



(51) International Patent Classification:

C40B 50/06 (2006.01) C12Q 1/6813 (2018.01)
C12N 15/10 (2006.01) C12Q 1/6869 (2018.01)
C12P 19/34 (2006.01) C12Q 1/6876 (2018.01)

(21) International Application Number:

PCT/US2024/032417

(22) International Filing Date:

04 June 2024 (04.06.2024)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

63/471,618 07 June 2023 (07.06.2023) US
63/535,439 30 August 2023 (30.08.2023) US

(71) Applicant: **CLARET BIOSCIENCE, LLC** [US/US]; 100 Enterprise Way, Suite A102, Scotts Valley, California 95066 (US).

(72) Inventor: **TROLL, Christopher J.**; 100 Enterprise Way, Suite A102, Scotts Valley, California 95066 (US).

(74) Agent: **DICKINSON, Kari A.**; Loza & Loza, LLP, 305 N. Second Ave., #127, Upland, California 91786 (US).

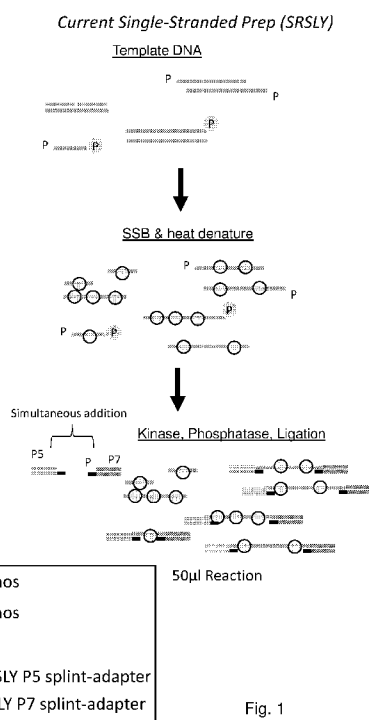
(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ,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

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

(54) Title: METHODS AND COMPOSITIONS FOR ANALYZING NUCLEIC ACID



(57) Abstract: The technology relates in part to methods and compositions for analyzing nucleic acid. In some aspects, the technology relates to methods and compositions for preparing a nucleic acid library from single-stranded nucleic acid fragments.

Fig. 1

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))*

METHODS AND COMPOSITIONS FOR ANALYZING NUCLEIC ACID

Related Patent Application(s)

This patent application claims the benefit of U.S. provisional patent application no. 63/471,618 filed on June 7, 2023, entitled METHODS AND COMPOSITIONS FOR ANALYZING NUCLEIC ACID, naming Christopher J. TROLL as inventor, and designated by attorney docket no. CBS-2010PROV1. This patent application also claims the benefit of U.S. provisional patent application no. 63/535,439 filed on August 30, 2023, entitled METHODS AND COMPOSITIONS FOR ANALYZING NUCLEIC ACID, naming Christopher J. TROLL as inventor, and designated by attorney docket no. CBS-2010PROV2. The entire content of the foregoing patent application is incorporated herein by reference for all purposes, including all text, tables and drawings.

Field

The technology relates in part to methods and compositions for analyzing nucleic acid. In some aspects, the technology relates to methods and compositions for preparing a nucleic acid library from single-stranded nucleic acid fragments.

Background

Genetic information of living organisms (e.g., animals, plants and microorganisms) and other forms of replicating genetic information (e.g., viruses) is encoded in nucleic acid (i.e., deoxyribonucleic acid (DNA) or ribonucleic acid (RNA)). Genetic information is a succession of nucleotides or modified nucleotides representing the primary structure of chemical or hypothetical nucleic acids.

A variety of high-throughput sequencing platforms are used for analyzing nucleic acid. The ILLUMINA platform, for example, involves clonal amplification of adaptor-ligated DNA fragments. Another platform is nanopore-based sequencing, which relies on the transition of nucleic acid molecules or individual nucleotides through a small channel. Library preparation for certain sequencing platforms often includes fragmentation of DNA, modification of fragment ends, and ligation of adapters, and may include amplification of nucleic acid fragments (e.g., PCR amplification).

