions with various cell numbers;

[0090] FIG. **15** provides an illustration of example yields from cDNA synthesis and real-time quantitative PCR at various input cell concentrations and also the effect of varying primer concentration on yield at a fixed cell input concentration;

[0091] FIG. **16** provides an illustration of example yields from in vitro transcription;

[0092] FIG. **17** shows an example computer control system that is programmed or otherwise configured to implement methods provided herein;

[0093] FIG. **18** provides a schematic illustration of an example barcoded oligonucleotide structure;

[0094] FIG. **19** shows example operations for performing RNA analysis (SEQ ID NOS 2-3, 2-3, 3, 3, and 3, respectively, in order of appearance);

[0095] FIG. **20** shows a method for characterizing a cell, according to embodiments;

[0096] FIG. **21** shows an oligonucleotide with modifications that may prevent extension by a polymerase;

[0097] FIG. **22** shows oligonucleotides comprising a U-excising element;

[0098] FIG. **23**A shows a bead coupled with an oligonucleotide comprising a target-specific primer and oligonucleotides with poly-T primers (SEQ ID NOS 4-5, and 4, respectively, in order of appearance); FIG. **23**B shows a bead coupled with a plurality of oligonucleotides, each of which comprises a target-specific primer (SEQ ID NOS 5, 5, 5, and 5, respectively, in order of appearance); FIG. **23**C shows a bead coupled with a plurality of oligonucleotides, each of which comprises a target-specific primer and a plurality of oligonucleotides, each of which comprises a poly-T primer (SEQ ID NOS 4-5, 5, 5, 5, and 4, respectively, in order of appearance);

[0099] FIG. **24** shows a bead coupled with a plurality of oligonucleotides, each of which comprises a target-specific primer and a plurality of oligonucleotides, each of which comprises a random N-mer primer for total RNA (SEQ ID NOS 6-7, 7, 7, 7, and 6, respectively, in order of appearance);

[0100] FIGS. **25**A-**25**C show exemplary oligonucleotides comprising adapters and assay primers (SEQ ID NOS 8-9, respectively, in order of appearance);

[0101] FIG. **26** shows an oligonucleotide with an adapter comprising a switch oligo (SEQ ID NO: 10);

[0102] FIG. **27**A shows oligonucleotides with backbones comprising P7 and R2 sequences and poly-T primers (SEQ ID NOS 11, 41, 12, 42, 13, 43, 14, 44, and 11, respectively, in order of appearance). FIG. **27**B shows y oligonucleotides with backbones comprising R1 sequences and poly-T primers (SEQ ID NOS 15, 45, 16, 46, 17, 47, 4, 48, and 4, respectively, in order of appearance). FIG. **27**C shows oligonucleotides with P5, R1, and R2 sequences and poly-T primers (SEQ ID NOS 18, 49, 18-19, 50, and 19, respectively, in order of appearance). FIG. **27**D shows oligonucleotides with R1 sequences and random N-mer primers (SEQ ID NOS 20, 51, 21, 51, 22, 51, 6, 51, and 6, respectively, in order of appearance).

[0103] FIG. **28** shows a workflow for conjugating a DNA barcode on an antibody using an antibody-binding protein;

[0104] FIG. **29** demonstrates swelling conditions and deswelling conditions in the process of making gel beads with magnetic particles;

[0105] FIG. **30** shows a unit cell comprising a scaffold and liquid immediately surrounding the scaffold;

[0106] FIG. **31** shows a microcapsule with a barcoded magnetic particle entrapped;

[0107] FIG. **32** shows a method for parallel sequencing DNA molecules and RNA molecules in a cell;

[0108] FIG. **33** shows various approaches for making antibody-reporter oligonucleotide conjugates;

[0109] FIG. **34** shows an antibody-reporter oligonucleotide conjugation;

[0110] FIGS. **35**A-**35**C show a method for analyzing mRNA molecules and proteins from a single cell ("AAAAAAAAAAAAAAAAAAAAA" disclosed as SEQ ID NO: 23);

[0111] FIG. **36**A shows a relationship between a diameter of a gel bead and a regent inside the gel bead; FIG. **36**B shows the relationship between the diameter of a gel bead and the number of droplets with more than one cell;

[0112] FIG. **37** shows analysis results of the CD3 protein-single-stranded DNA (ssDNA) conjugate;

[0113] FIG. **38** shows the fluorescence signals from the cells bound by labeled antibodies;

[0114] FIG. **39**A shows an approach for conjugating an oligonucleotide with an antibody; FIG. **39**B shows analysis results of barcoded antibodies;

[0115] FIG. **40** shows a conjugate of a functionalized antibody-binding protein and a functionalized oligonucleotide;

[0116] FIG. **41** shows a relationship between a degree of dibenzocyclooctyne (DBCO) incorporation and input dibenzocyclooctyne-N-hydroxysuccinimidyl ester (DBCO-NHS) concentrations;

[0117] FIG. **42** shows an example relationship between the degree of conjugation and oligonucleotide equivalence;

[0118] FIG. **43** shows fluorescence signals of labeled cells measured by flow cytometry;

[0119] FIG. **44** shows a method for producing a bead coupled with oligonucleotides with different primer sequences (SEQ ID NOS 24-26, 25-28, 15, 45, 29, and 45, respectively, in order of appearance);

[0120] FIG. **45**A shows a bead coupled with a plurality of oligonucleotides (SEQ ID NOS 30, 30, 30-31, 30, 30, 30, 30, 30, 30, 30-31, 30, and 30, respectively, in order of appearance). FIG. **45**B shows results from gel electrophoresis analysis of beads. On the beads, 0%, 5%, 15%, or 25% of coupled oligonucleotides contain antibody target primers;

[0121] FIGS. **46**A-**46**E schematically depict components of example multi-assay schemes described herein;

[0122] FIGS. **47**A-**47**B depicts data obtained from an example experiment described in Example XI;

[0123] FIG. **48** depicts data obtained from an example experiment described in Example XI;

[0124] FIGS. **49**A and **49**B depict data obtained from an example experiment described in Example XI;

[0125] FIG. **50**A schematically depicts an example bead comprising oligonucleotides having two different functional sequences (SEQ ID NOS 24 and 32, respectively, in order of appearance);

[0126] FIGS. **50**B and **50**C schematically depict example sequences that can be coupled to a bead (SEQ ID NOS 33, 52, 34, 53, 35, 54, 36, 55, 37, 56, 37, 56, 37, 56, 37, and 56, respectively, in order of appearance);

[0127] FIG. **51**A depicts sequences (SEQ ID NOS 38 and 39, respectively, in order of appearance) used in an example experiment described in Example XII; FIG. **51**B graphically depicts data from an example experiment described in Example XII;

[0128] FIG. **52**A depicts data obtained from an example experiment described in Example XIII;

[0129] FIG. **52**B schematically depicts example extension schemes to link barcodes;

[0130] FIGS. **53**A and **53**B provide data obtained from an example experiment described in Example XIII;

[0131] FIGS. **54** and **55** provide data obtained from example experiments described in Example XIV; and

[0132] FIGS. **56**A-**56**C schematically depict an example barcoding scheme that includes major histocompatibility complexes.

[0133] FIGS. **57**A-**57**B graphically depicts an exemplary barcoded streptavidin complex. (SEQ ID NOS 57-58 and 57, respectively, in order of appearance).

[0134] FIGS. **58**A-**58**B illustrates an exemplary analysis of barcoded streptavidin complexes. FIG. **58**A shows a

representative denaturing agarose gel while FIG. **58**B shows a representative SDS-PAGE gel.

[0135] FIG. **59** shows results of data obtained from an example barcoded MHC tetramer T-cell experiment as described in Example XV.

[0136] FIG. **60** shows results of data obtained from example EBV-expanded T-cell spike-in experiment as described in Example XV.

[0137] FIGS. **61**A-**61**D schematically depict an example barcoding scheme of CRISPR guide RNAs (SEQ ID NOS 59-60, 59-60, 59-60, 59, and 61, respectively, in order of appearance).

[0138] FIG. **62** shows an example of a microfluidic channel structure for co-partitioning biological particles and reagents.

[0139] FIG. **63** shows an example of a microfluidic channel structure for the controlled partitioning of beads into discrete droplets.

[0140] FIG. **64** shows an example of a microfluidic channel structure for increased droplet generation throughput.

[0141] FIG. **65** shows another example of a microfluidic channel structure for increased droplet generation throughput.

[0142] FIG. **66**A shows an example arrangement of nine sets of nucleic acid barcode molecules arranged in a two-dimensional configuration; FIG. **66**B shows an example of a sample overlaying a two-dimensional arrangement of nucleic acid barcode molecules.

[0143] FIG. **67** shows an exemplary lipophilic moiety-conjugated-feature barcode comprising a cholesterol, a linker, and a nucleic acid attachment region.

[0144] FIG. **68** schematically depicts representative lipophilic barcodes as well as exemplary nucleic acid extension schemes to couple cell barcodes to lipophilic barcodes.

[0145] FIGS. **69**A-**69**B show BioAnalyzer results of barcode libraries prepared from a first cell population (FIG. **69**A) and a second cell population (FIG. **69**B) incubated with ~1 uM of feature barcodes without a lipophilic moiety while FIGS. **69**C-**69**D show BioAnalyzer results of barcode libraries prepared from a first cell population (FIG. **69**C) and a second cell population (FIG. **69**D) incubated with ~1 uM of cholesterol-conjugated feature barcodes.

[0146] FIGS. **70**A-**70**J show representative graphs from pooled cell populations incubated with 0.1 μM cholesterol-conjugated feature barcodes showing the number of unique molecular identifier (UMI) counts on the x-axis versus number of cells on the y-axis. FIGS. **70**A-**70**B show log 10 UMI counts of a first feature barcode sequence ("BC1") identified from sequencing reads generated from sequencing libraries prepared from the pooled cell population (FIG. **70**A—replicate 1; FIG. **70**B—replicate 2). FIGS. **70**C-**70**D show log 10 UMI counts of a second feature barcode sequence ("BC2") identified from sequencing reads generated from sequencing libraries prepared from the pooled cell population (FIG. **70**C—replicate 1; FIG. **70**D—replicate 2). FIGS. **70**E-**70**F show log 10 UMI counts of a third feature barcode sequence ("BC3") identified from sequencing reads generated from sequencing libraries prepared from the pooled cell population (FIG. **70**E—replicate 1; FIG. **70**F—replicate 2). FIGS. **70**G-**70**H show log 10 UMI counts of a fourth feature barcode sequence ("BC4") identified from sequencing reads generated from sequencing libraries prepared from the pooled cell population (FIG. **70**G—replicate 1; FIG. **70**H—replicate 2). FIGS. **70**I-**70**J show 3D representations of UMI counts obtained from the pooled cell populations for replicate 1. Graphs depict UMI counts in linear (FIG. **70**I) and in log 10 scale (FIG. **70**J).

[0147] FIG. **71**A-**71**J show representative graphs from pooled cell populations incubated with 0.01 μM cholesterol-conjugated feature barcodes showing the number of unique molecular identifier (UMI) counts on the x-axis versus number of cells on the y-axis. FIGS. **71**A-**71**B show log 10 UMI counts of a first feature barcode sequence ("BC1") identified from sequencing reads generated from sequencing libraries prepared from the pooled cell population (FIG. **71**A—replicate 1; FIG. **71**B—replicate 2). FIGS. **71**C-**71**D show log 10 UMI counts of a second feature barcode sequence ("BC2") identified from sequencing reads generated from sequencing libraries prepared from the pooled cell population (FIG. **71**C—replicate 1; FIG. **71**D—replicate 2). FIGS. **71**E-**71**F show log 10 UMI counts of a third feature barcode sequence ("BC3") identified from sequencing reads generated from sequencing libraries prepared from the pooled cell population (FIG. **71**E—replicate 1; FIG. **71**F—replicate 2). FIGS. **71**G-**71**H show log 10 UMI counts of a fourth feature barcode sequence ("BC4") identified from sequencing reads generated from sequencing libraries prepared from the pooled cell population (FIG. **71**G—replicate 1; FIG. **71**H—replicate 2). FIGS. **71**I-**71**J show 3D representations of UMI counts obtained from the pooled cell populations for replicate 1. Graphs depict UMI counts in linear (FIG. **71**I) and in log 10 scale (FIG. **71**J).

[0148] FIGS. **72**A-**72**I show representative graphs from pooled cell populations incubated with antibody-conjugated feature barcodes showing the number of unique molecular identifier (UMI) counts on the x-axis versus number of cells on the y-axis. FIGS. **72**A-**72**B show UMI counts of a first feature barcode sequence ("BC18") identified from sequencing reads generated from sequencing libraries prepared from the pooled cell population (FIG. **72**A—replicate 1; FIG. **72**B—replicate 2). From these results, a clearly distinguished BC18-containing cell population can be distinguished **7201***a* (replicate 1) and **7201***b* (replicate 2). FIGS. **72**C-**72**D show UMI counts of a second feature barcode sequence ("BC19") identified from sequencing reads generated from sequencing libraries prepared from the pooled cell population (FIG. **72**C—replicate 1; FIG. **72**D—replicate 2). From these results, a clearly distinguished BC19-containing cell population can be distinguished **7202***a* (replicate 1) and **7202***b* (replicate 2). FIGS. **72**E-**72**F show UMI counts of a third feature barcode sequence ("BC20") identified from sequencing reads generated from sequencing libraries prepared from the pooled cell population (FIG. **72**E—replicate 1; FIG. **72**F—replicate 2). From these results, a clearly distinguished BC20-containing cell population can be distinguished **7203***a* (replicate 1) and **7203***b* (replicate 2). FIG. **72**G shows UMI counts of feature barcode sequences identified from sequencing reads generated from sequencing libraries prepared from the pooled cell population with log 10 UMI counts for BC18 on the y-axis and log 10 UMI counts for BC20 on the x-axis. FIG. **72**H shows UMI counts of feature barcode sequences identified from sequencing reads generated from sequencing libraries prepared from the pooled cell population with log 10 UMI counts for BC18 on the y-axis and log 10 UMI counts for BC19 on the x-axis. FIG. **72**I shows UMI counts of feature barcode sequences identified from sequencing reads generated from sequencing libraries prepared from the pooled cell

population with log 10 UMI counts for BC19 on the y-axis and log 10 UMI counts for BC20 on the x-axis.

[0149] FIGS. **73**A-**73**B show clustering of UMI counts prepared using antibody t-distributed stochastic neighbor embedding (t-SNE) (FIG. **73**A), as well as in gene expression (GEX) t-SNE analyses (FIG. **73**B).

[0150] FIG. **74** depicts an example of a tissue section with barcode staining using a fixed array of needles.

[0151] FIG. **75** depicts a diffusion map to spatially localize barcodes and associated cells.

[0152] FIG. **76** shows the position of cells (designated "C1" to "C7") defined by a barcode and its relative amount.

[0153] FIG. **77** depicts a three dimensional application of spatial mapping.

[0154] FIG. **78** depicts a three dimensional application of spatial mapping.

[0155] FIG. **79**A depicts regions of a mouse brain with delivery devices for delivering barcode molecules.

[0156] FIG. **79**B shows a pattern for injection of barcodes to a sample.

[0157] FIG. **80** shows a correlation between cell diameter and cell surface area.

[0158] FIG. **81** shows the uptake of lipophilic barcodes of given cell diameters (μm).

[0159] FIG. **82** shows an example graph of barcode counts vs. cell counts.

[0160] FIG. **83** shows a schematic for enriching V(D)J sequences from immune molecules such as TCRs, BCRs, and immunoglobulins.

[0161] FIGS. **84**A and **84**B show variations of a schematic for generating labeled polynucleotides.

[0162] FIG. **85** shows a schematic for enhanced cell multiplexing.

[0163] FIG. **86** shows an exemplary fluorophore-conjugated-feature barcode molecule.

[0164] FIG. **87** shows exemplary nucleic acid barcode molecules comprising different capture sequences.

[0165] FIG. **88** shows exemplary moiety conjugated oligonucleotides.

DETAILED DESCRIPTION

[0166] While various embodiments of the invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions may occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed.

[0167] Where values are described as ranges, it will be understood that such disclosure includes the disclosure of all possible sub-ranges within such ranges, as well as specific numerical values that fall within such ranges irrespective of whether a specific numerical value or specific sub-range is expressly stated.

[0168] The term "barcode," as used herein, generally refers to a label, or identifier, that can be part of an analyte to convey information about the analyte. A barcode can be a tag attached to an analyte (e.g., nucleic acid molecule) or a combination of the tag in addition to an endogenous characteristic of the analyte (e.g., size of the analyte or end sequence(s)). The barcode may be unique. Barcodes can have a variety of different formats, for example, barcodes can include: polynucleotide barcodes; random nucleic acid and/or amino acid sequences; and synthetic nucleic acid and/or amino acid sequences. A barcode can be attached to an analyte in a reversible or irreversible manner. A barcode can be added to, for example, a fragment of a deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) sample before, during, and/or after sequencing of the sample. Barcodes can allow for identification and/or quantification of individual sequencing-reads in real time.

[0169] The term "subject," as used herein, generally refers to an animal, such as a mammalian species (e.g., human) or avian (e.g., bird) species, or other organism, such as a plant. The subject can be a vertebrate, a mammal, a mouse, a primate, a simian or a human. Animals may include, but are not limited to, farm animals, sport animals, and pets. A subject can be a healthy individual, an individual that has or is suspected of having a disease or a pre-disposition to the disease, or an individual that is in need of therapy or suspected of needing therapy. A subject can be a patient.

[0170] The term "genome," as used herein, generally refers to an entirety of a subject's hereditary information. A genome can be encoded either in DNA or in RNA. A genome can comprise coding regions that code for proteins as well as non-coding regions. A genome can include the sequence of all chromosomes together in an organism. For example, the human genome has a total of 46 chromosomes. The sequence of all of these together may constitute a human genome.

[0171] The terms "adaptor(s)", "adapter(s)" and "tag(s)" may be used synonymously. An adaptor or tag can be coupled to a polynucleotide sequence to be "tagged" by any approach including ligation, hybridization, or other approaches.

[0172] The term "sequencing," as used herein, generally refers to methods and technologies for determining the sequence of nucleotide bases in one or more polynucleotides. The polynucleotides can be, for example, deoxyribonucleic acid (DNA) or ribonucleic acid (RNA), including variants or derivatives thereof (e.g., single stranded DNA). Sequencing can be performed by various systems currently available, such as, without limitation, a sequencing system by Illumina, Pacific Biosciences, Oxford Nanopore, or Life Technologies (Ion Torrent). Such devices may provide a plurality of raw genetic data corresponding to the genetic information of a subject (e.g., human), as generated by the device from a sample provided by the subject. In some situations, systems and methods provided herein may be used with proteomic information.

[0173] The term "variant," as used herein, generally refers to a genetic variant, such as a nucleic acid molecule comprising a polymorphism. A variant can be a structural variant or copy number variant, which can be genomic variants that are larger than single nucleotide variants or short indels. A variant can be an alteration or polymorphism in a nucleic acid sample or genome of a subject. Single nucleotide polymorphisms (SNPs) are a form of polymorphisms. Polymorphisms can include single nucleotide variations (SNVs), insertions, deletions, repeats, small insertions, small deletions, small repeats, structural variant junctions, variable length tandem repeats, and/or flanking sequences. Copy number variants (CNVs), transversions and other rearrangements are also forms of genetic variation. A genomic alteration may be a base change, insertion, deletion, repeat, copy number variation, or transversion.

[0174] The term "bead," as used herein, generally refers to a particle. The bead may be a solid or semi-solid particle. The bead may be a gel. The bead may be formed of a polymeric material. The bead may be magnetic or non-magnetic.

[0175] The term "sample," as used herein, generally refers to a biological sample of a subject. The sample may be a tissue sample, such as a biopsy, core biopsy, needle aspirate, or fine needle aspirate. The sample may be a fluid sample, such as a blood sample, urine sample, or saliva sample. The sample may be a skin sample. The sample may be a cheek swab. The sample may be a plasma or serum sample. The sample may be a cell-free (or cell free) sample. A cell-free sample may include extracellular polynucleotides. Extracellular polynucleotides may be isolated from a bodily sample that may be selected from a group consisting of blood, plasma, serum, urine, saliva, mucosal excretions, sputum, stool and tears.

[0176] The term "nucleic acid," as used herein, generally refers to a monomeric or polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs or variants thereof. A nucleic acid molecule may include one or more unmodified or modified nucleotides. Nucleic acid may have any three dimensional structure, and may perform any function. The following are non-limiting examples of nucleic acids: ribonucleic acid (RNA), deoxyribonucleic acid (DNA), coding or non-coding regions of a gene or gene fragment, loci (locus) defined from linkage analysis, exons, introns, messenger RNA (mRNA), transfer ribonucleic acid (RNA), ribosomal RNA, short interfering RNA (siRNA), short-hairpin RNA (shRNA), micro-RNA (miRNA), ribozymes, complementary deoxyribonucleic acid (cDNA), recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers. Nucleic acid may comprise one or more modified nucleotides, such as methylated nucleotides and nucleotide analogs, such as peptide nucleic acid (PNA), Morpholino and locked nucleic acid (LNA), glycol nucleic acid (GNA), threose nucleic acid (TNA), 2'-fluoro, 2'-OMe, and phosphorothiolated DNA. A nucleic acid may include one or more subunits selected from adenosine (A), cytosine (C), guanine (G), thymine (T) and uracil (U), or variants thereof. In some examples, a nucleic acid is DNA or RNA, or derivatives thereof. A nucleic acid may be single-stranded or double stranded. A nucleic acid may be circular.

[0177] The term "nucleotide," as used herein, generally refers to a nucleic acid subunit, which may include A, C, G, T or U, or variants or analogs thereof. A nucleotide can include any subunit that can be incorporated into a growing nucleic acid strand. Such subunit can be an A, C, G, T, or U, or any other subunit that is specific to one or more complementary A, C, G, T or U, or complementary to a purine (i.e., A or G, or variant or analogs thereof) or a pyrimidine (i.e., C, T or U, or variant or analogs thereof). A subunit can enable individual nucleic acid bases or groups of bases (e.g., AA, TA, AT, GC, CG, CT, TC, GT, TG, AC, CA, or uracil-counterparts thereof) to be resolved.

[0178] The term "analyte," as used herein, generally refers to a substance or one or more constituents thereof that is for identification, such as detection (e.g., detection via sequencing). Examples of analytes include, without limitation, DNA, RNA, a labelling agent, antibody, and protein. An analyte may be a cell or one or more constituents of a cell.

[0179] Analytes may be of different types. In some examples, in a plurality of analytes, a given analyte is of a different structural or functional class from other analytes of the plurality. Examples of different types of analytes include DNA and RNA; a nucleic acid molecule and a labelling agent; a transcript and genomic nucleic acid; a plurality of nucleic acid molecules, where each nucleic acid molecule has a different function, such as a different cellular function. A sample may have a plurality of analytes of different types, such as a mixture of DNA and RNA molecules, or a mixture of nucleic acid molecules and labelling agents. In some cases, different types of analytes do not include labelling agents directed to separate cell surface features of a cell.

[0180] The term "epitope binding fragment," as used herein generally refers to a portion of a complete antibody capable of binding the same epitope as the complete antibody, albeit not necessarily to the same extent. Although multiple types of epitope binding fragments are possible, an epitope binding fragment typically comprises at least one pair of heavy and light chain variable regions (VH and VL, respectively) held together (e.g., by disulfide bonds) to preserve the antigen binding site, and does not contain all or a portion of the Fc region. Epitope binding fragments of an antibody can be obtained from a given antibody by any suitable technique (e.g., recombinant DNA technology or enzymatic or chemical cleavage of a complete antibody), and typically can be screened for specificity in the same manner in which complete antibodies are screened. In some embodiments, an epitope binding fragment comprises an $F(ab')_2$ fragment, Fab' fragment, Fab fragment, Fd fragment, or Fv fragment. In some embodiments, the term "antibody" includes antibody-derived polypeptides, such as single chain variable fragments (scFv), diabodies or other multimeric scFvs, heavy chain antibodies, single domain antibodies, or other polypeptides comprising a sufficient portion of an antibody (e.g., one or more complementarity determining regions (CDRs)) to confer specific antigen binding ability to the polypeptide.

[0181] Nucleic acid sequencing technologies have yielded substantial results in sequencing biological materials, including providing substantial sequence information on individual organisms, and relatively pure biological samples. However, these systems have traditionally not been effective at being able to identify and characterize cells at the single cell level.

[0182] Nucleic acid sequencing technologies may derive the nucleic acids that they sequence from collections of cells obtained from tissue or other samples, such as biological fluids (e.g., blood, plasma, etc). The cells can be processed (e.g., all together in an ensemble approach) to extract the genetic material that represents an average of the population of cells, which can then be processed into sequencing ready DNA libraries that are configured for a given sequencing technology. Although often discussed in terms of DNA or nucleic acids, the nucleic acids derived from the cells may include DNA, or RNA, including, e.g., mRNA, total RNA, or the like, that may be processed to produce cDNA for sequencing.

[0183] In addition to the inability to attribute characteristics to particular subsets of cells or individual cells, such ensemble sample preparation methods can be, from the outset, predisposed to primarily identifying and characterizing the majority constituents in the sample of cells, and may not be designed to pick out the minority constituents,

e.g., genetic or proteomic material contributed by one cell, a few cells, or a small percentage of total cells in the sample. Likewise, where analyzing expression levels, e.g., of mRNA or cell surface proteins, an ensemble approach can be predisposed to presenting potentially inaccurate data from cell populations that are non-homogeneous in terms of expression levels. In some cases, where expression is high in a small minority of the cells in an analyzed population, and absent in the majority of the cells of the population, an ensemble method may indicate low level expression for the entire population.

[0184] These inaccuracies can be further magnified through processing operations used in generating the sequencing libraries from these samples. Some next generation sequencing technologies (e.g., massively parallel sequencing) may rely upon the geometric amplification of nucleic acid fragments, such as via polymerase chain reaction, in order to produce sufficient DNA for the sequencing library. However, such amplification can be biased toward amplification of majority constituents in a sample, and may not preserve the starting ratios of such minority and majority components. While some of these difficulties may be addressed by utilizing different sequencing systems, such as single molecule systems that do not require amplification, the single molecule systems, as well as the ensemble sequencing methods of other next generation sequencing (NGS) systems, can also have large input DNA requirements. Some single molecule sequencing systems, for example, can have sample input DNA requirements of from 500 nanograms (ng) to upwards of 10 micrograms (μg), which may not be obtainable from individual cells or small subpopulations of cells. Likewise, other NGS systems can be optimized for starting amounts of sample DNA in the sample of from approximately 50 nanograms (ng) to about 1 microgram (μg). Starting amounts of DNA may be at least about 1 ng, 10 ng, 20 ng, 30 ng, 40 ng, 50 ng, 100 ng, 500 ng, 1 μg, 10 μg, or 100 μg.

[0185] Disclosed herein are methods and systems for characterizing surface features, proteins, and nucleic acids of small populations of cells, and in some cases, for characterizing surface features, proteins, and nucleic acids of individual cells. The methods described herein may compartmentalize the analysis of individual cells or small populations of cells, including e.g., cell surface features, proteins, and nucleic acids of individual cells or small groups of cells, and then allow that analysis to be attributed back to the individual cell or small group of cells from which the cell surface features, proteins, and nucleic acids were derived. This can be accomplished regardless of whether the cell population represents a 50/50 mix of cell types, a 90/10 mix of cell types, or virtually any ratio of cell types, as well as a complete heterogeneous mix of different cell types, or any mixture between these. Differing cell types may include cells from different tissue types of an individual or the same tissue type from different individuals, or biological organisms such as microorganisms from differing genera, species, strains, variants, or any combination of any or all of the foregoing. For example, differing cell types may include normal and tumor tissue from an individual, various cell types obtained from a human subject such as a variety of immune cells (e.g., B cells, T cells, and the like), multiple different bacterial species, strains and/or variants from environmental, forensic, microbiome or other samples, or any of a variety of other mixtures of cell types.

[0186] In one aspect, the methods and systems described herein provide for the compartmentalization, depositing or partitioning of the nucleic acid contents of individual cells from a sample material containing cells, into discrete compartments or partitions (referred to interchangeably herein as partitions), where each partition maintains separation of its own contents from the contents of other partitions. In another aspect, the methods and system described herein provide for the compartmentalization, depositing or partitioning of individual cells from a sample material containing cells, into discrete partitions, where each partition maintains separation of its own contents from the contents of other partitions. In another aspect, the methods and system described herein provide for the compartmentalization, depositing or partitioning of individual cells from a sample material containing cells after at least one labelling agent has been bound to a cell surface feature of the cell, into discrete partitions, where each partition maintains separation of its own contents from the contents of other partitions. Unique identifiers, e.g., barcodes, may be previously, subsequently or concurrently delivered to the partitions that hold the compartmentalized or partitioned cells, in order to allow for the later attribution of the characteristics of the individual cells to the particular compartment. Further, unique identifiers, e.g., barcodes, may be coupled to labelling agents and previously, subsequently or concurrently delivered to the partitions that hold the compartmentalized or partitioned cells, in order to allow for the later attribution of the characteristics of the individual cells to the particular compartment. Barcodes may be delivered, for example on an oligonucleotide, to a partition via any suitable mechanism.

[0187] In some embodiments, barcoded oligonucleotides are delivered to a partition via a microcapsule. In some cases, barcoded oligonucleotides are initially associated with the microcapsule and then released from the microcapsule upon application of a stimulus which allows the oligonucleotides to dissociate or to be released from the microcapsule. In some embodiments, anchor oligonucleotides are delivered to a partition via a microcapsule. In some cases, anchor oligonucleotides are initially associated with the microcapsule and then released from the microcapsule upon application of a stimulus which allows the anchor oligonucleotides to dissociate or to be released from the microcapsule.

[0188] A microcapsule may be or may include a solid support or solid particle such as a bead. A solid support or a solid particle may be a bead. A microcapsule may be a droplet. A microcapsule, in some embodiments, may be or may comprise a bead. In some embodiments, a bead may be porous, non-porous, solid, semi-solid, semi-fluidic, or fluidic. In some embodiments, a bead may be dissolvable, disruptable, or degradable. In some cases, a bead may not be degradable. In some embodiments, the bead may be a gel bead. A gel bead may be a hydrogel bead. A gel bead may be formed from molecular precursors, such as a polymeric or monomeric species. A semi-solid bead may be a liposomal bead. Solid beads may comprise metals including iron oxide, gold, and silver. In some cases, the beads may be silica beads. In some cases, the beads may be rigid. In some cases, the beads may be flexible and/or compressible.

[0189] In some embodiments, the bead may contain molecular precursors (e.g., monomers or polymers), which may form a polymer network via polymerization of the precursors. In some cases, a precursor may be an already

polymerized species capable of undergoing further polymerization via, for example, a chemical cross-linkage. In some cases, a precursor comprises one or more of an acrylamide or a methacrylamide monomer, oligomer, or polymer. In some cases, the bead may comprise prepolymers, which are oligomers capable of further polymerization. For example, polyurethane beads may be prepared using prepolymers. In some cases, the bead may contain individual polymers that may be further polymerized together. In some cases, beads may be generated via polymerization of different precursors, such that they comprise mixed polymers, co-polymers, and/or block co-polymers.

[0190] A bead may comprise natural and/or synthetic materials. For example, a polymer can be a natural polymer or a synthetic polymer. In some cases, a bead may comprise both natural and synthetic polymers. Examples of natural polymers include proteins and sugars such as deoxyribonucleic acid, rubber, cellulose, starch (e.g., amylose, amylopectin), proteins, enzymes, polysaccharides, silks, polyhydroxyalkanoates, chitosan, dextran, collagen, carrageenan, ispaghula, acacia, agar, gelatin, shellac, sterculia gum, xanthan gum, Corn sugar gum, guar gum, gum karaya, agarose, alginic acid, alginate, or natural polymers thereof. Examples of synthetic polymers include acrylics, nylons, silicones, spandex, viscose rayon, polycarboxylic acids, polyvinyl acetate, polyacrylamide, polyacrylate, polyethylene glycol, polyurethanes, polylactic acid, silica, polystyrene, polyacrylonitrile, polybutadiene, polycarbonate, polyethylene, polyethylene terephthalate, poly(chlorotrifluoroethylene), poly(ethylene oxide), poly(ethylene terephthalate), polyethylene, polyisobutylene, poly(methyl methacrylate), poly(oxymethylene), polyformaldehyde, polypropylene, polystyrene, poly(tetrafluoroethylene), poly (vinyl acetate), poly(vinyl alcohol), poly(vinyl chloride), poly(vinylidene dichloride), poly(vinylidene difluoride), poly(vinyl fluoride) and combinations (e.g., co-polymers) thereof. Beads may also be formed from materials other than polymers, including lipids, micelles, ceramics, glass-ceramics, material composites, metals, other inorganic materials, and others.

[0191] In some cases, a chemical cross-linker may be a precursor used to cross-link monomers during polymerization of the monomers and/or may be used to attach oligonucleotides (e.g., barcoded oligonucleotides) to the bead. In some cases, polymers may be further polymerized with a cross-linker species or other type of monomer to generate a further polymeric network. Non-limiting examples of chemical cross-linkers (also referred to as a "crosslinker" or a "crosslinker agent" herein) include cystamine, gluteraldehyde, dimethyl suberimidate, N-Hydroxysuccinimide crosslinker BS3, formaldehyde, carbodiimide (EDC), SMCC, Sulfo-SMCC, vinylsilane, N,N'diallyltartardiamide (DATD), N,N'-Bis(acryloyl)cystamine (BAC), or homologs thereof. In some cases, the crosslinker used in the present disclosure contains cystamine.

[0192] Crosslinking may be permanent or reversible, depending upon the particular crosslinker used. Reversible crosslinking may allow for the polymer to linearize or dissociate under appropriate conditions. In some cases, reversible cross-linking may also allow for reversible attachment of a material bound to the surface of a bead. In some cases, a cross-linker may form disulfide linkages. In some cases, the chemical cross-linker forming disulfide linkages may be cystamine or a modified cystamine.

[0193] In some embodiments, disulfide linkages can be formed between molecular precursor units (e.g., monomers, oligomers, or linear polymers) or precursors incorporated into a bead and oligonucleotides. Cystamine (including modified cystamines), for example, is an organic agent comprising a disulfide bond that may be used as a cross-linker agent between individual monomeric or polymeric precursors of a bead. Polyacrylamide may be polymerized in the presence of cystamine or a species comprising cystamine (e.g., a modified cystamine) to generate polyacrylamide gel beads comprising disulfide linkages (e.g., chemically degradable beads comprising chemically-reducible cross-linkers). The disulfide linkages may permit the bead to be degraded (or dissolved) upon exposure of the bead to a reducing agent.

[0194] In some embodiments, chitosan, a linear polysaccharide polymer, may be crosslinked with glutaraldehyde via hydrophilic chains to form a bead. Crosslinking of chitosan polymers may be achieved by chemical reactions that are initiated by heat, pressure, change in pH, and/or radiation.

[0195] In some embodiments, the bead may comprise covalent or ionic bonds between polymeric precursors (e.g., monomers, oligomers, linear polymers), oligonucleotides, primers, and other entities. In some cases, the covalent bonds comprise carbon-carbon bonds or thioether bonds.

[0196] In some cases, a bead may comprise an acrydite moiety, which in certain aspects may be used to attach one or more oligonucleotides (e.g., barcode sequence, barcoded oligonucleotide, primer, or other oligonucleotide) to the bead. In some cases, an acrydite moiety can refer to an acrydite analogue generated from the reaction of acrydite with one or more species, such as, the reaction of acrydite with other monomers and cross-linkers during a polymerization reaction. Acrydite moieties may be modified to form chemical bonds with a species to be attached, such as an oligonucleotide (e.g., barcode sequence, barcoded oligonucleotide, primer, or other oligonucleotide). Acrydite moieties may be modified with thiol groups capable of forming a disulfide bond or may be modified with groups already comprising a disulfide bond. The thiol or disulfide (via disulfide exchange) may be used as an anchor point for a species to be attached or another part of the acrydite moiety may be used for attachment. In some cases, attachment is reversible, such that when the disulfide bond is broken (e.g., in the presence of a reducing agent), the attached species is released from the bead. In other cases, an acrydite moiety comprises a reactive hydroxyl group that may be used for attachment.

[0197] Functionalization of beads for attachment of oligonucleotides may be achieved through a wide range of different approaches, including activation of chemical groups within a polymer, incorporation of active or activatable functional groups in the polymer structure, or attachment at the pre-polymer or monomer stage in bead production.

[0198] For example, precursors (e.g., monomers, cross-linkers) that are polymerized to form a bead may comprise acrydite moieties, such that when a bead is generated, the bead also comprises acrydite moieties. The acrydite moieties can be attached to an oligonucleotide, such as a primer (e.g., a primer for amplifying target nucleic acids, barcoded oligonucleotide, etc) to be incorporated into the bead. In some cases, the primer comprises a P5 sequence for attachment to

a sequencing flow cell for Illumina sequencing. In some cases, the primer comprises a P7 sequence for attachment to a sequencing flow cell for Illumina sequencing. In some cases, the primer comprises a barcode sequence. In some cases, the primer further comprises a unique molecular identifier (UMI). In some cases, the primer comprises an R1 primer sequence for Illumina sequencing. In some cases, the primer comprises an R2 primer sequence for Illumina sequencing.

[0199] In some cases, precursors comprising a functional group that is reactive or capable of being activated such that it becomes reactive can be polymerized with other precursors to generate gel beads comprising the activated or activatable functional group. The functional group may then be used to attach additional species (e.g., disulfide linkers, primers, other oligonucleotides, etc.) to the gel beads. For example, some precursors comprising a carboxylic acid (COOH) group can co-polymerize with other precursors to form a gel bead that also comprises a COOH functional group. In some cases, acrylic acid (a species comprising free COOH groups), acrylamide, and bis(acryloyl)cystamine can be co-polymerized together to generate a gel bead comprising free COOH groups. The COOH groups of the gel bead can be activated (e.g., via 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) and N-Hydroxysuccinimide (NHS) or 4-(4,6-Dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMTMM)) such that they are reactive (e.g., reactive to amine functional groups where EDC/NHS or DMTMM are used for activation). The activated COOH groups can then react with an appropriate species (e.g., a species comprising an amine functional group where the carboxylic acid groups are activated to be reactive with an amine functional group) comprising a moiety to be linked to the bead.

[0200] Beads comprising disulfide linkages in their polymeric network may be functionalized with additional species via reduction of some of the disulfide linkages to free thiols. The disulfide linkages may be reduced via, for example, the action of a reducing agent (e.g., DTT, TCEP, etc.) to generate free thiol groups, without dissolution of the bead. Free thiols of the beads can then react with free thiols of a species or a species comprising another disulfide bond (e.g., via thiol-disulfide exchange) such that the species can be linked to the beads (e.g., via a generated disulfide bond). In some cases, free thiols of the beads may react with any other suitable group. For example, free thiols of the beads may react with species comprising an acrydite moiety. The free thiol groups of the beads can react with the acrydite via Michael addition chemistry, such that the species comprising the acrydite is linked to the bead. In some cases, uncontrolled reactions can be prevented by inclusion of a thiol capping agent such as N-ethylmalieamide or iodoacetate.

[0201] Activation of disulfide linkages within a bead can be controlled such that only a small number of disulfide linkages are activated. Control may be exerted, for example, by controlling the concentration of a reducing agent used to generate free thiol groups and/or concentration of reagents used to form disulfide bonds in bead polymerization. In some cases, a low concentration (e.g., molecules of reducing agent:gel bead ratios of less than or equal about 10000, 100000, 1000000, 10000000, 100000000, 1000000000, 10000000000, or 100000000000) of reducing agent may be used for reduction. Controlling the number of disulfide linkages that are reduced to free thiols may be useful in ensuring bead structural integrity during functionalization. In some cases, optically-active agents, such as fluorescent dyes may be may be coupled to beads via free thiol groups of the beads and used to quantify the number of free thiols present in a bead and/or track a bead.

[0202] In some cases, addition of moieties to a gel bead after gel bead formation may be advantageous. For example, addition of an oligonucleotide (e.g., barcoded oligonucleotide) after gel bead formation may avoid loss of the species during chain transfer termination that can occur during polymerization. Moreover, smaller precursors (e.g., monomers or cross linkers that do not comprise side chain groups and linked moieties) may be used for polymerization and can be minimally hindered from growing chain ends due to viscous effects. In some cases, functionalization after gel bead synthesis can minimize exposure of species (e.g., oligonucleotides) to be loaded with potentially damaging agents (e.g., free radicals) and/or chemical environments. In some cases, the generated gel may possess an upper critical solution temperature (UCST) that can permit temperature driven swelling and collapse of a bead. Such functionality may aid in oligonucleotide (e.g., a primer) infiltration into the bead during subsequent functionalization of the bead with the oligonucleotide. Post-production functionalization may also be useful in controlling loading ratios of species in beads, such that, for example, the variability in loading ratio is minimized. Species loading may also be performed in a batch process such that a plurality of beads can be functionalized with the species in a single batch.

[0203] In some cases, an acrydite moiety linked to precursor, another species linked to a precursor, or a precursor itself comprises a labile bond, such as chemically, thermally, or photo-sensitive bonds e.g., disulfide bonds, UV sensitive bonds, or the like. Once acrydite moieties or other moieties comprising a labile bond are incorporated into a bead, the bead may also comprise the labile bond. The labile bond may be, for example, useful in reversibly linking (e.g., covalently linking) species (e.g., barcodes, primers, etc.) to a bead. In some cases, a thermally labile bond may include a nucleic acid hybridization based attachment, e.g., where an oligonucleotide is hybridized to a complementary sequence that is attached to the bead, such that thermal melting of the hybrid releases the oligonucleotide, e.g., a barcode containing sequence, from the bead or microcapsule.

[0204] The addition of multiple types of labile bonds to a gel bead may result in the generation of a bead capable of responding to varied stimuli. Each type of labile bond may be sensitive to an associated stimulus (e.g., chemical stimulus, light, temperature, etc.) such that release of species attached to a bead via each labile bond may be controlled by the application of the appropriate stimulus. Such functionality may be useful in controlled release of species from a gel bead. In some cases, another species comprising a labile bond may be linked to a gel bead after gel bead formation via, for example, an activated functional group of the gel bead as described above. As will be appreciated, barcodes that are releasably, cleavably or reversibly attached to the beads described herein include barcodes that are released or releasable through cleavage of a linkage between the barcode molecule and the bead, or that are released through degradation of the underlying bead itself, allowing the barcodes to be accessed or accessible by other reagents, or both.

[0205] Species (e.g., oligonucleotides comprising barcodes) attached to a solid support (e.g., a bead) may comprise a U-excising element that allows the species to release from the bead. In some cases, the U-excising element may comprise a single-stranded DNA (ssDNA) sequence that contains at least one uracil. The species may be attached to a solid support via the ssDNA sequence. The species may be released by a combination of uracil-DNA glycosylase (e.g., to remove the uracil) and an endonuclease (e.g., to induce an ssDNA break). If the endonuclease generates a 5' phosphate group from the cleavage, then additional enzyme treatment may be included in downstream processing to eliminate the phosphate group, e.g., prior to ligation of additional sequencing handle elements, e.g., Illumina full P5 sequence, partial P5 sequence, full R1 sequence, and/or partial R1 sequence.

[0206] The barcodes that are releasable as described herein may sometimes be referred to as being activatable, in that they are available for reaction once released. Thus, for example, an activatable barcode may be activated by releasing the barcode from a bead (or other suitable type of partition described herein). Other activatable configurations are also envisioned in the context of the described methods and systems.

[0207] In addition to thermally cleavable bonds, disulfide bonds and UV sensitive bonds, other non-limiting examples of labile bonds that may be coupled to a precursor or bead include an ester linkage (e.g., cleavable with an acid, a base, or hydroxylamine), a vicinal diol linkage (e.g., cleavable via sodium periodate), a Diels-Alder linkage (e.g., cleavable via heat), a sulfone linkage (e.g., cleavable via a base), a silyl ether linkage (e.g., cleavable via an acid), a glycosidic linkage (e.g., cleavable via an amylase), a peptide linkage (e.g., cleavable via a protease), or a phosphodiester linkage (e.g., cleavable via a nuclease (e.g., DNAase)).

[0208] Species that do not participate in polymerization may also be encapsulated in beads during bead generation (e.g., during polymerization of precursors). Such species may be entered into polymerization reaction mixtures such that generated beads comprise the species upon bead formation. In some cases, such species may be added to the gel beads after formation. Such species may include, for example, oligonucleotides (e.g. barcoded oligonucleotides and/or anchor oligonucleotides), reagents for a nucleic acid amplification reaction (e.g., primers, polymerases, dNTPs, co-factors (e.g., ionic co-factors)) including those described herein, reagents for enzymatic reactions (e.g., enzymes, co-factors, substrates), or reagents for a nucleic acid modification reactions such as polymerization, ligation, or digestion. Trapping of such species may be controlled by the polymer network density generated during polymerization of precursors, control of ionic charge within the gel bead (e.g., via ionic species linked to polymerized species), or by the release of other species. Encapsulated species may be released from a bead upon bead degradation and/or by application of a stimulus capable of releasing the species from the bead.

[0209] Beads may be of uniform size or heterogeneous size. In some cases, the diameter of a bead may be about 1 micrometer (μm), 5 μm, 10 μm, 20 μm, 30 μm, 40 μm, 50 μm, 60 μm, 70 μm, 80 μm, 90 μm, 100 μm, 250 μm, 500 μm, or 1 mm. In some cases, a bead may have a diameter of at least about 1 μm, 5 μm, 10 μm, 20 μm, 30 μm, 40 μm, 50 μm, 60 μm, 70 μm, 80 μm, 90 μm, 100 μm, 250 μm, 500 μm, 1 mm, or more. In some cases, a bead may have a diameter of less than or equal to about 1 μm, 5 μm, 10 μm, 20 μm, 30 μm, 40 μm, 50 μm, 60 μm, 70 μm, 80 μm, 90 μm, 100 μm, 250 μm, 500 μm, or 1 mm. In some cases, a bead may have a diameter in the range of about 40-75 μm, 30-75 μm, 20-75 μm, 40-85 μm, 40-95 μm, 20-100 μm, 10-100 μm, 1-100 μm, 20-250 μm, or 20-500 μm.

[0210] In certain aspects, beads are provided as a population or plurality of beads having a relatively monodisperse size distribution. Such monodispersity can provide relatively consistent amounts of reagents within partitions and maintain relatively consistent bead characteristics. In particular, the beads described herein may have size distributions that have a coefficient of variation in their cross-sectional dimensions of less than or equal to about 50%, less than or equal to about 40%, less than or equal to about 30%, less than or equal to about 20%, less than or equal to about 15%, less than or equal to about 10%, or less than or equal to about 5%.

[0211] Beads may be of any suitable shape. Examples of bead shapes include, but are not limited to, spherical, non-spherical, oval, oblong, amorphous, circular, cylindrical, and variations thereof.

[0212] In addition to, or as an alternative to the cleavable linkages between the beads and the associated molecules, e.g., barcode containing oligonucleotides, described above, the beads may be degradable, disruptable, or dissolvable spontaneously or upon exposure to one or more stimuli (e.g., temperature changes, pH changes, exposure to particular chemical species or phase, exposure to light, reducing agent, etc.). In some cases, a bead may be dissolvable, such that material components of the beads are solubilized when exposed to a particular chemical species or an environmental change, such as a change temperature or a change in pH. In some cases, a gel bead is degraded or dissolved at elevated temperature and/or in basic conditions. In some cases, a bead may be thermally degradable such that when the bead is exposed to an appropriate change in temperature (e.g., heat), the bead degrades. Degradation or dissolution of a bead bound to a species (e.g., an oligonucleotide, e.g., barcoded oligonucleotide) may result in release of the species from the bead.

[0213] A degradable bead may comprise one or more species with a labile bond such that, when the bead/species is exposed to the appropriate stimuli, the bond is broken and the bead degrades. The labile bond may be a chemical bond (e.g., covalent bond, ionic bond) or may be another type of physical interaction (e.g., van der Waals interactions, dipole-dipole interactions, etc.). In some cases, a crosslinker used to generate a bead may comprise a labile bond. Upon exposure to the appropriate conditions, the labile bond can be broken and the bead degraded. For example, upon exposure of a polyacrylamide gel bead comprising cystamine crosslinkers to a reducing agent, the disulfide bonds of the cystamine can be broken and the bead degraded.

[0214] A degradable bead may be useful in more quickly releasing an attached species (e.g., an oligonucleotide, a barcode sequence, a primer, etc) from the bead when the appropriate stimulus is applied to the bead as compared to a bead that does not degrade. For example, for a species bound to an inner surface of a porous bead or in the case of an encapsulated species, the species may have greater mobility and accessibility to other species in solution upon degradation of the bead. In some cases, a species may also be

attached to a degradable bead via a degradable linker (e.g., disulfide linker). The degradable linker may respond to the same stimuli as the degradable bead or the two degradable species may respond to different stimuli. For example, a barcode sequence may be attached, via a disulfide bond, to a polyacrylamide bead comprising cystamine. Upon exposure of the barcoded-bead to a reducing agent, the bead degrades and the barcode sequence is released upon breakage of both the disulfide linkage between the barcode sequence and the bead and the disulfide linkages of the cystamine in the bead.

[0215] A degradable bead may be introduced into a partition, such as a droplet of an emulsion or a well, such that the bead degrades within the partition and any associated species (e.g., oligonucleotides) are released within the droplet when the appropriate stimulus is applied. The free species (e.g., oligonucleotides) may interact with other reagents contained in the partition. For example, a polyacrylamide bead comprising cystamine and linked, via a disulfide bond, to a barcode sequence, may be combined with a reducing agent within a droplet of a water-in-oil emulsion. Within the droplet, the reducing agent may break the various disulfide bonds resulting in bead degradation and release of the barcode sequence into the aqueous, inner environment of the droplet. In another example, heating of a droplet comprising a bead-bound barcode sequence in basic solution may also result in bead degradation and release of the attached barcode sequence into the aqueous, inner environment of the droplet.

[0216] As will be appreciated from the above disclosure, while referred to as degradation of a bead, degradation may refer to the disassociation of a bound or entrained species from a bead, both with and without structurally degrading the physical bead itself. For example, entrained species may be released from beads through osmotic pressure differences due to, for example, changing chemical environments. By way of example, alteration of bead pore sizes due to osmotic pressure differences can generally occur without structural degradation of the bead itself. In some cases, an increase in pore size due to osmotic swelling of a bead can permit the release of entrained species within the bead. In other cases, osmotic shrinking of a bead may cause a bead to better retain an entrained species due to pore size contraction.

[0217] Where degradable beads are provided, it can be useful to avoid exposing such beads to the stimulus or stimuli that cause such degradation prior to a given time, in order to avoid premature bead degradation and issues that arise from such degradation, including, for example poor flow characteristics and aggregation. By way of example, where beads comprise reducible cross-linking groups, such as disulfide groups, it can be useful to avoid contacting such beads with reducing agents, e.g., DTT or other disulfide cleaving reagents. In such cases, treatment to the beads described herein will, in some cases be provided free of reducing agents, such as DTT. Because reducing agents are often provided in commercial enzyme preparations, reducing agent free (or DTT free) enzyme preparations may be provided in treating the beads described herein. Examples of such enzymes include, e.g., polymerase enzyme preparations, reverse transcriptase enzyme preparations, ligase enzyme preparations, as well as many other enzyme preparations that may be used to treat the beads described herein. The terms "reducing agent free" or "DTT free" preparations can refer to a preparation having less than or equal to about 1/10th, less than or equal to about 1/50th, or less than or equal to about 1/100th of the lower ranges for such materials used in degrading the beads. For example, for DTT, the reducing agent free preparation will typically have less than or equal to about 0.01 mM, 0.005 mM, 0.001 mM DTT, 0.0005 mM DTT, or 0.0001 mM DTT. In some cases, the amount of DTT will be undetectable.

[0218] In some cases, a stimulus may be used to trigger degradation of the bead, which may result in the release of contents from the bead. Generally, a stimulus may cause degradation of the bead structure, such as degradation of the covalent bonds or other types of physical interaction. These stimuli may be useful in inducing a bead to degrade and/or to release its contents. Examples of stimuli that may be used include chemical stimuli, thermal stimuli, optical stimuli (e.g., light) and any combination thereof, as described more fully below.

[0219] Numerous chemical triggers may be used to trigger the degradation of beads. Examples of these chemical changes may include, but are not limited to pH-mediated changes to the integrity of a component within the bead, degradation of a component of a bead via cleavage of cross-linked bonds, and depolymerization of a component of a bead.

[0220] In some embodiments, a bead may be formed from materials that comprise degradable chemical crosslinkers, such as BAC or cystamine. Degradation of such degradable crosslinkers may be accomplished through a number of mechanisms. In some examples, a bead may be contacted with a chemical degrading agent that may induce oxidation, reduction or other chemical changes. For example, a chemical degrading agent may be a reducing agent, such as dithiothreitol (DTT). Additional examples of reducing agents may include β-mercaptoethanol, (2S)-2-amino-1,4-dimercaptobutane (dithiobutylamine or DTBA), tris(2-carboxyethyl) phosphine (TCEP), or combinations thereof. A reducing agent may degrade the disulfide bonds formed between gel precursors forming the bead, and thus, degrade the bead. In other cases, a change in pH of a solution, such as an increase in pH, may trigger degradation of a bead. In other cases, exposure to an aqueous solution, such as water, may trigger hydrolytic degradation, and thus degradation of the bead.

[0221] Beads may also be induced to release their contents upon the application of a thermal stimulus. A change in temperature can cause a variety of changes to a bead. For example, heat can cause a solid bead to liquefy. A change in heat may cause melting of a bead such that a portion of the bead degrades. In other cases, heat may increase the internal pressure of the bead components such that the bead ruptures or explodes. Heat may also act upon heat-sensitive polymers used as materials to construct beads.

[0222] The methods, compositions, devices, and kits of this disclosure may be used with any suitable agent to degrade beads. In some embodiments, changes in temperature or pH may be used to degrade thermo-sensitive or pH-sensitive bonds within beads. In some embodiments, chemical degrading agents may be used to degrade chemical bonds within beads by oxidation, reduction or other chemical changes. For example, a chemical degrading agent may be a reducing agent, such as DTT, wherein DTT may degrade the disulfide bonds formed between a crosslinker and gel precursors, thus degrading the bead. In some embodiments, a reducing agent may be added to degrade the

bead, which may or may not cause the bead to release its contents. Examples of reducing agents may include dithiothreitol (DTT), β-mercaptoethanol, (2S)-2-amino-1,4-dimercaptobutane (dithiobutylamine or DTBA), tris(2-carboxyethyl) phosphine (TCEP), or combinations thereof. The reducing agent may be present at a concentration of about 0.1 mM, 0.5 mM, 1 mM, 5 mM, or 10 mM. The reducing agent may be present at a concentration of at least about 0.1 mM, 0.5 mM, 1 mM, 5 mM, 10 mM, or greater. The reducing agent may be present at concentration of at most about 0.1 mM, 0.5 mM, 1 mM, 5 mM, or 10 mM.

[0223] Any suitable number of nucleic acid molecules (e.g., primer, barcoded oligonucleotide, anchor oligonucleotide) can be associated with a bead such that, upon release from the bead, the nucleic acid molecules (e.g., primer, barcoded oligonucleotide, anchor oligonucleotide) are present in the partition at a pre-defined concentration. Such pre-defined concentration may be selected to facilitate certain reactions for generating a sequencing library, e.g., amplification, within the partition. In some cases, the predefined concentration of the primer is limited by the process of producing oligonucleotide bearing beads.

[0224] In some aspects, the partitions refer to containers or vessels (such as wells, microwells, tubes, vials, through ports in nanoarray substrates, e.g., BioTrove nanoarrays, or other containers). In some aspects, the compartments or partitions comprise partitions that are flowable within fluid streams. These partitions may comprise, e.g., micro-vesicles that have an outer barrier surrounding an inner fluid center or core, or, in some cases, they may comprise a porous matrix that is capable of entraining and/or retaining materials within its matrix. In some aspects, partitions comprise droplets of aqueous fluid within a non-aqueous continuous phase, e.g., an oil phase. Examples of different vessels are described in U.S. Patent Application Publication No. 2014/0155295, which is entirely incorporated herein by reference for all purposes. Examples of emulsion systems for creating stable droplets in non-aqueous or oil continuous phases are described in detail in U.S. Patent Application Publication No. 2010/0105112, which is entirely incorporated herein by reference for all purposes.

[0225] In the case of droplets in an emulsion, allocating individual cells to discrete partitions may generally be accomplished by introducing a flowing stream of cells in an aqueous fluid into a flowing stream of a non-aqueous fluid, such that droplets are generated at the junction of the two streams. By providing the aqueous cell-containing stream at a certain concentration of cells, the occupancy of the resulting partitions (e.g., number of cells per partition) can be controlled. Where single cell partitions are implemented, the relative flow rates of the fluids can be selected such that, on average, the partitions contain less than one cell per partition, in order to ensure that those partitions that are occupied, are primarily singly occupied. In some embodiments, the relative flow rates of the fluids can be selected such that a majority of partitions are occupied, e.g., allowing for only a small percentage of unoccupied partitions. In some aspects, the flows and channel architectures are controlled as to ensure a number of singly occupied partitions, less than a certain level of unoccupied partitions and less than a certain level of multiply occupied partitions.

[0226] The systems and methods described herein can be operated such that a majority of occupied partitions include no more than one cell per occupied partition. In some cases, the partitioning process is conducted such that fewer than 25% of the occupied partitions contain more than one cell, and in some cases, fewer than 20% of the occupied partitions have more than one cell. In some cases, fewer than 10% or fewer than 5% of the occupied partitions include more than one cell per partition.

[0227] In some cases, it can be useful to avoid the creation of excessive numbers of empty partitions. For example, from a cost perspective and/or efficiency perspective, it may helpful to minimize the number of empty partitions. While this may be accomplished by providing sufficient numbers of cells into the partitioning zone, the Poissonian distribution may expectedly increase the number of partitions that may include multiple cells. As such, in accordance with aspects described herein, the flow of one or more of the cells, or other fluids directed into the partitioning zone are conducted such that, in some cases, no more than 50% of the generated partitions, no more than 25% of the generated partitions, or no more than 10% of the generated partitions are unoccupied. Further, in some aspects, these flows are controlled so as to present non-Poissonian distribution of single occupied partitions while providing lower levels of unoccupied partitions. The above ranges of unoccupied partitions can be achieved while still providing any of the single occupancy rates described above. For example, the use of the systems and methods described herein creates resulting partitions that have multiple occupancy rates of less than or equal to about 25%, 20%, 15%, 10%, or 5%, while having unoccupied partitions of less than or equal to about 50%, 40%, 30%, 20%, 10%, or 5%.

[0228] As will be appreciated, the above-described occupancy rates are also applicable to partitions that include both cells and additional reagents and agents, including, but not limited to, microcapsules carrying barcoded oligonucleotides, microcapsules carrying anchoring oligonucleotides, labelling agents, labelling agents comprising reporter oligonucleotides, labelling agents comprising reporter oligonucleotides comprising a nucleic barcode sequence, and cells with one or more labelling agents bound to one or more cell surface features. In some aspects, a substantial percentage of the overall occupied partitions can include a microcapsule (e.g., bead) comprising barcodes or anchoring oligonucleotides and a cell with or without bound labelling agents.

[0229] Although described in terms of providing substantially singly occupied partitions, above, in certain cases, it can be useful to provide multiply occupied partitions, e.g., containing two, three, four or more cells and/or microcapsules (e.g., beads) comprising barcoded oligonucleotides or anchor oligonucleotides within a single partition. Accordingly, the flow characteristics of the cell and/or bead containing fluids and partitioning fluids may be controlled to provide for such multiply occupied partitions. In particular, the flow parameters may be controlled to provide an occupancy rate at greater than or equal to about 50% of the partitions, greater than or equal to about 75%, or greater than or equal to about 80%, 90%, 95%, or higher.

[0230] In some cases, additional microcapsules are used to deliver additional reagents to a partition. In such cases, it may be advantageous to introduce different beads into a common channel or droplet generation junction, from different bead sources, i.e., containing different associated reagents, through different channel inlets into such common channel or droplet generation junction. In such cases, the

flow and frequency of the different beads into the channel or junction may be controlled to provide for a suitable ratio of microcapsules from each source, while ensuring the pairing or combination of such beads into a partition with the number of cells.

[0231] The partitions described herein may comprise small volumes, e.g., less than or equal to 10 □L, 5□L, 1 □L, 900 picoliters (μL), 800 pL, 700 pL, 600 pL, 500 pL, 400 pL, 300 pL, 200 pL, 100 pL, 50 pL, 20 pL, 10 pL, 1 pL, 500 nanoliters (nL), 100 nL, 50 nL, or less.

[0232] For example, in the case of droplet based partitions, the droplets may have overall volumes that are less than or equal to 1000 pL, 900 pL, 800 pL, 700 pL, 600 pL, 500 pL, 400 pL, 300 pL, 200 pL, 100 pL, 50 pL, 20 pL, 10 pL, or 1 pL. Where co-partitioned with microcapsules, it will be appreciated that the sample fluid volume, e.g., including co-partitioned cells, within the partitions may be less than or equal to 90%, 80%, 70%, 60%, 50%, 40%, 30%, 20%, 10%, or less than the above described volumes.

[0233] As is described elsewhere herein, partitioning species may generate a population or plurality of partitions. In such cases, any suitable number of partitions can be generated to generate the plurality of partitions. For example, in a method described herein, a plurality of partitions may be generated that comprises at least about 1,000 partitions, at least about 5,000 partitions, at least about 10,000 partitions, at least about 50,000 partitions, at least about 100,000 partitions, at least about 500,000 partitions, at least about 1,000,000 partitions, at least about 5,000,000 partitions at least about 10,000,000 partitions, at least about 50,000,000 partitions, at least about 100,000,000 partitions, at least about 500,000,000 partitions or at least about 1,000,000,000 partitions. Moreover, the plurality of partitions may comprise both unoccupied partitions (e.g., empty partitions) and occupied partitions

[0234] Microfluidic channel networks can be utilized to generate partitions as described herein. Alternative mechanisms may also be employed in the partitioning of individual cells, including porous membranes through which aqueous mixtures of cells are extruded into non-aqueous fluids.

[0235] An example of a simplified microfluidic channel structure for partitioning individual cells is illustrated in FIG. **1**. Cells may be partitioned with or without labelling agents bound to cell surface features, as described herein. As described herein, in some cases, the majority of occupied partitions include no more than one cell per occupied partition and, in some cases, some of the generated partitions are unoccupied. In some cases, though, some of the occupied partitions may include more than one cell. In some cases, the partitioning process may be controlled such that fewer than 25% of the occupied partitions contain more than one cell, and in some cases, fewer than 20% of the occupied partitions have more than one cell, while in some cases, fewer than 10% or fewer than 5% of the occupied partitions include more than one cell per partition. As shown, the channel structure can include channel segments **102**, **104**, **106** and **108** communicating at a channel junction **110**. In operation, a first aqueous fluid **112** that includes suspended cells **114**, may be transported along channel segment **102** into junction **110**, while a second fluid **116** that is immiscible with the aqueous fluid **112** is delivered to the junction **110** from channel segments **104** and **106** to create discrete droplets **118** of the aqueous fluid including individual cells **114**, flowing into channel segment **108**.

[0236] In some aspects, this second fluid **116** comprises an oil, such as a fluorinated oil, that includes a fluorosurfactant for stabilizing the resulting droplets, e.g., inhibiting subsequent coalescence of the resulting droplets. Examples of partitioning fluids and fluorosurfactants are described in U.S. Patent Application Publication No. 2010/0105112, which is entirely incorporated herein by reference for all purposes.

[0237] In other aspects, in addition to or as an alternative to droplet based partitioning, cells (with or without labelling agents bound to cell surface features, as described herein) may be encapsulated within a microcapsule that comprises an outer shell or layer or porous matrix in which is entrained one or more individual cells or small groups of cells, and may include other reagents. Encapsulation of cells may be carried out by a variety of processes. Such processes combine an aqueous fluid containing the cells to be analyzed with a polymeric precursor material that may be capable of being formed into a gel or other solid or semi-solid matrix upon application of a particular stimulus to the polymer precursor. Such stimuli include, e.g., thermal stimuli (either heating or cooling), photo-stimuli (e.g., through photocuring), chemical stimuli (e.g., through crosslinking, polymerization initiation of the precursor (e.g., through added initiators), or the like.

[0238] Preparation of microcapsules comprising cells may be carried out by a variety of methods. For example, air knife droplet or aerosol generators may be used to dispense droplets of precursor fluids into gelling solutions in order to form microcapsules that include individual cells or small groups of cells. Likewise, membrane based encapsulation systems may be used to generate microcapsules comprising encapsulated cells as described herein. In some aspects, microfluidic systems like that shown in FIG. **1** may be readily used in encapsulating cells as described herein. In particular, and with reference to FIG. **1**, the aqueous fluid comprising the cells and the polymer precursor material is flowed into channel junction **110**, where it is partitioned into droplets **118** comprising the individual cells **114**, through the flow of non-aqueous fluid **116**. In the case of encapsulation methods, non-aqueous fluid **116** may also include an initiator to cause polymerization and/or crosslinking of the polymer precursor to form the microcapsule that includes the entrained cells. Examples of polymer precursor/initiator pairs are described in U.S. Patent Application Publication No. 2014/0378345, which is entirely incorporated herein by reference for all purposes.

[0239] For example, in the case where the polymer precursor material comprises a linear polymer material, e.g., a linear polyacrylamide, PEG, or other linear polymeric material, the activation agent may comprise a cross-linking agent, or a chemical that activates a cross-linking agent within the formed droplets. Likewise, for polymer precursors that comprise polymerizable monomers, the activation agent may comprise a polymerization initiator. For example, in certain cases, where the polymer precursor comprises a mixture of acrylamide monomer with a N,N'-bis-(acryloyl) cystamine (BAC) co-monomer, an agent such as tetraethylmethylenediamine (TEMED) may be provided within the second fluid streams in channel segments **104** and **106**, which initiates the copolymerization of the acrylamide and BAC into a cross-linked polymer network or, hydrogel.

