



(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2017/030097

(22) International Filing Date:
28 April 2017 (28.04.2017)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/330,500 02 May 2016 (02.05.2016) US

(71) Applicant: **CELLULAR RESEARCH, INC.** [US/US];
4040 Campbell Avenue, Suite 110, Menlo Park, CA 94025
(US).

(72) Inventors: **SHUM, Eleen**; 4040 Campbell Avenue, Suite
110, Menlo Park, CA 94025 (US). **FU, Glenn**; 4040 Camp-
bell Avenue, Suite 110, Menlo Park, CA 94025 (US).

(74) Agent: **MALLON, Joseph, J.**; Knobbe, Martens, Olson &
Bear, Llp, 2040 Main Street, 14th Floor, Irvine, CA 92614
(US).

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

(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ,
UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ,

TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK,
EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,
MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,
TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,
KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

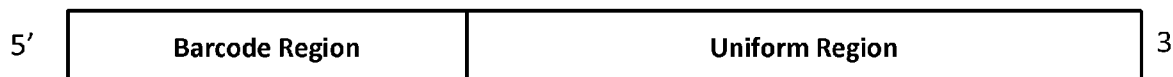
- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))
- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: ACCURATE MOLECULAR BARCODING

FIG. 1A



(57) Abstract: In accordance with some embodiments herein, compositions and methods for accurate barcoding of nucleic acids are described. The compositions and methods involve a plurality of unique oligonucleotide species comprising unique molecule barcodes. The molecular barcodes have a relatively low G (guanine) content, e.g. each of the barcodes contain less than 50% guanine or the overall set of barcodes contain in average less than 12.5% guanine. The set of barcodes is intended to exhibit reduced bias in amplification and analysis.



ACCURATE MOLECULAR BARCODING

REFERENCE TO RELATED APPLICATIONS

[0001] The present application claims the benefit of U.S. Provisional App. No. 62/330,500 filed May 2, 2016, which is hereby incorporated by reference in its entirety.

REFERENCE TO SEQUENCE LISTING

[0002] The present application is being filed along with a Sequence Listing in electronic format. The Sequence Listing is provided as a file entitled SEQUENCEBDCRI019WO.TXT, created and last modified April 18, 2017, which is 32,588 bytes in size. The information in the electronic format of the Sequence Listing is incorporated herein by reference in its entirety.

FIELD

[0003] Embodiments herein relate generally to compositions and methods for accurate barcoding of molecules, for example nucleic acid molecules.

SUMMARY

[0004] Some embodiments include a composition comprising at least 1000 unique oligonucleotide species is provided. Each unique oligonucleotide species can comprise a barcode region comprising a molecule barcode comprising at least 7 nucleotides, in which the unique oligonucleotide species comprise different nucleic acid sequences in their barcode regions, and in which at least one of: (a) the composition consists essentially of unique oligonucleotide species wherein each molecular barcode has a G content of less than 50%; or (b) the molecule barcodes of all of the unique oligonucleotide species in the composition collectively have a G content of no more than 12.5%. In some embodiments, the barcode region further comprises a sample barcode comprising at least 3 nucleotides. In some embodiments, each unique oligonucleotide species further comprises a uniform region 3' of the barcode region. In some embodiments, the uniform region further comprises a target-specific region 3' of the barcode region, the target-specific region comprising at least 10

nucleotides complementary to a target nucleic acid. In some embodiments, the composition consists essentially of unique oligonucleotide species wherein each molecular barcode has a G content of less than 50%. In some embodiments, the molecule barcodes of all of the unique oligonucleotide species in the composition collectively have a G content of no more than 12.5%. In some embodiments, the unique oligonucleotide species are disposed in at least two spatially isolated pools, each pool comprising at least 100 unique oligonucleotides of the unique oligonucleotide species, wherein unique oligonucleotides in the same pool comprise the same sample barcode sequence, and wherein different unique oligonucleotides of the same pool comprise a different molecule barcode sequences. In some embodiments, the sample barcode of each unique oligonucleotide species has a G content of 50% or less. In some embodiments, the barcode region of each unique oligonucleotide species has a G content of 50% or less. In some embodiments, the molecule barcodes of the unique oligonucleotide species collectively have a G content of less than 12.5%. In some embodiments, the barcode regions of the unique oligonucleotide species collectively have a G content of no more than 12.5%. In some embodiments, for at least 95% of the unique oligonucleotide species, any G in the molecule barcode is not adjacent to another G. In some embodiments, the composition consists essentially of unique oligonucleotide species for which any G in the molecule barcode is not adjacent to another G. In some embodiments, at least 95% of the molecule barcodes of the unique oligonucleotide species comprise a sequence totaling at least 6 alternating H's and N's, wherein each "H" is any one of A, C, or T, and wherein each "N" is any one of A, G, C, or T. In some embodiments, least 95% of the molecule barcodes of the unique oligonucleotide species comprise the sequence HNHNNHN, wherein each "H" is any one of A, C, or T, and wherein each "N" is any one of A, G, C, or T. In some embodiments, each molecule barcode of the unique oligonucleotide species comprises the sequence HNHNNHN, wherein each "H" is any one of A, C, or T, and wherein each "N" is any one of A, G, C, or T. In some embodiments, each molecule barcode of the unique oligonucleotide species comprises the sequence HHHHHHH, wherein each "H" is any one of A, C, or T. In some embodiments, each of the unique oligonucleotide species comprises a spacer 3' of the barcode region and 5' of the target specific region, said spacer comprising the sequence HNHNNHN, wherein each "H"

is any one of A, C, or T, and wherein each “N” is any one of A, G, C, or T. In some embodiments, each of the unique oligonucleotide species comprises a spacer 3’ of the barcode region and 5’ of the target specific region, said spacer comprising the sequence HHHHHHHH, wherein each “H” is any one of A, C, or T. In some embodiments, the target specific region comprises an oligo dT sequence. In some embodiments, for each unique oligonucleotide species, the molecule barcode is 3’ of the sample barcode. In some embodiments, for each unique oligonucleotide species, the sample barcode is 3’ of the molecule barcode. In some embodiments, each oligonucleotide species has a length of at least 24 nucleotides. In some embodiments, each oligonucleotide species has a length of 24-140 nucleotides. In some embodiments, the composition comprises at least 6,500 unique oligonucleotide species. In some embodiments, the composition comprises at least 65,000 unique oligonucleotide species. In some embodiments, the composition comprises at least two oligonucleotides of the same unique oligonucleotide species. In some embodiments, no two oligonucleotides of the composition are of the same unique oligonucleotide species. In some embodiments, the composition comprises at least 48 pools. In some embodiments, the unique oligonucleotide species of each pool are immobilized on a substrate, so that the sample barcodes but not the molecule barcodes are the same for the oligonucleotide species immobilized on each substrate. In some embodiments, the substrate comprises a discrete region of a surface, so that the surface can comprise two or more substrates. In some embodiments, the substrate comprises a bead. In some embodiments, each of the unique oligonucleotide species further comprises an adapter configured to immobilize the unique oligonucleotide on the substrate, wherein said barcode region is 3’ of the adapter. In some embodiments, the uniform region comprises a target-specific region comprising a sequence flanking an immune cell receptor or immunoglobulin variable region coding sequence. In some embodiments, the immune cell receptor variable region coding sequence is a T cell receptor variable region coding sequence, a B cell receptor variable region coding sequence, or a combination thereof. In some embodiments, a kit comprising the composition comprising at least 1000 unique oligonucleotide species is provided. The kit can further comprise a primer configured to hybridize on an opposite side of the variable region as the target-specific region and to hybridize to a complementary strand to a strand hybridized by

the target-specific region, so that the primer is configured to amplify the variable region in conjunction with the target-specific region. In some embodiments, the primer and target-specific region are configured to amplify a nucleic acid of at least 1 kb and comprising the variable region. In some embodiments, the primer of the kit is part of the composition comprising the unique oligonucleotide species. In some embodiments, the primer of the kit is part of an other composition that is separate from the composition comprising the unique oligonucleotide species. [0004] Some embodiments include a method of specifically barcoding a plurality of nucleic acids from two or more samples is provided. Each sample can comprising nucleic acid. The method can comprise contacting each sample with a pool comprising at least 100 unique oligonucleotide species, in which each sample is contacted in spatial isolation from the other samples. Each unique oligonucleotide species can comprise a barcode region comprising a molecule barcode comprising at least 7 nucleotides. The unique polynucleotide species of each pool can comprise the same sample barcode, and comprise different molecule barcodes. In the method, at least (a) the unique oligonucleotide species contacted with the sample consist essentially of unique oligonucleotide species wherein each molecular barcode has a G content of less than 50%; or (b) the molecule barcodes of all of the unique oligonucleotide species collectively have a G content of no more than 12.5% can apply. The method can include hybridizing target-specific regions of at least some oligonucleotides of the unique oligonucleotide species to at least some of the nucleic acids of the sample. The method can include extending the hybridized oligonucleotides, thereby producing strands comprising an oligonucleotide of the unique oligonucleotide species and a sequence complementary to the target, wherein for each sample, the strands comprise the same sample barcode and different molecule barcodes, and wherein for different samples, the molecule barcodes are different. In some embodiments, the barcode region further comprises a sample barcode comprising at least 3 nucleotides. In some embodiments, each unique oligonucleotide species further comprises a uniform region 3' of the barcode region. In some embodiments, the uniform region further comprises a target-specific region 3' of the barcode region, the target-specific region comprising at least 10 nucleotides complementary to a target nucleic acid. In some embodiments, the unique oligonucleotide species contacted with the sample consist essentially of unique oligonucleotide species wherein each molecular barcode

has a G content of less than 50%. In some embodiments, the molecule barcodes of all of the unique oligonucleotide species collectively have a G content of no more than 12.5%. In some embodiments, the method further comprises ascertaining nucleic acid sequences of the strands comprising the oligonucleotides of the unique oligonucleotide species and the sequence complementary to the target. In some embodiments, the at least 100 unique oligonucleotide species of each pool are immobilized on a substrate, so that the unique oligonucleotide species immobilized on a given substrate comprise the same sample barcode, and different unique oligonucleotide species immobilized on the substrate comprise different molecule barcodes. In some embodiments, each sample barcode has a G content of 50% or less. In some embodiments, the molecule barcodes of the unique oligonucleotide species collectively have a G content of less than 12.5%. In some embodiments, the barcode regions of the unique oligonucleotide species collectively have a G content of no more than 12.5%. In some embodiments, for at least 95% of the unique oligonucleotide species, any G in the molecule barcode is not adjacent to another G. In some embodiments, each pool consists essentially of unique oligonucleotide species for which any G in the molecule barcode is not adjacent to another G. In some embodiments, at least 95% of the molecule barcodes of the unique oligonucleotide species comprise a sequence totaling at least 6 alternating H's and N's, wherein each "H" is any one of A, C, or T, and wherein each "N" is any one of A, G, C, or T. In some embodiments, at least 95% of the molecule barcodes of the unique oligonucleotide species comprise the sequence HNHNNHNH, wherein each "H" is any one of A, C, or T, and wherein each "N" is any one of A, G, C, or T. In some embodiments, each molecule barcode of the unique oligonucleotide species comprises the sequence HNHNNHNH, wherein each "H" is any one of A, C, or T, and wherein each "N" is any one of A, G, C, or T. In some embodiments, each molecule barcode of the unique oligonucleotides comprises the sequence HHHHHHHH, wherein each "H" is any one of A, C, or T. In some embodiments, each unique oligonucleotide comprises a spacer 3' of the barcode region and 5' of the target specific region, said spacer comprising the sequence HNHNNHNH, wherein each "H" is any one of A, C, or T, and wherein each "N" is any one of A, G, C, or T. In some embodiments, each unique oligonucleotide comprises a spacer 3' of the barcode region and 5' of the target specific region, said spacer comprising the sequence

HHHHHHHH, wherein each "H" is any one of A, C, or T. In some embodiments, at least one pool comprises least two oligonucleotides of the same unique oligonucleotide species. In some embodiments, no pool comprises two oligonucleotides of the same unique oligonucleotide species. In some embodiments, the target specific region comprises an oligo dT sequence. In some embodiments, for each unique oligonucleotide species, the molecule barcode is 3' of the sample barcode. In some embodiments, for each unique oligonucleotide species, the sample barcode is 3' of the molecule barcode. In some embodiments, each unique oligonucleotide species has a length of at least 24 nucleotides. In some embodiments, each unique oligonucleotide species has a length of 24-140 nucleotides. In some embodiments, each pool comprises at least 6,500 unique oligonucleotide species. In some embodiments, each pool comprises at least 65,000 unique oligonucleotide species. In some embodiments, at least 48 unique samples are each contacted with a unique pool. In some embodiments, at least 99% of the samples comprise no more than one cell each. In some embodiments, the unique oligonucleotide species of each pool are immobilized on a substrate, so that the sample barcodes but not the molecule barcodes are the same for the unique oligonucleotide species immobilized on each substrate of the plurality. In some embodiments, the substrate comprises a spatially-isolated region of a surface, so that the substrates of different pools comprise the different spatially-isolated regions of the surface. In some embodiments, the substrate comprises a bead. In some embodiments, each of the unique oligonucleotide species further comprises an adapter configured to immobilize the unique oligonucleotide on the substrate, wherein said barcode region is 3' of the adapter. In some embodiments, the uniform region comprises a target-specific region comprising a sequence flanking a sequence encoding the variable region of an immune cell receptor or immunoglobulin. In some embodiments, the variable region is of a T cell receptor or B cell receptor, or a combination thereof. In some embodiments, the method further comprises contacting the extended strands comprising an oligonucleotide of the unique oligonucleotide species and a sequence complementary to the target with primer configured to hybridize on an opposite side of the variable region as the target-specific region, and to hybridize to a complementary strand to a strand hybridized by the target-specific region. As such, the method can comprise amplifying sequences encoding variable regions of a T cell receptor, B

cell receptor, or immunoglobulin. In some embodiments, the method amplifies a sequence of at least 1 kb, which comprises the variable region coding sequence.

[0005] Some embodiments include a method of making a composition comprising unique oligonucleotides is provided. The method can comprise providing a plurality of different sample barcodes comprising at least 3 nucleotides each. The method can comprise providing a plurality of different molecule barcodes comprising at least 7 nucleotides each. The methods can comprise synthesizing a plurality of unique oligonucleotide species, each unique oligonucleotide species comprising a barcode region comprising a sample barcode and a molecule barcode, in which at least one of: (a) the plurality of unique oligonucleotide species consists essentially of unique oligonucleotide species wherein each molecular barcode has a G content of less than 50%; or (b) the molecule barcodes of all of the unique oligonucleotide species in the plurality collectively have a G content of no more than 12.5%. The method can comprise disposing the unique oligonucleotide species in spatially-isolated pools, in which each pool comprises multiple unique oligonucleotide species, so that the unique oligonucleotide species of the same pool comprise the same sample barcode sequence, and wherein different unique oligonucleotide species of the same pool comprise different molecule barcode sequences, and in which each pool comprises at least 1000 unique oligonucleotide species. In some embodiments, each unique oligonucleotide species further comprises a uniform region 3' of the barcode region. In some embodiments, the uniform region further comprises a target-specific region 3' of the barcode region, the target-specific region comprising at least 10 nucleotides complementary to a target nucleic acid. In some embodiments, the plurality of unique oligonucleotide species consists essentially of unique oligonucleotide species wherein each molecular barcode has a G content of less than 50%. In some embodiments, the molecule barcodes of all of the unique oligonucleotide species in the plurality collectively have a G content of no more than 12.5%. In some embodiments, each molecule barcode has a G content of 50% or less. In some embodiments, each sample barcode has a G content of 50% or less. In some embodiments, the sample barcodes of the unique oligonucleotide species collectively have a G content of no more than 12.5%. In some embodiments, the molecule barcodes of the unique oligonucleotide species collectively have a G content of less than 12.5%. In some embodiments, the barcode regions of the

unique oligonucleotide species collectively have a G content of no more than 12.5%. In some embodiments, for at least 95% of the unique oligonucleotide species, any G in the molecule barcode is not adjacent to another G. In some embodiments, the plurality of the unique oligonucleotide species consists essentially of unique oligonucleotide species for which any G in the molecule barcode is not adjacent to another G. In some embodiments, at least 95% of the molecule barcodes of the unique oligonucleotide species comprise a sequence totaling at least 6 alternating H's and N's, wherein each "H" is any one of A, C, or T, and wherein each "N" is any one of A, G, C, or T. In some embodiments, at least 95% of the molecule barcodes of the unique oligonucleotides comprise the sequence HNHNNHNH, wherein each "H" is any one of A, C, or T, and wherein each "N" is any one of A, G, C, or T. In some embodiments, each molecule barcode of the unique oligonucleotides comprises the sequence HNHNNHNH, wherein each "H" is any one of A, C, or T, and wherein each "N" is any one of A, G, C, or T. In some embodiments, each molecule barcode of the unique oligonucleotides comprises the sequence HHHHHHHH, wherein each "H" is any one of A, C, or T. In some embodiments, each unique oligonucleotide species comprises a spacer 3' of the barcode region and 5' of the target specific region, said spacer comprising the sequence HNHNNHNH, wherein each "H" is any one of A, C, or T, and wherein each "N" is any one of A, G, C, or T. In some embodiments, each unique oligonucleotide species comprises a spacer 3' of the barcode region and 5' of the target specific region, said spacer comprising the sequence HHHHHHHH, wherein each "H" is any one of A, C, or T. In some embodiments, the target specific region comprises an oligo dT sequence. In some embodiments, for each unique oligonucleotide species, the molecule barcode is 3' of the sample barcode. In some embodiments, each unique oligonucleotide species, the sample barcode is 3' of the molecule barcode. In some embodiments, each unique oligonucleotide species has a length of at least 24 nucleotides. In some embodiments, each unique oligonucleotide species has a length of 24-140 nucleotides. In some embodiments, each pool comprises at least 6,500 unique oligonucleotide species. In some embodiments, each pool comprises at least 65,000 unique oligonucleotide species. In some embodiments, at least 48 pools are made. In some embodiments, the method further comprises immobilizing the unique oligonucleotide species of each pool onto a substrate, so that the sample barcodes but not the molecule

barcodes are the same for the oligonucleotide species immobilized on each substrate of the plurality. In some embodiments, the substrates comprise discrete regions of a surface. In some embodiments, the substrates comprise beads. In some embodiments, the unique oligonucleotide species are disposed in spatially-isolated pools concurrent with said synthesis. In some embodiments, the unique oligonucleotide species are disposed in spatially-isolated pools after said synthesis. In some embodiments, the uniform region comprises a target-specific region comprising a sequence flanking a variable region coding sequence of an immune cell receptor or immunoglobulin. In some embodiments, the immune cell receptor variable region coding sequence is a T cell receptor variable region coding sequence, a B cell receptor variable region coding sequence, or a combination thereof. In some embodiments, the kit further comprising a primer configured to hybridize on an opposite side of the variable region coding sequence as the target-specific region, and to hybridize to a complementary strand to a strand hybridized by the target-specific region, and is thus configured, in conjunction with the target-specific region, to amplify the variable region coding sequence.

[0006] Some embodiments include an oligonucleotide comprising a barcode region 3' of the adapter region. The barcode region can comprise a molecule barcode comprising at least 7 nucleotides, in which the molecule barcode has a G content of no more than 50%. In some embodiments, the oligonucleotide further comprises a sample barcode comprising at least 3 nucleotides. In some embodiments, the oligonucleotide further comprises a uniform region 3' of the barcode region. In some embodiments, the uniform region comprises a target-specific region comprising at least 10 nucleotides complementary to a target nucleic acid. In some embodiments, the oligonucleotide further comprises an adapter region 5' of the barcode region. In some embodiments, the uniform region comprises a target-specific region comprising a sequence flanking a variable region coding sequence of an immune cell receptor or immunoglobulin. In some embodiments, the immune cell receptor variable region coding sequence is a T cell receptor variable region coding sequence, a B cell receptor variable region coding sequence, or a combination thereof. In some embodiments, the kit further comprising a primer configured to hybridize on an opposite side of the variable region coding sequence as the target-specific region, and to hybridize to a

complementary strand to a strand hybridized by the target-specific region, and is thus configured to amplify the variable region coding sequence in conjunction with the target-specific region.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] **FIG. 1A** is a schematic diagram illustrating an oligonucleotide species comprising a barcode region 5' of a uniform region, in accordance with some embodiments herein.

[0008] **FIG. 1B** is a schematic diagram illustrating an oligonucleotide species comprising a barcode region 5' of an oligo dT sequence, in accordance with some embodiments herein.

[0009] **FIG. 1C** is a schematic diagram illustrating an oligonucleotide species comprising a barcode region 5' of a gene-specific region, in accordance with some embodiments herein.

[0010] **FIG. 2A** is a schematic diagram illustrating a barcode region comprising a molecule barcode 5' of a sample barcode, and which can comprise a barcode region of a unique oligonucleotide species in accordance with some embodiments herein.

[0011] **FIG. 2B** is a schematic diagram illustrating a barcode region comprising a sample barcode 5' of a molecule barcode, and which can comprise a barcode region of a unique oligonucleotide species in accordance with some embodiments herein.

[0012] **FIGs. 3A-3H** are diagrams illustrating a variety of configurations of oligonucleotide species in accordance with some embodiments herein. **FIG. 3A** illustrates a baseline oligonucleotide. **FIG. 3B** illustrates an oligonucleotide with a base spacer 3' of the barcode region. **FIG. 3C** illustrates an oligonucleotide with a 5-base spacer 3' of the barcode region. **FIG. 3D** illustrates an oligonucleotide in which the positions of the molecule barcode (indicated as "NNNNNNNN") and the sample barcode within the barcode region are switched in comparison to **FIG. 3A**, so that the molecule barcode is 5' of the sample barcode. **FIG. 3E** illustrates an oligonucleotide in which the positions of the molecule barcode (indicated as "NNNNNNNN") and the sample barcode within the barcode region are switched in comparison to **FIG. 3A**, so that the molecule barcode is 5' of the sample barcode,

and including a 3-base spacer. **FIG. 3F** illustrates an oligonucleotide in which the positions of the molecule barcode (indicated as “NNNNNNNN”) and the sample barcode within the barcode region are switched, so that the molecule barcode is 5’ of the sample barcode, and including a 5-base spacer. **FIG. 3G** illustrates an oligonucleotide in which the molecule barcode comprises the sequence HHHHHHHH (in which each H is an A, C, or T, and in which each H can be the same or different from any other H). **FIG. 3H** illustrates an oligonucleotide in which the molecule barcode comprises the sequence HNHNHNHN (in which each H is an A, C, or T, and in which each H can be the same or different from any other H).

[0013] **FIG. 4A** is a diagram illustrating nucleotide usage in a sampling of conventional unique oligonucleotide species.

[0014] **FIG. 4B** is a diagram illustrating nucleotide usage in a sampling of conventional unique oligonucleotide species.

[0015] **FIG. 4C** is a diagram illustrating nucleotide usage in a sampling of conventional unique oligonucleotide species.

[0016] **FIG. 5A** is a diagram illustrating nucleic acid sequences of oligonucleotide species that were constructed in accordance with some embodiments herein. The indicated SEQ ID NOs: in the right-most columns of **FIGs 5A-5D** refer to the polynucleotide sequence that includes the “anchor,” “1st bc,” “2nd bc,” “Spacer” (if any), and “5 to 3’ Recognition Sequence” as shown. Each “anchor” in **FIGs. 5A-5D** has the polynucleotide sequence of SEQ ID NO: 97.

[0017] **FIG. 5B** is a diagram illustrating nucleic acid sequences of oligonucleotide species that were constructed in accordance with some embodiments herein.

[0018] **FIG. 5C** is a diagram illustrating nucleic acid sequences of oligonucleotide species that were constructed in accordance with some embodiments herein.

[0019] **FIG. 5D** is a diagram illustrating nucleic acid sequences of oligonucleotide species that were constructed in accordance with some embodiments herein.

[0020] **FIG. 6** is a graph of error correction by individual for an analysis using unique oligonucleotide species in accordance with some embodiments herein.

[0021] **FIG. 7** is a graph of error correction by individual and filter MI crossover for an analysis using unique oligonucleotide species in accordance with some embodiments herein.

[0022] **FIG. 8** is a diagram schematically illustrating, without being limited by any theory, possible amplification errors that can result from oligonucleotides with high G-contents.