The selection of an appropriate sequencing platform for particular types of nucleic acid analysis requires a detailed understanding of the technologies available, including sources of error, error rate, as well as the speed and cost of sequencing. While sequencing costs have decreased, the throughput and costs of library preparation can be a limiting factor. One aspect of library preparation includes modification of the ends of nucleic acid fragments such that they are suitable for a particular sequencing platform. Nucleic acid ends may contain useful information. Accordingly, methods that modify nucleic acid ends (e.g., for library preparation) while preserving the

information contained in the nucleic acid ends would be useful for processing and analyzing nucleic acid.

Another aspect of library preparation includes capturing single stranded nucleic acid fragments. In certain instances, single-stranded library preparation methods can generate better and more complex libraries compared to traditional double-stranded DNA (dsDNA) preparation methods. For example, single-stranded library preparation methods can be useful for capturing single-stranded nucleic acid fragments in a mixture of single-stranded DNA fragments and single-stranded RNA fragments. Drawbacks to producing single-stranded DNA (ssDNA) libraries include labor intensive, expensive, and time-consuming protocols, and exotic or custom reagent requirements. Accordingly, methods that capture single-stranded nucleic acids (e.g., for library preparation), without requiring labor intensive, expensive, and time-consuming protocols, and/or exotic or custom reagents would be useful for processing and analyzing nucleic acid (e.g., single-stranded nucleic acid, denatured double-stranded nucleic acid, or mixtures containing single-stranded nucleic acid).

Summary

Provided in certain aspects are methods of producing a nucleic acid library, comprising (a) combining (i) a nucleic acid composition comprising single-stranded nucleic acid (ssNA), (ii) a first oligonucleotide, (iii) a plurality of first scaffold polynucleotide species, (iv) a second oligonucleotide, and (v) a plurality of second scaffold polynucleotide species, where (1) the first oligonucleotide comprises a 5' phosphate; (2) the second oligonucleotide comprises a 3' phosphate; (3) each polynucleotide in the plurality of first scaffold polynucleotide species comprises an ssNA hybridization region and a first oligonucleotide hybridization region; (4) each polynucleotide in the plurality of second scaffold polynucleotide species comprises an ssNA hybridization region and a second oligonucleotide hybridization region; and (5) the nucleic acid composition, the first oligonucleotide, the plurality of first scaffold polynucleotide species, the second oligonucleotide, and the plurality of second scaffold polynucleotide species, are combined under conditions where (i) a molecule of the first scaffold polynucleotide species is hybridized to a first ssNA terminal region and a molecule of the first oligonucleotide, and (ii) a molecule of the second scaffold polynucleotide species is hybridized to a second ssNA terminal region and a molecule of the second oligonucleotide, thereby forming a set of hybridization products in which an end of the molecule of the first oligonucleotide is adjacent to an end of the first ssNA terminal region and an end of the molecule of the second oligonucleotide is adjacent to an end of the second ssNA terminal region; (b) contacting the set of hybridization products with one or more agents comprising a ligase activity, thereby generating a first set of ligation products, where (i) the adjacent ends of the first oligonucleotide and the first ssNA terminal region are ligated; and (ii) the adjacent ends of the

second oligonucleotide and the second ssNA terminal region are not ligated; and (c) contacting the first set of ligation products with one or more agents comprising a 3' phosphatase activity, thereby generating a second set of ligation products where the adjacent ends of the second oligonucleotide and the second ssNA terminal region are ligated.

- 5 Also provided in certain aspects are kits comprising (a) a first oligonucleotide, a plurality of first scaffold polynucleotide species, a second oligonucleotide, and a plurality of second scaffold polynucleotide species, where (i) the first oligonucleotide comprises a 5' phosphate; (ii) the second oligonucleotide comprises a 3' phosphate; (iii) each polynucleotide in the plurality of first scaffold polynucleotide species comprises an ssNA hybridization region and a first oligonucleotide
10 hybridization region; and (iv) each polynucleotide in the plurality of second scaffold polynucleotide species comprises an ssNA hybridization region and a second oligonucleotide hybridization region; (b) an agent comprising a ligase activity; and (c) an agent comprising a 3' phosphatase activity.