[0240] Upon contact of the second fluid stream **116** with the first fluid stream **112** at junction **110** in the formation of

droplets, the TEMED may diffuse from the second fluid **116** into the aqueous first fluid **112** comprising the linear polyacrylamide, which will activate the crosslinking of the polyacrylamide within the droplets, resulting in the formation of the gel, e.g., hydrogel, microcapsules **118**, as solid or semi-solid beads or particles entraining the cells **114**. Although described in terms of polyacrylamide encapsulation, other 'activatable' encapsulation compositions may also be employed in the context of the methods and compositions described herein. For example, formation of alginate droplets followed by exposure to divalent metal ions, e.g., Ca2+, can be used as an encapsulation process using the described processes. Likewise, agarose droplets may also be transformed into capsules through temperature based gelling, e.g., upon cooling, or the like. In some cases, encapsulated cells can be selectively releasable from the microcapsule, e.g., through passage of time, or upon application of a particular stimulus, that degrades the microcapsule sufficiently to allow the cell, or its contents to be released from the microcapsule, e.g., into a partition, such as a droplet. For example, in the case of the polyacrylamide polymer described above, degradation of the microcapsule may be accomplished through the introduction of an appropriate reducing agent, such as DTT or the like, to cleave disulfide bonds that cross link the polymer matrix. See, e.g., U.S. Patent Application Publication No. 2014/0378345, which is entirely incorporated herein by reference for all purposes.

[0241] Encapsulated cells or cell populations provide certain potential advantages of being storable, and more portable than droplet based partitioned cells. Furthermore, in some cases, it may cells to be analyzed can be incubated for a select period of time, in order to characterize changes in such cells over time, either in the presence or absence of different stimuli. In such cases, encapsulation of individual cells may allow for longer incubation than partitioning in emulsion droplets, although in some cases, droplet partitioned cells may also be incubated for different periods of time, e.g., at least 10 seconds, at least 30 seconds, at least 1 minute, at least 5 minutes, at least 10 minutes, at least 30 minutes, at least 1 hour, at least 2 hours, at least 5 hours, or at least 10 hours or more. The encapsulation of cells may constitute the partitioning of the cells into which other reagents are co-partitioned. Alternatively, encapsulated cells may be readily deposited into other partitions, e.g., droplets, as described above.

[0242] In accordance with certain aspects, the cells may be partitioned along with lysis reagents in order to release the contents of the cells within the partition. In such cases, the lysis agents can be contacted with the cell suspension concurrently with, or immediately prior to the introduction of the cells into the partitioning junction/droplet generation zone, e.g., through an additional channel or channels upstream of channel junction **110**. Examples of lysis agents include bioactive reagents, such as lysis enzymes that are used for lysis of different cell types, e.g., gram positive or negative bacteria, plants, yeast, mammalian, etc., such as lysozymes, achromopeptidase, lysostaphin, labiase, kitalase, lyticase, and a variety of other lysis enzymes available from, e.g., Sigma-Aldrich, Inc. (St Louis, MO), as well as other commercially available lysis enzymes. Other lysis agents may additionally or alternatively be co-partitioned with the cells to cause the release of the cell's contents into the partitions. For example, in some cases, surfactant based lysis solutions may be used to lyse cells. In some cases, lysis solutions may include non-ionic surfactants such as, for example, TritonX-100 and Tween 20. In some cases, lysis solutions may include ionic surfactants such as, for example, sarcosyl and sodium dodecyl sulfate (SDS). Electroporation, thermal, acoustic or mechanical cellular disruption may also be used in certain cases, e.g., non-emulsion based partitioning such as encapsulation of cells that may be in addition to or in place of droplet partitioning, where any pore size of the encapsulate is sufficiently small to retain nucleic acid fragments of a suitable size, following cellular disruption.

[0243] In addition to the lysis agents co-partitioned with the cells described above, other reagents can also be co-partitioned with the cells, including, for example, DNase and RNase inactivating agents or inhibitors, such as proteinase K, chelating agents, such as EDTA, and other reagents employed in removing or otherwise reducing negative activity or impact of different cell lysate components on subsequent processing of nucleic acids. In addition, in the case of encapsulated cells, the cells may be exposed to an appropriate stimulus to release the cells or their contents from a co-partitioned microcapsule. For example, in some cases, a chemical stimulus may be co-partitioned along with an encapsulated cell to allow for the degradation of the microcapsule and release of the cell or its contents into the larger partition. In some cases, this stimulus may be the same as the stimulus described elsewhere herein for release of oligonucleotides from their respective microcapsule (e.g., bead). In alternative aspects, this may be a different and non-overlapping stimulus, in order to allow an encapsulated cell to be released into a partition at a different time from the release of oligonucleotides into the same partition.

[0244] Additional reagents may also be co-partitioned with the cells, such as endonucleases to fragment the cell's DNA, DNA polymerase enzymes and dNTPs used to amplify the cell's nucleic acid fragments and to attach the barcode oligonucleotides to the amplified fragments. Additional reagents may also include reverse transcriptase enzymes, including enzymes with terminal transferase activity, primers and oligonucleotides, and switch oligonucleotides (also referred to herein as "switch oligos" or "template switching oligonucleotides") which can be used for template switching. In some cases, template switching can be used to increase the length of a cDNA. In some cases, template switching can be used to append a predefined nucleic acid sequence to the cDNA. In one example of template switching, cDNA can be generated from reverse transcription of a template, e.g., cellular mRNA, where a reverse transcriptase with terminal transferase activity can add additional nucleotides, e.g., polyC, to the cDNA in a template independent manner. Switch oligos can include sequences complementary to the additional nucleotides, e.g., polyG. The additional nucleotides (e.g., polyC) on the cDNA can hybridize to the additional nucleotides (e.g., polyG) on the switch oligo, whereby the switch oligo can be used by the reverse transcriptase as template to further extend the cDNA. Template switching oligonucleotides may comprise a hybridization region and a template region. The hybridization region can comprise any sequence capable of hybridizing to the target. In some cases, as previously described, the hybridization region comprises a series of G bases to complement the overhanging C bases at the 3' end of a cDNA molecule. The series of G bases may comprise 1 G base, 2 G bases, 3 G bases, 4 G bases, 5 G bases or more than 5 G bases. The template sequence can comprise any

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

[0245] In some cases, the length of a switch oligo may be 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250 nucleotides or longer.

[0246] In some cases, the length of a switch oligo may be at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249 or 250 nucleotides or longer.

[0247] In some cases, the length of a switch oligo may be at most 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249 or 250 nucleotides.

[0248] Additional agents may also be co-partitioned with the cells, such as one or more labelling agents capable of binding to one or more cell surface features of the cell(s). Cell surface features may comprise a receptor, an antigen, a surface protein, a transmembrane protein, a cluster of differentiation protein, a protein channel, a protein pump, a carrier protein, a phospholipid, a glycoprotein, a glycolipid, a cell-cell interaction protein complex, an antigen-presenting complex, a major histocompatibility complex, an engineered T-cell receptor, a T-cell receptor, a B-cell receptor, a chimeric antigen receptor, a gap junction, and an adherens junction. The labelling agents may comprise an antibody, and antibody fragment, a cell surface receptor binding molecule, a receptor ligand, a small molecule, a bi-specific antibody, a bi-specific T-cell engager, a T-cell receptor engager, a B-cell receptor engager, a pro-body, an aptamer, a monobody, an affimer, a darpin, and a protein scaffold. The labelling agents may be coupled, through the coupling approaches as described herein, to a reporter oligonucleotide comprising a nucleic acid barcode sequence that permits identification of the labelling agent, as described herein. In some embodiments, the nucleic acid barcode sequence coupled to the labelling agent may comprise a unique molecular identifier (UMI) sequence segment, as described herein.

[0249] A labelling agent may comprise an antigen presenting particle. In some cases, an antigen presenting particle may comprise an antigen on or adjacent to its surface. The antigen presenting particle may bind to one or more molecules on the surface of a cell in a sample, e.g., through the antigen on the antigen presenting particle. In some cases, an antigen presenting particle may be used as a labelling agent for an immune cell, e.g., a T cell or a B cell. Such antigen presenting particle may bind to a T cell receptor and/or B cell receptor. In some cases, the antigen presenting particle comprise an antigen that is recognized (e.g., bound) by an immune cell. The antigen presenting particle may be a cell, e.g., a cancer cell or other antigen presenting cell. The antigen presenting particle may be a pathogen, e.g., a bacterium, a fungus, a microbe or a virus. In certain cases, the antigen presenting particle (e.g., a cell or a virus) may comprise an antigen expression vector that expresses the antigen on the surface of the particle. The antigen expression vector may comprise a barcode for identifying the nucleic acid or amino acid sequence of the antigen.

[0250] An example method for using an antigen presenting particle to analyze a cell may comprise one or more of the following operations. A sample comprising immune cells (e.g., blood or a fraction thereof) are mixed with a population of antigen presenting particles, and incubated to allow for the immune cells and antigen presenting particles to interact. The immune cells and antigen presenting particles bound to the immune cells are purified using an antibody that selectively binds to the immune cells. The bound immune cells and antigen presenting particles are partitioned

into droplets with beads (e.g., gel beads). Each of the beads comprises anchor oligonucleotide comprising a primer for mRNA molecules, a barcode and a UMI. At least one of the droplets contains an immune cell, an antigen presenting particle, and a gel bead. The immune cell and the antigen presenting particle in the droplet are lysed. The mRNA molecules from the immune cell and the antigen presenting particle are released. Reverse transcription is performed with the mRNA molecules and the anchor oligonucleotide from the bead. Thus, the resulting cDNA are tagged with the barcode and UMI from the anchor oligonucleotide. The resulting cDNA are then sequenced, e.g., to a high depth per cell on a sequencer (e.g., an Illumina sequencer). With the sequence reads, V(D)J regions of the immune cell are assembled and characteristics of the antigen presenting particle are also determined. When the antigen presenting particles are cancer cells, mutations and/or single-nucleotide polymorphisms (SNPs) may be determined with the sequence reads to identify a sub-populations of tumor cells that are targeted by an immune cell with the corresponding V(D)J sequences. When the antigen presenting particles are viruses, viral genome may be assembled to identify the sub-clone of viruses that are targeted by the immune cells with the corresponding V(D)J sequences. The method may yield pairs of V(D)J sequences and antigen-identifying sequences (e.g., mRNA of tumor cells or the genome of viruses) that are useful in developing personalized immunotherapies or vaccines against specific viral strains.

[0251] A protein labeled by a labelling agent (e.g., an antibody labeled by a barcode) may be used as a probe in a binding assay. The protein may be an antibody or a cell surface protein, e.g., a cell receptor such as a T-cell receptor and B-cell receptor. The labelling agent may comprise a barcode and/or a UMI. In some cases, another labelling agent comprising the same barcode and/or UMI may be used to analyze nucleic acids from the same cell as the protein. The nucleic acids and the protein from the same cell may be identified by the barcode and/or UMI. In some cases, the nucleic acid sequence of the cell surface protein may be determined using the labelling agent for analyzing nucleic acids, so that the amino acid sequence of the cell surface protein may also be determined. The labeled protein from the cell may then be used as a probe in a binding assay against a target molecule (e.g., a protein). For example, in the binding assay, whether the labeled cell surface protein can bind to the target protein may be determined. The label of the cell surface protein may be separated from the cell surface protein, e.g., by denaturation. Then the barcode and/or UMI on the label may be sequenced. The sequences of the barcode and/or UMI may be used to correlate the binding assay result with the sequence of the cell surface protein. Thus, the interaction of the protein with the target molecule may be correlated with the sequence of the protein. In some cases, the interaction between the protein and the target molecule may be quantified using the UMI.

[0252] Once the contents of the cells are released into their respective partitions, the nucleic acids contained therein may be further processed within the partitions. In accordance with the methods and systems described herein, the nucleic acid contents of individual cells can be provided with unique identifiers such that, upon characterization of those nucleic acids they may be attributed as having been derived from the same cell or cells. The ability to attribute characteristics to individual cells or groups of cells is provided by the assignment of unique identifiers specifically to an individual cell or groups of cells. Unique identifiers, e.g., in the form of nucleic acid barcodes can be assigned or associated with individual cells or populations of cells, in order to tag or label the cell's components (and as a result, its characteristics) with the unique identifiers. These unique identifiers can then be used to attribute the cell's components and characteristics to an individual cell or group of cells. In some aspects, this is carried out by co-partitioning the individual cells or groups of cells with the unique identifiers. In some aspects, the unique identifiers are provided in the form of oligonucleotides (also referred to herein as anchor oligonucleotides) that comprise nucleic acid barcode sequences that may be attached to or otherwise associated with the nucleic acid contents of individual cells, or to other components of the cells, and particularly to fragments of those nucleic acids. The oligonucleotides may be partitioned such that as between oligonucleotides in a given partition, the nucleic acid barcode sequences contained therein are the same, but as between different partitions, the oligonucleotides can, and do have differing barcode sequences, or at least represent a large number of different barcode sequences across all of the partitions in a given analysis. In some aspects, only one nucleic acid barcode sequence can be associated with a given partition, although in some cases, two or more different barcode sequences may be present.

[0253] The nucleic acid barcode sequences can include from 6 to about 20 or more nucleotides within the sequence of the oligonucleotides. In some cases, the length of a barcode sequence may be 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 nucleotides or longer. In some cases, the length of a barcode sequence may be at least 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 nucleotides or longer. In some cases, the length of a barcode sequence may be at most 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 nucleotides or shorter. These nucleotides may be completely contiguous, i.e., in a single stretch of adjacent nucleotides, or they may be separated into two or more separate subsequences that are separated by 1 or more nucleotides. In some cases, separated barcode subsequences can be from about 4 to about 16 nucleotides in length. In some cases, the barcode subsequence may be 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 nucleotides or longer. In some cases, the barcode subsequence may be at least 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 nucleotides or longer. In some cases, the barcode subsequence may be at most 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 nucleotides or shorter.

[0254] The co-partitioned oligonucleotides can also comprise other functional sequences useful in the processing of the nucleic acids from the co-partitioned cells. These sequences include, e.g., targeted or random/universal amplification primer sequences for amplifying the genomic DNA from the individual cells within the partitions while attaching the associated barcode sequences, sequencing primers or primer recognition sites, hybridization or probing sequences, e.g., for identification of presence of the sequences or for pulling down barcoded nucleic acids, or any of a number of other potential functional sequences. Other mechanisms of co-partitioning oligonucleotides may also be employed, including, e.g., coalescence of two or more droplets, where one droplet contains oligonucleotides, or microdispensing of oligonucleotides into partitions, e.g., droplets within microfluidic systems. Co-partitioning of

oligonucleotides and associated barcodes and other functional sequences or labels, along with sample materials as describe herein, may be performed, for example, as described in U.S. Patent Application Publication No. 2014/0227684, which is entirely incorporated herein by reference for all purposes.

[0255] Briefly, in one example, microcapsules, such as beads, are provided that each include large numbers of the above described barcoded oligonucleotides (also referred to herein as anchor oligonucleotides) releasably attached to the beads, where all of the oligonucleotides attached to a particular bead will include the same nucleic acid barcode sequence, but where a large number of diverse barcode sequences are represented across the population of beads used. In some embodiments, hydrogel beads, e.g., comprising polyacrylamide polymer matrices, are used as a solid support and delivery vehicle for the oligonucleotides into the partitions, as they are capable of carrying large numbers of oligonucleotide molecules, and may be configured to release those oligonucleotides upon exposure to a particular stimulus, as described elsewhere herein. In some cases, the population of beads will provide a diverse barcode sequence library that includes at least 1,000 different barcode sequences, at least 5,000 different barcode sequences, at least 10,000 different barcode sequences, at least at least 50,000 different barcode sequences, at least 100,000 different barcode sequences, at least 1,000,000 different barcode sequences, at least 5,000,000 different barcode sequences, or at least 10,000,000 different barcode sequences. Additionally, each bead can be provided with large numbers of oligonucleotide molecules attached. In particular, the number of molecules of oligonucleotides including the barcode sequence on an individual bead can be at least 1,000 oligonucleotide molecules, at least 5,000 oligonucleotide molecules, at least 10,000 oligonucleotide molecules, at least 50,000 oligonucleotide molecules, at least 100,000 oligonucleotide molecules, at least 500,000 oligonucleotides, at least 1,000,000 oligonucleotide molecules, at least 5,000,000 oligonucleotide molecules, at least 10,000,000 oligonucleotide molecules, at least 50,000,000 oligonucleotide molecules, at least 100,000,000 oligonucleotide molecules, and in some cases at least 1 billion oligonucleotide molecules.

[0256] Moreover, when the population of beads is partitioned, the resulting population of partitions can also include a diverse barcode library that includes at least 1,000 different barcode sequences, at least 5,000 different barcode sequences, at least 10,000 different barcode sequences, at least at least 50,000 different barcode sequences, at least 100,000 different barcode sequences, at least 1,000,000 different barcode sequences, at least 5,000,000 different barcode sequences, or at least 10,000,000 different barcode sequences. Additionally, each partition of the population can include at least 1,000 oligonucleotide molecules, at least 5,000 oligonucleotide molecules, at least 10,000 oligonucleotide molecules, at least 50,000 oligonucleotide molecules, at least 100,000 oligonucleotide molecules, at least 500,000 oligonucleotides, at least 1,000,000 oligonucleotide molecules, at least 5,000,000 oligonucleotide molecules, at least 10,000,000 oligonucleotide molecules, at least 50,000,000 oligonucleotide molecules, at least 100,000,000 oligonucleotide molecules, and in some cases at least 1 billion oligonucleotide molecules.

[0257] In some cases, multiple different barcodes can be incorporated within a given partition, either attached to a single or multiple beads within the partition. For example, in some cases, a mixed, but known barcode sequences set may provide greater assurance of identification in the subsequent processing, e.g., by providing a stronger address or attribution of the barcodes to a given partition, as a duplicate or independent confirmation of the output from a given partition.

[0258] The oligonucleotides may be releasable from the beads upon the application of a particular stimulus to the beads. In some cases, the stimulus may be a photo-stimulus, e.g., through cleavage of a photo-labile linkage that releases the oligonucleotides. In other cases, a thermal stimulus may be used, where elevation of the temperature of the beads environment will result in cleavage of a linkage or other release of the oligonucleotides form the beads. In still other cases, a chemical stimulus is used that cleaves a linkage of the oligonucleotides to the beads, or otherwise results in release of the oligonucleotides from the beads. In one case, such compositions include the polyacrylamide matrices described above for encapsulation of cells, and may be degraded for release of the attached oligonucleotides through exposure to a reducing agent, such as DTT. Examples of other systems and methods are described in U.S. Patent Application Publication No. 2014/0155295 and US. Patent Application Publication No. 2014/0378345, each of which is entirely incorporated herein by reference for all purposes.

[0259] In accordance with the methods and systems described herein, the beads including the attached oligonucleotides may be co-partitioned with the individual cells, such that a single bead and a single cell are contained within an individual partition. While single cell/single bead occupancy is one possible state, it will be appreciated that multiply occupied partitions (either in terms of cells, beads or both), or unoccupied partitions (either in terms of cells, beads or both) may often be present. An example of a microfluidic channel structure for co-partitioning cells and beads comprising barcode oligonucleotides is schematically illustrated in FIG. **2**. As described elsewhere herein, in some aspects, a substantial percentage of the overall occupied partitions may include both a bead and a cell and, in some cases, some of the partitions that are generated may be unoccupied. In some cases, some of the partitions may have beads and cells that are not partitioned **1**:**1**. In some cases, multiply occupied partitions may be provided, e.g., containing two, three, four or more cells and/or beads within a single partition. As shown, channel segments **202**, **204**, **206**, **208** and **210** are provided in fluid communication at channel junction **212**. An aqueous stream comprising the individual cells **214**, is flowed through channel segment **202** toward channel junction **212**. As described above, these cells may be suspended within an aqueous fluid, or may have been pre-encapsulated, prior to the partitioning process.

[0260] Concurrently, an aqueous stream comprising the barcode carrying beads **216**, is flowed through channel segment **204** toward channel junction **212**. A non-aqueous partitioning fluid **216** is introduced into channel junction **212** from each of side channels **206** and **208**, and the combined streams are flowed into outlet channel **210**. Within channel junction **212**, the two combined aqueous streams from channel segments **202** and **204** are combined, and partitioned into droplets **218**, that include co-partitioned cells

214 and beads **216**. By controlling the flow characteristics of each of the fluids combining at channel junction **212**, as well as controlling the geometry of the channel junction, partitioning can be optimized to achieve a suitable occupancy level of beads, cells or both, within the partitions **218** that are generated.

[0261] In some cases, lysis agents, e.g., cell lysis enzymes, may be introduced into the partition with the bead stream, e.g., flowing through channel segment **204**, such that the cell may be lysed at or after the time of partitioning. In some cases, cell membranes are maintained intact, such as to allow for the characterization of cell surface markers, as described later herein. Additional reagents may also be added to the partition in this configuration, such as endonucleases to fragment the cell's DNA, DNA polymerase enzyme and dNTPs used to amplify the cell's nucleic acid fragments and to attach the barcode oligonucleotides to the amplified fragments. A chemical stimulus, such as DTT, may be used to release the barcodes from their respective beads into the partition. In such cases, the chemical stimulus can be provided along with the cell-containing stream in channel segment **202**, such that release of the barcodes only occurs after the two streams have been combined, e.g., within the partitions **218**. Where the cells are encapsulated, however, introduction of a common chemical stimulus, e.g., that both releases the oligonucleotides form their beads, and releases cells from their microcapsules may generally be provided from a separate additional side channel (not shown) upstream of or connected to channel junction **212**.

[0262] A number of other reagents may be co-partitioned along with the cells, beads, lysis agents and chemical stimuli, including, for example, protective reagents, like proteinase K, chelators, nucleic acid extension, replication, transcription or amplification reagents such as polymerases, reverse transcriptases, transposases which can be used for transposon based methods (e.g., Nextera), nucleoside triphosphates or NTP analogues, primer sequences and additional cofactors such as divalent metal ions used in such reactions, ligation reaction reagents, such as ligase enzymes and ligation sequences, dyes, labels, or other tagging reagents.

[0263] The channel networks, e.g., as described herein, can be fluidly coupled to appropriate fluidic components. For example, the inlet channel segments, e.g., channel segments **202**, **204**, **206** and **208** are fluidly coupled to appropriate sources of the materials they are to deliver to channel junction **212**. For example, channel segment **202** may be fluidly coupled to a source of an aqueous suspension of cells **214** to be analyzed, while channel segment **204** may be fluidly coupled to a source of an aqueous suspension of beads **216**. Channel segments **206** and **208** may then be fluidly connected to one or more sources of the non-aqueous fluid. These sources may include any of a variety of different fluidic components, from simple reservoirs defined in or connected to a body structure of a microfluidic device, to fluid conduits that deliver fluids from off-device sources, manifolds, or the like. Likewise, the outlet channel segment **210** may be fluidly coupled to a receiving vessel or conduit for the partitioned cells. Again, this may be a reservoir defined in the body of a microfluidic device, or it may be a fluidic conduit for delivering the partitioned cells to a subsequent process operation, instrument or component.

[0264] FIG. **8** shows images of individual Jurkat cells co-partitioned along with barcode oligonucleotide containing beads in aqueous droplets in an aqueous in oil emulsion. As illustrated, individual cells may be readily co-partitioned with individual beads. As will be appreciated, optimization of individual cell loading may be carried out by a number of methods, including by providing dilutions of cell populations into the microfluidic system in order to achieve suitable cell loading per partition as described elsewhere herein.

[0265] In operation, once lysed, the nucleic acid contents of the individual cells are then available for further processing within the partitions, including, e.g., fragmentation, amplification and barcoding, as well as attachment of other functional sequences. Fragmentation may be accomplished through the co-partitioning of shearing enzymes, such as endonucleases, in order to fragment the nucleic acids into smaller fragments. These endonucleases may include restriction endonucleases, including type II and type Us restriction endonucleases as well as other nucleic acid cleaving enzymes, such as nicking endonucleases, and the like. In some cases, fragmentation may not be implemented, and full length nucleic acids may be retained within the partitions, or in the case of encapsulated cells or cell contents, fragmentation may be carried out prior to partitioning, e.g., through enzymatic methods, e.g., those described herein, or through mechanical methods, e.g., mechanical, acoustic or other shearing.

[0266] Once co-partitioned, and the cells are lysed to release their nucleic acids, the oligonucleotides disposed upon the bead may be used to barcode and amplify fragments of those nucleic acids. Briefly, in one aspect, the oligonucleotides present on the beads that are co-partitioned with the cells, are released from their beads into the partition with the cell's nucleic acids. The oligonucleotides can include, along with the barcode sequence, a primer sequence at its 5'end. This primer sequence may be a random oligonucleotide sequence intended to randomly prime numerous different regions on the cell's nucleic acids, or it may be a specific primer sequence targeted to prime upstream of a specific targeted region of the cell's genome.

[0267] Once released, the primer portion of the oligonucleotide can anneal to a complementary region of the cell's nucleic acid. Extension reaction reagents, e.g., DNA polymerase, nucleoside triphosphates, co-factors (e.g., Mg2+ or Mn2+), that may also be co-partitioned with the cells and beads, then extend the primer sequence using the cell's nucleic acid as a template, to produce a complementary fragment to the strand of the cell's nucleic acid to which the primer annealed, which complementary fragment includes the oligonucleotide and its associated barcode sequence. Annealing and extension of multiple primers to different portions of the cell's nucleic acids will result in a large pool of overlapping complementary fragments of the nucleic acid, each possessing its own barcode sequence indicative of the partition in which it was created. In some cases, these complementary fragments may themselves be used as a template primed by the oligonucleotides present in the partition to produce a complement of the complement that again, includes the barcode sequence. In some cases, this replication process is configured such that when the first complement is duplicated, it produces two complementary sequences at or near its termini, to allow formation of a hairpin structure or partial hairpin structure that may reduce the ability of the molecule to be the basis for producing further iterative copies. As described herein, the cell's nucleic acids may include any nucleic acids within the cell

including, for example, the cell's DNA, e.g., genomic DNA, RNA, e.g., messenger RNA, and the like. For example, in some cases, the methods and systems described herein are used in characterizing expressed mRNA, including, e.g., the presence and quantification of such mRNA, and may include RNA sequencing processes as the characterization process. Alternatively or additionally, the reagents partitioned along with the cells may include reagents for the conversion of mRNA into cDNA, e.g., reverse transcriptase enzymes and reagents, to facilitate sequencing processes where DNA sequencing is employed. In some cases, where the nucleic acids to be characterized comprise RNA, e.g., mRNA, schematic illustration of one example of this is shown in FIG. **3**.

[0268] As shown, oligonucleotides that include a barcode sequence are co-partitioned in, e.g., a droplet **302** in an emulsion, along with a sample nucleic acid **304**. The oligonucleotides **308** may be provided on a bead **306** that is co-partitioned with the sample nucleic acid **304**, which oligonucleotides are releasable from the bead **306**, as shown in panel A. The oligonucleotides **308** may include a barcode sequence **312**, in addition to one or more functional sequences, e.g., sequences **310**, **314** and **316**. For example, oligonucleotide **308** is shown as comprising barcode sequence **312**, as well as sequence **310** that may function as an attachment or immobilization sequence for a given sequencing system, e.g., a P5 sequence used for attachment in flow cells of an Illumina Hiseq® or Miseq® system. As shown, the oligonucleotides also include a primer sequence **316**, which may include a random or targeted N-mer for priming replication of portions of the sample nucleic acid **304**. Also included within oligonucleotide **308** is a sequence **314** which may provide a sequencing priming region, such as a "read1" or R1 priming region, that is used to prime polymerase mediated, template directed sequencing by synthesis reactions in sequencing systems. As will be appreciated, the functional sequences may be selected to be compatible with a variety of different sequencing systems, e.g., 454 Sequencing, Ion Torrent Proton or PGM, Illumina X10, etc., and the requirements thereof. In some cases, the barcode sequence **312**, immobilization sequence **310** and R1 sequence **314** may be common to all of the oligonucleotides attached to a given bead. The primer sequence **316** may vary for random N-mer primers, or may be common to the oligonucleotides on a given bead for certain targeted applications. Moreover, in some cases, barcoded oligonucleotides may be generated as described in U.S. Patent Publication No. 20160257984, which is herein incorporated by reference in its entirety.

[0269] An oligonucleotide of an anchor agent or a labelling agent may comprise modifications that render it non-extendable by a polymerase. When binding to a nucleic acid in a sample for a primer extension reaction, the oligonucleotide may serve as a template, not a primer. When the oligonucleotide also comprises a barcode (e.g., the oligonucleotide is a reporter oligonucleotide), such design may increase the efficiency of molecular barcoding by increasing the affinity between the oligonucleotide and the unbarcoded sample nucleic acids, and eliminate the potential formation of adaptor artifacts. In some cases, the oligonucleotide may comprise a random N-mer sequence that is capped with modifications that render it non-extendable by a polymerase. In some cases, the composition of the random N-mer sequence may be designed to maximize the binding efficiency to free, unbarcoded ssDNA molecules. The design may include a random sequence composition with a higher GC content, a partial random sequence with fixed G or C at specific positions, the use of guanosines, the use of locked nucleic acids, or any combination thereof.

[0270] A modification for blocking primer extension by a polymerase may be a carbon spacer group of different lengths or a dideoxynucleotide. In some cases, the modification may be an abasic site that has an apurine or apyrimidine structure, a base analog, or an analogue of a phosphate backbone, such as a backbone of N-(2-aminoethyl)-glycine linked by amide bonds, tetrahydrofuran, or 1', 2'-Dideoxyribose. The modification may also be a uracil base, 2'OMe modified RNA, C3-18 spacers (e.g., structures with 3-18 consecutive carbon atoms, such as C3 spacer), ethylene eglycol multimer spacers (e.g., spacer 18 (hexa-ethyleneglycol spacer), biotin, di-deoxynucleotide triphosphate, ethylene glycol, amine, or phosphate.

[0271] FIG. **21** shows an oligonucleotide with such modification. The double-stranded oligonucleotide **2110** comprises a single-stranded DNA (ssDNA) annealing region with a random N-mer sequence at its 3' end. The unbarcoded ssDNA **2120** from a sample binds to oligonucleotide **2110**. The random N-mer sequence of the oligonucleotide **2110** has modifications (shown as "X") on the 3' end. When oligonucleotide **2110** and unbarcoded ssDNA **2120** bind to each other in a primer extension reaction, only unbarcoded ssDNA **2120** can be extended using oligonucleotide **3310** as a template.

[0272] In some cases, the oligonucleotide with a random N-mer sequence may be coupled to a solid support (e.g., a bead) via a U-excising element, e.g., an ssDNA sequence with uracil. FIG. **22** shows an example of such oligonucleotide. Double-stranded oligonucleotide **2210** comprises an ssDNA annealing region that contains a random N-mer sequence at its 3' end. Oligonucleotide **2210** is coupled to a bead via an ssDNA **2211** that has a uracil. Oligonucleotide **2210** also comprises modifications preventing extension by a polymerase. Oligonucleotide **2210** may be released from the bead by uracil-DNA glycosylase (to remove the uracil) and an endonuclease (to induce the ssDNA break), resulting the released oligonucleotide **2230**. Oligonucleotide **2220** comprises an ssDNA priming region has similar design as Oligonucleotide **2210**. In some cases, the difference between an ssDNA annealing region and an ssDNA priming region is the presence or absence of a blocking group (e.g., "X"), respectively. Unblocked ssDNA can be extended and function as a primer, while blocked ssDNA can function as a passive annealing sequence.

[0273] As will be appreciated, in some cases, the functional sequences may include primer sequences useful for RNA-seq applications. For example, in some cases, the oligonucleotides may include poly-T primers for priming reverse transcription of RNA for RNA-seq. In still other cases, oligonucleotides in a given partition, e.g., included on an individual bead, may include multiple types of primer sequences in addition to the common barcode sequences, such as both DNA-sequencing and RNA sequencing primers, e.g., poly-T primer sequences included within the oligonucleotides coupled to the bead. In such cases, a single partitioned cell may be both subjected to DNA and RNA sequencing processes.

[0274] A primer on a labelling agent or an anchor agent (e.g., a primer for RNA-seq applications) may be a target-

specific primer. A target-specific primer may bind to a specific sequence in a RNA molecule or a DNA molecule (e.g., complementary DNA (cDNA) from RNA, or endogenous DNA from a cell). For example, the specific sequence may be a sequence that is not in the poly-A tail of an RNA molecule or its cDNA. In some cases, the target-specific primer may bind to RNA molecules such as mRNA molecules or non-coding RNA molecules, e.g., rRNA, tRNA, mRNA, or miRNA molecules. In some cases, the target-specific primer may bind to RNA molecules introduced to a cell. In some cases, the RNA molecules introduced to a cell may be RNA molecules used in gene editing methods (e.g., Clustered regularly interspaced short palindromic repeats (CRISPR) RNA (crRNA) or guide RNA for CRISPR gene editing). For example, the target-specific primer may bind to crRNA for identifying the crRNA introduced to a cell and/or determining the effect of the crRNA on the transcriptome of the cell. In some cases, the target-specific primer may be used to determine copy numbers of disease (e.g., cancer)-related genes while simultaneously analyzing the rest of the transcriptome. In other cases, the target-specific primer may be used to analyze RNA molecules from pathogens infecting the cell, e.g., for distinguishing pathogen infected cells from non-pathogen infected cells and/or determining how the pathogen alters the cells transcriptome. In some cases, a target-specific primer may bind to DNA molecules, e.g., endogenous DNA molecules from a cell, or synthetic DNA molecules. For example, a target-specific primer may bind to a barcode, e.g., a barcode of a cell (e.g., inside a cell or on the surface of a cell), a barcode of a protein (e.g., an antibody barcode), or a barcode of a nucleic acid (e.g., a CRISPR barcode).

[0275] A target-specific primer may be combined with one or more barcodes, one or more UMIs, one or more poly-T primers for mRNA, and/or one or more random N-mer primers (randomers) for total RNA in the same or different oligonucleotides. In some cases, a bead disclosed herein may comprise an oligonucleotide with a target-specific primer and one or more oligonucleotides with a poly-T primer, e.g., as shown in FIG. **23**A. In some cases, a bead may have a plurality of oligonucleotides, each of which comprises a target-specific primer, e.g., as shown in FIG. **23**B. In some cases, a bead may have a plurality of oligonucleotides, each of which comprises a target-specific primer and a plurality of oligonucleotides, each of which comprises a poly-T primer, e.g., as shown in FIG. **23**C. In some cases, a bead may have a plurality of oligonucleotides, each of which comprises a target-specific primer and a plurality of oligonucleotides, each of which comprises a random N-mer primer for total RNA, e.g., as shown in FIG. **24**.

[0276] On a bead, the ratio of oligonucleotides with target-specific primers to oligonucleotides with non-specific (poly-T or random N-mer) primers may be adjusted to match the needs of a specific application. In some cases, at least 0.1%, 1%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, or 100% of the oligonucleotides on a bead may comprise target-specific primers. In some cases, at least 0.1%, 1%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, or 100% of the oligonucleotides on a bead may comprise non-specific (poly-T or random N-mer) primers. The oligonucleotide may be made by attaching (e.g., by ligation) one or more oligonucleotide backbones on a bead and then attaching (e.g., by ligation) one or more primer sequences to the backbones.

[0277] An oligonucleotide of an anchor agent or a labelling agent may be a splint oligonucleotide. A splint oligonucleotide may comprise two or more different primers. The primers may have different functions. For example, a splint oligonucleotide may comprise two or more of the following: a poly-T primer, a random N-mer primer, and a target-specific primer.

[0278] An oligonucleotide of an anchor agent or a labelling agent may comprise an adapter that is capable of binding or ligating to an assay primer. The adapter may allow the anchor agent or the labelling agent to be attached to any suitable assay primers and used in any suitable assays. The assay primer may comprise a priming region and a sequence that is capable of binding or ligating to the adapter. In some cases, the adapter may be a non-specific primer (e.g., a 5' overhang) and the assay primer may comprise a 3' overhang that can be ligated to the 5' overhang. The priming region on the assay primer may be any primer described herein, e.g., a poly-T primer, a random N-mer primer, a target-specific primer, or a labelling agent capture sequence. FIG. **25**A shows exemplary adapters and assay primers. Oligonucleotide **2510** comprises an adapter **2511**, which is a 5' overhang comprising 10 nucleotides. The adapter **2511** can be ligated to the assay primers, each of which comprises a 3' overhang comprising 10 nucleotides that complementary to the 5' overhang of adapter **2511**. The anchor oligonucleotide may be used in any assay by attaching to the assay primer designed for that assay. FIG. **26**B shows exemplary adapters and assay primers that allows the anchor agent or the labelling agent to be attached to any suitable assay primers and used in any suitable assays. Barcoded adapter oligonucleotide **2561** is attached to a bead **2560**, such as a gel bead, and comprises a poly(dT) sequence **2562**. FIG. **26**C shows exemplary splint oligos comprising a poly-A sequence that facilitates coupling to the barcoded adapter oligonucleotide **2561** and a second sequence (shown as "XXX", "YYY", and "ZZZ") that facilitates coupling with an assay primer. Assay primers comprise a sequence complementary to the splint oligo second sequence (shown as "X'X'X'", "Y'Y'Y'", and "Z'Z'Z'") and an assay-specific sequence that determines assay primer functionality (e.g., a poly-T primer, a random N-mer primer, a target-specific primer, or a labelling agent capture sequence as described herein).

[0279] In some cases, the barcoded adapter comprises a switch oligo, e.g., with a 3' end 3rG. FIG. **26** shows a bead (such as a gel bead) comprising a barcoded adapter oligonucleotide functionalized with a 3rG sequence that enables template switching (e.g., reverse transcriptase template switching), but is not specific for any particular assay. Assay primers added to the reaction determine the particular assay by binding to targeted molecules and are extended by a reverse transcriptase enzyme/polymerase followed by template switching onto the barcoded adapter oligonucleotide to incorporate the barcode and other functional sequences. The priming region determines the assay and, in some embodiments, comprises a poly-T sequence for mRNA analysis, random primers for gDNA analysis, or a capture sequence that can bind a nucleic acid molecule coupled to a labelling agent (e.g., an antibody) or a nucleic acid molecule that can function in a CRISPR assay (e.g., CRISPR/Cas9) via a targeted priming sequence.

[0280] Based upon the presence of primer sequence **316**, the oligonucleotides can prime the sample nucleic acid as shown in panel B, which allows for extension of the oligonucleotides **308** and **308***a* using polymerase enzymes and other extension reagents also co-partitioned with the bead **306** and sample nucleic acid **304**. As shown in panel C, following extension of the oligonucleotides that, for random N-mer primers, may anneal to multiple different regions of the sample nucleic acid **304**; multiple overlapping complements or fragments of the nucleic acid are created, e.g., fragments **318** and **320**. Although including sequence portions that are complementary to portions of sample nucleic acid, e.g., sequences **322** and **324**, these constructs are generally referred to herein as comprising fragments of the sample nucleic acid **304**, having the attached barcode sequences.

[0281] The barcoded nucleic acid fragments may then be subjected to characterization, e.g., through sequence analysis, or they may be further amplified in the process, as shown in panel D. For example, additional oligonucleotides, e.g., oligonucleotide **308***b*, also released from bead **306**, may prime the fragments **318** and **320**. This shown for fragment **318**. In particular, again, based upon the presence of the random N-mer primer **316***b* in oligonucleotide **308***b* (which in some cases can be different from other random N-mers in a given partition, e.g., primer sequence **316**), the oligonucleotide anneals with the fragment **318**, and is extended to create a complement **326** to at least a portion of fragment **318** which includes sequence **328**, that comprises a duplicate of a portion of the sample nucleic acid sequence. Extension of the oligonucleotide **308***b* continues until it has replicated through the oligonucleotide portion **308** of fragment **318**. As illustrated in panel D, the oligonucleotides may be configured to prompt a stop in the replication by the polymerase at a given point, e.g., after replicating through sequences **316** and **314** of oligonucleotide **308** that is included within fragment **318**. As described herein, this may be accomplished by different methods, including, for example, the incorporation of different nucleotides and/or nucleotide analogues that are not capable of being processed by the polymerase enzyme used. For example, this may include the inclusion of uracil containing nucleotides within the sequence region **312** to prevent a non-uracil tolerant polymerase to cease replication of that region. As a result a fragment **326** is created that includes the full-length oligonucleotide **308***b* at one end, including the barcode sequence **312**, the attachment sequence **310**, the R1 primer region **314**, and the random N-mer sequence **316***b*. At the other end of the sequence may be included the complement **316'** to the random N-mer of the first oligonucleotide **308**, as well as a complement to all or a portion of the R1 sequence, shown as sequence **314'**. The R1 sequence **314** and its complement **314'** are then able to hybridize together to form a partial hairpin structure **328**. As will be appreciated because the random N-mers differ among different oligonucleotides, these sequences and their complements may not be expected to participate in hairpin formation, e.g., sequence **316'**, which is the complement to random N-mer **316**, may not be expected to be complementary to random N-mer sequence **316***b*. This may not be the case for other applications, e.g., targeted primers, where the N-mers may be common among oligonucleotides within a given partition.

[0282] By forming these partial hairpin structures, it allows for the removal of first level duplicates of the sample sequence from further replication, e.g., preventing iterative copying of copies. The partial hairpin structure also provides a useful structure for subsequent processing of the created fragments, e.g., fragment **326**.

[0283] In general, the amplification of the cell's nucleic acids is carried out until the barcoded overlapping fragments within the partition constitute at least 1× coverage of the particular portion or all of the cell's genome, at least 2×, at least 3×, at least 4×, at least 5×, at least 10×, at least 20×, at least 40× or more coverage of the genome or its relevant portion of interest. Once the barcoded fragments are produced, they may be directly sequenced on an appropriate sequencing system, e.g., an Illumina Hiseq®, Miseq® or X10 system, or they may be subjected to additional processing, such as further amplification, attachment of other functional sequences, e.g., second sequencing primers, for reverse reads, sample index sequences, and the like.

[0284] All of the fragments from multiple different partitions may then be pooled for sequencing on high throughput sequencers as described herein, where the pooled fragments comprise a large number of fragments derived from the nucleic acids of different cells or small cell populations, but where the fragments from the nucleic acids of a given cell will share the same barcode sequence. In particular, because each fragment is coded as to its partition of origin, and consequently its single cell or small population of cells, the sequence of that fragment may be attributed back to that cell or those cells based upon the presence of the barcode, which will also aid in applying the various sequence fragments from multiple partitions to assembly of individual genomes for different cells. This is schematically illustrated in FIG. **4**. As shown in one example, a first nucleic acid **404** from a first cell **400**, and a second nucleic acid **406** from a second cell **402** are each partitioned along with their own sets of barcode oligonucleotides as described above. The nucleic acids may comprise a chromosome, entire genome or other large nucleic acid from the cells.

[0285] Within each partition, each cell's nucleic acids **404** and **406** is then processed to separately provide overlapping set of second fragments of the first fragment(s), e.g., second fragment sets **408** and **410**. This processing also provides the second fragments with a barcode sequence that is the same for each of the second fragments derived from a particular first fragment. As shown, the barcode sequence for second fragment set **408** is denoted by "1" while the barcode sequence for fragment set **410** is denoted by "2". A diverse library of barcodes may be used to differentially barcode large numbers of different fragment sets. However, it is not necessary for every second fragment set from a different first fragment to be barcoded with different barcode sequences. In some cases, multiple different first fragments may be processed concurrently to include the same barcode sequence. Diverse barcode libraries are described in detail elsewhere herein.

[0286] The barcoded fragments, e.g., from fragment sets **408** and **410**, may then be pooled for sequencing using, for example, sequence by synthesis technologies available from Illumina or Ion Torrent division of Thermo-Fisher, Inc. Once sequenced, the sequence reads **412** can be attributed to their respective fragment set, e.g., as shown in aggregated reads **414** and **416**, at least in part based upon the included barcodes, and in some cases, in part based upon the sequence of the fragment itself. The attributed sequence reads for each fragment set are then assembled to provide the assembled

sequence for each cell's nucleic acids, e.g., sequences **418** and **420**, which in turn, may be attributed to individual cells, e.g., cells **400** and **402**.

[0287] While described in terms of analyzing the genetic material present within cells, the methods and systems described herein may have much broader applicability, including the ability to characterize other aspects of individual cells or cell populations, by allowing for the allocation of reagents and/or agents to individual cells, and providing for the attributable analysis or characterization of those cells in response to those reagents and/or agents. These methods and systems may be valuable in being able to characterize cells for, e.g., research, diagnostic, or pathogen identification. By way of example, a wide range of different cell surface features, e.g., cell surface proteins like cluster of differentiation or CD proteins, have significant diagnostic relevance in characterization of diseases like cancer.

[0288] In one particularly useful application, the methods and systems described herein may be used to characterize cell features, such as cell surface features. Cell surface features may include, but are not limited to, a receptor, an antigen, a surface protein, a transmembrane protein, a cluster of differentiation protein, a protein channel, a protein pump, a carrier protein, a phospholipid, a glycoprotein, a glycolipid, a cell-cell interaction protein complex, an antigen-presenting complex, a major histocompatibility complex, an engineered T-cell receptor, a T-cell receptor, a B-cell receptor, a chimeric antigen receptor, a gap junction, and an adherens junction. In particular, the methods described herein may be used to attach one or more labelling agents to these cell features, that when partitioned as described above, may be barcoded and analyzed, e.g., using DNA sequencing technologies, to ascertain the presence, and in some cases, relative abundance or quantity of such cell features of an individual cell or population of cells.

[0289] In a particular example, a library of potential cell surface feature labelling agents may be provided associated with a first set of nucleic acid reporter molecules, e.g., where a different reporter oligonucleotide sequence is associated with a specific labelling agent, and therefore capable of binding to a specific cell surface feature. Cell surface feature labelling agents may include, but are not limited to, an antibody, or an epitope binding fragment thereof, a cell surface receptor binding molecule, a receptor ligand, a small molecule, a bi-specific antibody, a bi-specific T-cell engager, a T-cell receptor engager, a B-cell receptor engager, a pro-body, an aptamer, a monobody, an affimer, a darpin, and a protein scaffold. In some aspects, different members of the library may be characterized by the presence of a different oligonucleotide sequence label, e.g., an antibody to a first type of cell surface protein or receptor may have associated with it a first known reporter oligonucleotide sequence, while an antibody to a second receptor protein may have a different known reporter oligonucleotide sequence associated with it. Prior to co-partitioning, the cells may be incubated with the library of labelling agents, that may represent antibodies to a broad panel of different cell surface features, e.g., receptors, proteins, etc., and which include their associated reporter oligonucleotides. Unbound labelling agents may be washed from the cells, and the cells may then be co-partitioned along with the barcode oligonucleotides described above. As a result, the partitions may include the cell or cells, as well as the bound labelling agents and their known, associated reporter oligonucleotides.

[0290] Without the need for lysing the cells within the partitions, one may then subject the reporter oligonucleotides to the barcoding operations described above for cellular nucleic acids, to produce barcoded, reporter oligonucleotides, where the presence of the reporter oligonucleotides can be indicative of the presence of the particular cell surface feature, and the barcode sequence will allow the attribution of the range of different cell surface features to a given individual cell or population of cells based upon the barcode sequence that was co-partitioned with that cell or population of cells. As a result, one may generate a cell-by-cell profile of the cell surface features within a broader population of cells. This aspect of the methods and systems described herein, is described in greater detail below.

[0291] This example is schematically illustrated in FIG. **5**. As shown, a population of cells, represented by cells **502** and **504** are incubated with a library of cell surface associated labelling agents, e.g., antibodies, antibody fragments, cell surface receptor binding molecules, receptor ligands, small molecules, bi-specific antibodies, bi-specific T-cell engagers, T-cell receptor engagers, B-cell receptor engagers, pro-bodies, aptamers, monobodies, affimers, darpins, protein scaffolds, or the like, where each different type of binding group includes an associated nucleic acid reporter molecule associated with it, shown as labelling agents and associated reporter oligonucleotide **506**, **508**, **510** and **512** (with the reporter oligonucleotides being indicated by the differently shaded circles). Where the cell expresses the surface features that are bound by the library of labelling agents, the labelling agents and their associated reporter oligonucleotides can become associated or coupled with the cell surface feature. Individual cells may then be partitioned into separate partitions, e.g., droplets **514** and **516**, as described herein, along with their associated labelling agents/reporter oligonucleotides, as well as a bead containing individual barcode oligonucleotides (e.g., anchor oligonucleotides) as described elsewhere herein, e.g., beads **518** and **520**, respectively. As with other examples described herein, the barcoded oligonucleotides may be released from the beads and used to attach the barcode sequence the reporter oligonucleotides present within each partition with a barcode that is common to a given partition, but which varies widely among different partitions. For example, as shown in FIG. **5**, the reporter oligonucleotides that associate with cell **502** in partition **514** are barcoded with barcode sequence **522**, while the reporter oligonucleotides associated with cell **504** in partition **516** are barcoded with barcode sequence **524**. As a result, one is provided with a library of oligonucleotides that reflects the surface features of the cell, as reflected by the reporter molecule, but which is substantially attributable to an individual cell by virtue of a common barcode sequence, allowing a single cell level profiling of the surface characteristics of the cell. As will be appreciated, this process is not limited to cell surface receptors but may be used to identify the presence of a wide variety of specific cell structures, chemistries or other characteristics.

[0292] Single cell processing and analysis methods and systems described herein can be utilized for a wide variety of applications, including analysis of specific individual cells, analysis of different cell types within populations of differing cell types, analysis and characterization of large populations of cells for environmental, human health, epidemiological forensic, or any of a wide variety of different applications.

[0293] A particularly valuable application of the single cell analysis processes described herein is in the sequencing and characterization of a diseased cell. A diseased cell can have altered metabolic properties, gene expression, protein expression, and/or morphologic features. Examples of diseases include inflammatory disorders, metabolic disorders, nervous system disorders, and cancer.

[0294] Of particular interest are cancer cells. In particular, conventional analytical techniques, including the ensemble sequencing processes alluded to above, are not highly adept at picking small variations in genomic make-up of cancer cells, particularly where those exist in a sea of normal tissue cells. Further, even as between tumor cells, wide variations can exist and can be masked by the ensemble approaches to sequencing (See, e.g., Patel, et al., Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma, Science DOI: 10.1126/science.1254257 (Published online Jun. 12, 2014). Cancer cells may be derived from solid tumors, hematological malignancies, cell lines, or obtained as circulating tumor cells, and subjected to the partitioning processes described above. Upon analysis, one can identify individual cell sequences as deriving from a single cell or small group of cells, and distinguish those over normal tissue cell sequences.

[0295] Non-limiting examples of cancer cells include cells of cancers such as Acanthoma, Acinic cell carcinoma, Acoustic neuroma, Acral lentiginous melanoma, Acrospiroma, Acute eosinophilic leukemia, Acute lymphoblastic leukemia, Acute megakaryoblastic leukemia, Acute monocytic leukemia, Acute myeloblastic leukemia with maturation, Acute myeloid dendritic cell leukemia, Acute myeloid leukemia, Acute promyelocytic leukemia, Adamantinoma, Adenocarcinoma, Adenoid cystic carcinoma, Adenoma, Adenomatoid odontogenic tumor, Adrenocortical carcinoma, Adult T-cell leukemia, Aggressive NK-cell leukemia, AIDS-Related Cancers, AIDS-related lymphoma, Alveolar soft part sarcoma, Ameloblastic fibroma, Anal cancer, Anaplastic large cell lymphoma, Anaplastic thyroid cancer, Angioimmunoblastic T-cell lymphoma, Angiomyolipoma, Angiosarcoma, Appendix cancer, Astrocytoma, Atypical teratoid rhabdoid tumor, Basal cell carcinoma, Basal-like carcinoma, B-cell leukemia, B-cell lymphoma, Bellini duct carcinoma, Biliary tract cancer, Bladder cancer, Blastoma, Bone Cancer, Bone tumor, Brain Stem Glioma, Brain Tumor, Breast Cancer, Brenner tumor, Bronchial Tumor, Bronchioloalveolar carcinoma, Brown tumor, Burkitt's lymphoma, Cancer of Unknown Primary Site, Carcinoid Tumor, Carcinoma, Carcinoma in situ, Carcinoma of the penis, Carcinoma of Unknown Primary Site, Carcinosarcoma, Castleman's Disease, Central Nervous System Embryonal Tumor, Cerebellar Astrocytoma, Cerebral Astrocytoma, Cervical Cancer, Cholangiocarcinoma, Chondroma, Chondrosarcoma, Chordoma, Choriocarcinoma, Choroid plexus papilloma, Chronic Lymphocytic Leukemia, Chronic monocytic leukemia, Chronic myelogenous leukemia, Chronic Myeloproliferative Disorder, Chronic neutrophilic leukemia, Clear-cell tumor, Colon Cancer, Colorectal cancer, Craniopharyngioma, Cutaneous T-cell lymphoma, Degos disease, Dermatofibrosarcoma protuberans, Dermoid cyst, Desmoplastic small round cell tumor, Diffuse large B cell lymphoma, Dysembryoplastic neuroepithelial tumor, Embryonal carcinoma, Endodermal sinus tumor, Endometrial cancer, Endometrial Uterine Cancer, Endometrioid tumor, Enteropathy-associated T-cell lymphoma, Ependymoblastoma, Ependymoma, Epithelioid sarcoma, Erythroleukemia, Esophageal cancer, Esthesioneuroblastoma, Ewing Family of Tumor, Ewing Family Sarcoma, Ewing's sarcoma, Extracranial Germ Cell Tumor, Extragonadal Germ Cell Tumor, Extrahepatic Bile Duct Cancer, Extramammary Paget's disease, Fallopian tube cancer, Fetus in fetu, Fibroma, Fibrosarcoma, Follicular lymphoma, Follicular thyroid cancer, Gallbladder Cancer, Gallbladder cancer, Ganglioglioma, Ganglioneuroma, Gastric Cancer, Gastric lymphoma, Gastrointestinal cancer, Gastrointestinal Carcinoid Tumor, Gastrointestinal Stromal Tumor, Gastrointestinal stromal tumor, Germ cell tumor, Germinoma, Gestational choriocarcinoma, Gestational Trophoblastic Tumor, Giant cell tumor of bone, Glioblastoma multiforme, Glioma, Gliomatosis cerebri, Glomus tumor, Glucagonoma, Gonadoblastoma, Granulosa cell tumor, Hairy Cell Leukemia, Hairy cell leukemia, Head and Neck Cancer, Head and neck cancer, Heart cancer, Hemangioblastoma, Hemangiopericytoma, Hemangiosarcoma, Hematological malignancy, Hepatocellular carcinoma, Hepatosplenic T-cell lymphoma, Hereditary breast-ovarian cancer syndrome, Hodgkin Lymphoma, Hodgkin's lymphoma, Hypopharyngeal Cancer, Hypothalamic Glioma, Inflammatory breast cancer, Intraocular Melanoma, Islet cell carcinoma, Islet Cell Tumor, Juvenile myelomonocytic leukemia, Kaposi Sarcoma, Kaposi's sarcoma, Kidney Cancer, Klatskin tumor, Krukenberg tumor, Laryngeal Cancer, Laryngeal cancer, Lentigo maligna melanoma, Leukemia, Leukemia, Lip and Oral Cavity Cancer, Liposarcoma, Lung cancer, Luteoma, Lymphangioma, Lymphangiosarcoma, Lymphoepithelioma, Lymphoid leukemia, Lymphoma, Macroglobulinemia, Malignant Fibrous Histiocytoma, Malignant fibrous histiocytoma, Malignant Fibrous Histiocytoma of Bone, Malignant Glioma, Malignant Mesothelioma, Malignant peripheral nerve sheath tumor, Malignant rhabdoid tumor, Malignant triton tumor, MALT lymphoma, Mantle cell lymphoma, Mast cell leukemia, Mediastinal germ cell tumor, Mediastinal tumor, Medullary thyroid cancer, Medulloblastoma, Medulloblastoma, Medulloepithelioma, Melanoma, Melanoma, Meningioma, Merkel Cell Carcinoma, Mesothelioma, Mesothelioma, Metastatic Squamous Neck Cancer with Occult Primary, Metastatic urothelial carcinoma, Mixed Mullerian tumor, Monocytic leukemia, Mouth Cancer, Mucinous tumor, Multiple Endocrine Neoplasia Syndrome, Multiple Myeloma, Multiple myeloma, Mycosis Fungoides, Mycosis fungoides, Myelodysplastic Disease, Myelodysplastic Syndromes, Myeloid leukemia, Myeloid sarcoma, Myeloproliferative Disease, Myxoma, Nasal Cavity Cancer, Nasopharyngeal Cancer, Nasopharyngeal carcinoma, Neoplasm, Neurinoma, Neuroblastoma, Neuroblastoma, Neurofibroma, Neuroma, Nodular melanoma, Non-Hodgkin Lymphoma, Non-Hodgkin lymphoma, Nonmelanoma Skin Cancer, Non-Small Cell Lung Cancer, Ocular oncology, Oligoastrocytoma, Oligodendroglioma, Oncocytoma, Optic nerve sheath meningioma, Oral Cancer, Oral cancer, Oropharyngeal Cancer, Osteosarcoma, Osteosarcoma, Ovarian Cancer, Ovarian cancer, Ovarian Epithelial Cancer, Ovarian Germ Cell Tumor, Ovarian Low Malignant Potential Tumor, Paget's disease of the breast, Pancoast tumor, Pancreatic Cancer, Pancreatic cancer, Papillary thyroid cancer, Papillomatosis, Paraganglioma, Paranasal Sinus Cancer, Parathyroid Cancer, Penile Cancer, Perivascular epithelioid cell tumor, Pharyngeal Cancer, Pheochromocytoma, Pineal Parenchymal Tumor of Intermediate Differen-

tiation, Pineoblastoma, Pituicytoma, Pituitary adenoma, Pituitary tumor, Plasma Cell Neoplasm, Pleuropulmonary blastoma, Polyembryoma, Precursor T-lymphoblastic lymphoma, Primary central nervous system lymphoma, Primary effusion lymphoma, Primary Hepatocellular Cancer, Primary Liver Cancer, Primary peritoneal cancer, Primitive neuroectodermal tumor, Prostate cancer, Pseudomyxoma peritonei, Rectal Cancer, Renal cell carcinoma, Respiratory Tract Carcinoma Involving the NUT Gene on Chromosome 15, Retinoblastoma, Rhabdomyoma, Rhabdomyosarcoma, Richter's transformation, Sacrococcygeal teratoma, Salivary Gland Cancer, Sarcoma, Schwannomatosis, Sebaceous gland carcinoma, Secondary neoplasm, Seminoma, Serous tumor, Sertoli-Leydig cell tumor, Sex cord-stromal tumor, Sezary Syndrome, Signet ring cell carcinoma, Skin Cancer, Small blue round cell tumor, Small cell carcinoma, Small Cell Lung Cancer, Small cell lymphoma, Small intestine cancer, Soft tissue sarcoma, Somatostatinoma, Soot wart, Spinal Cord Tumor, Spinal tumor, Splenic marginal zone lymphoma, Squamous cell carcinoma, Stomach cancer, Superficial spreading melanoma, Supratentorial Primitive Neuroectodermal Tumor, Surface epithelial-stromal tumor, Synovial sarcoma, T-cell acute lymphoblastic leukemia, T-cell large granular lymphocyte leukemia, T-cell leukemia, T-cell lymphoma, T-cell prolymphocytic leukemia, Teratoma, Terminal lymphatic cancer, Testicular cancer, Thecoma, Throat Cancer, Thymic Carcinoma, Thymoma, Thyroid cancer, Transitional Cell Cancer of Renal Pelvis and Ureter, Transitional cell carcinoma, Urachal cancer, Urethral cancer, Urogenital neoplasm, Uterine sarcoma, Uveal melanoma, Vaginal Cancer, Verner Morrison syndrome, Verrucous carcinoma, Visual Pathway Glioma, Vulvar Cancer, Waldenstrom's macroglobulinemia, Warthin's tumor, Wilms' tumor, and combinations thereof.

[0296] Where cancer cells are to be analyzed, primer sequences useful in any of the various operations for attaching barcode sequences and/or amplification reactions may comprise gene specific sequences which target genes or regions of genes associated with or suspected of being associated with cancer. For example, this can include genes or regions of genes where the presence of mutations (e.g., insertions, deletions, polymorphisms, copy number variations, and gene fusions) associated with a cancerous condition are suspected to be present in a cell population.

[0297] As with cancer cell analysis, the analysis and diagnosis of fetal health or abnormality through the analysis of fetal cells is a difficult task using conventional techniques. In particular, in the absence of relatively invasive procedures, such as amniocentesis obtaining fetal cell samples can employ harvesting those cells from the maternal circulation. As will be appreciated, such circulating fetal cells make up an extremely small fraction of the overall cellular population of that circulation. As a result complex analyses are performed in order to characterize what of the obtained data is likely derived from fetal cells as opposed to maternal cells. By employing the single cell characterization methods and systems described herein, however, one can attribute genetic make up to individual cells, and categorize those cells as maternal or fetal based upon their respective genetic make-up. Further, the genetic sequence of fetal cells may be used to identify any of a number of genetic disorders, including, e.g., aneuploidy such as Down syndrome, Edwards syndrome, and Patau syndrome. Further, the cell surface features of fetal cells may be used to identify any of a number of disorders or diseases.

[0298] Also of interest are immune cells. The methods, compositions, and systems disclosed herein can be utilized for sequence analysis of the immune repertoire, including genomic, proteomic, and cell surface features. Analysis of information underlying the immune repertoire can provide a significant improvement in understanding the status and function of the immune system.

[0299] Non-limiting examples of immune cells which can be analyzed utilizing the methods described herein include B cells, T cells (e.g., cytotoxic T cells, natural killer T cells, regulatory T cells, and T helper cells), natural killer cells, cytokine induced killer (CIK) cells; myeloid cells, such as granulocytes (basophil granulocytes, eosinophil granulocytes, neutrophil granulocytes/hypersegmented neutrophils), monocytes/macrophages, mast cell, thrombocytes/megakaryocytes, and dendritic cells. In some embodiments, individual T cells are analyzed using the methods disclosed herein. In some embodiments, individual B cells are analyzed using the methods disclosed herein.

[0300] Immune cells express various adaptive immunological receptors relating to immune function, such as T cell receptors and B cell receptors. T cell receptors and B cells receptors play a part in the immune response by specifically recognizing and binding to antigens and aiding in their destruction.

[0301] The T cell receptor, or TCR, is a molecule found on the surface of T cells that is generally responsible for recognizing fragments of antigen as peptides bound to major histocompatibility complex (MHC) molecules. The TCR is generally a heterodimer of two chains, each of which is a member of the immunoglobulin superfamily, possessing an N-terminal variable (V) domain, and a C terminal constant domain. In humans, in 95% of T cells the TCR consists of an alpha (α) and beta (β) chain, whereas in 5% of T cells the TCR consists of gamma and delta (γ/δ) chains. This ratio can change during ontogeny and in diseased states as well as in different species. When the TCR engages with antigenic peptide and MHC (peptide/MHC), the T lymphocyte is activated through signal transduction.

[0302] Each of the two chains of a TCR contains multiple copies of gene segments—a variable 'V' gene segment, a diversity 'D' gene segment, and a joining 'J' gene segment. The TCR alpha chain is generated by recombination of V and J segments, while the beta chain is generated by recombination of V, D, and J segments. Similarly, generation of the TCR gamma chain involves recombination of V and J gene segments, while generation of the TCR delta chain occurs by recombination of V, D, and J gene segments. The intersection of these specific regions (V and J for the alpha or gamma chain, or V, D and J for the beta or delta chain) corresponds to the CDR3 region that is important for antigen-MHC recognition. Complementarity determining regions (e.g., CDR1, CDR2, and CDR3), or hypervariable regions, are sequences in the variable domains of antigen receptors (e.g., T cell receptor and immunoglobulin) that can complement an antigen. Most of the diversity of CDRs is found in CDR3, with the diversity being generated by somatic recombination events during the development of T lymphocytes. A unique nucleotide sequence that arises during the gene arrangement process can be referred to as a clonotype.

[0303] The B cell receptor, or BCR, is a molecule found on the surface of B cells. The antigen binding portion of a BCR is composed of a membrane-bound antibody that, like most antibodies (e.g., immunoglobulins), has a unique and randomly determined antigen-binding site. The antigen binding portion of a BCR includes membrane-bound immunoglobulin molecule of one isotype (e.g., IgD, IgM, IgA, IgG, or IgE). When a B cell is activated by its first encounter with a cognate antigen, the cell proliferates and differentiates to generate a population of antibody-secreting plasma B cells and memory B cells. The various immunoglobulin isotypes differ in their biological features, structure, target specificity and distribution. A variety of molecular mechanisms exist to generate initial diversity, including genetic recombination at multiple sites.

[0304] The BCR is composed of two genes IgH and IgK (or IgL) coding for antibody heavy and light chains. Immunoglobulins are formed by recombination among gene segments, sequence diversification at the junctions of these segments, and point mutations throughout the gene. Each heavy chain gene contains multiple copies of three different gene segments—a variable 'V' gene segment, a diversity 'D' gene segment, and a joining 'J' gene segment. Each light chain gene contains multiple copies of two different gene segments for the variable region of the protein—a variable 'V' gene segment and a joining 'J' gene segment. The recombination can generate a molecule with one of each of the V, D, and J segments. Furthermore, several bases may be deleted and others added (called N and P nucleotides) at each of the two junctions, thereby generating further diversity. After B cell activation, a process of affinity maturation through somatic hypermutation occurs. In this process progeny cells of the activated B cells accumulate distinct somatic mutations throughout the gene with higher mutation concentration in the CDR regions leading to the generation of antibodies with higher affinity to the antigens. In addition to somatic hypermutation activated B cells undergo the process of isotype switching. Antibodies with the same variable segments can have different forms (isotypes) depending on the constant segment. Whereas all naïve B cells express IgM (or IgD), activated B cells mostly express IgG but also IgM, IgA and IgE. This expression switching from IgM (and/or IgD) to IgG, IgA, or IgE occurs through a recombination event causing one cell to specialize in producing a specific isotype. A unique nucleotide sequence that arises during the gene arrangement process can similarly be referred to as a clonotype.

[0305] In some embodiments, the methods, compositions and systems disclosed herein are utilized to analyze the various sequences of TCRs and BCRs from immune cells, for example various clonotypes. In some embodiments, methods, compositions and systems disclosed herein are used to analyze the sequence of a TCR alpha chain, a TCR beta chain, a TCR delta chain, a TCR gamma chain, or any fragment thereof (e.g., variable regions including VDJ or VJ regions, constant regions, transmembrane regions, fragments thereof, combinations thereof, and combinations of fragments thereof). In some embodiments, methods, compositions and systems disclosed herein are used to analyze the sequence of a B cell receptor heavy chain, B cell receptor light chain, or any fragment thereof (e.g., variable regions including VDJ or VJ regions, constant regions, transmembrane regions, fragments thereof, combinations thereof, and combinations of fragments thereof).