DETAILED DESCRIPTION

[0023] In accordance with some embodiments herein, methods and compositions for accurate barcoding and analysis of nucleic acids are described. In some embodiments, individual nucleic acids of a sample can be associated with a unique barcode (e.g. a “molecule barcode”), so that upon amplification and sequence analysis, individual nucleic acids of a sample can be quantified. Without being limited by any theory, it is contemplated that bias favoring or disfavoring the representation, amplification, or properties of certain kinds of barcode sequences can interfere with quantification and analysis of the individual nucleic acids of a sample (possible sources of bias in some amplification events are schematically illustrated in **FIG. 8**). Described in accordance with some embodiments herein are configurations and characteristics of unique oligonucleotide species comprising barcodes, in which the unique oligonucleotide species are configured to minimize barcode-related bias, and yield accurate analysis of nucleic acids. Without being limited by any theory, it is contemplated that barcode regions comprising features such as a guanosine (G) content below 50%, and/or no two consecutive “G’s” in the barcode region can minimize bias that could otherwise confound quantification and/or analysis of nucleic acids of a sample. Optionally, the individual nucleic acid molecules of a given sample can also be associated with a “sample barcode”, so that barcode-associated nucleic acids can be subsequently pooled for an efficient batch analysis of nucleic acids from two or more samples, for example by next-generation sequencing.

Nucleic Aids

[0024] Various nucleic acids are described in accordance with some embodiments herein. For example, oligonucleotide species, samples, and/or targets can comprise nucleic acids.

[0025] As used herein, a “nucleic acid” refers to a polynucleotide sequence, or fragment thereof. A nucleic acid can comprise nucleotides. A nucleic acid can be exogenous or endogenous to a cell. A nucleic acid can exist in a cell-free environment. A nucleic acid can comprise, consist of, or consist essentially of a gene or fragment thereof. A nucleic acid can comprise, consist of, or consist essentially of DNA. A nucleic acid can comprise, consist of, or consist essentially of RNA. A nucleic acid can comprise one or more analogs (e.g. altered backbone, sugar, or nucleobase). Some non-limiting examples of analogs include: 5-bromouracil, peptide nucleic acid, xeno nucleic acid, morpholinos, locked nucleic acids, glycol nucleic acids, threose nucleic acids, dideoxynucleotides, cordycepin, 7-deaza-GTP, fluorophores (e.g. rhodamine or fluorescein linked to the sugar), thiol containing nucleotides, biotin linked nucleotides, fluorescent base analogs, CpG islands, methyl-7-guanosine, methylated nucleotides, inosine, thiouridine, pseudouridine, dihydrouridine, queuosine, and wyosine. “Nucleic acid”, “polynucleotide,” “target polynucleotide”, and “target nucleic acid” can be used interchangeably.

[0026] As used herein, “upstream” (and variations of this root term) refers to a position that is relatively 5' on a nucleic acid (e.g., 5' in comparison to reference position). As used herein “downstream” (and variations of this root term) refers to a position that is relatively 3' on a nucleic acid (e.g., 3' in comparison to reference position). For example, as shown in **FIG. 2A**, the “sample barcode” is 3' of the “molecule barcode” and is understood to be downstream of the molecule “barcode.” For example, as shown in **FIG. 2B**, the “sample barcode” is 5' of the “molecule barcode” and is understood to be upstream of the molecule “barcode.”

[0027] A nucleic acid can comprise one or more modifications (e.g., a base modification, a backbone modification), to provide the nucleic acid with a new or enhanced feature (e.g., improved stability). A nucleic acid can comprise a nucleic acid affinity tag. A nucleoside can comprise, consist of, or consist essentially of a base-sugar combination. The base portion of the nucleoside can be a heterocyclic base. The two most common classes of

such heterocyclic bases are the purines and the pyrimidines. Nucleotides can comprise, consist of, or consist essentially of nucleosides that further include a phosphate group covalently linked to the sugar portion of the nucleoside. For those nucleosides that include a pentofuranosyl sugar, the phosphate group can be linked to the 2', the 3', or the 5' hydroxyl moiety of the sugar. In forming nucleic acids, the phosphate groups can covalently link adjacent nucleosides to one another to form a linear polymeric compound. In turn, the respective ends of this linear polymeric compound can be further joined to form a circular compound; however, linear compounds are generally suitable. In addition, linear compounds may have internal nucleotide base complementarity and may therefore fold in a manner as to produce a fully or partially double-stranded compound. Within nucleic acids, the phosphate groups can commonly be referred to as forming the internucleoside backbone of the nucleic acid. The linkage or backbone of the nucleic acid can be a 3' to 5' phosphodiester linkage.

[0028] A nucleic acid can comprise a modified backbone and/or modified internucleoside linkages. Modified backbones can include those that retain a phosphorus atom in the backbone and those that do not have a phosphorus atom in the backbone. Suitable modified nucleic acid backbones containing a phosphorus atom therein can include, for example, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates such as 3'-alkylene phosphonates, 5'-alkylene phosphonates, chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, phosphorodiamidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, selenophosphates, and boranophosphates having normal 3'-5' linkages, 2'-5' linked analogs, and those having inverted polarity wherein one or more internucleotide linkages is a 3' to 3', a 5' to 5' or a 2' to 2' linkage.

[0029] A nucleic acid can comprise polynucleotide backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These can include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and

thioformacetyl backbones; riboacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH₂ component parts.

[0030] A nucleic acid can comprise a nucleic acid mimetic. The term “mimetic” includes, for example, polynucleotides wherein only the furanose ring or both the furanose ring and the internucleotide linkage are replaced with non-furanose groups, replacement of only the furanose ring can also be referred as being a sugar surrogate. The heterocyclic base moiety or a modified heterocyclic base moiety can be maintained for hybridization with an appropriate target nucleic acid. One such nucleic acid can be a peptide nucleic acid (PNA). In a PNA, the sugar-backbone of a polynucleotide can be replaced with an amide containing backbone, in particular an aminoethylglycine backbone. The nucleotides can be retained and are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone. The backbone in PNA compounds can comprise two or more linked aminoethylglycine units which gives PNA an amide containing backbone. The heterocyclic base moieties can be bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone.

[0031] A nucleic acid can comprise a morpholino backbone structure. For example, a nucleic acid can comprise a 6-membered morpholino ring in place of a ribose ring. In some of these embodiments, a phosphorodiamidate or other non-phosphodiester internucleoside linkage can replace a phosphodiester linkage.

[0032] A nucleic acid can comprise linked morpholino units (i.e. morpholino nucleic acid) having heterocyclic bases attached to the morpholino ring. Linking groups can link the morpholino monomeric units in a morpholino nucleic acid. Non-ionic morpholino-based oligomeric compounds can have less undesired interactions with cellular proteins. Morpholino-based polynucleotides can be nonionic mimics of nucleic acids. A variety of compounds within the morpholino class can be joined using different linking groups. A further class of polynucleotide mimetic can be referred to as cyclohexenyl nucleic acids (CeNA). The furanose ring normally present in a nucleic acid molecule can be replaced with a cyclohexenyl ring. CeNA DMT protected phosphoramidite monomers can be prepared and used for oligomeric compound synthesis using phosphoramidite chemistry. The incorporation of CeNA monomers into a nucleic acid chain

can increase the stability of a DNA/RNA hybrid. CeNA oligoadenylates can form complexes with nucleic acid complements with similar stability to the native complexes. A further modification can include Locked Nucleic Acids (LNAs) in which the 2'-hydroxyl group is linked to the 4' carbon atom of the sugar ring thereby forming a 2'-C,4'-C-oxymethylene linkage thereby forming a bicyclic sugar moiety. The linkage can comprise, consist of, or consist essentially of a methylene (-CH₂-), group bridging the 2' oxygen atom and the 4' carbon atom wherein n is 1 or 2. LNA and LNA analogs can display very high duplex thermal stabilities with complementary nucleic acid ($T_m = +3$ to $+10^\circ$ C.), stability towards 3'-exonucleolytic degradation and good solubility properties.

[0033] A nucleic acid can also, in some embodiments, include nucleobase (often referred to simply as “base”) modifications or substitutions. As used herein, “unmodified” or “natural” nucleobases can include the purine bases, (e.g. adenine (A) and guanine (G)), and the pyrimidine bases, (e.g. thymine (T), cytosine (C) and uracil (U)). Modified nucleobases can include other synthetic and natural nucleobases such as 5-methylcytosine (5-me-C), 5-hydroxymethyl cytosine, xanthine, hypoxanthine, 2-aminoadenine, 6-methyl and other alkyl derivatives of adenine and guanine, 2-propyl and other alkyl derivatives of adenine and guanine, 2-thiouracil, 2-thiothymine and 2-thiocytosine, 5-halouracil and cytosine, 5-propynyl (—C≡C—CH₃) uracil and cytosine and other alkynyl derivatives of pyrimidine bases, 6-azo uracil, cytosine and thymine, 5-uracil (pseudouracil), 4-thiouracil, 8-halo, 8-amino, 8-thiol, 8-thioalkyl, 8-hydroxyl and other 8-substituted adenines and guanines, 5-halo particularly 5-bromo, 5-trifluoromethyl and other 5-substituted uracils and cytosines, 7-methylguanine and 7-methyladenine, 2-F-adenine, 2-aminoadenine, 8-azaguanine and 8-azaadenine, 7-deazaguanine and 7-deazaadenine and 3-deazaguanine and 3-deazaadenine. Modified nucleobases can include tricyclic pyrimidines such as phenoxazine cytidine (1H-pyrimido(5,4-b)(1,4)benzoxazin-2(3H)-one), phenothiazine cytidine (1H-pyrimido(5,4-b)(1,4)benzothiazin-2(3H)-one), G-clamps such as a substituted phenoxazine cytidine (e.g. 9-(2-aminoethoxy)-H-pyrimido(5,4-(b) (1,4)benzoxazin-2(3H)-one), carbazole cytidine (2H-pyrimido(4,5-b)indol-2-one), pyridoindole cytidine (Hpyrido(3',:4,5)pyrrolo[2,3-d]pyrimidin-2-one).

Samples

[0034] As used herein, the term “sample” refers to a composition comprising targets. Suitable samples for analysis by the disclosed methods, devices, and systems include, but are not limited to, cells, single cells, tissues, organs, or organisms. In some embodiments, a sample comprises raw or unprocessed samples, for example a whole cell, whole population of cells, or whole tissue. In some embodiments, a sample comprises an isolated cell or cell extract, or a nucleic-acid-containing fraction thereof, for example isolated nucleic acids, or a composition comprising enriched or isolated nucleic acids. In some embodiments, a sample comprises a fixed tissue, cell, or nucleic-acid-containing fraction thereof. In some embodiments, a sample comprises a frozen tissue, cell, or nucleic-acid-containing fraction thereof. In some embodiments, a sample comprises a solution comprising nucleic acids. In some embodiments, a sample comprises a solution comprising nucleic acids. In some embodiments, a sample comprises nucleic acids in a solid format, for example lyophilized nucleic acids and the like.

Unique oligonucleotide species

[0035] As used in compositions, methods, and oligonucleotides in accordance with some embodiments herein, a “unique oligonucleotide species” refers to an oligonucleotide, for example DNA or RNA, having a sequence that differs by at least one base from another unique oligonucleotide species. The unique oligonucleotide species of a composition in accordance with some embodiments herein can share certain structural features or formats, but can have different nucleic acid sequences from each other. The unique oligonucleotide species can be single-stranded or double-stranded. A composition can comprise two or more unique oligonucleotide species, for example a diversity of 100, 1000, 6500, or 65,000 unique oligonucleotide species. Optionally, the composition comprising unique oligonucleotide species can also comprise two or more oligonucleotides of the same unique oligonucleotide species. By way of example a composition can comprise two unique oligonucleotide species: ACTT-X and TCTT-X, in which “X” is a sequence that is the same for both unique oligonucleotide species. It would be possible for the composition to

comprise two copies of an oligonucleotide having the sequence ACTT-X, and one copy of an oligonucleotide having the sequence TCTT-X.

[0036] The oligonucleotide species of the compositions, methods, and oligonucleotides of some embodiments herein can comprise a barcode region and a uniform region as described herein. The barcode regions can differ between unique oligonucleotide species, so as to provide diversity in a population of unique oligonucleotide species, while the uniform regions remain the same. The barcode region can comprise a molecule index as described herein. The molecule index can be configured to minimize bias, for example by minimizing G content so that no unique oligonucleotide species in a population of unique oligonucleotide species has a molecule index with a G content greater than 50%, and/or so that the sequence “GG” does not appear in the molecule index (e.g. so that there are no two consecutive G’s). Optionally, the barcode region comprises a sample index. The sample index can be configured so that the unique oligonucleotides of a given pool can have the same sample index, but different molecule indices. As such, if multiple samples are analyzed, the sample index can indicate which sample each oligonucleotide corresponds to. As such, after unique oligonucleotide species bind to target, the unique oligonucleotide species can be pooled, and sequences can be analyzed. In some embodiments, the sample index is 5’ of the molecule index. In some embodiments, the molecule index is 5’ of the sample index. Optionally, the unique oligonucleotide species comprises an adapter. The adapter can be positioned 5’ of the barcode region. In some embodiments, the adapter is configured to immobilize the unique oligonucleotide species on a substrate.

[0037] Without being limited by any theory, it is contemplated that unique oligonucleotides species configured in accordance with some embodiments herein can yield accurate analysis and sequencing results with reduced, minimal, or no bias, for example by minimizing G content in the molecule barcode or barcode region, and/or minimizing G content near the uniform region. By way of example, a reduction in bias can be ascertained by a reduction in noise and/or an increase in sensitivity for detecting the number of different molecules of a target nucleic acid in a sample (*see*, e.g. **FIGs. 6-7**). In some embodiments, the reduction in bias can be ascertained as a decrease in noise (*see* **FIG. 6**), for example a smaller standard error in the number of target nucleic acid molecules detected in the sample.

In some embodiments, the reduction in bias can be ascertained as an increase in sensitivity (*see* **FIG. 7**), for example detecting a larger number of the different molecules of a target nucleic acid in a sample (e.g. fewer target nucleic acid molecules are “missed”. As such, in some embodiments, compositions and methods yield quantification of target nucleic acid molecules in a sample with low noise, for example a relative standard error less than 30%, for example, less than 30%, 25%, 20%, 15%, 10%, 5%, 4%, 3%, 2%, 1%, or .01%, including ranges between any two of the listed values. In some embodiments, compositions and methods yield quantification of target nucleic acid molecules in a sample with high sensitivity, for example a sensitivity (measured as a percent of the actual number of different target nucleic acids in a sample that are detected), for example, a sensitivity of at least 70%, 75%, 80%, 85%, 90%, 85%, 96%, 97%, 98%, 99%, or 99.9%, including ranges between any two of the listed values. In some embodiments, the methods or compositions described herein reduce bias by increasing sensitivity, decreasing relative standard error, or increasing sensitivity and reducing standard error relative to compositions or methods comprising unique oligonucleotide species in line with **FIG. 3A**.

[0038] **FIG 3A** illustrates a conventional oligonucleotide design. It is noted that the molecule index sequence NNNNNNNN (in which each N is A, G, C, or T, and in which any two of the N’s can be the same or different from each other) can comprise two or more consecutive G’s, and/or can have a G content greater than 50%. For example, a subset of the population of unique oligonucleotide species in line with **FIG. 3A** can be G-rich. Without being limited by any theory, it is contemplated that at least some unique oligonucleotide species in a population based on the configuration of **FIG. 3A** could be favored, leading to bias. Another example of “baseline” molecule index that can be subject to bias is the sequence ‘BBBBBBV’ (in which B is C, G, or T, and in which V is A, C, or G). Examples of configurations for unique oligonucleotide species in accordance with some embodiments herein are illustrated in **FIGs 3B-3H**. As shown in **FIGs. 3B-C**, addition of either TTT or TTTTT after the molecule index in accordance with some embodiments herein can reduce G-richness along the region (the length of non-G spacers can be variable, and can comprise any non-G nucleotide or nucleotides). In **FIG. 3D**, the sample index and molecule positions are switched (so that the molecule index is 5’ of the sample index), so that possible high G-rich

MIIs in at least a subset of unique oligonucleotide species would not be adjacent to potentially G-rich gene specific target regions. In FIGs. 3E-F, addition of either TTT or TTTTT to the configuration illustrated in FIG. 3D can further minimize potential G-rich regions. Without being limited by any theory, it is noted that that sample barcodes frequently can have low G contents (e.g. so as to be “non-G-rich”). It is further noted that the length of non-G spacers can be variable and can be any non-G nucleotide). In the configurations of FIGs. 3G-H, molecule barcodes comprise ‘HHHHHHHHH’ or ‘HNHNHNHN,’ (in which H is A, C, or T, and in which any two H’s can be the same or different from each other, and in which any two N’s can be the same or different from each other).

[0039] In some embodiments, each of a plurality of unique oligonucleotide species (e.g., each of the unique oligonucleotide species) in a composition or method has a length of at least 24 nucleotides, for example at least 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, 55, 60, 65, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, or 140 nucleotides in length, including ranges between any two of the listed values, for example 24-140, 24-135, 24-130, 24-125, 24-120, 24-115, 24-110, 24-105, 24-100, 24-95, 24-90, 24-85, 24-80, 24-75, 24-70, 24-65, 24-60, 24-55, 24-50, 24-45, 24-40, 25-140, 25-135, 25-130, 25-125, 25-120, 25-115, 25-110, 25-105, 25-100, 25-95, 25-90, 25-85, 25-80, 25-75, 25-70, 25-65, 25-60, 25-55, 25-50, 25-45, 25-40, 27-140, 27-135, 27-130, 27-125, 27-120, 27-115, 27-110, 27-105, 27-100, 27-95, 27-90, 27-85, 27-80, 27-75, 27-70, 27-65, 27-60, 27-55, 27-50, 27-45, 27-40, 30-140, 30-135, 30-130, 30-125, 30-120, 30-115, 30-110, 30-105, 30-100, 30-95, 30-90, 30-85, 30-80, 30-75, 30-70, 30-65, 30-60, 30-55, 30-50, 30-45, 30-40, 35-140, 35-135, 35-130, 35-125, 35-120, 35-115, 35-110, 35-105, 35-100, 35-95, 35-90, 35-85, 35-80, 35-75, 35-70, 35-65, 35-60, 35-55, 35-50, 35-45, 35-40, 40-140, 40-135, 40-130, 40-125, 40-120, 40-115, 40-110, 40-105, 40-100, 40-95, 40-90, 40-85, 40-80, 40-75, 40-70, 40-65, 40-60, 40-55, 40-50, or 40-45 nucleotides in length. Optionally, different unique oligonucleotide species in a composition or method have different lengths from each other. Optionally, all of the unique oligonucleotide species in a composition or method have the same length as each other. Optionally, some unique oligonucleotide species in a composition or method are the same

length as each other, while some unique oligonucleotide species are different lengths from each other.

[0040] In some embodiments, the uniform region comprises, consists of, or consists essentially of a 5' to 3' amplification sequence for a target nucleic acid, or class of target nucleic acids (this amplification sequence may also be referred to as a "target-specific" region). For example, if the target nucleic acids comprise mRNAs, the uniform region can comprise oligo dT. For example, if the target nucleic acids comprise variable regions of a T cell receptor, the uniform region can comprise sequences flanking variable regions of a T cell receptor mRNA. In some embodiments, the uniform region comprises at least 10 nucleotides that are complementary to the target nucleic acid, for example at least 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 nucleotides complementary to the target, including ranges between any two of the listed values, for example 10-30, 10-29, 10-28, 10-27, 10-26, 10-25, 10-24, 10-23, 10-22, 10-21, 10-20, 11-30, 11-29, 11-28, 11-27, 11-26, 11-25, 11-24, 11-23, 11-22, 11-21, 11-20, 12-30, 12-29, 12-28, 12-27, 12-26, 12-25, 12-24, 12-23, 12-22, 12-21, 12-20, 15-30, 15-29, 15-28, 15-27, 15-26, 15-25, 15-24, 15-23, 15-22, 15-21, 15-20, 20-30, 20-29, 20-28, 20-27, 20-26, 20-25, 20-24, 20-23, 20-22, or 20-21 nucleotides. In some embodiments, the uniform region comprises a target-specific region that hybridizes to a sequence flanking a sequence encoding a variable region of an immune cell receptor, for example a variable region of a T cell receptor, a B cell receptor, or immunoglobulin, for example an antibody. It is noted that because B cell receptors comprise membrane-bound immunoglobulin, target-specific regions specific for immunoglobulin variable region coding sequences are typically suitable for amplifying B cell receptors as well as secreted immunoglobulins (e.g. antibodies). Both options are noted herein to clarify that amplification of membrane-bound immunoglobulins (B cell receptors) and also secreted immunoglobulins (antibodies) is contemplated. As used herein, it will be understood that when primers or uniform regions comprise target-specific regions that comprise "flanking sequences" (and variations of this root term, such as "sequences flanking") of immune cell receptor and/or immunoglobulin variable regions, the target-specific regions will be understood to comprise at least one of (i) sequences that hybridize downstream (3') of the sequence encoding the variable region and in particular hybridize to strand of the coding

sequence, and thus are configured to produce a strand comprising the variable region coding sequence upon extension in the 5' to 3' direction; or (ii) sequences that hybridize upstream (5') of the sequence encoding the variable region and in particular hybridize to the strand complementary to that of the coding sequence, and thus are configured to produce a strand comprising the reverse complement of the variable region coding sequence upon extension in the 5' to 3' direction. Thus, a flanking sequence can be configured for amplification of the variable region coding sequences in conjunction with a suitable priming partner (e.g. a flanking sequence on the other side of the variable region). It will be understood that a flanking sequence does not necessarily need to stop or start exactly where the coding sequence starts or stops, and thus, it is permissible for there to be intervening sequences between a hybridization site of a flanking sequence and the variable region coding sequence itself. It will be understood that while a flanking sequence will generally hybridize to a sequence outside of the variable region coding sequence so as to amplify a broad range of possible variable region coding sequences, in some embodiments, the variable region "flanking sequence" further comprises some sequence of the variable region itself, for example if a subset of possible variable regions is of interest. However, a "flanking sequence" as used herein does not require a single sequence to flank both sides of the variable region. Rather, it will be understood that when flanking sequences are mentioned in conjunction with compositions, methods, and oligonucleotides of some embodiments herein, 5' and 3' sequences comprising suitable primer pairs to amplify the variable region coding sequence are also expressly contemplated.

[0041] In some embodiments, a unique oligonucleotide species comprises a barcode region as described herein and also comprises a uniform region comprising a target-specific region comprising a sequence flanking an immune cell receptor and/or immunoglobulin variable region coding sequence. A second oligonucleotide primer can also be provided for the other side of the variable region coding sequence, so as to amplify the variable region sequence in conjunction with the target-specific region of the uniform region. In some embodiments, the uniform region comprises a target-specific region, positioned 3' of the barcode region, and comprising a sequence flanking an immunoglobulin variable region (and thus flanking a B cell receptor variable region as well as a corresponding antibody

variable region), for example flanking the variable region of an immunoglobulin heavy chain locus, flanking the variable region of an immunoglobulin (light chain) kappa locus, or flanking the variable region of an immunoglobulin (light chain) lambda locus. In some embodiments, the uniform region comprises a target-specific region, positioned 3' of the barcode region, and comprising a sequence flanking at least one of a variable region of a T cell receptor alpha chain, a variable region of a T cell receptor beta chain, a variable region of a T cell receptor gamma chain, or a T cell receptor delta chain.

[0042] In some embodiments, a kit comprises a composition comprising unique oligonucleotide species as described herein, in which the unique oligonucleotide species each comprise a uniform sequence flanking an immune cell receptor or immunoglobulin variable region coding sequence as described herein. In some embodiments, the kit further comprises an oligonucleotide primer configured to hybridize to the opposite strand and on the other side of the variable region coding sequence compared to the uniform region, and is thus configured to amplify the variable region sequence in conjunction with the target-specific region of the uniform region. In some embodiments, the amplified sequence is at least 1kb and comprises the variable coding sequence, for example at least 1kb, 2kb, 3kb, 4kb, or 5kb, including ranges between any two of the listed values.

[0043] It is noted the nucleic acids encoding variable regions of some immune cell receptors or immunoglobulins can be more than 1kb long. For example, T cell receptor variable region sequences can comprise a CDR3 coding sequence that ends more than 1kb away from where the CDR1 coding sequence begins. Without being limited by theory, it is noted that some conventional and next-generation sequencing approaches, for example sequencing by synthesis, are limited to short reads that are typically considerably less than 1 kb. Accordingly, it is contemplated that methods and compositions and kits in accordance with some embodiments herein can be useful for barcoding and analyzing nucleic acids encoding immune cell receptor and/or immunoglobulin variable regions, which otherwise would not be amenable to single-read sequencing of less than 1 kb. Accordingly, in some embodiments, the unique oligonucleotide species comprises a sequence flanking an immune cell receptor or immunoglobulin variable region coding sequence, is configured to amplify a

sequence of at least 1kb and comprising the variable region coding sequence, for example at least 1kb, 2kb, 3kb, 4kb, or 5kb, including ranges between any two of the listed values.