- Also provided in certain aspects are methods of producing a nucleic acid library, comprising (a) combining (i) a nucleic acid composition comprising single-stranded nucleic acid (ssNA), (ii) a first
15 oligonucleotide, (iii) a plurality of first scaffold polynucleotide species, (iv) a second oligonucleotide, and (v) a plurality of second scaffold polynucleotide species, where (1) each polynucleotide in the plurality of first scaffold polynucleotide species comprises an ssNA hybridization region and a first oligonucleotide hybridization region; (2) each polynucleotide in the plurality of second scaffold polynucleotide species comprises an ssNA hybridization region and a second oligonucleotide
20 hybridization region; and (3) the nucleic acid composition, the first oligonucleotide, the plurality of first scaffold polynucleotide species, the second oligonucleotide, and the plurality of second scaffold polynucleotide species, are combined under conditions where (i) a molecule of the first scaffold polynucleotide species is hybridized to a first ssNA terminal region and a molecule of the first oligonucleotide, and (ii) a molecule of the second scaffold polynucleotide species is hybridized
25 to a second ssNA terminal region and a molecule of the second oligonucleotide, thereby forming a set of hybridization products; (b) contacting the set of hybridization products with one or more agents comprising a ligase activity, thereby generating a set of ligation products; (c) contacting under hybridization conditions the set of ligation products with one or more probe species, where each probe species comprises a sequence complementary to an RNA sequence, thereby
30 generating a mixture of hybridized ligation products and unhybridized ligation products; and (d) separating the hybridized ligation products from the unhybridized ligation products.

Also provided in certain aspects are kits comprising (a) a first oligonucleotide, a plurality of first scaffold polynucleotide species, a second oligonucleotide, and a plurality of second scaffold polynucleotide species, where (ii) each polynucleotide in the plurality of first scaffold polynucleotide

species comprises an ssNA hybridization region and a first oligonucleotide hybridization region; and (ii) each polynucleotide in the plurality of second scaffold polynucleotide species comprises an ssNA hybridization region and a second oligonucleotide hybridization region; (b) one or more agents comprising a ligase activity; and (c) one or more probe species, where each probe species
5 comprises a sequence complementary to an RNA sequence.

Also provided in certain aspects are methods of producing a nucleic acid library, comprising (a) combining (i) a nucleic acid composition comprising single-stranded nucleic acid (ssNA), (ii) a first oligonucleotide, (iii) a plurality of first scaffold polynucleotide species, (iv) a second oligonucleotide, and (v) a plurality of second scaffold polynucleotide species, where (1) each polynucleotide in the
10 plurality of first scaffold polynucleotide species comprises an ssNA hybridization region and a first oligonucleotide hybridization region; (2) each polynucleotide in the plurality of second scaffold polynucleotide species comprises an ssNA hybridization region and a second oligonucleotide hybridization region; and (3) the nucleic acid composition, the first oligonucleotide, the plurality of first scaffold polynucleotide species, the second oligonucleotide, and the plurality of second
15 scaffold polynucleotide species, are combined under conditions where (i) a molecule of the first scaffold polynucleotide species is hybridized to a first ssNA terminal region and a molecule of the first oligonucleotide, and (ii) a molecule of the second scaffold polynucleotide species is hybridized to a second ssNA terminal region and a molecule of the second oligonucleotide, thereby forming a set of hybridization products; (b) contacting the set of hybridization products with one or more
20 agents comprising a ligase activity, thereby generating a set of ligation products; (c) contacting under hybridization conditions the set of ligation products with one or more probe species, where each probe species comprises a sequence complementary to a ribosomal RNA (rRNA) sequence, thereby generating a mixture of hybridized ligation products and unhybridized ligation products; and (d) separating the hybridized ligation products from the unhybridized ligation products.