[0306] Where immune cells are to be analyzed, primer sequences useful in any of the various operations for attaching barcode sequences and/or amplification reactions may comprise gene specific sequences which target genes or regions of genes of immune cell proteins, for example immune receptors. Such gene sequences include, but are not limited to, sequences of various T cell receptor alpha variable genes (TRAV genes), T cell receptor alpha joining genes (TRAJ genes), T cell receptor alpha constant genes (TRAC genes), T cell receptor beta variable genes (TRBV genes), T cell receptor beta diversity genes (TRBD genes), T cell receptor beta joining genes (TRBJ genes), T cell receptor beta constant genes (TRBC genes), T cell receptor gamma variable genes (TRGV genes), T cell receptor gamma joining genes (TRGJ genes), T cell receptor gamma constant genes (TRGC genes), T cell receptor delta variable genes (TRDV genes), T cell receptor delta diversity genes (TRDD genes), T cell receptor delta joining genes (TRDJ genes), and T cell receptor delta constant genes (TRDC genes).

[0307] MHCs, including full or partial MHC-peptides, may be used as labelling agents that are coupled to oligonucleotides that comprise a barcode sequence that identifies its associated MHC (and, thus, for example, the MHC's TCR binding partner). In some cases, MHCs are used to analyze one or more cell-surface features of a T-cell, such as a TCR. In some cases, multiple MHCs are associated together in a larger complex to improve binding affinity of MHCs to TCRs via multiple ligand binding synergies.

[0308] For example, as shown in FIG. **56**A, the MHC peptides can individually be associated with biotin and bound to a streptavidin moiety such that the streptavidin moiety comprises multiple MHC moieties. Each of these moieties can bind to a TCR such that the streptavidin binds to the target T-cell via multiple MCH/TCR binding interactions. These multiple interactions synergize and can substantially improve binding affinity. Such improved affinity can improve labelling of T-cells and also reduce the likelihood that labels will dissociate from T-cell surfaces.

[0309] As shown in FIG. **56**B and continuing with this example, a barcoded oligonucleotide **5601** can be modified with streptavidin **5602** and contacted with multiple molecules of biotinylated MHC **5606** such that the biotinylated MHC **5606** molecules are coupled with the streptavidin conjugated barcoded oligonucleotide **5601**. The result is a barcoded MHC multimer complex **5608**. As shown in FIG. **56**B, the oligonucleotide **5601** barcode sequence **5602** can identify the MHC **5604** as its associated label and also includes sequences for hybridization with other oligonucleotides (e.g., sequence **5603** comprising a 'Spacer C C C' and sequence **5605** comprising a 'Spacer PCR handle'). One such other oligonucleotide is oligonucleotide **5611** that comprises a complementary sequence **5615** (e.g., rGrGrG corresponding to C C C), a barcode sequence **5613** and, such as, for example, a UMI **5614** as shown in FIG. **56**C. In some cases, oligonucleotide **5611** may at first be associated with a bead and released from the bead. In any case, though, oligonucleotide **5611** can hybridize with oligonucleotide **5601** of the MHC-oligonucleotide complex **5608**. The hybridized oligonucleotides **5611** and **5601** can then be extended in primer extension reactions such that constructs comprising sequences that correspond to each of the two barcode sequences **5613** and **5604** are generated. In some cases, one or both of these corresponding sequences may be

a complement of the original sequence in oligonucleotide **5611** or **5601**. One or both of the resulting constructs can be optionally further processed (e.g., to add any additional sequences and/or for clean-up) and subjected to sequencing. As described elsewhere herein, the sequence in such a construct derived from barcode sequence **5613** may be used to identify a partition or a cell within a partition and the sequence derived from barcode sequence **5604** may be used to identify the particular TCR on the surface of the cell, permitting a multi-assay analysis.

[0310] Furthermore, while the example shown in FIG. **56**B and FIG. **56**C shows streptavidin directly coupled to its oligonucleotide, the streptavidin may also be coupled to a hybridization oligonucleotide which then hybridizes with the identifying barcoded oligonucleotide, similar to the example scheme shown in FIG. **52**B (panel II) and described elsewhere herein.

[0311] The ability to characterize individual cells from larger diverse populations of cells is also of significant value in both environmental testing as well as in forensic analysis, where samples may, by their nature, be made up of diverse populations of cells and other material that "contaminate" the sample, relative to the cells for which the sample is being tested, e.g., environmental indicator organisms, toxic organisms, and the like for, e.g., environmental and food safety testing, victim and/or perpetrator cells in forensic analysis for sexual assault, and other violent crimes, and the like.

[0312] Additional useful applications of the above described single cell sequencing and characterization processes are in the field of neuroscience research and diagnosis. In particular, neural cells can include long interspersed nuclear elements (LINEs), or 'jumping' genes that can move around the genome, which cause each neuron to differ from its neighbor cells. Research has shown that the number of LINEs in human brain exceeds that of other tissues, e.g., heart and liver tissue, with between 80 and 300 unique insertions (See, e.g., Coufal, N. G. et al. Nature 460, 1127-1131 (2009)). These differences have been postulated as being related to a person's susceptibility to neuro-logical disorders (see, e.g., Muotri, A. R. et al. Nature 468, 443-446 (2010)), or provide the brain with a diversity with which to respond to challenges. As such, the methods described herein may be used in the sequencing and characterization of individual neural cells.

[0313] The single cell analysis methods described herein may also be useful in the analysis of gene expression, both in terms of identification of RNA transcripts and their quantitation. In particular, using the single cell level analysis methods described herein, one can isolate and analyze the RNA transcripts present in individual cells, populations of cells, or subsets of populations of cells. In particular, in some cases, the barcode oligonucleotides may be configured to prime, replicate and consequently yield barcoded fragments of RNA from individual cells. For example, in some cases, the barcode oligonucleotides may include mRNA specific priming sequences, e.g., poly-T primer segments that allow priming and replication of mRNA in a reverse transcription reaction or other targeted priming sequences. Alternatively or additionally, random RNA priming may be carried out using random N-mer primer segments of the barcode oligonucleotides.

[0314] FIG. **6** provides a schematic of one example method for RNA expression analysis in individual cells using the methods described herein. As shown, at operation **602** a cell containing sample is sorted for viable cells, which are quantified and diluted for subsequent partitioning. At operation **604**, the individual cells separately co-partitioned with gel beads bearing the barcoding oligonucleotides as described herein. The cells are lysed and the barcoded oligonucleotides released into the partitions at operation **606**, where they interact with and hybridize to the mRNA at operation **608**, e.g., by virtue of a poly-T primer sequence, which is complementary to the poly-A tail of the mRNA. Using the poly-T barcode oligonucleotide as a priming sequence, a reverse transcription reaction is carried out at operation **610** to synthesize a cDNA of the mRNA that includes the barcode sequence. The barcoded cDNAs are then subjected to additional amplification at operation **612**, e.g., using a PCR process, purification at operation **614**, before they are placed on a nucleic acid sequencing system for determination of the cDNA sequence and its associated barcode sequence(s). In some cases, as shown, operations **602** through **608** can occur while the reagents remain in their original droplet or partition, while operations **612** through **616** can occur in bulk (e.g., outside of the partition). In the case where a partition is a droplet in an emulsion, the emulsion can be broken and the contents of the droplet pooled in order to complete operations **612** through **616**. In some cases, barcode oligonucleotides may be digested with exonucleases after the emulsion is broken. Exonuclease activity can be inhibited by ethylenediaminetetraacetic acid (EDTA) following primer digestion. In some cases, operation **610** may be performed either within the partitions based upon co-partitioning of the reverse transcription mixture, e.g., reverse transcriptase and associated reagents, or it may be performed in bulk.

[0315] The structure of the barcode oligonucleotides may include a number of sequence elements in addition to the oligonucleotide barcode sequence. One example of a barcode oligonucleotide for use in RNA analysis as described above is shown in FIG. **7**. As shown, the overall oligonucleotide **702** is coupled to a bead **704** by a releasable linkage **706**, such as a disulfide linker. The oligonucleotide may include functional sequences that are used in subsequent processing, such as functional sequence **708**, which may include one or more of a sequencer specific flow cell attachment sequence, e.g., a P5 sequence for Illumina sequencing systems, as well as sequencing primer sequences, e.g., a R1 primer for Illumina sequencing systems. A barcode sequence **710** is included within the structure for use in barcoding the sample RNA. An mRNA specific priming sequence, such as poly-T sequence **712** is also included in the oligonucleotide structure. An anchoring sequence segment **714** may be included to ensure that the poly-T sequence hybridizes at the sequence end of the mRNA. This anchoring sequence can include a random short sequence of nucleotides, e.g., 1-mer, 2-mer, 3-mer or longer sequence, which will ensure that the poly-T segment is more likely to hybridize at the sequence end of the poly-A tail of the mRNA. An additional sequence segment **716** may be provided within the oligonucleotide sequence. In some cases, this additional sequence provides a unique molecular identifier (UMI) sequence segment, e.g., as a random sequence (e.g., such as a random N-mer sequence) that varies across individual oligonucleotides coupled to a single bead, whereas barcode sequence **710** can be constant among oligonucleotides tethered to an individual bead. This unique sequence serves to provide a unique identifier of the starting

mRNA molecule that was captured, in order to allow quantitation of the number of original expressed RNA. As will be appreciated, although shown as a single oligonucleotide tethered to the surface of a bead, individual bead can include tens to hundreds of thousands or millions of individual oligonucleotide molecules (e.g., at least about 10,000, 50,000, 100,000, 500,000, 1,000,000 or 10,000,000 oligonucleotide molecules), where the barcode segment can be constant or relatively constant for a given bead, but where the variable or unique sequence segment will vary across an individual bead. This unique molecular identifier (UMI) sequence segment may include from 5 to about 8 or more nucleotides within the sequence of the oligonucleotides. In some cases, the unique molecular identifier (UMI) sequence segment can be 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 nucleotides in length or longer. In some cases, the unique molecular identifier (UMI) sequence segment can be at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 nucleotides in length or longer. In some cases, the unique molecular identifier (UMI) sequence segment can be at most 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 nucleotides in length or shorter. In some cases, the oligonucleotide may comprise a target-specific primer. The target-specific primer may bind to specific sequence in a RNA molecule or a DNA molecule derived therefrom. For example, the specific sequence may be a sequence that is not in the poly-A tail.

[0316] In operation, and with reference to FIGS. **6** and **7**, a cell is co-partitioned along with a barcode bearing bead and lysed while the barcoded oligonucleotides are released from the bead. The poly-T portion of the released barcode oligonucleotide then hybridizes to the poly-A tail of the mRNA. The poly-T segment then primes the reverse transcription of the mRNA to produce a cDNA of the mRNA, but which includes each of the sequence segments **708-716** of the barcode oligonucleotide. Again, because the oligonucleotide **702** includes an anchoring sequence **714**, it will more likely hybridize to and prime reverse transcription at the sequence end of the poly-A tail of the mRNA. Within any given partition, all of the cDNA of the individual mRNA molecules will include a common barcode sequence segment **710**. However, by including the unique random N-mer sequence, the transcripts made from different mRNA molecules within a given partition will vary at this unique sequence. This provides a quantitation feature that can be identifiable even following any subsequent amplification of the contents of a given partition, e.g., the number of unique segments associated with a common barcode can be indicative of the quantity of mRNA originating from a single partition, and thus, a single cell. The transcripts may then be amplified, cleaned up and sequenced to identify the sequence of the cDNA of the mRNA, as well as to sequence the barcode segment and the unique sequence segment.

[0317] While a poly-T primer sequence is described, other targeted or random priming sequences may also be used in priming the reverse transcription reaction. Likewise, although described as releasing the barcoded oligonucleotides into the partition along with the contents of the lysed cells, it will be appreciated that in some cases, the gel bead bound oligonucleotides may be used to hybridize and capture the mRNA on the solid phase of the gel beads, in order to facilitate the separation of the RNA from other cell contents.

[0318] An additional example of a barcode oligonucleotide for use in RNA analysis, including messenger RNA (mRNA, including mRNA obtained from a cell) analysis, is shown in FIG. **9**A. As shown, the overall oligonucleotide **902** can be coupled to a bead **904** by a releasable linkage **906**, such as a disulfide linker. The oligonucleotide may include functional sequences that are used in subsequent processing, such as functional sequence **908**, which may include a sequencer specific flow cell attachment sequence, e.g., a P5 sequence for Illumina sequencing systems, as well as functional sequence **910**, which may include sequencing primer sequences, e.g., a R1 primer binding site for Illumina sequencing systems. A barcode sequence **912** is included within the structure for use in barcoding the sample RNA. An RNA specific (e.g., mRNA specific) priming sequence, such as poly-T sequence **914** is also included in the oligonucleotide structure. An anchoring sequence segment (not shown) may be included to ensure that the poly-T sequence hybridizes at the sequence end of the mRNA. An additional sequence segment **916** may be provided within the oligonucleotide sequence. This additional sequence can provide a unique molecular identifier (UMI) sequence segment, e.g., as a random N-mer sequence that varies across individual oligonucleotides coupled to a single bead, whereas barcode sequence **912** can be constant among oligonucleotides tethered to an individual bead. As described elsewhere herein, this unique sequence can serve to provide a unique identifier of the starting mRNA molecule that was captured, in order to allow quantitation of the number of original expressed RNA, e.g., mRNA counting. As will be appreciated, although shown as a single oligonucleotide tethered to the surface of a bead, individual beads can include tens to hundreds of thousands or millions of individual oligonucleotide molecules (e.g., at least about 10,000, 50,000, 100,000, 500,000, 1,000,000 or 10,000,000 oligonucleotide molecules), where the barcode segment can be constant or relatively constant for a given bead, but where the variable or unique sequence segment will vary across an individual bead.

[0319] In an example method of cellular RNA (e.g., mRNA) analysis and in reference to FIG. **9**A, a cell is co-partitioned along with a barcode bearing bead, switch oligo **924**, and other reagents such as reverse transcriptase, a reducing agent and dNTPs into a partition (e.g., a droplet in an emulsion). In operation **950**, the cell is lysed while the barcoded oligonucleotides **902** are released from the bead (e.g., via the action of the reducing agent) and the poly-T segment **914** of the released barcode oligonucleotide then hybridizes to the poly-A tail of mRNA **920** that is released from the cell. Next, in operation **952** the poly-T segment **914** is extended in a reverse transcription reaction using the mRNA as a template to produce a cDNA **922** complementary to the mRNA and also includes each of the sequence segments **908**, **912**, **910**, **916** and **914** of the barcode oligonucleotide. Terminal transferase activity of the reverse transcriptase can add additional bases to the cDNA (e.g., polyC). The switch oligo **924** may then hybridize with the additional bases added to the cDNA and facilitate template switching. A sequence complementary to the switch oligo sequence can then be incorporated into the cDNA **922** via extension of the cDNA **922** using the switch oligo **924** as a template. Within any given partition, all of the cDNAs of the individual mRNA molecules will include a common barcode sequence segment **912**. However, by including the unique

random N-mer sequence **916**, the transcripts made from different mRNA molecules within a given partition will vary at this unique sequence. As described elsewhere herein, this provides a quantitation feature that can be identifiable even following any subsequent amplification of the contents of a given partition, e.g., the number of unique segments associated with a common barcode can be indicative of the quantity of mRNA originating from a single partition, and thus, a single cell. Following operation **952**, the cDNA **922** is then amplified with primers **926** (e.g., PCR primers) in operation **954**. Next, the amplified product is then purified (e.g., via solid phase reversible immobilization (SPRI)) in operation **956**. At operation **958**, the amplified product is then sheared, ligated to additional functional sequences, and further amplified (e.g., via PCR). The functional sequences may include a sequencer specific flow cell attachment sequence **930**, e.g., a P7 sequence for Illumina sequencing systems, as well as functional sequence **928**, which may include a sequencing primer binding site, e.g., for a R2 primer for Illumina sequencing systems, as well as functional sequence **932**, which may include a sample index, e.g., an i7 sample index sequence for Illumina sequencing systems. In some cases, operations **950** and **952** can occur in the partition, while operations **954**, **956** and **958** can occur in bulk solution (e.g., in a pooled mixture outside of the partition). In the case where a partition is a droplet in an emulsion, the emulsion can be broken and the contents of the droplet pooled in order to complete operations **954**, **956** and **958**. In some cases, operation **954** may be completed in the partition. In some cases, barcode oligonucleotides may be digested with exonucleases after the emulsion is broken. Exonuclease activity can be inhibited by ethylenediaminetetraacetic acid (EDTA) following primer digestion. Although described in terms of specific sequence references used for certain sequencing systems, e.g., Illumina systems, it will be understood that the reference to these sequences is for illustration purposes only, and the methods described herein may be configured for use with other sequencing systems incorporating specific priming, attachment, index, and other operational sequences used in those systems, e.g., systems available from Ion Torrent, Oxford Nanopore, Genia, Pacific Biosciences, Complete Genomics, and the like.

[0320] In an alternative example of a barcode oligonucleotide for use in RNA (e.g., cellular RNA) analysis as shown in FIG. **9**A, functional sequence **908** may be a P7 sequence and functional sequence **910** may be a R2 primer binding site. Moreover, the functional sequence **930** may be a P5 sequence, functional sequence **928** may be a R1 primer binding site, and functional sequence **932** may be an i5 sample index sequence for Illumina sequencing systems. The configuration of the constructs generated by such a barcode oligonucleotide can help minimize (or avoid) sequencing of the poly-T sequence during sequencing.

[0321] Shown in FIG. **9**B is another example method for RNA analysis, including cellular mRNA analysis. In this method, the switch oligo **924** is co-partitioned with the individual cell and barcoded bead along with reagents such as reverse transcriptase, a reducing agent and dNTPs into a partition (e.g., a droplet in an emulsion). The switch oligo **924** may be labeled with an additional tag **934**, e.g., biotin. In operation **951**, the cell is lysed while the barcoded oligonucleotides **902** (e.g., as shown in FIG. **9**A) are released from the bead (e.g., via the action of the reducing agent). In some cases, sequence **908** is a P7 sequence and sequence **910** is a R2 primer binding site. In other cases, sequence **908** is a P5 sequence and sequence **910** is a R1 primer binding site. Next, the poly-T segment **914** of the released barcode oligonucleotide hybridizes to the poly-A tail of mRNA **920** that is released from the cell. In operation **953**, the poly-T segment **914** is then extended in a reverse transcription reaction using the mRNA as a template to produce a cDNA **922** complementary to the mRNA and also includes each of the sequence segments **908**, **912**, **910**, **916** and **914** of the barcode oligonucleotide. Terminal transferase activity of the reverse transcriptase can add additional bases to the cDNA (e.g., polyC). The switch oligo **924** may then hybridize with the cDNA and facilitate template switching. A sequence complementary to the switch oligo sequence can then be incorporated into the cDNA **922** via extension of the cDNA **922** using the switch oligo **924** as a template. Next, an isolation operation **960** can be used to isolate the cDNA **922** from the reagents and oligonucleotides in the partition. The additional tag **934**, e.g., biotin, can be contacted with an interacting tag **936**, e.g., streptavidin, which may be attached to a magnetic bead **938**. At operation **960** the cDNA can be isolated with a pull-down operation (e.g., via magnetic separation, centrifugation) before amplification (e.g., via PCR) in operation **955**, followed by purification (e.g., via solid phase reversible immobilization (SPRI)) in operation **957** and further processing (shearing, ligation of sequences **928**, **932** and **930** and subsequent amplification (e.g., via PCR)) in operation **959**. In some cases where sequence **908** is a P7 sequence and sequence **910** is a R2 primer binding site, sequence **930** is a P5 sequence and sequence **928** is a R1 primer binding site and sequence **932** is an i5 sample index sequence. In some cases where sequence **908** is a P5 sequence and sequence **910** is a R1 primer binding site, sequence **930** is a P7 sequence and sequence **928** is a R2 primer binding site and sequence **932** is an i7 sample index sequence. In some cases, as shown, operations **951** and **953** can occur in the partition, while operations **960**, **955**, **957** and **959** can occur in bulk solution (e.g., in a pooled mixture outside of the partition). In the case where a partition is a droplet in an emulsion, the emulsion can be broken and the contents of the droplet pooled in order to complete operation **960**. The operations **955**, **957**, and **959** can then be carried out following operation **960** after the transcripts are pooled for processing.

[0322] Shown in FIG. **9**C is another example method for RNA analysis, including cellular mRNA analysis. In this method, the switch oligo **924** is co-partitioned with the individual cell and barcoded bead along with reagents such as reverse transcriptase, a reducing agent and dNTPs in a partition (e.g., a droplet in an emulsion). In operation **961**, the cell is lysed while the barcoded oligonucleotides **902** (e.g., as shown in FIG. **9**A) are released from the bead (e.g., via the action of the reducing agent). In some cases, sequence **908** is a P7 sequence and sequence **910** is a R2 primer binding site. In other cases, sequence **908** is a P5 sequence and sequence **910** is a R1 primer binding site. Next, the poly-T segment **914** of the released barcode oligonucleotide then hybridizes to the poly-A tail of mRNA **920** that is released from the cell. Next, in operation **963** the poly-T segment **914** is then extended in a reverse transcription reaction using the mRNA as a template to produce a cDNA **922** complementary to the mRNA and also includes each of the sequence segments **908**, **912**, **910**, **916** and **914**

of the barcode oligonucleotide. Terminal transferase activity of the reverse transcriptase can add additional bases to the cDNA (e.g., polyC). The switch oligo **924** may then hybridize with the cDNA and facilitate template switching. A sequence complementary to the switch oligo sequence can then be incorporated into the cDNA **922** via extension of the cDNA **922** using the switch oligo **924** as a template. Following operation **961** and operation **963**, mRNA **920** and cDNA **922** are denatured in operation **962**. At operation **964**, a second strand is extended from a primer **940** having an additional tag **942**, e.g., biotin, and hybridized to the cDNA **922**. Also in operation **964**, the biotin labeled second strand can be contacted with an interacting tag **936**, e.g., streptavidin, which may be attached to a magnetic bead **938**. The cDNA can be isolated with a pull-down operation (e.g., via magnetic separation, centrifugation) before amplification (e.g., via polymerase chain reaction (PCR)) in operation **965**, followed by purification (e.g., via solid phase reversible immobilization (SPRI)) in operation **967** and further processing (shearing, ligation of sequences **928**, **932** and **930** and subsequent amplification (e.g., via PCR)) in operation **969**. In some cases where sequence **908** is a P7 sequence and sequence **910** is a R2 primer binding site, sequence **930** is a P5 sequence and sequence **928** is a R1 primer binding site and sequence **932** is an i5 sample index sequence. In some cases where sequence **908** is a P5 sequence and sequence **910** is a R1 primer binding site, sequence **930** is a P7 sequence and sequence **928** is a R2 primer binding site and sequence **932** is an i7 sample index sequence. In some cases, operations **961** and **963** can occur in the partition, while operations **962**, **964**, **965**, **967**, and **969** can occur in bulk (e.g., outside the partition). In the case where a partition is a droplet in an emulsion, the emulsion can be broken and the contents of the droplet pooled in order to complete operations **962**, **964**, **965**, **967** and **969**.

[0323] Shown in FIG. **9**D is another example method for RNA analysis, including cellular mRNA analysis. In this method, the switch oligo **924** is co-partitioned with the individual cell and barcoded bead along with reagents such as reverse transcriptase, a reducing agent and dNTPs. In operation **971**, the cell is lysed while the barcoded oligonucleotides **902** (e.g., as shown in FIG. **9**A) are released from the bead (e.g., via the action of the reducing agent). In some cases, sequence **908** is a P7 sequence and sequence **910** is a R2 primer binding site. In other cases, sequence **908** is a P5 sequence and sequence **910** is a R1 primer binding site. Next the poly-T segment **914** of the released barcode oligonucleotide then hybridizes to the poly-A tail of mRNA **920** that is released from the cell. Next in operation **973**, the poly-T segment **914** is then extended in a reverse transcription reaction using the mRNA as a template to produce a cDNA **922** complementary to the mRNA and also includes each of the sequence segments **908**, **912**, **910**, **916** and **914** of the barcode oligonucleotide. Terminal transferase activity of the reverse transcriptase can add additional bases to the cDNA (e.g., polyC). The switch oligo **924** may then hybridize with the cDNA and facilitate template switching. A sequence complementary to the switch oligo sequence can then be incorporated into the cDNA **922** via extension of the cDNA **922** using the switch oligo **924** as a template. In operation **966**, the mRNA **920**, cDNA **922** and switch oligo **924** can be denatured, and the cDNA **922** can be hybridized with a capture oligonucleotide **944** labeled with an additional tag **946**, e.g., biotin. In this operation, the biotin-labeled capture oligonucleotide **944**, which is hybridized to the cDNA, can be contacted with an interacting tag **936**, e.g., streptavidin, which may be attached to a magnetic bead **938**. Following separation from other species (e.g., excess barcoded oligonucleotides) using a pull-down operation (e.g., via magnetic separation, centrifugation), the cDNA can be amplified (e.g., via PCR) with primers **926** at operation **975**, followed by purification (e.g., via solid phase reversible immobilization (SPRI)) in operation **977** and further processing (shearing, ligation of sequences **928**, **932** and **930** and subsequent amplification (e.g., via PCR)) in operation **979**. In some cases where sequence **908** is a P7 sequence and sequence **910** is a R2 primer binding site, sequence **930** is a P5 sequence and sequence **928** is a R1 primer binding site and sequence **932** is an i5 sample index sequence. In other cases where sequence **908** is a P5 sequence and sequence **910** is a R1 primer binding site, sequence **930** is a P7 sequence and sequence **928** is a R2 primer binding site and sequence **932** is an i7 sample index sequence. In some cases, operations **971** and **973** can occur in the partition, while operations **966**, **975**, **977** (purification), and **979** can occur in bulk (e.g., outside the partition). In the case where a partition is a droplet in an emulsion, the emulsion can be broken and the contents of the droplet pooled in order to complete operations **966**, **975**, **977** and **979**.

[0324] Shown in FIG. **9**E is another example method for RNA analysis, including cellular RNA analysis. In this method, an individual cell is co-partitioned along with a barcode bearing bead, a switch oligo **990**, and other reagents such as reverse transcriptase, a reducing agent and dNTPs into a partition (e.g., a droplet in an emulsion). In operation **981**, the cell is lysed while the barcoded oligonucleotides (e.g., **902** as shown in FIG. **9**A) are released from the bead (e.g., via the action of the reducing agent). In some cases, sequence **908** is a P7 sequence and sequence **910** is a R2 primer binding site. In other cases, sequence **908** is a P5 sequence and sequence **910** is a R1 primer binding site. Next, the poly-T segment of the released barcode oligonucleotide then hybridizes to the poly-A tail of mRNA **920** released from the cell. Next at operation **983**, the poly-T segment is then extended in a reverse transcription reaction to produce a cDNA **922** complementary to the mRNA and also includes each of the sequence segments **908**, **912**, **910**, **916** and **914** of the barcode oligonucleotide. Terminal transferase activity of the reverse transcriptase can add additional bases to the cDNA (e.g., polyC). The switch oligo **990** may then hybridize with the cDNA and facilitate template switching. A sequence complementary to the switch oligo sequence and including a T7 promoter sequence, can be incorporated into the cDNA **922**. At operation **968**, a second strand is synthesized and at operation **970** the T7 promoter sequence can be used by T7 polymerase to produce RNA transcripts in in vitro transcription. At operation **985** the RNA transcripts can be purified (e.g., via solid phase reversible immobilization (SPRI)), reverse transcribed to form DNA transcripts, and a second strand can be synthesized for each of the DNA transcripts. In some cases, prior to purification, the RNA transcripts can be contacted with a DNase (e.g., DNAase I) to break down residual DNA. At operation **987** the DNA transcripts are then fragmented and ligated to additional functional sequences, such as sequences **928**, **932** and **930** and, in some cases, further amplified (e.g., via PCR). In some cases where sequence **908** is a P7 sequence and sequence **910** is a R2 primer binding site, sequence **930**

is a P5 sequence and sequence **928** is a R1 primer binding site and sequence **932** is an i5 sample index sequence. In some cases where sequence **908** is a P5 sequence and sequence **910** is a R1 primer binding site, sequence **930** is a P7 sequence and sequence **928** is a R2 primer binding site and sequence **932** is an i7 sample index sequence. In some cases, prior to removing a portion of the DNA transcripts, the DNA transcripts can be contacted with an RNase to break down residual RNA. In some cases, operations **981** and **983** can occur in the partition, while operations **968**, **970**, **985** and **987** can occur in bulk (e.g., outside the partition). In the case where a partition is a droplet in an emulsion, the emulsion can be broken and the contents of the droplet pooled in order to complete operations **968**, **970**, **985** and **987**.

[0325] Another example of a barcode oligonucleotide for use in RNA analysis, including messenger RNA (mRNA, including mRNA obtained from a cell) analysis is shown in FIG. **10**. As shown, the overall oligonucleotide **1002** is coupled to a bead **1004** by a releasable linkage **1006**, such as a disulfide linker. The oligonucleotide may include functional sequences that are used in subsequent processing, such as functional sequence **1008**, which may include a sequencer specific flow cell attachment sequence, e.g., a P7 sequence, as well as functional sequence **1010**, which may include sequencing primer sequences, e.g., a R2 primer binding site. A barcode sequence **1012** is included within the structure for use in barcoding the sample RNA. An RNA specific (e.g., mRNA specific) priming sequence, such as poly-T sequence **1014** may be included in the oligonucleotide structure. An anchoring sequence segment (not shown) may be included to ensure that the poly-T sequence hybridizes at the sequence end of the mRNA. An additional sequence segment **1016** may be provided within the oligonucleotide sequence. This additional sequence can provide a unique molecular identifier (UMI) sequence segment, as described elsewhere herein. An additional functional sequence **1020** may be included for in vitro transcription, e.g., a T7 RNA polymerase promoter sequence. As will be appreciated, although shown as a single oligonucleotide tethered to the surface of a bead, individual beads can include tens to hundreds of thousands or millions of individual oligonucleotide molecules (e.g., at least about 10,000, 50,000, 100,000, 500,000, 1,000,000 or 10,000,000 oligonucleotide molecules), where the barcode segment can be constant or relatively constant for a given bead, but where the variable or unique sequence segment will vary across an individual bead.

[0326] In an example method of cellular RNA analysis and in reference to FIG. **10**, a cell is co-partitioned along with a barcode bearing bead, and other reagents such as reverse transcriptase, reducing agent and dNTPs into a partition (e.g., a droplet in an emulsion). In operation **1050**, the cell is lysed while the barcoded oligonucleotides **1002** are released (e.g., via the action of the reducing agent) from the bead, and the poly-T segment **1014** of the released barcode oligonucleotide then hybridizes to the poly-A tail of mRNA **1020**. Next at operation **1052**, the poly-T segment is then extended in a reverse transcription reaction using the mRNA as template to produce a cDNA **1022** of the mRNA and also includes each of the sequence segments **1020**, **1008**, **1012**, **1010**, **1016**, and **1014** of the barcode oligonucleotide. Within any given partition, all of the cDNAs of the individual mRNA molecules will include a common barcode sequence segment **1012**. However, by including the unique random N-mer sequence, the transcripts made from different mRNA molecules within a given partition will vary at this unique sequence. As described elsewhere herein, this provides a quantitation feature that can be identifiable even following any subsequent amplification of the contents of a given partition, e.g., the number of unique segments associated with a common barcode can be indicative of the quantity of mRNA originating from a single partition, and thus, a single cell. At operation **1054** a second strand is synthesized and at operation **1056** the T7 promoter sequence can be used by T7 polymerase to produce RNA transcripts in in vitro transcription. At operation **1058** the transcripts are fragmented (e.g., sheared), ligated to additional functional sequences, and reverse transcribed. The functional sequences may include a sequencer specific flow cell attachment sequence **1030**, e.g., a P5 sequence, as well as functional sequence **1028**, which may include sequencing primers, e.g., a R1 primer binding sequence, as well as functional sequence **1032**, which may include a sample index, e.g., an i5 sample index sequence. At operation **1060** the RNA transcripts can be reverse transcribed to DNA, the DNA amplified (e.g., via PCR), and sequenced to identify the sequence of the cDNA of the mRNA, as well as to sequence the barcode segment and the unique sequence segment. In some cases, operations **1050** and **1052** can occur in the partition, while operations **1054**, **1056**, **1058** and **1060** can occur in bulk (e.g., outside the partition). In the case where a partition is a droplet in an emulsion, the emulsion can be broken and the contents of the droplet pooled in order to complete operations **1054**, **1056**, **1058** and **1060**.

[0327] In an alternative example of a barcode oligonucleotide for use in RNA (e.g., cellular RNA) analysis as shown in FIG. **10**, functional sequence **1008** may be a P5 sequence and functional sequence **1010** may be a R1 primer binding site. Moreover, the functional sequence **1030** may be a P7 sequence, functional sequence **1028** may be a R2 primer binding site, and functional sequence **1032** may be an i7 sample index sequence.

[0328] An additional example of a barcode oligonucleotide for use in RNA analysis, including messenger RNA (mRNA, including mRNA obtained from a cell) analysis is shown in FIG. **11**. As shown, the overall oligonucleotide **1102** is coupled to a bead **1104** by a releasable linkage **1106**, such as a disulfide linker. The oligonucleotide may include functional sequences that are used in subsequent processing, such as functional sequence **1108**, which may include a sequencer specific flow cell attachment sequence, e.g., a P5 sequence, as well as functional sequence **1110**, which may include sequencing primer sequences, e.g., a R1 primer binding site. In some cases, sequence **1108** is a P7 sequence and sequence **1110** is a R2 primer binding site. A barcode sequence **1112** is included within the structure for use in barcoding the sample RNA. An additional sequence segment **1116** may be provided within the oligonucleotide sequence. In some cases, this additional sequence can provide a unique molecular identifier (UMI) sequence segment, as described elsewhere herein. An additional sequence **1114** may be included to facilitate template switching, e.g., polyG. As will be appreciated, although shown as a single oligonucleotide tethered to the surface of a bead, individual beads can include tens to hundreds of thousands or millions of individual oligonucleotide molecules (e.g., at least about 10,000, 50,000, 100,000, 500,000, 1,000,000 or 10,000,000 oligo-

nucleotide molecules), where the barcode segment can be constant or relatively constant for a given bead, but where the variable or unique sequence segment will vary across an individual bead.

[0329] In an example method of cellular mRNA analysis and in reference to FIG. **11**, a cell is co-partitioned along with a microcapsule (e.g., bead bearing a barcoded oligonucleotide), polyT sequence, and other reagents such as a DNA polymerase, a reverse transcriptase, oligonucleotide primers, dNTPs, and reducing agent into a partition (e.g., a droplet in an emulsion). The partition can serve as a reaction volume. As described elsewhere herein, the partition serving as the reaction volume can comprise a container or vessel such as a well, a microwell, vial, a tube, through ports in nanoarray substrates, or micro-vesicles having an outer barrier surrounding an inner fluid center or core, emulsion, or a droplet. In some embodiments, the partition comprises a droplet of aqueous fluid within a non-aqueous continuous phase, e.g., an oil phase. Within the partition, the cell can be lysed and the barcoded oligonucleotides can be released from the bead (e.g., via the action of the reducing agent or other stimulus). Cell lysis and release of the barcoded oligonucleotides from the microcapsule may occur simultaneously in the partition (e.g., a droplet in an emulsion) or the reaction volume. In some embodiments, cell lysis precedes release of the barcoded oligonucleotides from the microcapsule. In some embodiments, release of the barcoded oligonucleotides from the microcapsule precedes cell lysis.

[0330] Subsequent to cell lysis and the release of barcoded oligonucleotides from the microcapsule, the reaction volume can be subjected to an amplification reaction to generate an amplification product. In an example amplification reaction, the polyT sequence hybridizes to the polyA tail of mRNA **1120** released from the cell as illustrated in operation **1150**. Next, in operation **1152**, the polyT sequence is then extended in a reverse transcription reaction using the mRNA as a template to produce a cDNA **1122** complementary to the mRNA. Terminal transferase activity of the reverse transcriptase can add additional bases to the cDNA (e.g., polyC) in a template independent manner. The additional bases added to the cDNA, e.g., polyC, can then hybridize with **1114** of the barcoded oligonucleotide. This can facilitate template switching and a sequence complementary to the barcoded oligonucleotide can be incorporated into the cDNA. In various embodiments, the barcoded oligonucleotide does not hybridize to the template polynucleotide.

[0331] The barcoded oligonucleotide, upon release from the microcapsule, can be present in the reaction volume at any suitable concentration. In some embodiments, the barcoded oligonucleotide is present in the reaction volume at a concentration of about 0.2 μM, 0.3 μM, 0.4 μM, 0.5 μM, 1 μM, 5 μM, 10 μM, 15 μM, 20 μM, 25 μM, 30 μM, 35 μM, 40 μM, 50 μM, 100 μM, 150 μM, 200 μM, 250 μM, 300 μM, 400 μM, or 500 μM. In some embodiments, the barcoded oligonucleotide is present in the reaction volume at a concentration of at least about 0.2 μM, 0.3 μM, 0.4 μM, 0.5 M, 1 μM, 5 μM, 10 μM, 15 μM, 20 μM, 25 μM, 30 μM, 35 μM, 40 μM, 50 μM, 100 μM, 150 μM, 200 μM, 250 μM, 300 μM, 400 μM, 500 μM or greater. In some embodiments, the barcoded oligonucleotide is present in the reaction volume at a concentration of at most about 0.2 M, 0.3 μM, 0.4 μM, 0.5 μM, 1 μM, 5 μM, 10 μM, 15 μM, 20 μM, 25 μM, 30 μM, 35 μM, 40 μM, 50 μM, 100 μM, 150 μM, 200 μM, 250 μM, 300 μM, 400 μM, or 500 μM.

[0332] The transcripts can be further processed (e.g., amplified, portions removed, additional sequences added, etc.) and characterized as described elsewhere herein. In some embodiments, the transcripts are sequenced directly. In some embodiments, the transcripts are further processed (e.g., portions removed, additional sequences added, etc) and then sequenced. In some embodiments, the reaction volume is subjected to a second amplification reaction to generate an additional amplification product. The transcripts or first amplification products can be used as the template for the second amplification reaction. In some embodiments, primers for the second amplification reaction comprise the barcoded oligonucleotide and polyT sequence. In some embodiments, primers for the second amplification reaction comprise additional primers co-partitioned with the cell. In some embodiments, these additional amplification products are sequenced directly. In some embodiments, these additional amplification products are further processed (e.g., portions removed, additional sequences added, etc) and then sequenced. The configuration of the amplification products (e.g., first amplification products and second amplification products) generated by such a method can help minimize (or avoid) sequencing of the poly-T sequence during sequencing.

[0333] An additional example of a barcode oligonucleotide for use in RNA analysis, including cellular RNA analysis is shown in FIG. **12**A. As shown, the overall oligonucleotide **1202** is coupled to a bead **1204** by a releasable linkage **1206**, such as a disulfide linker. The oligonucleotide may include functional sequences that are used in subsequent processing, such as functional sequence **1208**, which may include a sequencer specific flow cell attachment sequence, e.g., a P5 sequence, as well as functional sequence **1210**, which may include sequencing primer sequences, e.g., a R1 primer binding site. In some cases, sequence **1208** is a P7 sequence and sequence **1210** is a R2 primer binding site. A barcode sequence **1212** is included within the structure for use in barcoding the sample RNA. An additional sequence segment **1216** may be provided within the oligonucleotide sequence. In some cases, this additional sequence can provide a unique molecular identifier (UMI) sequence segment, as described elsewhere herein. As will be appreciated, although shown as a single oligonucleotide tethered to the surface of a bead, individual beads can include tens to hundreds of thousands or millions of individual oligonucleotide molecules (e.g., at least about 10,000, 50,000, 100,000, 500,000, 1,000,000 or 10,000,000 oligonucleotide molecules), where the barcode segment can be constant or relatively constant for a given bead, but where the variable or unique sequence segment will vary across an individual bead. In an example method of cellular RNA analysis using this barcode, a cell is co-partitioned along with a barcode bearing bead and other reagents such as RNA ligase and a reducing agent into a partition (e.g., a droplet in an emulsion). The cell is lysed while the barcoded oligonucleotides are released (e.g., via the action of the reducing agent) from the bead. The barcoded oligonucleotides can then be ligated to the 5' end of mRNA transcripts while in the partitions by RNA ligase. Subsequent operations may include purification (e.g., via solid phase reversible immobilization (SPRI)) and further processing (shearing, ligation of functional sequences, and subsequent amplification (e.g., via PCR)), and these operations may occur in bulk (e.g., outside the partition). In the case where a partition is a droplet in an

emulsion, the emulsion can be broken and the contents of the droplet pooled for the additional operations.

[0334] An additional example of a barcode oligonucleotide for use in RNA analysis, including cellular RNA analysis is shown in FIG. **12**B. As shown, the overall oligonucleotide **1222** is coupled to a bead **1224** by a releasable linkage **1226**, such as a disulfide linker. The oligonucleotide may include functional sequences that are used in subsequent processing, such as functional sequence **1228**, which may include a sequencer specific flow cell attachment sequence, e.g., a P5 sequence, as well as functional sequence **1230**, which may include sequencing primer sequences, e.g., a R1 primer binding site. In some cases, sequence **1228** is a P7 sequence and sequence **1230** is a R2 primer binding site. A barcode sequence **1232** is included within the structure for use in barcoding the sample RNA. A priming sequence **1234** (e.g., a random priming sequence) can also be included in the oligonucleotide structure, e.g., a random hexamer. An additional sequence segment **1236** may be provided within the oligonucleotide sequence. In some cases, this additional sequence provides a unique molecular identifier (UMI) sequence segment, as described elsewhere herein. As will be appreciated, although shown as a single oligonucleotide tethered to the surface of a bead, individual beads can include tens to hundreds of thousands or millions of individual oligonucleotide molecules (e.g., at least about 10,000, 50,000, 100,000, 500,000, 1,000,000 or 10,000,000 oligonucleotide molecules), where the barcode segment can be constant or relatively constant for a given bead, but where the variable or unique sequence segment will vary across an individual bead. In an example method of cellular mRNA analysis using the barcode oligonucleotide of FIG. **12**B, a cell is co-partitioned along with a barcode bearing bead and additional reagents such as reverse transcriptase, a reducing agent and dNTPs into a partition (e.g., a droplet in an emulsion). The cell is lysed while the barcoded oligonucleotides are released from the bead (e.g., via the action of the reducing agent). In some cases, sequence **1228** is a P7 sequence and sequence **1230** is a R2 primer binding site. In other cases, sequence **1228** is a P5 sequence and sequence **1230** is a R1 primer binding site. The priming sequence **1234** of random hexamers can randomly hybridize cellular mRNA. The random hexamer sequence can then be extended in a reverse transcription reaction using mRNA from the cell as a template to produce a cDNA complementary to the mRNA and also includes each of the sequence segments **1228**, **1232**, **1230**, **1236**, and **1234** of the barcode oligonucleotide. Subsequent operations may include purification (e.g., via solid phase reversible immobilization (SPRI)), further processing (shearing, ligation of functional sequences, and subsequent amplification (e.g., via PCR)), and these operations may occur in bulk (e.g., outside the partition). In the case where a partition is a droplet in an emulsion, the emulsion can be broken and the contents of the droplet pooled for additional operations. Additional reagents that may be co-partitioned along with the barcode bearing bead may include oligonucleotides to block ribosomal RNA (rRNA) and nucleases to digest genomic DNA and cDNA from cells. Alternatively, rRNA removal agents may be applied during additional processing operations. The configuration of the constructs generated by such a method can help minimize (or avoid) sequencing of the poly-T sequence during sequencing.

[0335] The single cell analysis methods described herein may also be useful in the analysis of the whole transcriptome. Referring back to the barcode of FIG. **12**B, the priming sequence **1234** may be a random N-mer. In some cases, sequence **1228** is a P7 sequence and sequence **1230** is a R2 primer binding site. In other cases, sequence **1228** is a P5 sequence and sequence **1230** is a R1 primer binding site. In an example method of whole transcriptome analysis using this barcode, the individual cell is co-partitioned along with a barcode bearing bead, poly-T sequence, and other reagents such as reverse transcriptase, polymerase, a reducing agent and dNTPs into a partition (e.g., droplet in an emulsion). In an operation of this method, the cell is lysed while the barcoded oligonucleotides are released from the bead (e.g., via the action of the reducing agent) and the poly-T sequence hybridizes to the poly-A tail of cellular mRNA. In a reverse transcription reaction using the mRNA as template, cDNAs of cellular mRNA can be produced. The RNA can then be degraded with an RNase. The priming sequence **1234** in the barcoded oligonucleotide can then randomly hybridize to the cDNAs. The oligonucleotides can be extended using polymerase enzymes and other extension reagents co-partitioned with the bead and cell similar to as shown in FIG. **3** to generate amplification products (e.g., barcoded fragments), similar to the example amplification product shown in FIG. **3** (panel F). The barcoded nucleic acid fragments may, in some cases subjected to further processing (e.g., amplification, addition of additional sequences, clean up processes, etc. as described elsewhere herein) characterized, e.g., through sequence analysis. In this operation, sequencing signals can come from full length RNA.

[0336] In an example method, the barcode sequence can be appended to the 3' end of the template polynucleotide sequence (e.g., mRNA). Such configuration may be useful, for example, if the sequence the 3' end of the template polynucleotide is to be analyzed. In some embodiments, the barcode sequence can be appended to the 5' end of a template polynucleotide sequence (e.g., mRNA). Such configuration may be useful, for example, if the sequence at the 5' end of the template polynucleotide is to be analyzed.

[0337] In another aspect, a partition comprises a cell co-partitioned with a primer having a sequence towards a 3' end that hybridizes to the template polynucleotide, a template switching oligonucleotide having a first predefined sequence towards a 5' end, and a microcapsule, such as a bead, having barcoded oligonucleotides releasably coupled thereto. In some embodiments, the oligonucleotides coupled to the bead include barcode sequences that are identical (e.g., all oligonucleotides sharing the same barcode sequence). In some aspects, the oligonucleotides coupled to the beads additionally include unique molecular identifier (UMI) sequence segments (e.g., all oligonucleotides having different unique molecular identifier sequences).

[0338] FIG. **18** shows a barcoded oligonucleotide coupled to a bead. As shown, the overall oligonucleotide **1802** is coupled to a bead **1804** by a releasable linkage **1806**, such as a disulfide linker. The oligonucleotide may include functional sequences that are useful for subsequent processing, such as functional sequence **1808**, which may include a sequencer specific flow cell attachment sequence, e.g., a P5 sequence, as well as functional sequence **1810**, which may include sequencing primer sequences, e.g., a R1 primer binding site. In some cases, sequence **1808** is a P7 sequence and sequence **1810** is a R2 primer binding site. A barcode

sequence **1812** can be included within the structure for use in barcoding the template polynucleotide. The functional sequences may be selected for compatibility with a variety of different sequencing systems, e.g., 454 Sequencing, Ion Torrent Proton or PGM, Illumina X10, etc., and the requirements thereof. In some cases, the barcode sequence **1812**, functional sequences **1808** (e.g., flow cell attachment sequence) and **1810** (e.g., sequencing primer sequences) may be common to all of the oligonucleotides attached to a given bead. The barcoded oligonucleotide can also comprise a sequence **1816** to facilitate template switching (e.g., a polyG sequence). In some cases, the additional sequence provides a unique molecular identifier (UMI) sequence segment, as described elsewhere herein.

[0339] Although shown as a single oligonucleotide tethered to the surface of a bead, individual beads can include tens to hundreds of thousands or millions of individual oligonucleotide molecules (e.g., at least about 10,000, 50,000, 100,000, 500,000, 1,000,000 or 10,000,000 oligonucleotide molecules), where the barcode sequence can be constant or relatively constant for a given bead.

[0340] In an example method of cellular polynucleotide analysis using the barcode oligonucleotide of FIG. **18**, a cell is co-partitioned along with a bead bearing a barcoded oligonucleotide and additional reagents such as reverse transcriptase, primers, oligonucleotides (e.g., template switching oligonucleotides), dNTPs, and reducing agent into a partition (e.g., a droplet in an emulsion). Within the partition, the cell can be lysed to yield a plurality of template polynucleotides (e.g., DNA such as genomic DNA, RNA such as mRNA, etc). In some cases, the cell is lysed using lysis reagents that are co-partitioned with the cell.

[0341] Where the bead is a degradable or disruptable bead, the barcoded oligonucleotide can be released from the bead following the application of stimulus as previously described. Following release from the bead, the barcoded oligonucleotide can be present in the partition at any suitable concentration. In some embodiments, the barcoded oligonucleotide is present in the partition at a concentration that is suitable for generating a sufficient yield of amplification products for downstream processing and analysis, including, but not limited to, sequencing adaptor attachment and sequencing analysis. In some embodiments, the concentration of the barcoded oligonucleotide is limited by the loading capacity of the barcode bearing bead, or the amount of oligonucleotides deliverable by the bead.

[0342] The template switching oligonucleotide, which can be co-partitioned with the cell, bead bearing barcoded oligonucleotides, etc, can be present in the partition at any suitable concentration. In some embodiments, the template switching oligonucleotide is present in the partition at a concentration that is suitable for efficient template switching during an amplification reaction. The concentration of the template switching oligonucleotide can be dependent on the reagents used for droplet generation. In some embodiments, the template switching oligonucleotide is among a plurality of template switching oligonucleotides.

[0343] In some embodiments, the barcoded oligonucleotide and template switching oligonucleotide are present in the partition at similar concentrations. In some embodiments, the barcoded oligonucleotide and template switching oligonucleotides may be present in proportions reflective of the amount of amplification products to be generated using each oligonucleotide. In some embodiments, the template switching oligonucleotide is present in the partition at a greater concentration than the barcoded oligonucleotide. This difference in concentration can be due to limitations on the capacity of the barcode bearing bead. In some embodiments, the concentration of the template switching oligonucleotide in the reaction volume is at least 2, 5, 10, 20, 50, 100, or 200 times that of the concentration of the barcoded oligonucleotide in the same reaction volume when the barcoded oligonucleotide is free in the partition (e.g., not attached to the bead).

[0344] As illustrated in FIG. **19**, a reaction mixture comprising a template polynucleotide from a cell **1920** and (i) the primer **1924** having a sequence towards a 3' end that hybridizes to the template polynucleotide (e.g., polyT) and (ii) a template switching oligonucleotide **1926** that comprises a first predefined sequence **1810** towards a 5' end can be subjected to an amplification reaction to yield a first amplification product. In some cases, the template polynucleotide is an mRNA with a polyA tail and the primer that hybridizes to the template polynucleotide comprises a polyT sequence towards a 3' end, which is complementary to the polyA segment. The first predefined sequence can comprise at least one of an adaptor sequence, a barcode sequence, a unique molecular identifier (UMI) sequence, a primer binding site, and a sequencing primer binding site or any combination thereof. In some cases, the first predefined sequence **1810** is a sequence that can be common to all partitions of a plurality of partitions. For example, the first predefined sequence may comprise a flow cell attachment sequence, an amplification primer binding site, or a sequencing primer binding site and the first amplification reaction facilitates the attachment the predefined sequence to the template polynucleotide from the cell. In some embodiments, the first predefined sequence comprises a primer binding site. In some embodiments, the first predefined sequence comprises a sequencing primer binding site. As illustrated in operation **1950**, the sequence towards a 3' end (e.g., polyT) of the primer **1924** hybridizes to the template polynucleotide **1920**. In a first amplification reaction, extension reaction reagents, e.g., reverse transcriptase, nucleoside triphosphates, co-factors (e.g., Mg2+ or Mn2+), that are also co-partitioned, can extend the primer **1924** sequence using the cell's nucleic acid as a template, to produce a transcript, e.g., cDNA, **1922** having a fragment complementary to the strand of the cell's nucleic acid to which the primer annealed. In some cases, the reverse transcriptase has terminal transferase activity and the reverse transcriptase adds additional nucleotides, e.g., polyC, to the cDNA in a template independent manner. As illustrated in operation **1952**, the template switching oligonucleotide **1926**, for example a template switching oligonucleotide which includes a polyG sequence, can hybridize to the cDNA **1922** and facilitate template switching in the first amplification reaction. The transcript, therefore, may comprise the sequence of the primer **1924**, a sequence complementary to the template polynucleotide from the cell, and a sequence complementary to the template switching oligonucleotide.

[0345] Among a plurality of partitions, each partition containing one or more cells or no cells, the primer and template switching oligonucleotide may be universal to all partitions. Where analysis of mRNA is conducted, for example, the primer may comprise at least a polyT segment capable of hybridizing and priming an extension reaction from the polyA segment of an mRNA. Where analysis of a

variety of polynucleotides is conducted, the primer may comprise a random sequence capable of hybridizing to and priming extension reactions randomly on various polynucleotide templates. As template switching can occur with the use of an enzyme having terminal transferase activity, a template switching oligonucleotide having a sequence capable of hybridizing to the appended bases can be used for template switching in manner that is independent of the sequence of the polynucleotide templates to be analyzed. In some embodiments, the template switching oligonucleotide can comprise a first predefined sequence towards a 5' end that does not specifically hybridize to the template. In some embodiments, analysis of particular genes is conducted. In such cases, the primer may comprise a gene specific sequence capable of hybridizing to and priming extension reactions from templates comprising specific genes. In some embodiments, multiple genes are analyzed and a primer is among a plurality of primers. Each of the plurality of primers may have a sequence for a particular gene of interest.

[0346] Subsequent to the first amplification reaction, the first amplification product or transcript can be subjected to a second amplification reaction to generate a second amplification product. In some cases, additional sequences (e.g., functional sequences such as flow cell attachment sequence, sequencing primer binding sequences, barcode sequences, etc) are to be attached. The first and second amplification reactions can be performed in the same volume, such as for example in a droplet. In some cases, the first amplification product is subjected to a second amplification reaction in the presence of a barcoded oligonucleotide to generate a second amplification product having a barcode sequence. The barcode sequence can be unique to a partition, that is, each partition has a unique barcode sequence. The barcoded oligonucleotide may comprise a sequence of at least a segment of the template switching oligonucleotide and at least a second predefined sequence. The segment of the template switching oligonucleotide on the barcoded oligonucleotide can facilitate hybridization of the barcoded oligonucleotide to the transcript, e.g., cDNA, to facilitate the generation of a second amplification product. In addition to a barcode sequence, the barcoded oligonucleotide may comprise a second defined sequence such as at least one of an adaptor sequence, a unique molecular identifier (UMI) sequence, a primer binding site, and a sequencing primer binding site or any combination thereof.

[0347] In some embodiments, the second amplification reaction uses the first amplification product as a template and the barcoded oligonucleotide as a primer. As illustrated in operation **1954**, the segment of the template switching oligonucleotide on the barcoded oligonucleotide **1928** can hybridize to the portion of the cDNA or complementary fragment **1922** having a sequence complementary to the template switching oligonucleotide or that which was copied from the template switching oligonucleotide. In the second amplification reaction, extension reaction reagents, e.g., polymerase, nucleoside triphosphates, co-factors (e.g., Mg2+ or Mn2+), that are also co-partitioned, can extend the primer sequence using the first amplification product as template as illustrated in operation **1956**. The second amplification product can comprise a second predefined sequence (e.g., **1808**, **1812**, and **1810**), a sequence of a segment of the template polynucleotide (e.g., mRNA), and a sequence complementary to the primer (e.g., **1924**).

[0348] In some embodiments, the second amplification product uses the barcoded oligonucleotide as a template and at least a portion of the first amplification product as a primer. As illustrated in operation **1954**, the segment of the first amplification product (e.g., cDNA) having a sequence complementary to the template switching oligonucleotide can hybridize to the segment of the barcoded oligonucleotide comprising a sequence of at least a segment of the template switching oligonucleotide. In the second amplification reaction, extension reaction reagents, e.g., polymerase, nucleoside triphosphates, co-factors (e.g., Mg2+ or Mn2+), that are also co-partitioned, can extend the primer sequence (e.g., first amplification product) using the barcoded oligonucleotide as template as illustrated in operation **1958**. The second amplification product may comprise the sequence of the primer (e.g., **1924**), a sequence which is complementary to the sequence of the template polynucleotide (e.g., mRNA), and a sequence complementary to the second predefined sequence (e.g., **1808**, **1812**, and **1810**).

[0349] In some embodiments, the second amplification reaction is performed subsequent to the first amplification reaction in the presence of an intervening purification operation. An intervening purification operation can be used, for example, to purify the template (e.g., first amplification product) from excess reagents, including excess primers such as template switching oligonucleotides. In some embodiments, the amplification reaction is performed in the absence of an intervening purification operation. In certain embodiments, an intervening purification operation is not performed so that all sample preparation is performed in a same reaction volume. In the absence of an intervening purification operation, the template switching oligonucleotide may compete with barcoded oligonucleotide in the second amplification reaction as the barcoded oligonucleotide comprises at least a segment of the template switching oligonucleotide. Competition between the template switching oligonucleotide and barcoded oligonucleotide in the second amplification reaction to generate additional amplification product may result in a second amplification product lacking a barcode sequence. In some embodiments, the template switching oligonucleotide may out-compete the barcoded oligonucleotide in the second amplification reaction if the template switching oligonucleotide is present at a higher concentration in the reaction volume than the barcoded oligonucleotide. Various approaches can be utilized to favor the use of the barcoded oligonucleotide in the second amplification reaction to generate amplification products having a barcode sequence in situations where the barcoded oligonucleotide is present at a lower concentration than the template switching oligonucleotide in the reaction volume.

[0350] In some embodiments, the template switching oligonucleotide is not available for primer extension during the second amplification reaction. In some embodiments, the template switching oligonucleotide is degraded prior to the second amplification reaction. In some embodiments, the template switching oligonucleotide is degraded during the second amplification reaction. The template switching oligonucleotide may comprise ribonucleic acids (RNA). A template switching oligonucleotide comprising RNA can be degraded, for example, by elevated temperatures or alkaline conditions. In some embodiments, the template switching oligonucleotide comprises at least 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95% RNA. In some embodiments, the

template switching oligonucleotide comprises 100% RNA. In some embodiments, a first reaction rate of the second amplification reaction using the barcoded oligonucleotide is greater than a second reaction rate of the second amplification using the template switching oligonucleotide.

[0351] In some embodiments, the barcoded oligonucleotide can hybridize to the first amplification product at a higher annealing temperature as compared to the template switching oligonucleotide. For example, the first amplification product and the barcoded oligonucleotide can have a higher melting temperature as compared to a melting temperature of the first amplification product and the template switching oligonucleotide. In such cases, the second amplification reaction may be performed with an annealing temperature at which the barcoded oligonucleotide is able to hybridize to the first amplification product and initiation primer extension and at which the template switching oligonucleotide is unable to hybridize to the first amplification product and initiate primer extension. In some embodiments, the primer annealing temperature of the second amplification reaction is at least about 0.5° C., 1° C., 2° C., 3° C., 4° C., 5° C., 6° C., 7° C., 8° C., 9° C., 10° C. or greater than a primer annealing temperature of the first amplification reaction. The difference in melting temperatures can result from the presence of modified nucleotides in the template switching oligonucleotide. In some embodiment, the template switching oligonucleotide comprises at least about 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95% modified nucleotides. In some embodiments, the template switching oligonucleotide comprises 100% modified oligonucleotides. In some embodiments, the difference in melting temperature can be the result of the presence of modified nucleotides in the barcoded oligonucleotide. In some embodiment, the barcoded oligonucleotide comprises at least 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90°/a, or 95% modified nucleotides. In some embodiments, the barcoded oligonucleotide comprises 100% modified oligonucleotides. Modified nucleotides include, but are not limited to, 2-Aminopurine, 2,6-Diaminopurine (2-Amino-dA), inverted dT, 5-Methyl dC, 2'-deoxyInosine, Super T (5-hydroxybutynl-2'-deoxyuridine), Super G (8-aza-7-deazaguanosine), locked nucleic acids (LNAs), unlocked nucleic acids (UNAs, e.g., UNA-A, UNA-U, UNA-C, UNA-G), Iso-dG, Iso-dC, and 2' Fluoro bases (e.g., Fluoro C, Fluoro U, Fluoro A, and Fluoro G).

[0352] In various embodiments, the first amplification reaction is facilitated using an enzyme comprising polymerase activity. For example, the first amplification reaction can be facilitated by a DNA-dependent polymerase or a reverse-transcriptase (e.g., RNA dependent). In some embodiments, the first amplification reaction comprises polymerase chain reaction. In some embodiments, the first amplification reaction comprises reverse transcription. In various embodiments, the second amplification reaction is facilitated using an enzyme comprising polymerase activity. For example, the second amplification reaction can be facilitated by a DNA-dependent polymerase. In some embodiments, the second amplification reaction comprises polymerase chain reaction.

[0353] Following the generation of amplification products, subsequent operations may include purification (e.g., via solid phase reversible immobilization (SPRI)), further processing (e.g., shearing, ligation of functional sequences, and subsequent amplification (e.g., via PCR)). These operations may occur in bulk (e.g., outside the partition). In the case where a partition is a droplet in an emulsion, the emulsion can be broken and the contents of the droplet pooled for additional operations. Additional reagents that may be co-partitioned along with the barcode bearing bead may include oligonucleotides to block ribosomal RNA (rRNA) and nucleases to digest genomic DNA from cells. Alternatively, rRNA removal agents may be applied during additional processing operations. The configuration of the constructs generated by such a method can help minimize (or avoid) sequencing of the poly-T sequence during sequencing and/or sequence the 5' end of a polynucleotide sequence. The amplification products, for example first amplification products and/or second amplification products, may be subject to sequencing for sequence analysis.

[0354] Although operations with various barcode designs have been discussed individually, individual beads can include barcode oligonucleotides of various designs for simultaneous use.

[0355] In addition to characterizing individual cells or cell sub-populations from larger populations, the processes and systems described herein may also be used to characterize individual cells as a way to provide an overall profile of a cellular, or other organismal population. A variety of applications require the evaluation of the presence and quantification of different cell or organism types within a population of cells, including, for example, microbiome analysis and characterization, environmental testing, food safety testing, epidemiological analysis, e.g., in tracing contamination or the like. In particular, the analysis processes described above may be used to individually characterize, sequence and/or identify large numbers of individual cells within a population. This characterization may then be used to assemble an overall profile of the originating population, which can provide important prognostic and diagnostic information.

[0356] For example, shifts in human microbiomes, including, e.g., gut, buccal, epidermal microbiomes, etc., have been identified as being both diagnostic and prognostic of different conditions or general states of health. Using the single cell analysis methods and systems described herein, one can again, characterize, sequence and identify individual cells in an overall population, and identify shifts within that population that may be indicative of diagnostic ally relevant factors. By way of example, sequencing of bacterial **16**S ribosomal RNA genes has been used as a highly accurate method for taxonomic classification of bacteria. Using the targeted amplification and sequencing processes described above can provide identification of individual cells within a population of cells. One may further quantify the numbers of different cells within a population to identify current states or shifts in states over time. See, e.g., Morgan et al, PLoS Comput. Biol., Ch. 12, December 2012, 8(12):e1002808, and Ram et al., Syst. Biol. Reprod. Med., June 2011, 57(3):162-170, each of which is entirely incorporated herein by reference for all purposes. Likewise, identification and diagnosis of infection or potential infection may also benefit from the single cell analyses described herein, e.g., to identify microbial species present in large mixes of other cells or other biological material, cells and/or nucleic acids, including the environments described above, as well as any other diagnostically relevant environments, e.g., cerebrospinal fluid, blood, fecal or intestinal samples, or the like.

[0357] The foregoing analyses may also be particularly useful in the characterization of potential drug resistance of different cells or pathogens, e.g., cancer cells, bacterial pathogens, etc., through the analysis of distribution and profiling of different resistance markers/mutations across cell populations in a given sample. Additionally, characterization of shifts in these markers/mutations across populations of cells over time can provide valuable insight into the progression, alteration, prevention, and treatment of a variety of diseases characterized by such drug resistance issues.

[0358] Although described in terms of cells, it will be appreciated that any of a variety of individual biological organisms, or components of organisms are encompassed within this description, including, for example, cells, viruses, organelles, cellular inclusions, vesicles, or the like. Additionally, where referring to cells, it will be appreciated that such reference includes any type of cell, including without limitation prokaryotic cells, eukaryotic cells, bacterial, fungal, plant, mammalian, or other animal cell types, mycoplasmas, normal tissue cells, tumor cells, or any other cell type, whether derived from single cell or multicellular organisms.

[0359] Similarly, analysis of different environmental samples to profile the microbial organisms, viruses, or other biological contaminants that are present within such samples, can provide important information about disease epidemiology, and potentially aid in forecasting disease outbreaks, epidemics an pandemics.

[0360] As described above, the methods, systems and compositions described herein may also be used for analysis and characterization of other aspects of individual cells or populations of cells.

[0361] A method **2000** for characterizing a cell is shown in FIG. **20**. The method **2000** may comprise, as shown in operation **2010**, providing a partition comprising a cell and at least one labelling agent, all as described herein. The labelling agent may be capable of binding to a cell surface feature of the cell, and may be coupled to a reporter oligonucleotide comprising a nucleic acid barcode sequence that permits identification of the labelling agent. Further, the partition may comprise one or more anchor oligonucleotides (also referred to herein as oligonucleotides and barcoded oligonucleotides) that are capable of interacting with the reporter oligonucleotide barcode, as described in detail herein. Next, in operation **2020**, within the partition a nucleic acid molecule comprising at least a portion of the nucleic acid barcode sequence or a complement thereof may be synthesized, as described herein. Next, in operation **2030**, the nucleic acid molecule may be sequenced to identify the labelling agent or the cell. In some cases, the labelling agent and/or the reporter oligonucleotide may be delivered into the cell, e.g., by transfection (e.g., using transfectamine), by lipid (e.g., 1,2-Dioleoyl-sn-glycero-3-phosphocholine (DOPC)), or by transporter proteins.

[0362] As described herein, a labelling agent may comprise an antibody, or an epitope binding fragment thereof, a cell surface receptor binding molecule, a receptor ligand, a small molecule, a bi-specific antibody, a bi-specific T-cell engager, a T-cell receptor engager, a B-cell receptor engager, a pro-body, an aptamer, a monobody, an affimer, a darpin, a protein scaffold, and the like, and any combination thereof. As described herein, a cell surface feature may comprise a receptor, an antigen, a surface protein, a transmembrane protein, a cluster of differentiation protein, a protein channel, a protein pump, a carrier protein, a phospholipid, a glycoprotein, a glycolipid, a cell-cell interaction protein complex, an antigen-presenting complex, a major histocompatibility complex, an engineered T-cell receptor, a T-cell receptor, a B-cell receptor, a chimeric antigen receptor, a gap junction, an adherens junction, and the like, and any combination thereof.

[0363] In some instances, prior to operation **2010**, labelling agents may be subjected to conditions suitable for binding the labelling agents to cell surface features. In some instances, prior to operation **2010**, labelling agents may be subjected to conditions suitable for binding the labelling agents to cell surface features when the cell and the labelling agents are free from the partition (e.g., prior to partitioning). In some instances, prior to operation **2010**, the reporter oligonucleotide may be coupled to the labelling agent. In some instances, in operation **2010**, at least one labelling agent is bound to the cell surface feature.