Barcode regions

[0044] In accordance with compositions, methods, and oligonucleotides of some embodiments herein, a barcode region comprises a nucleic acid sequence that is useful in identifying a nucleic acid, for example a target nucleic acid from a sample, or an amplicon or reverse-transcript derived from a single target nucleic acid of a sample. For example, two mRNA transcripts from a sample can be reverse-transcribed and barcoded so that nucleic acids corresponding to the first mRNA comprise a first barcode, and nucleic acids corresponding to the second mRNA comprise a second barcode. Upon sequencing (or other analysis), information about the individual mRNAs in the sample, for example copy number can be ascertained even after amplification. However, if a large population of mRNAs is stochastically labeled and some barcodes are represented more favorably (for example due to stability, amplification efficiency, etc.), bias can result, skewing the ability to quantify nucleic acids of a sample. As such, in accordance with some embodiments herein, each unique oligonucleotide species in a population can comprise a unique barcode region. The greater the diversity of barcodes, the greater the diversity of unique oligonucleotide species, and the greater the probability that a particular barcode sequence will be associated with only one target nucleic acid of a sample. The barcode region can be, for example, positioned 5' of a uniform region on an oligonucleotide species. In some embodiments, a barcode region comprises a molecule barcode as described herein. Optionally, a barcode region comprises a molecule barcode and a sample barcode as described herein. Optionally, a barcode region comprises a molecule barcode 5' of a sample barcode. In some embodiments, methods or compositions comprising unique oligonucleotide species comprising molecule barcodes as described herein reduce bias by increasing sensitivity, decreasing relative standard error, or increasing sensitivity and reducing standard error. As used herein, a "molecule barcode" can also be referred to as a "molecular barcode," "Molecular Index (MI)", or Unique Molecular Identifier (UMI). As used herein, a "sample barcode," can also be referred to as a "Sample Index (SI)."

[0045] A barcode region can comprise a molecule barcode. The molecule barcode can comprise a unique sequence, so that when multiple sample nucleic acids (which can be the same and/or different from each other) are associated one-to-one with molecule barcodes, different sample nucleic acids can be differentiated from each other by the molecule barcodes. As such, even if a sample comprises two nucleic acids having the same sequence, each of these two nucleic acids can be labeled with a different molecule barcode, so that nucleic acids in the population can be quantified, even after amplification. The molecule barcode can comprise a nucleic acid sequence of at least 5 nucleotides, for example at least 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, or 50 nucleotides, including ranges between any two of the listed values, for example 5-50, 5-45, 5-40, 5-35, 5-30, 5-25, 5-20, 5-15, 5-14, 5-13, 5-12, 5-11, 5-10, 5-9, 5-8, 5-7, 5-6, 6-50, 6-45, 6-40, 6-35, 6-30, 6-25, 6-20, 6-15, 6-14, 6-13, 6-12, 6-11, 6-10, 6-9, 6-8, 6-7, 7-50, 7-45, 7-40, 7-35, 7-30, 7-25, 7-20, 7-15, 7-14, 7-13, 7-12, 7-11, 7-10, 7-9, 7-8, 8-50, 8-45, 8-40, 8-35, 8-30, 8-25, 8-20, 8-15, 8-14, 8-13, 8-12, 8-11, 8-10, 8-9, 9-50, 9-45, 9-40, 9-35, 9-30, 9-25, 9-20, 9-15, 9-14, 9-13, 9-12, 9-11, 9-10, 10-50, 10-45, 10-40, 10-35, 10-30, 10-25, 10-20, 10-15, 10-14, 10-13, 10-12, or 10-11 nucleotides. In some embodiments, the nucleic acid sequence of the molecule barcode comprises a unique sequence, for example, so that each unique oligonucleotide species in a composition comprises a different molecule barcode. In some embodiments, two or more unique oligonucleotide species can comprise the same molecule barcode, but still differ from each other. For example, if the unique oligonucleotide species include sample barcodes, each unique oligonucleotide species with a particular sample barcode can comprise a different molecule barcode. In some embodiments, a composition comprising unique oligonucleotide species comprises a molecule barcode diversity of at least 1000 different molecule barcodes, and thus at least 1000 unique oligonucleotide species. In some embodiments, a composition comprising unique oligonucleotide species comprises a molecule barcode diversity of at least 6,500 different molecule barcodes, and thus at least 6,500 unique oligonucleotide species. In some embodiments, a composition comprising unique oligonucleotide species comprises a molecule barcode diversity of at least 65,000 different molecule barcodes, and thus at least 65,000 unique oligonucleotide species. **[0042]**

Without being limited by any theory, it is contemplated that a molecule barcode comprising a low G content (e.g., 50% G or less, for example, less than 50% G, 45% G, 40% G, 35% G, 30% G, 25% G, 20% G, 15% G, 12.5% G, 10% G, 7.5% G, 5% G, or 2.5% G) can minimize bias for a composition or pool of unique oligonucleotide species being used to barcode a population of nucleic acids (e.g. minimize bias that would preferentially amplify barcodes comprising a higher G content). It is noted that conventional barcoding approaches typically exhibit a bias in favor of higher G content. For example **FIGs. 4A, 4B, and 4C** illustrate samples of nucleotide usage in conventional molecule barcodes of compositions comprising numerous unique molecule barcodes for ES32, TRAC (**FIG. 4A**), ES32 TRBC (**FIG. 4B**), and ES32 OligodT (**FIG. 4C**). That is, if anything, conventional molecule barcodes and barcode regions, designed without respect to certain guidance provided herein, can comprise a higher G-content that would be expected by random chance. In some embodiments, methods or compositions comprising unique oligonucleotide species comprising molecule barcodes as described herein reduce bias by increasing sensitivity, decreasing relative standard error, or increasing sensitivity and reducing standard error.

[0046] In some embodiments, all of the molecule barcodes of a composition or composition used in the methods described herein comprise unique oligonucleotide species collectively have a G content of 50% G or less, for example, less than 50% G, 45% G, 40% G, 35% G, 30% G, 25% G, 20% G, 15% G, 12.5% G, 10% G, 7.5% G, 5% G, or 2.5% G, or 0% G, including ranges between any two of the listed values, for example 2.5-50% G, 2.5-45% G, 2.5-40% G, 2.5-35% G, 2.5-30% G, 2.5-25% G, 2.5-20% G, 2.5-15% G, 2.5-10% G, 2.5-7.5% G, 2.5-5% G, 5-50% G, 5-45% G, 5-40% G, 5-35% G, 5-30% G, 5-25% G, 5-20% G, 5-15% G, 5-10% G, 5-7.5% G, 7.5-50% G, 7.5-45% G, 7.5-40% G, 7.5-35% G, 7.5-30% G, 7.5-25% G, 7.5-20% G, 7.5-15% G, 7.5-10% G, 10-50% G, 10-45% G, 10-40% G, 10-35% G, 10-30% G, 10-25% G, 10-20% G, 10-15% G, 10 – 12.5% G, 12.5-50% G, 12.5-45% G, 12.5-40% G, 12.5-35% G, 12.5-30% G, 12.5-25% G, 12.5-20% G, 12.5-15% G, 15-50% G, 15-45% G, 15-40% G, 15-35% G, 15-30% G, 15-25% G, 15-20% G, 20-50% G, 20-45% G, 20-40% G, 20-35% G, 20-30% G, or 20-25% G. By “all of the molecule barcodes of a composition of unique oligonucleotide species collectively have a G content of...”, it is meant that if the total G content among all of molecule barcodes in the whole composition

was calculated (e.g., a population of at least 10, 50, 100, 200, 500, 1000, 2000, 5000, 6500, 10,000, 20,000, 30,000, 40,000, 50,000, 60,000, or 65,000 unique oligonucleotide species), this total G content of the sum total of the barcodes would fall below the recited values or within the recited ranges. While it would still be possible for an individual unique oligonucleotide species to have a molecule barcode with a G content above the indicated value or outside the indicated range, the collective nucleotide content of the unique oligonucleotide species of the composition would be below the indicated value or within the indicated range. In some embodiments, all of the molecule barcodes in a composition comprising at least 1000 unique oligonucleotide species collectively have a G content of 50% G or less, for example, less than 50% G, 45% G, 40% G, 35% G, 30% G, 25% G, 20% G, 15% G, 12.5% G, 10% G, 7.5% G, 5% G, or 2.5% G, or 0% G, including ranges between any two of the listed values. In some embodiments, all of the molecule barcodes in a composition comprising at least 6500 unique oligonucleotide species collectively have a G content of 50% G or less, for example, less than 50% G, 45% G, 40% G, 35% G, 30% G, 25% G, 20% G, 15% G, 12.5% G, 10% G, 7.5% G, 5% G, or 2.5% G, or 0% G, including ranges between any two of the listed values. In some embodiments, all of the molecule barcodes in a composition comprising at least 65,000 unique oligonucleotide species collectively have a G content of 50% G or less, for example, less than 50% G, 45% G, 40% G, 35% G, 30% G, 25% G, 20% G, 15% G, 12.5% G, 10% G, 7.5% G, 5% G, or 2.5% G, or 0% G, including ranges between any two of the listed values. In some embodiments, all of the barcode regions of the composition as described herein collectively have a G content of less than 50% as described herein.

[0047] In some embodiments, for a composition comprising unique oligonucleotide species (or such a composition used in a method), the composition consists of or consists essentially of unique oligonucleotide species that each comprise a molecule barcode G content of 50% G or less, for example, less than 50% G, 45% G, 40% G, 35% G, 30% G, 25% G, 20% G, 15% G, 12.5% G, 10% G, 7.5% G, 5% G, or 2.5% G, or 0% G, including ranges between any two of the listed values, for example 2.5-50% G, 2.5-45% G, 2.5-40% G, 2.5-35% G, 2.5-30% G, 2.5-25% G, 2.5-20% G, 2.5-15% G, 2.5-10% G, 2.5-7.5% G, 2.5-5% G, 5-50% G, 5-45% G, 5-40% G, 5-35% G, 5-30% G, 5-25% G, 5-20% G,

5-15% G, 5-10% G, 5-7.5% G, 7.5-50% G, 7.5-45% G, 7.5-40% G, 7.5-35% G, 7.5-30% G, 7.5-25% G, 7.5-20% G, 7.5-15% G, 7.5-10% G, 10-50% G, 10-45% G, 10-40% G, 10-35% G, 10-30% G, 10-25% G, 10-20% G, 10-15% G, 10 – 12.5% G, 12.5-50% G, 12.5-45% G, 12.5-40% G, 12.5-35% G, 12.5-30% G, 12.5-25% G, 12.5-20% G, 12.5-15% G, 15-50% G, 15-45% G, 15-40% G, 15-35% G, 15-30% G, 15-25% G, 15-20% G, 20-50% G, 20-45% G, 20-40% G, 20-35% G, 20-30% G, or 20-25% G. By “the composition consists of or consists essentially of unique oligonucleotide species that each have a molecule barcodes G content less than...”, it is mean that each or essentially each of the unique oligonucleotide species in a composition, population, or pool have a molecule barcode G content below the indicated value, or outside the indicated range. That is, for a composition, population, or pool “consisting essentially of” unique oligonucleotide species that each have the indicated G content, it would be possible for an analytically insignificant portion of the unique oligonucleotides in the composition to have molecule barcodes with a G content above the indicated value or outside the recited range. For example, the analytically insignificant portion of the unique oligonucleotides can be, or can be no more than, 5%, 4%, 3%, 2%, 1%, 0.5%, 0.4%, 0.3%, 0.2%, 0.1%, or less of the unique oligonucleotide in a composition. In some embodiments, less than 1% of the unique oligonucleotide species in a composition comprising at least 1000 unique oligonucleotides comprise a molecule barcode having a G content of greater than 50%, for example less than 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, 0.1%, 0.05%, 0.01%, 0.001%, or 0.0001% of the unique oligonucleotide species have the a G content of greater than 50%. In some embodiments, less than 1% of the unique oligonucleotide species in a composition comprising at least 6500 unique oligonucleotides comprise a molecule barcode having a G content of greater than 50%, for example less than 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, 0.1%, 0.05%, 0.01%, 0.001%, or 0.0001% of the unique oligonucleotide species have a G content of greater than 50%. In some embodiments, less than 1% of the unique oligonucleotide species in a composition comprising at least 65,000 unique oligonucleotides comprise a molecule barcode having a G content of greater than 50%, for example less than 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, 0.1%, 0.05%, 0.01%, 0.001%, or 0.0001% of the unique oligonucleotide species have a G content of greater than 50%. In some embodiments, less than 1% of the unique

oligonucleotide species in a composition comprising at least 1000 unique oligonucleotides comprise a molecule barcode having a G content of greater than 25%, for example less than 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, 0.1%, 0.05%, 0.01%, 0.001%, or 0.0001% of the unique oligonucleotide species have a G content of greater than 25%. In some embodiments, less than 1% of the unique oligonucleotide species in a composition of at least 6500 unique oligonucleotides comprise a molecule barcode having a G content of greater than 25%, for example less than 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, 0.1%, 0.05%, 0.01%, 0.001%, or 0.0001% of the unique oligonucleotide species have a G content of greater than 25%. In some embodiments, less than 1% of the unique oligonucleotide species in a composition, population, or pool of at least 65,000 unique oligonucleotides comprise a molecule barcode having a G content of greater than 25%, for example less than 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, 0.1%, 0.05%, 0.01%, 0.001%, or 0.0001% of the unique oligonucleotide species have a G content of greater than 25%. Optionally, none of the molecule barcodes of the unique oligonucleotide species of the composition collectively have a G content of greater than 50% G, for example, all of the molecule barcodes of the unique oligonucleotide species have a G content less than 50% G, 45% G, 40% G, 35% G, 30% G, 25% G, 20% G, 15% G, 12.5% G, 10% G, 7.5% G, 5% G, or 2.5% G, or 0% G, including ranges between any two of the listed values.

[0048] In some embodiments, the composition as described herein (or such a composition as used in a method) consists of or consists essentially of unique oligonucleotide species that each comprise a barcode region G content of less than 50% as described herein. In some embodiments, for a composition comprising unique oligonucleotide species, the composition consists of or consists essentially of unique oligonucleotide species that each comprise a barcode region G content of 50% G or less, for example, less than 50% G, 45% G, 40% G, 35% G, 30% G, 25% G, 20% G, 15% G, 12.5% G, 10% G, 7.5% G, 5% G, or 2.5% G, or 0% G, including ranges between any two of the listed values, for example 2.5-50% G, 2.5-45% G, 2.5-40% G, 2.5-35% G, 2.5-30% G, 2.5-25% G, 2.5-20% G, 2.5-15% G, 2.5-10% G, 2.5-7.5% G, 2.5-5% G, 5-50% G, 5-45% G, 5-40% G, 5-35% G, 5-30% G, 5-25% G, 5-20% G, 5-15% G, 5-10% G, 5-7.5% G, 7.5-50% G, 7.5-45% G, 7.5-40% G, 7.5-35% G, 7.5-30% G, 7.5-25% G, 7.5-20% G, 7.5-15% G, 7.5-10% G, 10-50% G, 10-45%

G, 10-40% G, 10-35% G, 10-30% G, 10-25% G, 10-20% G, 10-15% G, 10 – 12.5% G, 12.5-50% G, 12.5-45% G, 12.5-40% G, 12.5-35% G, 12.5-30% G, 12.5-25% G, 12.5-20% G, 12.5-15% G, 15-50% G, 15-45% G, 15-40% G, 15-35% G, 15-30% G, 15-25% G, 15-20% G, 20-50% G, 20-45% G, 20-40% G, 20-35% G, 20-30% G, or 20-25% G. By “the composition consists of or consists essentially of unique oligonucleotide species that each have a barcode regions G content less than...”, it is mean that each or essentially each of the unique oligonucleotide species in a composition, population, or pool have a barcode region G content below the indicated value, or outside the indicated range. That is, for a composition, population, or pool “consisting essentially of” unique oligonucleotide species that each have the indicated G content, it would be possible for an analytically insignificant portion of the unique oligonucleotides in the composition to have barcode regions with a G content above the indicated value or outside the recited range. In some embodiments, less than 1% of the unique oligonucleotide species in a composition comprising at least 1000 unique oligonucleotides comprise a barcode region having a G content of greater than 50%, for example less than 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, 0.1%, 0.05%, 0.01%, 0.001%, or 0.0001% of the unique oligonucleotide species have a G content greater than 50%, including ranges between any two of the listed values. In some embodiments, less than 1% of the unique oligonucleotide species in a composition comprising at least 6500 unique oligonucleotides comprise a barcode region having a G content of greater than 50%, for example less than 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, 0.1%, 0.05%, 0.01%, 0.001%, or 0.0001% of the unique oligonucleotide species have a G content greater than 50%, including ranges between any two of the listed values. In some embodiments, less than 1% of the unique oligonucleotide species in a composition comprising at least 65,000 unique oligonucleotides comprise a barcode region having a G content of greater than 50%, for example less than 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, 0.1%, 0.05%, 0.01%, 0.001%, or 0.0001% of the unique oligonucleotide species have a G content greater than 50%, including ranges between any two of the listed values. In some embodiments, less than 1% of the unique oligonucleotide species in a composition comprising at least 1000 unique oligonucleotides comprise a barcode region having a G content of greater than 25%, for example less than 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, 0.1%, 0.05%,

0.01%, 0.001%, or 0.0001% of the unique oligonucleotide species have a G content greater than 25%, including ranges between any two of the listed values. In some embodiments, less than 1% of the unique oligonucleotide species in a composition of at least 6500 unique oligonucleotides comprise a barcode region having a G content of greater than 25%, for example less than 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, 0.1%, 0.05%, 0.01%, 0.001%, or 0.0001% of the unique oligonucleotide species have a G content greater than 50%, including ranges between any two of the listed values. In some embodiments, less than 1% of the unique oligonucleotide species in a composition, population, or pool of at least 65,000 unique oligonucleotides comprise a barcode region having a G content of greater than 25%, for example less than 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, 0.1%, 0.05%, 0.01%, 0.001%, or 0.0001% of the unique oligonucleotide species have a G content greater than 25%, including ranges between any two of the listed values. Optionally, none of the barcode regions of the unique oligonucleotide species of the composition collectively have a G content of greater than 50% G, for example, all of the barcode regions of the unique oligonucleotide species have a G content less than 50% G, 45% G, 40% G, 35% G, 30% G, 25% G, 20% G, 15% G, 12.5% G, 10% G, 7.5% G, 5% G, or 2.5% G, or 0% G, including ranges between any two of the listed values.

[0049] In some embodiments, the composition (or such a composition as used in a method) consist of or consists essentially of unique oligonucleotide species that each have a molecule barcode comprising the sequence of at least three repeats of the doublet “HN” (in which each “H” is any of A, C, or T, and in which “N” is any of A, G, C, or T), for example at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 repeats, including ranges between any two of the listed values. Examples of multiple repeats of the doublet “HN” include HN, HNHN, HNHNHN, and HNHNHNHN. It is noted that while the formula “HN” describes constraints on the base content, not every H or every N has to be the same or different. For example, if the molecule barcodes of unique oligonucleotide species in a composition comprised HNHNHN, one molecule barcode can comprise the sequence ACTGCA, while another molecule barcode can comprise the sequence TAACTA, while another molecule barcode could comprise the sequence AGACAC. It is noted that any number of repeats of the doublet “HN” would have a G content of no more than 50%. In

some embodiments, at least 95% of the unique oligonucleotide species of a composition comprising at least 1000 unique oligonucleotide species comprise molecule barcodes comprising at least three repeats of the doublet “HN,” for example at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 repeats, including ranges between any two of the listed values. In some embodiments, at least 99% of the unique oligonucleotide species of a composition comprising at least 1000 unique oligonucleotide species comprise molecule barcodes comprising at least three repeats of the doublet “HN,” for example at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 repeats, including ranges between any two of the listed values. In some embodiments, at least 99.9% of the unique oligonucleotide species of a composition comprising at least 1000 unique oligonucleotide species comprise molecule barcodes comprising at least three repeats of the doublet “HN,” for example at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 repeats, including ranges between any two of the listed values. In some embodiments, at least 95% of the unique oligonucleotide species of a composition comprising at least 6500 unique oligonucleotide species comprise molecule barcodes comprising at least three repeats of the doublet “HN,” for example at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 repeats, including ranges between any two of the listed values. In some embodiments, at least 99% of the unique oligonucleotide species of a composition comprising at least 6500 unique oligonucleotide species comprise molecule barcodes comprising at least three repeats of the doublet “HN,” for example at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 repeats, including ranges between any two of the listed values. In some embodiments, at least 99.9% of the unique oligonucleotide species of a composition comprising at least 6500 unique oligonucleotide species comprise molecule barcodes comprising at least three repeats of the doublet “HN,” for example at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 repeats, including ranges between any two of the listed values. In some embodiments, at least 95% of the unique oligonucleotide species of a composition comprising at least 65,000 unique oligonucleotide species comprise molecule barcodes comprising at least three repeats of the doublet “HN,” for example at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 repeats, including ranges between any two of the listed values. In some embodiments, at least 99% of the unique oligonucleotide species of a

of composition comprising at least 65,000 unique oligonucleotide species comprise molecule barcodes comprising at least three repeats of the doublet “HN,” for example at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 repeats, including ranges between any two of the listed values. In some embodiments, at least 99.9% of the unique oligonucleotide species of a composition comprising at least 65,000 unique oligonucleotide species comprise molecule barcodes comprising at least three repeats of the doublet “HN,” for example at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 repeats, including ranges between any two of the listed values. In some embodiments, the composition consists of or consists essentially of at least 1000, 6500, or 65,000 unique oligonucleotide species that each have a molecule barcode comprising the sequence HNHNNH. In some embodiments, the composition consists of or consists essentially of of at least 1000, 6500, or 65,000 unique oligonucleotide species that each has a molecule barcode comprising the sequence HNHNNHNN. In some embodiments, at least 95%, 99%, or 99.9% of the barcode regions of the composition as described herein comprise at least three repeats of the doublet “HN” as described herein. Without being limited by any theory, it is noted that having a relatively large number of available nucleotides sequences for molecule barcodes can be helpful when barcoding a population of target nucleic acids from a sample, for example to increase the diversity of barcodes within a given sequence length along with the probability that each target nucleic acid will be uniquely labeled, while minimizing oligonucleotide species size. It is noted that limiting the G content of molecule barcodes and/or barcode regions can limit the diversity of these barcodes and barcode regions by decreasing the number of available nucleotides from which barcodes can be constructed (and the number of available different sequences per length of nucleic acid). As such, having some G’s in molecule barcodes or barcode regions in accordance with various embodiments herein can be helpful in increasing diversity, while limiting the G content can be helpful minimizing bias. It is noted, and has been observed (*see Example 2 and FIGs. 6-7*) that sequences comprising repeated “HN” doublets can yield low bias, while providing a compromise between reducing bias and maintaining a relatively large quantity of available nucleotide sequences, so that relatively high diversity can be obtained in a relatively short sequence, while still minimizing bias. In some embodiments, methods or compositions comprising unique oligonucleotide species

comprising molecule barcodes comprising repeated “HN” doublets as described herein reduce bias by increasing sensitivity, decreasing relative standard error, or increasing sensitivity and reducing standard error.

[0050] In some embodiments, the composition (or such a composition as used in a method) comprises, consists of, or consists essentially of unique oligonucleotide species that each comprise a molecule barcode comprising at least six consecutive “H’s” (in which each “H” is any of A, C, or T), for example at least 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 consecutive H’s, including ranges between any two of the listed values. It is noted that while the formula “H” describes constraints on the base content, not every H has to be the same (or different). For example, if the molecule barcodes of unique oligonucleotide species in a population each comprised the sequence HHHH, one molecule barcode of a unique oligonucleotide species could comprise ACTA, one molecule barcode of another unique oligonucleotide species could comprise TTAC, and one molecule barcode of another unique oligonucleotide species could comprise ACAT. In some embodiments, a composition comprises, consists of, or consists essentially of at least 1000 unique oligonucleotide species, of which at least 95% comprise a molecule barcode that comprises at least six consecutive H’s, for example at least 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 consecutive H’s, including ranges between any two of the listed values. In some embodiments, a composition comprises, consists of, or consists essentially of at least 1000 unique oligonucleotide species, of which at least 99% comprise a molecule barcode that comprises at least six consecutive H’s, for example at least 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 consecutive H’s, including ranges between any two of the listed values. In some embodiments, a composition comprises, consists of, or consists essentially of at least 1000 unique oligonucleotide species, of which at least 99.9% comprise a molecule barcode that comprises at least six consecutive H’s, for example at least 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 consecutive H’s, including ranges between any two of the listed values. In some embodiments, a composition comprises, consists of, or consists essentially of at least 6500 unique oligonucleotide species, of which at least 95% comprise a molecule barcode that comprises at least six consecutive H’s, for example at least 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 consecutive H’s, including ranges between any two of the listed values.