Also provided in certain aspects are kits comprising (a) a first oligonucleotide, a plurality of first scaffold polynucleotide species, a second oligonucleotide, and a plurality of second scaffold polynucleotide species, where (i) each polynucleotide in the plurality of first scaffold polynucleotide species comprises an ssNA hybridization region and a first oligonucleotide hybridization region; and (ii) each polynucleotide in the plurality of second scaffold polynucleotide species comprises an
30 ssNA hybridization region and a second oligonucleotide hybridization region; (b) one or more agents comprising a ligase activity; and (c) one or more probe species, where each probe species comprises a sequence complementary to a ribosomal RNA (rRNA) sequence.

Certain implementations are described further in the following description, examples and claims, and in the drawings.

Brief Description of the Drawings

The drawings illustrate certain implementations of the technology and are not limiting. For clarity and ease of illustration, the drawings are not made to scale and, in some instances, various aspects may be shown exaggerated or enlarged to facilitate an understanding of particular implementations.

Fig. 1 shows a current single-stranded prep (SRSLY) workflow.

Fig. 2 shows a modified single-stranded prep (SRSLY 2.0) workflow.

Fig. 3A shows a single-stranded prep (SRSLY 2.0) workflow with RNA depletion. Fig. 3B shows a single-stranded prep (SRSLY 2.0) workflow with ribosomal RNA (rRNA) depletion.

Fig. 4A and Fig 4B show representative RNA-seq results from libraries generated from single-stranded prep (SRSLY 2.0) workflows having certain RNA and/or rRNA depletion steps, starting with 100 ng total RNA input. Mitochondrial rRNA and cytoplasmic rRNA generally are the two main sources of rRNA. Fig. 4A shows how well the workflow performs on each type, and Fig. 4B shows the total performance looking at both types. For each bar graph in Fig. 4A and 4B, samples are from left to right: 1) no depletion, 2) without RNase and without mitochondrial ribosomal RNA (mtrRNA), 3) with RNase and without mtrRNA, and 4) with RNase and with mtrRNA.

Fig. 5 shows rRNA depletion subunit percent mapped reads from RNA-seq results from libraries generated from single-stranded prep (SRSLY 2.0) workflows having no rRNA depletion, 2 rounds of rRNA depletion, or 1 round of rRNA depletion. For each bar graph, samples/libraries are from left to right: 1) 50 ng un-depleted control (74% rRNA), 2) 100 ng un-depleted control (69% rRNA), 3) 500 ng un-depleted control (46% rRNA), 4) 50 ng 2-round depletion (0.33% rRNA), 5) 100 ng 2-round depletion (0.42% rRNA), 6) 500 ng 2-round depletion (2.81% rRNA), 7) 50 ng 1-round depletion (45% rRNA), 8) 100 ng 1-round depletion (46% rRNA), and 9) 500 ng 1-round depletion (37% rRNA).

Fig. 6 shows rRNA depletion RN7SL transcript percent mapped reads from RNA-seq results from libraries generated from single-stranded prep (SRSLY 2.0) workflows having no rRNA depletion, 2 rounds of rRNA depletion, or 1 round of rRNA depletion. For each bar graph, samples/libraries are from left to right: 1) 50 ng un-depleted control (74% rRNA), 2) 100 ng un-depleted control (69% rRNA), 3) 500 ng un-depleted control (46% rRNA), 4) 50 ng 2-round depletion (0.33% rRNA), 5) 100 ng 2-round depletion (0.42% rRNA), 6) 500 ng 2-round depletion (2.81% rRNA), 7) 50 ng 1-round depletion (45% rRNA), 8) 100 ng 1-round depletion (46% rRNA), and 9) 500 ng 1-round depletion (37% rRNA).

Fig. 7 provides a table showing rRNA depletion result data for the samples/libraries shown in Fig. 5 and Fig. 6.

Fig. 8 shows rRNA depletion metrics for two libraries: 1) 100 ng no depletion, and 2) 100 ng with 2-round depletion.