[0364] In some instances, in operation **2020**, the reporter oligonucleotide coupled to the labelling agent may be subjected to a primer extension reaction that generates the nucleic acid molecule. In some instances, in operation **2020**, the anchor oligonucleotide may be coupled to a bead also partitioned with the cell and labelling agent(s), as described herein, and the method further comprises releasing the anchor oligonucleotide from the bead prior to synthesizing.

[0365] As described herein, the bead may comprise a gel bead. Further, as described herein, the bead may comprise a diverse library of anchor oligonucleotides. In some instances, the bead may comprise at least about 1,000 copies of an anchor oligonucleotide, at least about 10,000 copies of an anchor oligonucleotide, at least about 100,000 copies of an anchor oligonucleotide, at least about 100,000 copies of an anchor oligonucleotide, at least about 1,000,000 copies of an anchor oligonucleotide, at least about 5,000,000 copies of an anchor oligonucleotide, or at least about 10,000,000 copies of an anchor oligonucleotide. In some instances, the bead may comprise at least about 1,000 copies of diverse anchor oligonucleotides, at least about 10,000 copies of diverse anchor oligonucleotides, at least about 100,000 copies of diverse anchor oligonucleotides, at least about 100,000 copies of diverse anchor oligonucleotides, at least about 1,00,000 copies of diverse anchor oligonucleotides, at least about 5,000,000 copies of diverse anchor oligonucleotides, or at least about 10,000,000 copies of diverse anchor oligonucleotides. In some instances, and as described herein, releasing anchor oligonucleotides from the bead may comprise subjecting the bead to a stimulus that degrades the bead. In some instances, as described herein, releasing anchor oligonucleotides from the bead may comprise subjecting the bead to a chemical stimulus that degrades the bead.

[0366] A solid support (e.g., a bead) may comprise different types of anchor oligonucleotides for analyzing both intrinsic and extrinsic information of a cell. For example, a solid support may comprise one or more of the following: 1) an anchor oligonucleotide comprising a primer that binds to one or more endogenous nucleic acids in the cell; 2) an anchor oligonucleotide comprising a primer that binds to one or more exogenous nucleic acids in the cell, e.g., nucleic acids from a microorganism (e.g., a virus, a bacterium) that infects the cell, nucleic acids introduced into the cell (e.g., such as plasmids or nucleic acid derived therefrom), nucleic acids for gene editing (e.g., CRISPR-related RNA such as

crRNA, guide RNA); 3) an anchor oligonucleotide comprising a primer that binds to a barcode (e.g., a barcode of a nucleic acid, of a protein, or of a cell); and 4) an anchor oligonucleotide comprising a sequence (e.g., a primer) that binds to a protein, e.g., an exogenous protein expressed in the cell, an protein from a microorganism (e.g., a virus, a bacterium) that infects the cell, or an binding partner for a protein of the cell (e.g., an antigen for an immune cell receptor).

[0367] In some cases, the methods may be used to screen cells carrying mutations, e.g., mutations generated by gene editing such as CRISPR technology. For example, a bead comprising a first anchor oligonucleotide with a primer for CRISPR RNA (e.g., crRNA or guide RNA) or its complementary DNA and a second anchor oligonucleotide with a primer endogenous nucleic acid in the cell, e.g., total mRNA or a specific mRNA. The bead may be made into a partition with a cell transfected with CRISPR RNA or a plasmid expressing CRISPR RNA. In some cases, the expressed CRISPR RNA or the plasmid may have a barcode (CRISPR barcode) or a capture sequence. The primers on the bead may be used to amplify and sequence the CRISPR RNA (e.g., using a barcoded adapter oligonucleotide comprising a sequence complementary to the CRISPR capture sequence, see FIGS. **61**A-**61**D) and endogenous mRNA (e.g., using a barcoded adapter oligonucleotide comprising an oligo(dT) sequence), thus determining the mutations generated by in the cell (see FIG. **61**D). In some cases, the methods may be used to perform single cell RNA sequencing, e.g., as described in Dixit, et al., Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. Cell; Dec. 15, 2016; 167(7):1853-1866. e17, which is incorporated herein by reference in its entirety.

[0368] An oligonucleotide of an anchor agent or a labelling agent may comprise a backbone. The backbone may comprise one or more of the following elements: a sequencer primer, a barcode, and a UMI. In addition to the backbone, the oligonucleotide may also comprise a primer as described herein, e.g., a poly-T primer, a random N-mer primer, and/or a target-specific primer. Examples of oligonucleotides comprising various backbones and primer sequences are shown in FIGS. **27**A-**27**D.

[0369] An example work flow for the methods herein may include inputting fixed reference (e.g., known transcripts from a cell with intrinsic information), reference templates (e.g., design of synthetic barcodes (random or target-specific) with extrinsic information, and sequence reads; and outputting classification of sequence reads as originating from intrinsic or extrinsic sequences, counts of detected copies per transcript/gene per partition, and list and counts of detected barcodes from extrinsic sequences per partition. In some cases, the example workflow may be implemented with software.

[0370] In some instances, prior to operation **2030**, the method **2000** may comprise releasing the nucleic acid molecule from the partition (e.g., by disruption of the partition). In some instances, operation **2030** may comprise identifying the labelling agent (e.g., the labelling agent bound to a cell surface feature). In some instances, operation **2030** may comprise identifying the cell surface feature from identifying the labelling agent. In some instances, operation **2030** comprises determining an abundance of the given cell surface feature on the cell. In some instances, operation **2030** comprises identifying the cell. In some instances, operation **2030** comprises identifying the labelling agent and the cell.

[0371] In method **2000**, the reporter oligonucleotide that may be coupled to the labelling agent may comprise a unique molecular identification (UMI) sequence, as described herein. The UMI sequence may permit identification of the cell, the labelling agent, or both. In some instances, operation **2030** of method **2000** may comprise determining a sequence of the UMI sequence and identifying the cell.

[0372] In method **2000**, the anchor oligonucleotide may comprise a unique molecular identification (UMI) sequence, as described herein. In these instances, the UMI sequence of the anchor oligonucleotide may permit identification of the cell. In some instances, operation **2030** of method **2000** may comprise determining a sequence of the UMI sequence from the reporter oligonucleotide bound to the labelling agent, and a sequence of the UMI sequence from the anchor oligonucleotide, to identify the cell and the cell surface feature.

[0373] In method **2000**, and as described herein, the partition may comprise a droplet in an emulsion. In some instances, the partition comprises only one cell. In some instances, the cell is bound to at least one labelling agent. In some instances, the labelling agent may comprise at least two of the same labelling agent. In some instances, the labelling agent may comprise at least two different labelling agents. In some instances, the cell may be bound to at least about 5 different labelling agents, at least about 10 different labelling agents, at least about 50 different labelling agents, at least about 100 different labelling agents, at least about 500 different labelling agents, at least about 1,000 different labelling agents, at least about 5,000 different labelling agents, at least about 10,000 different labelling agents, or at least about 50,000 different labelling agents. In some instances, the cell may be bound to between about 2 and 5 different labelling agents, between about 5 and 10 different labelling agents, between about 10 and 100 different labelling agents, between about 100 and 500 different labelling agents, between about 500 and 1,000 different labelling agents, between about 1,000 and 5,000 different labelling agents, between about 5,000 and 10,000 different labelling agents, between about 10,000 and 50,000 different labelling agents, or between about 2 and 50,000 different labelling agents, or any range in-between. In some instances, operation **2030** of method **2000** may comprise determining an identity of at least a subset of the different labelling agents.

[0374] In one example process, a sample is provided that contains cells that are to be analyzed and characterized as to their cell surface features. A cell surface feature may include, but is not limited to, a receptor, an antigen, a surface protein, a transmembrane protein, a cluster of differentiation protein, a protein channel, a protein pump, a carrier protein, a phospholipid, a glycoprotein, a glycolipid, a cell-cell interaction protein complex, an antigen-presenting complex, a major histocompatibility complex, an engineered T-cell receptor, a T-cell receptor, a B-cell receptor, a chimeric antigen receptor, a gap junction, an adherens junction, or any combination thereof. Also provided is at least one labelling agent, such as a library of labelling agents, capable of binding to a cell surface feature of interest. A labelling agent may include, but is not limited to, an antibody, or an epitope binding fragment thereof, a cell surface receptor binding

molecule, a receptor ligand, a small molecule, a bi-specific antibody, a bi-specific T-cell engager, a T-cell receptor engager, a B-cell receptor engager, a pro-body, an aptamer, a monobody, an affimer, a darpin, and a protein scaffold, or any combination thereof. The labelling agents can include a reporter oligonucleotide that is indicative of the cell surface feature to which the binding group binds. In particular, a labelling agent that is specific to one type of cell surface feature may have coupled thereto a first reporter oligonucleotide, while a labelling agent that is specific to a different cell surface feature may have a different reporter oligonucleotide coupled thereto. In some aspects, these reporter oligonucleotides may comprise nucleic acid barcode sequences that permit identification of the labelling agent which the reporter oligonucleotide is coupled to. The selection of oligonucleotides as the reporter may provide advantages of being able to generate significant diversity in terms of sequence, while also being readily attachable to most biomolecules, e.g., antibodies, etc., as well as being readily detected, e.g., using sequencing or array technologies. In some embodiments, the labelling agents may include reporter oligonucleotides attached to them. Thus, a first labelling agent, e.g., an antibody to a first cell surface feature, may have associated with it a reporter oligonucleotide that has a first nucleic acid sequence. Different labelling agents, e.g., antibodies having binding affinity for other, different cell surface features, may have associated therewith reporter oligonucleotides that comprise different nucleic acid sequences, e.g., having a partially or completely different nucleic acid sequence. In some cases, for each type of cell surface feature labelling agent, e.g., antibody or antibody fragment, the reporter oligonucleotide sequence may be known and readily identifiable as being associated with the known cell surface feature labelling agent. These reporter oligonucleotides may be directly coupled to the labelling agent, or they may be attached to a bead, molecular lattice, e.g., a linear, globular, cross-slinked, or other polymer, or other framework that is attached or otherwise associated with the labelling agent, which allows attachment of multiple reporter oligonucleotides to a single labelling agent.

[0375] In the case of multiple reporter oligonucleotides coupled to a single labelling agent, such reporter oligonucleotides can comprise the same sequence, or a particular labelling agent may include a known set of reporter oligonucleotide sequences. As between different labelling agents, e.g., specific for different cell surface features, the reporter oligonucleotides may be different and attributable to the particular labelling agent.

[0376] Attachment (coupling) of the reporter oligonucleotides to the labelling agents may be achieved through any of a variety of direct or indirect, covalent or non-covalent associations or attachments. For example, in the case of oligonucleotide reporter oligonucleotides associated with antibody based labelling agents, such oligonucleotides may be covalently attached to a portion of an antibody or antibody fragment using chemical conjugation techniques (e.g., Lightning-Link® antibody labelling kits available from Innova Biosciences), as well as other non-covalent attachment mechanisms, e.g., using biotinylated antibodies and oligonucleotides (or beads that include one or more biotinylated linker, coupled to oligonucleotides) with an avidin or streptavidin linker. Antibody and oligonucleotide biotinylation techniques are available. See, e.g., Fang, et al., "Fluoride-Cleavable Biotinylation Phosphoramidite for 5'-end-Labelling and Affinity Purification of Synthetic Oligonucleotides," Nucleic Acids Res. Jan. 15, 2003; 31(2): 708-715, which is entirely incorporated herein by reference for all purposes. Likewise, protein and peptide biotinylation techniques have been developed and are readily available. See, e.g., U.S. Pat. No. 6,265,552, which is entirely incorporated herein by reference for all purposes. Furthermore, click reaction chemistry such as a Methyltetrazine-PEG5-NHS Ester reaction, a TCO-PEG4-NHS Ester reaction, or the like, may be used to couple reporter oligonucleotides to labelling agents. In the case that the labelling agent is a primary antibody, a reporter oligonucleotide may be coupled to the labelling agent through a secondary antibody coupling interaction. Commercially available kits, such as those from Thunderlink and Abcam, and techniques common in the art may be used to couple reporter oligonucleotides to labelling agents as appropriate.

[0377] In some cases, a reporter oligonucleotide may be associated (e.g., covalently linked such as conjugated or non-covalently bound through a binding interaction) to an antibody via an antibody-binding protein. For example, a reporter oligonucleotide and an antibody-binding protein may form a complex. The complex may bind to a respective antibody through the antibody-binding protein. FIG. **28** shows an example workflow for associating a nucleic acid (e.g., DNA) barcode on an antibody using an antibody-binding protein. An antibody binding protein **2810**, e.g., Protein A or Protein G, and an oligonucleotide comprising a nucleic acid (e.g., DNA) barcode **2820** are conjugated to the Fc region of an antibody, forming a complex **2830** comprising the antibody, the antibody-binding protein **2810**, and the DNA barcode **2820**. The complex **2830** is incubated with cells and unbound antibody is washed out. When the complex **2830** binds to a cell, the complex and the cell are partitioned into a droplet for further analysis.

[0378] An antibody-binding protein may have fast adsorption kinetics, slow desorption kinetics, and/or a low binding equilibrium constant. Any methods for adding chemical functionality to peptides or proteins may be used. Some methods may include attaching a reporter oligonucleotide to specific amino acids or chemical groups (e.g., chemical groups present in multiple types of proteins) on the antibody-binding protein. The conjugation of antibody-binding proteins and oligonucleotides may be performed using methods for forming antibody-nucleic acid conjugation described herein, e.g., using click chemistry. Dissociation of the antibody-binding protein/oligonucleotide complexes may be prevented by crosslinking (e.g., using a crosslinker such as formaldehyde), protein engineering, or adding the protein-binding proteins in excess.

[0379] Examples of antibody-binding proteins include proteins that bind to the constant (Fc) region of antibodies, such as Protein A, Protein G, Protein L, or fragments thereof. Other binding proteins (e.g., streptavidin) may be expressed as fusion proteins with antibody-binding proteins, and used to associate oligonucleotides (e.g., by binding of biotinylated oligonucleotides to a streptavidin-Protein A fusion protein). Other antibody-binding proteins or domains may provide additional binding affinity for various antibody classes. In some cases, the antibody-binding protein may be an antibody, e.g., a secondary antibody for the antibody targeting the sample. The secondary antibody may comprise

an oligonucleotide described here, e.g., an oligonucleotide with a barcode and a poly-A or poly T terminated sequence.

[0380] The antibody-binding proteins may be engineered to introduce additional functionalities. Antibody-binding proteins may be engineered to contain amino acids with functional groups amenable to conjugation with oligonucleotide. For example, the antibody-binding proteins may naturally have or be engineered to have cysteine residues, e.g., for controlling stoichiometry and/or attachment location of the oligonucleotides. The antibody-binding proteins may be engineered to have non-natural amino acid residues, e.g., for targeted crosslinking of binding proteins and antibodies. The antibody-binding proteins may be engineered to have tags, e.g., fluorescent tags (e.g., by fusing with a fluorescent protein such as green fluorescence protein (GFP), red fluorescence protein (RFP), yellow fluorescence protein (YFP)) and/or affinity tags for purification and visualization. The fluorescent tags and/or the affinity tags may be cleavable. In some cases, the antigen-binding protein may be engineered to have one or more (e.g., only one) barcode attachment sites per protein.

[0381] Also provided herein are kits comprising antibody-binding proteins conjugated with reporter oligonucleotides, e.g., in well plates. Antibody for an assay may be incubated with the antibody-binding proteins conjugated with reporter oligonucleotides at a specified concentration without interfering with the antibody's binding site and/or without the need for any chemistry to be carried out in the customer's hands to conjugate the reporter oligonucleotide to the antibody.

[0382] The reporter oligonucleotides may be provided having any of a range of different lengths, depending upon the diversity of reporter oligonucleotides suitable for a given analysis, the sequence detection scheme employed, and the like. In some cases, these reporter oligonucleotides can be greater than or equal to about 5 nucleotides in length, greater than or equal to about 10 nucleotides in length, greater than or equal to about 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 150, 200 or 250 nucleotides in length. In some cases, these reporter oligonucleotides may be less than or equal to about 250, 200, 180, 150, 120 100, 90, 80, 70, 60, 50, 40, 30, 20, 10 or 5 nucleotides in length. In some cases, the reporter oligonucleotides may be selected to provide barcoded products that are already sized, and otherwise configured to be analyzed on a sequencing system. For example, these sequences may be provided at a length that ideally creates sequenceable products of a suitable length for particular sequencing systems. Likewise, these reporter oligonucleotides may include additional sequence elements, in addition to the reporter sequence, such as sequencer attachment sequences, sequencing primer sequences, amplification primer sequences, or the complements to any of these.

[0383] In operation, a cell-containing sample may be incubated with the labelling agents and their associated reporter oligonucleotides, for any of the cell surface features to be analyzed. Following incubation, the cells may be washed to remove unbound labelling agents. Following washing, the cells may be partitioned into separate partitions, e.g., droplets, along with the barcode (also referred to as anchor oligonucleotides) carrying beads described above, where each partition includes a limited number of cells, e.g., a single cell. Upon releasing of the barcodes (or anchor oligonucleotides) from the beads, they may prime the amplification and barcoding of the reporter oligonucleotides coupled to the labelling agents. The barcoded replicates of the reporter oligonucleotides may additionally include functional sequences, such as primer sequences, attachment sequences or the like.

[0384] The barcoded reporter oligonucleotides may then subjected to sequence analysis to identify which reporter oligonucleotides were bound to the cells (i.e., cell surface features) within the partitions. Further, by also sequencing the associated barcode sequence, one can identify that a given cell surface feature likely came from the same cell as other, different cell surface features, whose reporter sequences include the same barcode sequence, i.e., they were derived from the same partition.

[0385] In some embodiments, anchor oligonucleotides within the partition may interact with the reporter oligonucleotides coupled to labelling agents bound to cell surface features and lead to the synthesizing of a nucleic acid molecule as described herein, where the synthesized nucleic acid molecule may comprise at least a portion of the nucleic acid barcode sequence(s), or complement(s) thereof, that comprise the reporter oligonucleotide, or the anchor oligonucleotide, or both. These synthesized nucleic acid molecules may then be subjected to amplification and sequencing, as described herein.

[0386] In some embodiments, more than one labelling agent may be bound to a single cell surface feature, and proximity between the labelling agents may allow the 3' ends of the reporter oligonucleotides coupled thereto to hybridize (wherein this hybridization is discouraged by the melting temperature when unbound in solution). By an extension reaction as described herein, a nucleic acid molecule may be synthesized, amplified, and subjected to sequencing, as described herein.

[0387] Based upon the reporter oligonucleotides that emanate from an individual partition based upon the presence of the barcode sequence, one may then create a cell surface feature profile of individual cells from a population of cells. Profiles of individual cells or populations of cells may be compared to profiles from other cells, e.g., 'normal' cells, to identify variations in cell surface features, which may provide diagnostically relevant information. In particular, these profiles may be particularly useful in the diagnosis of a variety of disorders that are characterized by variations in cell surface receptors, such as cancer and other disorders.

[0388] In some embodiments, the genomic, proteomic, and cell surface information of cells characterized by the methods and systems described herein may be sequenced individually. In some embodiments, the genomic, proteomic, and cell surface information of cells characterized by the methods and systems described herein may be pooled and sequenced together. In some embodiments, the genomic, proteomic, and cell surface information of cells characterized by the methods and systems described herein may be sequenced sequentially (i.e., cell surface information characterized first, then proteomic and genomic information).

[0389] Also provided herein are compositions and methods for screening a chemical compound library. The methods may comprise providing a partition comprising at least one chemical compound and an identifier of the partition. The identifier may be an oligonucleotide comprising a nucleic acid barcode sequence as described in the application. The identifier oligonucleotide may be amplified and subject to sequence. The sequence read of the identifier oligonucleotide or a fragment thereof may be used to

identify the partition and the at least one chemical compound in the partition. The methods may be used for screening a chemical compound library in a reaction of small volumes, e.g., on the scale of nanoliters. Multiple reactions may be performed in different partitions with the same substrate and/or reagent. The reaction may be multiplexed to decrease the effort and time needed to process the same number of compounds in reactions of larger scale, e.g., on the scale of microliters. The methods and compositions may allow high throughput screening of a chemical compound library with low noise and/or false-positive results. In some cases, a method for screening a chemical compound library may comprise one or more of the following operations: (1) providing a plurality of partitions, wherein a given partition of the plurality of partitions (i) has or is suspected of having at least one chemical compound and (ii) comprises an identifier oligonucleotide comprising a nucleic acid barcode sequence that permits identification of the given partition; (2) subjecting the plurality of partitions to screening under conditions sufficient to select a subset of the plurality of partitions from a remainder of the plurality of partitions, which subset comprises the given partition having or suspected of having the at least one chemical compound; (3) subjecting the subset of the plurality of partitions, including the given partition, to conditions sufficient to generate a nucleic acid molecule comprising at least a portion of the nucleic acid barcode sequence or a complement thereof, and (4) sequencing the nucleic acid molecule to generate sequence reads, which sequence reads permit identification of the at least one chemical compound.

[0390] The methods may comprise building combinatorial chemical and identifier oligonucleotide libraries on a solid support, e.g., a monodispersed polymeric bead. The oligonucleotide barcoding may be intrinsically linked to a chemical synthesis path unique for that monodispersed polymer bead. Upon partitioning this polymeric bead, the population of compounds may be released from the substrate to interact with the target molecule unencumbered by the identifier oligonucleotides. Partitions may then be sorted based on positive/negative interactions as indicated by a traditional reporter assay. Positives partitions may then be homogenized and pooled. The identifier oligonucleotides in the positive partitions may be amplified for sequencing. The methods may allow for large quantities of single compounds to be packaged into nanoliter partitions individually and for the subsequent deconvolution of partitions with positive interactions that may be pooled and processed in a multiplexed format.

[0391] In some cases, the methods comprise synthesizing a controlled number of chemical compounds on a solid support (e.g., a bead) while simultaneously synthesizing a controlled number of identifier oligonucleotides unique to the compounds on the solid support. The combinatorial libraries of the chemical compounds and identifier oligonucleotides may be made through sequential additions of chemical compound subunits that concord with simultaneous or subsequent sequential additions of identifier oligonucleotides on the solid matrix. The methods may be multiplexed in a single vessel for additions of chemical compounds and identifier oligonucleotides in a massively parallel way. The quantity of the chemical compounds to be screened may be normalized.

[0392] The number of chemical compounds and/or identifier oligonucleotides synthesized on a solid support may be controlled by adjusting the number of attachment points. An attachment point may be a location on a solid support where a chemical compound or identifier oligonucleotide may be attached to. Attachment points may include multiple types of chemistries for the cleavage of chemical compounds and/or identifier oligonucleotides. This allows for selective release of chemical compounds and/or identifier oligonucleotides in a controlled fashion. The solid may have a single or multiple attachment points.

[0393] The solid support may act as a covalent linker between chemical compounds and identifier oligonucleotides. A single type of solid support or multiple types of solid support may be used in the screening. If multiple types of solid support are used, they may be covalently linked to form a single solid support. In certain cases, if multiple types of solid support are used, they may be comingled (but not covalently linked) and occupy the same physical space. A solid support may have two or more matrices intermingled. In these cases, chemical compounds and the identifier oligonucleotides may be on the same matrix or on separate matrices of the solid support. In the latter case, the chemical compounds and the identifier oligonucleotides are comingled (and not covalently linked) and occupy the same physical space. In some cases, the solid support may be permeable or non-permeable. In certain cases, the solid support may be dissolvable or non-dissolvable.

[0394] A chemical compound may be a protein (e.g., an antibody or a fragment thereof, or an antigen or a fragment thereof), a nucleic acid molecule. In some cases, a chemical compound may be a small molecule compound. A small molecule compound may be a low molecular weight (e.g., no greater than 1000 daltons) organic compound that may help regulate a biological process. A small compound may have a size on the order of 1 nm. For example, a small molecule compound may be a small molecule drug.

[0395] Screening of a chemical compound library may be performed using methods for screening small molecules for drug discovery. For example, the screening may be performed using high-throughput screening or high-content analysis in drug discovery. A high-throughput screening may be a screening that identifies active compounds, antibodies, or genes that modulate a particular biomolecular pathway. A high-content analysis may be a screening that identifies substances such as small molecules, peptides, or RNAi that alter the phenotype of a cell in certain manner. In some cases, a screening may be an immunoassay, e.g., enzyme-linked immunosorbent assay (ELISA).

[0396] Also provided herein are scaffolds for delivery of one or more reagents. In some cases, a reagent is not covalently bound to the solid scaffold. For example, the reagent may be inside the scaffold and hindered (e.g., through steric interaction with the scaffold) from diffusing out of the scaffold. The reagent may be released from the scaffold when the scaffold is dissolved. In some cases, the scaffold may be a microcapsule described herein, such as a gel bead.

[0397] The scaffold may be used in a method for characterizing a cell. The method may comprise providing a partition comprising a cell, a scaffold, and an reagent in the scaffold. To characterize the cell in the partition, the scaffold may be dissolved to release the reagent. The reagent then contacts with the cell for determining one or more characteristics of the cell. In some cases, the partition may com-

prise a plurality of reagents. Any reagent described in the disclosure may be used in this method.

[0398] The scaffold may be used to deliver two or more reagents. In some cases, a first reagent be non-covalently bound to the scaffold, and the second reagent may be covalently bound to the scaffold. In other cases, multiple scaffolds may be used to deliver multiple reagents. In these cases, a first reagent may be covalently bound to a first scaffold, and a second reagent may be non-covalently bound to a second scaffold. The first scaffold and the second scaffold may be encapsulated in the same partition with a cell.

[0399] The reagent that is non-covalently bound to the scaffold may be released when the scaffold is dissolved. A scaffold is dissolved when at least 0.01%, 0.1, 1%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, or 100% of the volume of the scaffold is dissolved in the solution around it.

[0400] The scaffold may comprise one or more pores and the reagent non-covalently bound to the scaffold may be in the one or more pores. The diameter of the one or more pores may be up to 0.01 nm, 0.1 nm, 1 nm, 5 nm, 10 nm, 50 nm, 100 nm, 200 nm, 400 nm, 600 nm, 800 nm, 1 μm, or 10 μm.

[0401] A scaffold loaded with a non-covalently bound reagent may be made using any method of incorporating an agent in a solid substance. In some cases, the scaffold loaded with a non-covalently bound reagent may be made using the one or more of following operations: 1) Placing the scaffold (e.g., gel bead) and the reagent under a condition that causes the scaffold to swell and the pores defined by the polymer scaffold to enlarge. Such condition may include: in a thermodynamically-favorable solvent, at higher or lower temperatures (e.g., for temperature-responsive hydrogel materials), in a solvent with higher or lower ion concentration and/or in the presence or absence of an electric field for electric charge-/field-responsive hydrogel materials; 2) Allowing sufficient time for the reagent to diffuse into the interior of the scaffold; 3) Transferring the scaffold into a condition that causes the pores to shrink. The reagent molecules within the scaffold are then hindered from diffusing out of the scaffold by steric interactions with the polymer scaffold. The transfer in operation 3) may be achieved microfluidically, e.g., by moving the scaffold from one co-flowing solvent stream to another. FIG. **29** demonstrates examples of swelling conditions and de-swelling conditions in the process. The swellability and pore sizes of the scaffold may be adjusted by changing the polymer composition.

[0402] In a partition comprising a scaffold loaded with non-covalently bound reagent, the composition of the partition may be adjusted by including a scaffold of a certain volume. For example, when a partition has a fixed volume, the concentration of the reagent in the partition may be upregulated by including a reagent-loaded scaffold of a larger volume. In some cases, the adjustment may be performed without changing the initial concentration of the components in the partition. In certain cases, the adjustment may be performed without changing the total volume of the partition. Such methods are useful for delivering a reagent that interferes with the partition generation, e.g., a cell lysis agent.

[0403] A partition with the scaffold may be generated using methods described in the disclosure. In certain cases, during the partition generation, both the scaffold and the liquid immediately surrounding the scaffold are encapsulated in a single partition as shown in FIG. **30**. The volume of the scaffold and the surrounding liquid comprise a "unit cell". Unit cells may be defined by the geometry of the microchannel in which scaffolds flow and by the pressure applied. For example, higher pressures may compress the scaffold, which are deformable, thereby reducing the volume of the unit cell.

[0404] The composition of a partition may be determined by the volume of scaffold suspension (Z1) and the volume of the sample (Z2) encapsulated in that partition. The characteristic of the composition may be described by the ratio of these two volumes (Z1/Z2). The maximum Z1 possible for single-scaffold encapsulations is equal to the volume of the unit cell. Thus, to increase the concentration of a reagent delivered by the scaffold in a partition of a fixed volume without increasing the concentration of the reagent in the scaffold suspension, the dimensions of the scaffold may be increased. Thus, the encapsulated unit cell may occupy a greater volume of the partition (at higher Z1/Z2 ratio). In a microchannel for making the partitions, the dimension of the microchannel may or may not have to be increased to accommodate the larger partitions, depending on the mechanical properties of the scaffolds. When higher pressures are applied, the scaffold may compress, the volume of the unit cell may decrease, and a lower Z1/Z2 ratio may be achieved.

[0405] Also provided herein are methods and compositions for sequencing DNA (e.g., genomic DNA) molecules and RNA (e.g., mRNA) molecules from a cell in parallel and/or simultaneously. In some cases, the methods and compositions may be used for sequencing the genome and transcriptome from a single cell in parallel. The methods may be useful to dissect the functional consequences of genetic variations.

[0406] A microcapsule (e.g., a bead) entrapping one or more magnetic particles may be used in the methods. The magnetic particles may not diffuse out of the microcapsule until the microcapsule is dissolved. The microcapsule may comprise an oligonucleotide comprising a DNA primer. For example, the DNA primer may be a genomic DNA primer. The DNA primer may bind to DNA molecules from a cell. The DNA primer may be used to amplify and/or sequence DNA molecules from a cell. DNA primers may be entrapped and/or bound to the microcapsule and released when the microcapsule is dissolved.

[0407] The magnetic particles entrapped within the microcapsule may comprise an oligonucleotide comprising an RNA primer. The RNA primer may bind to RNA molecules from a cell. In some cases, the RNA primer is an mRNA primer that binds to the mRNA molecules from the cell. For example, the mRNA primer may comprise a poly-T sequence that binds to the poly-A sequence of the mRNA molecules from the cell. FIG. **31** shows a microcapsule with a barcoded magnetic particle entrapped.

[0408] The magnetic particles may be made from materials such as iron oxide (e.g., superparamagnetic iron oxide), ferromagnetic, ferrimagnetic, or paramagnetic materials. Ferromagnetic materials may be strongly susceptible to magnetic fields and capable of retaining magnetic properties when the field can be removed. Ferromagnetic materials include, but are not limited to, iron, cobalt, nickel, alloys

thereof, and combinations thereof. Other ferromagnetic rare earth metals or alloys thereof can also be used to make the magnetic particles.

[0409] The oligonucleotides on both the microcapsule and the magnetic particle may comprise the same barcode sequence. The barcode sequence may allow matching the information (e.g., sequence reads) of DNA and RNA from the same cell.

[0410] In some cases, the barcode sequence may comprise a unique identifier of the cell. For example, the unique identifier may distinguish a cell from other cells in a sample. Thus, the unique identifier may allow parallel analysis of DNA molecules and RNA molecules in a plurality of cells, e.g., at least 10, 50, 100, 200, 300, 400, 500, 600, 800, or 1000 cells. For example, the unique identifier may allow parallel analysis of DNA molecules and RNA molecules in a plurality of cells, e.g., at least 200, or 500 cells.

[0411] In some cases, the microcapsule may also contain one or more reagents for analyzing cells. For example, the microcapsule may contain a lysis agent. When the microcapsule is dissolved, the lysis agent may be released and lyse the cell in the same partition with the microcapsule.

[0412] In some cases, the microcapsule may be a gel bead. An example method for making a gel bead with one or more magnetic particles may comprise one or more of the following operations: 1) Magnetic particles are added to the aqueous phase of the material for making the gel beads, e.g., the gel beads monomer mixture; 2) The gel beads are made using a microfluidic approach, e.g., by forming droplets that polymerize to form the gel beads. When the droplets polymerize, the magnetic particles are entrapped within; 3) The same barcode sequence is added to the gel bead and the magnetic particles entrapped within, e.g., using dual ligation strategy.

[0413] Once a partition is generated to include a cell, a microcapsule, and a magnetic particle entrapped in the microcapsule, the partition may be incubated with one or more reagents (e.g., a lysis agent) to lyse the cell and dissolve the microcapsule. The incubation may be performed on a microfluidic chip device, e.g., with a delay line device as described in Frenz et al., Reliable microfluidic on-chip incubation of droplets in delay-lines. Lab Chip. 2009 May 21; 9(10):1344-8, which is incorporated herein by reference in its entirety. After the incubation, the partition may be collected and placed in a container e.g., a strip tube or plate.

[0414] The incubation may be performed for a period that allows sufficient time for the cell to lyse and the magnetic particles to be released from the microcapsule. The incubation time may also allow sufficient binding of the RNA primers on the magnetic particles with the RNA molecules from the cell. In some cases, the incubation time may be from 1 minute to 100 minutes, from 5 minutes to 50 minutes, from 10 minutes to 30 minutes, or from 10 minutes to 20 minutes.

[0415] One or more RNA molecules bound to the RNA primers on the magnetic particles may be separated from other components in the partition. The separation may be performed by concentrating the magnetic particles. The magnetic particles may be concentrated by a magnetic field. The separation may be performed on a microfluidic device, e.g., a device as described in Gao et al., Wash-free magnetic immunoassay of the PSA cancer marker using SERS and droplet microfluidics, Lab Chip, 2016, 16, 1022-1029; Brouzes et al., Rapid and continuous magnetic separation in droplet microfluidic devices. Lab Chip. 2015 Feb. 7; 15(3): 908-19; or Lombardi et al., Droplet microfluidics with magnetic beads: a new tool to investigate drug-protein interactions. Anal Bioanal Chem. 2011 January; 399(1):347-52, which are incorporated herein by reference in their entireties. In some cases, the one or more RNA molecules may be separated from DNA molecules. The separated RNA molecules and DNA molecules from a single cell may be analyzed using approaches described herein, e.g., sequencing, to determine a characteristic of the cell.

[0416] FIG. **32** shows a method for parallel sequencing DNA (e.g., genomic DNA) and RNA (e.g., mRNA) in a cell. In operation **3210**, single cell partitions are prepared by mixing gel beads with magnetic particles, cells and reaction reagents, e.g., a lysis agent. Droplets are generated from the mixture. A single droplet **3220** contains one cell, a gel bead with magnetic particles, and reaction reagents. The gel bead has genomic DNA primers and the magnetic particles have mRNA primers. The gel bead and the magnetic particles in the partition have the same barcode sequence. In **3230**, the gel bead is dissolved to release the magnetic particles and genomic DNA primers. The cell is also lysed to release the genomic DNA molecules and mRNA molecules. The mRNA molecules are captured on the magnetic particles by binding with the mRNA primers. In operation **3240**, on a microfluidic device, the partition split into two daughter droplets. The magnetic particles with the captured mRNA molecules are collected in only one of the daughter droplets, thus being separated from other components, e.g., genomic DNA in the other daughter droplet. Thus, the genomic DNA molecules and mRNA molecules from a single cell are separated and may be used for further analysis.

[0417] Also provided herein are methods and compositions for analyzing one or more proteins and one or more nucleic acids from a sample (e.g., a single cell). For example, the methods and compositions may be used for analyzing the proteome, the genome and/or the transcriptome in a single cell. The methods may comprise generating a partition that contains the sample, a labelling agent for proteins and a labelling agent for nucleic acids. In some cases, the labelling agent for proteins may interact with one or more proteins in the sample. For example, the labelling agent for proteins may comprise an antibody. In other cases, the labelling agent for proteins may be coupled with a protein probe that interacts with one or more proteins in the sample. For example, the labelling agent for proteins may be coupled with an antibody. The labelling agent for nucleic acids may interact with one or more nucleic acids in the sample. The labelling agent for nucleic acids may comprise a primer, e.g., a primer that bind to a DNA molecule and/or RNA molecule. The labelling agent for proteins and the labelling agent for nucleic acids may comprise the same reporter oligonucleotide. The reporter oligonucleotide may comprise a barcode and/or a UMI. The barcode and/or the UMI may allow for matching proteins with nucleic acids from the same sample. When bound to the labelling agent for nucleic acids, the nucleic acids from a sample may be sequenced. The reporter oligonucleotide or a portion thereof may also be sequenced. In some cases, the methods may comprise one or more of the following operations: a) providing a partition comprising a biological sample comprising a protein and a first nucleic acid molecule, a labelling agent that is (i) capable of binding to the protein and (ii) is

coupled to a reporter oligonucleotide comprising a nucleic acid barcode sequence that permits identification of the labelling agent, a first anchor oligonucleotide coupled to a support, which first anchor oligonucleotide is capable of interacting with the reporter oligonucleotide; and a second anchor oligonucleotide coupled to the support, which second anchor oligonucleotide is capable of interacting with the first nucleic acid molecule; (b) in the partition, synthesizing a second nucleic acid molecule comprising at least a portion of the nucleic acid barcode sequence or a complement thereof, and (c) subjecting the first nucleic acid molecule and the second nucleic acid molecule to sequencing. When the labelling agent for proteins and a protein probe is separate molecules, the protein probe may be incubated with the sample before making the partition in operation (a).

[0418] Two anchor agents may be used in the methods described herein. The first anchor agent may interact with one or more nucleic acids from a sample. Additionally or alternatively, the first anchor agent may be coupled with a labelling agent for nucleic acids. For example, the first anchor agent may comprise an oligonucleotide that bind to a labelling agent for nuclei acid. The second anchor agent may interact with one or more proteins from a sample. Additionally or alternatively, the second anchor agent may interact be coupled with a labelling agent for proteins. For example, the second anchor agent may comprise an element that interacts with the labelling agent for proteins. In some cases, the second anchor agent may comprise a nucleic acid sequence that interacts with an oligonucleotide sequence coupled to a labelling agent for proteins.

[0419] The labelling agent for proteins may comprise one or more elements. The labelling agent may comprise an element (e.g., an oligonucleotide sequence) that interacts with an anchor agent. The labelling agent may comprise a reporter oligonucleotide, e.g., an oligonucleotide comprising a barcode that allows for identifying the protein targeted by the labelling agent. For example, in the cases where the labelling agent for proteins comprises an antibody, the reporter oligonucleotide may allow for identifying the antibody, thereby identifying the protein bound by the antibody.

[0420] The labelling agent for proteins may comprise a reactive moiety that allows the labelling agent to be coupled with a protein probe, e.g., antibody. The labelling agent may be coupled with a protein probe by any chemistry descried herein for attaching a reporter oligonucleotide to a labelling agent. In some cases, the reactive moiety may include a click chemistry linker, such as Methyltetrazine-PEG5-NHS Ester or TCO-PEG4-NHS Ester. The reactive moiety on the labelling agent may also include amine for targeting aldehydes, amine for targeting maleimide (e.g., free thiols), azide for targeting click chemistry compounds (e.g., alkynes), biotin for targeting streptavidin, phosphates for targeting EDC, which in turn targets active ester (e.g., NH_2). The reactive moiety on the protein probe may be a chemical compound or group that binds to the reactive moiety on the labelling agent. Example strategies to conjugate the protein probe to the labelling agent include using of commercial kits (e.g., Solulink, Thunder link), conjugation of mild reduction of hinge region and maleimide labelling, stain-promoted click chemistry reaction to labeled amides (e.g., copper-free), and conjugation of periodate oxidation of sugar chain and amine conjugation. In the cases where the protein probe is an antibody, the antibody may be modified for conjugating the reporter oligonucleotide. For example, the antibody may be glycosylated with a substrate-permissive mutant of β-1, 4-galactosyltransferase, GalT (Y289L) and azide-bearing uridine diphosphate-N-acetylgalactosamine analog uridine diphosphate—GalNAz. The modified antibody may be conjugated with a reporter oligonucleotide with a dibenzocyclooctyne-PEG4-NHS group. FIG. **33** shows example strategies for antibody-reporter oligonucleotide conjugation. In some cases, some strategy (e.g., COOH activation (e.g., EDC) and homobifunctional cross linkers) may be avoided to prevent the protein probes from conjugating to themselves.

[0421] The two anchor agents may be coupled to a solid support, e.g., a microcapsule. For example, the microcapsule may be a bead, e.g., a gel bead. In some cases, the two anchor agents are coupled to the same solid support. In other cases, the two anchor agents are coupled to different solid supports. The two anchor agent may comprise the same reporter oligonucleotide.

[0422] FIG. **34** shows example reagents used in the methods. An anchor agent **3420** is coupled to a bead **3410**. The anchor agent comprises a barcode sequence **3422** and a UMI **3423**. The anchor agent also comprises an oligonucleotide sequence **3424** that allows binding to the labelling agent **3430**. The labelling agent **3430** comprises an oligonucleotide **3431** for binding to the anchor agent. The labelling agent **3430** also comprises a barcode **3432** that allows identifying the antibody it is coupled to. The labelling agent **3430** further comprises a reactive moiety **3434** that allows the labelling agent to couple with an antibody **3440**.

[0423] An additional example of reagents and schemes suitable for analysis of barcoded labelling agents is shown in panels I and II of FIG. **52**B. As shown in FIG. **52**B (panel I), a labelling agent (e.g., antibody, an MHC moiety) **5201** is directly (e.g., covalently bound, bound via a protein-protein interaction, such as with Protein G) coupled to an oligonucleotide **5202** comprising a barcode sequence **5203** that identifies the label agent **5201**. Oligonucleotide **5202** also includes additional sequences (sequence **5204** comprising a reverse complement of a template switch oligo and sequence **5205** comprising a PCR handle) suitable for downstream reactions. FIG. **52**B (panel I) also shows an additional oligonucleotide **5206** (e.g., which may have been released from a bead as described elsewhere herein) comprising a barcode sequence **5208**, a UMI sequence **5209** and additional sequences (sequence **5207** comprising a sequencing read primer binding site 'pR1' and sequence **5210** comprising a template switch oligo) suitable for downstream reactions. During analysis, the labelling agent is bound to its target cell surface feature and the rGrGrG sequence of sequence **5210** hybridizes with sequence **5204** and both oligonucleotides **5202** and **5206** are extended via the action of a polymerizing enzyme (e.g., a reverse transcriptase, a polymerase), where oligonucleotide **5206** then comprises complement sequences to oligonucleotide **5202** at its 3' end. These constructs can then be optionally processed as described elsewhere herein and subject to sequencing to, for example, identify the target cell surface feature (via the complementary barcode sequence generated from oligonucleotide **5202**) and associate it with the cell, identified by the barcode sequence of oligonucleotide **5206**.

[0424] In another example, shown in FIG. **52**B (panel II), a labelling agent (e.g., antibody) **5221** is indirectly (e.g., via hybridization) coupled to an oligonucleotide **5222** comprising a barcode sequence **5223** that identifies the label agent

5221. Labelling agent **5221** is directly (e.g., covalently bound, bound via a protein-protein interaction, such as with Protein G) coupled to a hybridization oligonucleotide **5232** that hybridizes with sequence **5231** of oligonucleotide **5222**. Hybridization of oligonucleotide **5232** to oligonucleotide **5231** couples label agent **5221** to oligonucleotide **5222**. Oligonucleotide **5222** also includes additional sequences (sequence **5224** comprising a reverse complement of a template switch oligo and sequence **5225** comprising a PCR handle) suitable for downstream reactions. FIG. **52**B (panel II) also shows an additional oligonucleotide **5226** (e.g., which may have been released from a bead as described elsewhere herein) comprising a barcode sequence **5228**, a UMI sequence **5229** and additional sequences (sequence **5227** comprising a sequencing read primer binding site 'pR1' and sequence **5220** comprising a template switch oligo) suitable for downstream reactions. During analysis, the labelling agent is bound to its target cell surface feature and the rGrGrG sequence of sequence **5220** hybridizes with sequence **5224** and both oligonucleotides **5222** and **5226** are extended via the action of a polymerizing enzyme (e.g., a reverse transcriptase, a polymerase), where oligonucleotide **5226** then comprises complement sequences to oligonucleotide **5222** at its 3' end. These constructs can then be optionally processed as described elsewhere herein and subject to sequencing to, for example, identify the target cell surface feature (via the complementary barcode sequence generated from oligonucleotide **5222**) and associate it with the cell, identified by the barcode sequence of oligonucleotide **5226**.

[0425] An example of the methods for analyzing mRNA molecules and proteins from a single cell is shown in FIGS. **35**A and **35**B. The method uses a barcoded antibody **3510** containing an antibody **3511** conjugated with an oligonucleotide **3512**. The oligonucleotide **3512** can bind to a first anchor oligonucleotide **3520** coupled to a bead. The barcoded antibody **3510** is incubated with cells such that the antibody binds to an antigen on the cell, and form antibody-cell complexes (FIG. **35**A). Unbound antibodies are washed out. The antibody-cell complexes are made into emulsion partitions. Each partition contains an antibody-cell complex, the first anchor oligonucleotide **3520**, and a second anchor oligonucleotide **3530** that binds to mRNA molecules from the cell. The cell is lysed and the mRNA molecules are released from the cell. As shown in FIG. **35**B, the mRNA molecules are reverse transcribed to cDNA and amplified with the help of the second anchor oligonucleotide. The amplified cDNA molecules have the barcode and UMI that are the same as the barcode and UMI on the first anchor oligonucleotide **3520**. Primer extension is performed on the complex of the first anchor oligonucleotide **3520** and the oligonucleotide **3512**, thus generating a reporter oligonucleotide **3550** comprising the barcode and UMI the same as those on the second anchor oligonucleotide. The reporter oligonucleotide **3550** also comprises an antibody identifier (antibody barcode (AbBC)) that identifies the antibody and the antigen bound by the antibody. When the cDNA molecules are sequenced, the sequence reads are correlated to the antigen in the same cell using the barcode and UMI. FIG. **35**C shows the primer extension of the first anchor oligonucleotide and oligonucleotide **3512** conjugated with the antibody. The resulting oligonucleotides may be separated from cDNA synthesized from mRNA from the cell (e.g., by size-based selection). The first anchor oligonucleotide and the complex of the second anchor oligonucleotide with oligonucleotide **3512** may be processed and/or sequenced separately or jointly. In some cases, the anchor agents **3520** and **3530** may be coupled to the same bead.

[0426] Also provided herein are the microfluidic devices used for partitioning the cells as described above. Such microfluidic devices can comprise channel networks for carrying out the partitioning process like those set forth in FIGS. **1** and **2**. Briefly, these microfluidic devices can comprise channel networks, such as those described herein, for partitioning cells into separate partitions, and co-partitioning such cells with oligonucleotide barcode library members, e.g., disposed on beads. These channel networks can be disposed within a solid body, e.g., a glass, semiconductor or polymer body structure in which the channels are defined, where those channels communicate at their termini with reservoirs for receiving the various input fluids, and for the ultimate deposition of the partitioned cells, etc., from the output of the channel networks. By way of example, and with reference to FIG. **2**, a reservoir fluidly coupled to channel **202** may be provided with an aqueous suspension of cells **214**, while a reservoir coupled to channel **204** may be provided with an aqueous suspension of beads **216** carrying the oligonucleotides. Channel segments **206** and **208** may be provided with a non-aqueous solution, e.g., oil, into which the aqueous fluids are partitioned as droplets at the channel junction **212**. An outlet reservoir may be fluidly coupled to channel **210** into which the partitioned cells and beads can be delivered and from which they may be harvested. As will be appreciated, while described as reservoirs, it will be appreciated that the channel segments may be coupled to any of a variety of different fluid sources or receiving components, including tubing, manifolds, or fluidic components of other systems.

[0427] Also provided are systems that control flow of these fluids through the channel networks e.g., through applied pressure differentials, centrifugal force, electrokinetic pumping, capillary or gravity flow, or the like.

[0428] Also provided herein are kits for analyzing individual cells or small populations of cells. The kits may include one, two, three, four, five or more, up to all of partitioning fluids, including both aqueous buffers and non-aqueous partitioning fluids or oils, nucleic acid barcode libraries that are releasably associated with beads, as described herein, labelling agents, as described herein, anchor oligonucleotides, as described herein, microfluidic devices, reagents for disrupting cells amplifying nucleic acids, and providing additional functional sequences on fragments of cellular nucleic acids or replicates thereof, as well as instructions for using any of the foregoing in the methods described herein.

[0429] Another aspect of the disclosure provides a composition for characterizing a plurality of analytes, comprising a partition comprising a plurality of barcode molecules and the plurality of analytes. The plurality of barcode molecules can also include at least 1,000 barcode molecules. Moreover, (i) a first individual barcode molecule of the plurality of barcode molecules can comprise a first nucleic acid barcode sequence that is capable of coupling to a first analyte of the plurality of analytes; and (ii) a second individual barcode molecule of the plurality of barcoded molecules can comprise a second nucleic acid barcode sequence that is capable of coupling to a second analyte of the plurality of analytes, where the first analyte and the second

analyte are different types of analytes (e.g., DNA and RNA, DNA and protein, RNA and protein, or DNA, RNA and protein). In some cases, the composition comprises a plurality of partitions comprising the partition.

[0430] An additional aspect of the disclosure provides a method for analyte characterization. The method comprises: (a) providing a plurality of partitions, where a given partition of the plurality of partitions comprises a plurality of barcode molecules and a plurality of analytes. The plurality of barcode molecules can comprise at least 1,000 barcode molecules. Moreover, (i) a first individual barcode molecule of the plurality of barcode molecules can comprise a first nucleic acid barcode sequence that is capable of coupling to a first analyte of the plurality of analytes; and (ii) a second individual barcode molecule of the plurality of barcoded molecules can comprise a second nucleic acid barcode sequence that is capable of coupling to a second analyte of the plurality of analytes; where the first analyte and the second analyte are different types of analytes. The method also includes (b) in said given partition (i) synthesizing a first nucleic acid molecule comprising at least a portion of the first nucleic acid barcode sequence or complement thereof, and (ii) synthesizing a second nucleic acid molecule comprising at least a portion of the second nucleic acid barcode sequence or complement thereof, and (c) removing said first nucleic acid molecule and said second molecule from said given partition. In some cases, the method further comprises subjecting the first nucleic acid molecule and the second nucleic acid molecule, or a derivative of the first nucleic acid molecule and/or second nucleic acid molecule, to sequencing to characterize the first and/or the second analyte.

[0431] Characterizing the first analyte and/or the second analyte generally provides information regarding the first analyte and/or second analyte. This information can be used to select first and/or second analytes for one or more additional cycles of (a)-(c). Accordingly, the method may further comprise repeating (a)-(c) based on a characterization of the first analyte or the second analyte from sequencing. In some cases, the method further comprises selecting the first analyte and/or the second analyte based on a characterization of the first analyte or the second analyte obtained from the sequencing a subsequent sequencing upon repeating (a)-(c).

[0432] Moreover, in some cases, (b) further comprises: (1) synthesizing the first nucleic acid molecule comprising at least a portion of the first nucleic acid barcode sequence or complement thereof, and (2) synthesizing the second nucleic acid molecule comprising at least a portion of the second nucleic acid barcode sequence or complement thereof. For example, the first nucleic acid molecule and/or the second nucleic acid molecules may be synthesized with the aid of one or more primer extension reactions that make use of a primer that hybridizes with a first or second analyte. Such a primer may comprise a barcode sequence and/or a UMI sequence as described elsewhere herein. In some cases, the first nucleic acid molecule and/or the second nucleic acid molecule may be synthesized with the aid of ligation between two nucleic acid molecules.

[0433] In some cases, the method further comprises performing one or more reactions subsequent to removing the first nucleic acid molecule and the second nucleic acid molecule from the given partition. Such reactions can include the addition of additional nucleic acid sequences (e.g., sample index sequences, a sequence for function in a particular sequencing platform) via additional primer extension reactions, nucleic acid amplification schemes (e.g., PCR) or ligation. In some cases, portions of the first and/or second nucleic acid molecules may be removed (e.g., via restriction enzymes, via shearing) prior to or after the addition of additional nucleic acid sequences. Moreover, these reactions can be performed in bulk, such that processing of the first and second nucleic acid molecules and first and second nucleic acid molecules from other partitions are processed simultaneously in bulk. Such processing can be completed in a single pot reaction. Examples of such one or more other reactions are provided in U.S. Patent Publication No. 2015/0376609, which is entirely incorporated herein by reference.

[0434] An additional aspect of the disclosure provides a system for characterizing a plurality of analytes. The system comprises a partitioning unit for providing a partition comprising a plurality of barcode molecules and the plurality of analytes, where: (i) a first individual barcode molecule of the plurality of barcode molecules comprises a first nucleic acid barcode sequence and is capable of coupling to a first analyte of the plurality of analytes; and (ii) a second individual barcode molecule of the plurality of barcode molecules comprises a second nucleic acid barcode sequence and is capable of coupling to a second analyte of the plurality of analytes, where the first analyte and the second analyte are different types of analytes. The system also can include a controller coupled to the partitioning unit, where the controller is programmed to: (i) direct the partitioning unit to provide the partition; subject the partition to conditions that are sufficient to: (1) synthesize a first nucleic acid molecule comprising at least a portion of the first nucleic acid barcode sequence or complement thereof, and (2) synthesize a second nucleic acid molecule comprising at least a portion of the second nucleic acid barcode sequence or complement thereof. Sequencing of the first nucleic acid molecule and the second nucleic acid molecule, or derivatives thereof, can characterize the first analyte or the second analyte. In some cases, the partitioning unit can provides a plurality of partitions comprising the partition.

[0435] In some cases, the partitioning unit comprises a multi-well plate. In some cases, the partitioning unit comprises a plurality of channels, which may be microfluidic channels. The plurality of channels may come together to form at least one channel junction that provides the partition. In some cases, a partitioning unit may comprise a first (i) a first channel fluidically connected to the at least one channel junction and configured to provide a first fluid to the at least one channel junction; (ii) and a second channel fluidically connected to the at least one channel junction and configured to provide a second fluid, immiscible with the first fluid, to the at least one channel junction. In an example, then first channel may be configured to provide an aqueous phase comprising aqueous phase reagents (e.g., nucleic acids, including barcoded nucleic acids, labelling agents, beads, an agent that can degrade beads, amplification/primer extension reagents, sample nucleic acids, cells, cell lysis reagents, etc.) and the second channel may be configured to provide an oil phase comprising an oil (e.g., an oil comprising a fluorosurfactant) that is immiscible with the aqueous phase. Upon contact of the aqueous phase with the oil phases, aqueous phase droplets comprising aqueous phase reagents are generated.

[0436] In various aspects, the partition or the given partition may comprise at least 1,000 barcode molecules, at least 2,500 barcode molecules at least 5,000 barcode molecules, at least 7,500 barcode molecules, at least 10,000 barcode molecules, at least 20,000 barcode molecules, at least 30,000 barcode molecules, at least 50,000 barcode molecules, at least 60,000 barcode molecules, at least 70,000 barcode molecules, at least 80,000 barcode molecules, at least 90,000 barcode molecules, at least 100,000 barcode molecules, at least 200,000 barcode molecules, at least 300,000 barcode molecules, at least 400,000 barcode molecules, at least 500,000 barcode molecules, at least 600,000 barcode molecules, at least 700,000 barcode molecules, at least 800,000 barcode molecules, at least 900,000 barcode molecules, at least 1,000,000 barcode molecules, at least 2,500,000 barcode molecules, at least 5,000,000 barcode molecules, at least 7,500,000 barcode molecules at least 10,000,000 barcode molecules, at least 50,000,000 barcode molecules, at least 100,000,000 barcode molecules or more.

[0437] In various aspects, at least one of the first individual barcode molecule and the second individual barcode molecule may be coupled (e.g., via a covalent bond, via non-covalent interactions, via a labile bond, etc.) to a bead. In some cases, the bead comprises a gel bead and/or is degradable as described elsewhere herein. In methods described herein, the first or second barcode molecule can be released from the bead after a partition or partitions are provided. In some cases, release of a barcode molecule may occur prior to, simultaneous to, or following its use in barcoding a respective nucleic acid molecule. Where release happens after barcoding, barcoded constructs are initially coupled to the bead. Moreover, a partition may comprise an agent capable of degrading the bead. In some cases, such a reagent is a reducing agent that can reduce disulfide bonds of the bead and/or any disulfide linkages between species coupled to the bead and the bead itself. Moreover, in various aspects, the partition or a given partition can be any suitable partition such as a droplet among a plurality of droplets (e.g., droplets in an emulsion) or a well among a plurality of wells. Furthermore, in various aspects, the first nucleic acid barcode sequence and the second nucleic acid barcode sequence are identical.

[0438] In various aspects, the first analyte or the second analyte can be a nucleic acid molecule, including any type of nucleic acid molecule described elsewhere herein. For example, the nucleic acid molecule may be genomic deoxyribonucleic acid (gDNA). In another example, the nucleic acid molecule is messenger ribonucleic acid (mRNA).

[0439] Moreover, in various aspects, the first analyte or the second analyte is a labelling agent capable of coupling to a cell surface feature of a cell. The partition or the given partition can comprise the cell or one or more components of the cell (e.g., such as free cellular surface features remaining after cell lysis). In some cases, the partition or given partition comprises a single cell. The labelling agent can be any labelling agent, including a type of labelling agent described elsewhere herein including an antibody, or an epitope binding fragment thereof, a cell surface receptor binding molecule, a receptor ligand, a small molecule, a bi-specific antibody, a bi-specific T-cell engager, a T-cell receptor engager, a B-cell receptor engager, a pro-body, an aptamer, a monobody, an affimer, a darpin, a protein scaffold, an antigen, an antigen presenting particle and a major histocompatibility complex (MHC). Examples of cell surface features include a receptor, an antigen, a surface protein, a transmembrane protein, a cluster of differentiation protein, a protein channel, a protein pump, a carrier protein, a phospholipid, a glycoprotein, a glycolipid, a cell-cell interaction protein complex, an antigen-presenting complex, a major histocompatibility complex, a T-cell receptor, a B-cell receptor, a chimeric antigen receptor, a gap junction, an adherens junction and any other cell surface feature described elsewhere herein.

[0440] In some cases, cells are incubated in bulk with one or more labelling agents prior to partitioning of cells. The one or more labelling agents can be chosen such that they are directed to particular cell surface features of interest in a given assay. Upon binding of the one or more labeling agents to respective cell surface features, where present, the cells can then be washed to remove unbound labelling agents and the resulting cells then subject to partitioning.

[0441] Moreover, in some cases, the first individual barcode molecule or the second individual barcode molecule may be capable of coupling to the labelling agent via a third nucleic acid molecule coupled to the labelling agent. The third nucleic acid molecule can be coupled to the labelling agent and comprise a third nucleic acid barcode sequence that identifies the coupled labelling agent (and, thus, a cell surface feature to which the labelling agent is bound). In a primer extension reaction, the first individual barcode molecule or the second individual barcode molecule can be extended such that a complement of the third barcode sequence is added to the first or second individual barcode molecule. During sequencing, the first or second barcode sequence of these molecules can identify the partition from which the molecules were synthesized and, where a partition comprises a single cell, the third barcode sequence can associate a particular cell surface feature with that single cell.

[0442] In various aspects, the first analyte and second analyte can be different types of nucleic acid molecules. For example, the first analyte may be a deoxyribonucleic acid molecule (e.g., gDNA) and the second analyte may be ribonucleic acid molecule (e.g., mRNA), such as, for example, a transcript. Where implemented, a cell's genomic DNA and also the cell's transcriptome can be analyzed and characterized.

[0443] Moreover, where the first and second analytes are nucleic acid molecules, the first individual barcode molecule and/or the second individual barcode molecule may comprise a priming sequence capable of hybridizing to the first analyte and/or second analyte respectively. In addition to the first nucleic acid barcode molecule or the second nucleic acid barcode molecule, may also include a UMI sequence, that can be useful for identifying (and even quantifying) particular molecules that are barcoded within a given partition, as is described elsewhere herein.

[0444] In an example, schematically depicted in FIG. **46**A, a partition (e.g., a droplet, a well or any other type of partition described herein) comprises a bead **4601**, which is coupled (e.g., reversibly coupled) to barcoded oligonucleotides **4602** and **4603**. The bead **4601** and barcoded oligonucleotides **4602** and **4603** are schematically depicted in FIG. **46**A. Barcoded oligonucleotide **4602** comprises a first nucleic acid barcode sequence and a poly-T priming sequence **4604** that can hybridize with the poly-A tail of an mRNA transcript. Barcoded oligonucleotide **4602** may also comprise a UMI sequence that can uniquely identify a given

transcript. Barcoded oligonucleotide **4603** comprises a second nucleic acid barcode sequence and a random N-mer priming sequence **4605** that is capable of randomly hybridizing with gDNA. In this configuration, barcoded oligonucleotides **4602** and **4603** comprise the same nucleic acid barcode sequence, which permits association of downstream sequencing reads with the partition. In some cases, though, the first nucleic acid barcode sequence and the second nucleic acid barcode sequence are different.

[0445] The partition also comprises a cell (not shown) and lysis agents that aid in releasing nucleic acids from the cell and can also include an agent (e.g., a reducing agent) that can degrade the bead and/or break a covalent linkage between the barcoded oligonucleotides **4602** and **4603** and bead **4601**, releasing them into the partition. The released barcoded oligonucleotide **4602** can hybridize with mRNA released from the cell and the released barcoded oligonucleotide **4603** can hybridize with gDNA released from the cell. Barcoded constructs A and B can then be generated for each of the mRNA and barcoded oligonucleotide **4623** as described elsewhere herein, such as via the action of a polymerase (and/or reverse transcriptase) and/or primer extension. Barcoded construct A can comprises a sequence corresponding to the original barcode sequence from the bead and a sequence corresponding to a transcript from the cell. Barcoded construct B can comprise a sequence corresponding to the original barcode sequence from the bead and a sequence corresponding to genomic DNA from the cell. The barcoded constructs can then be released/removed from the partition and, in some cases, further processed to add any additional sequences. The resulting constructs are then sequenced, sequencing data processed, and the results used to characterize the mRNA and the gDNA from the cell. Analysis can be completed, for example, as described elsewhere herein. The information received from the characterization can then be used in a subsequent analysis of another cell in a partition. Moreover, barcoded oligonucleotides **4602** and **4603** can be designed to prime any particular type of nucleic acid, including those that are not derived from a cell. Moreover, the priming sequences shown in FIG. **46**A are for example purposes only and are not meant to be limiting.

[0446] In various aspects, the first analyte may be a nucleic acid molecule (e.g., deoxyribonucleic acid (e.g., gDNA), ribonucleic acid (e.g., mRNA), a transcript) and the second analyte a labelling agent capable of coupling to a cell surface feature. In such a case, the first individual barcode molecule may comprise a priming sequence capable of hybridizing to the nucleic acid molecule and may also include a UMI sequence. Moreover, the second individual barcode molecule may comprise a priming sequence capable of hybridizing with a third nucleic acid molecule coupled to the labelling agent. As noted elsewhere herein, this third nucleic acid molecule can include a barcode sequence that identifies the labelling agent. It may also include a UMI sequence. The labelling agent can be any suitable labelling agent, including a type of example labelling agents described elsewhere herein, and may be targeted to any suitable cell surface feature to which it can selectively bind. Non-limiting examples of such cell surface features are provided elsewhere herein. Furthermore, in some cases, the partition comprises a cell having the cell surface feature and, in some cases, may comprise only one cell.

[0447] In an example, schematically depicted in FIG. **46**B, a partition (e.g., a droplet, a well, a microcapsule, or any other type of partition described herein) comprises a bead **4611**, which is coupled (e.g., reversibly coupled) to barcoded oligonucleotides **4612** and **4613**. The bead **4611** and barcoded oligonucleotides **4612** and **4613** are schematically depicted in FIG. **46**B. Barcoded oligonucleotide **4612** comprises a first nucleic acid barcode sequence and a poly-T priming sequence **4614** that can hybridize with the poly-A tail of an mRNA transcript. Barcoded oligonucleotide **4612** may also comprise a UMI sequence that can uniquely identify a given transcript. Barcoded oligonucleotide **4613** comprises a second nucleic acid barcode sequence and a targeted priming sequence that is capable of specifically hybridizing with a barcoded oligonucleotide **4623** (via a complementary portion **4624** of barcoded oligonucleotide **4623** coupled to an antibody **4621** that is bound to the surface of a cell **4622**. Barcoded oligonucleotide **4623** comprises a barcode sequence that uniquely identifies the antibody **4621** (and thus, the particular cell surface feature to which it is bound). In this configuration, barcoded oligonucleotides **4612** and **4613** comprise the same nucleic acid barcode sequence, which permit downstream association of barcoded nucleic acids with the partition. In some cases, though, the first nucleic acid barcode sequence and the second nucleic acid barcode sequence are different. Furthermore, barcoded labelling agents, including antibodies, may be produced by any suitable route, including via example coupling schemes described elsewhere herein.

[0448] As shown in FIG. **46**B, the partition also comprises cell **4622**, lysis agents that aid in releasing nucleic acids from the cell **4622** and can also include an agent (e.g., a reducing agent) that can degrade the bead and/or break a covalent linkage between the barcoded oligonucleotides **4612** and **4613** and bead **4611**, releasing them into the partition. The released barcoded oligonucleotide **4612** can hybridize with mRNA released from the cell and the released barcoded oligonucleotide **4613** can hybridize with barcoded oligonucleotide **4623**. Barcoded constructs A and B can then be generated for each of the mRNA and barcoded oligonucleotide **4623** as described elsewhere herein, such as via the action of a polymerase (and/or reverse transcriptase) and/or primer extension. Barcoded construct A may comprise a sequence corresponding to the original barcode sequence from the bead and a sequence corresponding to a transcript from the cell. Barcoded construct B may comprise a sequence corresponding to the original barcode sequence from the bead and an additional sequence corresponding to the barcode sequence coupled to the labelling agent. The barcoded constructs can then be released/removed from the partition and, in some cases, further processed to add any additional sequences. The resulting constructs are then sequenced, sequencing data processed, and the results used to characterize the mRNA and cell surface feature of the cell. Analysis, for example, can be completed as described elsewhere herein. The information received from the characterization can then be used in a subsequent analysis of another cell in a partition. Moreover, the priming sequences shown in FIG. **46**B are for example purposes only and are not meant to be limiting. In addition, the scheme shown in FIG. **46**B may also be used for concurrent analysis of genomic DNA and cell surface features. In some cases, the partition comprises only one cell.