In some embodiments, a composition comprises, consists of, or consists essentially of at least 6500 unique oligonucleotide species, of which at least 99% comprise a molecule barcode that comprises at least six consecutive H's, for example at least 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 consecutive H's, including ranges between any two of the listed values. In some embodiments, a composition comprises, consists of, or consists essentially of at least 6500 unique oligonucleotide species, of which at least 99.9% comprise a molecule barcode that comprises at least six consecutive H's, for example at least 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 consecutive H's, including ranges between any two of the listed values. In some embodiments, a composition comprises, consists of, or consists essentially of at least 65,000 unique oligonucleotide species, of which at least 95% comprise a molecule barcode that comprises at least six consecutive H's, for example at least 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 consecutive H's, including ranges between any two of the listed values. In some embodiments, a composition comprises, consists of, or consists essentially of at least 65,000 unique oligonucleotide species, of which at least 99% comprise a molecule barcode that comprises at least six consecutive H's, for example at least 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 consecutive H's, including ranges between any two of the listed values. In some embodiments, a composition comprises, consists of, or consists essentially of at least 65,000 unique oligonucleotide species, of which at least 99.9% comprise a molecule barcode that comprises at least six consecutive H's, for example at least 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 consecutive H's, including ranges between any two of the listed values. In some embodiments, at least 95% of the molecule barcodes of the unique oligonucleotide species comprise a sequence totaling at least 6 alternating H's and N's, for example at least 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 alternating H's and N's, including ranges between any two of the listed values. In some embodiments, at least 95% of the molecule barcodes of the unique oligonucleotide species comprise the sequence HNHNNHNH, wherein each "H" is any one of A, C, or T, and wherein each "N" is any one of A, G, C, or T. In some embodiments, each molecule barcode of the unique oligonucleotide species comprises the sequence HNHNNHNH, wherein each "H" is any one of A, C, or T, and wherein each "N" is any one of A, G, C, or T. In some

embodiments, each molecule barcode of the unique oligonucleotide species comprises the sequence HHHHHHHH, wherein each “H” is any one of A, C, or T.

[0051] In some embodiments, at least 95% (e.g., 95%, 96%, 97%, 98%, 99% or more) of the unique oligonucleotide species of the composition as described herein (or such a composition as used in a method) comprise barcode regions comprising at least 6 consecutive H’s, for example at least 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 consecutive H’s, including ranges between any two of the listed values as described herein. In some embodiments, at least 99% of the unique oligonucleotide species of the composition as described herein comprise barcode regions comprising at least 6 consecutive H’s, for example at least 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 consecutive H’s, including ranges between any two of the listed values as described herein. In some embodiments, at least 99.9% of the unique oligonucleotide species of the composition as described herein comprise barcode regions comprising at least 6 consecutive H’s, for example at least 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 consecutive H’s, including ranges between any two of the listed values as described herein.

[0052] In some embodiments, the sample barcode of each unique oligonucleotide species has a G content of 50% or less, for example 50% or less, 40% or less, 25% or less, 20% or less, 12.5% or less, 10% or less, or 5% or less, including ranges between any two of the listed values.

[0053] In some embodiments, the barcode region of each unique oligonucleotide species has a G content of 50% or less, for example 50% or less, 40% or less, 25% or less, 20% or less, 12.5% or less, 10% or less, or 5% or less, including ranges between any two of the listed values.

[0054] In some embodiments, the molecule barcodes of the unique oligonucleotide species collectively have a G content of less than 12.5%, for example less than 12.5%, less than 10%, less than 7.5%, less than 5%, less than 2.5%, or less than 1%, including ranges between any two of the listed values.

[0055] In some embodiments, the barcode regions of the unique oligonucleotide species collectively have a G content of less than 12.5%, for example less than 12.5%, less

than 10%, less than 7.5%, less than 5%, less than 2.5%, or less than 1%, including ranges between any two of the listed values.

[0056] In some embodiments, for at least 95% of the unique oligonucleotide species, any G in the molecule barcode is not adjacent to another G. In some embodiments, for at least 99% of the unique oligonucleotide species, any G in the molecule barcode is not adjacent to another G. In some embodiments, for all or substantially all of the unique oligonucleotide species, any G in the molecule barcode is not adjacent to another G.

[0057] Each barcode region can optionally comprise a sample barcode. In accordance with compositions, methods, and oligonucleotides of some embodiments herein, each unique oligonucleotide species in a pool can comprise the same sample barcode, but there can be two or more pools that are each associated with different sample barcodes. As such, all or essentially all of the unique oligonucleotide species in pool #1 can comprise sample barcode #1, and all or essentially all of the unique oligonucleotide species in pool #2 can comprise sample barcode #2. Nucleic acids from a first sample can be associated with the unique oligonucleotide species in pool #1, and nucleic acids from a second sample can be associated with unique oligonucleotide species in pool #2, for example by hybridization and amplification. As such, all or essentially all of the amplified nucleic acids corresponding to the first sample will comprise sample barcode #1 (but can comprise different molecule barcodes), and all of the amplified nucleic acids corresponding to the second sample will comprise sample barcode #2. In some embodiments, there are at least 24, 48, 96, or 192 pools.

[0058] The sample barcode can comprise a nucleic acid sequence of at least 3 nucleotides, for example at least 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 nucleotides, including ranges between any two of the listed values, for example 3-50, 3-45, 3-40, 3-35, 3-30, 3-25, 3-20, 3-15, 3-14, 3-13, 3-12, 3-11, 3-10, 3-9, 3-8, 3-7, 3-6, 3-5, 3-4, 4-50, 4-45, 4-40, 4-35, 4-30, 4-25, 4-20, 4-15, 4-14, 4-13, 4-12, 4-11, 4-10, 4-9, 4-8, 4-7, 4-6, 4-5, 5-50, 5-45, 5-40, 5-35, 5-30, 5-25, 5-20, 5-15, 5-14, 5-13, 5-12, 5-11, 5-10, 5-9, 5-8, 5-7, 5-6, 6-50, 6-45, 6-40, 6-35, 6-30, 6-25, 6-20, 6-15, 6-14, 6-13, 6-12, 6-11, 6-10, 6-9, 6-8, 6-7, 7-50, 7-45, 7-40, 7-35, 7-30, 7-25, 7-20, 7-15, 7-14, 7-13, 7-12, 7-11, 7-

10, 7-9, 7-8, 8-50, 8-45, 8-40, 8-35, 8-30, 8-25, 8-20, 8-15, 8-14, 8-13, 8-12, 8-11, 8-10, 8-9, 9-50, 9-45, 9-40, 9-35, 9-30, 9-25, 9-20, 9-15, 9-14, 9-13, 9-12, 9-11, 9-10, 10-50, 10-45, 10-40, 10-35, 10-30, 10-25, 10-20, 10-15, 10-14, 10-13, 10-12, or 10-11 nucleotides. In some embodiments, the nucleic acid sequence of the sample barcode comprises a unique sequence, for example, so that each unique oligonucleotide species in a population comprises a different molecule barcode.

[0059] Without being limited by any theory, it is contemplated that a sample barcode comprising a low G content (e.g., less than 50% G, for example, less than 45% G, 40% G, 35% G, 30% G, 25% G, 20% G, 15% G, 12.5% G, 10% G, 7.5% G, 5% G, or 2.5% G) can be positioned 3' of a sample barcode that comprises a relatively higher G content and 5' of a uniform region (e.g. a target-specific sequence or oligo dT sequence), so as to minimize bias by separating the relatively G-rich sample barcode from the uniform region. In some embodiments, the barcode region comprises a sample barcode comprising 50% G content or less, for example, less than 50% G, 45% G, 40% G, 35% G, 30% G, 25% G, 20% G, 15% G, 12.5% G, 10% G, 7.5% G, 5% G, or 2.5% G, or 0% G, including ranges between any two of the listed values, for example 2.5-50% G, 2.5-45% G, 2.5-40% G, 2.5-35% G, 2.5-30% G, 2.5-25% G, 2.5-20% G, 2.5-15% G, 2.5-10% G, 2.5-7.5% G, 2.5-5% G, 5-50% G, 5-45% G, 5-40% G, 5-35% G, 5-30% G, 5-25% G, 5-20% G, 5-15% G, 5-10% G, 5-7.5% G, 7.5-50% G, 7.5-45% G, 7.5-40% G, 7.5-35% G, 7.5-30% G, 7.5-25% G, 7.5-20% G, 7.5-15% G, 7.5-10% G, 10-50% G, 10-45% G, 10-40% G, 10-35% G, 10-30% G, 10-25% G, 10-20% G, 10-15% G, 10 – 12.5% G, 12.5-50% G, 12.5-45% G, 12.5-40% G, 12.5-35% G, 12.5-30% G, 12.5-25% G, 12.5-20% G, 12.5-15% G, 15-50% G, 15-45% G, 15-40% G, 15-35% G, 15-30% G, 15-25% G, 15-20% G, 20-50% G, 20-45% G, 20-40% G, 20-35% G, 20-30% G, or 20-25% G.

[0060] In some embodiments, for a composition comprising unique oligonucleotide species (or such a composition as used in a method), at least 95% of the sample barcodes of the unique oligonucleotides of the composition each have less than 50% G content, for example, less than 45% G, 40% G, 35% G, 30% G, 25% G, 20% G, 15% G, 12.5% G, 10% G, 7.5% G, 5% G, or 2.5% G, or 0% G, including ranges between any two of the listed values, for example 2.5-50% G, 2.5-45% G, 2.5-40% G, 2.5-35% G, 2.5-30% G,

2.5-25% G, 2.5-20% G, 2.5-15% G, 2.5-10% G, 2.5-7.5% G, 2.5-5% G, 5-50% G, 5-45% G, 5-40% G, 5-35% G, 5-30% G, 5-25% G, 5-20% G, 5-15% G, 5-10% G, 5-7.5% G, 7.5-50% G, 7.5-45% G, 7.5-40% G, 7.5-35% G, 7.5-30% G, 7.5-25% G, 7.5-20% G, 7.5-15% G, 7.5-10% G, 10-50% G, 10-45% G, 10-40% G, 10-35% G, 10-30% G, 10-25% G, 10-20% G, 10-15% G, 10 – 12.5% G, 12.5-50% G, 12.5-45% G, 12.5-40% G, 12.5-35% G, 12.5-30% G, 12.5-25% G, 12.5-20% G, 12.5-15% G, 15-50% G, 15-45% G, 15-40% G, 15-35% G, 15-30% G, 15-25% G, 15-20% G, 20-50% G, 20-45% G, 20-40% G, 20-35% G, 20-30% G, or 20-25% G.

[0061] In some embodiments, for a composition comprising unique oligonucleotide species (or such a composition as used in a method), at least 99% of the sample barcodes of the unique oligonucleotides of the composition each have 50% G content or less, for example, less than 50% G, 45% G, 40% G, 35% G, 30% G, 25% G, 20% G, 15% G, 12.5% G, 10% G, 7.5% G, 5% G, or 2.5% G, or 0% G, including ranges between any two of the listed values, for example 2.5-50% G, 2.5-45% G, 2.5-40% G, 2.5-35% G, 2.5-30% G, 2.5-25% G, 2.5-20% G, 2.5-15% G, 2.5-10% G, 2.5-7.5% G, 2.5-5% G, 5-50% G, 5-45% G, 5-40% G, 5-35% G, 5-30% G, 5-25% G, 5-20% G, 5-15% G, 5-10% G, 5-7.5% G, 7.5-50% G, 7.5-45% G, 7.5-40% G, 7.5-35% G, 7.5-30% G, 7.5-25% G, 7.5-20% G, 7.5-15% G, 7.5-10% G, 10-50% G, 10-45% G, 10-40% G, 10-35% G, 10-30% G, 10-25% G, 10-20% G, 10-15% G, 10 – 12.5% G, 12.5-50% G, 12.5-45% G, 12.5-40% G, 12.5-35% G, 12.5-30% G, 12.5-25% G, 12.5-20% G, 12.5-15% G, 15-50% G, 15-45% G, 15-40% G, 15-35% G, 15-30% G, 15-25% G, 15-20% G, 20-50% G, 20-45% G, 20-40% G, 20-35% G, 20-30% G, or 20-25% G. That is, within the population of unique oligonucleotides, less than 1% of the sample barcodes have G content greater than 50%. In some embodiments, the unique oligonucleotide species of the composition consist of or consists essentially of unique oligonucleotide species that each have sample barcodes having less than 50% G content, as described herein.

[0062] In some embodiments, a barcode region is positioned 5' of the uniform region without any intervening sequences between the barcode region and the uniform region. In some embodiments, the barcode region is positioned 5' of a spacer, which is positioned 5' of the the uniform region, so that a spacer is between the barcode region and the uniform

region. Optionally, the spacer has a low G content (e.g., 50% G or less, for example, less than 50% G, 45% G, 40% G, 35% G, 30% G, 25% G, 20% G, 15% G, 12.5% G, 10% G, 7.5% G, 5% G, or 2.5% G). The spacer can have a length of at least 1 nucleotide, for example 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, or 50 nucleotides, including ranges between any two of the listed values, for example 1-50, 1-45, 1-35, 1-30, 1-25, 1-20, 1-15, 1-10, 1-9, 1-8, 1-7, 1-6, 1-5, 1-4, 1-3, 1-2, 2-50, 2-45, 2-35, 2-30, 2-25, 2-20, 2-15, 2-10, 2-9, 2-8, 2-7, 2-6, 2-5, 2-4, 2-3, 3-50, 3-45, 3-35, 3-30, 3-25, 3-20, 3-15, 3-10, 3-9, 3-8, 3-7, 3-6, 3-5, 3-4, 4-50, 4-45, 4-35, 4-30, 4-25, 4-20, 4-15, 4-10, 4-9, 4-8, 4-7, 4-6, 4-5, 5-50, 5-45, 5-35, 5-30, 5-25, 5-20, 5-15, 5-10, 5-9, 5-8, 5-7, 5-6, 6-50, 6-45, 6-35, 6-30, 6-25, 6-20, 6-15, 6-10, 6-9, 6-8, 6-7, 7-50, 7-45, 7-35, 7-30, 7-25, 7-20, 7-15, 7-10, 7-9, 7-8, 8-50, 8-45, 8-35, 8-30, 8-25, 8-20, 8-15, 8-10, 8-9, 9-50, 9-45, 9-35, 9-30, 9-25, 9-20, 9-15, 9-10, 10-50, 10-45, 10-35, 10-30, 10-25, 10-20, or 10-15 nucleotides. In some embodiments, the spacer can comprise at least two consecutive non-G nucleotides (denoted as “H”, in which “H” is A, C, or T), for example at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 consecutive H’s, including ranges between any two of the listed values, for example 2-20, 2-15, 2-10, 2-9, 2-8, 2-7, 2-6, 2-5, 2-4, 2-3, 3-20, 3-15, 3-10, 3-9, 3-8, 3-7, 3-6, 3-5, 3-4, 4-20, 4-15, 4-10, 4-9, 4-8, 4-7, 4-6, 4-5, 5-20, 5-15, 5-10, 5-9, 5-8, 5-7, 5-6, 6-20, 6-15, 6-10, 6-9, 6-8, or 6-7. In some embodiments, the nucleotides of the spacers are Ts. For example, the space can comprise a poly-T sequence, such as TT, TTT, TTTT, TTTTT, TTTTTT, TTTTTTT, or TTTTTTTTTT. It noted that the spacers do not necessarily provide diversity, and as such, in a composition comprising unique oligonucleotide species as described herein, some or all of the unique oligonucleotide species can have the same spacer sequences. Optionally, all of the unique oligonucleotide species of the composition comprise the same spacer sequence.

Pools and pooling

[0063] Compositions comprising unique oligonucleotide species as described in accordance with compositions and methods of some embodiments herein can be disposed in spatially isolated pools, for example so that multiple sample can be analyzed, with one sample per pool. In some embodiments, the unique oligonucleotide species are disposed in

spatially isolated pools, each pool comprising a plurality of unique oligonucleotides of the unique oligonucleotide species, so that unique oligonucleotides in the same pool comprise the same sample barcode sequence, and so that different unique oligonucleotides of the same pool comprise a different molecule barcode sequences. As used herein, “spatial isolation” (and variations of this root term) means that target nucleic acids of a sample can hybridize to the unique oligonucleotide species of the pool without substantially cross-reacting with other pools, and without the unique oligonucleotide species of the pool substantially hybridizing to target nucleic acids of other samples. As such, the sample barcode can identify which pool a given unique oligonucleotide species came from. Moreover, upon barcoding target nucleic acid sequences with unique oligonucleotide species, the sample barcode can identify which pool a barcoded target nucleic acid sequence (or reverse transcript or amplicon thereof) came from.

[0064] In some embodiments, a substrate organizes the pools so that they are spatially isolated from each other. For example, a multi-well plate can organize spatially isolated pools, so that each pool is in a separate well. For example, each pool can be immobilized on a different bead. Optionally, each well of a multi-well plate can contain a single bead on which the pool of unique oligonucleotide species is immobilized, so that a different pool of unique oligonucleotide species is positioned in each well of the multi-well plate.

[0065] In some embodiments, the unique oligonucleotide species are disposed in at least 2 pools, for example, at least 2, 3, 4, 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, 60, 70, 72, 96, 99, 100, 110, 120, 144, 168, or 192, pools, including ranges between any two of the listed values. Optionally, there can be at least 100 unique oligonucleotides per pool.

[0066] In some embodiments, the unique oligonucleotide species are disposed in at least 2 pools, and there are at least 100 unique oligonucleotides per pool for example at least 100, 200, 300, 400, 500, 1000, 2000, 3000, 4000, 5000, 6000, 6500, 10,000, or 65,000 unique oligonucleotide species per pool.

[0067] In some embodiments, the unique oligonucleotide species are disposed in at least 24 pools, and there are at least 100 unique oligonucleotides per pool for example at least 100, 200, 300, 400, 500, 1000, 2000, 3000, 4000, 5000, 6000, 6500, 10,000, or 65,000 unique oligonucleotide species per pool.

[0068] In some embodiments, the unique oligonucleotide species are disposed in at least 48 pools, and there are at least 100 unique oligonucleotides per pool for example at least 100, 200, 300, 400, 500, 1000, 2000, 3000, 4000, 5000, 6000, 6500, 10,000, or 65,000 unique oligonucleotide species per pool.

[0069] In some embodiments, the unique oligonucleotide species are disposed in at least 96 pools, and there are at least 100 unique oligonucleotides per pool for example at least 100, 200, 300, 400, 500, 1000, 2000, 3000, 4000, 5000, 6000, 6500, 10,000, or 65,000 unique oligonucleotide species per pool.

Substrates

[0070] Unique oligonucleotide species in accordance with compositions and methods of some embodiments herein can be immobilized on substrates, for example beads, wells of multi-well plates, or arrays. For example, unique oligonucleotide species having the same sample barcode, but different molecule barcodes can all be immobilized on a substrate, such as on a single bead, or in a single well of a multi-well plate. As such, when the immobilized unique oligonucleotide species of a particular substrate are contacted with a particular sample, such as a single cell, the immobilized unique oligonucleotide species immobilized on that substrate will be associated with target nucleic acids of the same sample. If the unique oligonucleotide species immobilized on the same substrate have the same sample barcode, while those on other substrates have different sample barcodes, the sample associated with each target nucleic acid can be readily identified (e.g. all of the unique oligonucleotide species on substrate #1 have sample barcode #1, all of the unique oligonucleotide species on substrate #2 have sample barcode #2).

[0071] In some embodiments, such configurations in which unique oligonucleotide species are immobilized on a substrate can be helpful in facilitating efficiency and throughput. For example, single cells can be added to wells of a multi-well

plate, so that no more than one single cell is in each well. If the unique oligonucleotide species immobilized within a given well (for example, on the well or on a bead) comprise a unique sample barcode, barcoded target nucleic acids and reverse-transcripts and amplicons thereof corresponding to the single cell can be identified and quantified, even if barcoded reverse transcripts or amplicons from multiple samples are pooled.

[0072] A substrate can comprise a type of solid support. A substrate can refer to a continuous solid or semi-solid surface on which the methods of the disclosure may be performed. A substrate can refer to an array, a cartridge, a chip, a device, and a slide, for example. As such, “solid support” and “substrate” can be used interchangeably.

[0073] A substrate or solid support in accordance with some embodiments herein can encompass any type of solid, porous, or hollow sphere, ball, bearing, cylinder, or other similar configuration composed of plastic, ceramic, metal, or polymeric material (e.g., hydrogel) onto which a nucleic acid may be immobilized (e.g., covalently or non-covalently). A substrate or solid support can comprise a discrete particle that may be spherical (e.g., microspheres) or have a non-spherical or irregular shape, such as cubic, cuboid, pyramidal, cylindrical, conical, oblong, or disc-shaped, and the like. A plurality of solid supports spaced in an array may not comprise a substrate. A solid support may be used interchangeably with the term “bead.”

[0074] In some embodiments, a plurality of substrates is provided, and the diversity of sample barcodes can be represented in the plurality of substrates, so that unique substrates can comprise unique sample barcodes.

Methods of barcoding

[0075] In accordance with some embodiments herein, methods of specifically barcoding nucleic acids from two or more samples are described. Each sample can comprise nucleic acids. The method can comprise contacting each sample with a pool comprising a plurality of unique oligonucleotide species, as described herein. Each sample can be contacted in spatial isolation from the other samples. The unique polynucleotide species of each pool can comprise the same sample barcode and comprise different molecule barcodes. The method can include hybridizing target specific regions of at least some oligonucleotides

of the unique oligonucleotide species to at least some of the nucleic acids of the sample. The method can include extending the hybridized oligonucleotides so as to produce strands comprising an oligonucleotide of the unique oligonucleotide species and a sequence complementary to the target region. Thus, for each sample, the produced strands can comprise the same sample barcode and different molecule barcodes. For different samples, the sample barcodes can be different. In some embodiments, the unique oligonucleotide species contacted with the sample consist of or consist essentially of unique oligonucleotide species wherein each molecule barcode has a G content of less than 50%. In some embodiments, the molecule barcodes of all of the unique oligonucleotide species collectively have a G content of no more than 12.5%. Optionally each pool comprises at least 100 unique oligonucleotide species, for example at least 100, 500, 100, 500, 1000, 2000, 6500, or 65,000 unique oligonucleotide species. Such methods in accordance with some embodiments herein can barcode nucleic acids of two or more different samples, such as single-cell sample, so as to permit analysis and quantification of the nucleic acids of a single cell.

[0076] Optionally, each sample is contacted with a single pool of unique oligonucleotide species in spatial isolation from the other pools and samples. In some embodiments, the pools are spatially isolated by being in unique contained spaces, for example different wells of a multi-well plate, different test tubes, different microfluidic channels, or the like. In some embodiments, the pools are spatially isolated by being on different regions of a surface, for example reaction dots on an array.

[0077] The oligonucleotides can be contacted with target nucleic acids of the sample under standard hybridization conditions (e.g., in standard buffers, and at a temperature lower than the T_m of a portion of the uniform region that is complementary to the target nucleic acid). Optionally, for example if the target nucleic acids comprise RNAs, the target nucleic acids of the sample are reverse-transcribed after hybridization to the unique oligonucleotide species, so as to generate DNAs (e.g. cDNAs). Optionally, the products of the hybridization or reverse transcription reaction are amplified, so as to generate a library of DNAs. The amplification can be via PCR under standard conditions, or via other suitable methods, for example isothermal amplification, rolling circle amplification, and the like. Optionally, the amplification can comprise amplifying reverse-transcripts or hybridized

nucleic acids with a polymerase having 5' to 3' activity. Optionally, the products of the amplification can be further analyzed, for example by sequence. Optionally, reverse transcripts or amplicons from different pools can be pooled for sequencing, as different sample barcodes will indicate the corresponding pool (and sample) of each individual barcoded nucleic acid. It is noted that there may be bias in favor of G-rich molecule barcode or unique oligonucleotide species, and methods in accordance with some embodiments herein can minimize or eliminate such bias. By way of example, when quantifying nucleic acids of the sample at the single nucleic acid level, it can be useful to minimize or eliminate such bias that could favor representation of some nucleic acids and skew quantitative results in favor of certain kinds of barcodes rather than representative target nucleic acids of the sample. In some embodiments, methods comprising unique oligonucleotide species as described herein reduce bias by increasing sensitivity, decreasing relative standard error, or increasing sensitivity and reducing standard error.