5 Detailed Description

Provided herein are methods and compositions useful for analyzing nucleic acid. Also provided herein are methods and compositions useful for producing nucleic acid libraries. Also provided herein are methods and compositions useful for analyzing single-stranded nucleic acid fragments. In certain aspects, the methods include combining sample nucleic acid comprising single-stranded
10 nucleic acid fragments and specialized adapters. In some embodiments, the specialized adapters include a scaffold polynucleotide capable of hybridizing and ligating to an end of a single-stranded nucleic acid. Products of such hybridization and ligation may be useful for producing a nucleic acid library and/or further analysis or processing, for example. In certain configurations, methods described herein may provide better yields, more uniform ligation, fewer adapter dimers, and other
15 workflow advantages compared to previous single-stranded nucleic acid library preparation methods. Certain methods and compositions useful for analyzing nucleic acid and/or producing nucleic acid libraries are described in International PCT Publication No. WO2019/140201, International PCT Publication No. WO2020/206143, and International PCT Publication No. WO2021/262805, each of which is incorporated by reference in its entirety.

20 *Scaffold adapters*

Certain methods herein comprise combining single stranded nucleic acid (ssNA) with scaffold adapters, or components thereof. Scaffold adapters generally include a scaffold polynucleotide and an oligonucleotide. Accordingly, a "component" of a scaffold adapter may refer to a scaffold polynucleotide and/or an oligonucleotide, or a subcomponent or region thereof. The oligonucleotide
25 and/or the scaffold polynucleotide can be composed of pyrimidine (C, T, U) and/or purine (A, G) nucleotides. Additional components or subcomponents may include one or more of an index polynucleotide, a unique molecular identifier (UMI), one or more regions that flank a unique molecular identifier (UMI), primer binding site (e.g., sequencing primer binding site, P5 primer binding site, P7 primer binding site), flow cell binding region, and the like, and complements
30 thereto. Scaffold adapters comprising a P5 primer binding site may be referred to as P5 adapters or P5 scaffold adapters. Scaffold adapters comprising a P7 primer binding site may be referred to as P7 adapters or P7 scaffold adapters.

A scaffold polynucleotide is a single-stranded component of a scaffold adapter. A polynucleotide herein generally refers to a single-stranded multimer of nucleotide from 5 to 500 nucleotides, e.g., 5 to 100 nucleotides. Polynucleotides may be synthetic or may be made enzymatically, and, in some embodiments, are about 5 to 50 nucleotides in length. Polynucleotides may contain ribonucleotide monomers (i.e., may be polyribonucleotides or “RNA polynucleotides”), deoxyribonucleotide monomers (i.e., may be polydeoxyribonucleotides or “DNA polynucleotides”), or a combination thereof. Polynucleotides may be 10 to 20, 20 to 30, 30 to 40, 40 to 50, 50 to 60, 60 to 70, 70 to 80, 80 to 100, 100 to 150 or 150 to 200, or up to 500 nucleotides in length, for example. The terms polynucleotide and oligonucleotide may be used interchangeably.

A scaffold polynucleotide may include an ssNA hybridization region (also referred to as a scaffold, scaffold region, single-stranded scaffold, single-stranded scaffold region, or splint) and an oligonucleotide hybridization region. An ssNA hybridization region and an oligonucleotide hybridization region may be referred to as subcomponents of a scaffold polynucleotide. An ssNA hybridization region typically comprises a polynucleotide that hybridizes, or is capable of hybridizing, to an ssNA terminal region. An oligonucleotide hybridization region typically comprises a polynucleotide that hybridizes, or is capable of hybridizing, to all or a portion of the oligonucleotide component of the scaffold adapter.

An ssNA hybridization region of a scaffold polynucleotide may comprise a polynucleotide that is complementary, or substantially complementary, to an ssNA terminal region (e.g., an ssDNA terminal region, an sscDNA terminal region, an ssRNA terminal region). In some embodiments, an ssNA hybridization region is an ssDNA hybridization region, an sscDNA hybridization region, or an ssRNA hybridization region. In some embodiments, an sscDNA hybridization region of a scaffold polynucleotide comprises a polynucleotide or subcomponent that is complementary, or substantially complementary, to an RNA specific tag.

In some embodiments, an ssRNA hybridization region of a scaffold polynucleotide comprises a polynucleotide or subcomponent that is complementary, or substantially complementary, to an RNA specific tag. In some embodiments, an ssDNA hybridization region of a scaffold polynucleotide comprises a polynucleotide or subcomponent that is complementary, or substantially complementary, to a DNA specific tag.