[0449] Furthermore, in various aspects, the first analyte may comprise a nucleic acid molecule with a nucleic acid sequence (mRNA, complementary DNA derived from reverse transcription of mRNA) encoding at least a portion of a V(D)J sequence of an immune cell receptor. Accordingly, a first barcode molecule may comprise a priming sequence that can prime such a nucleic acid sequence, as is described elsewhere herein. In some cases, the nucleic acid molecule with a nucleic acid sequence encoding at least a portion of a V(D)J sequence of an immune cell receptor is cDNA first generated from reverse transcription of the corresponding mRNA, using a poly-T containing primer. The cDNA that is generated can then be barcoded using a primer, comprising a barcode sequence (and optionally, a UMI sequence) that hybridizes with at least a portion of the cDNA that is generated. In some cases, a template switching oligonucleotide in conjunction a terminal transferase or a reverse transcriptase having terminal transferase activity may be employed to generate a priming region on the cDNA to which a barcoded primer can hybridize during cDNA generation. Terminal transferase activity can, for example, add a poly-C tail to a 3' end of the cDNA such that the template switching oligonucleotide can bind via a poly-G priming sequence and the 3' end of the cDNA can be further extended. The original mRNA template and template switching oligonucleotide can then be denatured from the cDNA and the barcoded primer comprising a sequence complementary to at least a portion of the generated priming region on the cDNA can then hybridize with the cDNA and a barcoded construct comprising the barcode sequence (and any optional UMI sequence) and a complement of the cDNA generated. Additional methods and compositions suitable for barcoding cDNA generated from mRNA transcripts including those encoding V(D)J regions of an immune cell receptor and/or barcoding methods and composition including a template switch oligonucleotide are described in U.S. Provisional Patent Application Ser. No. 62/410,326, filed Oct. 19, 2016 and U.S. Provisional Patent Application Ser. No. 62/490,546, filed Apr. 26, 2017, both of which applications are herein incorporated by reference in their entireties. In one example, the scheme described elsewhere herein and schematically depicted in FIG. **19** may be used for V(D)J analysis.

[0450] V(D)J analysis may also be completed with the use of one or more labelling agents that bind to particular surface features of immune cells and are associated with barcode sequences as described elsewhere herein. In some cases, the one or more labelling agents comprise an MHC.

[0451] In some cases, different types of analytes do not include labelling agents directed to separate cell surface features of a cell.

[0452] Moreover, in various aspects, the first analyte may comprise a nucleic acid capable of functioning as a component of a gene editing reaction, such as, for example, clustered regularly interspaced short palindromic repeats (CRISPR)-based gene editing. Accordingly, the first barcode molecule may comprise a priming sequence that can prime such a nucleic acid sequence, as is described elsewhere herein.

[0453] While the examples described with respect to FIGS. **46**A and **46**B involve the analysis of two different types of analytes, these examples are not meant to be limiting. Any suitable number of analytes may be evaluated. Accordingly, in various aspects, there may be at least about 2, at least about 3, at least about 4, at least about 5, at least about 6, at least about 7, at least about 8, at least about 9, at least about 10, at least about 11, at least about 12, at least about 13, at least about 14, at least about 15, at least about 20, at least about 25, at least about 30, at least about 40, at least about 50, at least about 100 or more different analytes present in a partition, that can be subject to barcoded sequencing analysis. Higher number, multi-assay analysis can be completed by including primer species (one or more of which may be barcoded) that are capable of generating barcoded constructs and capable of specifically hybridizing with a particular analyte or oligonucleotide coupled to a labelling agent that is itself coupled to a particular analyte in the partition and subjecting the partition to suitable conditions for barcoding.

[0454] An example reagent for multi-assay analysis is schematically depicted in FIG. **46**C. As shown in FIG. **46**C, a partition can include a bead **4651** that is coupled to barcoded primers that can each participate in an assay of a different analyte. The bead **4651** is coupled (e.g., reversibly coupled) to a barcoded oligonucleotide **4652** that comprises a poly-T priming sequence **4654** for mRNA analysis and is also coupled (e.g., reversibly coupled) to barcoded oligonucleotide **4653** that comprises a random N-mer priming sequence **4655** for gDNA analysis. Moreover, bead **4651** is also coupled (e.g., reversibly coupled) to a barcoded oligonucleotide **4656** that can specifically bind an oligonucleotide coupled to a labelling agent (e.g., an antibody), via its targeted priming sequence **4657**. Bead **4651** is also coupled to a barcoded oligonucleotide **4658** that can specifically bind a nucleic acid molecule that can function in a CRISPR assay (e.g., CRISPR/Cas9), via its targeted priming sequence **4659**. In this example, each of the various barcoded primers comprises the same barcode sequence. Each barcoded oligonucleotide can be released from the bead **4651** within the partition and subject to conditions suitable for analysis of its respective analyte. In some cases, one or more of the analytes is associated with or derived from a cell, which itself, may be in the partition. In some cases, the partition comprises only one cell. Barcoded constructs A, B, C and D can be generated as described elsewhere herein and analyzed. Barcoded construct A may comprise a sequence corresponding to the barcode sequence from the bead and a DNA sequence corresponding to a target mRNA. Barcoded construct B may comprise a sequence corresponding to the barcode sequence from the bead and a sequence corresponding to genomic DNA. Barcoded construct C comprises a sequence corresponding to the barcode sequence from the bead and a sequence corresponding to barcode sequence associated with an antibody labelling agent. Barcoded construct D comprises a sequence corresponding to the barcode sequence from the bead and a sequence corresponding to a CRISPR nucleic acid (which, in some embodiments, also comprises a barcode sequence). Each construct can be analyzed via sequencing and the results associated with the given cell from which the various analytes originated. While only four different barcoded constructs are shown in FIG. **46**C, barcoded (or even non-barcoded) constructs can be tailored for analyses of any given analyte associated with a nucleic acid and capable of binding with such a construct.

[0455] For example, a partition can include a bead (e.g., a gel bead) that is coupled (e.g., reversibly coupled) to barcoded oligonucleotides that can participate in an assay of at least two different analytes. See FIG. **46**A for an exemplary

bead coupled to a barcoded oligonucleotide **4602** that comprises a poly-T priming sequence **4604** for mRNA analysis and a barcoded oligonucleotide **4603** that comprises a random N-mer priming sequence **4605** for gDNA analysis. See FIG. **46**B for an exemplary bead coupled to a barcoded oligonucleotide **4612** that comprise a poly-T priming sequence **4614** for mRNA analysis and a barcoded oligonucleotide **4613** that comprises a capture sequence **4615** that can specifically bind an oligonucleotide coupled to a labelling agent (e.g., an antibody), via its targeted priming sequence **4624**.

[0456] Additional exemplary assays for measuring at least two different analytes include a bead coupled to a barcoded oligonucleotide (e.g., **4602**) that comprises a poly-T priming sequence (e.g., **4604**) for mRNA analysis and a barcoded oligonucleotide (e.g., **4658**) that comprises a capture sequence **4659** that can specifically bind a nucleic acid molecule that can function in a CRISPR assay (e.g., CRISPR/Cas9), via its targeted priming sequence (see, e.g., FIGS. **61**A-**61**D). Further exemplary assays for measuring at least two different analytes include a bead coupled to a barcoded oligonucleotide (e.g., **4613**) that comprises a capture sequence (e.g., **4615**) that can specifically bind an oligonucleotide coupled to a labelling agent (e.g., an antibody), via its targeted priming sequence (e.g., **4624**) and a barcoded oligonucleotide (e.g., **4603**) that comprises a random N-mer priming sequence (e.g., **4605**) for gDNA analysis. Additional exemplary assays for measuring at least two different analytes include a bead coupled a barcoded oligonucleotide (e.g., **4613**) that comprises a capture sequence (e.g., **4615**) that can specifically bind an oligonucleotide coupled to a labelling agent (e.g., an antibody), via its targeted priming sequence (e.g., **4624**) and a barcoded oligonucleotide (e.g., **4658**) that comprises a capture sequence (e.g., **4659**) that can specifically bind a nucleic acid molecule that can function in a CRISPR assay (e.g., CRISPR/Cas9), via its targeted priming sequence (see, e.g., FIGS. **61**A-**61**D). Further exemplary assays for measuring at least two different analytes include a bead coupled a barcoded oligonucleotide (e.g., **4603**) that comprises a random N-mer priming sequence (e.g., **4605**) for gDNA analysis and a barcoded oligonucleotide (e.g., **4658**) that comprises a capture sequence (e.g., **4659**) that can specifically bind a nucleic acid molecule that can function in a CRISPR assay (e.g., CRISPR/Cas9), via its targeted priming sequence (see, e.g., FIGS. **61**A-**61**D).

[0457] For example, a partition can include a bead (e.g., a gel bead) that is coupled (e.g., reversibly coupled) to barcoded oligonucleotides that can participate in an assay of at least three different analytes. See FIG. **46**D for an exemplary bead **4660** coupled to a barcoded oligonucleotide **4661** that comprises a poly-T priming sequence **4662** for mRNA analysis; a barcoded oligonucleotide **4663** that comprises a random N-mer priming sequence **4664** for gDNA analysis; and a barcoded oligonucleotide **4665** that comprises a capture sequence **4666** that can specifically bind an oligonucleotide coupled to a labelling agent (e.g., an antibody), via its targeted priming sequence (e.g., **4624**). See FIG. **46**E for an exemplary bead **4667** coupled to a barcoded oligonucleotide **4661** that comprises a poly-T priming sequence **4662** for mRNA analysis; a barcoded oligonucleotide **4665** that comprises a capture sequence **4666** that can specifically bind an oligonucleotide coupled to a labelling agent (e.g., an antibody), via its targeted priming sequence (e.g., **4624**); and a barcoded oligonucleotide **4672** that comprises a capture sequence **4673** that can specifically bind a nucleic acid molecule that can function in a CRISPR assay (e.g., CRISPR/Cas9), via its targeted priming sequence (see, e.g., FIGS. **61**A-**61**D).

[0458] Additional exemplary assays for measuring at least three different analytes include a bead coupled to a barcoded oligonucleotide (e.g., **4661**) that comprises a poly-T priming sequence (e.g., **4662**) for mRNA analysis; a barcoded oligonucleotide (e.g., **4663**) that comprises a random N-mer priming sequence (e.g., **4664**) for gDNA analysis; and a barcoded oligonucleotide (e.g., **4672**) that comprises a capture sequence (e.g., **4673**) that can specifically bind a nucleic acid molecule that can function in a CRISPR assay (e.g., CRISPR/Cas9), via its targeted priming sequence (see, e.g., FIGS. **61**A-**61**D).

Parallel Analysis of Cell Samples

[0459] Provided herein are methods, systems, and compositions for analysis of a plurality of samples in parallel. The samples can comprise cells, cell beads, or in some cases, cellular derivatives (e.g., components of cells, such as cell nuclei, or matrices comprising cells or components thereof, such as cell beads). A cell bead can be a biological particle and/or one or more of its macromolecular constituents encased inside of a gel or polymer matrix, such as via polymerization of a droplet containing the biological particle and precursors capable of being polymerized or gelled. In an aspect, the present disclosure provides a method of analyzing nucleic acids (e.g., deoxyribonucleic acids (DNAs) or ribonucleic acid (RNAs)) of a plurality of different cell samples. The method may comprise labeling cells and/or cell beads of one or more different cell samples using a plurality of nucleic acid barcode molecules to yield a plurality of labeled cell samples, wherein an individual nucleic acid barcode molecule of the plurality of nucleic acid barcode molecules comprises a sample barcode sequence (e.g., a moiety-conjugated barcode molecule, also referred to herein as a feature barcode), and wherein nucleic acid barcode molecules of a given labeled cell sample are distinguishable from nucleic acid barcode molecules of another labeled cell sample by the sample barcode sequence. Nucleic acid molecules of the plurality of labeled cell samples may then be subjected to one or more reactions to yield a plurality of nucleic acid barcode products, wherein an individual nucleic acid barcode product of the plurality of nucleic acid barcode products comprises (i) a sample barcode sequence (e.g., a nucleic acid barcode sequence) and (ii) a sequence corresponding to a nucleic acid molecule of the plurality of labeled cell samples. The sequence corresponding to the nucleic acid molecule of the plurality of labeled cell samples may be, for example, a partition nucleic acid barcode molecule. The plurality of nucleic acid barcode products may be subjected to a sequencing reaction to yield a plurality of sequencing reads, which sequencing reads may be associated with individual labeled cell samples based on the sample barcode sequence, thereby analyzing nucleic acids of the plurality of different cell samples. In some embodiments, individual cells of a cell sample are labeled with two or more nucleic acid barcode molecules. In some cases, each of the two or more nucleic acid barcode molecules have unique barcode sequences (e.g., unique nucleic acid barcode sequences). In some cases, the barcode sequences of the two or more nucleic acid barcode mol-

ecules are not unique amongst the different cell samples but the combination of the barcode sequences of the two or more nucleic acid barcode molecules is a unique combination.

[0460] A nucleic acid barcode molecule can be used to label individual cells and/or cell beads of a cell sample. The label can be used in downstream processes, for example in sequencing analysis, as a mechanism to associate a cell and/or cell bead and a particular cell sample. For example, a plurality of cell samples (e.g., a plurality of cell samples from a plurality of different subjects (e.g., human or animal subjects), or a plurality of cell samples from a plurality of different biological fluids or tissues of a given subject, or a plurality of cell samples taken at different times from the same subject) can be uniquely labeled with nucleic acid barcode molecules such that the cells of a particular sample can be identified as originating from the particular sample, even if the particular cell sample was mixed with other cell samples and subjected to nucleic acid processing and/or sequencing in parallel. Accordingly, the present methods provide means of deconvoluting complex samples and enable massively parallel, high throughput sequencing.

[0461] Cells and/or cell beads of a given sample may be labeled with the same or different labels. For example, a first cell of a cell sample may be labeled with a first label and a second cell of the cell sample may be labeled with a second label. In some cases, the first and second labels may be the same. In other cases, the first and second labels may be different. Labels may differ in different aspects. For example, a first label and a second label used to label cells of the same sample may comprise the same nucleic acid barcode sequence but differ in another aspect, such as a unique molecular identifier sequence. Alternatively or in addition, a first label and a second label may both comprise a first nucleic acid barcode sequence and a second nucleic acid barcode sequence, where the first nucleic acid barcode sequences are the same and the second nucleic acid barcode sequences are different. Similarly, labels applied to different cellular samples may have one or more common features. For example, labels for cells of a first sample from a given subject may include a first common barcode sequence (e.g., identical nucleic acid barcode sequence) and a second common barcode sequence, while labels for cells of a second sample from the same subject may include a third common barcode sequence and a fourth common barcode sequence, which first common barcode sequence and third common barcode sequence are identical and which second common barcode sequence and fourth common barcode sequence are different.

[0462] The methods provided herein may comprise labeling and/or analysis of cell beads. Cell beads may comprise biological particles and/or their macromolecular constituents encased in a gel or polymer matrix. For example, a cell bead may comprise an entrapped cell. A cell bead may be generated prior to labeling of the cell bead, or components thereof. Alternatively, a cell bead may be generated after labeling and partitioning of a cell. For example, a labeled cell may be co-partitioned with polymerizable materials, and a cell bead comprising the labeled cell may be generated within the partition. A stimulus may be used to promote polymerization of the polymerizable materials within the partition.

[0463] Labeling individual cells and/or cell beads of a cell sample with nucleic acid barcode molecules for different cell samples can yield a plurality of labeled cell samples. An individual nucleic acid barcode molecule for labeling a cell and/or cell bead (e.g., a moiety-conjugated barcode molecule) can comprise a sample barcode sequence (also referred to as a feature barcode). Individual cell samples of a plurality of cell samples can each be labeled with nucleic acid barcode molecules having a barcode sequence unique to the cell sample. In embodiments herein, nucleic acid barcode molecules of a given labeled cell sample are distinguishable from nucleic acid barcode molecules of another labeled cell sample by the sample barcode sequence. In some instances, labeled cell samples can be combined and subjected to downstream sample processing in bulk. Sample barcode sequences can later be used to determine from which cell sample a particular cell originated.

[0464] Individual nucleic acid barcode molecules may form a part of a barcoded oligonucleotide. A barcoded oligonucleotide (e.g., a moiety-conjugated barcode molecule) can comprise sequence elements (e.g., functional sequences) in addition to the nucleic acid barcode molecule or sample barcode sequence. The additional sequence elements may be useful for a variety of downstream applications, including, but not limited to, sample preparation for sequencing analysis, e.g., next-generation sequence analysis. Non-limiting examples of additional sequence elements that can be present on barcoded oligonucleotides in embodiments herein include amplification primer annealing sequences or complements thereof, sequencing primer annealing sequences or complements thereof, common sequences shared among multiple different barcoded oligonucleotides; restriction enzyme recognition sites; probe binding sites or sequencing adapters (e.g., for attachment to a sequencing platform, such as a flow cell for parallel sequencing); molecular identifier sequences, e.g., unique molecular identifiers (UMIs); lipophilic molecules; and antibodies or epitope fragments thereof. For example, the barcoded oligonucleotide may comprise an amplification primer binding sequence. In another example, the barcoded oligonucleotide may comprise a sequencing primer binding sequence. In another example, the barcoded oligonucleotide may comprise a lipophilic molecule. In another example, the barcoded oligonucleotide may comprise an antibody or epitope fragment thereof. A sequence element may include a label, such as an optical label. Such a label may, for example, enable detection of a moiety with which the sequence element is associated. For example, a sequence element such as a lipophilic molecule may comprise a fluorescent moiety. The fluorescent moiety may permit optical detection of the lipophilic molecule and moieties with which it is associated.

[0465] A nucleic acid barcode molecule or a barcoded oligonucleotide comprising the nucleic acid barcode molecule may be linked to a moiety ("barcoded moiety") such as an antibody or an epitope binding fragment thereof, a cell surface receptor binding molecule, a receptor ligand, a small molecule, a pro-body, an aptamer, a monobody, an affimer, a darpin, or a protein scaffold. The moiety to which a nucleic acid barcode molecule or barcoded oligonucleotide can be linked may bind a molecule expressed on the surface of individual cells of the plurality of cell samples. A labeled cell sample may refer to a sample in which the cells and/or cell beads are bound to barcoded moieties.

[0466] A molecule of a cell and/or cell bead to which a moiety (e.g., barcoded moiety) may bind may be common to all cells of a given sample and/or all cells and/or cell beads

of a plurality of different cell samples. Such a molecule may be a protein. For example, a protein to which a moiety may bind may be a transmembrane receptor, major histocompatibility complex protein, cell-surface protein, glycoprotein, glycolipid, protein channel, or protein pump. A non-limiting example of a cell-surface protein can be a cell adhesion molecule. A molecule to which a moiety (e.g., barcoded moiety) may bind may be expressed at similar levels for all cells and/or cell beads of a given sample and/or all cells of a plurality of different cell samples. The expression of the molecule for all cells and/or cell beads of a sample and/or all cells of a plurality of different cell samples may be within biological variability. Alternatively, the molecule may be differentially expressed for certain cells and/or cell beads of the cell sample or a plurality of different cell samples. For example, the expression of the molecule for all cells and/or cell beads of a sample or a plurality of different cell samples may not be within biological variability, and/or some of the cells and/or cell beads of a cell sample or a plurality of different cell sample may be abnormal cells. A barcoded moiety may bind a molecule that is present on a majority of the cells and/or cell beads of a cell sample and/or a plurality of different cell samples. The molecule may be present on at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% of the cells and/or cell beads in a cell sample and/or a plurality of different cell samples.

[0467] A nucleic acid barcode molecule or barcoded oligonucleotide comprising the nucleic acid barcode molecule may be linked to an antibody or an epitope binding fragment thereof, and labeling cells and/or cell beads may comprise subjecting the antibody-linked barcode molecule or the epitope binding fragment-linked barcode molecule to conditions suitable for binding the antibody to a molecule present on a cell surface. The binding affinity between the antibody or the epitope binding fragment thereof and the molecule present on the cell surface may be within a desired range to ensure that the antibody or the epitope binding fragment thereof remains bound to the molecule. For example, the binding affinity may be within a desired range to ensure that the antibody or the epitope binding fragment thereof remains bound to the molecule during various sample processing steps, such as partitioning and/or nucleic acid amplification or extension. A dissociation constant (Kd) between the antibody or an epitope binding fragment thereof and the molecule to which it binds may be less than about 100 μM, 90 μM, 80 μM, 70 μM, 60 μM, 50 μM, 40 μM, 30 μM, 20 μM, 10 μM, 9 μM, 8 μM, 7 μM, 6 μM, 5 μM, 4 μM, 3 μM, 2 μM, 1 μM, 900 nM, 800 nM, 700 nM, 600 nM, 500 nM, 400 nM, 300 nM, 200 nM, 100 nM, 90 nM, 80 nM, 70 nM, 60 nM, 50 nM, 40 nM, 30 nM, 20 nM, 10 nM, 9 nM, 8 nM, 7 nM, 6 nM, 5 nM, 4 nM, 3 nM, 2 nM, 1 nM, 900 pM, 800 pM, 700 pM, 600 pM, 500 pM, 400 pM, 300 pM, 200 pM, 100 pM, 90 pM, 80 pM, 70 pM, 60 pM, 50 pM, 40 pM, 30 pM, 20 pM, 10 pM, 9 pM, 8 pM, 7 pM, 6 pM, 5 pM, 4 pM, 3 pM, 2 pM, or 1 pM. For example, the dissociation constant may be less than about 10 μM.

[0468] A nucleic acid barcode molecule or barcoded oligonucleotide comprising the nucleic acid barcode molecule may be coupled to a cell-penetrating peptide (CPP), and labeling cells may comprise delivering the CPP coupled nucleic acid barcode molecule into a cell and/or cell bead by the cell-penetrating peptide. The nucleic acid barcode molecule or barcoded oligonucleotide comprising the nucleic acid barcode molecule may be conjugated to a cell-penetrating peptide (CPP), and labeling cells and/or cell beads may comprise delivering the CPP conjugated nucleic acid barcode molecule into a cell and/or cell bead by the cell-penetrating peptide. A cell-penetrating peptide that can be used in the methods provided herein can comprise at least one non-functional cysteine residue, which may be either free or derivatized to form a disulfide link with an oligonucleotide that has been modified for such linkage. Non-limiting examples of cell-penetrating peptides that can be used in embodiments herein include penetratin, transportan, plsl, TAT(48-60), pVEC, MTS, and MAP. Cell-penetrating peptides useful in the methods provided herein can have the capability of inducing cell penetration for at least about 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% of cells of a cell population. The cell-penetrating peptide may be an arginine-rich peptide transporter. The cell-penetrating peptide may be Penetratin or the Tat peptide.

[0469] A nucleic acid barcode molecule or barcoded oligonucleotide comprising a nucleic acid barcode molecule may be coupled to a fluorophore or dye, and labeling cells may comprise subjecting the fluorophore-linked barcode molecule to conditions suitable for binding the fluorophore to the cell surface. See, e.g., FIG. **86**. In some instances, fluorophores can interact strongly with lipid bilayers and labeling cells may comprise subjecting the fluorophore-linked barcode molecule to conditions such that the fluorophore binds to or is inserted into the cell membrane. In some cases, the fluorophore is a water-soluble, organic fluorophore. In some instances, the fluorophore is Alexa 532 maleimide, tetramethylrhodamine-5-maleimide (TMR maleimide), BODIPY-TMR maleimide, Sulfo-Cy3 maleimide, Alexa 546 carboxylic acid/succinimidyl ester, Atto 550 maleimide, Cy3 carboxylic acid/succinimidyl ester, Cy3B carboxylic acid/succinimidyl ester, Atto 565 biotin, Sulforhodamine B, Alexa 594 maleimide, Texas Red maleimide, Alexa 633 maleimide, Abberior STAR 635P azide, Atto 647N maleimide, Atto 647 SE, or Sulfo-Cy5 maleimide. See, e.g., Hughes L D, et al. PLoS One. 2014 Feb. 4; 9(2):e87649, which is hereby incorporated by reference in its entirety for a description of organic fluorophores.

[0470] A nucleic acid barcode molecule or barcoded oligonucleotide comprising the nucleic acid barcode molecule may be coupled to a lipophilic molecule, and labeling cells and/or cell beads may comprise delivering the nucleic acid barcode molecule to a cell membrane or a nuclear membrane by the lipophilic molecule. Lipophilic molecules can associate with and/or insert into lipid membranes such as cell membranes and nuclear membranes. In some cases, the insertion can be reversible. In some cases, the association between the lipophilic molecule and the cell and/or cell bead may be such that the cell and/or cell bead retains the lipophilic molecule (e.g., and associated components, such as nucleic acid barcode molecules, thereof) during subsequent processing (e.g., partitioning, cell permeabilization, amplification, pooling, etc.). The nucleic acid barcode molecule or barcoded oligonucleotide comprising the nucleic acid barcode molecule may enter into the intracellular space and/or a cell nucleus. Non-limiting examples of lipophilic molecules that can be used in the methods provided herein include sterol lipids such as cholesterol, tocopherol, and derivatives thereof, steryl lipids, lignoceric acid, and palmitic acid. Other lipophilic molecules that may be used in the methods provided herein comprise amphiphilic molecules

wherein the headgroup (e.g., charge, aliphatic content, and/or aromatic content) and/or fatty acid chain length (e.g., C12, C14, C16, or C18) can be varied. For instance, fatty acid side chains (e.g., C12, C14, C16, or C18) can be coupled to glycerol or glycerol derivatives (e.g., 3-t-butyldiphenylsilylglycerol), which can also comprise, e.g., a cationic head group. The nucleic acid feature barcode molecules disclosed herein can then be coupled (either directly or indirectly) to these amphiphilic molecules. An amphiphilic molecule may associate with and/or insert into a membrane (e.g., a cell/cell bead or nuclear membrane). In some cases, an amphiphilic or lipophilic moiety may cross a cell membrane and provide a nucleic acid barcode molecule to an internal region of a cell and/or cell bead.

[0471] A nucleic acid barcode molecule may be attached to a lipophilic moiety (e.g., a cholesterol molecule). A nucleic acid barcode molecule may be attached to the lipophilic moiety via a linker, such as a tetra-ethylene glycol (TEG) linker. Other exemplary linkers include, but are not limited to, Amino Linker C6, Amino Linker C12, Spacer C3, Spacer C6, Spacer C12, Spacer 9, Spacer 18. A nucleic acid barcode molecule may be attached to the lipophilic moiety or the linker on the 5' end of the nucleic acid barcode molecule. Alternatively, a nucleic acid barcode molecule may be attached to the lipophilic moiety or the linker on the 3' end of the nucleic acid barcode molecule. In some instances, a first nucleic acid barcode molecule is attached to the lipophilic moiety or the linker at the 5' end of the nucleic acid barcode molecule and a second nucleic acid barcode molecule is attached to the lipophilic moiety or the linker at the 3' of the nucleic acid barcode molecule. The linker may be a glycol or derivative thereof. For example, the linker may be tetra-ethylene glycol (TEG) or polyethylene glycol (PEG). A nucleic acid barcode molecule may be releasably attached to the linker or lipophilic moiety (e.g., as described elsewhere herein for releasable attachment of nucleic acid molecules) such that the nucleic acid barcode molecule or a portion thereof can be released from the lipophilic molecule.

[0472] In some cases, a lipophilic molecule may comprise a label, such as an optical label. Such a label may, for example, enable detection of a moiety with which the lipophilic molecule is associated. For example, a lipophilic molecule may comprise a fluorescent moiety. The fluorescent moiety may permit optical detection of the lipophilic molecule and moieties with which it is associated.

[0473] An example of reagents and schemes suitable for analysis of barcoded lipophilic molecules is shown in panels I and II of FIG. **68**. Although a lipophilic moiety is shown in FIG. **68**, any moiety described herein (e.g., an antibody) can be conjugated to barcode oligonucleotides as described below. As shown in FIG. **68** (panel I), a lipophilic moiety (e.g., a cholesterol) **6801** is directly (e.g., covalently bound, bound via a protein-protein interaction, etc.) coupled to an oligonucleotide **6802** comprising a feature barcode sequence **6803** that functions to identify a cell or cell population. In some embodiments, oligonucleotide **6802** also includes additional sequences suitable for downstream reactions (e.g., sequence **6804** comprising a reverse complement of a sequence on second nucleic acid molecule **6806** and optionally sequence **6805** comprising a sequence configured to function as a PCR primer binding site). FIG. **68** (panel I) also shows an additional oligonucleotide **6806** (e.g., which in some instances, may be attached to a bead as described elsewhere herein) comprising a cell barcode sequence **6808** (also referred to herein as a bead barcode sequence or a nucleic acid barcode sequence), and a sequence **6810** complementary to a sequence **6804** on oligonucleotide **6802**. See also FIGS. **87** and **88** for exemplary sequences (e.g., **6810**, **6830**) complementary to moiety bound oligonucleotides (e.g., **6802**, **6822**). In some instances, oligonucleotide **6806** also comprises additional functional sequences suitable for downstream reactions such as a UMI sequence **6809** and an adapter sequence **6807** (e.g., a sequence **6807** comprising a sequencing primer binding site, e.g., a Read 1 ("R1") or a Read 2 ("R2") sequence, and in some instances, a P5 or P7 flow cell attachment sequence). Sequence **6810** represents a sequence that is complementary to complementary sequence **6804**. In some instances, sequence **6804** comprises a poly-A sequence and sequence **6810** comprises a poly-T sequence. In some instances, sequence **6810** comprises a poly-A sequence and sequence **6804** comprises a poly-T sequence. In some instances, sequence **6804** comprises a GGG-containing sequence and sequence **6810** comprises a complementary CCC-containing sequence. In some instances, sequence **6810** comprises a GGG-containing sequence and sequence **6804** comprises a complementary CCC-containing sequence. In some instances, the CCC-containing or GGG-containing sequences comprise one or more ribonucleotides. During analysis, sequence **6810** hybridizes with sequence **6804** and oligonucleotides **6802** and/or **6806** are extended via the action of a polymerizing enzyme (e.g., a reverse transcriptase, a polymerase), where oligonucleotide **6806** then comprises complement sequences to oligonucleotide **6802** at its 3' end. These constructs can then be optionally processed as described elsewhere herein and subjected to nucleic acid sequencing to, for example, identify cells associated with a specific feature barcode **6803** and a specific cell barcode **6808**. While the sequences included in panel I of FIG. **68** are presented in a given order, the sequences may be included in a different order, and/or with additional sequences or nucleotides disposed between one or more of the sequences. For example, the UMI **6809** and the barcode sequence **6808** may be transposed.

[0474] In another example, shown in FIG. **68** (panel II), a lipophilic moiety (e.g., a cholesterol) **6821** is indirectly (e.g., via hybridization or ligand-ligand interactions, such as biotin-streptavidin) coupled to an oligonucleotide **6822** comprising a feature barcode sequence **6823** that functions to identify a cell or cell population. Lipophilic molecule **6821** is directly (e.g., covalently bound, bound via a protein-protein interaction) coupled to a hybridization oligonucleotide **6832** that hybridizes with sequence **6831** of oligonucleotide **6822**, thereby indirectly coupling oligonucleotide **6822** to the lipophilic moiety. In some embodiments, oligonucleotide **6822** includes additional sequences suitable for downstream reactions (e.g., sequence **6824** comprising a reverse complement of a sequence on second nucleic acid molecule **6826** and optionally sequence **6825** comprising a sequence configured to function as a PCR primer binding site). FIG. **68** (panel II) also shows an additional oligonucleotide **6826** (e.g., which in some instances, may be attached to a bead as described elsewhere herein) comprising a cell barcode sequence **6828** (e.g., a nucleic acid barcode sequence), and a sequence **6830** complementary to a sequence **6824** on oligonucleotide **6822**. In some instances, oligonucleotide **6826** also comprises additional functional sequences suitable for downstream reactions such as a UMI sequence **6829** and an adapter sequence **6827** (e.g., a

sequence **6827** comprising a sequencing primer binding site, e.g., a Read 1 ("R1") or a Read 2 ("R2") sequence, and in some instances, a P5 or P7 flow cell attachment sequence). Sequence **6810** represents a sequence that is complementary to complementary sequence **6804**. In some instances, sequence **6824** comprises a poly-A sequence and sequence **6830** comprises a poly-T sequence. In some instances, sequence **6830** comprises a poly-A sequence and sequence **6824** comprises a poly-T sequence. In some instances, sequence **6824** comprises a GGG-containing sequence and sequence **6830** comprises a complementary CCC-containing sequence. In some instances, sequence **6830** comprises a GGG-containing sequence and sequence **6824** comprises a complementary CCC-containing sequence. In some instances, the CCC-containing or GGG-containing sequences comprise one or more ribonucleotides. During analysis, sequence **6830** hybridizes with sequence **6824** and oligonucleotides **6822** and/or **6826** are extended via the action of a polymerizing enzyme (e.g., a reverse transcriptase, a polymerase), where oligonucleotide **6826** then comprises complement sequences to oligonucleotide **6822** at its 3' end. These constructs can then be optionally processed as described elsewhere herein and subjected to nucleic acid sequencing to, for example, identify cells associated with a specific feature barcode **6823** and a specific cell barcode **6828**. While the sequences included in panel II of FIG. **68** are presented in a given order, the sequences may be included in a different order, and/or with additional sequences or nucleotides disposed between one or more of the sequences. For example, the UMI **6829** and the barcode sequence **6828** may be transposed. See, e.g., FIG. **88** for additional exemplary oligonucleotides suitable for use with the labeling moieties (e.g., lipophilic, antibody, fluorophore, etc.) described herein.

[0475] In an example, a method provided herein may be used to label cells using feature barcodes linked to cell surfaces. A cell surface feature (e.g., a lipophilic moiety, such as a cholesterol) of a plurality of cells may be linked (e.g., conjugated) to a feature barcode. The feature barcode may include, for example, a sequence configured to hybridize to a nucleic acid barcode molecule, such as a sequence comprising multiple cytosine nucleotides (e.g., a CCC sequence). Each feature barcode may comprise a barcode sequence and/or a unique molecular identifier sequence. A plurality of beads (e.g., gel beads) each comprising a plurality of nucleic acid barcode molecules may be provided. The nucleic acid barcode molecules of each bead (e.g., releasably attached to each bead) may comprise a barcode sequence (e.g., cell barcode sequence), a unique molecular identifier sequence, and a sequence configured to hybridize to a feature barcode linked to a cell surface. Nucleic acid barcode molecules of each different bead may comprise the same barcode sequence, which barcode sequence differs from barcode sequences of nucleic acid barcode molecules of other beads of the plurality of beads. The feature barcode-linked cells may be partitioned with the plurality of beads into a plurality of partitions (e.g., droplets, such as aqueous droplets in an emulsion) such that at least a subset of the plurality of partitions each comprise a single cell and a single bead. One or more nucleic acid barcode molecules of the bead of each partition may attach (e.g., hybridize or ligate) to one or more feature barcodes of the cell of the same partition. The one or more nucleic acid barcode molecules of the bead may be released (e.g., via application of a stimulus, such as a chemical stimulus) from the bead within the partition prior to attachment of the one or more nucleic acid barcode molecules to the one or more feature barcodes of the cell. The cell may be lysed or permeabilized within the partition to provide access to analytes therein, such as nucleic acid molecules therein (e.g., deoxyribonucleic acid (DNA) molecules and/or ribonucleic acid (RNA) molecules). One or more analytes (e.g., nucleic acid molecules) of the cell may also be barcoded within the partition with one or more nucleic acid barcode molecules of the bead to provide a plurality of barcoded analytes (e.g., barcoded nucleic acid molecules). The plurality of partitions comprising barcoded analytes and barcoded cell surface features may be combined (e.g., pooled). Additional processing may be performed to, for example, prepare the barcoded analytes and barcoded cell surface features for subsequent analysis. For example, barcoded nucleic acid molecules may be derivatized with flow cell adapters to facilitate nucleic acid sequencing. Barcodes of barcoded analytes may be detected (e.g., using nucleic acid sequencing) and used to identify the barcoded analytes as deriving from particular cells or cell types of the plurality of cells.

[0476] In another example, a method provided herein may be used to label cells using lipophilic feature barcodes. Feature barcodes comprising a lipophilic moiety (e.g., a cholesterol moiety) may be incubated with a plurality of cells. The feature barcodes may comprise an optical label such as a fluorescent moiety. The feature barcodes may include, for example, a sequence configured to hybridize to a nucleic acid barcode molecule, such as a sequence comprising multiple cytosine nucleotides (e.g., a CCC sequence). Each feature barcode may also comprise a barcode sequence and/or a unique molecular identifier sequence. A plurality of beads (e.g., gel beads) each comprising a plurality of nucleic acid barcode molecules may be provided. The nucleic acid barcode molecules of each bead (e.g., releasably attached to each bead) may comprise a barcode sequence (e.g., cell barcode sequence), a unique molecular identifier sequence, and a sequence configured to hybridize to a feature barcode. Nucleic acid barcode molecules of each different bead may comprise the same barcode sequence, which barcode sequence differs from barcode sequences of nucleic acid barcode molecules of other beads of the plurality of beads. The cells incubated with feature barcodes may be partitioned (e.g., subsequent to one or more washing processes) with the plurality of beads into a plurality of partitions (e.g., droplets, such as aqueous droplets in an emulsion) such that at least a subset of the plurality of partitions each comprise a single cell and a single bead. Within each partition of the at least a subset of the plurality of partitions, one or more nucleic acid barcode molecules of the bead may attach (e.g., hybridize or ligate) to one or more feature barcodes of the cell. The one or more nucleic acid barcode molecules of the bead may be released (e.g., via application of a stimulus, such as a chemical stimulus) from the bead within the partition prior to attachment of the one or more nucleic acid barcode molecules to the one or more feature barcodes of the cell to provide a barcoded feature barcode. The cell may be lysed or permeabilized within the partition to provide access to analytes therein, such as nucleic acid molecules therein (e.g., deoxyribonucleic acid (DNA) molecules and/or ribonucleic acid (RNA) molecules), and/or to the feature barcode therein (e.g., if the feature barcode has permeated the cell mem-

brane). One or more analytes (e.g., nucleic acid molecules) of the cell may also be barcoded within the partition with one or more nucleic acid barcode molecules of the bead to provide a plurality of barcoded analytes (e.g., barcoded nucleic acid molecules). The plurality of partitions comprising barcoded analytes and barcoded feature barcodes may be combined (e.g., pooled). Additional processing may be performed to, for example, prepare the barcoded analytes and barcoded feature barcodes for subsequent analysis. For example, barcoded nucleic acid molecules and/or barcoded feature barcodes may be derivatized with flow cell adapters to facilitate nucleic acid sequencing. Barcodes of barcoded analytes and barcoded feature barcodes may be detected (e.g., using nucleic acid sequencing) and used to identify the barcoded analytes and barcoded feature barcodes as deriving from particular cells or cell types of the plurality of cells.

[0477] Cells and/or cell beads may be contacted with one or more additional agents along with moiety-conjugated feature barcodes (e.g., the lipophilic molecules described herein). For example, cells and/or cell beads may be contacted with a lipophilic moiety-conjugated barcode molecule and one or more additional moiety (e.g., lipophilic moiety) conjugated "anchor" molecules (see, e.g., FIG. **67**). In some instances, a cell and/or cell bead is contacted with (1) a lipophilic-moiety conjugated to a first nucleic acid molecule comprising a capture sequence (e.g., a poly-A sequence), a feature barcode sequence, and a primer sequence; and (2) an anchor molecule comprising a lipophilic moiety conjugated to a second nucleic acid molecule comprising a sequence complementary to the primer sequence. In other instances, a cell and/or cell bead is contacted with (1) a lipophilic-moiety conjugated to a first nucleic acid molecule comprising a capture sequence (e.g., a poly-A sequence), a feature barcode sequence, and a primer sequence; (2) an anchor molecule comprising a lipophilic moiety conjugated to a second nucleic acid molecule comprising an anchor sequence and a sequence complementary to the primer sequence; and (3) a co-anchor molecule comprising a lipophilic moiety conjugated to a third nucleic acid molecule comprising a sequence complementary to the anchor sequence. Moiety-conjugated oligonucleotides can comprise any number of modifications, such as modifications which prevent extension by a polymerase and other such modifications described elsewhere herein.

[0478] The structure of the moiety-attached barcode oligonucleotides may include a number of sequence elements in addition to the feature barcode sequence. The oligonucleotide may include functional sequences that are used in subsequent processing, which may include one or more of a sequencer specific flow cell attachment sequence, e.g., a P5 or P7 sequence for Illumina sequencing systems, as well as sequencing primer sequences, e.g., a R1 or R2 sequencing primer sequence for Illumina sequencing systems. A specific priming and/or capture sequence, such as poly-A sequence, may be also included in the oligonucleotide structure.

[0479] As described above, moiety-attached barcode oligonucleotides can be processed to attach a cell barcode sequence. Cell barcode oligonucleotides (which can be attached to a bead) may comprise a poly-T sequence designed to hybridize and capture poly-A containing moiety-attached barcode oligonucleotides. A poly-T cell barcode molecule may comprise an anchoring sequence segment to ensure that the poly-T sequence hybridizes to the poly-A sequence of the moiety-attached barcode oligonucleotides. This anchoring sequence can include a random short sequence of nucleotides, e.g., 1-mer, 2-mer, 3-mer or longer sequence. An additional sequence segment may be included within the cell barcode oligonucleotide molecules. This additional sequence may provide a unique molecular identifier (UMI) sequence segment, e.g., as a random sequence (e.g., such as a random N-mer sequence) that varies across individual oligonucleotides (e.g., cell barcode molecules coupled to a single bead), whereas the cell barcode sequence is constant among the oligonucleotides (e.g., cell barcode molecules coupled to a single bead). This unique sequence may serve to provide a unique identifier of the starting nucleic acid molecule that was captured, in order to allow quantitation of the number of original molecules present (e.g., the number of moiety-conjugated nucleic acid barcode molecules).

[0480] Nucleic acid barcode molecules or barcoded oligonucleotides comprising the nucleic acid barcode molecules may be coupled to a plurality of beads, such as a plurality of gel beads. An individual bead of a plurality of beads can include tens to hundreds of thousands or millions of individual oligonucleotide molecules (e.g., at least about 10,000, 50,000, 100,000, 500,000, 1,000,000 or 10,000,000 oligonucleotide molecules), where a barcode segment of the oligonucleotide molecules can be constant or relatively constant for all of the oligonucleotide molecules coupled to a given bead. Oligonucleotide molecules coupled to a given bead may also comprise a variable or unique sequence segment that may vary across the oligonucleotide molecules coupled to the given bead. The variable or unique sequence segment may be a unique molecular identifier (UMI) sequence segment that may include from 5 to about 8 or more nucleotides within the sequence of the oligonucleotides. In some cases, the unique molecular identifier (UMI) sequence segment can be 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 nucleotides in length or longer. In some cases, the unique molecular identifier (UMI) sequence segment can be at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 nucleotides in length or longer. In some cases, the unique molecular identifier (UMI) sequence segment can be at most 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 nucleotides in length. In some cases, the sample oligonucleotide (e.g., partition nucleic acid barcode molecule) may comprise a target-specific primer (e.g., a primer sequence specific for a sequence in the moiety-conjugated oligonucleotides). For example, the specific sequence may be a sequence that is not in the capture sequence (e.g., not the poly-A or CCC-containing capture sequence).

[0481] Labeling cells and/or cell beads may comprise delivering a nucleic acid barcode molecule or barcoded oligonucleotide comprising the nucleic acid barcode molecule into a cell and/or cell bead using a physical force or chemical compound. A labeled cell sample may refer to a sample in which one or more cells and/or cell beads have nucleic acid barcode molecules introduced to the cells and/or cell beads (e.g., coupled to the surface of the cells and/or cell beads) and/or within the cells and/or cell beads.

[0482] Use of physical force (e.g., to deliver a nucleic acid barcode molecule or barcoded oligonucleotide to a cell and/or cell bead) can refer to the use of a physical force to counteract the cell membrane barrier in facilitating intracellular delivery of oligonucleotides. Examples of physical methods that can be used in embodiments herein include the

use of a needle, ballistic DNA, electroporation, sonoporation, photoporation, magnetofection, and hydroporation.

[0483] Labeling cells and/or cell beads may comprise the use of a needle, for example for injection (e.g., microinjection). Alternatively or in addition, labeling cells and/or cell beads may comprise particle bombardment. With particle bombardment, nucleic acid barcode molecules can be coated on heavy metal particles and delivered to a cell and/or cell bead at a high speed. Labeling cells and/or cell beads may comprise electroporation. With electroporation, nucleic acid barcode molecules can enter a cell and/or cell bead through one or more pores in the cellular membrane formed by applied electricity. The pore of the membrane can be reversible based on the applied field strength and pulse duration. Labeling cells and/or cell beads may comprise sonoporation. Cell membranes can be temporarily permeabilized using sound waves, allowing cellular uptake of nucleic acid barcode molecules. Labeling cells and/or cell beads may comprise photoporation. A transient pore in a cell membrane can be generated using a laser pulse, allowing cellular uptake of nucleic acid barcode molecules. Labeling individual cells and/or cell beads may comprise magnetofection. Nucleic acid barcode molecules can be coupled to a magnetic particle (e.g., magnetic nanoparticle, nanowires, etc.) and localized to a target cell and/or cell bead via an applied magnetic field. Labeling cells and/or cell beads may comprise hydroporation. Nucleic acid barcode molecules can be delivered to cells and/or cell beads via hydrodynamic pressure.

[0484] Various chemical compounds can be used in embodiments herein to deliver nucleic acid barcode molecules into a cell and/or cell bead. Chemical vectors can include inorganic particles, lipid-based vectors, polymer-based vectors and peptide-based vectors. Non-limiting examples of inorganic particles that can be used in embodiments herein to deliver nucleic acid barcode molecules into a cell and/or cell bead include inorganic nanoparticles prepared from metals, (e.g., iron, gold, and silver), inorganic salts, and ceramics (e.g, phosphate or carbonate salts of calcium, magnesium, or silicon). The surface of a nanoparticle can be coated to facilitate nucleic acid molecule binding or chemically modified to facilitate nucleic acid molecule attachment. Magnetic nanoparticles (e.g., supermagnetic iron oxide), fullerenes (e.g., soluble carbon molecules), carbon nanotubes (e.g., cylindrical fullerenes), quantum dots and supramolecular systems may be used.

[0485] Labeling cells and/or cell beads may comprise use of a cationic lipid, such as a liposome. Various types of lipids can be used in liposome delivery. In some cases, a nucleic acid barcode molecule is delivered to a cell via a lipid nano emulsion. A lipid emulsion refers to a dispersion of one immiscible liquid in another stabilized by emulsifying agent. Labeling cells and/or cell beads may comprise use of a solid lipid nanoparticle.

[0486] Labeling cells and/or cell beads may comprise use of a peptide based chemical vector. Cationic peptides may be rich in basic residues like lysine and/or arginine. Labeling cells and/or cell beads may comprise use of polymer based chemical vector. Cationic polymers, when mixed with nucleic acid molecules, can form nanosized complexes called polypexes. Polymer based vectors may comprise natural proteins, peptides and/or polysaccharides. Polymer based vectors may comprise synthetic polymers. Labeling cells may comprise use of a polymer based vector comprising polyethylenimine (PEI). PEI can condense DNA into positively charged particles which bind to anionic cell surface residues and are brought into the cell via endocytosis. Labeling cells and/or cell beads may comprise use of polymer based chemical vector comprising poly-L-lysine (PLL), poly (DL-lactic acid) (PLA), poly (DL-lactide-co-glycoside) (PLGA), polyornithine, polyarginine, histones, or protamines. Polymer based vectors may comprise a mixture of polymers, for example PEG and PLL. Other polymers include dendrimers, chitosans, synthetic amino derivatives of dextran, and cationic acrylic polymers.

[0487] Following cell labeling, a majority of the cells and/or cell beads of individual cell samples can be labeled with nucleic acid barcode molecules having a sample barcode sequence (e.g., a moiety-conjugated barcode molecule, also referred to herein as a feature barcode). At least 50%, 60%, 70%, 75%, 80%, 85%, 90%, or 95% of cells of a cell sample may be labeled. In some cases, not all of the cells are labeled. For example, less than 100%, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, or 50% of cells of a cell sample may be labeled.

[0488] The plurality of labeled cell samples may be subjected to one or more reactions. The one or more reactions may comprise one or more nucleic acid extension reactions. The one or more reactions may comprise one or more nucleic acid amplification reactions. Alternatively or in addition, the one or more reactions may comprise one or more ligation reactions.

[0489] Individual labeled cells and/or cell beads of the plurality of labeled cell samples may be co-partitioned into a plurality of partitions (e.g., a plurality of wells or droplets). For example, labeled cells and/or cell beads may be partitioned into a plurality of partitions prior to undergoing one or more reactions. Labeled cells may be partitioned into partitions with one or more polymerizable materials such that labeled cell beads may be generated within the partitions. One or more labeled cells and/or cell beads may be included in a given partition of the plurality of partitions. Subjecting the nucleic acid molecules of the plurality of labeled cell samples one or more reactions may comprise partitioning individual cells and/or cell beads of the plurality of labeled cell samples into partitions and within individual partitions, synthesizing a nucleic acid molecule comprising (i) a sample barcode sequence and (ii) a sequence corresponding to a nucleic acid molecule. By partitioning the labeled cell samples into a plurality of partitions, the one or more reactions can be performed for individual cells and/or cell beads in isolated environments. Individual partitions may comprise at most a single cell and/or cell bead. Alternatively, a subset of partitions may contain at least a single cell and/or cell bead.

[0490] A partition may be an aqueous droplet in a non-aqueous phase such as oil. For example, a partition may comprise droplets, such as a droplet in an emulsion. Alternatively or in addition, partitions comprise wells or tubes.

[0491] A partition may contain a bead comprising a reagent for synthesizing a nucleic acid molecule. The reagent may be releasably attached to the bead. The reagent may comprise a nucleic acid, such as a nucleic acid primer. The nucleic acid may comprise a partition-specific barcode sequence. Two cells from a given cell sample may have an identical sample (e.g., cell) barcode sequence but different partition-specific barcode sequences (e.g., if the two cells are partitioned in two different partitions comprising the differ-

ent partition-specific barcode sequences). In an example, a first cell from a first cell sample has a first sample barcode sequence and a first partition-specific barcode sequence and a second cell from a second cell sample has a second sample barcode sequence and a second partition-specific barcode sequence. The first sample barcode sequence and the second sample barcode sequence may be different. The first partition-specific barcode sequence and the second partition-specific barcode sequence may also be different (e.g., if the two cells are partitioned in two different partitions comprising the different partition-specific barcode sequences). Alternatively, the first partition-specific barcode sequence and the second partition-specific barcode sequence may be the same (e.g., if the two cells are partitioned in the same partition).

[0492] A bead to which one or more oligonucleotides or nucleic acid barcode molecules may be degradable upon application of a stimulus. The stimulus may comprise a chemical stimulus. A bead may be degraded within a partition. Where a bead comprises a reagent for synthesizing a nucleic acid molecule, the reagent may be released, e.g., into a partition comprising the bead, upon degradation of the bead.

[0493] A plurality of nucleic acid barcode products can be subjected to nucleic acid sequencing to yield a plurality of sequencing reads. Individual sequencing reads can be associated with individual labeled cell samples based on a sample barcode sequence. Individual reads can be associated with individual labeled cell samples based on the sample barcode sequence.

[0494] A method of the present disclosure may comprise pooling a plurality of nucleic acid barcode products from partitions prior to subjecting the nucleic acid barcode products, or derivatives thereof, to an assay such as nucleic acid sequencing. Nucleic acid barcode products may be subjected to processing such as nucleic acid amplification. In some cases, one or more features such as one or more functional sequences (e.g., sequencing primers and/or flow cell adapter sequences) may be added to nucleic acid barcode products, e.g., after pooling of nucleic acid barcode products from the partitions. For example, pooled amplification products may be subjected to one or more reactions prior to sequencing. For example, the pooled nucleic acid barcode products may be subjected to one or more additional reactions (e.g., nucleic acid extension, polymerase chain reaction, or adapter ligation). Adapter ligation may include, for example, fragmenting the nucleic acid barcode products (e.g., by mechanical shearing or enzymatic digestion) and enzymatic ligation.

[0495] A cell sample may comprise a plurality of cells and/or cell beads. A cell sample may comprise constituents in addition to cells and/or cell beads. For example, a cell sample can contain at least one of proteins, cell-free polynucleotides (e.g., cell-free DNA), cell stabilizing agents, protein stabilizing agents, enzyme inhibitors, cell nuclei, and ions.

[0496] Cell samples can be obtained from any of a variety of sources. For example, cell samples can be obtained from tissue samples. A tissue sample can be obtained from any suitable tissue source. Tissue samples can be obtained from components of the circulatory system, the digestive system, the endocrine system, the immune system, the lymphatic system, the nervous system, the muscular system, the reproductive system, the skeletal system, the respiratory system, the urinary system, and the integumentary system. A cell sample may be obtained from a tissue sample of the circulatory system such as the heart or blood vessels (e.g., arteries, veins, etc). A cell sample may be obtained from a tissue sample of the digestive system (e.g., mouth, esophagus, stomach, small intestine, large intestine, rectum, and anus). A cell sample may be obtained from a tissue sample of the endocrine system (e.g., pituitary gland, pineal gland, thyroid gland, parathyroid gland, adrenal gland, and pancreas). A cell sample may be obtained from a tissue sample of the immune system (e.g., lymph nodes, spleen, and bone marrow). A cell sample may be obtained from a tissue sample of the lymphatic system (e.g., lymph nodes, lymph ducts, and lymph vessels). In some embodiments, a cell sample is obtained from a tissue sample of the nervous system (e.g., brain and spinal cord). In some embodiments, a cell sample is obtained from a tissue sample of the muscular system (e.g., skeletal muscle, smooth muscle, and cardiac muscle). In some embodiments, a cell sample is obtained from a tissue sample of the reproductive system (e.g., penis, testes, vagina, uterus, and ovaries). In some embodiments, a cell sample is obtained from a tissue sample of the skeletal system (e.g., tendons, ligaments, and cartilage). In some embodiments, a cell sample is obtained from a tissue sample of the respiratory system (e.g., trachea, diaphragm, and lungs). In some embodiments, a cell sample is obtained from a tissue sample of the urinary system (e.g., kidneys, ureters, bladder, sphincter muscle, and urethra). In some embodiments, a cell sample is obtained from a tissue sample of the integumentary system (e.g., skin).

[0497] A tissue sample can be obtained by invasive, minimally invasive, or non-invasive procedures. Tissues samples can be obtained, for example, by surgical excision, biopsy, cell scraping, or swabbing. A tissue sample may be a tissue sample obtained during a surgical procedure or a sample obtained for diagnostic purposes. A tissue sample can be a fresh tissue sample, a frozen tissue sample, or a fixed tissue sample.

[0498] In some cases, a tissue and/or cell sample may be embedded, embalmed, preserved, and/or fixed. For example, a tissue and/or cell sample may be both fixed and embedded. A tissue and/or cell sample may comprise one or more fixed cells. Fixation is a process that preserves biological tissue or a cell from decay, thereby preventing autolysis or putrefaction. A fixed tissue may preserve its cells, its tissue components, or both. Fixation may be done through a crosslinking fixative by forming covalent bonds between proteins in the tissue or cell to be fixed. Fixation may anchor soluble proteins to the cytoskeleton of a cell. Fixation may form a rigid cell, a rigid tissue, or both. Fixation may be achieved through use of chemicals such as formaldehyde (e.g. formalin), gluteraldehyde, ethanol, methanol, acetic acid, osmium tetraoxide, potassium dichromate, chromic acid, potassium permanganate, Zenker's fixative, picrates, Hepes-glutamic acid buffer-mediated organic solvent protection effect (HOPE), or any combination thereof. Formaldehyde may be used as a mixture of about 37% formaldehyde gas in aqueous solution on a weight by weight basis. The aqueous formaldehyde solution may additionally comprise about 10-15% of an alcohol (e.g. methanol), forming a solution termed "formalin." A fixative-strength (10%) solution would equate to a 3.7% solution of formaldehyde gas in water. Formaldehyde may be used as at least 5%, 8%, 10%, 12% or 15% Neutral Buffered Formalin (NBF) solution (i.e. fixative strength). Formaldehyde may be used as 3.7% to

4.0% formaldehyde in phosphate buffered saline (i.e. formalin). In some instances, fixation is performed using at least 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8.0, 8.5, 9.0, 9.5, 10, 10.5, 11.0, 11.5, 12.0, 12.5, 13.0, 13.5, 14.0, 14.5, or 15.0 percent (%) or more formalin flush or immersion. In some instances, fixation is performed using about 10% formalin flush. Fixative volume can be 10, 15, 20, 25 or 30 times that of tissue on a weight per volume. Subsequent to fixation in formaldehyde, the tissue or cell may be submerged in alcohol for long term storage. In some cases, the alcohol is methanol, ethanol, propanol, butanol, an alcohol containing five or more carbon atoms, or any combination thereof. The alcohol may be linear or branched. The alcohol may be at least 50%, 60%, 70%, 80% or 90% alcohol in aqueous solution. In some examples, the alcohol is 70% ethanol in aqueous solution.

[0499] Cell samples can be obtained from biological fluids. A biological fluid can be obtained from any suitable source. Exemplary biological fluid sources from which cell samples can be obtained include amniotic fluid, bile, blood, cerebral spinal fluid, lymph fluid, pericardial fluid, peritoneal fluid, pleural fluid, saliva, seminal fluid, sputum, sweat, tears, and urine. Biological fluids can be obtained by invasive, minimally invasive, or non-invasive procedures. A biological fluid comprising blood can be obtained, for example, by venipuncture, pinprick, or aspiration.

[0500] The plurality of different cell samples analyzed by methods provided herein may be a plurality of samples from a single subject. The plurality of different cell samples may be obtained from the single subject at different time points over the course of a pre-defined or un-defined length of time. For example, the plurality of cell samples may be obtained from a subject a multiple time points before and/or after the administration of a therapeutic treatment. The plurality of cell samples can be analyzed to assess and/or monitor the subject's response to the therapeutic treatment. In some embodiments, the plurality of different cell samples are cell samples obtained from different sources from the single subject. For example, the subject may be diagnosed with cancer and cell samples from a plurality of tissue sources are examined to determine the extent of cancer metastasis. The plurality of different cell samples may be obtained from different regions of a tissue sample. For example, a subject may undergo surgical treatment to excise a tumorous region. A plurality of different cell samples from different regions of a tissue sample can be assessed to identify the boundary between normal and abnormal tissue. The plurality of different cell samples may comprise cancerous and non-cancerous cell samples.

[0501] The plurality of different cell samples analyzed by methods provided herein may be a plurality of samples from a plurality of subjects. Alternatively or in addition, the plurality of different cell samples may comprise a plurality of different cell samples from the same subject. For example, different cell samples may be taken from the same subject at different times (e.g., at different time points in during a treatment regimen). In another example, different cell samples may be taken from different areas or features of the same subject. For instance, a first cell sample may be a blood sample, and a second cell sample may be a tissue sample. For parallel processing, a plurality of samples (e.g., from a plurality of subjects) can be combined for simultaneous processing. In some cases, at least two different cell samples from at least two different subjects are processed simultaneously (e.g., at least 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, or 25 samples) are combined and processed in parallel.

Spatial Mapping

[0502] In an aspect, the present disclosure provides methods and compositions for spatial mapping. A plurality of nucleic acid barcode molecules can be arranged according to a spatial relationship. The method of spatially mapping a plurality of cells in a sample may comprise spotting or otherwise distributing a plurality of nucleic acid barcode molecules comprising a labelling barcode sequence onto a cell sample comprising cells and/or cell beads (e.g., a three-dimensional tissue sample or a tissue section on a substrate) to yield a plurality of labeled cells in said cell sample. The plurality of nucleic acid barcode molecules may be modified to penetrate the cell membrane of cells and/or cell beads in said cell sample. The nucleic acid barcode molecules may be modified with a lipophilic moiety. In some instances, the cell sample is spotted with the plurality of nucleic acid barcode molecules according to a pre-defined spatial configuration or pattern. For example, nine sets of nucleic acid barcode molecules (e.g., 9 sets of nucleic acid barcode molecules having 9 unique sample barcode sequences) can be arranged in square grid of 3×3. All sample barcodes located in a particular square of the grid (e.g., #1) can have the same sample barcode sequence (e.g., sample barcode sequence #1). The sample barcode sequence in a given square may be different from all other sample barcode sequences in other squares. The sample barcodes and corresponding sample barcode sequences of the various sets can have a pre-defined spatial relationship. For example, with reference to FIG. **66**A, a sample barcode sequence #1 can be positioned in proximity to sample barcode sequence #2 and #4; sample barcode sequence #2 can be positioned in proximity to sample barcode sequence #1, #3 and #5; sample barcode sequence #3 can be positioned in proximity to sample barcode sequence #2 and #6; sample barcode sequence #4 can be positioned in proximity to sample barcode sequence #1, #5 and #7; sample barcode sequence #5 can be positioned in proximity to sample barcode sequence #2, #4, #6, and #8; sample barcode sequence #6 can be positioned in proximity to sample barcode sequence #3, #5 and #9; sample barcode sequence #7 can be positioned in proximity to sample barcode sequence #4 and #8; sample barcode sequence #8 can be positioned in proximity to sample barcode sequence #5, #7 and #9; and sample barcode sequence #9 can be positioned in proximity to sample barcode sequence #6 and #8. Other spatial arrangements and relationships are contemplated herein. A plurality of nucleic acid barcode molecules can be arranged in any suitable configuration, for example deposited onto a planar or non-planar two-dimensional surface.

[0503] In some instances, the modified nucleic acid barcode molecule is coupled to a lipophilic molecule which enables the delivery of the nucleic acid molecule across the cell membrane or the nuclear membrane. Non-limiting examples of lipophilic molecules that can be used in embodiments described herein include sterol lipids such as cholesterol, tocopherol, and derivatives thereof. In other instances, the modified nucleic acid barcode molecule is coupled to a cell-penetrating peptide which can enable the molecule to penetrate the cell in the sample. In other cases, the modified nucleic acid barcode molecules are delivered into the cells and/or cell beads using liposomes, nanopar-

ticles, or electroporation. In some cases, the modified nucleic acid barcode molecule may be delivered into the cells and/or cell beads by mechanical force (e.g. nanowires, or microinjection). In some examples, the unique sample barcode sequences are generated using antibodies, which may bind to proteins coupled to cells and/or cell beads in each of the regions in which the sample is located. The antibodies or sequences derived from the antibodies may then be used to identify the regions within which the sample is located. In yet another embodiment, the modified nucleic acid barcode molecule is coupled to a fluorophore or dye, as further described herein. In one other embodiment, the modified nucleic acid barcode molecule is coupled to an inorganic nanoparticle, as further described herein.

[0504] In some instances, nucleic acid barcode molecules are spotted or otherwise distributed onto a cell sample comprising cells and/or cell beads present in the cell sample in at least two dimensions. Nucleic acid barcode molecules may be spotted onto the cell sample in known locations or in a regular pattern, e.g., in a grid pattern as described above and as shown in FIG. **66**A. In some cases, nucleic acid barcode molecules spotted into a known location are distributed radially from the spotting location. The spotting or distribution pattern of nucleic acid barcode molecules may be such that some cells and/or cell beads will comprise two or more different nucleic acid barcode molecules, each comprising a unique barcode sequence. For example, nucleic acid barcode molecules (e.g., nucleic acid barcode molecules conjugated to a lipophilic moiety) are spotted onto a cell sample in a 3×3 grid pattern (see, e.g., FIG. **66**A) such that a different set of nucleic acid barcode molecules are deposited onto each "square" of the grid (i.e., each "square" of the grid has a unique barcode sequence). In some cases, the nucleic acid barcode molecules diffuse out (e.g. radially) from the spotting or distribution point creating a concentration gradient of nucleic acid barcode molecules such that cells and/or cell beads closer to the spotting position will have relatively more nucleic acid barcode molecules compared to cells further from the spotting point. Furthermore, in some instances, a labeled cell and/or cell bead will comprise nucleic acid barcode molecules comprising 2 or more different nucleic acid barcode sequences. A cell and/or cell bead can then be analyzed for particular barcode sequences to infer the special relationship of cells (or the relative spatial relationship of a cell to another cell) within the cell sample. For example, cells and/or cell beads present in grid #1 of FIG. **66**A are labelled by a set nucleic acid barcode molecules, each comprising a common barcode sequence (e.g., barcode sequence #1), while cells and/or cell beads present in grid #2 are labelled by a different set nucleic acid barcode molecules each comprising a common barcode sequence (e.g., barcode sequence #2). The labelling procedure is repeated for each area of the grid or pattern such that a different set of nucleic acid barcode molecules is distributed across the relevant portions of the cell sample. Dependent upon their position in the cell sample, cells and/or cell beads can be labelled with one or more unique barcode sequences (e.g., a cell can be labelled with both barcode sequence #1 and barcode sequence #2, etc.). Individual cells and/or cell beads are then dissociated from the cell sample and analyzed for the presence of nucleic acid barcode molecules comprising one or more barcode sequences. In some instances, cells and/or cell beads are analyzed for both the presence of specific barcode sequences and also the amount of each nucleic acid barcode molecule associated with each cell and/or cell bead (e.g., using a UMI). Thus, in some instances, the known spotting pattern of the nucleic acid barcode molecules, the presence of particular barcode sequences, and the amount of each nucleic acid barcode molecule is utilized to determine the spatial position of a cell and/or cell bead in the cell sample or the relative spatial position of a cell and/or cell bead to another cell and/or cell bead in the cell sample.

[0505] A sample **6600** having at least two dimensions, for example a tissue sample or a cross-section of a tissue, may be labeled with a plurality of nucleic acid barcode molecules, for example, as shown in FIG. **66**B. In some cases, cells and/or cell beads present in different locations of a tissue sample or a cross-section of a tissue can be labeled with different sample barcode sequences (e.g., a moiety-conjugated barcode molecule, also referred to herein as a feature barcode). Nucleic acid analysis, for example sequencing analysis, can utilize the sample barcode sequences and spatial relationship of the barcode sequences to analyze various differences among subpopulations of cells and/or cell beads in the sample.

[0506] In some examples, a method for spatially mapping a plurality of cells and/or cell beads comprises labeling cells and/or cell beads of a different cell samples using nucleic acid barcode molecules to yield a plurality of labeled cell samples. An individual nucleic acid barcode molecule may comprise a sample barcode sequence, and nucleic acid barcode molecules of a given labeled cell sample can be distinguished from nucleic acid barcode molecules of another labeled cell sample by the sample barcode sequence. The nucleic acid barcode molecules may be arranged in at least a pre-defined two-dimensional configuration.

[0507] Next, nucleic acid molecules of the plurality of labeled cell samples may be subjected to one or more reactions to yield a plurality of barcoded nucleic acid products. Individual nucleic acid barcode products can comprise (i) a sample barcode sequence and (ii) a sequence corresponding to a nucleic acid molecule.

[0508] Next, the plurality of nucleic acid barcode products (or derivatives thereof) may be sequenced to yield sequencing reads. Spatial relationships may then be inferred between individual cell samples based on the sample barcode sequence and the pre-defined two-dimensional arrangement of nucleic acid barcode molecules, thereby spatially mapping a plurality of cell samples to at least a two dimensional configuration.