[0078] Optionally, the method can further comprise ascertaining nucleic acid sequences of the strands comprising the oligonucleotides of the unique oligonucleotide species and the sequence complementary to the target. For example, reverse transcripts, amplicons, and/or cDNA libraries generated from the hybridization and extension of the unique oligonucleotide probes can be sequenced. Any suitable sequencing method known in the art can be used, preferably high-throughput approaches. For example, cyclic array sequencing using platforms such as Roche 454, Illumina Solexa, ABI-SOLiD, ION Torrent, Complete Genomics, Pacific Bioscience, Helicos, or the Polonator platform, may also be utilized. Sequencing may comprise MiSeq sequencing. Sequencing may comprise HiSeq sequencing. If applicable, the unique oligonucleotide species can comprise adapters to mediate high-throughput sequencing, for example universal priming sites. Optionally, barcoded nucleic acids corresponding to two or more different samples (e.g. from two or more different pools) can be pooled or combined for sequencing. Optionally, all of the pools are combined or pooled for sequencing. Without being limited by any theory, it is noted that sample barcodes can identify the corresponding sample (or pool) of the combined barcoded nucleic acids, that combining or pooling the nucleic acids can facilitate throughput and/or resource utilization for sequencing, and that accurate barcoding in accordance with some

embodiments herein can minimize or eliminate bias, even when a large number of different nucleic acids are analyzed.

[0079] In some embodiments, the barcode region of each pool comprises a sample barcode comprising at least 3 nucleotides as described herein, for example at least 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, or 50 nucleotides, including ranges between any two of the listed values. Optionally, the sample barcode is positioned 3' of the molecule barcode. Optionally, the molecule barcode is positioned 3' of the sample barcode. Optionally, the unique oligonucleotide species comprise the same sample barcode, which can be a different sample barcode from the unique oligonucleotide species of other pools. Optionally, the sample barcode has a low G content as described herein, for example less than 50% G, for example, less than 45% G, 40% G, 35% G, 30% G, 25% G, 20% G, 15% G, 12.5% G, 10% G, 7.5% G, 5% G, or 2.5% G. Optionally, at least 95%, 99%, or 99.9% of the sample barcodes have a low G content as described herein. Optionally, each unique oligonucleotide comprises a spacer 3' of the molecule barcode and 5' of the uniform region as described herein.

[0080] In some embodiments, the unique oligonucleotide species contacted with the sample consist essentially of unique oligonucleotide species in which each molecule barcode has a G content of less than 50% as described herein, for example, less than 45% G, 40% G, 35% G, 30% G, 25% G, 20% G, 15% G, 12.5% G, 10% G, 7.5% G, 5% G, or 2.5% G content, or 0% G content. In some embodiments, the molecule barcodes of at least 95% of the unique oligonucleotide species contacted with the sample comprise molecule barcodes having a G content of less than 50% as described herein, for example, less than 45% G, 40% G, 35% G, 30% G, 25% G, 20% G, 15% G, 12.5% G, 10% G, 7.5% G, 5% G, or 2.5% G content, or 0% G content. In some embodiments, the molecule barcodes of at least 99% of the unique oligonucleotide species contacted with the sample comprise molecule barcodes having a G content of less than 50% as described herein, for example, less than 45% G, 40% G, 35% G, 30% G, 25% G, 20% G, 15% G, 12.5% G, 10% G, 7.5% G, 5% G, or 2.5% G content, or 0% G content. In some embodiments, all of the molecule barcodes of the unique oligonucleotide species contacted with the sample comprise molecule barcodes having a G

content of less than 50% as described herein, for example, less than 45% G, 40% G, 35% G, 30% G, 25% G, 20% G, 15% G, 12.5% G, 10% G, 7.5% G, 5% G, or 2.5% G content, or 0% G content.

[0081] In some embodiments, the molecule barcodes of all of the unique oligonucleotide species collectively have a G content of less than 50% G, for example, less than 45% G, 40% G, 35% G, 30% G, 25% G, 20% G, 15% G, 12.5% G, 10% G, 7.5% G, 5% G, or 2.5% G, or 0% G, including ranges between any two of the listed values. In some embodiments, the molecule barcodes of all of the unique oligonucleotide species collectively have a G content of no more than 12.5%.

[0082] In some embodiments, each unique oligonucleotide species further comprises a uniform region 3' of the barcode region. Optionally, the uniform region comprises at least 10 nucleotides complementary to a target nucleic acid sequence (e.g., at least 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 nucleotides complementary to the target, including ranges between any two of the listed values), and configured for 5' to 3' amplification of the target nucleic acid sequence. Optionally, the uniform region comprises an oligo dT sequence.

[0083] In some embodiments, the unique oligonucleotide species of each pool are immobilized on a substrate as described herein, so that the unique oligonucleotide species immobilized on a given substrate comprise the same sample barcode, and different unique oligonucleotide species immobilized on the substrate comprise different molecule barcodes. By way of example, substrates can include wells in multi-well plates (e.g. 24-, 48, and 96-well plates), spots on an array, beads, and the like. In some embodiments, at least 100 unique oligonucleotides are immobilized on each substrate. Optionally, all of the unique oligonucleotide species immobilized on a given substrate (and thus in the same pool) comprise the same sample barcode. In some embodiments, each sample barcode has a G content of 50% or less, for example 50% or less, 40% or less, 25% or less, 20% or less, 12.5% or less, 10% or less, or 5% or less.

[0084] In some embodiments, the molecule barcodes of the unique oligonucleotide species collectively have a G content of less than 12.5%, for example less than 12.5%, less than 10%, less than 7.5%, less than 5%, less than 2.5%, or less than 1%.

[0085] In some embodiments, the barcode regions of the unique oligonucleotide species collectively have a G content of no more than 12.5%, for example no more than 12.5%, no more than 10%, no more than 7.5%, no more than 5%, no more than 2.5%, or no more than 1%.

[0086] In some embodiments, for at least 95% of the unique oligonucleotide species, any G in the molecule barcode is not adjacent to another G. For example, for at least 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 99.99%, of the unique oligonucleotide species, any G in the molecule barcode can be not-adjacent to another G.

[0087] In some embodiments, each pool consists of, or consists essentially of unique oligonucleotide species for which any G in the molecule barcode is not adjacent to another G.

[0088] In some embodiments, at least 95% of the molecule barcodes of the unique oligonucleotide species comprise a sequence totaling at least 6 alternating H's and N's, for example at least 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 alternating H's and N's, including ranges between any two of the listed values. By way of example, 6 alternating H's and N's can be depicted as HNHNNHN. It is noted that each H can be the same as, or different from any other H, and that each N can be the same as, or different from any other N.

[0089] In some embodiments, at least 95% of the molecule barcodes of the unique oligonucleotide species comprise the sequence HNHNNHNHN, in which each "H" is any one of A, C, or T, and wherein each "N" is any one of A, G, C, or T. For example at least 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 99.99% of the molecule barcodes of the unique oligonucleotide species can comprise the sequence HNHNNHNHN. It is noted that each H can be the same as, or different from any other H, and that each N can be the same as, or different from any other N. In some embodiments, each molecule barcode of the unique oligonucleotide species comprises the sequence HNHNNHNHN.

[0090] In some embodiments, at least 99% of the molecule barcodes of the unique oligonucleotide species comprise the sequence HHHHHHHHH, in which each "H" is any one of A, C, or T. For example at least 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 99.99% of the molecule barcodes of the unique oligonucleotide species can comprise the sequence HHHHHHHHH. It is noted that each H can be the same as, or different from any other H. In

some embodiments, each molecule barcode of the unique oligonucleotide species comprises the sequence HHHHHHHH.

[0091] In some embodiments, each unique oligonucleotide comprises a spacer 3' of the barcode region and 5' of the target specific region. Optionally, the spacer can comprise the sequence HNH, in which each "H" is any one of A, C, or T, and in which each "N" is any one of A, G, C, or T. Optionally, the spacer can comprise the sequence HHH. Optionally, the spacer can comprise the sequence HNHNHNH. Optionally, the spacer can comprise the sequence HHHHHHHH. It is noted that each H can be the same as, or different from any other H, and that each N can be the same as, or different from any other N. In some embodiments, each unique oligonucleotide comprises a spacer 3' of the barcode region and 5' of the target specific region, said spacer comprising the sequence HNHNHNH, in which each "H" is any one of A, C, or T, and in which each "N" is any one of A, G, C, or T. In some embodiments, unique oligonucleotide comprises a spacer 3' of the barcode region and 5' of the target specific region, said spacer comprising the sequence HHHHHHHH, in which each "H" is any one of A, C, or T.

[0092] In some embodiments, at least one pool comprises at least two oligonucleotides of the same unique oligonucleotide species. In some embodiments, no pool comprises two oligonucleotides of the same unique oligonucleotide species.

[0093] In some embodiments, the target specific region can comprise a sequence for 5' to 3' amplification of a target nucleic acid. In some embodiments, the target specific region comprises an oligo dT sequence. In some embodiments, the target specific region comprises a sequence flanking a sequence encoding a variable region of an immune cell receptor or immunoglobulin.

[0094] In some embodiments, for each unique oligonucleotide species, the molecule barcode is 3' of the sample barcode. In some embodiments, for each unique oligonucleotide species, the sample barcode is 3' of the molecule barcode.

[0095] In some embodiments, each unique oligonucleotide species has a length of at least 24 nucleotides. In some embodiments, each unique oligonucleotide species has a length of 24-140 nucleotides.

[0096] In some embodiments, each pool comprises at least 1000 unique oligonucleotide species. In some embodiments, each pool comprises at least 6,500 unique oligonucleotide species. In some embodiments, each pool comprises at least 65,000 unique oligonucleotide species.

[0097] In some embodiments, at least 48 unique samples are each contacted with a different pool of unique oligonucleotide species. For example, at least 48, 72, 96, 120, 144, 168, or 192 samples can each be contacted with a different pool. For example each sample can be contacted with a different pool of unique oligonucleotide species in a well of a multi-well plate.

[0098] In some embodiments, at least 99% of the samples comprise no more than one cell each. For example, at least 99%, 99.5%, 99.9%, or 99.99% of the samples can comprise no more than one cell. By way of example, a solution comprising multiple cells can be diluted to an appropriate concentration so that there is a suitably high probability that each sample comprises no more than one cell.

[0099] In some embodiments, the unique oligonucleotide species of each pool are immobilized on a substrate, so that the sample barcodes but not the molecule barcodes are the same for the unique oligonucleotide species immobilized on each substrate. As such, each pool can be identified by a particular sample barcode. It is noted that different substrates (and as such, different pools) can be associated with different sample barcodes. In some embodiments, the substrate comprises a spatially isolated region of a surface, so that the substrates of different pools comprise the different spatially isolated regions of the surface. In some embodiments, the substrate comprises a well in a multi-well plate. In some embodiments, the substrate comprises a bead.

[0100] In some embodiments, the unique oligonucleotide species further comprises an adapter configured to immobilize the unique oligonucleotide on the substrate, wherein said barcode region is 3' of the adapter. Optionally, each adapter can comprise a universal priming site, for example for use in sequencing.

Methods of making compositions comprising unique oligonucleotide species

[0101] In accordance with some embodiments herein, methods of making a composition comprising unique oligonucleotides are described. The method can comprise providing a plurality of different sample barcodes as described herein. The method can comprise providing a plurality of different molecule barcodes as described herein. The method can comprise synthesizing a plurality of unique oligonucleotide species as described herein, in which each unique oligonucleotide species comprises a barcode region comprising a sample barcode and a molecule barcode as described herein. The method can comprise disposing the plurality of unique oligonucleotide species in spatially-isolated pools. Each pool can comprise multiple unique oligonucleotide species, so that the unique oligonucleotide species of the same pool comprise the same sample barcode sequence, and so that different unique oligonucleotide species of the same pool comprise different molecule barcode sequences. Optionally, the unique oligonucleotide species are disposed in the spatially-isolated pools concurrent with synthesis. Optionally, the unique oligonucleotide species are disposed in the spatially-isolated pools after synthesis. Optionally, the composition consists essentially of unique oligonucleotide species in which each molecular barcode has a G content of less than 50%. Optionally, the molecule barcodes of all of the unique oligonucleotide species collectively have a G content of no more than 12.5%. Optionally, the composition consists essentially of unique oligonucleotide species in which each molecular barcode has a G content of less than 50 and the molecule barcodes of all of the unique oligonucleotide species collectively have a G content of no more than 12.5%. In some embodiments, each spatially-isolated pool comprises at least 100 unique oligonucleotide species, for example, at least 100, 200, 300, 400, 500, 1000, 2000, 3000, 5000, 6500, or 65,000 unique oligonucleotide species. In some embodiments, each spatially-isolated pool comprises at least 1000 unique oligonucleotide species.

[0102] The unique oligonucleotides can be synthesized using any of a number of suitable methods. In some embodiments, the sample barcode sequences and molecule barcode sequences are generated *in silico*, and unique oligonucleotide species comprising the sample barcode sequences and molecule barcode sequences together are synthesized, for example using chemical oligonucleotide synthesis. In some embodiments, a plurality of oligonucleotides comprising the same sample barcode sequence are pooled in spatial

isolation, and joined with a plurality of oligonucleotides comprising molecule barcodes, for example via hybridization and extension, or via ligation. Such an approach can be performed in multiple spatially isolated environments in parallel, or sequentially, to achieve multiple unique oligonucleotide species.

[0103] Optionally, each unique oligonucleotide species further comprises a uniform region 3' of the barcode region as described herein. Optionally, the uniform region comprises a target-specific region. In some embodiments, the target-specific region comprises at least 10 nucleotides complementary to a target nucleic acid (e.g., at least 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 nucleotides complementary to the target, including ranges between any two of the listed values), and configured for 5' to 3' amplification of the target nucleic acid sequence. Optionally, the uniform region comprises an oligo dT sequence.

[0104] In some embodiments, the plurality of molecule barcodes (and as such, the plurality of unique oligonucleotide species) consists essentially of unique oligonucleotide species in which each molecular barcode has a G content of no more than 50% as described herein, for example, less than 50% G, 45% G, 40% G, 35% G, 30% G, 25% G, 20% G, 15% G, 12.5% G, 10% G, 7.5% G, 5% G, or 2.5% G content, or 0% G content. In some embodiments, the molecule barcodes of at least 95% of the unique oligonucleotide species have a G content of no more than 50% as described herein, for example, less than 50% G, 45% G, 40% G, 35% G, 30% G, 25% G, 20% G, 15% G, 12.5% G, 10% G, 7.5% G, 5% G, or 2.5% G content, or 0% G content. In some embodiments, the molecule barcodes of at least 99% of the unique oligonucleotide species have a G content of no more than 50% as described herein, for example, less than 50% G, 45% G, 40% G, 35% G, 30% G, 25% G, 20% G, 15% G, 12.5% G, 10% G, 7.5% G, 5% G, or 2.5% G content, or 0% G content. In some embodiments, all of the molecule barcodes of the unique oligonucleotide species have molecule barcodes having a G content of no more than 50% as described herein, for example, less than 50% G, 45% G, 40% G, 35% G, 30% G, 25% G, 20% G, 15% G, 12.5% G, 10% G, 7.5% G, 5% G, or 2.5% G content, or 0% G content. In some embodiments, each molecule barcode of each unique oligonucleotide species has a G content of no more than 50%, for

example, less than 50% G, 45% G, 40% G, 35% G, 30% G, 25% G, 20% G, 15% G, 12.5% G, 10% G, 7.5% G, 5% G, or 2.5% G content, or 0% G content.

[0105] In some embodiments, the molecule barcodes of all of the unique oligonucleotide species in the plurality of unique oligonucleotide species collectively have a G content of no more than 50% G, for example, less than 50%, 45% G, 40% G, 35% G, 30% G, 25% G, 20% G, 15% G, 12.5% G, 10% G, 7.5% G, 5% G, or 2.5% G, or 0% G, including ranges between any two of the listed values. In some embodiments, the molecule barcodes of all of the unique oligonucleotide species in the plurality of unique oligonucleotide species collectively have a G content of no more than 12.5%.

[0106] In some embodiments, each sample barcode has a G content of 50% or less as described herein, for example, less than 50%, 45% G, 40% G, 35% G, 30% G, 25% G, 20% G, 15% G, 12.5% G, 10% G, 7.5% G, 5% G, or 2.5% G content, or 0% G content.

[0107] In some embodiments, the sample barcodes of all of the unique oligonucleotide species in the plurality of unique oligonucleotide species collectively have a G content of no more than 50% G, for example, less than 50%, 45% G, 40% G, 35% G, 30% G, 25% G, 20% G, 15% G, 12.5% G, 10% G, 7.5% G, 5% G, or 2.5% G, or 0% G, including ranges between any two of the listed values. In some embodiments, the sample barcodes of all of the unique oligonucleotide species in the plurality of unique oligonucleotide species collectively have a G content of no more than 12.5%.

[0108] In some embodiments, for at least 95% (e.g., at least 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 99.99%) of the of the unique oligonucleotide species, any G in the molecule barcode is not adjacent to another G. In some embodiments, the plurality of unique oligonucleotide species consists essentially of unique oligonucleotide species for which any G in the molecule barcode is not adjacent to another G.

[0109] In some embodiments, at least 95% of the molecule barcodes of the unique oligonucleotide species comprise the sequence HNHNNHNH, in which each "H" is any one of A, C, or T, and in which each "N" is any one of A, G, C, or T. For example at least 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 99.99% of the molecule barcodes of the unique oligonucleotide species can comprise the sequence HNHNNHNH. It is noted that each H can be the same as, or different from any other H, and that each N can be the same as, or different

from any other N. In some embodiments, each molecule barcode of the unique oligonucleotide species comprises the sequence HNHNNHNH.

[0110] In some embodiments, at least 99% of the molecule barcodes of the unique oligonucleotide species comprise molecule barcodes comprising the sequence HHHHHHHH, in which each “H” is any one of A, C, or T. For example at least 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 99.99% of the molecule barcodes of the unique oligonucleotide species can comprise the sequence HHHHHHHH. It is noted that each H can be the same as, or different from any other H. In some embodiments, each molecule barcode of the unique oligonucleotide species comprises the sequence HHHHHHHH.

[0111] In some embodiments, each unique oligonucleotide species comprises a spacer 3’ of the barcode region and 5’ of the target specific region. Optionally, the spacer can comprise the sequence HNH, in which each “H” is any one of A, C, or T, and in which each “N” is any one of A, G, C, or T. Optionally, the spacer can comprise the sequence HHH. Optionally, the spacer can comprise the sequence HNHNNHNH. Optionally, the spacer can comprise the sequence HHHHHHHH. It is noted that each H can be the same as, or different from any other H, and that each N can be the same as, or different from any other N. In some embodiments, each unique oligonucleotide species comprises a spacer 3’ of the barcode region and 5’ of the target specific region, said spacer comprising the sequence HNHNNHNH, in which each “H” is any one of A, C, or T, and in which each “N” is any one of A, G, C, or T. In some embodiments, unique oligonucleotide comprises a spacer 3’ of the barcode region and 5’ of the target specific region, the spacer comprising the sequence HHHHHHHH, in which each “H” is any one of A, C, or T.

[0112] In some embodiments, at least one pool comprises least two oligonucleotides of the same unique oligonucleotide species. In some embodiments, no pool comprises two oligonucleotides of the same unique oligonucleotide species.

[0113] In some embodiments, the target-specific region can comprise a sequence for 5’ to 3’ amplification of a target nucleic acid. In some embodiments, the target specific region comprises an oligo dT sequence. In some embodiments, the target specific region comprises a sequence flanking a sequence encoding a variable region of an immune cell receptor or immunoglobulin.

[0114] In some embodiments, for each unique oligonucleotide species, the molecule barcode is 3' of the sample barcode. In some embodiments, for each unique oligonucleotide species, the sample barcode is 3' of the molecule barcode.

[0115] In some embodiments, each unique oligonucleotide species has a length of at least 24 nucleotides. In some embodiments, each unique oligonucleotide species has a length of 24-140 nucleotides.

[0116] In some embodiments, each pool comprises at least 6,500 unique oligonucleotide species. In some embodiments, each pool comprises at least 65,000 unique oligonucleotide species.

[0117] In some embodiments, the method further comprises immobilizing the unique oligonucleotide species of each spatially-distinct pool on a substrate, so that the sample barcodes but not the molecule barcodes are the same for the unique oligonucleotide species immobilized on each substrate. As such, each pool can be identified by a particular sample barcode. It is noted that different substrates (and as such, different pools) can be associated with different sample barcodes. In some embodiments, the substrate comprises a discrete region of a surface, so that the substrates of different pools comprise the different discrete regions of the surface. In some embodiments, the substrate comprises a well in a multi-well plate. In some embodiments, the substrate comprises a bead. In some embodiments, the unique oligonucleotide species are immobilized on the substrate by a covalent bond. In some embodiments, the unique oligonucleotide species are immobilized on the substrate by a magnetic or electromagnetic force. In some embodiments, the unique oligonucleotide species are imbedded in the substrate, so as to immobilize them on the substrate.

[0118] In some embodiments, at least 48 spatially-distinct pools are made, for example at least 48, 72, 96, 120, 144, 168, or 192 spatially-distinct pools.

[0119] In some embodiments, the unique oligonucleotide species are disposed in spatially-isolated pools concurrent with said synthesis. In some embodiments, the unique oligonucleotide species are disposed in spatially-isolated pools after said synthesis.

Oligonucleotides

[0120] In some embodiments, an oligonucleotide comprising a barcode region 3' of the adapter region is described. The barcode region can comprise a molecule barcode as described herein, in which the molecule barcode has a G content of no more than 50%. The molecule barcode can comprise at least 7 nucleotides.

[0121] In some embodiments, the oligonucleotide further comprises a sample barcode as described herein. The sample barcode can comprise at least 3 nucleotides.

[0122] In some embodiments, the oligonucleotide further comprises a uniform region 3' of the barcode region. In some embodiments, the uniform region comprises a target-specific region comprising at least 10 nucleotides complementary to a target nucleic acid as described herein. In some embodiments, the uniform region comprises an oligo dT sequence. In some embodiments, the oligonucleotide further comprises an adapter region 5' of the barcode region. In some embodiments, the oligonucleotide further comprises a spacer 3' of the barcode region and 5' of the uniform region as described herein.

Alternative embodiments

[0123] In some embodiments, designs for accurate and unbiased barcoding of nucleic acid transcripts are provided. These primer designs are composed of nucleic acid sequences (DNA, RNA, or LNA) that can target specific nucleic acid (DNA/RNA) transcripts via a gene-specific approach, or target a large set of nucleic acid transcripts (e.g. via poly-A tail of messenger RNAs or other consensus sequences) (**FIGs. 1A-C**). Along with the target sequences, these primers include two categories of nucleic acid barcodes: 1) molecule barcodes, also called "molecular barcodes," Molecular Index (MI) or Unique Molecular Identifier (UMI) and 2) sample barcodes, also called Sample Index (SI) (**FIGs. 1A-C**). The MI and SI barcodes are used to uniquely label for identification of its original target transcript via the MI, and its sample origin (SI) after library preparation for next generation sequencing or other sequence readout and barcode readout methods. Some embodiments detail different arrangements of MI and SIs within the primer that ensure unbiased and random labelling of barcodes to target transcripts for accurate gene expression profiling (*see, e.g.* **FIGs. 3A-3H**) (see Fu et al. 2011 PNAS 108: 9026-9032, hereby incorporated by reference in its entirety for discussion of random labelling).

[0124] In some embodiments, designs unique primer designs that can reduce ‘barcode bias,’ which is a phenomenon that is observed in the prior art designs, whereby with barcodes with high G-nucleotide content is more oftenly used compared to other barcode compositions (**FIG. 4**). In certain cases where conventional primers are used to target and label mRNA transcripts using a gene-specific sequence, MIs with high % G content and multiple G’s are preferably observed (**FIG. 4**). Moreover, without being limited by any theory, in conventional designs, SIs associated with no target genes observes small amounts of PCR errors due to these primers cross reacting during a sample multiplex PCR (**FIGs. 4-5**). These crossover events often happen with primers with high % G content and multiple G’s, possibly due unique secondary structure called “G-Quadplexes” in G-rich oligonucleotides. Because these G-rich MI/SI primers can form inter- and intra-molecular complexes with one another, conventional DNA cleanup steps to remove primers (e.g. Ampure XP beads) may not be sufficient. To minimize these events from happening, this invention entails several designs that reduces primer G content to minimize ‘barcode bias’ and the primer crossover are provided (**FIGs. 3A-3H**).