In some embodiments, an ssNA hybridization region comprises a random sequence. In some embodiments, an ssNA hybridization region comprises a sequence complementary to an ssNA terminal region sequence of interest (e.g., targeted sequence). In certain embodiments, an ssNA hybridization region comprises one or more nucleotides that are all capable of non-specific base pairing to bases in the ssNA. Nucleotides capable of non-specific base pairing may be referred to

as universal bases. A universal base is a base capable of indiscriminately base pairing with each of the four standard nucleotide bases: A, C, G and T. Universal bases that may be incorporated into the ssNA hybridization region include, but are not limited to, inosine, deoxyinosine, 2'-deoxyinosine (dl, dInosine), nitroindole, 5-nitroindole, and 3-nitropyrrole. In certain embodiments, an ssNA hybridization region comprises one or more degenerate/wobble bases which can replace two or three (but not all) of the four typical bases (e.g., non-natural base P and K).

An ssNA hybridization region of a scaffold polynucleotide may have any suitable length and sequence. In some embodiments, the length of the ssNA hybridization region is 10 nucleotides or less. In certain aspects, the ssNA hybridization region is from 4 to 100 nucleotides in length, e.g., about 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 nucleotides in length. In certain aspects, the ssNA hybridization region is from 4 to 20 nucleotides in length, e.g., from 5 to 15, 5 to 10, 5 to 9, 5 to 8, or 5 to 7 (e.g., 6 or 7) nucleotides in length. In some embodiments, the ssNA hybridization region is 7 nucleotides in length. In some embodiments, the ssNA hybridization region comprises or consists of a random nucleotide sequence, such that when a plurality of heterogeneous scaffold polynucleotides having various random ssNA hybridization regions are employed, the collection is capable of acting as scaffold polynucleotides for a heterogeneous population of ssNAs irrespective of the sequences of the terminal regions of the ssNAs. Each scaffold polynucleotide having a unique ssNA hybridization region sequence may be referred to as a scaffold polynucleotide species and a collection of multiple scaffold polynucleotide species may be referred to as a plurality of scaffold polynucleotide species (e.g., for a scaffold polynucleotide designed to have 7 random bases in the ssNA hybridization region, a plurality of scaffold polynucleotide species would include 4^7 unique ssNA hybridization region sequences). Accordingly, each scaffold adapter having a unique scaffold polynucleotide (i.e., comprising a unique ssNA hybridization region sequence) may be referred to as a scaffold adapter species and a collection of multiple scaffold adapter species may be referred to as a plurality of scaffold adapter species. A species of scaffold polynucleotide generally contains a feature that is unique with respect to other scaffold polynucleotide species. For example, a scaffold polynucleotide species may contain a unique sequence feature. A unique sequence feature may include a unique sequence length, a unique nucleotide sequence (e.g., a unique random sequence, a unique targeted sequence), or a combination of a unique sequence length and nucleotide sequence.

A scaffold polynucleotide may comprise one or more additional subcomponents including an index polynucleotide, a unique molecular identifier (UMI), one or more regions that flank a unique molecular identifier (UMI), primer binding site (e.g., P5 primer binding site, P7 primer binding site), flow cell binding region, and the like, or complementary polynucleotides thereof. A scaffold polynucleotide may comprise a primer binding site (or a polynucleotide complementary to a primer

binding site). Scaffold polynucleotides comprising a P5 primer binding site (or complement thereof) may be referred to as P5 scaffolds or P5 scaffold polynucleotides. Scaffold polynucleotides comprising a P7 primer binding site (or complement thereof) may be referred to as P7 scaffolds or P7 scaffold polynucleotides.