[0509] For example, a cell sample having at least two dimensions (e.g., a tissue section on a slide or a three-dimensional tissue sample from a subject, such as a fixed tissue sample) may be spotted with labelling nucleic acid barcode molecules comprising a labeling barcode sequence in a predefined pattern as described above. Cells are then dissociated from the cell sample and partitioned into a plurality of partitions, each partition comprising (1) a single cell from the cell sample, the single cell comprising at least one labelling nucleic acid barcode molecule comprising a labeling barcode sequence; and (2) a plurality of sample nucleic acid barcode molecules comprising a sample barcode sequence, wherein each partition comprises sample nucleic acid barcode molecules comprising a different sample barcode sequence. The plurality of sample nucleic acid barcode molecules further may comprise a unique molecular identifier (UMI) sequence. The plurality of

sample nucleic acid barcode molecules may be attached to a bead (e.g., a gel bead) and each partition comprises a single bead. In some cases, the labelling nucleic acid barcode molecules comprise one or more functional sequences, such as a primer sequence or a UMI sequence. In some instances, cells are lysed to release the labelling nucleic acid barcode molecule or other analytes present in or associated with the cells. In each partition, the labelling nucleic acid barcode molecules associated with each cell are barcoded by the sample nucleic acid barcode molecule to generate a nucleic acid molecule comprising the labeling barcode sequence and the sample barcode sequence. In addition to the barcoding of the labelling nucleic acid barcode molecules, another analyte such as RNA or DNA molecules may also be barcoded with a sample barcode sequence. Nucleic acid molecules barcoded with a sample barcode sequence can then be processed as necessary to generate a library suitable for sequencing as described elsewhere herein.

Three-Dimensional Spatial Mapping

[0510] Barcoded molecules (e.g., oligonucleotide-lipophilic moiety conjugates) may be used to target or label cells in suspension. In one aspect, cells within an intact tissue sample (e.g., a solid tissue sample) are contacted with these barcode molecules for spatial analysis. The present invention concerns methods and devices or instruments for injecting barcode molecules in situ into a tissue sample and subsequently identifying positions that correspond to uptake of the barcode molecules by cells within the tissue sample. In one aspect, oligonucleotide-lipophilic moiety conjugates (e.g., oligonucleotide-cholesterol conjugates) are used to label cells in a tissue sample. In one embodiment, the conjugates are injected into a tissue sample with a very fine needle (or array of needles). The location of each barcode molecule would have a defined position, e.g., in two dimensions (2D in one plane) or in three dimensions (3D in several planes). After injection of the conjugate, the barcode molecules insert into the plasma membrane of cells (e.g., via the lipophilic moiety) and diffuse within the tissue. At the point of injection, the concentration of the barcode would be the highest, and as it diffuses in the tissue its concentration would decrease. Considering this diffusion, the uptake of the barcode would define its location to the point of injection. With an array of needles (e.g., FIG. **74**), it would be possible to reconstruct cell position as cells take up different barcodes at different concentrations, thereby indicating the relative position of cells to each other. The barcoded molecules may also be applied to cells within a tissue sample using microarray nucleic acid printing methods known to those of ordinary skill in the art.

[0511] FIG. **74** depicts an example of a tissue section with barcode staining using one fixed array of needles (one 2-dimensional plane). x, y z may be determined depending on diffusion of the barcode. By way of example, a cell diameter of 10 m means the diffusion of barcodes will be on a scale of about 10-15 cells or about 100 μm-150 μm. A very fine needle can be used to infuse barcodes with or without pressure where the infusion can be in a skewer-like pattern separated by x μm apart in all directions (defined by desired diffusion of barcode). Each needle can infuse a different barcode.

[0512] FIG. **75** depicts a diffusion map to localize spatially barcodes and associated cells (one plane in 2D view). FIG. **76** shows the position of cells (designated "C1" to "C7") defined by the barcode and its relative amount (higher amount at the point of infusion, lower as cells are away from the point of diffusion). The amount of the different barcode in each cell defines its position in the tissue spatially. The following table illustrates this for cells C1 to C7 in a hypothetical scenario.

TABLE 1

Distribution of barcodes throughout cells.

Cell #	BC level: solid line	BC level: dashed line	BC level: dotted line
C1	++	–	–
C2	+++	+	–
C3	++	++	–
C4	+	+++	+
C5	–	++	++
C6	–	+	+++
C7	–	–	++

[0513] FIG. **77** depicts a three dimensional application. A fused needle at 3 levels is used to deliver 3 different barcodes. FIG. **78** depicts a three dimensional application to maximize 3D space with barcode staining.

[0514] In one embodiment, the present disclosure provides methods and compositions for spatial mapping where different barcode molecules are contacted with different regions of a 3D biological sample (e.g., a solid tissue sample). In one other embodiment, the biological sample comprises different regions of interest that may be contacted with barcode molecules. For instance, FIG. **79**A depicts regions of a mouse brain (P0-P8) with delivery devices (e.g., needles including fused or multipoint needles) for delivering barcode molecules (e.g., oligonucleotide-lipophilic moiety conjugates). The tissue sample (e.g., mouse brain or other solid tissue sample) is washed with a suitable media such as Hibernate Medium or HEB medium (Thermo Fisher Scientific), removed from the media, and any excess media allowed to drain before application of the barcode molecules. Multiple syringes (e.g., 2-3 L volume, mounted with 30 to 31 gauge needle) loaded with oligonucleotide-lipophilic moiety conjugates at a suitable concentration (e.g., about 0.1 M) for injection into the tissue sample at a depth of about 1 mm. At a fixed injection volume, the concentration of the conjugate can be adjusted depending on the resulting labeling of cells and the diffusion speed within the tissue. As depicted in FIG. **79**B, a first conjugate is injected at position A, a second conjugate at position B, a third conjugate at position C, and a fourth conjugate at position D according to a pattern. In one embodiment, position B is a first distance away from position A, position C is a second distance away from positions A and B, and position D is a third distance away from positions A and B. In other embodiments, the first distance is less than the second distance and/or greater than the third distance (e.g., Pattern 1 in FIG. **79**B).

[0515] In another embodiment, positions A-D are injected in a linear pattern, wherein each position is the same distance from the other in sequence. For example, position A is a first distance away from position B and a second distance away from position C, wherein the first distance is half of the second distance (e.g., Pattern 2 in FIG. **79**B). Those of ordinary skill in the art will appreciate that different

conjugates can be injected into a tissue sample according to the patterns shown in FIG. **79**B or any other suitable pattern.

[0516] Following injection, the tissue sample is incubated at room temperature or any other suitable temperature to allow the conjugates to diffuse into the tissue at their respective points of injection. After incubation, the tissue sample is placed in a 15 mL conical tube and washed again in HEB medium (e.g., washed twice). Following removal of the medium, the tissue sample is dissociated according to a suitable sample preparation protocol for single cell sequencing (e.g., 10× Genomics Sample Preparation Demonstrated Protocol—Dissociation of Mouse Embryonic Neural Tissue for Single Cell RNA Sequencing CG00055). Following dissociation, the suspension of cells from the tissue sample is processed to generate a sequencing library. As described herein, single cells (with the oligonucleotide-lipophilic moiety (e.g., cholesterol) conjugates inserted into their cell membranes) from the suspension of cells are provided in individual partitions with reagents for one or more additional barcoding reactions that involve analytes from the same single cells. Analytes from the suspension of cells are processed to provide nucleic acid libraries for sequencing (see, e.g., U.S. Pat. Nos. 10,011,872, 9,951,386, 10,030,267, and 10,041,116, which are incorporated herein by reference in their entireties). In one embodiment, barcode sequences of the plurality of oligonucleotide-lipophilic moiety conjugates are identified via sequencing along with barcode sequences associated with the analyte(s) processed from the single cells in suspension. In one embodiment, one or more barcode sequences from the plurality of oligonucleotide-lipophilic moiety conjugates are associated with one or more spatial positions corresponding to one or more cells within the tissue sample (see FIGS. **79**A-**79**B). In another embodiment, the spatial position corresponds to one or more cells where a particular oligonucleotide-lipophilic moiety conjugate diffused into the tissue sample (as determined by the pattern by which the oligonucleotide-lipophilic moiety conjugates were delivered to the tissue). In other embodiments, the one or more spatial positions are then associated with the analyte(s) detected and identified in the cell or cells into which the oligonucleotide-lipophilic moiety conjugate diffused. In one additional embodiment, a method of spatial analysis (e.g., three dimensional spatial analysis) using oligonucleotide-lipophilic moiety conjugates is provided. In one embodiment, the method comprises contacting a tissue sample (e.g., a solid tissue sample) with a plurality of oligonucleotide-lipophilic moiety conjugates at a plurality of locations within the sample. In another embodiment, the plurality of oligonucleotide-lipophilic moiety conjugates comprises a first, second, third, fourth, fifth, sixth, etc. types of oligonucleotide-lipophilic moiety conjugates. The type of oligonucleotide-lipophilic moiety conjugate may differ as to the sequence of the barcode and/or the type of lipophilic moiety. In one other embodiment, the method comprises allowing the plurality of oligonucleotide-lipophilic moiety conjugates to diffuse into the tissue sample, such that the plurality of oligonucleotide-lipophilic moiety conjugates insert into cell membranes of the cells within the tissue sample. In additional embodiments, the method comprises providing a suspension of cells (e.g., single cells) that are derived from the tissue sample (containing the diffused oligonucleotide-lipophilic moiety conjugates), such that the suspension comprises one or more cells that retain one or more oligonucleotide-lipophilic moiety conjugates of the plurality of oligonucleotide-lipophilic moiety conjugates. In one more embodiment, the method comprises providing a nucleic acid library for sequencing from the suspension of cells. In one embodiment, the nucleic acid library comprises nucleic acid barcode molecules corresponding to an oligonucleotide-lipophilic moiety conjugate and an analyte (as described herein), including without limitation, a nucleic acid analyte, a metabolite analyte, and a protein analyte.

[0517] In one aspect, the present invention provides methods of processing a tissue sample for spatial analysis. In one embodiment, the method comprises the step of delivering a plurality of spatial oligonucleotides to a location in a tissue sample, wherein a spatial oligonucleotide of the plurality of spatial oligonucleotides comprises (i) a spatial barcode sequence and (ii) a cell membrane labeling (or targeting) agent to label a cell at the location in the tissue sample. In one embodiment, the cell membrane labeling agent interacts with or associates with the cell membrane as further described herein (e.g., lipophilic molecules, fluorophores, dyes, etc.). In another embodiment, the spatial oligonucleotide further comprises a cleavable linker (such as a linker described herein) to allow separation of the spatial barcode sequence from the cell membrane labeling agent. In another embodiment, the plurality of spatial oligonucleotides may be delivered to the tissue sample in a pattern as described herein. In another embodiment, the method further comprises the step of dissociating the tissue sample into a plurality of cells, wherein a cell of the plurality of cells is a single cell that comprises the spatial oligonucleotide and an analyte of interest. In another embodiment, the single cell comprises the spatial oligonucleotide via the cell membrane labeling agent. In another embodiment, the method further comprises the step of partitioning the single cell with a (i) plurality of cell barcode nucleic acid molecules each comprising a cell barcode sequence and configured to couple to the analyte and (ii) a plurality of spatial barcode nucleic acid molecules configured to couple to the spatial oligonucleotide. In another embodiment, the method further comprises the step of in the partition, lysing the single cell and using the spatial oligonucleotide and the analyte of interest to generate (i) a first barcoded nucleic acid molecule comprising the spatial barcode sequence or a complement thereof, and (ii) a second barcoded nucleic acid molecule comprising the cell barcode sequence or a complement thereof. In other embodiments, the method further comprises the step of sequencing (i) the first barcoded nucleic acid molecule to determine the spatial barcode sequence, and (ii) the second barcoded nucleic acid molecule to determine the cell barcode sequence. In further embodiments, the method also comprises the step of using (i) the determined spatial barcode sequence to identify the location in the tissue sample at which the single cell was labelled and/or from which the single cell originated, and (ii) the determined cell barcode sequence to identify the analyte as originating from the single cell. In another embodiment, the cell membrane labeling agent is selected from the group consisting of a lipid (e.g., a lipophilic moiety), a fluorophore, a dye, a peptide, and a nanoparticle. In another embodiment, the analyte is a nucleic acid molecule or a protein labelling agent capable of specifically binding to a surface protein on the cell. In another embodiment, each cell barcode nucleic acid molecule further comprises a cleavable linker (such as a linker described herein) to allow separation of the cell barcode sequence from the protein labeling agent. In other embodi-

ments, the method is suitable for processing tissue samples for two dimensional (e.g., tissue section or sample on a slide) and three dimensional (e.g., biopsy from a subject) spatial analysis.

Doublet Reduction and Detection

[0518] The present disclosure also provides methods and compositions for doublet reduction. In an aspect, a method of analyzing polynucleotides may comprise labeling cells and/or cell beads of different cell samples (e.g., cell samples from different subjects, such as different humans or animals; cell samples from the same subject taken at different times; and/or cell samples from the same subject taken from different areas or features of a subject, such as from different tissues) using nucleic acid barcode molecules or oligonucleotides comprising the nucleic acid barcode molecules to yield a plurality of labeled cell samples, wherein an individual nucleic acid barcode molecule comprises a sample barcode sequence (e.g., a moiety-conjugated barcode molecule, also referred to herein as a feature barcode), and wherein nucleic acid barcode molecules of a given labeled cell sample are distinguishable from nucleic acid barcode molecules of another labeled cell sample by the sample barcode sequence. Different cells and/or cell beads from the same cell sample may have the same sample barcode sequence. Labeled cells and/or cell beads of the plurality of cell samples may be co-into a plurality of partitions. The labeled cells and/or cell beads may be co-partitioned with a plurality of beads, such as a plurality of gel beads. Beads of the plurality of beads may comprise a plurality of bead nucleic acid barcode molecules attached (e.g., releasably coupled) thereto, wherein an individual bead nucleic acid barcode molecule attached to a bead comprises a bead barcode sequence. Bead nucleic acid barcode molecules of a given bead may e distinguishable from bead nucleic acid barcode molecules of another bead by their bead barcode sequence(s). Nucleic acid molecules of the at least one labeled cell and/or cell bead of a given partition may be subjected to one or more reactions to yield nucleic acid barcode products comprising (i) a sample barcode sequence, (ii) a bead barcode sequence, and (iii) a sequence corresponding to a nucleic acid molecule of the nucleic acid molecules of the at least one labeled cell and/or cell bead. Nucleic acid barcode products may be subjected to sequencing to yield a plurality of sequencing reads. In some cases, contents of a plurality of partitions may be pooled to provide a plurality of nucleic acid barcode products corresponding to the plurality of partitions. Sequencing reads may be processed to identify bead and sample barcode sequences, which sequences may be used to identify the cell and/or cell bead to which a sequencing read corresponds. For example, sequencing reads corresponding to two different cells and/or cell beads from different cell samples that are co-partitioned in the same partition may be identified as having identical bead barcode sequences and different sample barcode sequences. Sequencing reads corresponding to two different cells and/or cell beads from the same cell sample partitioned in different partitions may be identified as having different bead barcode sequences and identical sample barcode sequences.

[0519] As described elsewhere herein, a sample barcode sequence which is used to label individual cells and/or cell beads of a cell sample can later be used as a mechanism to associate a cell and/or cell bead and a given cell sample. For example, a plurality of cell samples can be uniquely labeled with nucleic acid barcode molecules such that the cells and/or cell beads of a particular sample can be identified as originating from the particular sample, even if the particular cell sample were mixed with additional cell samples and subjected to nucleic acid processing in bulk.

[0520] Individual nucleic acid barcode molecules may form a part of a barcoded oligonucleotide. A barcoded oligonucleotide, as described elsewhere herein, can comprise sequence elements in addition to a sample barcode sequence that may serve a variety of purposes, for example in sample preparation for sequencing analysis, e.g., next-generation sequence analysis.

[0521] Cells and/or cell beads can be labeled with nucleic acid barcode molecules by any of a variety of suitable mechanisms described elsewhere herein. A nucleic acid barcode molecule or a barcoded oligonucleotide comprising the nucleic acid barcode molecule may be linked to a moiety ("barcoded moiety") such as an antibody or an epitope binding fragment thereof, a cell surface receptor binding molecule, a receptor ligand, a small molecule, a pro-body, an aptamer, a monobody, an affimer, a darpin, or a protein scaffold. The moiety to which a nucleic acid barcode molecule or barcoded oligonucleotide can be linked may bind a molecule expressed on the surface of individual cells of the plurality of cell samples. A labeled cell sample may refer to a sample in which the cells and/or cell beads are bound to barcoded moieties. A labeled cell sample may refer to a sample in which the cells have nucleic acid barcode molecules within the cells and/or cell beads.

[0522] A molecule (e.g., a molecule expressed on the surface of individual cells of the plurality of cell samples) may be common to all cells and/or cell beads of the plurality of the different cell samples. The molecule may be a protein. Exemplary proteins in embodiments herein include, but are not limited to, transmembrane receptors, major histocompatibility complex proteins, cell-surface proteins, glycoproteins, glycolipids, protein channels, and protein pumps. A non-limiting example of a cell-surface protein can be a cell adhesion molecule. The molecule may be expressed at similar levels for all cells and/or cell beads of the sample. The expression of the molecule for all cells and/or cell beads of a sample may be within biological variability. The molecule may be differentially expressed in cells and/or cell beads of the cell sample. The expression of the molecule for all cells and/or cell beads of a sample may not be within biological variability, and some of the cells and/or cell beads of a cell sample may be and/or comprise abnormal cells. A moiety linked to a nucleic acid barcode molecule or barcoded oligonucleotide may bind a molecule that is present on a majority of the cells and/or cell beads of a cell sample. The molecule may be present on at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% of the cells and/or cell beads in a cell sample.

[0523] Cells and/or cell beads can be labeled in (a) by any suitable mechanism, including those described elsewhere herein. The nucleic acid barcode molecule or barcoded oligonucleotide comprising the nucleic acid barcode molecule may be linked to an antibody or an epitope binding fragment thereof, and labeling cells and/or cell beads may comprise subjecting the antibody-linked nucleic acid barcode molecule or the epitope binding fragment-linked nucleic acid barcode molecule to conditions suitable for binding the antibody or the epitope binding fragment thereof

to a molecule present on a cell surface. The nucleic acid barcode molecule or barcoded oligonucleotide comprising the nucleic acid barcode molecule may be coupled to a cell-penetrating peptide (CPP), and labeling cells and/or cell beads may comprise delivering the CPP coupled nucleic acid barcode molecule into a cell and/or cell bead by the CPP. The nucleic acid barcode molecule or barcoded oligonucleotide comprising the nucleic acid barcode molecule may be conjugated to a cell-penetrating peptide (CPP), and labeling cells and/or cell beads may comprise delivering the CPP conjugated nucleic acid barcode molecule into a cell and/or cell bead by the CPP. The nucleic acid barcode molecule or barcoded oligonucleotide comprising the nucleic acid barcode molecule may be coupled to a lipophilic molecule, and labeling cells and/or cell beads may comprise delivering the nucleic acid barcode molecule to a cell membrane by the lipophilic molecule. The nucleic acid barcode molecule or barcoded oligonucleotide comprising the nucleic acid barcode molecule may enter into the intracellular space. The nucleic acid barcode molecule or barcoded oligonucleotide comprising the nucleic acid barcode molecule may be coupled to a lipophilic molecule, and labeling cells may comprise delivering the nucleic acid barcode molecule to a nuclear membrane by the lipophilic molecule. The nucleic acid barcode molecule or barcoded oligonucleotide comprising the nucleic acid barcode molecule may enter into a cell nucleus. Labeling cells and/or cell beads may comprise use of a physical force or chemical compound to deliver the nucleic acid barcode molecule or barcoded oligonucleotide into the cell and/or cell bead. Examples of physical methods that can be used in the methods provided herein include the use of a needle, ballistic DNA, electroporation, sonoporation, photoporation, magnetofection, and hydroporation. Various chemical compounds can be used in the methods provided herein to deliver nucleic acid barcode molecules to a cell. Chemical vectors, as previously described herein, can include inorganic particles, lipid-based vectors, polymer-based vectors and peptide-based vectors. In some cases, labeling cells and/or cell beads may comprise use of a cationic lipid, such as a liposome. A labeled cell sample may refer to a sample in which the cells and/or cell beads have nucleic acid barcode molecules within the cells and/or cell beads.

[0524] Following labeling of cells and/or cell beads, a majority of the cells and/or cell beads of a particular cell sample can be labeled with nucleic acid barcode molecules having a sample specific barcode sequence. At least 50%, 60%, 70%, 75%, 80%, 85%, 90%, or 95% of cells of a cell sample may be labeled. In some cases, not all of the cells and/or cell beads of a given cell sample of a plurality of cell samples are labeled. Less than 100%, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, or 50% of cells and/or cell beads of a cell sample may be labeled. In some cases, cells and/or cell beads of multiple different cell samples of the plurality cell samples may not be labeled.

[0525] The plurality of labeled cell samples can be co-partitioned with a plurality of beads into a plurality of partitions. Individual beads can comprise a plurality of bead nucleic acid barcode molecules attached thereto. Bead nucleic acid barcode molecules of a given bead can be distinguishable from bead nucleic acid barcode molecules of another bead by a bead barcode sequence. The bead nucleic acid barcode molecule may be releasably attached to the bead. The bead may be degradable upon application of a stimulus. The stimulus may comprise a chemical stimulus.

[0526] By partitioning the labeled cell samples into a plurality of partitions, one or more reactions can be performed individually for single cells in isolated partitions. In some cases, the partition is an aqueous droplet in a non-aqueous phase such as oil. The partitions comprise droplets. For example, a partition can be a droplet in an emulsion. Alternatively, the partitions may comprise wells or tubes.

[0527] Individual partitions may comprise a single cell and/or cell bead. Alternatively or in addition, a subset of partitions may contain more than a single cell and/or cell bead.

[0528] Nucleic acids generated in partitions having more than a single cell and/or cell bead may undesirably assign the same bead barcode sequence to two different cells and/or cell beads. While the nucleic acids may share the same bead barcode sequence, the two different cells and/or cell beads can be distinguished by different sample barcode sequences if the two cells and/or cell beads originated from different cell samples. By using both a sample barcode sequence (e.g., a moiety-conjugated barcode molecule) and a bead (or partition) barcode sequence, sequencing reads from partitions comprising more than one labeled cell and/or cell bead can be identified.

[0529] A method of the present disclosure may comprise pooling a plurality of nucleic acid barcode products from partitions prior to subjecting the nucleic acid barcode products, or derivatives thereof, to an assay such as nucleic acid sequencing. Nucleic acid barcode products may be subjected to processing such as nucleic acid amplification. In some cases, one or more features such as one or more functional sequences (e.g., sequencing primers and/or flow cell adapter sequences) may be added to nucleic acid barcode products, e.g., after pooling of nucleic acid barcode products from the partitions. For example, pooled amplification products may be subjected to one or more reactions prior to sequencing. For example, the pooled nucleic acid barcode products may be subjected to one or more additional reactions (e.g., nucleic acid extension, polymerase chain reaction, or adapter ligation). Adapter ligation may include, for example, fragmenting the nucleic acid barcode products (e.g., by mechanical shearing or enzymatic digestion) and enzymatic ligation.

Cell Characterization

[0530] In an aspect, the methods provided herein may be useful in identifying and/or characterizing cells and/or cell beads. For example, the present disclosure provides a method of identifying a size of a cell and/or cell bead. By identifying the size of the cell, other properties, such as its type and/or tissue of origin may also be determined.

[0531] Cells of different sizes (e.g., diameters) will have different associated cell surfaces. For example, a first cell of a first size may have a different surface area and surface features than a second cell of a second size that is larger than the first size. As described herein, lipophilic or amphiphilic moieties (e.g., coupled to nucleic acid barcode molecules) may associate with and/or insert into membranes of cells and/or cell beads. At a non-saturating concentration of lipophilic or amphiphilic moieties (e.g., coupled to nucleic acid barcode molecules), uptake of the lipophilic or amphiphilic moieties by a cell or cell bead may be proportional to the surface of the cell or cell bead. Accordingly, a second

cell or cell bead that is larger than a first cell or cell bead (e.g., has a larger diameter and, accordingly, a larger surface area, than the first cell or cell bead) may uptake more lipophilic or amphiphilic moieties than the first cell or cell bead (see, e.g., FIGS. **80** and **81**).

[0532] Identifying or characterizing cells and/or cell beads may comprise measuring uptake of lipophilic or amphiphilic moieties (e.g., coupled to nucleic acid barcode molecules) by the cells and/or cell beads. A known amount of lipophilic and/or amphiphilic moieties (e.g., coupled to nucleic acid barcode molecules) may be provided to a cell or cell bead or a collection of cells or cell beads and the uptake of such moieties may be measured. Uptake of such moieties by cells may be measured by, for example, measuring a residual amount of such moieties that are not taken up by cells and subtracting this amount from the initial known amount. In another example, lipophilic and/or amphiphilic moieties may be labeled (e.g., with optically detectable labels such as fluorescent moieties) and the labels may be used to determine a relative uptake of the lipophilic and/or amphiphilic moieties by the cell/cell bead and/or cells/cell beads (e.g., using an optical detection method). In another example, the amount of lipophilic/amphiphilic moieties (e.g., coupled to nucleic acid barcode molecules) taken up by cells and/or cell beads may be determined by measuring the amount of nucleic acid barcode molecules associated with the cells and/or cell beads (e.g., using nucleic acid sequencing). Such a method may provide an alternative to other methods of determining cell size, such as flow cytometry.

[0533] In an example, a plurality of cells may be labeled with lipophilic or amphiphilic feature barcodes (e.g., as described herein). Feature barcodes comprising a lipophilic moiety (e.g., a cholesterol moiety) may be incubated with the plurality of cells. The feature barcodes may comprise an optical label such as a fluorescent moiety. The feature barcodes may include, for example, a sequence configured to hybridize to a nucleic acid barcode molecule, such as a sequence comprising multiple cytosine nucleotides (e.g., a CCC sequence). Each feature barcode may also comprise a barcode sequence and/or a unique molecular identifier (UMI) sequence. Each lipophilic or amphiphilic moiety may be coupled to a different UMI sequence. For example, where about 1 million lipophilic or amphiphilic moieties will be used, about 1 million different UMI sequences may be used. Alternatively, each lipophilic or amphiphilic moiety may be coupled to a different combination of UMI and barcode sequences. For example, where about 1 million lipophilic or amphiphilic moieties will be used, about 1 million different combinations may be used. Cells may be partitioned into a plurality of partitions (e.g., a plurality of droplets, such as aqueous droplets in an emulsion) with a plurality of partition nucleic acid barcode molecules, where each nucleic acid barcode molecule of the plurality of partition nucleic acid barcode molecules comprises a barcode sequence. Each partition may comprise at most one cell. The plurality of partition nucleic acid barcode molecules may be distributed throughout the partitions such that each partition includes nucleic acid barcode molecules having a different barcode sequence, where a given partition of the plurality of partitions may include multiple nucleic acid barcode molecules having the same barcode sequence. Nucleic acid barcode molecules may be coupled (e.g., releasably coupled) to beads (e.g., gel beads). In addition to barcode sequences, nucleic acid barcode molecules may further comprise unique molecule identifier sequences and/or sequences configured to hybridize to feature barcodes coupled to the lipophilic or amphiphilic moieties (e.g., GGG sequences). Within each partition comprising a cell, partition nucleic acid barcode molecules may couple to feature barcodes coupled to lipophilic or amphiphilic moieties, such that cells comprise a plurality of lipophilic or amphiphilic moieties coupled to i) feature barcodes and ii) partition nucleic acid barcode molecules. The barcode sequences of the partition nucleic acid barcode molecules are uniform across the plurality of lipophilic or amphiphilic moieties and identify the cell as corresponding to a given partition, while the diversity of barcode and/or UMI sequences of the feature barcodes is proportional to the uptake of lipophilic or amphiphilic moieties by the cell, and thus to the cell size. Accordingly, upon sequencing the feature barcodes coupled to the partition nucleic acid barcode molecules (e.g., subsequent to derivitization of the feature barcodes coupled to the partition nucleic acid barcode molecules with, e.g., flow cell adapters), a plurality of sequencing reads may be obtained that may be associated with the cells to which the feature barcodes and partition nucleic acid barcode molecules corresponded. The number of barcode and/or UMI sequences of the feature barcodes may be used to determine a relative size of the cells with which they are associated (e.g., a larger cell will have more barcode and/or UMI sequences associated therewith than a smaller cell) (see, e.g., FIG. **82**).

[0534] In another example, a plurality of cells may be labeled with lipophilic or amphiphilic feature barcodes (e.g., as described herein). Feature barcodes comprising a lipophilic moiety (e.g., a cholesterol moiety) may be incubated with a plurality of cells. The feature barcodes may comprise an optical label such as a fluorescent moiety. The feature barcodes may include, for example, a sequence configured to hybridize to a nucleic acid barcode molecule, such as a sequence comprising multiple cytosine nucleotides (e.g., a CCC sequence). Each feature barcode may also comprise a barcode sequence and/or a unique molecular identifier sequence. A plurality of beads (e.g., gel beads) each comprising a plurality of nucleic acid barcode molecules may be provided. The nucleic acid barcode molecules of each bead (e.g., releasably attached to each bead) may comprise a barcode sequence (e.g., cell barcode sequence), a unique molecular identifier sequence, and a sequence configured to hybridize to a feature barcode. Nucleic acid barcode molecules of each different bead may comprise the same barcode sequence, which barcode sequence differs from barcode sequences of nucleic acid barcode molecules of other beads of the plurality of beads. The cells incubated with feature barcodes may be partitioned (e.g., subsequent to one or more washing processes) with the plurality of beads into a plurality of partitions (e.g., droplets, such as aqueous droplets in an emulsion) such that at least a subset of the plurality of partitions each comprise a single cell and a single bead. Within each partition of the at least a subset of the plurality of partitions, one or more nucleic acid barcode molecules of the bead may attach (e.g., hybridize or ligate) to one or more feature barcodes of the cell. The one or more nucleic acid barcode molecules of the bead may be released (e.g., via application of a stimulus, such as a chemical stimulus) from the bead within the partition prior to attachment of the one or more nucleic acid barcode molecules to the one or more feature barcodes of the cell to provide a barcoded feature barcode. The cell may be lysed or permea-

bilized within the partition to provide access to analytes therein, such as nucleic acid molecules therein (e.g., deoxyribonucleic acid (DNA) molecules and/or ribonucleic acid (RNA) molecules), and/or to the feature barcode therein (e.g., if the feature barcode has permeated the cell membrane). One or more analytes (e.g., nucleic acid molecules) of the cell may also be barcoded within the partition with one or more nucleic acid barcode molecules of the bead to provide a plurality of barcoded analytes (e.g., barcoded nucleic acid molecules). The plurality of partitions comprising barcoded analytes and barcoded feature barcodes may be combined (e.g., pooled). Additional processing may be performed to, for example, prepare the barcoded analytes and barcoded feature barcodes for subsequent analysis. For example, barcoded nucleic acid molecules and/or barcoded feature barcodes may be derivatized with flow cell adapters to facilitate nucleic acid sequencing. Barcodes of barcoded analytes and barcoded feature barcodes may be detected using nucleic acid sequencing and used to identify the barcoded analytes and barcoded feature barcodes as deriving from particular cells or cell types of the plurality of cells. The relative abundance of a given sequence (e.g., barcode or UMI sequence) measured in a sequencing assay may provide an estimate of the size of various cells of the plurality of cells. For example, a first barcode sequence associated with a first cell (e.g., via a feature barcode and/or a partition nucleic acid barcode sequence of a nucleic acid barcode molecule of a bead co-partitioned with the first cell) may appear in greater number than a second barcode sequence associated with a second cell, indicating that the first cell is larger than the second cell. Barcode sequences and UMIs associated with cellular debris (e.g., cellular components and/or damaged cells) may have few lipophilic or amphiphilic moieties associated therewith and may therefore contribute only minimally to distributions of barcode sequences vs. cell counts (see, e.g., FIG. **82**).

Cell Multiplexing and Hashing

[0535] As described herein, in an aspect, the present disclosure provides methods for simultaneously processing multiple analytes derived from the same or different samples. Such a method may comprise, for example, providing a first nucleic acid barcode sequence (e.g., as a component of a cell nucleic acid barcode molecule) to a first sample and a second nucleic acid barcode sequence to a second sample such that cells or other analytes associated with the first sample are labeled with the first nucleic acid barcode sequence and cells or other analytes associated with the second sample are labeled with the second nucleic acid barcode sequence. The nucleic acid barcode sequences may be components of nucleic acid barcode molecules that also comprise lipophilic moieties (such as cholesterol moieties, e.g., as described herein). Cells may be labeled by, for example, binding cell binding moieties coupled to nucleic acid barcode sequences to the cells. Such cell binding moieties may be, for example, antibodies, cell surface receptor binding molecules, receptor ligands, small molecules, pro-bodies, aptamers, monobodies, affimers, darpins, or protein scaffolds (e.g., as described herein). Cell binding moieties may bind to a protein and/or a cell surface species of the cells. Alternatively, cells may be labeled by delivering nucleic acid barcode molecules (e.g., as described herein) to the cells, optionally using cell-penetrating peptides, liposomes, nanoparticles, electroporation, or mechanical force (e.g., as described herein). Nucleic acid barcode molecules may comprise barcode sequences unique to a cell sample and/or to an individual cell within a cellular sample. Labeled cells (and/or other analytes) may be partitioned between a plurality of partitions (e.g., as described herein), which partitions may comprise one or more reagents, such as one or more partition nucleic acid barcode sequences. Each partition may comprise a different partition nucleic acid barcode sequence. Some partitions may comprise more than one labeled cell (e.g., as described herein). For example, partitions (e.g., droplets or wells) may be intentionally loaded in such a manner that more partitions including more than one cell than would be achieved according to Poisson statistics (e.g., partitions may be overloaded). At least two labeled cells may be identified as originating from a same partition using the nucleic acid barcode sequences with which the cells are labeled, or complements thereof, and the partition nucleic acid barcode sequences associated with the partition, or complements thereof. Such identification may be facilitated by synthesizing barcoded nucleic acid products from the plurality of labeled cells (e.g., as described herein), which a given barcoded nucleic acid product may comprise a cell identification sequence comprising a cell nucleic acid barcode sequence or complement thereof and a partition identification sequence comprising a partition nucleic acid barcode sequence or complement thereof. Synthesizing the barcoded nucleic acid products may comprise hybridizing a sequence of a partition nucleic acid barcode molecule to a cell nucleic acid barcode molecule and performing an extension reaction (e.g., as described herein). Such methods may facilitate assignation of cells to their samples of origin, as well as the identification of multiplets originating from multiple samples (e.g., as described herein).

[0536] Single cell processing and analysis methods and systems such as those described herein can be utilized for a wide variety of applications, including analysis of specific individual cells, analysis of different cell types within populations of differing cell types, analysis and characterization of large populations of cells for environmental, human health, epidemiological forensic, or any of a wide variety of different applications.

[0537] One application of the methods described herein is in the sequencing and characterization of immune cells. Methods and compositions disclosed herein can be utilized for sequence analysis of the immune repertoire. Analysis of sequence information underlying the immune repertoire can provide a significant improvement in understanding the status and function of the immune system.

[0538] Non-limiting examples of immune cells which can be analyzed utilizing the methods described herein include B cells, T cells (e.g., cytotoxic T cells, natural killer T cells, regulatory T cells, and T helper cells), natural killer cells, cytokine induced killer (CIK) cells; myeloid cells, such as granulocytes (basophil granulocytes, eosinophil granulocytes, neutrophil granulocytes/hypersegmented neutrophils), monocytes/macrophages, mast cell, thrombocytes/megakaryocytes, and dendritic cells. In some embodiments, individual T cells are analyzed using the methods disclosed herein. In some embodiments, individual B cells are analyzed using the methods disclosed herein.

[0539] Immune cells express various adaptive immunological receptors relating to immune function, such as T cell receptors and B cell receptors. T cell receptors and B cells

receptors play a part in the immune response by specifically recognizing and binding to antigens and aiding in their destruction.

[0540] The T cell receptor, or TCR, is a molecule found on the surface of T cells that is generally responsible for recognizing fragments of antigen as peptides bound to major histocompatibility complex (MHC) molecules. The TCR is generally a heterodimer of two chains, each of which is a member of the immunoglobulin superfamily, possessing an N-terminal variable (V) domain, and a C terminal constant domain. In humans, in 95% of T cells the TCR consists of an alpha (?) and beta (?) chain, whereas in 5% of T cells the TCR consists of gamma and delta (?/?) chains. This ratio can change during ontogeny and in diseased states as well as in different species. When the TCR engages with antigenic peptide and MHC (peptide/MHC), the T lymphocyte is activated through signal transduction.

[0541] Each of the two chains of a TCR contains multiple copies of gene segments—a variable 'V' gene segment, a diversity 'D' gene segment, and a joining 'J' gene segment. The TCR alpha chain is generated by recombination of V and J segments, while the beta chain is generated by recombination of V, D, and J segments. Similarly, generation of the TCR gamma chain involves recombination of V and J gene segments, while generation of the TCR delta chain occurs by recombination of V, D, and J gene segments. The intersection of these specific regions (V and J for the alpha or gamma chain, or V, D and J for the beta or delta chain) corresponds to the CDR3 region that is important for antigen-MHC recognition. Complementarity determining regions (e.g., CDR1, CDR2, and CDR3), or hypervariable regions, are sequences in the variable domains of antigen receptors (e.g., T cell receptor and immunoglobulin) that can complement an antigen. Most of the diversity of CDRs is found in CDR3, with the diversity being generated by somatic recombination events during the development of T lymphocytes. A unique nucleotide sequence that arises during the gene arrangement process can be referred to as a clonotype.

[0542] The B cell receptor, or BCR, is a molecule found on the surface of B cells. The antigen binding portion of a BCR is composed of a membrane-bound antibody that, like most antibodies (e.g., immunoglobulins), has a unique and randomly determined antigen-binding site. The antigen binding portion of a BCR includes membrane-bound immunoglobulin molecule of one isotype (e.g., IgD, IgM, IgA, IgG, or IgE). When a B cell is activated by its first encounter with a cognate antigen, the cell proliferates and differentiates to generate a population of antibody-secreting plasma B cells and memory B cells. The various immunoglobulin isotypes differ in their biological features, structure, target specificity and distribution. A variety of molecular mechanisms exist to generate initial diversity, including genetic recombination at multiple sites.

[0543] The BCR is composed of two genes IgH and IgK (or IgL) coding for antibody heavy and light chains. Immunoglobulins are formed by recombination among gene segments, sequence diversification at the junctions of these segments, and point mutations throughout the gene. Each heavy chain gene contains multiple copies of three different gene segments—a variable 'V' gene segment, a diversity 'D' gene segment, and a joining 'J' gene segment. Each light chain gene contains multiple copies of two different gene segments for the variable region of the protein—a variable 'V' gene segment and a joining 'J' gene segment. The recombination can generate a molecule with one of each of the V, D, and J segments. Furthermore, several bases may be deleted and others added (called N and P nucleotides) at each of the two junctions, thereby generating further diversity. After B cell activation, a process of affinity maturation through somatic hypermutation occurs. In this process progeny cells of the activated B cells accumulate distinct somatic mutations throughout the gene with higher mutation concentration in the CDR regions leading to the generation of antibodies with higher affinity to the antigens. In addition to somatic hypermutation activated B cells undergo the process of isotype switching. Antibodies with the same variable segments can have different forms (isotypes) depending on the constant segment. Whereas all naïve B cells express IgM (or IgD), activated B cells mostly express IgG but also IgM, IgA and IgE. This expression switching from IgM (and/or IgD) to IgG, IgA, or IgE occurs through a recombination event causing one cell to specialize in producing a specific isotype. A unique nucleotide sequence that arises during the gene arrangement process can similarly be referred to as a clonotype.

[0544] In some embodiments, the methods, compositions and systems disclosed herein are utilized to analyze the various sequences of TCRs and BCRs from immune cells, for example various clonotypes. In some embodiments, methods, compositions and systems disclosed herein are used to analyze the sequence of a TCR alpha chain, a TCR beta chain, a TCR delta chain, a TCR gamma chain, or any fragment thereof (e.g., variable regions including VDJ or VJ regions, constant regions, transmembrane regions, fragments thereof, combinations thereof, and combinations of fragments thereof). In some embodiments, methods, compositions and systems disclosed herein are used to analyze the sequence of a B cell receptor heavy chain, B cell receptor light chain, or any fragment thereof (e.g., variable regions including VDJ or VJ regions, constant regions, transmembrane regions, fragments thereof, combinations thereof, and combinations of fragments thereof).

[0545] Where immune cells are to be analyzed, primer sequences useful in any of the various operations for attaching barcode sequences and/or extension/amplification reactions may comprise gene specific sequences which target genes or regions of genes of immune cell proteins, for example immune receptors. Such gene sequences include, but are not limited to, sequences of various T cell receptor alpha variable genes (TRAV genes), T cell receptor alpha joining genes (TRAJ genes), T cell receptor alpha constant genes (TRAC genes), T cell receptor beta variable genes (TRBV genes), T cell receptor beta diversity genes (TRBD genes), T cell receptor beta joining genes (TRBJ genes), T cell receptor beta constant genes (TRBC genes), T cell receptor gamma variable genes (TRGV genes), T cell receptor gamma joining genes (TRGJ genes), T cell receptor gamma constant genes (TRGC genes), T cell receptor delta variable genes (TRDV genes), T cell receptor delta diversity genes (TRDD genes), T cell receptor delta joining genes (TRDJ genes), and T cell receptor delta constant genes (TRDC genes).

[0546] Additionally the methods and compositions disclosed herein, allow the determination of not only the immune repertoire and different clonotypes, but the functional characteristics (e.g., the transcriptome) of the cells associated with a clonotype or plurality of clonotypes that

bind to the same or similar antigen. These functional characteristics can comprise transcription of cytokine, chemokine, or cell-surface associated molecules, such as, costimulatory molecules, checkpoint inhibitors, cell surface maturation markers, or cell-adhesion molecules. Such analysis allows a cell or cell population expressing a particular T cell receptor, B cell receptor, or immunoglobulin to be associated with certain functional characteristics. For example, for any given antigen there will be multiple clonotypes of T cell receptor, B cell receptor, or immunoglobulin that specifically bind to that antigen. Multiple clonotypes that bind to the same antigen are known as the idiotype.

[0547] The present disclosure also provides methods for reducing nonspecific priming in a single-cell 5? gene expression assay. In generating an assay that allows measurement of 1) a cell barcode sequence (barcode), 2) a unique molecular identifier sequence (UMI) and 3) the 5? sequence of an mRNA transcript simultaneously, one strategy is to place these sequences on a sequence that attaches to the 5? end of an mRNA transcript—in the present disclosure, this may be accomplished by placing the barcode and UMI on a template switching oligonucleotide (TSO). This oligonucleotide may be attached to the first strand cDNA via a template switching reaction where the reverse transcription (RT) enzyme 1) reverse transcribes a messenger RNA (mRNA) sequence into first-strand complementary DNA (cDNA) from a primer targeting the 3? end of the mRNA, 2) adds nontemplated cytidines to the 5? end of the first-strand cDNA, 3) switches template to the TSO, which may contain 3? guanidines or guanidine-derivatives that hybridize to the added cytidines. The result is a first-strand cDNA molecule that is complementary to the TSO sequence: cell-barcode, UMI, guanidines, and the 5? end of the mRNA.

[0548] In some cases, the TSO may co-exist in solution with the RT enzyme and the total RNA contents of a cell. If the TSO is a single stranded DNA (ssDNA) molecule, it can participate as an RT primer rather than as a template-switching substrate. Given, for example, that the over 90% of the total RNA contents of a cell include noncoding ribosomal RNA (rRNA), this may produce barcoded off products that do not contribute to the 5? gene expression or V(D)J sequencing assay but do consume sequencing reads, increasing the cost required to achieve the same sequencing depth. In addition, if the UMI is implemented as a randomer, the presence of this randomer at the 3? end of the TSO greatly increases its ability to serve as a primer on rRNA template.

[0549] In some cases, a TSO that is less likely to serve as an RT primer via the introduction of a particular spacer sequence between the UMI and terminal riboGs may be used. Another approach is to design and include a set of auxiliary blocking oligonucleotides that may hybridize to rRNA and prevent binding of the TSO.

[0550] The spacer sequence can be optimized by selecting a sequence that minimizes the predicted melting temperature of the (spacer-GGG):rRNA duplex against all human ribosomal RNA molecules.

[0551] The blocker sequences can be optimized by selecting sequences that maximize the predicted melting temperature of the (blocker):rRNA duplex against all human ribosomal RNA molecules.

[0552] Provided herein are TSO that are less likely to serve as an RT primer via the introduction of a particular spacer sequence between the UMI and terminal riboGs. Additionally, described herein are auxiliary blocking oligonucleotides that hybridize to rRNA and prevent binding of the TSO.

[0553] Examples of spacer sequences, blocker sequences, and full construct barcodes that may of use in the methods provided herein can be found in at least U.S. Patent Publication No. 201801058008, which is herein incorporated by reference in its entirety.

[0554] In some examples, a cell barcode may be a 16 base sequence that is a random choice from about 737,000 sequences. The length of the barcode (16) can be altered. The diversity of potential barcode sequences (737 k) can be alterable. The defined nature of the barcode can be altered, for example, it may also be completely random (16 Ns) or semi-random (16 bases that come from a biased distribution of nucleotides).

[0555] The canonical UMI sequence may be a 10 nucleotide randomer. The length of the UMI can be altered. The random nature of the UMI can be altered, for example, it may be semi-random (bases that come from a biased distribution of nucleotides.) In a certain case, the distribution of UMI nucleotide(s) may be biased; for example, UMI sequences that do not contain Gs or Cs may be less likely to serve as primers.

[0556] The spacer may alterable within given or predetermined parameters. For example one method may give an optimal sequence of TTTCTTATAT, but using a slightly different optimization strategy results in a sequence that is likely just as or nearly as good.

[0557] The selected template switching region can comprise 3 consecutive riboGs or more. The selected template switching region can comprise 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 consecutive riboGs or more. Alternative nucleotide may be used such as deoxyribo Gs, LNA G's, and potentially any combination thereof.

[0558] The present disclosure also provides methods of enriching cDNA sequences. Enrichment may be useful for TCR, BCR, and immunoglobulin gene analysis since these genes may possess similar yet polymorphic variable region sequences. These sequences can be responsible for antigen binding and peptide-MHC interactions. For example, due to gene recombination events in individual developing T cells, a single human or mouse will naturally express many thousands of different TCR genes. This T cell repertoire can exceed 100,000 or more different TCR rearrangements occurring during T cell development, yielding a total T cell population that is highly polymorphic with respect to its TCR gene sequences especially for the variable region. For immunoglobulin genes, the same may apply, except even greater diversity may be present. As previously noted, each distinct sequence may correspond to a clonotype. In certain embodiments, enrichment increases accuracy and sensitivity of methods for sequencing TCR, BCR and immunoglobulin genes at a single cell level. In certain embodiments, enrichment increases the number of sequencing reads that map to a TCR, BCR, or immunoglobulin gene. In some embodiments, enrichment leads to greater than or equal to 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% or more of total sequencing reads mapping to a TCR, BCR or immunoglobulin gene. In some embodiments, enrichment leads to greater than or equal to 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%,

75%, 80%, 85%, 90%, 95% or more of total sequencing reads mapping to a variable region of a TCR, BCR or immunoglobulin gene.

[0559] In order to aide in sequencing, detection, and analysis of sequences of interest, an enrichment step can be employed. Enrichment may be useful for the sequencing and analysis of genes that may be related yet highly polymorphic. In some embodiments, an enriched gene comprises a TCR sequence, a BCR sequence, or an immunoglobulin sequence. In some embodiments, an enriched gene comprises a mitochondrial gene or a cytochrome family gene. In some embodiments, enrichment is employed after an initial round of reverse transcription (e.g., cDNA production). In some embodiments, enrichment is employed after an initial round of reverse transcription and cDNA amplification for at least 5, 10, 15, 20, 25, 30, 40 or more cycles. In some embodiments, enrichment is employed after a cDNA amplification. In some embodiments, the amplified cDNA can be subjected to a clean-up step before the enrichment step using a column, gel extraction, or beads in order to remove unincorporated primers, unincorporated nucleotides, very short or very long nucleic acid fragments and enzymes. In some embodiments, enrichment is followed by a clean-up step before sequencing library preparation.

[0560] Enrichment of gene or cDNA sequences can be facilitated by a primer that anneals within a known sequence of the target gene. In some embodiments, for enrichment of a TCR, BCR, or immunoglobulin gene, a primer that anneals to a constant region of the gene or cDNA can be paired with a sequencing primer that anneals to a TSO functional sequence. In some embodiments, the enriched cDNA falls into a length range that approximately corresponds to that genes variable region. In some embodiments, greater than about 50%, 60%, 70%, 80%, 85%, 90%, 95% or more cDNA or cDNA fragments fall within a range of about 300 base pairs to about 900 base pairs, of about 400 base pairs to about 800 base pairs, of about 500 base pairs to about 700 base pairs, or of about 500 base pairs to about 600 base pairs.

[0561] FIG. **83** shows an example enrichment scheme. In operation **2001**, an oligonucleotide with a poly-T sequence **2014**, and in some cases an additional sequence **2016** that binds to, for example, a sequencing or PCR primer, anneals to a target RNA **2020**. In operation **2002** the oligonucleotide is extended yielding an anti-sense strand **2022** which is appended by multiple cytidines on the 3? end. A barcode oligonucleotide attached to a bead **2038** (such as a gel bead) is provided and a riboG of the barcode oligonucleotide **2008** pairs with the cytidines of the sense strand and is extended to create a sense and an antisense strand. In some cases, the barcode oligonucleotide is released from the gel bead during extension. In some cases, the barcode oligonucleotide is released from the gel bead prior to extension. In some cases, the barcode oligonucleotide is released from the gel bead after extension. In addition to the riboG sequence, the barcode oligonucleotide comprises a barcode **2012** sequence (which, in some instances may also comprise a unique molecular index) and one or more additional functional sequences **2010**. The additional functional sequences can comprise a primer/primer binding sequence (such as a sequencing primer sequence, e.g., R1 or R2, or partial sequences thereof), a sequence for attachment to an Illumina sequencing flow cell (such as a P5 or P7 sequence), etc. Operations **2001** and **2002** may be performed in a partition (e.g., droplet or well). Subsequent to operation **2002**, the nucleic acid product from operations **2001** and **2002** may be removed from the partition and in some cases pooled with other products from other partitions for subsequent processing. In some cases, the barcode oligonucleotide may be a template switching oligonucleotide.

[0562] Next, additional functional sequences can be added that allow for amplification or sample identification. This may occur in a partition or in bulk. This reaction yields amplified cDNA molecules as in **2003** comprising a barcode and, e.g., sequencing primers. In some cases, not all of these cDNA molecules will comprise a target variable region sequence (e.g., from a TCR or immunoglobulin). In one enrichment scheme, shown in operation **2004**, a primer **2018** that anneals to a sequence 3? of a TCR, BCR or immunoglobulin variable region **2020** specifically amplifies the variable region comprising cDNAs yielding products as shown in operation **2005**. Such enrichment may be performed for various approaches described herein.

[0563] In certain aspects, primer **2018** anneals in a constant region of a TCR (e.g., TCR-alpha or TCR-beta), BCR or immunoglobulin gene. After amplification the products are sheared, adaptors ligated and amplified a second time to add additional functional sequences **2007** and **2011** and a sample index **2009** as shown in operation **2006**. The additional functional sequences can be, for example a primer/primer binding sequence (such as a sequencing primer sequence, e.g., R1 or R2, or partial sequences thereof), a sequence for attachment to an Illumina sequencing flow cell (such as a P5 or P7 sequence), etc. In some embodiments, the initial poly-T primer, comprising sequences **2016** and **2014** can be attached to a gel bead as opposed to the barcode oligonucleotide or template switching oligonucleotide (TSO). In some embodiments, the poly-T comprising primer comprises functional sequences and barcode sequences **2008**, **2010**, **2012**, and the barcode oligonucleotide (e.g., TSO, which, in some instances, is free in solution) comprises sequence **2016**. Operations **2003-2006** may be performed in bulk.

[0564] In some embodiments, clonotype information derived from next-generation sequencing data of cDNA prepped from cellular RNA is combined with other targeted on non-targeted cDNA enrichment to illuminate functional and ontological aspects of B-cell and T cells that express a given TCR, BCR, or immunoglobulin. In some embodiments, clonotype information is combined with analysis of expression of an immunologically relevant cDNA. In some embodiments, the cDNA encodes a cell lineage marker, a cell surface functional marker, immunoglobulin isotype, a cytokine and/or chemokine, an intracellular signaling polypeptide, a cell metabolism polypeptide, a cell-cycle polypeptide, an apoptosis polypeptide, a transcriptional activator/inhibitor, an miRNA or lncRNA.

[0565] Also disclosed herein are methods and systems for reference-free clonotype identification. Such methods may be implemented by way of software executing algorithms. Tools for assembling T-cell Receptor (TCR) sequences may use known sequences of V and C regions to "anchor" assemblies. This may make such tools only applicable to organisms with well characterized references (human and mouse). However, most mammalian T cell receptors have similar amino acid motifs and similar structure. In the absence of a reference, a method can scan assembled transcripts for regions that are diverse or semi-diverse, find the junction region which should be highly diverse, then scan

for known amino acid motifs. In some cases, it may not be critical that the complementary CDRs, such as the CDR1, CDR2, or CDR3, region be accurately delimited, only that a diverse sequence is found that can uniquely identify the clonotype. One advantage of this method is that the software may not require a set of reference sequences and can operate fully de novo, thus this method can enable immune research in eukaryotes with poorly characterized genomes/transcriptomes.

[0566] The methods described herein allow simultaneously obtaining single-cell gene expression information with single-cell immune receptor sequences (TCRs/BCRs). This can be achieved using the methods described herein, such as by amplifying genes relevant to lymphocyte function and state (either in a targeted or unbiased way) while simultaneously amplifying the TCR/BCR sequences for clonotyping. This can allow such applications as 1) interrogating changes in lymphocyte activation/response to an antigen, at the single clonotype or single cell level; or 2) classifying lymphocytes into subtypes based on gene expression while simultaneously sequencing their TCR/BCRs. UMIs are typically ignored during TCR (or generally transcriptome) assembly.

[0567] Key analytical operations involved in clonotype sequencing according to the methods described herein include: 1) Assemble each UMI separately, then merge highly similar assembled sequences. High depth per molecule in TCR sequencing makes this feasible. This may result in a reduced chance of "chimeric" assemblies; 2) Assemble all UMIs from each cell together but use UMI information to choose paths in the assembly graph. This is analogous to using barcode and read-pair information to resolve "bubbles" in WGS assemblies; 3) Base quality estimation. UMI information and alignment of short reads may be used to assemble contigs to compute per-base quality scores. Base quality scoring may be important as a few base differences in a CDR sequence may differentiate one clonotype from another. This may be in contrast to other methods that rely on using long-read sequencing.

[0568] Thus, base quality estimates for assembled contigs can inform clonotype inference. Errors can make cells with the same (real) clonotype have mismatching assembled sequences. Further, combining base-quality estimates and clonotype abundances to correct clonotype assignments. For example, if 10 cells have clonotype X and one cell has a clonotype that differs by X in only a few bases and these bases have low quality, then this cell may be assigned to clonotype X. In some embodiments, clonotypes that differ by a single amino acid or nucleic acid may be discriminated. In some embodiments, clonotypes that differ by less than 50, 40, 30, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, or 2 amino acids or nucleic acids may be discriminated.

[0569] The present disclosure provides methods combining cell multiplexing methods and immune cell analysis methods. In an example, the present disclosure provides a method for analyzing a cell, which cell may be an immune cell such as a T cell or B cell. The cell may comprise a plurality of nucleic acid molecules (e.g., RNA molecules and/or DNA molecules). The plurality of nucleic acid molecules may comprise a plurality of nucleic acid sequences corresponding to a V(D)J region of the genome of the cell. The V(D)J region of the genome of the cell may comprise a T cell receptor variable region sequence, a B cell receptor variable region sequence, or an immunoglobulin variable region sequence. The cell may be labeled with a cell nucleic acid barcode sequence to generate a labeled cell. The cell nucleic acid barcode sequence may be a component of a cell nucleic acid barcode molecule. The cell nucleic acid barcode molecule may also comprise a cell labeling agent that may couple to the cell, such as to a cell surface feature. The cell labeling agent may be, for example, a lipophilic moiety (e.g., a cholesterol), a fluorophore, a dye, a peptide, a nanoparticle, an antibody, or another moiety. The cell nucleic acid barcode sequence may identify a sample from which the cell originates. The sample may be derived from a biological fluid, such as a biological fluid comprising blood or saliva. The cell nucleic acid barcode molecule may be at least partially disposed within the labeled cell.

[0570] The labeled cell may be partitioned in a partition (e.g., a droplet or well) with a plurality of partition nucleic acid barcode molecules. Each partition nucleic acid barcode molecule of the plurality of partition nucleic acid barcode molecules may comprise a partition nucleic acid barcode sequence. Each partition nucleic acid barcode molecule of the plurality of partition nucleic acid barcode molecules may comprise a priming sequence, such as a targeted priming sequence or a random N-mer sequence. Each partition nucleic acid barcode molecule of the plurality of partition nucleic acid barcode molecules may comprise a TSO sequence as described elsewhere herein. The priming sequence may be capable of hybridizing to a sequence of at least a subset of the plurality of nucleic acid molecules. The priming sequence may be capable of hybridizing to a sequence of the cell nucleic acid barcode molecule. The TSO sequence may be capable of facilitating a template switching reaction and/or serve as a priming/hybridization sequence for a cell nucleic acid molecule present in a labeled cell (e.g., a lipophilic or other moiety as described elsewhere herein). The partition nucleic acid barcode molecules may be coupled to a bead, such as a gel bead. The gel bead may be dissolvable or degradable. The partition nucleic acid barcode molecules may be releasably coupled to the bead. Some or all of the partition nucleic acid barcode molecules may be released from the bead within the partition (e.g., upon application of a stimulus, such as a chemical stimulus). Within the partition, the cell may be lysed or permeabilized to provide access to the plurality of nucleic acid molecules therein. The partition may also include a primer molecule, which primer molecule may comprise a sequence complementary to a sequence of the plurality of nucleic acid molecules. Where the plurality of nucleic acid molecules is a plurality of messenger RNA (mRNA) molecules, such a sequence may be a poly(A) sequence.

[0571] A barcoded nucleic acid molecule comprising the cell nucleic acid barcode sequence, or a complement thereof, and the partition nucleic acid barcode sequence, or a complement thereof may be generated within the partition. A plurality of barcoded nucleic acid products each comprising a sequence of a nucleic acid molecule of the plurality of nucleic acid molecules and the partition nucleic acid barcode sequence, or a complement thereof may also be generated within the partition. The plurality of barcoded nucleic acid products may comprise a plurality of complementary DNA (cDNA) molecules, or derivatives thereof. Generating the plurality of barcoded nucleic acid products may comprise hybridizing a sequence of a primer molecule within the partition to a sequence (e.g., a poly(A) sequence) of a nucleic acid molecule of the plurality of nucleic acid mol-

ecules (e.g., mRNA molecules) and using an enzyme (e.g., a reverse transcriptase) to extend the sequence of the primer molecule to provide a nucleic acid product comprising a cDNA sequence corresponding to a sequence of the nucleic acid molecule. The enzyme may have terminal transferase activity and may incorporate a sequence at an end of the nucleic acid product. Such a sequence may be, for example, a poly(C) sequence. Some or all of the plurality of partition nucleic acid barcode molecules may comprise a sequence complementary to the poly(C) sequence (e.g., a poly(riboG) sequence). Generating the plurality of barcoded nucleic acid products may comprise using the nucleic acid product and a partition nucleic acid barcode molecule to generate a barcoded nucleic acid product. The barcoded nucleic acid molecule and the plurality of barcoded nucleic acid products may be synthesized via one or more primer extension reactions, ligation reactions, or nucleic acid amplification reactions. The barcoded nucleic acid molecule and the barcoded nucleic acid products, or derivatives thereof (e.g., the barcoded nucleic acid molecule and the barcoded nucleic acid products having functional sequences appended thereto, such as flow cell sequences and sequencing primers) to yield a plurality of sequencing reads. Each sequencing read of the plurality of sequencing reads may be associated with the partition via its partition nucleic acid barcode sequence. The plurality of nucleic acid molecules may subsequently be identified as originating from the cell.

[0572] Such a method may be extended to a plurality of labeled cells. Each cell of the plurality of labeled calls may be labeled with a cell nucleic acid barcode sequence of a plurality of cell nucleic acid barcode sequences. A plurality of cell nucleic acid barcode molecules may comprise the plurality of cell nucleic barcode sequences, wherein each cell nucleic acid barcode molecule of the plurality of cell nucleic acid barcode molecules may comprise (i) a single cell nucleic acid barcode sequence of the plurality of cell nucleic acid barcode sequences and (ii) a cell labeling agent. The cell labeling agent may be, for example, a lipophilic moiety, a nanoparticle, a fluorophore, a dye, a peptide, an antibody, or another moiety. A lipophilic moiety of each nucleic acid barcode molecule of the plurality of nucleic acid barcode molecules may comprise cholesterol. The cell labeling agent may be linked to the plurality of cell nucleic acid barcode molecules via a linker. The cell labeling agent may be linked to a cell via a cell surface feature, such as a protein. Each labeled cell of the plurality of labeled cells may comprise a target nucleic acid molecule of a plurality of target nucleic acid molecules. The plurality of target nucleic acid molecules may comprise a plurality of messenger RNA (mRNA) molecules. The plurality of target nucleic acid molecules may comprise a plurality of nucleic acid sequences corresponding to V(D)J regions of genomes of the plurality of labeled cells. The V(D)J regions of the genomes of the plurality of labeled cells may comprise T cell receptor variable region sequences, B cell receptor variable region sequences, immunoglobulin variable region sequences, or a combination thereof. The plurality of labeled cells may be a plurality of immune cells, such as a plurality of T cells or B cells. The plurality of labeled cells may derive from a plurality of cellular samples. A given cell nucleic acid barcode sequence of the plurality of cell nucleic acid barcode sequences may identify a cellular sample from which an associated cell of the plurality of labeled cells originates, such as a sample derived from a biological fluid (e.g., a biological fluid comprising saliva or blood). The plurality of cells may be labeled according to the methods provided herein. For example, cells may be labeled using cell binding moieties (e.g., antibodies, cell surface receptor binding molecules, receptor ligands, small molecules, pro-bodies, aptamers, monobodies, affimers, darpins, or protein scaffolds) that may bind to a protein, cell surface species, or other feature of the cells. Cells may alternatively be labeled by delivering nucleic acid barcode molecules to cells using cell-penetrating peptides, liposomes, nanoparticles, electroporation, or mechanical force (e.g., nanowires or microinjection). The cell nucleic acid barcode molecules utilized to label cells may comprise a barcode sequence and one or more functional sequences including a unique molecular index, a primer/primer binding sequence (such as a sequencing primer sequence, e.g., R1, R2, or partial sequences thereof), a sequence configured to attach to the flow cell of a sequencer (such as P5 or P7), an adapter sequence (such as a sequence configured to be complementary or hybridize to a sequence on a partition barcode molecule, e.g., attached to a bead), etc.

[0573] The plurality of labeled cells and a plurality of nucleic acid barcode molecules may be co-partitioned within a plurality of partitions (e.g., droplets or wells). Each partition of the plurality of partitions may comprise at least one labeled cell of the plurality of labeled cells and a partition nucleic acid barcode molecule of a plurality of partition nucleic acid barcode molecules. At least a subset of the plurality of partitions may comprise at least two labeled cells of the plurality of labeled cells. Each partition nucleic acid barcode molecule of the plurality of partition nucleic acid barcode molecules may comprise a partition nucleic acid barcode sequence of a plurality of partition nucleic acid barcode sequences, and each partition of the plurality of partitions may comprise a different partition nucleic acid barcode sequence. The plurality of partition nucleic acid barcode molecules may be coupled to a plurality of beads, such as a plurality of gel beads. Each bead of the plurality of beads may comprise at least 10,000 partition nucleic acid barcode molecules of the plurality of partition nucleic acid barcode molecules coupled thereto. The plurality of gel beads may be dissolvable or degradable. Each partition of the plurality of partitions may comprise a single bead of the plurality of beads. The plurality of partition nucleic acid barcode molecules may be releasably coupled to the plurality of beads. The plurality of partition nucleic acid barcode molecules may be releasable from the beads upon application of a stimulus, such as a chemical stimulus. Partition nucleic acid barcode molecules of the plurality of partition nucleic acid barcode molecules may be released from each bead of the plurality of beads within the plurality of partitions. Each partition nucleic acid barcode molecule of the plurality of partition nucleic acid barcode molecules may comprise a common partition nucleic acid barcode sequence. Each partition nucleic acid barcode molecule of the plurality of partition nucleic acid barcode molecules may comprise a common partition nucleic acid barcode sequence and one or more functional sequences including a unique molecular index, a primer/primer binding sequence (such as a sequencing primer sequence, e.g., R1, R2, or partial sequences thereof), a sequence configured to attach to the flow cell of a sequencer (such as P5 or P7), an adapter sequence (such as a sequence configured to be complementary or hybridize to a sequence on a cell barcode molecule,

e.g., coupled to a labeled cell, such as via a lipophilic moiety), etc. A given bead may comprise multiple different types of partition nucleic acid barcode molecules. For example, the given bead may comprise a first set of partition nucleic acid barcode molecules and a second set of partition nucleic acid barcode molecules. The first set of partition nucleic acid barcode molecules may comprise a sequence complementary to a sequence of the cell nucleic acid barcode sequence of a given partition comprising the given bead, while the second set of partition nucleic acid barcode molecules may comprise a sequence useful in processing target nucleic acid molecules of a labeled cell of the given partition.