[0125] **FIGs. 3A-3H** show the design of prior art barcode layout (**FIG. 3A**) and novel designs to minimize bias as described (**FIGs. 3B-H**). In **FIG. 3A**, the conventional design allows for either the MI composition of ‘BBBBBBV’ or ‘NNNNNNNN’ (B is C,G, or T; V is A,C, or G; N is A,T,G, or C). A subset of the prior design MI barcode would be G-rich. In panel 3B-C, addition of either TTT or TTTTT after the MI to reduce G-richness along the region (length of non-G spacers can be variable, and can be any non-G nucleotide). In **FIG. 3D**, SI and MI positions are switched, so that high G-rich MIs would not be adjacent to potentially G-rich gene specific target regions. In **FIGs. 3E-F**, addition of either TTT or TTTTT to **FIG. 3D** design to further reduce G-rich regions (note that SIs are not G-rich, and the length of non-G spacers can be variable and can be any non-G nucleotide). In **FIGs. 3G-H** designs, MIs are designed to not be G-rich by using ‘HHHHHHHH’ or ‘HNHNHNHN,’ where H is A, C, or T.

[0126] The G-rich drawbacks of conventional designs are most observed gene-specific targeting method is used. Some embodiments include the use of the new primer designs in gene-specific target panels, such as T cell receptors (TCR), where primers are used

to target TCR-specific sequences. Designs in accordance with some embodiments herein can alleviate SI barcode crossover, thereby reducing the ‘noise’ of a TCR molecular barcoding assay.

Example 1: Accurate analysis of sample target nucleic acids using unique oligonucleotide species

[0127] 96 formats of unique oligonucleotide species were designed, the sequences of which are summarized in **FIGs. 5A-5E** (SEQ ID NOs: 1-96). Unique oligonucleotide species formats A01-A12 (SEQ ID NOs: 1-12) were provided as controls, in which the molecule barcode (“2nd bc”) was NNNNNNNN. As such, the molecule barcodes of these control unique oligonucleotide species did not have any constraints or limits on the G content. Unique oligonucleotide species formats B01-B12 (SEQ ID NOs: 13-24) comprised a spacer of the sequence TTT 3’ of the molecule barcode (“2nd bc”) and 5’ of the uniform region (“5 to 3’ recognition sequence”). Unique oligonucleotide species formats C01-C12 (SEQ ID NOs: 25-36) comprised a spacer of the sequence TTTTT 3’ of the molecule barcode (“2nd bc”) and 5’ of the uniform region (“5 to 3’ recognition sequence”). Unique oligonucleotide species formats D01-D12 (SEQ ID NOs: 37-48) comprised a molecule barcode (“1st bc”) 5’ of the sample barcode (“2nd bc”). Unique oligonucleotide species formats E01-E12 (SEQ ID NOs: 49-60) comprised a molecule barcode (“1st bc”) 5’ of the sample barcode (“2nd bc”) and a spacer of the sequence TTT 3’ of the barcode region and 4’ of the uniform region. Unique oligonucleotide species formats F01-F12 (SEQ ID NOs: 61-72) comprised a molecule barcode (“1st bc”) 5’ of the sample barcode (“2nd bc”) and a spacer of the sequence TTTTT 3’ of the barcode region and 4’ of the uniform region. Unique oligonucleotide species formats G01-G12 (SEQ ID NOs: 73-84) comprised molecule barcodes (“2nd bc”) each comprising the sequence HHHHHHHH (in which each “H” is an A, C, or T, and in which any two “H” nucleotides can be the same or different). Unique oligonucleotide species formats H01-H12 (SEQ ID NOs: 85-96) comprised molecule barcodes (“2nd bc”) each comprising the sequence HNHNNHN (in which each “H” is an A, C, or T, and in which any two “H” nucleotides can be the same or different, and in which

each “N” can be A, G, C, or, T, and in which any two “N” nucleotides can be the same or different).

[0128] Amplification and analysis of target T cell receptor nucleic acid sequences was performed as follows: the Precise™ TCR Encoding Plate (BD Cellular Research Inc.) was thawed at room temperature and spun briefly to collect the 5 μ L reagents in each well. The plate was placed on ice in a 96 well rack and very carefully remove the seal, while avoiding disturbing reagents in the wells. Cells were sorted or added directly into the 96 well encoding plate (1 cell per well). The plate was sealed and vortexed briefly (5-10s) to mix the reagents, and then spun briefly (~1000 rpm x 10 s). The sorted sample plates can optionally be stored at -80°C until ready to start the Precise™ protocol. For reverse transcription, the plate was incubated at 65°C for 3 min, cooled to 4°C, and then placed on ice for 5 min. The reaction master mix was as follows:

Table 1

Reagent	Per well	1 Plate MM
Nuclease Free Water	2.7	324.0
5X Precise Reverse Transcription Buffer	2.0	240
Precise RNase Inhibitor	0.1	12.0
Precise Reverse Transcriptase	0.2	24

[0129] 5 μ L of RT MM were pipetted to each well. Each plate was sealed, and a reverse-transcriptase program was run: 42°C, 30 min; 80C, 5 min; 4C pause.

[0130] cDNA purification was performed using the following protocol: Spin the plate (1000 rpm x 10 s). Combine all reactions into a single 2 mL tube. Add 8 μ L of Precise DBP Mix into the pooled cDNA tube. Vortex gently to get buffer into solution. Add an equivalent volume of AMPure XP beads to the single tube containing the pooled RT reaction product (1X volume of pooled PCR product). Pipette up and down to mix the sample and beads. Incubate bead mix at room temperature for 5 minutes. Open the cap to make sure the beads aren't disturbed. Place tube on magnet and wait until the liquid and beads separate. The liquid should be entirely clear before it is removed. This may take up to 5 min. Keeping the tube on the magnet, carefully remove only the liquid and discard. Keeping the

tube on the magnet stand, gently rinse the beads once using 2 mL of 70% EtOH /30% H₂O. Prepare fresh EtOH solution each day to avoid evaporation. Care can be taken not to disturb or re-suspend the beads during the rinse step. Immediately remove all residual EtOH solution with a pipette. While the tube is on the magnet, tap to collect as much of the EtOH as possible at the bottom of the tube. To elute the product from the beads, remove the tube from the magnet. Transfer 68 µL of Elution Buffer into the tube containing the beads. Wet the beads with the elution buffer by tilting the tube. Vortex well to mix and wait 1 min. Return the tube to the magnet. Wait until the solution is clear, typically <5 min. Recover the purified PCR product solution by carefully pipetting the solution away from the beads. Transfer the clear liquid product (68 µL) to a new 1.5 mL tube.

[0131] PCR N1 amplification (target amplification) was performed using the following protocol: Prepare PCR mix (200uL) in pre-amplification area according to Table 2:

Table 2

Reagent	Per sample
RT Product	68
2X OneTaq HotStart	100
Precise TCR N1 Primer Pool	12
10uM N1R Primer	20
Total	200

[0132] Split this 200uL reaction mix into 4 PCR tubes. Run 20 cycles of the following PCR conditions: 3-minute annealing time at 60C, 1-minute extension time at 68C (About 2 hours). Combine all PCR products into one sample tube. Add 160uL of Ampure beads (0.8X), pipette to mix well, and incubate for 5 minutes at RT. Place tube on magnet and wait until the liquid and beads separate. The liquid should be entirely clear before it is removed. This may take up to 5 min. Remove only the liquid and discard. Gently rinse the tube once while on the magnet stand with 1 mL of 70% EtOH/30% H₂O. Do not re-suspend the beads during the rinse step. Immediately remove all residual EtOH with a pipette. While the tube is on the magnet, tap to collect as much of the EtOH as possible at the bottom of the tube. Elute PCR product in 50uL Elution Buffer.

[0133] PCR N1 amplification (adapter and indexing PCR) was performed using the following protocol: Prepare PCR mix (50uL) according to Table 3:

Table 3

Reagent	Per sample
Water	7.75
2X OneTaq HotStart	25
Precise Indexing Primer Set (D70x)	1.25
Precise TCR N2 Primer Pool	6
PCR N1 Product	10

[0134] Run 25 cycles of the following PCR conditions: 3-minute annealing at 60C, 1 min extension time at 68C. Purify at 0.8X Ampure twice, elute in 30 uL of Elution Buffer.

[0135] FIG. 6 shows error correction for the average of sequence results (by individual TCRA CDR3). In FIG. 6, the Y axis plots the number of different molecule barcodes detected for TCRA.

[0136] Results are shown for reactions with 50 pg of target RNA (denoted as “50” in FIG. 6, e.g., A50, B50, C50, D50, E50, F50, G50) and negative controls that received 0 pg of target RNA (denoted as “0” in FIG. 6, e.g., A0, B0, C0, D0, E0, F0, G0). As such, 0 pg serves as a negative control and detector of assay noise that is leaked from 50 pg sample. For the reactions that received target RNA, unique oligonucleotide designs in accordance with some embodiments herein B01-B12 (“B50”), C01-C12 (“C50”), D01-D12 (“D50”), E01-E12 (“E50”), F01-F12 (“F50”), G01-G12 (“G50”), and H01-H12 (“H50”) had lower error rates than conventional formats A01-A12 (“A50”). In particular, for for TRAC in this experiment, format “H” had the highest sensitivity with low noise and performed better than the original barcoded primers (A).

[0137] Thus, compositions comprising unique oligonucleotide species comprising barcode regions, and molecule barcodes having a relatively low G content, and/or unique oligonucleotides comprising a spacer 3’ of the molecule barcode and 5’ of the uniform region as described herein yielded amplification with a lower error rate, and thus higher

accuracy than compositions comprising general polynucleotide species that did not comprise such features.

[0138] While various aspects and embodiments have been disclosed herein, other aspects and embodiments will be apparent to those skilled in the art. The various aspects and embodiments disclosed herein are for purposes of illustration and are not intended to be limiting, with the true scope and spirit being indicated by the following claims.

[0139] One skilled in the art will appreciate that, for this and other processes and methods disclosed herein, the functions performed in the processes and methods can be implemented in differing order. Furthermore, the outlined steps and operations are only provided as examples, and some of the steps and operations can be optional, combined into fewer steps and operations, or expanded into additional steps and operations without detracting from the essence of the disclosed embodiments.

[0140] With respect to the use of substantially any plural and/or singular terms herein, those having skill in the art can translate from the plural to the singular and/or from the singular to the plural as is appropriate to the context and/or application. The various singular/plural permutations may be expressly set forth herein for sake of clarity.

[0141] It will be understood by those within the art that, in general, terms used herein, and especially in the appended claims (e.g., bodies of the appended claims) are generally intended as "open" terms (e.g., the term "including" should be interpreted as "including but not limited to," the term "having" should be interpreted as "having at least," the term "includes" should be interpreted as "includes but is not limited to," etc.). It will be further understood by those within the art that if a specific number of an introduced claim recitation is intended, such an intent will be explicitly recited in the claim, and in the absence of such recitation no such intent is present. For example, as an aid to understanding, the following appended claims may contain usage of the introductory phrases "at least one" and "one or more" to introduce claim recitations. However, the use of such phrases should not be construed to imply that the introduction of a claim recitation by the indefinite articles "a" or "an" limits any particular claim containing such introduced claim recitation to embodiments containing only one such recitation, even when the same claim includes the introductory

phrases "one or more" or "at least one" and indefinite articles such as "a" or "an" (e.g., "a" and/or "an" should be interpreted to mean "at least one" or "one or more"); the same holds true for the use of definite articles used to introduce claim recitations. In addition, even if a specific number of an introduced claim recitation is explicitly recited, those skilled in the art will recognize that such recitation should be interpreted to mean at least the recited number (e.g., the bare recitation of "two recitations," without other modifiers, means at least two recitations, or two or more recitations). Furthermore, in those instances where a convention analogous to "at least one of A, B, and C, etc." is used, in general such a construction is intended in the sense one having skill in the art would understand the convention (e.g., "a system having at least one of A, B, and C" would include but not be limited to systems that have A alone, B alone, C alone, A and B together, A and C together, B and C together, and/or A, B, and C together, etc.). In those instances where a convention analogous to "at least one of A, B, or C, etc." is used, in general such a construction is intended in the sense one having skill in the art would understand the convention (e.g., "a system having at least one of A, B, or C" would include but not be limited to systems that have A alone, B alone, C alone, A and B together, A and C together, B and C together, and/or A, B, and C together, etc.). It will be further understood by those within the art that virtually any disjunctive word and/or phrase presenting two or more alternative terms, whether in the description, claims, or drawings, should be understood to contemplate the possibilities of including one of the terms, either of the terms, or both terms. For example, the phrase "A or B" will be understood to include the possibilities of "A" or "B" or "A and B."

[0142] In addition, where features or aspects of the disclosure are described in terms of Markush groups, those skilled in the art will recognize that the disclosure is also thereby described in terms of any individual member or subgroup of members of the Markush group.

[0143] As will be understood by one skilled in the art, for any and all purposes, such as in terms of providing a written description, all ranges disclosed herein also encompass any and all possible subranges and combinations of subranges thereof. Any listed range can be easily recognized as sufficiently describing and enabling the same range being broken down into at least equal halves, thirds, quarters, fifths, tenths, etc. As a non-limiting

example, each range discussed herein can be readily broken down into a lower third, middle third and upper third, etc. As will also be understood by one skilled in the art all language such as “up to,” “at least,” and the like include the number recited and refer to ranges which can be subsequently broken down into subranges as discussed above. Finally, as will be understood by one skilled in the art, a range includes each individual member. Thus, for example, a group having 1-3 cells refers to groups having 1, 2, or 3 cells. Similarly, a group having 1-5 cells refers to groups having 1, 2, 3, 4, or 5 cells, and so forth.

[0144] From the foregoing, it will be appreciated that various embodiments of the present disclosure have been described herein for purposes of illustration, and that various modifications may be made without departing from the scope and spirit of the present disclosure. Accordingly, the various embodiments disclosed herein are not intended to be limiting, with the true scope and spirit being indicated by the following claims.

WHAT IS CLAIMED IS:

1. A composition comprising at least 1000 unique oligonucleotide species, each unique oligonucleotide species comprising a barcode region comprising a molecule barcode comprising at least 7 nucleotides,

wherein the unique oligonucleotide species comprise different nucleic acid sequences in their barcode regions, and

wherein at least one of:

(a) the composition consists essentially of unique oligonucleotide species wherein each molecular barcode has a G content of less than 50%; or

(b) the molecule barcodes of all of the unique oligonucleotide species in the composition collectively have a G content of no more than 12.5%.

2. The composition of claim 1, wherein the barcode region further comprises a sample barcode comprising at least 3 nucleotides.

3. The composition of claim 1 or claim 2, wherein each unique oligonucleotide species further comprises a uniform region 3' of the barcode region.

4. The composition of claim 3, wherein the uniform region further comprises a target-specific region 3' of the barcode region, the target-specific region comprising at least 10 nucleotides complementary to a target nucleic acid.

5. The composition of any one of claims 1-4, wherein the composition consists essentially of unique oligonucleotide species wherein each molecular barcode has a G content of less than 50%.

6. The composition of any one of claims 1-5, wherein the molecule barcodes of all of the unique oligonucleotide species in the composition collectively have a G content of no more than 12.5%.

7. The composition of any one of claims 1-6, wherein the unique oligonucleotide species are disposed in at least two spatially isolated pools, each pool comprising at least 100 unique oligonucleotides of the unique oligonucleotide species,

wherein unique oligonucleotides in the same pool comprise the same sample barcode sequence, and

wherein different unique oligonucleotides of the same pool comprise a different molecule barcode sequences.

8. The composition of any one of claims 2-7, wherein the sample barcode of each unique oligonucleotide species has a G content of 50% or less.

9. The composition of any one of claims 1-8, wherein the barcode region of each unique oligonucleotide species has a G content of 50% or less.

10. The composition of any one of claims 1- 9, wherein the molecule barcodes of the unique oligonucleotide species collectively have a G content of less than 12.5%.

11. The composition of any one of claims 1- 10, wherein the barcode regions of the unique oligonucleotide species collectively have a G content of no more than 12.5%.

12. The composition any one of claims 1-11, wherein for at least 95% of the unique oligonucleotide species, any G in the molecule barcode is not adjacent to another G.

13. The composition any one of claims 1-12, wherein the composition consists essentially of unique oligonucleotide species for which any G in the molecule barcode is not adjacent to another G.

14. The composition of any one of claims 1-13, wherein at least 95% of the molecule barcodes of the unique oligonucleotide species comprise a sequence totaling at least 6 alternating H's and N's, wherein each "H" is any one of A, C, or T, and wherein each "N" is any one of A, G, C, or T.

15. The composition of any one of claims 1-14, wherein at least 95% of the molecule barcodes of the unique oligonucleotide species comprise the sequence HNHNHNHN, wherein each "H" is any one of A, C, or T, and wherein each "N" is any one of A, G, C, or T.

16. The composition of any one of claims 1-15, wherein each molecule barcode of the unique oligonucleotide species comprises the sequence HNHNHNHN, wherein each "H" is any one of A, C, or T, and wherein each "N" is any one of A, G, C, or T.

17. The composition of any one of claims 1-16, wherein each molecule barcode of the unique oligonucleotide species comprises the sequence HHHHHHHH, wherein each "H" is any one of A, C, or T.

18. The composition of any one of claims 1-17, wherein each of the unique oligonucleotide species comprises a spacer 3' of the barcode region and 5' of the target

specific region, said spacer comprising the sequence HNHNNHNH, wherein each “H” is any one of A, C, or T, and wherein each “N” is any one of A, G, C, or T.

19. The composition of any one of claims 1-17, wherein each of the unique oligonucleotide species comprises a spacer 3’ of the barcode region and 5’ of the target specific region, said spacer comprising the sequence HHHHHHHH, wherein each “H” is any one of A, C, or T.

20. The composition of any one of claims 4-19, wherein the target specific region comprises an oligo dT sequence.

21. The composition of any one of claims 1-20, wherein for each unique oligonucleotide species, the molecule barcode is 3’ of the sample barcode.

22. The composition of any one of claims 1-20, wherein for each unique oligonucleotide species, the sample barcode is 3’ of the molecule barcode.

23. The composition of any one of claims 1-22, wherein each oligonucleotide species has a length of at least 24 nucleotides.

24. The composition of any one of claims 1-22, wherein each oligonucleotide species has a length of 24-140 nucleotides.

25. The composition of any one of claims 1-24, wherein the composition comprises at least 6,500 unique oligonucleotide species.

26. The composition of any one of claims 1-24, wherein the composition comprises at least 65,000 unique oligonucleotide species.

27. The composition of any one of claims 1-26, wherein the composition comprises at least two oligonucleotides of the same unique oligonucleotide species.

28. The composition of any one of claims 1-27, wherein no two oligonucleotides of the composition are of the same unique oligonucleotide species.

29. The composition of any one of claims 7-28, wherein the composition comprises at least 48 pools.

30. The composition of any one of claims 7-29, wherein the unique oligonucleotide species of each pool are immobilized on a substrate, so that the sample barcodes but not the molecule barcodes are the same for the oligonucleotide species immobilized on each substrate.

31. The composition of claim 30, wherein the substrate comprises a discrete region of a surface, so that the surface can comprise two or more substrates.

32. The composition of claim 30, wherein the substrate comprises a bead.

33. The composition of any one of claims 27-29, wherein each of the unique oligonucleotide species further comprises an adapter configured to immobilize the unique oligonucleotide on the substrate, wherein said barcode region is 3' of the adapter.

34. A method of specifically barcoding a plurality of nucleic acids from two or more samples, each sample comprising nucleic acids:

contacting each sample with a pool comprising at least 100 unique oligonucleotide species, wherein each sample is contacted in spatial isolation from the other samples,

each unique oligonucleotide species comprising a barcode region comprising a molecule barcode comprising at least 7 nucleotides,

wherein the unique polynucleotide species of each pool comprise the same sample barcode, and comprise different molecule barcodes, and

wherein at least one of:

(a) the unique oligonucleotide species contacted with the sample consist essentially of unique oligonucleotide species wherein each molecular barcode has a G content of less than 50%; or

(b) the molecule barcodes of all of the unique oligonucleotide species collectively have a G content of no more than 12.5%;

hybridizing target-specific regions of at least some oligonucleotides of the unique oligonucleotide species to at least some of the nucleic acids of the sample; and

extending the hybridized oligonucleotides, thereby producing strands comprising an oligonucleotide of the unique oligonucleotide species and a sequence complementary to the target, wherein for each sample, the strands comprise the same sample barcode and different molecule barcodes, and wherein for different samples, the molecule barcodes are different.

35. The method of claim 34, wherein the barcode region further comprises a sample barcode comprising at least 3 nucleotides.

36. The method of claim 34 or claim 35, wherein each unique oligonucleotide species further comprises a uniform region 3' of the barcode region.

37. The method of claim 36, wherein the uniform region further comprises a target-specific region 3' of the barcode region, the target-specific region comprising at least 10 nucleotides complementary to a target nucleic acid.

38. The method of any one of claims 34-37, wherein the unique oligonucleotide species contacted with the sample consist essentially of unique oligonucleotide species wherein each molecular barcode has a G content of less than 50%.

39. The method of any one of claims 34-38, wherein the molecule barcodes of all of the unique oligonucleotide species collectively have a G content of no more than 12.5%.

40. The method of any one of claims 34-39, further comprising ascertaining nucleic acid sequences of the strands comprising the oligonucleotides of the unique oligonucleotide species and the sequence complementary to the target.

41. The method of any one of claims 34-40, wherein the at least 100 unique oligonucleotide species of each pool are immobilized on a substrate, so that the unique oligonucleotide species immobilized on a given substrate comprise the same sample barcode, and different unique oligonucleotide species immobilized on the substrate comprise different molecule barcodes.

42. The method of any one of claims 34-41, wherein each sample barcode has a G content of 50% or less.

43. The method of any one of claims 34-42, wherein the molecule barcodes of the unique oligonucleotide species collectively have a G content of less than 12.5%.

44. The method of any one of claims 34-43, wherein the barcode regions of the unique oligonucleotide species collectively have a G content of no more than 12.5%.

45. The method any one of claims 34-44, wherein for at least 95% of the unique oligonucleotide species, any G in the molecule barcode is not adjacent to another G.

46. The method any one of claims 34-44, wherein the each pool consists essentially of unique oligonucleotide species for which any G in the molecule barcode is not adjacent to another G.

47. The method of any one of claims 34-46, wherein at least 95% of the molecule barcodes of the unique oligonucleotide species comprise a sequence totaling at least 6 alternating H's and N's, wherein each "H" is any one of A, C, or T, and wherein each "N" is any one of A, G, C, or T.

48. The method of any one of claims 34-47, wherein at least 95% of the molecule barcodes of the unique oligonucleotide species comprise the sequence HNHNNHNH, wherein each "H" is any one of A, C, or T, and wherein each "N" is any one of A, G, C, or T.

49. The method of any one of claims 34-47, wherein each molecule barcode of the unique oligonucleotide species comprises the sequence HNHNNHNH, wherein each "H" is any one of A, C, or T, and wherein each "N" is any one of A, G, C, or T.

50. The method of any one of claims 34-49, wherein each molecule barcode of the unique oligonucleotides comprises the sequence HHHHHHHH, wherein each "H" is any one of A, C, or T.

51. The method of any one of claims 34-50, wherein each unique oligonucleotide comprises a spacer 3' of the barcode region and 5' of the target specific region, said spacer comprising the sequence HNHNNHNH, wherein each "H" is any one of A, C, or T, and wherein each "N" is any one of A, G, C, or T.

52. The method of any one of claims 34-50, wherein each unique oligonucleotide comprises a spacer 3' of the barcode region and 5' of the target specific region, said spacer comprising the sequence HHHHHHHH, wherein each "H" is any one of A, C, or T.

53. The method of any one of claims 34-52, wherein at least one pool comprises least two oligonucleotides of the same unique oligonucleotide species.

54. The method of any one of claims 34-52, wherein no pool comprises two oligonucleotides of the same unique oligonucleotide species.

55. The method of any one of claims 34-54, wherein the target specific region comprises an oligo dT sequence.

56. The method of any one of claims 34-55, wherein for each unique oligonucleotide species, the molecule barcode is 3' of the sample barcode.

57. The method of any one of claims 34-55, wherein for each unique oligonucleotide species, the sample barcode is 3' of the molecule barcode.

58. The method of any one of claims 34-57, wherein each unique oligonucleotide species has a length of at least 24 nucleotides.

59. The method of any one of claims 34-58, wherein each unique oligonucleotide species has a length of 24-140 nucleotides.

60. The method of any one of claims 34-59, wherein each pool comprises at least 6,500 unique oligonucleotide species.

61. The method of any one of claims 34-59, wherein each pool comprises at least 65,000 unique oligonucleotide species.

62. The method of any one of claims 34-61, wherein at least 48 unique samples are each contacted with a unique pool.

63. The method of any one of claims 34-62, wherein at least 99% of the samples comprise no more than one cell each.

64. The method of any one of claims 34-63, wherein the unique oligonucleotide species of each pool are immobilized on a substrate, so that the sample barcodes but not the molecule barcodes are the same for the unique oligonucleotide species immobilized on each substrate of the plurality.