[0574] Within the partitions, the plurality of labeled cells may be subjected to conditions sufficient to provide access to the plurality of target nucleic acid molecules therein. For example, the plurality of labeled cells may be lysed or permeabilized. The plurality of partition nucleic acid barcode molecules may be used to synthesize (i) a first plurality of barcoded nucleic acid products comprising a cell nucleic acid barcode sequence of the plurality of cell nucleic acid barcode sequences, or a complement thereof, and a partition nucleic acid barcode sequence of the plurality of partition nucleic acid barcode sequences, or a complement thereof; and (ii) a second plurality of barcoded nucleic acid products comprising a sequence of a target nucleic acid molecule (e.g., a V(D)J sequence as described herein) of the plurality of target nucleic acid molecules, or a complement thereof, and the partition nucleic acid barcode sequence of the plurality of partition nucleic acid barcode sequences, or a complement thereof. This process may comprise reverse transcribing mRNA molecules to generate cDNA molecules (e.g., as described herein). A reverse transcriptase, such as a reverse transcriptase having terminal transferase activity, may be used to reverse transcribe mRNA. Template switching may be performed (e.g., using partition nucleic acid barcode molecules comprising terminal poly(riboG) sequences) to generate the second plurality of barcoded nucleic acid products (e.g., as described herein). In some cases, multiplet reduction techniques such as those described herein may also be employed. For example, at least two labeled cells of the plurality of labeled cells may be identified as originating from a same partition of the plurality of partitions using (i) cell nucleic acid barcode sequences of the plurality of cell nucleic acid barcode sequences, or complements thereof, and (ii) partition nucleic acid barcode sequences of the plurality of partition nucleic acid barcode sequences, or complements thereof. Relative cell sizes of the plurality of labeled cells may also be determined (e.g., as described herein).

[0575] In some instances, different cell barcode sequences may be attached to different samples of cells, which are then pooled for partition barcoding. For example, in some embodiments, (1) a first population of cells is labeled with a first cell barcode sequence using, e.g., a lipophilic moiety as described herein and (2) a second population of cells is labeled with a second cell barcode sequence using, e.g., a lipophilic moiety as described herein. The labeled first and second population of cells may then be pooled and co-partitioned with partition barcode molecules (e.g., attached to a bead, such as a gel bead) for barcoding as described elsewhere herein. Any suitable number of samples (e.g., population of cells) may be labeled with cell barcodes as described herein and pooled (e.g., multiplexed) for analysis thereby increasing the throughput and reducing the cost of sample analysis.

Enhanced Cell Multiplexing

[0576] The methods provided herein may make use of multiple cellular barcodes or tags (e.g., multiple different cell nucleic acid barcode sequences for a given cell). The use of multiple tags may facilitate higher level multiplexing with a reduced number of reagents. Accordingly, the present disclosure provides a method comprising the use of multiple (e.g., two or more) different tags to label a single population of cells. Cell identification in such a scheme is based on a combination of tags, rather than a single tag. Such a method may be referred to as "combinatorial tagging."

[0577] In some cases, the combinatorial tagging methods provided herein may be used to specifically label different populations and conditions. For example, a first set of tags may be used for sample identification, while a second set of tags may be used to associate cells with a given condition. Multiple additional layers of tagging may be incorporated. For example, a first set of tags may be used to indicate a subject from which a cellular sample derives, a second set of tags may be used to indicate a bodily area of the subject from which a cellular sample or portion thereof derives, a third set of tags may be used to indicate a first processing or storage condition, a fourth set of tags may be used to indicate a second processing or storage condition, etc. Tagging of cells may be performed simultaneously or sequentially. For example, a first tag may be provided to a cell prior to provision of a second tag. Alternatively, the first and second tags may be provided at the same time (e.g., in a mixture of tags). In some cases, a matrix-based method may be used for staining. For example, FIG. **85** shows tagging of cells assigned to specific spatial positions (e.g., wells within a well plate). For a microwell plate having 96 microwells, a total of 20 barcodes (8 for 8 rows and 12 for 12 columns) may be used to provide 96 unique cell identifier combinations. Accordingly, many more cell identifiers may be generated with fewer total reagents.

[0578] In addition to providing for greater levels of multiplexing, the use of multiple tags may also provide greater confidence in sample identification, which may be particularly relevant for clinical samples. For example, if each tag is assumed to be about 95% sensitive (e.g., binds to 95% of the intended cells) and 1% non-specific (e.g., binds to 1% of the wrong cells, possibly after pooling and prior to partitioning of cells), using just 2 tags per sample would result in much better specificity (0.01%) without significant loss of sensitivity (net sensitivity 90.2%). Using 2 tags per sample, $N(N-1)/2$ pairs can be achieved from N tags. Using 3 tags per sample, this increases to $O(N^3)$. Additional schemes may also be used.

[0579] In some cases, first tags and second tags may be provided to a population of cells simultaneously (e.g., within a mixture). In other cases, a cell may be labeled with a first tag (e.g., as described herein) prior to provision of the second tag. Subsequent to labeling with the first tag, the cell may be labeled with the second tag (e.g., as described herein). In some cases, the second tag may couple to the first tag (e.g., via hybridization of complementary sequences of the first and second tags, ligation, chemical binding (e.g.,

formation of a covalent bond), or another process). In other cases, the second tag may not be directly coupled to the first tag.

[0580] First and second tags may label cells according to the same or different mechanisms. The present disclosure provides numerous examples of labeling of cells with tags (e.g., cell nucleic acid barcode molecules comprising cell nucleic acid barcode sequences). In an example, first and second tags may each include lipophilic moieties capable of coupling to cells (e.g., as described herein).

[0581] First and second tags may have the same or different characteristics. For example, first tags may comprise barcode sequences having a first length (e.g., between 6-20 nucleotides) while second tags may comprise barcode sequences having a second length (e.g., between 6-20 nucleotides) that is different than the first length. In another example, first tags may comprise nucleic acid barcode sequences (e.g., as described herein) while second tags may comprise optical labels. Optical labels may be distinguished by, for example, the intensity and wavelength of fluorescence emission upon excitation. Optical labels may comprise fluorescent labels such as fluorescent dyes.

[0582] In an example, the present disclosure provides a method of analyzing a plurality of cells, comprising providing a first plurality of cell nucleic acid barcode molecules comprising a first plurality of cell nucleic acid barcode sequences and a second plurality of cell nucleic acid barcode molecules comprising a second plurality of cell nucleic acid barcode sequences. Each cell nucleic acid barcode molecule of the first plurality of cell nucleic acid barcode molecules and the second plurality of cell nucleic acid barcode molecules may comprise a single cell nucleic acid barcode sequence of the first plurality of cell nucleic acid barcode sequences or the second plurality of cell nucleic acid barcode sequences. In some cases, each cell nucleic acid barcode molecule of the first plurality of cell nucleic acid barcode molecules or the second plurality of cell nucleic acid barcode molecules comprises a lipophilic moiety. The lipophilic moiety may comprise cholesterol. The lipophilic moiety may be linked to the first plurality of cell nucleic acid barcode molecules or the second plurality of cell nucleic acid barcode molecules via a linker.

[0583] The plurality of cells may be labeled with the first plurality of cell nucleic acid barcode sequences and the second plurality of cell nucleic acid barcode sequences (e.g., as described herein) to generate a plurality of labeled cells. Each labeled cell of the plurality of labeled cells may comprise (i) a different cell nucleic acid barcode sequence of the first plurality of cell nucleic acid barcode sequences and (ii) a different cell nucleic acid barcode sequence of the second plurality of cell nucleic acid barcode sequences. In some cases, the plurality of cells may be labeled with the first plurality of cell nucleic acid barcode sequences and the second plurality of cell nucleic acid barcode sequences simultaneously. In other cases, the plurality of cells are labeled with the first plurality of cell nucleic acid barcode sequences prior to the second plurality of cell nucleic acid barcode sequences. A cell nucleic acid barcode molecule of the second plurality of cell nucleic acid barcode sequences may be coupled to a cell nucleic acid barcode molecule of the first plurality of cell nucleic acid barcode sequences coupled to a given cell of the plurality of cells. In some cases, the second plurality of cell nucleic acid barcode sequences may comprise a sequence complementary to a sequence of the first plurality of cell nucleic acid barcode sequences. The plurality of cells may be labeled with the first plurality of cell nucleic acid barcode sequences and/or the second plurality of cell nucleic acid barcode sequences by binding cell binding moieties, each coupled to a given cell nucleic acid barcode sequence of the first plurality of cell nucleic acid barcode sequences and/or the second plurality of cell nucleic acid barcode sequences, to each cell of the plurality of cells. The cell binding moieties may be, for example, antibodies, cell surface receptor binding molecules, receptor ligands, small molecules, pro-bodies, aptamers, monobodies, affimers, darpins, or protein scaffolds. The cell binding moieties may bind to a protein or a cell surface species of cells of the plurality of cells. In some cases, the cell binding moieties may bind to a species common to each cell of the plurality of cells. In some cases, the plurality of cells may be labeled with the first plurality of cell nucleic acid barcode sequences and/or the second plurality of cell nucleic acid barcode sequences by delivering nucleic acid barcode molecules each comprising an individual cell nucleic acid barcode sequence of the first plurality of cell nucleic acid barcode sequences and/or the second plurality of cell nucleic acid barcode sequences to each cell of the plurality of cells with the aid of a cell-penetrating peptide. Alternatively, the plurality of cells may be labeled with the first plurality of cell nucleic acid barcode sequences and/or the second plurality of cell nucleic acid barcode sequences with the aid of liposomes, nanoparticles, electroporation, or mechanical force (e.g., using nanowires or microinjection).

[0584] A plurality of partitions (e.g., droplets or wells) comprising the plurality of labeled cells and a plurality of partition nucleic acid barcode sequences may be generated (e.g., as described herein). Each partition of the plurality of partitions may comprise a different partition nucleic barcode sequence of the plurality of partition nucleic acid barcode sequences. The plurality of partition nucleic acid barcode sequences may be components a plurality of partition nucleic acid barcode molecules, which plurality of partition nucleic acid barcode molecules may be coupled to a plurality of beads (e.g., gel beads that may be dissolvable or degradable). Each partition of the plurality of partitions may comprise a single bead of the plurality of beads. The plurality of partition nucleic acid barcode molecules may be releasably coupled to the plurality of beads. The plurality of partition nucleic acid barcode molecules may be releasable from the bead upon application of a stimulus (e.g., a chemical stimulus). In some cases, subsequent to partitioning, partition nucleic acid barcode molecules of the plurality of partition nucleic acid barcode molecules may be released from each bead of the plurality of beads. Each partition nucleic acid barcode molecule of the plurality of partition nucleic acid barcode molecules coupled to a given bead may comprise a common partition nucleic acid barcode sequence. Each partition nucleic acid barcode molecule of the plurality of partition nucleic acid barcode molecules may comprise a unique molecular identifier sequence and/or a priming sequence (e.g., a targeted priming sequence or a random priming sequence). In some cases, the plurality of labeled cells may be lysed or permeabilized after partitioning, e.g., to provide access to nucleic acid molecules therein.

[0585] A plurality of barcoded nucleic acid products may be synthesized from the plurality of labeled cells, wherein a given barcoded nucleic acid product of the plurality of

barcoded nucleic acid products comprises (i) a cell identification sequence comprising a given cell nucleic acid barcode sequence of the first plurality of cell nucleic acid barcode sequences or the second plurality of cell nucleic acid barcode sequences, or a complement of the given cell nucleic acid barcode sequence; and (ii) a partition identification sequence comprising a given partition nucleic acid barcode sequence of the plurality of partition nucleic acid barcode sequences, or a complement of the given partition nucleic acid barcode sequence.

[0586] The plurality of labeled cells may be derived from a plurality of cellular samples. A given cell nucleic acid barcode sequence of the first plurality of cell nucleic acid barcode sequences or the second plurality of cell nucleic acid barcode sequences may identify a cellular sample from which an associated cell of the plurality of labeled cells originates. The sample may be derived from a biological fluid (e.g., blood or saliva). In some cases, the first plurality of cell nucleic acid barcode sequences may identify the cellular sample. In some cases, the second plurality of cell nucleic acid barcode sequences may identify a condition to which an associated cell of the plurality of labeled cells is subjected. In some cases, the first plurality of cell nucleic acid barcode sequences and the second plurality of cell nucleic acid barcode sequences may identify a spatial position of an associated cell of the plurality of labeled cells prior to cell partitioning.

[0587] In some cases, at least a subset of the plurality of partitions may comprise at least two labeled cells of the plurality of labeled cells. The method may further comprise identifying at least two labeled cells of the plurality of labeled cells as originating from a same partition of the plurality of partitions using (i) cell nucleic acid barcode sequences of the first plurality of cell nucleic acid barcode sequences, or complements thereof, (ii) cell nucleic acid barcode sequences of the second plurality of cell nucleic acid barcode sequences, or complements thereof, and/or (iii) partition nucleic acid barcode sequences of the plurality of partition nucleic acid barcode sequences, or complements thereof. The method may further comprise identifying the first plurality of barcoded nucleic acid products and the second plurality of barcoded nucleic acid products as originating from labeled cells of the plurality of labeled cells.

Reagents

[0588] In accordance with certain aspects, biological particles may be partitioned along with lysis reagents in order to release the contents of the biological particles within the partition. In such cases, the lysis agents can be contacted with the biological particle suspension concurrently with, or immediately prior to, the introduction of the biological particles into the partitioning junction/droplet generation zone (e.g., junction **210**), such as through an additional channel or channels upstream of the channel junction. In accordance with other aspects, additionally or alternatively, biological particles may be partitioned along with other reagents, as will be described further below.

[0589] FIG. **62** shows an example of a microfluidic channel structure **6200** for co-partitioning biological particles and reagents. The channel structure **6200** can include channel segments **6201**, **6202**, **6204**, **6206** and **6208**. Channel segments **6201** and **6202** communicate at a first channel junction **6209**. Channel segments **6202**, **6204**, **6206**, and **6208** communicate at a second channel junction **6210**.

[0590] In an example operation, the channel segment **6201** may transport an aqueous fluid **6212** that includes a plurality of biological particles **6214** along the channel segment **6201** into the second junction **6210**. As an alternative or in addition to, channel segment **6201** may transport beads (e.g., gel beads). The beads may comprise barcode molecules.

[0591] For example, the channel segment **6201** may be connected to a reservoir comprising an aqueous suspension of biological particles **6214**. Upstream of, and immediately prior to reaching, the second junction **6210**, the channel segment **6201** may meet the channel segment **6202** at the first junction **6209**. The channel segment **6202** may transport a plurality of reagents **6215** (e.g., lysis agents) suspended in the aqueous fluid **6212** along the channel segment **6202** into the first junction **6209**. For example, the channel segment **6202** may be connected to a reservoir comprising the reagents **6215**. After the first junction **6209**, the aqueous fluid **6212** in the channel segment **6201** can carry both the biological particles **6214** and the reagents **6215** towards the second junction **6210**. In some instances, the aqueous fluid **6212** in the channel segment **6201** can include one or more reagents, which can be the same or different reagents as the reagents **6215**. A second fluid **6216** that is immiscible with the aqueous fluid **6212** (e.g., oil) can be delivered to the second junction **6210** from each of channel segments **6204** and **6206**. Upon meeting of the aqueous fluid **6212** from the channel segment **6201** and the second fluid **6216** from each of channel segments **6204** and **6206** at the second channel junction **6210**, the aqueous fluid **6212** can be partitioned as discrete droplets **6218** in the second fluid **6216** and flow away from the second junction **6210** along channel segment **6208**. The channel segment **6208** may deliver the discrete droplets **6218** to an outlet reservoir fluidly coupled to the channel segment **6208**, where they may be harvested.

[0592] The second fluid **6216** can comprise an oil, such as a fluorinated oil, that includes a fluorosurfactant for stabilizing the resulting droplets, for example, inhibiting subsequent coalescence of the resulting droplets **6218**.

[0593] A discrete droplet generated may include an individual biological particle **6214** and/or one or more reagents **6215**. In some instances, a discrete droplet generated may include a barcode carrying bead (not shown), such as via other microfluidics structures described elsewhere herein. In some instances, a discrete droplet may be unoccupied (e.g., no reagents, no biological particles).

[0594] Beneficially, when lysis reagents and biological particles are co-partitioned, the lysis reagents can facilitate the release of the contents of the biological particles within the partition. The contents released in a partition may remain discrete from the contents of other partitions.

[0595] As will be appreciated, the channel segments described herein may be coupled to any of a variety of different fluid sources or receiving components, including reservoirs, tubing, manifolds, or fluidic components of other systems. As will be appreciated, the microfluidic channel structure **6200** may have other geometries. For example, a microfluidic channel structure can have more than two channel junctions. For example, a microfluidic channel structure can have 2, 3, 4, 5 channel segments or more each carrying the same or different types of beads, reagents, and/or biological particles that meet at a channel junction. Fluid flow in each channel segment may be controlled to control the partitioning of the different elements into droplets. Fluid may be directed flow along one or more channels

or reservoirs via one or more fluid flow units. A fluid flow unit can comprise compressors (e.g., providing positive pressure), pumps (e.g., providing negative pressure), actuators, and the like to control flow of the fluid. Fluid may also or otherwise be controlled via applied pressure differentials, centrifugal force, electrokinetic pumping, vacuum, capillary or gravity flow, or the like.

[0596] Examples of lysis agents include bioactive reagents, such as lysis enzymes that are used for lysis of different cell types, e.g., gram positive or negative bacteria, plants, yeast, mammalian, etc., such as lysozymes, achromopeptidase, lysostaphin, labiase, kitalase, lyticase, and a variety of other lysis enzymes available from, e.g., Sigma-Aldrich, Inc. (St Louis, MO), as well as other commercially available lysis enzymes. Other lysis agents may additionally or alternatively be co-partitioned with the biological particles to cause the release of the biological particles's contents into the partitions. For example, in some cases, surfactant-based lysis solutions may be used to lyse cells, although these may be less desirable for emulsion based systems where the surfactants can interfere with stable emulsions. In some cases, lysis solutions may include non-ionic surfactants such as, for example, TritonX-100 and Tween 20. In some cases, lysis solutions may include ionic surfactants such as, for example, sarcosyl and sodium dodecyl sulfate (SDS). Electroporation, thermal, acoustic or mechanical cellular disruption may also be used in certain cases, e.g., non-emulsion based partitioning such as encapsulation of biological particles that may be in addition to or in place of droplet partitioning, where any pore size of the encapsulate is sufficiently small to retain nucleic acid fragments of a given size, following cellular disruption.

[0597] In addition to the lysis agents co-partitioned with the biological particles described above, other reagents can also be co-partitioned with the biological particles, including, for example, DNase and RNase inactivating agents or inhibitors, such as proteinase K, chelating agents, such as EDTA, and other reagents employed in removing or otherwise reducing negative activity or impact of different cell lysate components on subsequent processing of nucleic acids. In addition, in the case of encapsulated biological particles, the biological particles may be exposed to an appropriate stimulus to release the biological particles or their contents from a co-partitioned microcapsule. For example, in some cases, a chemical stimulus may be co-partitioned along with an encapsulated biological particle to allow for the degradation of the microcapsule and release of the cell or its contents into the larger partition. In some cases, this stimulus may be the same as the stimulus described elsewhere herein for release of nucleic acid molecules (e.g., oligonucleotides) from their respective microcapsule (e.g., bead). In alternative aspects, this may be a different and non-overlapping stimulus, in order to allow an encapsulated biological particle to be released into a partition at a different time from the release of nucleic acid molecules into the same partition.

[0598] Additional reagents may also be co-partitioned with the biological particles, such as endonucleases to fragment a biological particle's DNA, DNA polymerase enzymes and dNTPs used to amplify the biological particle's nucleic acid fragments and to attach the barcode molecular tags to the amplified fragments. Other enzymes may be co-partitioned, including without limitation, polymerase, transposase, ligase, proteinase K, DNAse, etc. Additional reagents may also include reverse transcriptase enzymes, including enzymes with terminal transferase activity, primers and oligonucleotides, and switch oligonucleotides (also referred to herein as "switch oligos" or "template switching oligonucleotides") which can be used for template switching. In some cases, template switching can be used to increase the length of a cDNA. In some cases, template switching can be used to append a predefined nucleic acid sequence to the cDNA. In an example of template switching, cDNA can be generated from reverse transcription of a template, e.g., cellular mRNA, where a reverse transcriptase with terminal transferase activity can add additional nucleotides, e.g., polyC, to the cDNA in a template independent manner. Switch oligos can include sequences complementary to the additional nucleotides, e.g., polyG. The additional nucleotides (e.g., polyC) on the cDNA can hybridize to the additional nucleotides (e.g., polyG) on the switch oligo, whereby the switch oligo can be used by the reverse transcriptase as template to further extend the cDNA. Template switching oligonucleotides may comprise a hybridization region and a template region. The hybridization region can comprise any sequence capable of hybridizing to the target. In some cases, as previously described, the hybridization region comprises a series of G bases to complement the overhanging C bases at the 3' end of a cDNA molecule. The series of G bases may comprise 1 G base, 2 G bases, 3 G bases, 4 G bases, 5 G bases or more than 5 G bases. The template sequence can comprise any sequence to be incorporated into the cDNA. In some cases, the template region comprises at least 1 (e.g., at least 2, 3, 4, 5 or more) tag sequences and/or functional sequences. Switch oligos may comprise deoxyribonucleic acids; ribonucleic acids; modified nucleic acids including 2-Aminopurine, 2,6-Diaminopurine (2-Amino-dA), inverted dT, 5-Methyl dC, 2'-deoxyInosine, Super T (5-hydroxybutynl-2'-deoxyuridine), Super G (8-aza-7-deazaguanosine), locked nucleic acids (LNAs), unlocked nucleic acids (UNAs, e.g., UNA-A, UNA-U, UNA-C, UNA-G), Iso-dG, Iso-dC, 2' Fluoro bases (e.g., Fluoro C, Fluoro U, Fluoro A, and Fluoro G), or any combination.

[0599] In some cases, the length of a switch oligo may be at least about 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249 or 250 nucleotides or longer.

[0600] In some cases, the length of a switch oligo may be at most about 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32,

33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249 or 250 nucleotides.

[0601] Once the contents of the cells are released into their respective partitions, the macromolecular components (e.g., macromolecular constituents of biological particles, such as RNA, DNA, or proteins) contained therein may be further processed within the partitions. In accordance with the methods and systems described herein, the macromolecular component contents of individual biological particles can be provided with unique identifiers such that, upon characterization of those macromolecular components they may be attributed as having been derived from the same biological particle or particles. The ability to attribute characteristics to individual biological particles or groups of biological particles is provided by the assignment of unique identifiers specifically to an individual biological particle or groups of biological particles. Unique identifiers, e.g., in the form of nucleic acid barcodes can be assigned or associated with individual biological particles or populations of biological particle, in order to tag or label the biological particle's macromolecular components (and as a result, its characteristics) with the unique identifiers. These unique identifiers can then be used to attribute the biological particle's components and characteristics to an individual biological particle or group of biological particles.

[0602] In some aspects, this is performed by co-partitioning the individual biological particle or groups of biological particles with the unique identifiers, such as described above (with reference to FIG. **2**). In some aspects, the unique identifiers are provided in the form of nucleic acid molecules (e.g., oligonucleotides) that comprise nucleic acid barcode sequences that may be attached to or otherwise associated with the nucleic acid contents of individual biological particle, or to other components of the biological particle, and particularly to fragments of those nucleic acids. The nucleic acid molecules are partitioned such that as between nucleic acid molecules in a given partition, the nucleic acid barcode sequences contained therein are the same, but as between different partitions, the nucleic acid molecule can, and do have differing barcode sequences, or at least represent a large number of different barcode sequences across all of the partitions in a given analysis. In some aspects, only one nucleic acid barcode sequence can be associated with a given partition, although in some cases, two or more different barcode sequences may be present.

[0603] The nucleic acid barcode sequences can include from about 6 to about 20 or more nucleotides within the sequence of the nucleic acid molecules (e.g., oligonucleotides). In some cases, the length of a barcode sequence may be about 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 nucleotides or longer. In some cases, the length of a barcode sequence may be at least about 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 nucleotides or longer. In some cases, the length of a barcode sequence may be at most about 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 nucleotides or shorter. These nucleotides may be completely contiguous, i.e., in a single stretch of adjacent nucleotides, or they may be separated into two or more separate subsequences that are separated by 1 or more nucleotides. In some cases, separated barcode subsequences can be from about 4 to about 16 nucleotides in length. In some cases, the barcode subsequence may be about 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 nucleotides or longer. In some cases, the barcode subsequence may be at least about 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 nucleotides or longer. In some cases, the barcode subsequence may be at most about 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 nucleotides or shorter.

[0604] The co-partitioned nucleic acid molecules can also comprise other functional sequences useful in the processing of the nucleic acids from the co-partitioned biological particles. These sequences include, e.g., targeted or random/universal amplification/extension primer sequences for amplifying or extending the genomic DNA from the individual biological particles within the partitions while attaching the associated barcode sequences, sequencing primers or primer recognition sites, hybridization or probing sequences, e.g., for identification of presence of the sequences or for pulling down barcoded nucleic acids, or any of a number of other potential functional sequences. Other mechanisms of co-partitioning oligonucleotides may also be employed, including, e.g., coalescence of two or more droplets, where one droplet contains oligonucleotides, or microdispensing of oligonucleotides into partitions, e.g., droplets within microfluidic systems.

[0605] In an example, microcapsules, such as beads, are provided that each include large numbers of the above described barcoded nucleic acid molecules (e.g., barcoded oligonucleotides) releasably attached to the beads, where all of the nucleic acid molecules attached to a particular bead will include the same nucleic acid barcode sequence, but where a large number of diverse barcode sequences are represented across the population of beads used. In some embodiments, hydrogel beads, e.g., comprising polyacrylamide polymer matrices, are used as a solid support and delivery vehicle for the nucleic acid molecules into the partitions, as they are capable of carrying large numbers of nucleic acid molecules, and may be configured to release those nucleic acid molecules upon exposure to a particular stimulus, as described elsewhere herein. In some cases, the population of beads provides a diverse barcode sequence library that includes at least about 1,000 different barcode sequences, at least about 5,000 different barcode sequences, at least about 10,000 different barcode sequences, at least about 50,000 different barcode sequences, at least about 100,000 different barcode sequences, at least about 1,000,000 different barcode sequences, at least about 5,000,000 different barcode sequences, or at least about 10,000,000 different barcode sequences, or more. Additionally, each bead can be provided with large numbers of nucleic acid (e.g., oligonucleotide) molecules attached. In particular, the number of molecules of nucleic acid molecules including the barcode sequence on an individual bead can be at least about 1,000 nucleic acid molecules, at least about 5,000 nucleic

acid molecules, at least about 10,000 nucleic acid molecules, at least about 50,000 nucleic acid molecules, at least about 100,000 nucleic acid molecules, at least about 500,000 nucleic acids, at least about 1,000,000 nucleic acid molecules, at least about 5,000,000 nucleic acid molecules, at least about 10,000,000 nucleic acid molecules, at least about 50,000,000 nucleic acid molecules, at least about 100,000,000 nucleic acid molecules, at least about 250,000,000 nucleic acid molecules and in some cases at least about 1 billion nucleic acid molecules, or more. Nucleic acid molecules of a given bead can include identical (or common) barcode sequences, different barcode sequences, or a combination of both. Nucleic acid molecules of a given bead can include multiple sets of nucleic acid molecules. Nucleic acid molecules of a given set can include identical barcode sequences. The identical barcode sequences can be different from barcode sequences of nucleic acid molecules of another set.

[0606] Moreover, when the population of beads is partitioned, the resulting population of partitions can also include a diverse barcode library that includes at least about 1,000 different barcode sequences, at least about 5,000 different barcode sequences, at least about 10,000 different barcode sequences, at least at least about 50,000 different barcode sequences, at least about 100,000 different barcode sequences, at least about 1,000,000 different barcode sequences, at least about 5,000,000 different barcode sequences, or at least about 10,000,000 different barcode sequences. Additionally, each partition of the population can include at least about 1,000 nucleic acid molecules, at least about 5,000 nucleic acid molecules, at least about 10,000 nucleic acid molecules, at least about 50,000 nucleic acid molecules, at least about 100,000 nucleic acid molecules, at least about 500,000 nucleic acids, at least about 1,000,000 nucleic acid molecules, at least about 5,000,000 nucleic acid molecules, at least about 10,000,000 nucleic acid molecules, at least about 50,000,000 nucleic acid molecules, at least about 100,000,000 nucleic acid molecules, at least about 250,000,000 nucleic acid molecules and in some cases at least about 1 billion nucleic acid molecules.

[0607] In some cases, it may be desirable to incorporate multiple different barcodes within a given partition, either attached to a single or multiple beads within the partition. For example, in some cases, a mixed, but known set of barcode sequences may provide greater assurance of identification in the subsequent processing, e.g., by providing a stronger address or attribution of the barcodes to a given partition, as a duplicate or independent confirmation of the output from a given partition.

[0608] The nucleic acid molecules (e.g., oligonucleotides) are releasable from the beads upon the application of a particular stimulus to the beads. In some cases, the stimulus may be a photo-stimulus, e.g., through cleavage of a photolabile linkage that releases the nucleic acid molecules. In other cases, a thermal stimulus may be used, where elevation of the temperature of the beads environment will result in cleavage of a linkage or other release of the nucleic acid molecules form the beads. In still other cases, a chemical stimulus can be used that cleaves a linkage of the nucleic acid molecules to the beads, or otherwise results in release of the nucleic acid molecules from the beads. In one case, such compositions include the polyacrylamide matrices described above for encapsulation of biological particles, and may be degraded for release of the attached nucleic acid molecules through exposure to a reducing agent, such as DTT.

[0609] In some aspects, provided are systems and methods for controlled partitioning. Droplet size may be controlled by adjusting certain geometric features in channel architecture (e.g., microfluidics channel architecture). For example, an expansion angle, width, and/or length of a channel may be adjusted to control droplet size.

[0610] FIG. **63** shows an example of a microfluidic channel structure for the controlled partitioning of beads into discrete droplets. A channel structure **6300** can include a channel segment **6302** communicating at a channel junction **6306** (or intersection) with a reservoir **6304**. The reservoir **6304** can be a chamber. Any reference to "reservoir," as used herein, can also refer to a "chamber." In operation, an aqueous fluid **6308** that includes suspended beads **6312** may be transported along the channel segment **6302** into the junction **6306** to meet a second fluid **6310** that is immiscible with the aqueous fluid **6308** in the reservoir **6304** to create droplets **6316**, **6318** of the aqueous fluid **6308** flowing into the reservoir **6304**. At the juncture **6306** where the aqueous fluid **6308** and the second fluid **6310** meet, droplets can form based on factors such as the hydrodynamic forces at the juncture **6306**, flow rates of the two fluids **6308**, **6310**, fluid properties, and certain geometric parameters (e.g., w, h0, ?, etc.) of the channel structure **6300**. A plurality of droplets can be collected in the reservoir **6304** by continuously injecting the aqueous fluid **6308** from the channel segment **6302** through the juncture **6306**.

[0611] A discrete droplet generated may include a bead (e.g., as in occupied droplets **6316**). Alternatively, a discrete droplet generated may include more than one bead. Alternatively, a discrete droplet generated may not include any beads (e.g., as in unoccupied droplet **6318**). In some instances, a discrete droplet generated may contain one or more biological particles, as described elsewhere herein. In some instances, a discrete droplet generated may comprise one or more reagents, as described elsewhere herein.

[0612] In some instances, the aqueous fluid **6308** can have a substantially uniform concentration or frequency of beads **6312**. The beads **6312** can be introduced into the channel segment **6302** from a separate channel (not shown in FIG. **63**). The frequency of beads **6312** in the channel segment **6302** may be controlled by controlling the frequency in which the beads **6312** are introduced into the channel segment **6302** and/or the relative flow rates of the fluids in the channel segment **6302** and the separate channel. In some instances, the beads can be introduced into the channel segment **6302** from a plurality of different channels, and the frequency controlled accordingly.

[0613] In some instances, the aqueous fluid **6308** in the channel segment **6302** can comprise biological particles (e.g., described with reference to FIGS. **1** and **2**). In some instances, the aqueous fluid **6308** can have a substantially uniform concentration or frequency of biological particles. As with the beads, the biological particles can be introduced into the channel segment **6302** from a separate channel. The frequency or concentration of the biological particles in the aqueous fluid **6308** in the channel segment **6302** may be controlled by controlling the frequency in which the biological particles are introduced into the channel segment **6302** and/or the relative flow rates of the fluids in the channel segment **6302** and the separate channel. In some

instances, the biological particles can be introduced into the channel segment **6302** from a plurality of different channels, and the frequency controlled accordingly. In some instances, a first separate channel can introduce beads and a second separate channel can introduce biological particles into the channel segment **6302**. The first separate channel introducing the beads may be upstream or downstream of the second separate channel introducing the biological particles.

[0614] The second fluid **6310** can comprise an oil, such as a fluorinated oil, that includes a fluorosurfactant for stabilizing the resulting droplets, for example, inhibiting subsequent coalescence of the resulting droplets.

[0615] In some instances, the second fluid **6310** may not be subjected to and/or directed to any flow in or out of the reservoir **6304**. For example, the second fluid **6310** may be substantially stationary in the reservoir **6304**. In some instances, the second fluid **6310** may be subjected to flow within the reservoir **6304**, but not in or out of the reservoir **6304**, such as via application of pressure to the reservoir **6304** and/or as affected by the incoming flow of the aqueous fluid **6308** at the juncture **6306**. Alternatively, the second fluid **6310** may be subjected and/or directed to flow in or out of the reservoir **6304**. For example, the reservoir **6304** can be a channel directing the second fluid **6310** from upstream to downstream, transporting the generated droplets.

[0616] The channel structure **6300** at or near the juncture **6306** may have certain geometric features that at least partly determine the sizes of the droplets formed by the channel structure **6300**. The channel segment **6302** can have a height, h_0 and width, w, at or near the juncture **6306**. By way of example, the channel segment **6302** can comprise a rectangular cross-section that leads to a reservoir **6304** having a wider cross-section (such as in width or diameter). Alternatively, the cross-section of the channel segment **6302** can be other shapes, such as a circular shape, trapezoidal shape, polygonal shape, or any other shapes. The top and bottom walls of the reservoir **6304** at or near the juncture **6306** can be inclined at an expansion angle, a. The expansion angle, a, allows the tongue (portion of the aqueous fluid **6308** leaving channel segment **6302** at junction **6306** and entering the reservoir **6304** before droplet formation) to increase in depth and facilitate decrease in curvature of the intermediately formed droplet. Droplet size may decrease with increasing expansion angle. The resulting droplet radius, R_d, may be predicted by the following equation for the aforementioned geometric parameters of h_0, w, and α:

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

[0617] By way of example, for a channel structure with w=21 μm, h=21 μm, and α=3°, the predicted droplet size is 121 μm. In another example, for a channel structure with w=25 μm, h=25 μm, and α=5°, the predicted droplet size is 123 μm. In another example, for a channel structure with w=28 μm, h=28 μm, and α=7°, the predicted droplet size is 124 μm.

[0618] In some instances, the expansion angle, α, may be between a range of from about 0.5° to about 4°, from about 0.1° to about 10°, or from about 0° to about 90°. For example, the expansion angle can be at least about 0.01°, 0.1°, 0.2°, 0.3°, 0.4°, 0.5°, 0.6°, 0.7°, 0.8°, 0.9°, 1°, 2°, 3°, 4°, 5°, 6°, 7°, 8°, 9°, 10°, 15°, 20°, 25°, 30°, 35°, 40°, 45°, 50°, 55°, 60°, 65°, 70°, 75°, 80°, 85°, or higher. In some instances, the expansion angle can be at most about 89°, 88°, 87°, 86°, 85°, 84°, 83°, 82°, 81°, 80°, 75°, 70°, 65°, 60°, 55°, 50°, 45°, 40°, 35°, 30°, 25°, 20°, 15°, 10°, 9°, 8°, 7°, 6°, 5°, 4°, 3°, 2°, 1°, 0.1°, 0.01°, or less. In some instances, the width, w, can be between a range of from about 100 micrometers (μm) to about 500 μm. In some instances, the width, w, can be between a range of from about 10 μm to about 200 μm. Alternatively, the width can be less than about 10 μm. Alternatively, the width can be greater than about 500 μm. In some instances, the flow rate of the aqueous fluid **6308** entering the junction **6306** can be between about 0.04 microliters (μL)/minute (min) and about 40 μL/min. In some instances, the flow rate of the aqueous fluid **6308** entering the junction **6306** can be between about 0.01 microliters (μL)/minute (min) and about 100 μL/min. Alternatively, the flow rate of the aqueous fluid **6308** entering the junction **6306** can be less than about 0.01 μL/min. Alternatively, the flow rate of the aqueous fluid **6308** entering the junction **6306** can be greater than about 40 μL/min, such as 45 μL/min, 50 μL/min, 55 μL/min, 60 μL/min, 65 μL/min, 70 μL/min, 75 μL/min, 80 μL/min, 85 μL/min, 90 μL/min, 95 μL/min, 100 μL/min, 110 μL/min, 120 μL/min, 130 μL/min, 140 μL/min, 150 μL/min, or greater. At lower flow rates, such as flow rates of about less than or equal to 10 microliters/minute, the droplet radius may not be dependent on the flow rate of the aqueous fluid **6308** entering the junction **6306**.

[0619] In some instances, at least about 50% of the droplets generated can have uniform size. In some instances, at least about 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or greater of the droplets generated can have uniform size. Alternatively, less than about 50% of the droplets generated can have uniform size.

[0620] The throughput of droplet generation can be increased by increasing the points of generation, such as increasing the number of junctions (e.g., junction **6306**) between aqueous fluid **6308** channel segments (e.g., channel segment **6302**) and the reservoir **6304**. Alternatively or in addition, the throughput of droplet generation can be increased by increasing the flow rate of the aqueous fluid **6308** in the channel segment **6302**.

[0621] FIG. **64** shows an example of a microfluidic channel structure for increased droplet generation throughput. A microfluidic channel structure **6400** can comprise a plurality of channel segments **6402** and a reservoir **6404**. Each of the plurality of channel segments **6402** may be in fluid communication with the reservoir **6404**. The channel structure **6400** can comprise a plurality of channel junctions **6406** between the plurality of channel segments **6402** and the reservoir **6404**. Each channel junction can be a point of droplet generation. The channel segment **6302** from the channel structure **6300** in FIG. **63** and any description to the components thereof may correspond to a given channel segment of the plurality of channel segments **6402** in channel structure **6400** and any description to the corresponding components thereof. The reservoir **6304** from the channel structure **6300** and any description to the components thereof may correspond to the reservoir **6404** from the channel structure **6400** and any description to the corresponding components thereof.

[0622] Each channel segment of the plurality of channel segments **6402** may comprise an aqueous fluid **6408** that includes suspended beads **6412**. The reservoir **6404** may

comprise a second fluid **6410** that is immiscible with the aqueous fluid **6408**. In some instances, the second fluid **6410** may not be subjected to and/or directed to any flow in or out of the reservoir **6404**. For example, the second fluid **6410** may be substantially stationary in the reservoir **6404**. In some instances, the second fluid **6410** may be subjected to flow within the reservoir **6404**, but not in or out of the reservoir **6404**, such as via application of pressure to the reservoir **6404** and/or as affected by the incoming flow of the aqueous fluid **6408** at the junctures. Alternatively, the second fluid **6410** may be subjected and/or directed to flow in or out of the reservoir **6404**. For example, the reservoir **6404** can be a channel directing the second fluid **6410** from upstream to downstream, transporting the generated droplets.

[0623] In operation, the aqueous fluid **6408** that includes suspended beads **6412** may be transported along the plurality of channel segments **6402** into the plurality of junctions **6406** to meet the second fluid **6410** in the reservoir **6404** to create droplets **6416**, **6418**. A droplet may form from each channel segment at each corresponding junction with the reservoir **6404**. At the juncture where the aqueous fluid **6408** and the second fluid **6410** meet, droplets can form based on factors such as the hydrodynamic forces at the juncture, flow rates of the two fluids **6408**, **6410**, fluid properties, and certain geometric parameters (e.g., w, h_0, α, etc.) of the channel structure **6400**, as described elsewhere herein. A plurality of droplets can be collected in the reservoir **6404** by continuously injecting the aqueous fluid **6408** from the plurality of channel segments **6402** through the plurality of junctures **6406**. Throughput may significantly increase with the parallel channel configuration of channel structure **6400**. For example, a channel structure having five inlet channel segments comprising the aqueous fluid **6408** may generate droplets five times as frequently than a channel structure having one inlet channel segment, provided that the fluid flow rate in the channel segments are substantially the same. The fluid flow rate in the different inlet channel segments may or may not be substantially the same. A channel structure may have as many parallel channel segments as is practical and allowed for the size of the reservoir. For example, the channel structure may have at least about 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 500, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 1500, 5000 or more parallel or substantially parallel channel segments.

[0624] The geometric parameters, w, h_0, and α, may or may not be uniform for each of the channel segments in the plurality of channel segments **6402**. For example, each channel segment may have the same or different widths at or near its respective channel junction with the reservoir **6404**. For example, each channel segment may have the same or different height at or near its respective channel junction with the reservoir **6404**. In another example, the reservoir **6404** may have the same or different expansion angle at the different channel junctions with the plurality of channel segments **6402**. When the geometric parameters are uniform, beneficially, droplet size may also be controlled to be uniform even with the increased throughput. In some instances, when it is desirable to have a different distribution of droplet sizes, the geometric parameters for the plurality of channel segments **6402** may be varied accordingly.

[0625] In some instances, at least about 50% of the droplets generated can have uniform size. In some instances, at least about 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or greater of the droplets generated can have uniform size. Alternatively, less than about 50% of the droplets generated can have uniform size.

[0626] FIG. **65** shows another example of a microfluidic channel structure for increased droplet generation throughput. A microfluidic channel structure **6500** can comprise a plurality of channel segments **6502** arranged generally circularly around the perimeter of a reservoir **6504**. Each of the plurality of channel segments **6502** may be in fluid communication with the reservoir **6504**. The channel structure **6500** can comprise a plurality of channel junctions **6506** between the plurality of channel segments **6502** and the reservoir **6504**. Each channel junction can be a point of droplet generation. The channel segment **6302** from the channel structure **6300** in FIG. **2** and any description to the components thereof may correspond to a given channel segment of the plurality of channel segments **6502** in channel structure **6500** and any description to the corresponding components thereof. The reservoir **6304** from the channel structure **6300** and any description to the components thereof may correspond to the reservoir **6504** from the channel structure **6500** and any description to the corresponding components thereof.

[0627] Each channel segment of the plurality of channel segments **6502** may comprise an aqueous fluid **6508** that includes suspended beads **6512**. The reservoir **6504** may comprise a second fluid **6510** that is immiscible with the aqueous fluid **6508**. In some instances, the second fluid **6510** may not be subjected to and/or directed to any flow in or out of the reservoir **6504**. For example, the second fluid **6510** may be substantially stationary in the reservoir **6504**. In some instances, the second fluid **6510** may be subjected to flow within the reservoir **6504**, but not in or out of the reservoir **6504**, such as via application of pressure to the reservoir **6504** and/or as affected by the incoming flow of the aqueous fluid **6508** at the junctures. Alternatively, the second fluid **6510** may be subjected and/or directed to flow in or out of the reservoir **6504**. For example, the reservoir **6504** can be a channel directing the second fluid **6510** from upstream to downstream, transporting the generated droplets.

[0628] In operation, the aqueous fluid **6508** that includes suspended beads **6512** may be transported along the plurality of channel segments **6502** into the plurality of junctions **6506** to meet the second fluid **6510** in the reservoir **6504** to create a plurality of droplets **6516**. A droplet may form from each channel segment at each corresponding junction with the reservoir **6504**. At the juncture where the aqueous fluid **6508** and the second fluid **6510** meet, droplets can form based on factors such as the hydrodynamic forces at the juncture, flow rates of the two fluids **6508**, **6510**, fluid properties, and certain geometric parameters (e.g., widths and heights of the channel segments **6502**, expansion angle of the reservoir **6504**, etc.) of the channel structure **6500**, as described elsewhere herein. A plurality of droplets can be collected in the reservoir **6504** by continuously injecting the aqueous fluid **6508** from the plurality of channel segments **6502** through the plurality of junctures **6506**. Throughput may significantly increase with the substantially parallel channel configuration of the channel structure **6500**. A channel structure may have as many substantially parallel channel segments as is practical and allowed for by the size of the reservoir. For example, the channel structure may have at least about 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500,

600, 700, 800, 900, 1000, 1500, 5000 or more parallel or substantially parallel channel segments. The plurality of channel segments may be substantially evenly spaced apart, for example, around an edge or perimeter of the reservoir. Alternatively, the spacing of the plurality of channel segments may be uneven.

[0629] The reservoir **6504** may have an expansion angle, a (not shown in FIG. **65**) at or near each channel juncture. Each channel segment of the plurality of channel segments **6502** may have a width, w, and a height, h_0, at or near the channel juncture. The geometric parameters, w, h_0, and α, may or may not be uniform for each of the channel segments in the plurality of channel segments **6502**. For example, each channel segment may have the same or different widths at or near its respective channel junction with the reservoir **6504**. For example, each channel segment may have the same or different height at or near its respective channel junction with the reservoir **6504**.

[0630] The reservoir **6504** may have the same or different expansion angle at the different channel junctions with the plurality of channel segments **6502**. For example, a circular reservoir (as shown in FIG. **65**) may have a conical, dome-like, or hemispherical ceiling (e.g., top wall) to provide the same or substantially same expansion angle for each channel segments **6502** at or near the plurality of channel junctions **6506**. When the geometric parameters are uniform, beneficially, resulting droplet size may be controlled to be uniform even with the increased throughput. In some instances, when it is desirable to have a different distribution of droplet sizes, the geometric parameters for the plurality of channel segments **6502** may be varied accordingly.

[0631] In some instances, at least about 50% of the droplets generated can have uniform size. In some instances, at least about 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or greater of the droplets generated can have uniform size. Alternatively, less than about 50% of the droplets generated can have uniform size. The beads and/or biological particle injected into the droplets may or may not have uniform size.

[0632] The channel networks, e.g., as described above or elsewhere herein, can be fluidly coupled to appropriate fluidic components. For example, the inlet channel segments are fluidly coupled to appropriate sources of the materials they are to deliver to a channel junction. These sources may include any of a variety of different fluidic components, from simple reservoirs defined in or connected to a body structure of a microfluidic device, to fluid conduits that deliver fluids from off-device sources, manifolds, fluid flow units (e.g., actuators, pumps, compressors) or the like. Likewise, the outlet channel segment (e.g., channel segment **208**, reservoir **6504**, etc.) may be fluidly coupled to a receiving vessel or conduit for the partitioned cells for subsequent processing. Again, this may be a reservoir defined in the body of a microfluidic device, or it may be a fluidic conduit for delivering the partitioned cells to a subsequent process operation, instrument or component.

[0633] The methods and systems described herein may be used to greatly increase the efficiency of single cell applications and/or other applications receiving droplet-based input. For example, following the sorting of occupied cells and/or appropriately-sized cells, subsequent operations that can be performed can include generation of amplification products, purification (e.g., via solid phase reversible immobilization (SPRI)), further processing (e.g., shearing, ligation of functional sequences, and subsequent amplification (e.g., via PCR)). These operations may occur in bulk (e.g., outside the partition). In the case where a partition is a droplet in an emulsion, the emulsion can be broken and the contents of the droplet pooled for additional operations. Additional reagents that may be co-partitioned along with the barcode bearing bead may include oligonucleotides to block ribosomal RNA (rRNA) and nucleases to digest genomic DNA from cells. Alternatively, rRNA removal agents may be applied during additional processing operations. The configuration of the constructs generated by such a method can help minimize (or avoid) sequencing of the poly-T sequence during sequencing and/or sequence the 5' end of a polynucleotide sequence. The amplification products, for example, first amplification products and/or second amplification products, may be subject to sequencing for sequence analysis. In some cases, amplification may be performed using the Partial Hairpin Amplification for Sequencing (PHASE) method.

[0634] A variety of applications require the evaluation of the presence and quantification of different biological particle or organism types within a population of biological particles, including, for example, microbiome analysis and characterization, environmental testing, food safety testing, epidemiological analysis, e.g., in tracing contamination or the like.

Computer Systems

[0635] The present disclosure provides computer control systems that are programmed to implement methods of the disclosure, i.e., protocols of the disclosure. For example, the present disclosure provides computer control systems programmed to implement method **2000** of the present disclosure. FIG. **17** shows a computer system **1701** that is programmed or otherwise configured to implement methods of the disclosure including nucleic acid sequencing methods, cell surface feature identification methods, interpretation of nucleic acid sequencing data and analysis of cellular nucleic acids, such as RNA (e.g., mRNA), interpretation of nucleic acid sequencing data and analysis of nucleic acids derived from the characterization of cell surface features, and characterization of cells from sequencing data. The computer system **1701** can be an electronic device of a user or a computer system that is remotely located with respect to the electronic device. The electronic device can be a mobile electronic device.

[0636] The computer system **1701** includes a central processing unit (CPU, also "processor" and "computer processor" herein) **1705**, which can be a single core or multi core processor, or a plurality of processors for parallel processing. The computer system **1701** also includes memory or memory location **1710** (e.g., random-access memory, read-only memory, flash memory), electronic storage unit **1715** (e.g., hard disk), communication interface **1720** (e.g., network adapter) for communicating with one or more other systems, and peripheral devices **1725**, such as cache, other memory, data storage and/or electronic display adapters. The memory **1710**, storage unit **1715**, interface **1720** and peripheral devices **1725** are in communication with the CPU **1705** through a communication bus (solid lines), such as a motherboard. The storage unit **1715** can be a data storage unit (or data repository) for storing data. The computer system **1701** can be operatively coupled to a computer network ("network") **1730** with the aid of the communication interface

1720. The network **1730** can be the Internet, an internet and/or extranet, or an intranet and/or extranet that is in communication with the Internet. The network **1730** in some cases is a telecommunication and/or data network. The network **1730** can include one or more computer servers, which can enable distributed computing, such as cloud computing. The network **1730**, in some cases with the aid of the computer system **1701**, can implement a peer-to-peer network, which may enable devices coupled to the computer system **1701** to behave as a client or a server.

[0637] The CPU **1705** can execute a sequence of machine-readable instructions, which can be embodied in a program or software. The instructions may be stored in a memory location, such as the memory **1710**. The instructions can be directed to the CPU **1705**, which can subsequently program or otherwise configure the CPU **1705** to implement methods of the present disclosure. Examples of operations performed by the CPU **1705** can include fetch, decode, execute, and writeback.

[0638] The CPU **1705** can be part of a circuit, such as an integrated circuit. One or more other components of the system **1701** can be included in the circuit. In some cases, the circuit is an application specific integrated circuit (ASIC).

[0639] The storage unit **1715** can store files, such as drivers, libraries and saved programs. The storage unit **1715** can store user data, e.g., user preferences and user programs. The computer system **1701** in some cases can include one or more additional data storage units that are external to the computer system **1701**, such as located on a remote server that is in communication with the computer system **1701** through an intranet or the Internet.

[0640] The computer system **1701** can communicate with one or more remote computer systems through the network **1730**. For instance, the computer system **1701** can communicate with a remote computer system of a user. Examples of remote computer systems include personal computers (e.g., portable PC), slate or tablet PC's (e.g., Apple® iPad, Samsung® Galaxy Tab), telephones, Smart phones (e.g., Apple® iPhone, Android-enabled device, Blackberry®), or personal digital assistants. The user can access the computer system **1701** via the network **1730**.

[0641] Methods as described herein can be implemented by way of machine (e.g., computer processor) executable code stored on an electronic storage location of the computer system **1701**, such as, for example, on the memory **1710** or electronic storage unit **1715**. The machine executable or machine readable code can be provided in the form of software. During use, the code can be executed by the processor **1705**. In some cases, the code can be retrieved from the storage unit **1715** and stored on the memory **1710** for ready access by the processor **1705**. In some situations, the electronic storage unit **1715** can be precluded, and machine-executable instructions are stored on memory **1710**.

[0642] The code can be pre-compiled and configured for use with a machine having a processer adapted to execute the code, or can be compiled during runtime. The code can be supplied in a programming language that can be selected to enable the code to execute in a pre-compiled or as-compiled fashion.

[0643] Aspects of the systems and methods provided herein, such as the computer system **1701**, can be embodied in programming. Various aspects of the technology may be thought of as "products" or "articles of manufacture" typically in the form of machine (or processor) executable code and/or associated data that is carried on or embodied in a type of machine readable medium. Machine-executable code can be stored on an electronic storage unit, such as memory (e.g., read-only memory, random-access memory, flash memory) or a hard disk. "Storage" type media can include any or all of the tangible memory of the computers, processors or the like, or associated modules thereof, such as various semiconductor memories, tape drives, disk drives and the like, which may provide non-transitory storage at any time for the software programming. All or portions of the software may at times be communicated through the Internet or various other telecommunication networks. Such communications, for example, may enable loading of the software from one computer or processor into another, for example, from a management server or host computer into the computer platform of an application server. Thus, another type of media that may bear the software elements includes optical, electrical and electromagnetic waves, such as used across physical interfaces between local devices, through wired and optical landline networks and over various air-links. The physical elements that carry such waves, such as wired or wireless links, optical links or the like, also may be considered as media bearing the software. As used herein, unless restricted to non-transitory, tangible "storage" media, terms such as computer or machine "readable medium" refer to any medium that participates in providing instructions to a processor for execution.

[0644] Hence, a machine readable medium, such as computer-executable code, may take many forms, including but not limited to, a tangible storage medium, a carrier wave medium or physical transmission medium. Non-volatile storage media include, for example, optical or magnetic disks, such as any of the storage devices in any computer(s) or the like, such as may be used to implement the databases, etc. shown in the drawings. Volatile storage media include dynamic memory, such as main memory of such a computer platform. Tangible transmission media include coaxial cables; copper wire and fiber optics, including the wires that comprise a bus within a computer system. Carrier-wave transmission media may take the form of electric or electromagnetic signals, or acoustic or light waves such as those generated during radio frequency (RF) and infrared (IR) data communications. Common forms of computer-readable media therefore include for example: a floppy disk, a flexible disk, hard disk, magnetic tape, any other magnetic medium, a CD-ROM, DVD or DVD-ROM, any other optical medium, punch cards paper tape, any other physical storage medium with patterns of holes, a RAM, a ROM, a PROM and EPROM, a FLASH-EPROM, any other memory chip or cartridge, a carrier wave transporting data or instructions, cables or links transporting such a carrier wave, or any other medium from which a computer may read programming code and/or data. Many of these forms of computer readable media may be involved in carrying one or more sequences of one or more instructions to a processor for execution.

[0645] The computer system **1701** can include or be in communication with an electronic display screen **1735** that comprises a user interface (UI) **1740** for providing, for example, results of nucleic acid sequencing, analysis of nucleic acid sequencing data, characterization of nucleic acid sequencing samples, cell characterizations, etc. Examples of UI's include, without limitation, a graphical

user interface (GUI) and web-based user interface. The system **1701** may comprise an electronic display screen **1735** comprising a user interface **1740** that displays a graphical element that is accessible by a user to execute a protocol per the methods described herein, (e.g. to characterize cells), and a computer processor coupled to the electronic display screen and programmed to execute the protocol upon selection of the graphical element by the user.

[0646] Methods and systems of the present disclosure can be implemented by way of one or more algorithms. An algorithm can be implemented by way of software upon execution by the central processing unit **1705**. The algorithm can, for example, initiate nucleic acid sequencing, process nucleic acid sequencing data, interpret nucleic acid sequencing results, characterize nucleic acid samples, characterize cells, etc.

[0647] Barcoded oligonucleotides as described elsewhere herein may be generated in any suitable manner and comprise one or more sequences in addition to a barcode sequence. As noted elsewhere herein, one such sequence can be a priming sequence that can aid in barcoding analytes. Moreover, a barcoded oligonucleotide may also comprise one or more additional functional sequences that may, for example, aid in rendering the barcoded oligonucleotide compatible with a given sequencing platform (e.g., functional sequences may be flow cell adaptor immobilization sequences (such as, for example, P7 and P5 from an Illumina platform), sequencing primer binding site sequences (such as, for example, R1 from an Illumina platform), and other priming sites for downstream amplification, such as, for example, a Nextera functional sequence or a TruSeq functional sequence.

[0648] In some cases, barcoded oligonucleotides are coupled to beads and beads may comprise oligonucleotides having a first type functional sequence at a given position and oligonucleotides having a second, different type of functional sequence at the given position. An example is depicted in FIG. **50**A. As shown in FIG. **50**A, a bead may be coupled to oligonucleotides comprising a TruSeq functional sequence and also to oligonucleotides comprising a Nextera functional sequence. Onto each of these sequences additional sequences can be added to generate a full oligonucleotide also comprising a nucleic acid barcode sequence, an optional UMI sequence and a priming sequence. Attachment of these sequences can be via ligation (including via splint ligation as is described in U.S. Patent Publication No. 20140378345, which is herein incorporated by reference in its entirety) or any other suitable route. Sequences of example barcoded oligonucleotides comprising a TruSeq functional group are shown in FIG. **50**B and sequences of example barcoded oligonucleotides comprising a Nextera functional group are shown in FIG. **50**C. Each of the example barcoded oligonucleotides shown in FIG. **50**B and FIG. **50**B (top sequence for each construct) are shown hybridized with splint sequences (bottom sequence for each construct) that can be helpful in constructing complete barcoded oligonucleotides.

[0649] In some aspects, methods provided herein may also be used to prepare polynucleotide contained within cells in a manner that enables cell-specific information to be obtained. The methods enable detection of genetic variations (e.g., SNPs, mutations, indels, copy number variations, transversions, translocations, inversions, etc.) from very small samples, such as from samples comprising about 10-100 cells. In some cases, about 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90 or 100 cells may be used in the methods described herein. In some cases, at least about 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90 or 100 cells may be used in the methods described herein. In other cases, at most about 5, 10, 20, 30, 40, 50, 60, 70, 80, 90 or 100 cells may be used in the methods described herein.

[0650] In an example, a method comprises partitioning a cellular sample (or crude cell extract) such that at most one cell (or extract of one cell) is present per partition, lysing the cells, fragmenting the polynucleotides contained within the cells by any of the methods described herein, attaching the fragmented polynucleotides to barcodes, pooling, and sequencing.

[0651] The barcodes and other reagents may be contained within a microcapsule. These microcapsules may be loaded into a partition (e.g., a microwell, a droplet) before, after, or concurrently with the loading of the cell, such that each cell is contacted with a different microcapsule. This technique may be used to attach a unique barcode to polynucleotides obtained from each cell. The resulting tagged polynucleotides may then be pooled and sequenced, and the barcodes may be used to trace the origin of the polynucleotides. For example, polynucleotides with identical barcodes may be determined to originate from the same cell, while polynucleotides with different barcodes may be determined to originate from different cells.

[0652] The methods described herein may be used to detect the distribution of oncogenic mutations across a population of cancerous tumor cells. For example, some tumor cells may have a mutation, or amplification, of an oncogene (e.g., HER2, BRAF, EGFR, KRAS) in both alleles (homozygous), others may have a mutation in one allele (heterozygous), and still others may have no mutation (wild-type). The methods described herein may be used to detect these differences, and also to quantify the relative numbers of homozygous, heterozygous, and wild-type cells. Such information may be used, for example, to stage a particular cancer and/or to monitor the progression of the cancer and its treatment over time.

[0653] In some examples, this disclosure provides methods of identifying mutations in two different oncogenes (e.g., KRAS and EGFR). If the same cell comprises genes with both mutations, this may indicate a more aggressive form of cancer. In contrast, if the mutations are located in two different cells, this may indicate that the cancer is more benign, or less advanced.

EXAMPLES

Example I: Cellular RNA Analysis Using Emulsions

[0654] In an example, reverse transcription with template switching and cDNA amplification (via PCR) is performed in emulsion droplets with operations as shown in FIG. **9**A. The reaction mixture that is partitioned for reverse transcription and cDNA amplification (via PCR) includes 1,000 cells or 10,000 cells or 10 ng of RNA, beads bearing barcoded oligonucleotides/0.2% Tx-100/5× Kapa buffer, 2× Kapa HS HiFi Ready Mix, 4 μM switch oligo, and Smartscribe. Where cells are present, the mixture is partitioned such that a majority or all of the droplets comprise a single cell and single bead. The cells are lysed while the barcoded oligonucleotides are released from the bead, and the poly-T

segment of the barcoded oligonucleotide hybridizes to the poly-A tail of mRNA that is released from the cell as in operation **950**. The poly-T segment is extended in a reverse transcription reaction as in operation **952** and the cDNA is amplified as in operation **954**. The thermal cycling conditions are 42° C. for 130 minutes; 98° C. for 2 min; and 35 cycles of the following 98° C. for 15 sec, 60° C. for 20 sec, and 72° C. for 6 min. Following thermal cycling, the emulsion is broken and the transcripts are purified with Dynabeads and 0.6×SPRI as in operation **956**.

[0655] The yield from template switch reverse transcription and PCR in emulsions is shown for 1,000 cells in FIG. **13**A and 10,000 cells in FIG. **13**C and 10 ng of RNA in FIG. **13**B (Smartscribe line). The cDNAs from RT and PCR performed in emulsions for 10 ng RNA is sheared and ligated to functional sequences, cleaned up with 0.8×SPRI, and is further amplified by PCR as in operation **958**. The amplification product is cleaned up with 0.8×SPRI. The yield from this processing is shown in FIG. **13**B (SSII line).

Example II: Cellular RNA Analysis Using Emulsions

[0656] In another example, reverse transcription with template switching and cDNA amplification (via PCR) is performed in emulsion droplets with operations as shown in FIG. **9**A. The reaction mixture that is partitioned for reverse transcription and cDNA amplification (via PCR) includes Jurkat cells, beads bearing barcoded oligonucleotides/0.2% TritonX-100/5×Kapa buffer, 2×Kapa HS HiFi Ready Mix, 4 μM switch oligo, and Smartscribe. The mixture is partitioned such that a majority or all of the droplets comprise a single cell and single bead. The cells are lysed while the barcoded oligonucleotides are released from the bead, and the poly-T segment of the barcoded oligonucleotide hybridizes to the poly-A tail of mRNA that is released from the cell as in operation **950**. The poly-T segment is extended in a reverse transcription reaction as in operation **952** and the cDNA is amplified as in operation **954**. The thermal cycling conditions are 42° C. for 130 minutes; 98° C. for 2 min; and 35 cycles of the following 98° C. for 15 sec, 60° C. for 20 sec, and 72° C. for 6 min. Following thermal cycling, the emulsion is broken and the transcripts are cleaned-up with Dynabeads and 0.6×SPRI as in operation **956**. The yield from reactions with various cell numbers (625 cells, 1,250 cells, 2,500 cells, 5,000 cells, and 10,000 cells) is shown in FIG. **14**A. These yields are confirmed with GADPH qPCR assay results shown in FIG. **14**B.

Example III: RNA Analysis Using Emulsions

[0657] In another example, reverse transcription is performed in emulsion droplets and cDNA amplification is performed in bulk in a manner similar to that as shown in FIG. **9**C. The reaction mixture that is partitioned for reverse transcription includes beads bearing barcoded oligonucleotides, 10 ng Jurkat RNA (e.g., Jurkat mRNA), 5× First-Strand buffer, and Smartscribe. The barcoded oligonucleotides are released from the bead, and the poly-T segment of the barcoded oligonucleotide hybridizes to the poly-A tail of the RNA as in operation **961**. The poly-T segment is extended in a reverse transcription reaction as in operation **963**. The thermal cycling conditions for reverse transcription are one cycle at 42° C. for 2 hours and one cycle at 70° C. for 10 min. Following thermal cycling, the emulsion is broken and RNA and cDNAs are denatured as in operation **962**. A second strand is then synthesized by primer extension with a primer having a biotin tag as in operation **964**. The reaction conditions for this primer extension include cDNA as the first strand and biotinylated extension primer ranging in concentration from 0.5-3.0 μM. The thermal cycling conditions are one cycle at 98° C. for 3 min and one cycle of 98° C. for 15 sec, 60° C. for 20 sec, and 72° C. for 30 min. Following primer extension, the second strand is pulled down with Dynabeads MyOne Streptavidin C1 and T1, and cleaned-up with Agilent SureSelect XT buffers. The second strand is pre-amplified via PCR as in operation **965** with the following cycling conditions—one cycle at 98° C. for 3 min and one cycle of 98° C. for 15 sec, 60° C. for 20 sec, and 72° C. for 30 min. The yield for various concentrations of biotinylated primer (0.5 μM, 1.0 M, 2.0 μM, and 3.0 μM) is shown in FIG. **15**.

Example IV: RNA Analysis Using Emulsions

[0658] In another example, in vitro transcription by T7 polymerase is used to produce RNA transcripts as shown in FIG. **10**. The mixture that is partitioned for reverse transcription includes beads bearing barcoded oligonucleotides which also include a T7 RNA polymerase promoter sequence, 10 ng human RNA (e.g., human mRNA), 5× First-Strand buffer, and Smartscribe. The mixture is partitioned such that a majority or all of the droplets comprise a single bead. The barcoded oligonucleotides are released from the bead, and the poly-T segment of the barcoded oligonucleotide hybridizes to the poly-A tail of the RNA as in operation **1050**. The poly-T segment is extended in a reverse transcription reaction as in operation **1052**. The thermal cycling conditions are one cycle at 42° C. for 2 hours and one cycle at 70° C. for 10 min. Following thermal cycling, the emulsion is broken and the remaining operations are performed in bulk. A second strand is then synthesized by primer extension as in operation **1054**. The reaction conditions for this primer extension include cDNA as template and extension primer. The thermal cycling conditions are one cycle at 98° C. for 3 min and one cycle of 98° C. for 15 sec, 60° C. for 20 sec, and 72° C. for 30 min. Following this primer extension, the second strand is purified with 0.6×SPRI. As in operation **1056**, in vitro transcription is then performed to produce RNA transcripts. In vitro transcription is performed overnight, and the transcripts are purified with 0.6×SPRI. The RNA yields from in vitro transcription are shown in FIG. **16**.

Example V: Delivering Lysis Agent to a Partition Using Gel Beads

[0659] A lysis agent is introduced into the partition (GEM) via the gel bead suspension (GBS). The lysis agent is a surfactant that causes wetting failures (uncontrolled droplet formation) to occur when its concentration in the GBS exceeds a threshold.

[0660] A larger gel bead can be used to increase the in-partition concentration of the lysis agent, without increasing the in-GBS concentration (to avoid wetting failures) and without decreasing the total volume of the partition (which may not be reduced without decreasing the sensitivity of the assay) (FIG. **36**A). Alternatively, a larger gel bead can be used to increase the volume of the partition (which increases

the sensitivity of the assay) and preserve the existing in-partition lysis agent concentration without increasing the in-GBS concentration.

[0661] The size of the gel bead can also affect how cells are partitioned. By replacing a portion of the sample volume (Z2) with the gel bead suspension volume (Z1), larger gel beads decrease the in-partition concentration of cells, which, according to Poisson statistics, results in a lower probability of the unfavorable encapsulation of more than one cell per partition (FIG. **36**B).

Example VI: Producing CD3 Protein Conjugated with Short ssDNA Molecules

[0662] The CD3 protein and the ssDNA molecule are first activated for click chemistry reaction. The CD3 protein is activated with 5-(methacrylamido)tetrazole (MTet) and the ssDNA molecule is activated with trans-cyclooctene (TCO). The ssDNA molecule comprises a biotin group. The activated CD3 protein and ssDNA molecule are mixed for conjugation by click chemistry reactions. The ssDNA molecule concentration is 5 times excess over the CD3 protein concentration to avoid multiple barcode copies conjugating on the same protein molecule. In some cases, the ssDNA concentration is 10 times excess over the CD3 protein to maximize barcode attachment. A biotin group may also be incorporated in the activated CD3-ssDNA conjugate for purification. The CD3 protein and ssDNA conjugate is purified and tested as shown in FIG. **37**.

Example VII: Labelling Jurkat Cells with Human CD3 and Mouse CD3

[0663] The impact of DNA conjugation on the binding of CD3 on Jurkat cells is tested. Human CD3 (hCD3, MCA463) and mouse CD3 (mCD3, MCA500) are incubated with AF488-NHS, where the concentration of AF499-NHS is 1×, 2×, 5×, and 10× excess over the CD3 protein, in order to generate labeled CD3, where the AF999 is coupled to an amine of the CD3. The conjugated hCD3 and mCD3 are incubated with Jurkat cells. Unbound CD3 proteins are washed away. The fluorescence signals from the labeled cells are determined (FIG. **38**). The fluorescent signals are normalized by comparing to commercial Jurkat cells control. The data show that Jurkat cells specifically bind to hCD3 over mCD3, indicating that the conjugation of dye/DNA does not affect the binding of CD3 proteins with Jurkat cells. Blocking reagents (e.g., FBS, 5% BSA) may be added to improve specificity.

Example VIII: Conjugating a DNA Barcode to IgG of an Antibody

[0664] An antibody is incubated with Methyltetrazine-PEG5-NHS Ester at room temperature for 1 hour and desalted. A DNA barcode of about 65 nt long is incubated with TCO-PEG4-NHS Ester at room temperature for an hour and desalted. The resulting antibody and DNA barcode are incubated at room temperature for 2 hours for conjugation. FIG. **39**A shows the conjugation strategy. The conjugated antibody-DNA complex is subject to protein gel analysis. As shown in FIG. **39**B, protein gel shifts of about 20 kDa indicates successful conjugation of the DNA barcode to IgG of the antibody. Multiple viable chemistries for primary antibody barcoding are validated (e.g., mTet, dibenzocyclooctyne (DBCO), SiteClick). The conjugated antibody-DNA complex is incubated with cells for labelling.

Example IX: Conjugating Oligonucleotides to Antibodies Using Antibody-Binding Proteins

[0665] Antibody-binding proteins Protein X (Protein A or Protein G) are functionalized with dibenzocyclooctyne-N-hydroxysuccinimidyl ester (DBCO-NHS). Fluorescein amidite (FAM)-labeled oligoX22-azide (3 eq) is used as the oligonucleotides to be conjugated with the antibody-binding proteins. The functionalized antibody-binding proteins and the oligonucleotides are conjugated as shown in FIG. **40**. The degree of conjugation between the dibenzocyclooctyne (DBCO) and Protein G may be controlled based on Gong et al., Simple Method To Prepare Oligonucleotide-Conjugated Antibodies and Its Application in Multiplex Protein Detection in Single Cells. Bioconjugate Chem., 2016, which is incorporated herein by reference in its entirety. Degree of DBCO incorporation may be controlled by adjusting input DBCO-NHS concentration as shown in FIG. **41**.

[0666] Moreover, the degree of conjugation may be controlled through oligonucleotide equivalence as shown in FIG. **42**. A crude protein-oligonucleotide conjugation reaction was analyzed by gel electrophoresis (SDS-PAGE) to determine conjugation efficiency and the number of oligonucleotides conjugated. Increase of oligonucleotide equivalence with respect to the protein leads to a higher degree of conjugation as shown in FIG. **42**. Because the oligonucleotide contains a fluorescent molecule, the unused oligonucleotide can easily be visualized with in-gel fluorescence imaging (black panel in FIG. **42**).

[0667] The oligonucleotide-Protein X conjugates are incubated with CD47 antibodies to form labeled antibodies. The labeled antibodies are incubated with Jurkat cells and washed twice to make labeled cells. The labelling of cells is measured by fluorescence signals using flow cytometry (FIG. **43**).

Example X: Producing a Bead Coupled with Oligonucleotides with Different Primer Sequences

[0668] This example shows a method for producing a bead coupled with oligonucleotides with different primer sequences. The work flow is shown in FIG. **44**. A barcode sequence **4421** is ligated to a sequence primer R1 **4411** coupled to a bead. The R1 primer **4411** and barcode sequence **4421** form the backbone **4420** of the oligonucleotides on the bead. A plurality of backbone oligonucleotides **4420** are coupled to the same bead. Different primers sequences are then ligated to the backbone oligonucleotides **4420**. The primers include a poly-T primer **4431** that targets the poly-A of mRNA molecules. The primers also include a target specific primer, e.g., an antibody target primer that binds to a barcode on an antibody. After the second ligation, the bead comprises oligonucleotides with poly-T primers (**4430**) and oligonucleotides with antibody target primers (**4440**). The resulting product from the method is a bead coupled with a plurality of oligonucleotides (FIG. **45**A). All of the oligonucleotides comprise the same backbone. Some of the oligonucleotide comprises poly-T primers and some comprises the antibody target primers. Beads with 0%, 5%, 15%, and 25% of coupled oligonucleotides containing antibody target primers are analyzed by gel electrophoresis (FIG. **45**B)

Example XI: Barcoding Antibody Labelling Agents and Cell Surface Feature Analysis

[0669] In a first set of experiments, a barcoded oligonucleotide comprising an azide functional group and a FAM dye was conjugated to a Protein G labelling agent using a click chemistry reaction scheme. The barcoded oligonucleotide included a barcode sequence that may be used to identify Protein G and also a sequence that may be used as a priming site. Protein G was mixed with increasingly higher molar equivalents of DBCO-NHS (0×, 1×, 2×, 4× and 6×) in a series of mixtures. The DBCO-NHS was used to activate amine groups to become reactive to azide. Also included were varying equivalents of azide oligonucleotide to DBCO (0×, 1×, 1.5× and 2×) in the mixtures. Reactions were then allowed to proceed for 4 hours and the reaction mixtures evaluated with gel electrophoresis on a 4-12% bis-Tris gel. The results of the analysis are graphically depicted in FIGS. **47**A-**47**B. Protein G having up to 6 oligonucleotides linked were observed.

[0670] The various labeled Protein G moieties were then mixed with CD47 antibody to bind the labeled Protein G moieties to CD47 antibodies. The resulting Protein G-CD47 complexes were then incubated with 293T cells such that the complexes may bind CD47 on the surface of cells. Cells were washed to remove unbound complex and then subject to flow cytometry to observe binding of antibodies via the oligo-bound FAM dye. Results of flow cytometry are graphically depicted in FIG. **48**.

[0671] Next, labeled cells were mixed with a bead coupled to an oligonucleotide comprising a nucleic acid barcode sequence, a UMI and a poly-T sequence capable of binding the poly-A sequence of mRNA transcripts in a cell. Also included was a barcoded primer having a priming sequence capable of specifically hybridizing the barcoded oligonucleotide coupled to CD47 antibodies via the barcoded oligonucleotide's priming site. The mixture was then partitioned into a droplets in an emulsion. The emulsion was then subject to conditions suitable for priming sequences to hybridize with their respective targets (mRNA or barcoded antibody oligonucleotide) and for extension of primers via the action of a polymerase or reverse transcriptase. Extension generated barcoded constructs. Following reactions, the emulsion was broken. Barcoded transcript constructs still attached to beads were removed by removing beads and the supernatant subject to 2×SPRI separation to recover the ~110 bp antibody barcode. The recovered products were then analyzed, with results shown in FIGS. **49**A and **49**B.

Example XII: Coupling of Barcodes

[0672] In a bulk experiment, two oligonucleotides shown in FIG. **51**A, **5101** and **5102**, were linked together via extension reactions. Oligonucleotide **5101** represented an oligonucleotide comprising a barcode sequence that may be used to identify a partition comprising the oligonucleotide **5101** and oligonucleotide **5102** represented an oligonucleotide comprising a barcode sequence that may be used to identify a labelling agent, such as an antibody coupled to oligonucleotide **5102**. Oligonucleotide **5102** also included a FAM dye and a 3' reverse complement of a template switch oligonucleotide spacer-rGrGrG region included on oligonucleotide **5101**. In the experiment, 50 nM AbBC of oligonucleotide **5102** was mixed with oligonucleotide **5101** in two separate mixtures. Included in the mixture were reagents for conducting a primer extension reaction, including one of two reverse transcriptases capable of facilitating a primer extension reaction and dNTPs. Extension products were then analyzed via capillary electrophoresis.

[0673] The results of the experiment are graphically shown in FIG. **51**B. As shown, expected extension products having both a sequence corresponding to the barcode sequence of oligonucleotide **5101** (or a complement of the barcode sequence) and a sequence corresponding to the barcode sequence of oligonucleotide **5102** (or a complement of the barcode sequence) were detected. These results confirm that the reverse transcriptases tested may be used to generate extension products having sequences corresponding to both barcode sequences of oligonucleotides **5101** and **5102**.

Example XIII: Single-Cell Barcode Behavior

[0674] Anti-CD47 and Anti-CD99 antibodies were obtained and both types were coupled to an oligonucleotide comprising a barcode sequence that was suitable for identifying its respective antibody and also comprising a unique molecular identification (UMI) sequence and a template switch oligonucleotide reverse complement sequence (e.g., C C C). The antibody-oligonucleotide constructs were generated by linking the oligonucleotides to protein G and then binding the protein G-oligonucleotide constructs to the antibodies. The oligonucleotides were linked to protein G by modifying protein G with a single cysteine residue and linking it to oligonucleotides via the cysteine residue. Protein G also included a His×6 tag (SEQ ID NO: 40) which may be used to separate unconjugated oligonucleotides from those coupled to Protein G. Sample data from gel electrophoresis analysis of generated constructs is shown in FIG. **52**A. The lanes in FIG. **52**A show expression of a cysteine-containing protein G antibody binding protein. The culture lane depicts a homogenized cell culture, the flow through lane depicts is all proteins that did not bind to a nickel-NTA column, and the two elution lanes are eluted purified protein G.

[0675] Jurkat cells were then incubated with antibody-oligonucleotide constructions to bind antibodies to the surface of cells via their respective cell surface feature targets. The cells were then partitioned into aqueous droplets in an emulsion, along with beads linked to oligonucleotides comprising a barcode sequence, a UMI sequence, a priming sequences capable of hybridizing with antibody-bound oligonucleotides (e.g., primer sequence include a template switch sequence, such as rGrGrG). A reducing agent, capable of disrupting disulfide linkages of beads and linkages between beads and its oligonucleotides was also included in the partitions. The reducing agent released the bead's oligonucleotides and the droplets were then subjected to conditions suitable for hybridizing the previously bead-bound oligonucleotides to cell-bound antibody oligonucleotides via an interaction of sequences of the two oligonucleotides, including via an rGrGrG/CCC interaction. While a particular sequence is shown, hybridization may be achieved via any constant sequence at the ends of the two oligonucleotides.

[0676] The two hybridized oligonucleotides were then extended in primer extension reactions to generate constructs comprising sequences corresponding to both bead oligonucleotide and antibody barcode sequences, similar to the example scheme shown in FIG. **52**B (panel I). The

emulsion was then broken, the extended products further processed and then subject to sequencing. Sequencing results for Jurkat+CD47 and Jurkat+CD47/CD99 runs are graphically depicted in panels I and II, respectively, of FIG. **53**A and tabulated in FIG. **53**B. The data shown in FIG. **53**A and FIG. **53**B indicate that the antibody-oligonucleotide constructions comprising barcode sequences were able to show single cell behavior, as evidenced, for example, by an approximately 2-log enrichment of antibody-oligonucleotide UMIs in bead-originating barcode constructs corresponding to cells.

Example XIV: Antibody Barcode Staining Parameters

[0677] Various parameters associated with methods described herein were evaluated in the context of their effects on antibody-barcode construct binding, including a reverse transcription deactivation process and the concentration of reducing agent in partitions (e.g., reducing agent used to degrade barcoded beads as described elsewhere herein).

[0678] Reverse transcription can be deactivated by elevating the temperature of reverse transcription reaction mixtures to relatively high temperatures (a "heat kill"). However, such high temperatures may result in antibody-barcode constructs precipitating out of reaction mixtures, resulting in an inability to bind to cells. Various anti-CD3 barcode construct samples were tested against cells, with some samples subject to heat kill and others not subjected to heat kill. Sequencing data for the experiments is tabulated in FIG. **54**. As shown in FIG. **54**, a number of sequencing metrics are improved when no heat kill is used, including reads mapped confidently and complexity.

[0679] Moreover, high concentrations of reducing agents can also degrade antibodies used to label cell-surface features. Accordingly, the effect of lower reducing agent (e.g., DTT) concentration by 10-fold was tested on overall efficiency of reverse transcription in partitions. As show in FIG. **55**, traces are similar for all samples tested (22 mM DTT vs. 2.2 mM DTT), suggesting that reverse transcription, as described elsewhere herein, can effectively proceed at substantially reduced DTT concentrations. In another experiment, 0.15 mM DTT was also shown to be effective.

Example XV: Linking T-Cell Receptor Sequence to Antigen Binding Phenotype Using Barcoded MHC-Antigen Multimers

[0680] Many TCRs can bind a particular antigen (with varying affinity) and identifying individual clonotypes specific to a particular antigen is difficult. While flow cytometry and bead-based enrichment schemes allow physical sorting of antigen-binding cells, when cells are rare or samples are limited, cell losses associated with traditional methodologies can be unacceptable. Moreover, traditional approaches based on fluorescent detection have important limitations with regard to multiplexing (the ability to simultaneously assay the binding properties of multiple independent antigens/ligands in single experiment) due to the small number of spectrally distinguishable fluorescent labels that can be effectively used in combination. Furthermore, multiple antigen-binding clonotypes may be present in a heterogeneous sample, which makes identifying specific antigen-binding TCR complexes difficult, even when the cells expressing antigen-binding clonotypes are physically sorted.

[0681] The compositions, methods, and systems described herein allow functionalization of MHC-peptide multimers with an oligonucleotide (DNA or RNA) that includes a unique peptide barcode sequence specific to the MHC-peptide identity (e.g., Barcode 1 associated with peptide EGALIYWPN (SEQ ID NO: 62), Barcode 2 associated with peptide AHMRDSQQ (SEQ ID NO: 63), etc). A single peptide-MHC complex or peptide-MHC library can be exposed to a cell population (e.g., T-cells) to produce cells "tagged" with barcoded MHC multimers. These cells can then be partitioned and processed as described herein to assemble TCR sequences and quantify the number of MHC-peptide barcodes associated with each cell. Clonotypes with low levels of MHC-peptide derived UMIs have a low affinity for the MHC-peptide while clonotypes with high levels of the MHC-peptide UMIs have a high affinity for the antigen.

[0682] Barcoded, peptide-bound MHC tetramers bound to a streptavidin core were generated generally as depicted in FIG. **54**A and as described below. Although Class I MHC-tetramers were utilized in the following series of experiments, there are many possible configurations of Class I and/or Class II MHC-antigen multimers that can be utilized with the compositions, methods, and systems disclosed herein, e.g., MHC pentamers (MHC assembled via a coiled-coil domain, e.g., Pro5® MHC Class I Pentamers, (ProImmune, Ltd.), MHC decorated dextran molecules (e.g., MHC Dextramer® (Immudex)), etc.

[0683] Streptavidin molecules (**5701**) were conjugated to a hybridization oligonucleotide (**5702**) using general lysine chemistry (streptavidin modified via lysine residues with NHS-DBCO; subsequently an azide-modified oligonucleotide was attached via the DBCO functional group) to produce streptavidin-conjugated oligonucleotides (**5703**) as depicted in FIG. **57**A. Streptavidin-conjugated oligonucleotides (**5703**) were then analyzed on a TBE-urea denaturing agarose gel. As shown in FIG. **58**A, 0.6 μM, 1.2 μM, 1.8 μM, 2.4 μM, and 3 μM of unmodified oligonucleotide were all observed to have bands of a similar size while streptavidin-conjugated oligonucleotides exhibited a clear shift in molecular weight indicating successful streptavidin conjugation. The multiple bands observed in the streptavidin-conjugated oligonucleotide lane correspond to conjugated streptavidin molecules with increasing numbers of oligonucleotides attached (e.g., 1 oligo, 2 oligos, 3 oligos, etc.). As seen in FIG. **58**A, streptavidin-conjugated oligonucleotides are produced with minimal excess non-conjugated oligonucleotide.

[0684] Streptavidin-conjugated oligonucleotides (**5703**) were also analyzed on an SDS-PAGE protein gel. As shown in FIG. **58**B, 0.25 μg, 0.5 μg, and 1.0 μg of unmodified streptavidin exhibit a similar molecular weight while streptavidin-conjugated oligonucleotides exhibit a molecular weight shift indicative of streptavidin conjugated with 0, 1, 2, 3, 4 (or more) oligonucleotides. Quantification of the conjugated oligonucleotide can be estimated by comparing the density of the conjugated oligonucleotide bands with the density of the 0.25 μg, 0.5 μg, and 1.0 μg unmodified streptavidin bands. From this comparison, the overall degree of conjugation is approximately 1 oligonucleotide per each streptavidin subunit (resulting in approximately 4 oligonucleotides per each MHC tetramer).

[0685] Following quantification of the degree of conjugation, barcode oligonucleotides (**5708**) are hybridized to the streptavidin-conjugated oligonucleotides (**5703**) via the reverse complement (**5704**) of the hybridization oligo sequence (**5702**) at a stoichiometry of between 0.25:1 to 1:1 of barcode oligonucleotides (**5708**) to streptavidin-conjugated oligonucleotides (**5703**). Here, the barcode oligonucleotides (**5708**) comprise a sequence that is the reverse complement (**5704**) of the hybridization oligo sequence (**5702**), a TruSeq R2 sequencing primer sequence (**5705**), a unique molecular identification (UMI) (series of any "N" nucleotides) and a barcode sequence (**5706**), and an adapter sequence (**5707**) that is complementary to a sequence on a gel bead. Alternatively, the barcode oligonucleotide can be directly conjugated to the streptavidin.

[0686] After hybridization, the barcoded streptavidin (**5709**) is added to a pool of biotinylated HLA-A-02:01 MHC monomers (see, e.g., **5606**) displaying an Epstein-Barr Virus (EBV) peptide antigen (GLCTLVAML (SEQ ID NO: 64)) to produce barcoded MHC tetramers (see, e.g., **5608**). The barcoded streptavidin (**5709**) is added until a 1:1 ratio of biotinylated EBV MHC monomers to biotin binding sites is achieved (**4** biotinylated MHC monomers/streptavidin complex).

[0687] Barcoded MHC tetramers (0.4 μg or 4.0 μg) are then incubated for 30 minutes with ~200,000 (100 μL) EBV antigen-expanded T-cells (Astarte Biologics) and/or ~200,000 (100 μL) of naïve T cells. Cells were washed three times with PBS/1% FBS to remove unbound multimers. The cells were then resuspended in PBS+0.04% BSA and partitioned into droplets comprising a barcoded MHC bound T-cell and a barcoded gel bead (see, e.g., FIG. **11**). Barcoded MHC tetramers are then generally processed as described herein (see, e.g., FIG. **56**C and accompanying text). T-cells are then lysed and released mRNA molecules are generally processed as described herein (see, e.g., FIG. **11** and accompanying text). The droplet emulsion was then broken and bulk PCR-amplification used to enrich for barcoded, full-length V(D)J segments from TCR cDNA. A second library was prepared to quantify the number of MHC-EBV peptide UMIs associated with each cell. The fully constructed sequencing libraries were then sequenced using an Illumina sequencer. T-cell receptor clonotypes were assembled bioinformatically and the number of UMI counts from barcoded MHC tetramers were quantified per cell and per clonotype.

[0688] FIG. **59** shows the number of UMI counts from barcoded MHC tetramers vs. the clonotype frequency as measured by the number of barcodes. For each clonotype detected, the average number of MHC multimer-derived UMI counts per cell-barcode was computed for all cell-associated cell-barcodes corresponding to that clonotype, and the log 10 of one plus its mean UMI counts per cell value is plotted on the y-axis. The number of cell-associated cell-barcodes detected with each clonotype is plotted on the x-axis. For visualization purposes, a random amount of Gaussian noise was added to each point's x and y coordinate values to avoid overplotting. Feature **5901** shows the mean y-axis value of log 10 (1+UMI counts per cell) averaged across all clonotypes from EBV-expanded T-cells incubated with 4 μg MHC multimer ("1 k EBC+4 ug tet"); feature **5902** shows the mean y-axis value of log 10 (1+UMI counts per cell) averaged across all clonotypes from EBV-expanded T-cells incubated with 0.4 μg MHC multimer ("1 k EBC+0.4 ug tet"); feature **5903** shows the mean y-axis value of log 10 (1+UMI counts per cell) averaged across all clonotypes from naïve T-cells incubated with 4 μg MHC multimer ("1 k T+4 ug tet"); and feature **5904** shows the mean y-axis value of log 10 (1+UMI counts per cell) averaged across all clonotypes from naïve T-cells incubated with 0.4 μg MHC multimer ("1 k T+0.4 ug tet"). As seen in FIG. **59**, the EBV-expanded cell types have the most UMI counts associated with the tetramer (Features **5901** and **5902**) as compared to the values obtained for the naïve T cell populations (Features **5903** and **5904**). Moreover, clonotypes from the EBV-expanded cells that occur at high frequency within the EBV-expanded cell population (bounded circle, feature **5905**) exhibited even greater values of MHC-tetramer UMIs, indicating their enriched frequency in the EBV-expanded population is associated with preferential MHC-tetramer binding. Conversely, naïve T-cells are not expected to preferentially bind the antigen and all have low background levels of tetramer-associated UMIs.

[0689] In another experiment, EBV-expanded T-cells were spiked-into a naïve T cell background prior to incubation with the barcoded MHC tetramer described above. Cells were then processed, sequenced, and analyzed as previously described. FIG. **60** shows the number of UMI counts from barcoded MHC tetramers vs. the clonotype frequency from the mixed T-cell population (following the axes and plotting conventions used in FIG. **59**). Feature **6001** shows the mean y-axis value of log 10 (1+UMI counts per cell) averaged across all clonotypes from cells containing clonotypes which were previously observed to occur in at least one sample of independently processed EBV-expanded cells ("EBV (n=1)"); feature **6002** shows the mean y-axis value of log 10 (1+UMI counts per cell) averaged across all clonotypes from cells containing clonotypes which were previously observed to occur in more than one sample of independently processed EBV-expanded cells ("EBV (n>1)"); while feature **6003** shows the mean y-axis value of log 10 (1+UMI counts per cell) averaged across all clonotypes from all cells detected in the experiment ("Other"). As seen in FIG. **60**, while the precise number of cells originating from the EBV spike-in is unknown (due to differences in cell recovery during washing between naïve T cells and EBV-expanded cells), two clonotypes representing a total of four cells (bounded circle, feature **6004**) were detected in this mixed sample that exhibited very high tetramer-associated UMI counts (~1000× greater than background). These four cells were determined to correspond to the clonotype of the most frequently detected cell in the EBV-expanded sample and corresponded to the EBV spike-in cells. Thus, particular clonotypes of interest can be distinguished from a mixed population of cells containing a complex distribution of clonotypes.

Example XVI. Cells Incubated with Cholesterol-Conjugated Feature Barcodes can be Detected in Sequencing Libraries

[0690] Single cell sequencing libraries were prepared and analyzed from cells incubated with and without a cholesterol conjugated-feature barcode to assess the ability to detect the feature barcode in processed libraries.

[0691] Briefly, cells were washed in medium followed by a wash in PBS. The cells were counted and separated into 2 mL Eppendorf tubes and incubated for five minutes at room temperature with: (1) cholesterol-conjugated feature barcodes at a concentration of 1 uM; or (2) 1 uM of feature

barcodes only (i.e., barcodes not conjugated to a cholesterol moiety). Following the incubation, the cells were washed three times in medium. The cells were then pooled and counted. The pooled cell population was then partitioned into droplets as generally described elsewhere herein to generate droplets comprising: (1) a single cell; and (2) a single gel bead comprising releasable nucleic acid barcode molecules attached thereto. The nucleic acid barcode molecules attached to the gel bead comprise a barcode sequence, a UMI sequence, and a GGG-containing capture sequence. The cholesterol-conjugated feature barcodes comprise a CCC-containing sequence complementary to the gel bead oligonucleotide capture sequence.

[0692] Cells in each droplet were then lysed and the cellular nucleic acids (including feature barcodes if present) were barcoded with the cell barcode sequences. Cell barcoded nucleic acids were then pooled and processed to complete library preparation. Fully constructed barcode libraries were analyzed on a BioAnalyzer to detect the presence of the feature barcode.

[0693] FIGS. **69**A-**69**D show BioAnalyzer results for sequencing libraries prepared from four different cell populations (two cell populations incubated with cholesterol-conjugated feature barcodes "oligo133" and two cell populations incubated with feature barcodes only "oligo131" i.e., no cholesterol conjugation). As seen in FIGS. **69**A-**69**B, the signal (as measured by fluorescent units (FU, y-axis)) at ~150 basepairs (the expected size of feature barcodes—see x-axis) was about 500 FU (see arrow FIGS. **69**A-B) for the two cell populations incubated with feature barcodes that were not conjugated to a cholesterol moiety. In contrast, as seen in FIGS. **69**C-**69**D, a signal of over 5,000 FU (FIG. **69**C—see arrow) and ~10,000 FU (FIG. **69**D—see arrow) was observed in libraries prepared from cells incubated with the cholesterol-conjugated feature barcodes. These results indicate that feature barcodes were successfully introduced into the cell populations and that the feature barcodes can be successfully detected when present in a mixed cell, pooled population.

Example XVII. DNA Sequencing Results of Cholesterol-Conjugated Feature Barcode Libraries

[0694] Jurkat cells were washed in medium followed by a wash in PBS, and then counted. 100,000 such cells were split into 5 Eppendorf tubes (2 mL) to generate **5** different cell populations. Individual cell populations (four in total) were then incubated with 0.1 uM or 0.01 uM cholesterol-conjugated feature barcodes (four in total, one for each cell population) for five minutes at room temperature to yield one cell population "tagged" with a first barcode (BC1), one cell population "tagged" with a second barcode (BC2), one cell population "tagged" with a third barcode (BC3), and one cell population "tagged" with a fourth barcode (BC4). One cell population was not incubated with a cholesterol-conjugated feature barcode (background population). The 5 cell populations were then washed in media, pooled into a single tube, and then counted to determine cell numbers. The pooled cell population was then partitioned into single-cell containing droplets for single-cell barcoding as described above. Fully constructed barcode libraries were then sequenced on an Illumina sequencer to detect the presence of the cell and feature barcodes.

[0695] A summary of the analysis of the sequencing results are presented in Table 2. As seen in Table 2, sequencing reads corresponding to cells containing feature barcodes BC1, BC2, BC3, and BC4 were successfully detected from the pooled cell sample at both the 0.1 uM and 0.01 uM concentration of cholesterol-conjugated feature barcodes tested. The "#background" indicates the number of cells associated with the unlabeled population. Two replicates were performed at each concentration (replicate 1 and replicate 2).

TABLE 2

Sequence Analysis of Pooled Cell Populations

Description	Total cells	# BC1 cells	# BC2 cells	# BC3 cells	# BC4 cells	# doublets	# back-ground	mean purity BC1 cells	mean purity BC2 cells	mean purity BC3 cells	mean purity BC4 cells
5'Chol-BC 0.1 uM (Replicate 1)	1593	285	314	303	344	8	339	0.953	0.966	0.961	0.923
5'Chol-BC 0.1 uM (Replicate 2)	1776	303	335	373	361	15	389	0.951	0.964	0.956	0.908
5'Chol-BC 0.01 uM (Replicate 1)	1676	325	337	348	313	11	342	0.936	0.945	0.951	0.871
5'Chol-BC 0.01 uM (Replicate 2)	1602	292	330	326	320	12	322	0.939	0.949	0.955	0.876

[0696] FIGS. **70**A-**70**L show graphs from pooled cell populations incubated with 0.1 μM cholesterol-conjugated feature barcodes showing the number of unique molecular identifier (UMI) counts on the x-axis versus number of cells on the y-axis. FIGS. **70**A-**70**B show log 10 UMI counts of a first feature barcode sequence ("BC1") identified from sequencing reads generated from sequencing libraries prepared from the pooled cell population (FIG. **70**A—replicate 1; FIG. **70**B—replicate 2). From these results, a clearly distinguished BC1-containing cell population can be distinguished **7001***a* (replicate 1) and **7001***b* (replicate 2). FIGS. **70**C-**70**D show log 10 UMI counts of a second feature barcode sequence ("BC2") identified from sequencing reads generated from sequencing libraries prepared from the pooled cell population (FIG. **70**C—replicate 1; FIG. **70**D—replicate 2). From these results, a clearly distinguished BC2-containing cell population can be distinguished **7002***a* (replicate 1) and **7002***b* (replicate 2). FIGS. **70**E-**70**F show log 10 UMI counts of a third feature barcode sequence ("BC3") identified from sequencing reads generated from sequencing libraries prepared from the pooled cell population (FIG. **70**E—replicate 1; FIG. **70**F—replicate 2). From

these results, a clearly distinguished BC3-containing cell population can be distinguished **7003***a* (replicate 1) and **7003***b* (replicate 2). FIGS. **70**G-**70**H show log 10 UMI counts of a fourth feature barcode sequence ("BC4") identified from sequencing reads generated from sequencing libraries prepared from the pooled cell population (FIG. **70**G—replicate 1; FIG. **70**H—replicate 2). From these results, a clearly distinguished BC4-containing cell population can be distinguished **7004***a* (replicate 1) and **7004***b* (replicate 2).

[0697] FIGS. **70**I-**70**J show 3D representations of UMI counts obtained from the pooled cell populations barcoded with 0.1 uM cholesterol-conjugated feature barcodes for replicate 1. Graphs depict UMI counts in linear (FIG. **70**I) and log 10 scale (FIG. **70**J). The three axes of the graphs show UMI counts corresponding to sequencing reads found to contain BC1 (**7005**, **7009**), BC2 (**7006**, **7010**), or BC3 (**7007**, **7011**). UMI counts associated with sequencing reads containing BC4 and unlabeled cells (**7008**, **7070**) are clustered together.

[0698] FIGS. **71**A-**71**L show graphs from pooled cell populations incubated with 0.01 μM cholesterol-conjugated feature barcodes showing the number of unique molecular identifier (UMI) counts on the x-axis versus number of cells on the y-axis. FIGS. **71**A-**71**B show log 10 UMI counts of a first feature barcode sequence ("BC1") identified from sequencing reads generated from sequencing libraries prepared from the pooled cell population (FIG. **71**A—replicate 1; FIG. **71**B—replicate 2). From these results, a clearly distinguished BC1-containing cell population can be distinguished **7101***a* (replicate 1) and **7101***b* (replicate 2). FIGS. **71**C-**71**D show log 10 UMI counts of a second feature barcode sequence ("BC2") identified from sequencing reads generated from sequencing libraries prepared from the pooled cell population (FIG. **71**C—replicate 1; FIG. **71**D—replicate 2). From these results, a clearly distinguished BC2-containing cell population can be distinguished **7102***a* (replicate 1) and **7102***b* (replicate 2). FIGS. **71**E-**71**F show log 10 UMI counts of a third feature barcode sequence ("BC3") identified from sequencing reads generated from sequencing libraries prepared from the pooled cell population (FIG. **71**E—replicate 1; FIG. **71**F—replicate 2). From these results, a clearly distinguished BC3-containing cell population can be distinguished **7103***a* (replicate 1) and **7103***b* (replicate 2). **71**G-**71**H show log 10 UMI counts of a fourth feature barcode sequence ("BC4") identified from sequencing reads generated from sequencing libraries prepared from the pooled cell population (FIG. **71**G—replicate 1; FIG. **71**H—replicate 2). From these results, a clearly distinguished BC4-containing cell population can be distinguished **7104***a* (replicate 1) and **7104***b* (replicate 2).

[0699] FIGS. **71**I-**71**J show 3D representations of UMI counts obtained from the pooled cell populations barcoded with 0.01 uM cholesterol-conjugated feature barcodes for replicate 1. Graphs depict UMI counts in linear (FIG. **71**I) and log 10 scale (FIG. **71**J). The three axes of the graphs show UMI counts corresponding to sequencing reads found to contain BC1 (**7105**, **7109**), BC2 (**7106**, **7110**), or BC3 (**7107**, **7111**). UMI counts associated with sequencing reads containing BC4 and unlabeled cells (**7108**, **7112**) are clustered together.

Example XVIII. DNA Sequencing Results of Antibody-Conjugated Feature Barcode Libraries

[0700] BioLegend "hashing" antibodies that broadly target cell surface proteins across human cell types were provided. The antibodies included a mixture of clones LNH94 (anti-CD298) and 2M2 (anti-?2-microglobulin). The antibodies were pooled into different populations and barcoded with different feature barcodes. Jurkat, Raji, and 293T cells were provided in separate populations and incubated with different antibody-associated feature barcodes. Jurkat cells were stained with antibodies barcoded with Barcode #18 (BC18); Raji cells were stained with antibodies barcoded with Barcode #19 (BC19); and 293T cells were stained with antibodies barcoded with Barcode #20 (BC20). A total of 9,000 cells were loaded. The separate cell populations were subsequently pooled. The pooled mixture was expected to include Jurkat cells comprising feature barcode BC18, Raji cells comprising feature barcode BC19, and 293T cells comprising feature barcode BC20. The number of cells in the pooled mixture was counted to determine cell numbers. The pooled cell population was then partitioned into single-cell containing droplets for single-cell barcoding as described above. Fully constructed barcode libraries were then sequenced on an Illumina sequencer to detect the presence of the cell and feature barcodes.

[0701] Feature barcode UMI counts were used to group cells after pooling and library preparation. Barcode purity was calculated as (target barcode UMIs)/(sum of all barcode UMIs). Multiplets were identified by high UMI count for more than 1 barcode.

[0702] A summary of the analysis of the sequencing results are presented in Table 3. As seen in Table 3, sequencing reads corresponding to cells containing feature barcodes BC1, BC2, BC3, and BC4 were successfully detected from the pooled cell sample at both the 0.1 uM and 0.01 uM concentration of cholesterol-conjugated feature barcodes tested. The "#background" indicates the number of cells associated with the unlabeled population. Two replicates were performed at each concentration (replicate 1 and replicate 2).

TABLE 3

Sequence Analysis of Pooled Cell Populations

Description	Total cells	# BC18 cells	# BC19 cells	# BC20 cells	# doublets	# background	mean purity BC18 cells	mean purity BC19 cells	mean purity BC20 cells
Cell multiplexing_9000_rep1_3'ver_meta	8595	2866	2338	2800	506	85	0.985	0.99	0.813

TABLE 3-continued

Sequence Analysis of Pooled Cell Populations

Description	Total cells	# BC18 cells	# BC19 cells	# BC20 cells	# doublets	# back-ground	mean purity BC18 cells	mean purity BC19 cells	mean purity BC20 cells
Cell multiplexing_9000_rep2_3'ver_meta	8175	2582	2407	2613	513	60	0.984	0.99	0.822

[0703] FIGS. **72**A-**72**I show graphs from pooled cell populations incubated with antibody-conjugated feature barcodes showing the number of unique molecular identifier (UMI) counts on the x-axis versus number of cells on the y-axis. FIGS. **72**A-**72**B show UMI counts of a first feature barcode sequence ("BC18") identified from sequencing reads generated from sequencing libraries prepared from the pooled cell population (FIG. **72**A—replicate 1; FIG. **72**B—replicate 2). From these results, a clearly distinguished BC18-containing cell population can be distinguished **7201***a* (replicate 1) and **7201***b* (replicate 2). FIGS. **72**C-**72**D show UMI counts of a second feature barcode sequence ("BC19") identified from sequencing reads generated from sequencing libraries prepared from the pooled cell population (FIG. **72**C—replicate 1; FIG. **72**D—replicate 2). From these results, a clearly distinguished BC19-containing cell population can be distinguished **7202***a* (replicate 1) and **7202***b* (replicate 2). FIGS. **72**E-**72**F show UMI counts of a third feature barcode sequence ("BC20") identified from sequencing reads generated from sequencing libraries prepared from the pooled cell population (FIG. **72**E—replicate 1; FIG. **72**F—replicate 2). From these results, a clearly distinguished BC20-containing cell population can be distinguished **7203***a* (replicate 1) and **7203***b* (replicate 2).

[0704] FIGS. **72**G-**72**I show graphs from pooled cell populations incubated with antibody-conjugated feature barcodes showing the number of unique molecular identifier (UMI) counts against populations of various barcode sequences. Cells enriched for one, two (cell doublets), and three (cell triplets) are categorized. FIG. **72**G shows UMI counts of feature barcode sequences identified from sequencing reads generated from sequencing libraries prepared from the pooled cell population with log 10 UMI counts for BC18 on the y-axis and log 10 UMI counts for BC20 on the x-axis. The graph shows clustered UMI counts in which the majority of sequencing reads were found to contain BC18 (**7204**), BC19 (**7205**), BC20 (**7206**), and BC18 and BC20 (**7207**). FIG. **72**H shows UMI counts of feature barcode sequences identified from sequencing reads generated from sequencing libraries prepared from the pooled cell population with log 10 UMI counts for BC18 on the y-axis and log 10 UMI counts for BC19 on the x-axis. The graph shows clustered UMI counts in which the majority of sequencing reads were found to contain BC18 (**7208**), BC19 (**7210**), BC20 (**7209**), and BC18 and BC19 (**7211**). FIG. **72**I shows UMI counts of feature barcode sequences identified from sequencing reads generated from sequencing libraries prepared from the pooled cell population with log 10 UMI counts for BC19 on the y-axis and log 10 UMI counts for BC20 on the x-axis. The graph shows clustered UMI counts in which the majority of sequencing reads were found to contain BC18 (**7213**), BC19 (**7212**), BC20 (**7272**), and BC19 and BC20 (**7215**). Additional UMI counts corresponding to other doublets and to triplets for each of FIGS. **72**G-**72**I are less pronounced in these visualizations.

[0705] Cell types and multiplets are identifiable using feature barcode UMI counts. As shown in FIGS. **73**A-**73**B, doublets identified by antibody UMI counts cluster together in antibody t-distributed stochastic neighbor embedding (t-SNE) (FIG. **73**A), as well as in gene expression (GEX) t-SNE analyses (FIG. **73**B). Clustering is driven by cell type in GEX t-SNE, and by antibody label in antibody t-SNE. Overlap between clusters shows that antibody-based doublet identification matches the expected gene expression profiles. FIG. **73**A shows clusters corresponding to single barcodes BC18, BC19, and BC20 (**7303**, **7302**, **7301**, respectively); doublets including BC18 and BC19 (**7305**), BC18 and BC20 (**7304**), and BC19 and BC20 (**7306**); triplets including BC18, BC19, and BC20 (**7307**); and absence of any barcode (**7308**). FIG. **73**B shows clusters corresponding to single barcodes BC18, BC19, and BC20 (**7313**, **7312**, **7311**, respectively); doublets including BC18 and BC19 (**7373**), BC18 and BC20 (**7314**), and BC19 and BC20 (**7316**); and absence of any barcode (**7318**). A cluster corresponding to triplets including BC18, BC19, and BC20 is not pronounced in FIG. **73**B.

Example XIX: Generating Labeled Polynucleotides

[0706] In this example, and with reference to FIGS. **84**A and **84**B, individual cells are lysed in partitions comprising gel bead emulsions (GEMs). GEMs, for example, can be aqueous droplets comprising gel beads. Within GEMs, a template polynucleotide comprising an mRNA molecule can be reverse transcribed by a reverse transcriptase and a primer comprising a poly(dT) region. A template switching oligo (TSO) present in the GEM, for example a TSO delivered by the gel bead, can facilitate template switching so that a resulting polynucleotide product or cDNA transcript from reverse transcription comprises the primer sequence, a reverse complement of the mRNA molecule sequence, and a sequence complementary to the template switching oligo. The template switching oligo can comprise additional sequence elements, such as a unique molecular identifier (UMI), a barcode sequence (BC), and a Read1 sequence. See FIG. **84**A. In some cases, a plurality of mRNA molecules from the cell is reverse transcribed within the GEM, yielding a plurality of polynucleotide products having various nucleic acid sequences. Following reverse transcription, the polynucleotide product can be subjected to target enrichment in bulk. Prior to target enrichment, the polynucleotide product can be optionally subjected to additional reaction(s) to yield double-stranded polynucleotides. The target may comprise VDJ sequences of a T cell and/or B cell receptor gene sequence. As shown at the top of the right panel of FIG. **84**A, the polynucleotide product (shown as a double-stranded molecule, but can optionally be a single-stranded transcript)

can be subjected to a first target enrichment polymerase chain reaction (PCR) using a primer that hybridizes to the Read 1 region and a second primer that hybridizes to a first region of the constant region (C) of the receptor sequence (e.g., TCR or BCR). The product of the first target enrichment PCR can be subjected to a second, optional target enrichment PCR. In the second target enrichment PCR, a second primer that hybridizes to a second region of the constant region (C) of the receptor can be used. This second primer can, in some cases, hybridize to a region of the constant region that is closer to the VDJ region that the primer used in the first target enrichment PCR. Following the first and second (optional) target enrichment PCR, the resulting polynucleotide product can be further processed to add additional sequences useful for downstream analysis, for example sequencing. The polynucleotide products can be subjected to fragmentation, end repair, A-tailing, adapter ligation, and one or more clean-up/purification operations.

[0707] In some cases, a first subset of the polynucleotide products from cDNA amplification can be subjected to target enrichment (FIG. **84**B, right panel) and a second subset of the polynucleotide products from cDNA amplification is not subjected to target enrichment (FIG. **84**B, bottom left panel). The second subset can be subjected to further processing without enrichment to yield an unenriched, sequencing ready population of polynucleotides. For example, the second subset can be subjected to fragmentation, end repair, A-tailing, adapter ligation, and one or more clean-up/purification operations.

[0708] The labeled polynucleotides can then be subjected to sequencing analysis. Sequencing reads of the enriched polynucleotides can yield sequence information about a particular population of the mRNA molecules in the cell whereas the enriched polynucleotides can yield sequence information about various mRNA molecules in the cell.

Example XX: Multiplexing Immune Samples

[0709] The multiplexing and sample pooling described herein may be applied to the analysis of immune cells (e.g., T cells and B cells) and immune receptors (e.g., TCRs, BCRs, and immunoglobulins). For example, a first cell population of cells comprising immune cells (such as peripheral blood mononuclear cells (PBMCs) or immune cells isolated from PBMCs) are labeled with a plurality of nucleic acid label molecules comprising a first cell barcode sequence and a universal capture sequence. A second cell population of cells comprising immune cells (such as peripheral blood mononuclear cells (PBMCs) or immune cells isolated from PBMCs) are labeled with a plurality of nucleic acid label molecules comprising a second cell barcode sequence and the universal capture sequence. Additional populations of cells (e.g., from additional samples or treatment conditions) can be labeled with additional cell barcode sequences as necessary. Additional labels can also be added to the cells, such as in a "combinatorial tagging" scheme as described elsewhere herein. Further, in some instances, the labels on cell populations can be stabilized through use of one or more anchor oligonucleotides (e.g., attached to a lipophilic moiety) as described herein.

[0710] Labeled cell populations are then pooled and partitioned into a plurality of partitions (e.g., a plurality of aqueous droplets or wells of a microwell array) such that at least some partitions of the plurality of partitions comprise a single labelled cell and a single bead (e.g., a gel bead) comprising a plurality of nucleic acid barcode molecules comprising a common partition barcode sequence and a template switch oligonucleotide (TSO) sequence. The TSO sequence is configured to facilitate a template switching reaction as described herein to generate barcoded molecules comprising a sequence corresponding to an immune transcript (e.g., TCR, BCR, immunoglobulin). In some instances, the TSO sequence is also complementary to and/or capable of hybridizing to the universal capture sequence of the label molecules. In other instances, the nucleic acid barcode molecules comprise (1) a first plurality of nucleic acid barcode molecules comprising (i) a common partition barcode sequence; and (ii) a TSO sequence configured to facilitate a template switching reaction; and (2) a second plurality of nucleic acid barcode molecules comprising (i) the common partition barcode sequence and (ii) a capture sequence complementary to and/or capable of hybridizing to the universal capture sequence of the label molecules. See, e.g., FIG. **83**.

[0711] Subsequent to partitioning, cells are lysed to release mRNA, which is then barcoded, e.g., as described in Example XIX. Nucleic acid label molecules are then hybridized to the partition barcode molecules and a nucleic acid molecule is generated comprising the label barcode and the partition barcode. Barcoded products may then be pooled and subjected to one or more reactions to generate a sequencing library, such as a library suitable for an Illumina sequencer.

[0712] While some embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. It is not intended that the invention be limited by the specific examples provided within the specification. While the invention has been described with reference to the aforementioned specification, the descriptions and illustrations of the embodiments herein are not meant to be construed in a limiting sense. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. Furthermore, it shall be understood that all aspects of the invention are not limited to the specific depictions, configurations or relative proportions set forth herein which depend upon a variety of conditions and variables. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is therefore contemplated that the invention shall also cover any such alternatives, modifications, variations or equivalents. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

SEQUENCE LISTING

```
Sequence total quantity: 64
SEQ ID NO: 1            moltype = DNA  length = 16
```

-continued

```
FEATURE                 Location/Qualifiers
source                  1..16
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 1
aaaaaaaaaa aaaaaa                                                   16

SEQ ID NO: 2            moltype = DNA  length = 11
FEATURE                 Location/Qualifiers
source                  1..11
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 2
aaaaaaaaaa a                                                        11

SEQ ID NO: 3            moltype = DNA  length = 11
FEATURE                 Location/Qualifiers
source                  1..11
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 3
tttttttttt t                                                        11

SEQ ID NO: 4            moltype = DNA  length = 78
FEATURE                 Location/Qualifiers
source                  1..78
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 4
ctacacgacg ctcttccgat ctnnnnnntc nnnnnnnnnn nnnnnntttt tttttttttt  60
tttttttttt ttttttvn                                                 78

SEQ ID NO: 5            moltype = DNA  length = 46
FEATURE                 Location/Qualifiers
source                  1..46
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 5
ctacacgacg ctcttccgat ctnnnnnntc nnnnnnnnnn nnnnnn                 46

SEQ ID NO: 6            moltype = DNA  length = 52
FEATURE                 Location/Qualifiers
source                  1..52
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 6
ctacacgacg ctcttccgat ctnnnnnnnn nnnnnnnntc nnnnnnnnnn nn          52

SEQ ID NO: 7            moltype = DNA  length = 44
FEATURE                 Location/Qualifiers
source                  1..44
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 7
ctacacgacg ctcttccgat ctnnnnnnnn nnnnnntcnn nnnn                   44

SEQ ID NO: 8            moltype = DNA  length = 53
FEATURE                 Location/Qualifiers
source                  1..53
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 8
nnnnnnnnnn nnnnnnnnnn ntttttttttt tttttttttt tttttttttt tvn         53

SEQ ID NO: 9            moltype = DNA  length = 52
FEATURE                 Location/Qualifiers
source                  1..52
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 9
nnnnnnnnnn nnnnnnnnnn tttttttttt tttttttttt tttttttttt vn          52

SEQ ID NO: 10           moltype = DNA  length = 32
FEATURE                 Location/Qualifiers
source                  1..32
                        mol_type = other DNA
```

-continued

```
                        organism = synthetic construct
SEQUENCE: 10
tttttttttt tttttttttt tttttttttt vn                                32

SEQ ID NO: 11           moltype = DNA  length = 114
FEATURE                 Location/Qualifiers
source                  1..114
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 11
caagcagaag acggcatacg agatnnnnnn gtnnnnnngt gactggagtt cagacgtgtg  60
ctcttccgat ctnnnnnnnn nnttttttttt tttttttttt tttttttttt ttvn        114

SEQ ID NO: 12           moltype = DNA  length = 114
FEATURE                 Location/Qualifiers
source                  1..114
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 12
caagcagaag acggcatacg agatnnnnnn cannnnnngt gactggagtt cagacgtgtg  60
ctcttccgat ctnnnnnnnn nnttttttttt tttttttttt tttttttttt ttvn        114

SEQ ID NO: 13           moltype = DNA  length = 114
FEATURE                 Location/Qualifiers
source                  1..114
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 13
caagcagaag acggcatacg agatnnnnnn agnnnnnngt gactggagtt cagacgtgtg  60
ctcttccgat ctnnnnnnnn nnttttttttt tttttttttt tttttttttt ttvn        114

SEQ ID NO: 14           moltype = DNA  length = 114
FEATURE                 Location/Qualifiers
source                  1..114
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 14
caagcagaag acggcatacg agatnnnnnn tcnnnnnngt gactggagtt cagacgtgtg  60
ctcttccgat ctnnnnnnnn nnttttttttt tttttttttt tttttttttt ttvn        114

SEQ ID NO: 15           moltype = DNA  length = 78
FEATURE                 Location/Qualifiers
source                  1..78
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 15
ctacacgacg ctcttccgat ctnnnnnngt nnnnnnnnnn nnnnnnttttt tttttttttt  60
tttttttttt tttttvn                                                  78

SEQ ID NO: 16           moltype = DNA  length = 78
FEATURE                 Location/Qualifiers
source                  1..78
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 16
ctacacgacg ctcttccgat ctnnnnnnca nnnnnnnnnn nnnnnnttttt tttttttttt  60
tttttttttt tttttvn                                                  78

SEQ ID NO: 17           moltype = DNA  length = 78
FEATURE                 Location/Qualifiers
source                  1..78
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 17
ctacacgacg ctcttccgat ctnnnnnnag nnnnnnnnnn nnnnnnttttt tttttttttt  60
tttttttttt tttttvn                                                  78

SEQ ID NO: 18           moltype = DNA  length = 118
FEATURE                 Location/Qualifiers
source                  1..118
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 18
aatgatacgg cgaccaccga gatctacacn nnnnngtnnn nnnacactct ttccctacac  60
gacgctcttc cgatctnnnn nnnnnnttttt tttttttttt tttttttttt tttttvn    118
```

-continued

```
SEQ ID NO: 19          moltype = DNA  length = 78
FEATURE                Location/Qualifiers
source                 1..78
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 19
cagacgtgtg ctcttccgat ctnnnnnngt nnnnnnnnnn nnnnnntttt tttttttttt  60
tttttttttt ttttttvn                                                78

SEQ ID NO: 20          moltype = DNA  length = 52
FEATURE                Location/Qualifiers
source                 1..52
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 20
ctacacgacg ctcttccgat ctnnnnnnnn nnnnnnnngt nnnnnnnnnn nn          52

SEQ ID NO: 21          moltype = DNA  length = 52
FEATURE                Location/Qualifiers
source                 1..52
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 21
ctacacgacg ctcttccgat ctnnnnnnnn nnnnnnnnca nnnnnnnnnn nn          52

SEQ ID NO: 22          moltype = DNA  length = 52
FEATURE                Location/Qualifiers
source                 1..52
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 22
ctacacgacg ctcttccgat ctnnnnnnnn nnnnnnnnag nnnnnnnnnn nn          52

SEQ ID NO: 23          moltype = DNA  length = 21
FEATURE                Location/Qualifiers
source                 1..21
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 23
aaaaaaaaaa aaaaaaaaaa a                                            21

SEQ ID NO: 24          moltype = DNA  length = 15
FEATURE                Location/Qualifiers
source                 1..15
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 24
ctacacgacg ctctt                                                   15

SEQ ID NO: 25          moltype = DNA  length = 28
FEATURE                Location/Qualifiers
source                 1..28
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 25
ctacacgacg ctcttccgat ctnnnnnn                                     28

SEQ ID NO: 26          moltype = DNA  length = 22
FEATURE                Location/Qualifiers
source                 1..22
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 26
acnnnnnnag atcggaagag cg                                           22

SEQ ID NO: 27          moltype = DNA  length = 50
FEATURE                Location/Qualifiers
source                 1..50
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 27
gtnnnnnnnn nnnnnnnntt tttttttttt tttttttttt tttttttvn              50

SEQ ID NO: 28          moltype = DNA  length = 41
FEATURE                Location/Qualifiers
source                 1..41
```

-continued

```
                      mol_type = other DNA
                      organism = synthetic construct
SEQUENCE: 28
gtnnnnnnnn nnnnnnnnta cgctagtttc gcgtacgaag c                      41

SEQ ID NO: 29         moltype = DNA  length = 69
FEATURE               Location/Qualifiers
source                1..69
                      mol_type = other DNA
                      organism = synthetic construct
SEQUENCE: 29
ctacacgacg ctcttccgat ctnnnnnngt nnnnnnnnnn nnnnnntacg ctagtttcgc  60
gtacgaagc                                                          69

SEQ ID NO: 30         moltype = DNA  length = 80
FEATURE               Location/Qualifiers
source                1..80
                      mol_type = other DNA
                      organism = synthetic construct
SEQUENCE: 30
ctacacgacg ctcttccgat ctnnnnnnnn nnnnnnnnnn nnnnnnnntt tttttttttt  60
tttttttttt tttttttvn                                               80

SEQ ID NO: 31         moltype = DNA  length = 70
FEATURE               Location/Qualifiers
source                1..70
                      mol_type = other DNA
                      organism = synthetic construct
SEQUENCE: 31
ctacacgacg ctcttccgat ctnnnnnnnn nnnnnnnnnn nnnnnnnnac gctagtttcg  60
cgtacgaagc                                                         70

SEQ ID NO: 32         moltype = DNA  length = 15
FEATURE               Location/Qualifiers
source                1..15
                      mol_type = other DNA
                      organism = synthetic construct
SEQUENCE: 32
gtcagatgtg tataa                                                   15

SEQ ID NO: 33         moltype = DNA  length = 80
FEATURE               Location/Qualifiers
source                1..80
                      mol_type = other DNA
                      organism = synthetic construct
SEQUENCE: 33
ctacacgacg ctcttccgat ctnnnnnnng tnnnnnnnnn nnnnnnnntt tttttttttt  60
tttttttttt tttttttvn                                               80

SEQ ID NO: 34         moltype = DNA  length = 80
FEATURE               Location/Qualifiers
source                1..80
                      mol_type = other DNA
                      organism = synthetic construct
SEQUENCE: 34
ctacacgacg ctcttccgat ctnnnnnnnc annnnnnnnn nnnnnnnntt tttttttttt  60
tttttttttt tttttttvn                                               80

SEQ ID NO: 35         moltype = DNA  length = 80
FEATURE               Location/Qualifiers
source                1..80
                      mol_type = other DNA
                      organism = synthetic construct
SEQUENCE: 35
ctacacgacg ctcttccgat ctnnnnnnna gnnnnnnnnn nnnnnnnntt tttttttttt  60
tttttttttt tttttttvn                                               80

SEQ ID NO: 36         moltype = DNA  length = 80
FEATURE               Location/Qualifiers
source                1..80
                      mol_type = other DNA
                      organism = synthetic construct
SEQUENCE: 36
ctacacgacg ctcttccgat ctnnnnnnnt cnnnnnnnnn nnnnnnnntt tttttttttt  60
tttttttttt tttttttvn                                               80
```

-continued

```
SEQ ID NO: 37          moltype = DNA  length = 70
FEATURE                Location/Qualifiers
source                 1..70
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 37
gtcagatgtg tataagagac agnnnnnnng tnnnnnnnnn nnnnnnnngc ttcgtacgcg  60
aaactagcgt                                                         70

SEQ ID NO: 38          moltype = DNA  length = 60
FEATURE                Location/Qualifiers
source                 1..60
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 38
ctacacgacg ctcttccgat ctnnnnnnnn nnnnnnnnnn nnnnnnnttt cttatatggg  60

SEQ ID NO: 39          moltype = DNA  length = 43
FEATURE                Location/Qualifiers
source                 1..43
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 39
caactttagc ggtccaaggt gcagtcagat cccatataag aaa                    43

SEQ ID NO: 40          moltype = AA  length = 6
FEATURE                Location/Qualifiers
source                 1..6
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 40
HHHHHH                                                             6

SEQ ID NO: 41          moltype = DNA  length = 61
FEATURE                Location/Qualifiers
source                 1..61
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 41
agatcggaag agcacacgtc tgaactccag tcacnnnnnn acnnnnnnat ctcgtatgcc  60
g                                                                  61

SEQ ID NO: 42          moltype = DNA  length = 61
FEATURE                Location/Qualifiers
source                 1..61
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 42
agatcggaag agcacacgtc tgaactccag tcacnnnnnn tgnnnnnnat ctcgtatgcc  60
g                                                                  61

SEQ ID NO: 43          moltype = DNA  length = 61
FEATURE                Location/Qualifiers
source                 1..61
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 43
agatcggaag agcacacgtc tgaactccag tcacnnnnnn ctnnnnnnat ctcgtatgcc  60
g                                                                  61

SEQ ID NO: 44          moltype = DNA  length = 61
FEATURE                Location/Qualifiers
source                 1..61
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 44
agatcggaag agcacacgtc tgaactccag tcacnnnnnn gannnnnnat ctcgtatgcc  60
g                                                                  61

SEQ ID NO: 45          moltype = DNA  length = 28
FEATURE                Location/Qualifiers
source                 1..28
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 45
nnnnnnacnn nnnnagatcg gaagagcg                                     28
```

-continued

```
SEQ ID NO: 46          moltype = DNA  length = 28
FEATURE                Location/Qualifiers
source                 1..28
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 46
nnnnnntgnn nnnnagatcg gaagagcg                                    28

SEQ ID NO: 47          moltype = DNA  length = 28
FEATURE                Location/Qualifiers
source                 1..28
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 47
nnnnnnctnn nnnnagatcg gaagagcg                                    28

SEQ ID NO: 48          moltype = DNA  length = 28
FEATURE                Location/Qualifiers
source                 1..28
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 48
nnnnnngann nnnnagatcg gaagagcg                                    28

SEQ ID NO: 49          moltype = DNA  length = 62
FEATURE                Location/Qualifiers
source                 1..62
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 49
agatcggaag agcgtcgtgt agggaaagag tgtnnnnnna cnnnnnngtg tagatctcgg  60
tg                                                                 62

SEQ ID NO: 50          moltype = DNA  length = 27
FEATURE                Location/Qualifiers
source                 1..27
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 50
nnnnnnacnn nnnnagatcg gaagagc                                     27

SEQ ID NO: 51          moltype = DNA  length = 14
FEATURE                Location/Qualifiers
source                 1..14
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 51
agatcggaag agcg                                                   14

SEQ ID NO: 52          moltype = DNA  length = 30
FEATURE                Location/Qualifiers
source                 1..30
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 52
nnnnnnnacn nnnnnnagat cggaagagcg                                  30

SEQ ID NO: 53          moltype = DNA  length = 30
FEATURE                Location/Qualifiers
source                 1..30
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 53
nnnnnnntgn nnnnnnagat cggaagagcg                                  30

SEQ ID NO: 54          moltype = DNA  length = 30
FEATURE                Location/Qualifiers
source                 1..30
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 54
nnnnnnnctn nnnnnnagat cggaagagcg                                  30

SEQ ID NO: 55          moltype = DNA  length = 30
FEATURE                Location/Qualifiers
```

-continued

```
source                  1..30
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 55
nnnnnnngan nnnnnnagat cggaagagcg                                  30

SEQ ID NO: 56           moltype = DNA  length = 31
FEATURE                 Location/Qualifiers
source                  1..31
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 56
nnnnnnnacn nnnnnnctgt ctcttataca c                                31

SEQ ID NO: 57           moltype = DNA  length = 24
FEATURE                 Location/Qualifiers
source                  1..24
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 57
caactttagc ggtccaaggt gcat                                        24

SEQ ID NO: 58           moltype = DNA  length = 88
FEATURE                 Location/Qualifiers
source                  1..88
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 58
caccttggac cgctaaagtt ggtgactgga gttcagacgt gtgctcttcc gatctnnnnn  60
nnnnncagat ttgaccccat ataagaaa                                    88

SEQ ID NO: 59           moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 59
aaaaaaaaaa aaaaaaaaaa                                             20

SEQ ID NO: 60           moltype = DNA  length = 10
FEATURE                 Location/Qualifiers
source                  1..10
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 60
atcctagcaa                                                        10

SEQ ID NO: 61           moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 61
tgatcctagc aa                                                     12

SEQ ID NO: 62           moltype = AA  length = 9
FEATURE                 Location/Qualifiers
source                  1..9
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 62
EGALIYWPN                                                         9

SEQ ID NO: 63           moltype = AA  length = 8
FEATURE                 Location/Qualifiers
source                  1..8
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 63
AHMRDSQQ                                                          8

SEQ ID NO: 64           moltype = AA  length = 9
FEATURE                 Location/Qualifiers
source                  1..9
                        mol_type = protein
                        organism = synthetic construct
```

-continued

SEQUENCE: 64	
GLCTLVAML	9

1-20. (canceled)

21. A composition comprising:

a partition comprising a plurality of barcode molecules, wherein the plurality of barcode molecules comprises:

(a) a first barcode molecule comprising a nucleic acid barcode sequence, wherein the first barcode molecule is hybridized to a first analyte, wherein the first analyte is not messenger ribonucleic acid (mRNA) and is not hybridized to the first barcode molecule via a polythymine (poly-T) sequence;

(b) a second barcode molecule comprising the nucleic acid barcode sequence, wherein the second barcode molecule is hybridized to a second analyte, wherein the second analyte is mRNA and is hybridized to the second barcode molecule via a poly-T sequence.

22. The composition of claim **21**, wherein the first analyte comprises clustered regularly interspaced short palindromic repeat (CRISPR) ribonucleic acid (crRNA) or single guide ribonucleic acid (sgRNA).

23. The composition of claim **21**, wherein the first analyte comprises a deoxyribonucleic acid (DNA) oligonucleotide conjugated to an antibody.

24. The composition of claim **23**, wherein the DNA oligonucleotide comprises a barcode sequence.

25. The composition of claim **21**, wherein the plurality of barcode molecules are attached to a bead.

26. The composition of claim **25**, wherein the bead is a gel bead.

27. The composition of claim **25**, wherein the bead comprises polyacrylamide.

28. The composition of claim **25**, wherein the plurality of barcode molecules are releasably attached to the bead.

29. The composition of claim **21**, wherein the partition is a partition among a plurality of partitions.

30. The composition of claim **29**, wherein the plurality of partitions comprises at least about 1,000; 5,000; 10,000; 50,000; 100,000; 500,000; 1,000,000; 5,000,000; 10,000,000; 50,000,000; 100,000,000; 500,000,000; or 1,000,000,000 partitions.

31. The composition of claim **29**, wherein the partition is a droplet, and the plurality of partitions is a plurality of droplets.

32. The composition of claim **29**, wherein the partition is a well, and the plurality of partitions is a plurality of wells.

33. The composition of claim **21**, wherein the first barcode molecule and the second barcode molecule further comprise a random N-mer sequence.

34. A composition comprising:

a plurality of droplets, wherein a droplet of the plurality of droplets comprises: a gel bead having a plurality of barcode molecules attached thereto,

wherein the plurality of barcode molecules comprises:

(a) a first barcode molecule comprising a nucleic acid barcode sequence, wherein the first barcode molecule is hybridized to a first analyte, wherein the first analyte is not mRNA and is not hybridized to the first barcode molecule via a poly-T sequence;

(b) a second barcode molecule comprising the nucleic acid barcode sequence, wherein the second barcode molecule is hybridized to a second analyte, wherein the second analyte is mRNA and is hybridized to the second barcode molecule via a poly-T sequence,

wherein the first barcode molecule and the second molecule further comprise a random N-mer sequence.

35. The composition of claim **34**, wherein the first analyte comprises a clustered regularly interspaced short palindromic repeat (CRISPR) ribonucleic acid (crRNA) or a single guide ribonucleic acid (sgRNA).

36. The composition of claim **34**, wherein the first analyte comprises a DNA oligonucleotide conjugated to an antibody, wherein the DNA oligonucleotide comprises a barcode sequence.

37. A composition comprising:

a partition comprising a bead having a plurality of barcode molecules attached thereto, wherein the plurality of barcode molecules comprises:

(a) a first population of barcode molecules each hybridized to mRNA via poly-T sequences, wherein the first population of barcode molecules each comprise a barcode sequence and a random sequence;

(b) a second population of barcode molecules each hybridized to an analyte comprising crRNA or sgRNA via a sequence that is not poly-T, wherein the second population of barcode molecules each comprise a barcode sequence and a random sequence;

wherein the barcode sequences are constant among the plurality of barcode molecules and the random sequences vary among the plurality of barcode molecules.

38. The composition of claim **37**, wherein the bead is a gel bead.

39. The composition of claim **37**, wherein the bead comprises polyacrylamide.

40. The composition of claim **37**, wherein the partition is a partition among a plurality of partitions.

41. The composition of claim **40**, wherein the plurality of partitions comprises at least about 1,000; 5,000; 10,000; 50,000; 100,000; 500,000; 1,000,000; 5,000,000; 10,000,000; 50,000,000; 100,000,000; 500,000,000; or 1,000,000,000 partitions.

42. The composition of claim **40**, wherein the partition is a droplet, and the plurality of partitions is a plurality of droplets.

43. The composition of claim **40**, wherein the partition is a well, and the plurality of partitions is a plurality of wells.

44. A composition comprising:

a partition comprising a bead having a plurality of barcode molecules attached thereto, wherein the plurality of barcode molecules comprises:

(a) a first population of barcode molecules each hybridized to mRNA via a poly-T sequence, wherein the first population of barcode molecules each comprise a barcode sequence and a random sequence;

(b) a second population of barcode molecules each hybridized to a DNA-antibody conjugate via a

sequence that is not poly-T, wherein the second population of barcode molecules each comprise a barcode sequence and a random sequence;

wherein the barcode sequences are constant among the plurality of barcode molecules and the random sequences vary among the plurality of barcode molecules.

45. The composition of claim **44**, wherein the bead is a gel bead.

46. The composition of claim **44**, wherein the bead comprises polyacrylamide.

47. The composition of claim **44**, wherein the partition is a partition among a plurality of partitions.

48. The composition of claim **47**, wherein the plurality of partitions comprises at least about 1,000; 5,000; 10,000; 50,000; 100,000; 500,000; 1,000,000; 5,000,000; 10,000,000; 50,000,000; 100,000,000; 500,000,000; or 1,000,000,000 partitions.

49. The composition of claim **47**, wherein the partition is a droplet, and the plurality of partitions is a plurality of droplets.

50. The composition of claim **47**, wherein the partition is a well, and the plurality of partitions is a plurality of wells.

* * * * *