65. The method of claim 64, wherein the substrate comprises a spatially-isolated region of a surface, so that the substrates of different pools comprise the different spatially-isolated regions of the surface.

66. The method of claim 64, wherein the substrate comprises a bead.

67. The method of any one of claims 64-66, wherein each of the unique oligonucleotide species further comprises an adapter configured to immobilize the unique oligonucleotide on the substrate, wherein said barcode region is 3' of the adapter.

68. A method of making a composition comprising unique oligonucleotides, the method comprising:

providing a plurality of different sample barcodes comprising at least 3 nucleotides each,

providing a plurality of different molecule barcodes comprising at least 7 nucleotides each;

synthesizing a plurality of unique oligonucleotide species, each unique oligonucleotide species comprising a barcode region comprising a sample barcode and a molecule barcode, wherein at least one of;

(a) the plurality of unique oligonucleotide species consists essentially of unique oligonucleotide species wherein each molecular barcode has a G content of less than 50%; or

(b) the molecule barcodes of all of the unique oligonucleotide species in the plurality collectively have a G content of no more than 12.5%; and

disposing the unique oligonucleotide species in spatially-isolated pools, wherein each pool comprises multiple unique oligonucleotide species, so that the unique oligonucleotide species of the same pool comprise the same sample barcode sequence, and wherein different unique oligonucleotide species of the same pool comprise different molecule barcode sequences,

wherein each pool comprises at least 1000 unique oligonucleotide species.

69. The method of claim 68, wherein each unique oligonucleotide species further comprises a uniform region 3' of the barcode region.

70. The method of claim 69, wherein the uniform region further comprises a target-specific region 3' of the barcode region, the target-specific region comprising at least 10 nucleotides complementary to a target nucleic acid.

71. The method of any one of claims 68-70, wherein the plurality of unique oligonucleotide species consists essentially of unique oligonucleotide species wherein each molecular barcode has a G content of less than 50%.

72. The method of claim 68 or 69, wherein the molecule barcodes of all of the unique oligonucleotide species in the plurality collectively have a G content of no more than 12.5%

73. The method of claim 68, wherein each molecule barcode has a G content of 50% or less.

74. The method of any one of claims 68-69, wherein each sample barcode has a G content of 50% or less.

75. The method of any one of claims 68-72, wherein the sample barcodes of the unique oligonucleotide species collectively have a G content of no more than 12.5%.

76. The method of any one of claims 68-72, wherein the molecule barcodes of the unique oligonucleotide species collectively have a G content of less than 12.5%.

77. The method of any one of claims 68-76, wherein the barcode regions of the unique oligonucleotide species collectively have a G content of no more than 12.5%.

78. The method any one of claims 68-77, wherein for at least 95% of the unique oligonucleotide species, any G in the molecule barcode is not adjacent to another G.

79. The method any one of claims 68-77, wherein the plurality of the unique oligonucleotide species consists essentially of unique oligonucleotide species for which any G in the molecule barcode is not adjacent to another G.

80. The method any one of claims 68-79, wherein at least 95% of the molecule barcodes of the unique oligonucleotide species comprise a sequence totaling at least 6 alternating H's and N's, wherein each "H" is any one of A, C, or T, and wherein each "N" is any one of A, G, C, or T.

81. The method of any one of claims 68-80, wherein at least 95% of the molecule barcodes of the unique oligonucleotides comprise the sequence HNHNNHNH, wherein each "H" is any one of A, C, or T, and wherein each "N" is any one of A, G, C, or T.

82. The method of any one of claims 68-81, wherein each molecule barcode of the unique oligonucleotides comprises the sequence HNHNNHNH, wherein each "H" is any one of A, C, or T, and wherein each "N" is any one of A, G, C, or T.

83. The method of any one of claims 68-82, wherein each molecule barcode of the unique oligonucleotides comprises the sequence HHHHHHHH, wherein each "H" is any one of A, C, or T.

84. The method of any one of claims 68-83, wherein each unique oligonucleotide species comprises a spacer 3' of the barcode region and 5' of the target specific region, said spacer comprising the sequence HNHNNHNH, wherein each "H" is any one of A, C, or T, and wherein each "N" is any one of A, G, C, or T.

85. The method of any one of claims 68-84, wherein each unique oligonucleotide species comprises a spacer 3' of the barcode region and 5' of the target specific region, said spacer comprising the sequence HHHHHHHH, wherein each "H" is any one of A, C, or T.

86. The method of any one of claims 70-85, wherein the target specific region comprises an oligo dT sequence.

87. The method of any one of claims 68-86, wherein for each unique oligonucleotide species, the molecule barcode is 3' of the sample barcode.

88. The method of any one of claims 68-86, wherein for each unique oligonucleotide species, the sample barcode is 3' of the molecule barcode.

89. The method of any one of claims 68-88, wherein each unique oligonucleotide species has a length of at least 24 nucleotides.

90. The method of any one of claims 68-89, wherein each unique oligonucleotide species has a length of 24-140 nucleotides.

91. The method of any one of claims 68-90, wherein each pool comprises at least 6,500 unique oligonucleotide species.

92. The method of any one of claims 68-91, wherein each pool comprises at least 65,000 unique oligonucleotide species.

93. The method of any one of claims 68-92, wherein at least 48 pools are made.

94. The method of any one of claims 68-93, further comprising immobilizing the unique oligonucleotide species of each pool onto a substrate, so that the sample barcodes but not the molecule barcodes are the same for the oligonucleotide species immobilized on each substrate of the plurality.

95. The method of claim 94, wherein the substrates comprise discrete regions of a surface.

96. The method of claim 94, wherein the substrates comprise beads

97. The method of any one of claims 68-96, wherein the unique oligonucleotide species are disposed in spatially-isolated pools concurrent with said synthesis.

98. The method of any one of claims 68-96, wherein the unique oligonucleotide species are disposed in spatially-isolated pools after said synthesis.

99. An oligonucleotide comprising a barcode region 3' of the adapter region, the barcode region comprising a molecule barcode comprising at least 7 nucleotides, wherein the molecule barcode has a G content of no more than 50%.

100. The oligonucleotide of claim 99, further comprising a sample barcode comprising at least 3 nucleotides.

101. The oligonucleotide of claim 99 or claim 100, further comprising a uniform region 3' of the barcode region.

102. The oligonucleotide of claim 101, wherein the uniform region comprises a target-specific region comprising at least 10 nucleotides complementary to a target nucleic acid.

103. The oligonucleotide of any one of claims 99-102, further comprising an adapter region 5' of the barcode region.

104. The composition of any one of claims 3-19 or 21-33, wherein the uniform region comprises a target-specific region comprising a sequence flanking an immune cell receptor or immunoglobulin variable region coding sequence.

105. The composition of claim 104, wherein the immune cell receptor variable region coding sequence is selected from the group consisting of: a T cell receptor variable region coding sequence, a B cell receptor variable region coding sequence, and a combination thereof.

106. A kit comprising the composition of claim 104 or 105, the kit further comprising a primer configured to hybridize on an opposite side of the variable region as the target-specific region, and to hybridize to a complementary strand to a strand hybridized by the target-specific region, and is thereby configured to amplify the variable region in conjunction with the target-specific region.

107. The kit of claim 106, wherein the primer and target-specific region are configured to amplify a nucleic acid of at least 1 kb comprising the variable region.

108. The method of any one of claims 36-54, 56-67, 69-85, or 87-98, wherein the uniform region comprises a target-specific region comprising a sequence flanking an immune cell receptor or immunoglobulin variable region coding sequence.

109. The method of claim 108, wherein the immune cell receptor variable region coding sequence is selected from the group consisting of: a T cell receptor variable region coding sequence, a B cell receptor variable region coding sequence, and a combination thereof.

110. The oligonucleotide of any one of claims 101-103, wherein the uniform region comprises a target-specific region comprising a sequence flanking a variable region of an immune cell receptor or immunoglobulin.

111. The oligonucleotide of claim 110, wherein the immune cell receptor variable region coding sequence is selected from the group consisting of: a T cell receptor variable region coding sequence, a B cell receptor variable region coding sequence, and a combination thereof.

FIG. 1A



FIG. 1B

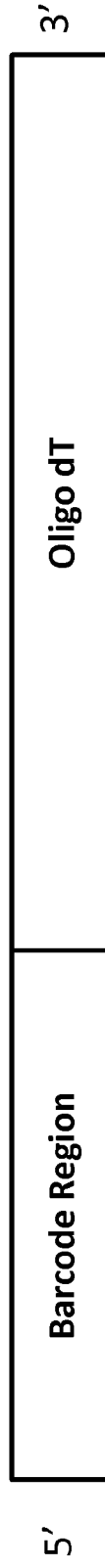


FIG. 1C

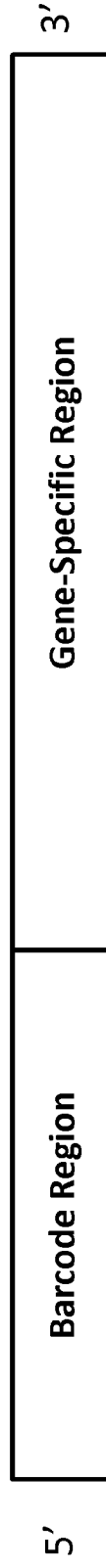


FIG. 2A

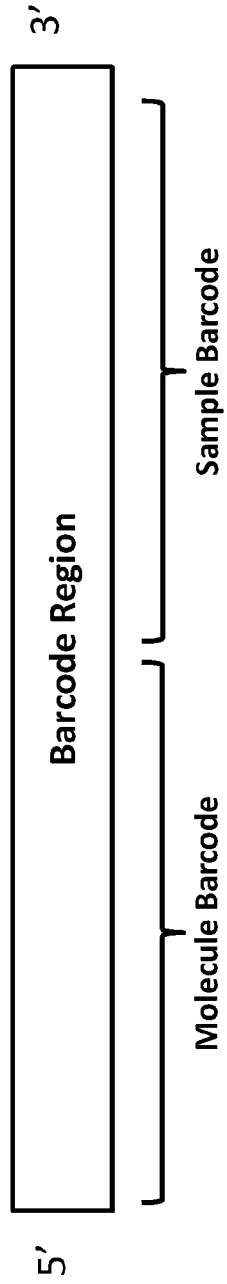


FIG. 2B

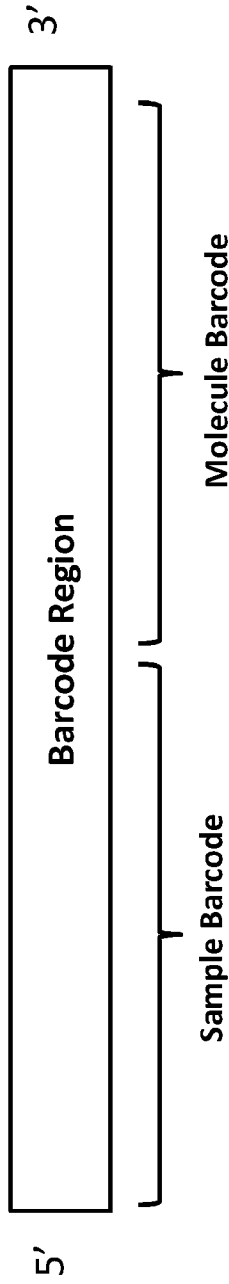


FIG. 3A

Baseline

5' [Adapter] [Sample Barcode] NNNNNNNN [Uniform Region, Gene Specific Region, or oligo dT] 3'

FIG. 3B

Add 3 base spacer before MI/SI

5' [Adapter] [Sample Barcode] NNNNNNNN **TTT** [Uniform Region, Gene Specific Region, or oligo dT] 3'

FIG. 3C 3/14

Add 5 base spacer before MI/SI

5' [Adapter] [Sample Barcode] NNNNNNNN **TTTTT** [Uniform Region, Gene Specific Region, or oligo dT] 3'

FIG. 3D

Switch SI and MI positions

5' [Adapter] NNNNNNNN [Sample Barcode] [Uniform Region, Gene Specific Region, or oligo dT] 3'

FIG. 3E

Switch SI and MI position, add 3 base spacer
 5' [Adapter]NNNNNNNN[Sample Barcode]**TTT**[Uniform Region, Gene Specific Region, or oligo dT] 3'

FIG. 3F

Switch SI and MI position, add 5 base spacer
 5' [Adapter]NNNNNNNN[Sample Barcode]**TTTTT**[Uniform Region, Gene Specific Region, or oligo dT] 3'

FIG. 3G

Switch MI format from 'NNNNNNNN' to 'HHHHHHHH'
 5' [Adapter]NNNNNNNN[Sample Barcode]HHHHHHHH[Uniform Region, Gene Specific Region, or oligo dT] 3'

4/14

FIG. 3H

Switch MI format from 'NNNNNNNN' to 'HNNHNNHN'
 5' [Adapter]NNNNNNNN[Sample Barcode]HNNHNNHN[Uniform Region, Gene Specific Region, or oligo dT] 3'

FIG. 4A

ES32, TRBC, Entire MI Nucleotide Usage

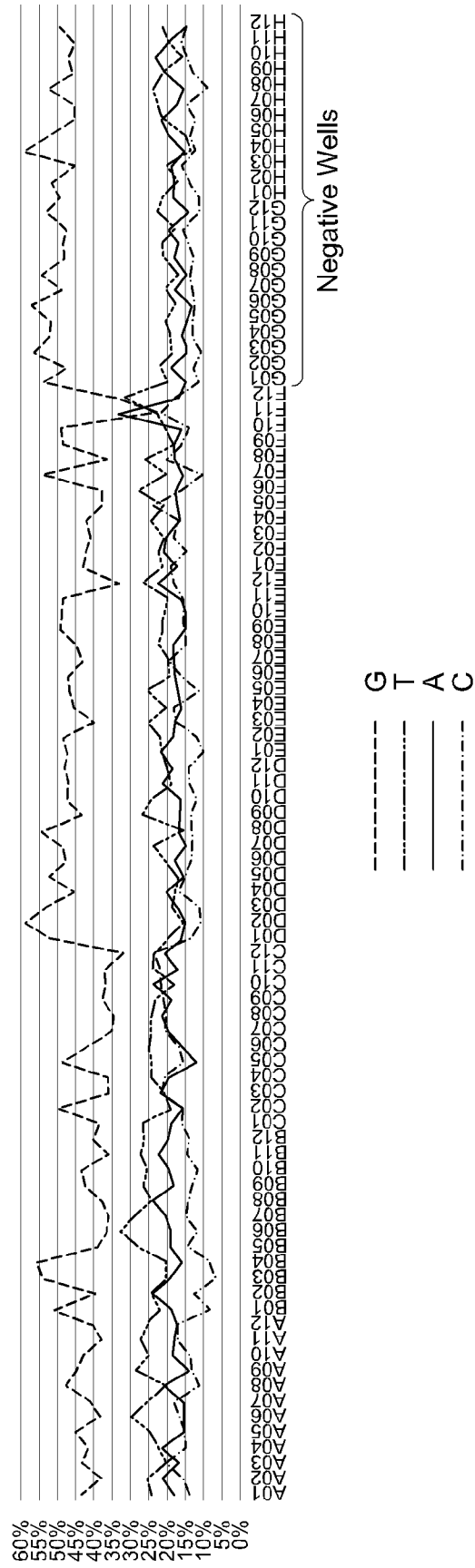


FIG. 4B
ES32, TRBC, Entire MI Nucleotide Usage

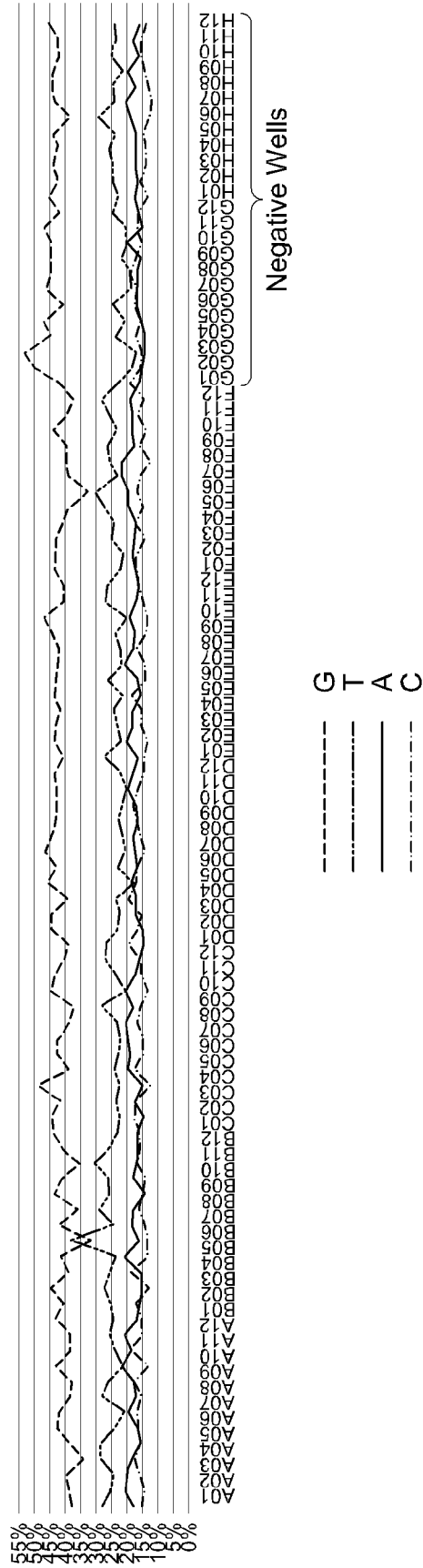


FIG. 5A

Sample	Anchor (SEQ ID NO: 97)	1st bc	2nd bc	Spacer	5' to 3' Recognition Sequence	SEQ ID NO: (anchor + 1 st & 2 nd bc + spacer + 5' to 3' rec. sequence)
A01	ACACGACCGCTCTTCCGATCT	ACAGTGCT	NNNNNNNN		CAGACAGACTTGT	SEQ ID NO. 1
A02	ACACGACCGCTCTTCCGATCT	ACATGCGT	NNNNNNNN		CAGACAGACTTGT	SEQ ID NO. 2
A03	ACACGACCGCTCTTCCGATCT	ACCGATTG	NNNNNNNN		CAGACAGACTTGT	SEQ ID NO. 3
A04	ACACGACCGCTCTTCCGATCT	ACCTTGAG	NNNNNNNN		CAGACAGACTTGT	SEQ ID NO. 4
A05	ACACGACCGCTCTTCCGATCT	ACGATGTC	NNNNNNNN		CAGACAGACTTGT	SEQ ID NO. 5
A06	ACACGACCGCTCTTCCGATCT	ACGTATGC	NNNNNNNN		CAGACAGACTTGT	SEQ ID NO. 6
A07	ACACGACCGCTCTTCCGATCT	ACTCTGGA	NNNNNNNN		CAGACAGACTTGT	SEQ ID NO. 7
A08	ACACGACCGCTCTTCCGATCT	ACTGGCTA	NNNNNNNN		CAGACAGACTTGT	SEQ ID NO. 8
A09	ACACGACCGCTCTTCCGATCT	AGACCTTG	NNNNNNNN		CAGACAGACTTGT	SEQ ID NO. 9
A10	ACACGACCGCTCTTCCGATCT	AGATTCCG	NNNNNNNN		CAGACAGACTTGT	SEQ ID NO. 10
A11	ACACGACCGCTCTTCCGATCT	AGCCGACC	NNNNNNNN		CAGACAGACTTGT	SEQ ID NO. 11
A12	ACACGACCGCTCTTCCGATCT	AGCGCTGC	NNNNNNNN		CAGACAGACTTGT	SEQ ID NO. 12
B01	ACACGACCGCTCTTCCGATCT	AGCTAGTC	NNNNNNNN	TTT	CAGACAGACTTGT	SEQ ID NO. 13
B02	ACACGACCGCTCTTCCGATCT	AGGCTCTA	NNNNNNNN	TTT	CAGACAGACTTGT	SEQ ID NO. 14
B03	ACACGACCGCTCTTCCGATCT	AGGTCCTA	NNNNNNNN	TTT	CAGACAGACTTGT	SEQ ID NO. 15
B04	ACACGACCGCTCTTCCGATCT	AGTACTAT	NNNNNNNN	TTT	CAGACAGACTTGT	SEQ ID NO. 16
B05	ACACGACCGCTCTTCCGATCT	AGTCAGCT	NNNNNNNN	TTT	CAGACAGACTTGT	SEQ ID NO. 17
B06	ACACGACCGCTCTTCCGATCT	AGTGTCCG	NNNNNNNN	TTT	CAGACAGACTTGT	SEQ ID NO. 18
B07	ACACGACCGCTCTTCCGATCT	ATACGTGC	NNNNNNNN	TTT	CAGACAGACTTGT	SEQ ID NO. 19
B08	ACACGACCGCTCTTCCGATCT	ATAGCGTC	NNNNNNNN	TTT	CAGACAGACTTGT	SEQ ID NO. 20
B09	ACACGACCGCTCTTCCGATCT	AATCGGTC	NNNNNNNN	TTT	CAGACAGACTTGT	SEQ ID NO. 21
B10	ACACGACCGCTCTTCCGATCT	ATCTCGGA	NNNNNNNN	TTT	CAGACAGACTTGT	SEQ ID NO. 22
B11	ACACGACCGCTCTTCCGATCT	ATGACGCT	NNNNNNNN	TTT	CAGACAGACTTGT	SEQ ID NO. 23
B12	ACACGACCGCTCTTCCGATCT	ATTCCGAG	NNNNNNNN	TTT	CAGACAGACTTGT	SEQ ID NO. 24

FIG. 5B

Sample	Anchor (SEQ ID NO: 97)	1st bc	2nd bc	Spacer	5 to 3' Recognition Sequence	SEQ ID NO: (anchor + 1 st & 2 nd bc + spacer + 5' to 3' rec. sequence)
C01	ACACGACGCTCTTCCGATCT	ATTGACCG	NNNNNNNN	TTTTT	CAGACAGACTTGT	SEQ ID NO. 25
C02	ACACGACGCTCTTCCGATCT	CAGAGTTC	NNNNNNNN	TTTTT	CAGACAGACTTGT	SEQ ID NO. 26
C03	ACACGACGCTCTTCCGATCT	CAGCCGAC	NNNNNNNN	TTTTT	CAGACAGACTTGT	SEQ ID NO. 27
C04	ACACGACGCTCTTCCGATCT	CAGTTAGC	NNNNNNNN	TTTTT	CAGACAGACTTGT	SEQ ID NO. 28
C05	ACACGACGCTCTTCCGATCT	CATCGTGA	NNNNNNNN	TTTTT	CAGACAGACTTGT	SEQ ID NO. 29
C06	ACACGACGCTCTTCCGATCT	CATGCCGA	NNNNNNNN	TTTTT	CAGACAGACTTGT	SEQ ID NO. 30
C07	ACACGACGCTCTTCCGATCT	CCAGTTAG	NNNNNNNN	TTTTT	CAGACAGACTTGT	SEQ ID NO. 31
C08	ACACGACGCTCTTCCGATCT	CCATGGCG	NNNNNNNN	TTTTT	CAGACAGACTTGT	SEQ ID NO. 32
C09	ACACGACGCTCTTCCGATCT	CCGATTGA	NNNNNNNN	TTTTT	CAGACAGACTTGT	SEQ ID NO. 33
C10	ACACGACGCTCTTCCGATCT	CGACCAGC	NNNNNNNN	TTTTT	CAGACAGACTTGT	SEQ ID NO. 34
C11	ACACGACGCTCTTCCGATCT	CGAGATTC	NNNNNNNN	TTTTT	CAGACAGACTTGT	SEQ ID NO. 35
C12	ACACGACGCTCTTCCGATCT	CGATFGAC	NNNNNNNN	TTTTT	CAGACAGACTTGT	SEQ ID NO. 36
D01	ACACGACGCTCTTCCGATCT	NNNNNNNN	CGCATGTA		CAGACAGACTTGT	SEQ ID NO. 37
D02	ACACGACGCTCTTCCGATCT	NNNNNNNN	CGCTATGA		CAGACAGACTTGT	SEQ ID NO. 38
D03	ACACGACGCTCTTCCGATCT	NNNNNNNN	CGTACATG		CAGACAGACTTGT	SEQ ID NO. 39
D04	ACACGACGCTCTTCCGATCT	NNNNNNNN	CGTCATAG		CAGACAGACTTGT	SEQ ID NO. 40
D05	ACACGACGCTCTTCCGATCT	NNNNNNNN	CTACGACA		CAGACAGACTTGT	SEQ ID NO. 41
D06	ACACGACGCTCTTCCGATCT	NNNNNNNN	CTAGCTGA		CAGACAGACTTGT	SEQ ID NO. 42
D07	ACACGACGCTCTTCCGATCT	NNNNNNNN	CTATAGTA		CAGACAGACTTGT	SEQ ID NO. 43
D08	ACACGACGCTCTTCCGATCT	NNNNNNNN	CTGACTAG		CAGACAGACTTGT	SEQ ID NO. 44
D09	ACACGACGCTCTTCCGATCT	NNNNNNNN	CTGCAGCG		CAGACAGACTTGT	SEQ ID NO. 45
D10	ACACGACGCTCTTCCGATCT	NNNNNNNN	CTGTGATG		CAGACAGACTTGT	SEQ ID NO. 46
D11	ACACGACGCTCTTCCGATCT	NNNNNNNN	CTTAGAGC		CAGACAGACTTGT	SEQ ID NO. 47
D12	ACACGACGCTCTTCCGATCT	NNNNNNNN	CTTGAGAC		CAGACAGACTTGT	SEQ ID NO. 48

FIG. 5C

Sample	Anchor (SEQ ID NO: 97)	1st bc	2nd bc	Spacer	5 to 3' Recognition Sequence	SEQ ID NO: (anchor + 1 st & 2 nd bc + spacer + 5' to 3' rec. sequence)
E01	ACACGACGCTCTTCCGATCT	NNNNNNNN	GACCAGCC	TTT	CAGACAGACTTGT	SEQ ID NO. 49
E02	ACACGACGCTCTTCCGATCT	NNNNNNNN	GACGTGCG	TTT	CAGACAGACTTGT	SEQ ID NO. 50
E03	ACACGACGCTCTTCCGATCT	NNNNNNNN	GACTGATC	TTT	CAGACAGACTTGT	SEQ ID NO. 51
E04	ACACGACGCTCTTCCGATCT	NNNNNNNN	GAGCCTTA	TTT	CAGACAGACTTGT	SEQ ID NO. 52
E05	ACACGACGCTCTTCCGATCT	NNNNNNNN	GAGTTCCA	TTT	CAGACAGACTTGT	SEQ ID NO. 53
E06	ACACGACGCTCTTCCGATCT	NNNNNNNN	GATATCAT	TTT	CAGACAGACTTGT	SEQ ID NO. 54
E07	ACACGACGCTCTTCCGATCT	NNNNNNNN	GATCGACT	TTT	CAGACAGACTTGT	SEQ ID NO. 55
E08	ACACGACGCTCTTCCGATCT	NNNNNNNN	GATGCTGT	TTT	CAGACAGACTTGT	SEQ ID NO. 56
E09	ACACGACGCTCTTCCGATCT	NNNNNNNN	GCAGTATC	TTT	CAGACAGACTTGT	SEQ ID NO. 57
E10	ACACGACGCTCTTCCGATCT	NNNNNNNN	GCATGTAC	TTT	CAGACAGACTTGT	SEQ ID NO. 58
E11	ACACGACGCTCTTCCGATCT	NNNNNNNN	GCCAGTTA	TTT	CAGACAGACTTGT	SEQ ID NO. 59
E12	ACACGACGCTCTTCCGATCT	NNNNNNNN	GCCGACCA	TTT	CAGACAGACTTGT	SEQ ID NO. 60
F01	ACACGACGCTCTTCCGATCT	NNNNNNNN	GCCTTAGA	TTTTT	CAGACAGACTTGT	SEQ ID NO. 61
F02	ACACGACGCTCTTCCGATCT	NNNNNNNN	GCGATACT	TTTTT	CAGACAGACTTGT	SEQ ID NO. 62
F03	ACACGACGCTCTTCCGATCT	NNNNNNNN	GCGTACAT	TTTTT	CAGACAGACTTGT	SEQ ID NO. 63
F04	ACACGACGCTCTTCCGATCT	NNNNNNNN	GGCATTGT	TTTTT	CAGACAGACTTGT	SEQ ID NO. 64
F05	ACACGACGCTCTTCCGATCT	NNNNNNNN	GGCGCCAT	TTTTT	CAGACAGACTTGT	SEQ ID NO. 65
F06	ACACGACGCTCTTCCGATCT	NNNNNNNN	GGTGTTAC	TTTTT	CAGACAGACTTGT	SEQ ID NO. 66
F07	ACACGACGCTCTTCCGATCT	NNNNNNNN	GTACGCAT	TTTTT	CAGACAGACTTGT	SEQ ID NO. 67
F08	ACACGACGCTCTTCCGATCT	NNNNNNNN	GTCGGCTG	TTTTT	CAGACAGACTTGT	SEQ ID NO. 68
F09	ACACGACGCTCTTCCGATCT	NNNNNNNN	GTGACATC	TTTTT	CAGACAGACTTGT	SEQ ID NO. 69
F10	ACACGACGCTCTTCCGATCT	NNNNNNNN	GTGCATAC	TTTTT	CAGACAGACTTGT	SEQ ID NO. 70
F11	ACACGACGCTCTTCCGATCT	NNNNNNNN	GTGTGCGC	TTTTT	CAGACAGACTTGT	SEQ ID NO. 71
F12	ACACGACGCTCTTCCGATCT	NNNNNNNN	GTTAGCCA	TTTTT	CAGACAGACTTGT	SEQ ID NO. 72

FIG. 5D

Sample	Anchor (SEQ ID NO: 97)	1st bc	2nd bc	Spacer	5 to 3' Recognition Sequence	SEQ ID NO: (anchor + 1 st & 2 nd bc + spacer + 5' to 3' rec. sequence)
G01	ACACGACCGCTCTTCCGATCT	GTTCAG	HHHHHHH		CAGACAGACTTGT	SEQ ID NO. 73
G02	ACACGACCGCTCTTCCGATCT	TACGTGCA	HHHHHHH		CAGACAGACTTGT	SEQ ID NO. 74
G03	ACACGACCGCTCTTCCGATCT	TACTCGCA	HHHHHHH		CAGACAGACTTGT	SEQ ID NO. 75
G04	ACACGACCGCTCTTCCGATCT	TAGAGCCT	HHHHHHH		CAGACAGACTTGT	SEQ ID NO. 76
G05	ACACGACCGCTCTTCCGATCT	TAGCCAGT	HHHHHHH		CAGACAGACTTGT	SEQ ID NO. 77
G06	ACACGACCGCTCTTCCGATCT	TATCGCAG	HHHHHHH		CAGACAGACTTGT	SEQ ID NO. 78
G07	ACACGACCGCTCTTCCGATCT	TCAGTCGA	HHHHHHH		CAGACAGACTTGT	SEQ ID NO. 79
G08	ACACGACCGCTCTTCCGATCT	TCATGATA	HHHHHHH		CAGACAGACTTGT	SEQ ID NO. 80
G09	ACACGACCGCTCTTCCGATCT	TCCAGAGT	HHHHHHH		CAGACAGACTTGT	SEQ ID NO. 81
G10	ACACGACCGCTCTTCCGATCT	TCCGAGAT	HHHHHHH		CAGACAGACTTGT	SEQ ID NO. 82
G11	ACACGACCGCTCTTCCGATCT	TCGATCAG	HHHHHHH		CAGACAGACTTGT	SEQ ID NO. 83
G12	ACACGACCGCTCTTCCGATCT	TCGCGACG	HHHHHHH		CAGACAGACTTGT	SEQ ID NO. 84
H01	ACACGACCGCTCTTCCGATCT	TCGGCTGG	HNHNHNH		CAGACAGACTTGT	SEQ ID NO. 85
H02	ACACGACCGCTCTTCCGATCT	TCGTAGTG	HNHNHNH		CAGACAGACTTGT	SEQ ID NO. 86
H03	ACACGACCGCTCTTCCGATCT	TGACCGAT	HNHNHNH		CAGACAGACTTGT	SEQ ID NO. 87
H04	ACACGACCGCTCTTCCGATCT	TGAGACCT	HNHNHNH		CAGACAGACTTGT	SEQ ID NO. 88
H05	ACACGACCGCTCTTCCGATCT	TGCATACG	HNHNHNH		CAGACAGACTTGT	SEQ ID NO. 89
H06	ACACGACCGCTCTTCCGATCT	TGCGCGTG	HNHNHNH		CAGACAGACTTGT	SEQ ID NO. 90
H07	ACACGACCGCTCTTCCGATCT	TGCTACAG	HNHNHNH		CAGACAGACTTGT	SEQ ID NO. 91
H08	ACACGACCGCTCTTCCGATCT	TGTACGCA	HNHNHNH		CAGACAGACTTGT	SEQ ID NO. 92
H09	ACACGACCGCTCTTCCGATCT	TTACGGTG	HNHNHNH		CAGACAGACTTGT	SEQ ID NO. 93
H10	ACACGACCGCTCTTCCGATCT	TTAGCCAG	HNHNHNH		CAGACAGACTTGT	SEQ ID NO. 94
H11	ACACGACCGCTCTTCCGATCT	TTGACCGA	HNHNHNH		CAGACAGACTTGT	SEQ ID NO. 95
H12	ACACGACCGCTCTTCCGATCT	TTGTGGCA	HNHNHNH		CAGACAGACTTGT	SEQ ID NO. 96

FIG. 6

Error Correction by Individual CDR3

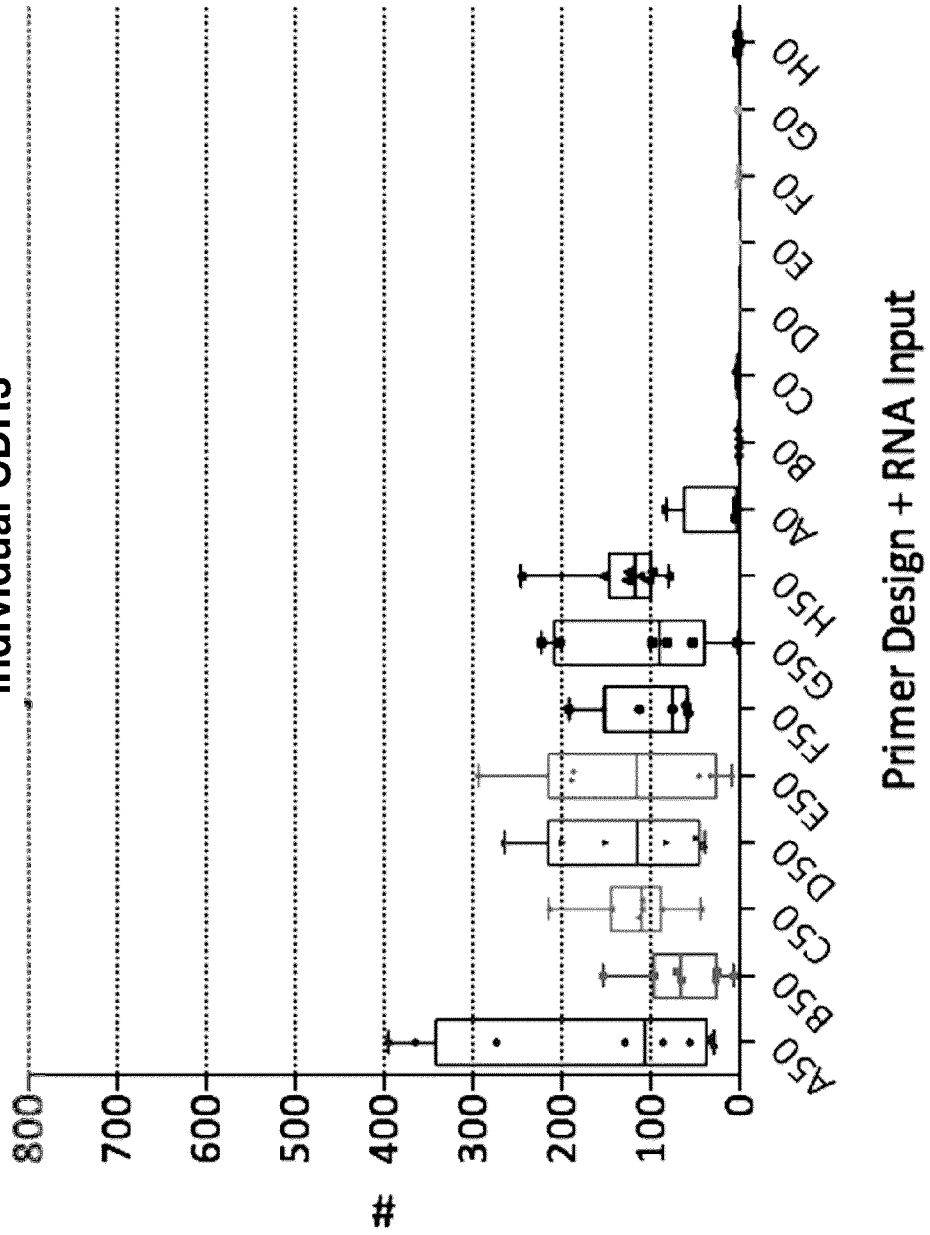


FIG. 7

**Error Correction by Individual
CDR3 and Filter MI Crossover**

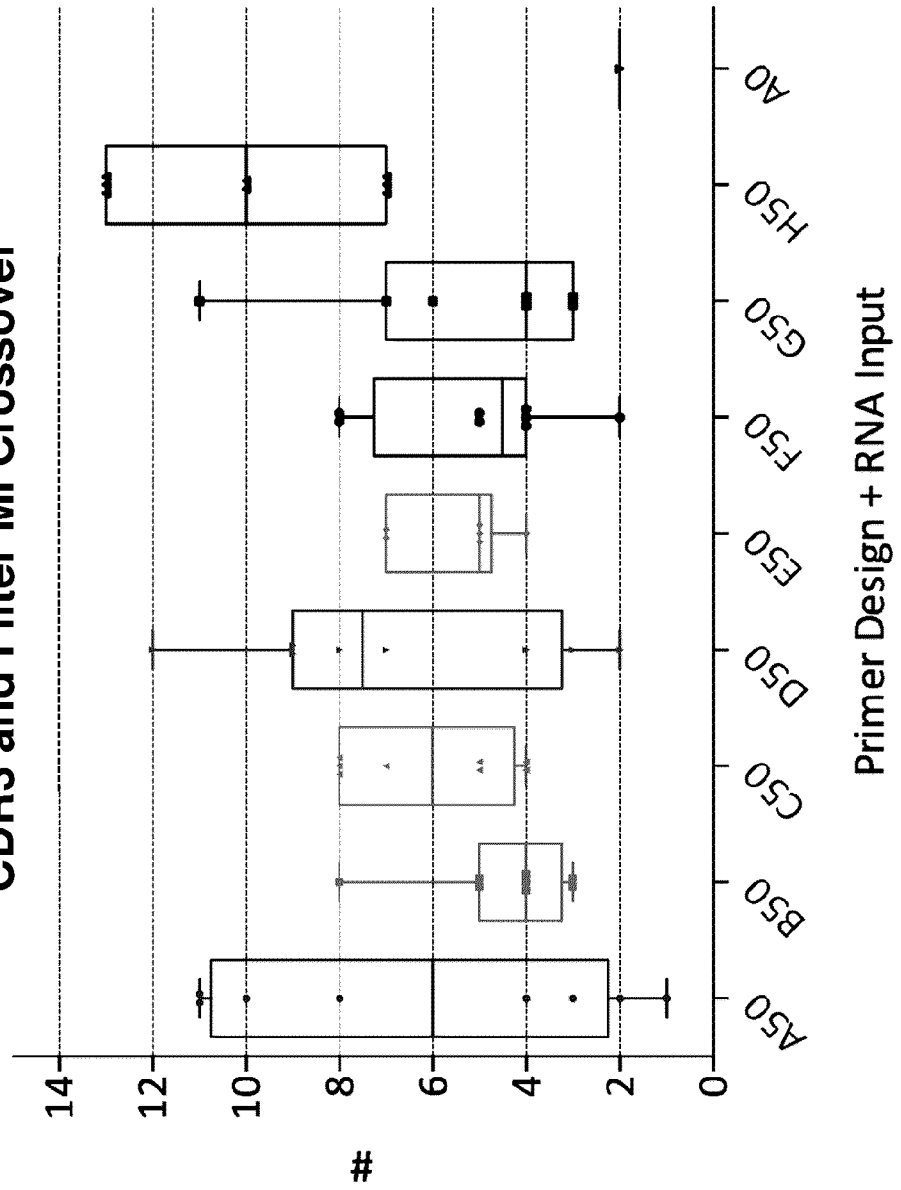
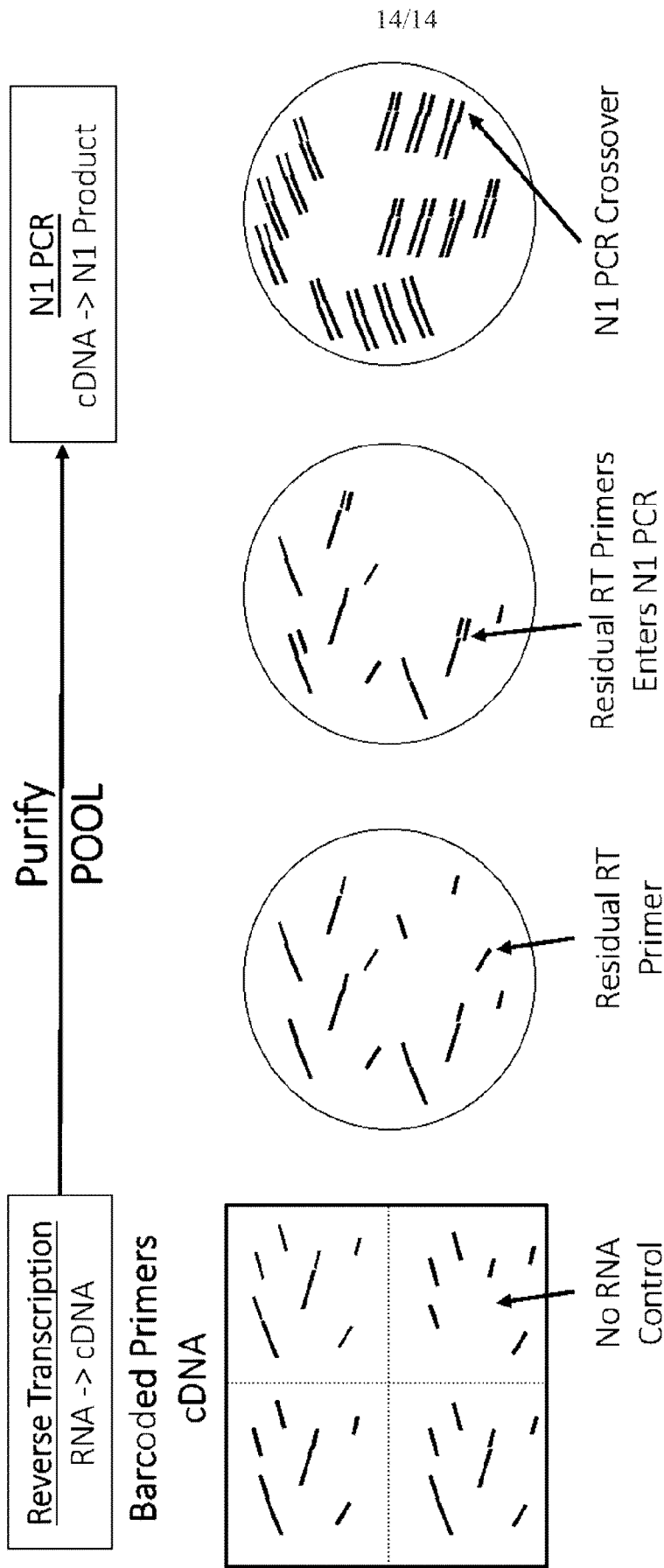


FIG. 8



INTERNATIONAL SEARCH REPORT

International application No PCT/US2017/030097

A. CLASSIFICATION OF SUBJECT MATTER INV. C12Q1/68 ADD.		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols) C12Q		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) EPO-Internal, BIOSIS, EMBASE, WPI Data		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 02/059355 A2 (TM BIOSCIENCE CORP [CA]; KOBLER DANIEL [CA]; FIELDHOUSE DANIEL [CA]) 1 August 2002 (2002-08-01)	1,5,6, 12,13, 17,23, 24,27, 34,38, 39, 42-46, 50,53, 58,59
Y	pages 30-31 pages 39-41, pages 47-48	2-4, 7-11, 14-16, 18-22, 25,26, 28-33, 35-37, 40,41, 47-49, 51,52,
-/--		
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C.		
<input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents :		
"A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family	
Date of the actual completion of the international search	Date of mailing of the international search report	
28 August 2017	08/09/2017	
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Authorized officer Knudsen, Henrik	

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2017/030097

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
		56,57, 60-98
X	----- WO 2015/031691 A1 (CELLULAR RES INC [US]) 5 March 2015 (2015-03-05)	99-111
Y	paragraphs [0335], [0754]; figures 1,2A67-68; example 1	2-4, 7-11, 14-16, 18-22, 25,26, 28-33, 35-37, 40,41, 47-49, 51,52, 56,57, 60-98
X	----- WO 2015/134787 A2 (ADAPTIVE BIOTECHNOLOGIES CORP [US]) 11 September 2015 (2015-09-11) paragraph [0233]; claim 1	1-111
X	----- TILO BUSCHMANN ET AL: "Enhancing the detection of barcoded reads in high throughput DNA sequencing data by controlling the false discovery rate", BMC BIOINFORMATICS, BIOMED CENTRAL, LONDON, GB, vol. 15, no. 1, 7 August 2014 (2014-08-07) , page 264, XP021191527, ISSN: 1471-2105, DOI: 10.1186/1471-2105-15-264 page 3, right-hand column, last paragraph - page 4, left-hand column, paragraph 1 & Thilo Buschmann ET AL: "Additional file 2", BMC Bioinformatics, 7 August 2014 (2014-08-07), page 264, XP055401307, Retrieved from the Internet: URL:https://www.ncbi.nlm.nih.gov/pmc/artic les/PMC4133078/bin/12859_2013_6528_MOESM2_ ESM.pdf [retrieved on 2017-08-25]	1-111
X,P	----- WO 2016/138500 A1 (CELLULAR RES INC [US]) 1 September 2016 (2016-09-01) claims 1,4; figure 1 ----- -/--	99-111

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2017/030097

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	TEEMU KIVIOJA ET AL: "Counting absolute numbers of molecules using unique molecular identifiers", NATURE METHODS, vol. 9, no. 1, 20 November 2011 (2011-11-20), pages 72-74, XP055401382, ISSN: 1548-7091, DOI: 10.1038/nmeth.1778 abstract	1-111
A	----- US 2010/159533 A1 (LIPSON DORON [US] ET AL) 24 June 2010 (2010-06-24) paragraph [0016] -----	1-111

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2017/030097

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 02059355	A2	01-08-2002	AT 546545 T 15-03-2012
			AU 2008201349 A1 17-04-2008
			BR 0206746 A 16-11-2004
			BR 0206747 A 30-11-2004
			CA 2435551 A1 01-08-2002
			CA 2435612 A1 01-08-2002
			DK 1364065 T3 14-05-2012
			EP 1356117 A2 29-10-2003
			EP 1364065 A2 26-11-2003
			EP 2325336 A1 25-05-2011
			EP 2327794 A2 01-06-2011
			ES 2382542 T3 11-06-2012
			JP 4422963 B2 03-03-2010
			JP 4588976 B2 01-12-2010
			JP 5189569 B2 24-04-2013
			JP 2004522440 A 29-07-2004
			JP 2004526433 A 02-09-2004
			JP 2010000086 A 07-01-2010
			US 2005191625 A1 01-09-2005
			US 2010311957 A1 09-12-2010
			WO 02059354 A2 01-08-2002
WO 02059355 A2 01-08-2002			
WO 2015031691	A1	05-03-2015	AU 2014312208 A1 10-03-2016
			CN 105705659 A 22-06-2016
			EP 3039158 A1 06-07-2016
			GB 2525104 A 14-10-2015
			GB 2546833 A 02-08-2017
			JP 2016533187 A 27-10-2016
			KR 20160055168 A 17-05-2016
			SG 11201601188T A 30-03-2016
			US 2015299784 A1 22-10-2015
			US 2016010151 A1 14-01-2016
			US 2016145683 A1 26-05-2016
			US 2016265069 A1 15-09-2016
			WO 2015031691 A1 05-03-2015
WO 2015134787	A2	11-09-2015	AU 2015227054 A1 22-09-2016
			CA 2941612 A1 11-09-2015
			EP 3114240 A2 11-01-2017
			WO 2015134787 A2 11-09-2015
WO 2016138500	A1	01-09-2016	US 2016257993 A1 08-09-2016
			WO 2016138500 A1 01-09-2016
US 2010159533	A1	24-06-2010	NONE