5 An oligonucleotide can be a further single-stranded component of a scaffold adapter. An oligonucleotide herein generally refers to a single-stranded multimer of nucleotides from 5 to 500 nucleotides, e.g., 5 to 100 nucleotides. Oligonucleotides may be synthetic or may be made enzymatically, and, in some embodiments, are 5 to 50 nucleotides in length. Oligonucleotides may contain ribonucleotide monomers (i.e., may be oligoribonucleotides or “RNA oligonucleotides”),
 10 deoxyribonucleotide monomers (i.e., may be oligodeoxyribonucleotides or “DNA oligonucleotides”), or a combination thereof. Oligonucleotides may be 10 to 20, 20 to 30, 30 to 40, 40 to 50, 50 to 60, 60 to 70, 70 to 80, 80 to 100, 100 to 150 or 150 to 200, or up to 500 nucleotides in length, for example. The terms oligonucleotide and polynucleotide may be used interchangeably.

An oligonucleotide component of a scaffold adapter generally comprises a nucleic acid sequence
 15 that is complementary or substantially complementary to an oligonucleotide hybridization region of a scaffold polynucleotide. An oligonucleotide component of a scaffold adapter may include one or more subcomponents useful for one or more downstream applications such as, for example, PCR amplification of the ssNA fragment or derivative thereof, sequencing of the ssNA or derivative thereof, and the like. In some embodiments, a subcomponent of an oligonucleotide is a sequencing
 20 adapter. Sequencing adapter generally refers to one or more nucleic acid domains that include at least a portion of a nucleotide sequence (or complement thereof) utilized by a sequencing platform of interest, such as a sequencing platform provided by Illumina® (e.g., the HiSeq™, MiSeq™ and/or Genome Analyzer™ sequencing systems); Oxford Nanopore™ Technologies (e.g., the MinION™ sequencing system), Ion Torrent™ (e.g., the Ion PGM™ and/or Ion Proton™ sequencing
 25 systems); Pacific Biosciences (e.g., a Sequel or PACBIO RS II sequencing system); Life Technologies™ (e.g., a SOLiD™ sequencing system); Roche (e.g., the 454 GS FLX+ and/or GS Junior sequencing systems); Genapsys; BGI; or any sequencing platform of interest.

In some embodiments, an oligonucleotide component of a scaffold adapter is, or comprises, a nucleic acid domain selected from: a domain (e.g., a “capture site” or “capture sequence”) that
 30 specifically binds to a surface-attached sequencing platform oligonucleotide (e.g., a P5 or P7 oligonucleotide attached to the surface of a flow cell in an Illumina® sequencing system); a sequencing primer binding domain (e.g., a domain to which the Read 1 or Read 2 primers of the Illumina® platform may bind); a unique identifier or index (e.g., a barcode or other domain that uniquely identifies the sample source of the ssNA being sequenced to enable sample multiplexing

by marking every molecule from a given sample with a specific barcode or “tag”); a barcode sequencing primer binding domain (a domain to which a primer used for sequencing a barcode binds); a molecular identification domain or unique molecular identifier (UMI) (e.g., a molecular index tag, such as a randomized tag of 4, 6, or other number of nucleotides) for uniquely marking molecules of interest, e.g., to determine expression levels based on the number of instances a unique tag is sequenced; a complement of any such domains; or any combination thereof. In some embodiments an oligonucleotide comprises one or more regions that flank a unique molecular identifier (UMI). In some embodiments, a barcode domain (e.g., sample index tag) and a molecular identification domain (e.g., a molecular index tag; UMI) may be included in the same nucleic acid. Sequencing platform oligonucleotides, sequencing primers, and their corresponding binding domains can be designed to be compatible with a variety of available sequencing platforms and technologies, including but not limited to those discussed herein.

When an oligonucleotide component of a scaffold adapter includes one or a portion of a sequencing adapter, one or more additional sequencing adapters and/or a remaining portion of the sequencing adapter may be added using a variety of approaches. For example, additional and/or remaining portions of sequencing adapters may be added by any one of ligation, reverse transcription, PCR amplification, and the like. In the case of PCR, an amplification primer pair may be employed that includes a first amplification primer that includes a 3' hybridization region (e.g., for hybridizing to an adapter region of the oligonucleotide) and a 5' region including an additional and/or remaining portion of a sequencing adapter, and a second amplification primer that includes a 3' hybridization region (e.g., for hybridizing to an adapter region of a second oligonucleotide added to the opposite end of an ssNA molecule) and optionally a 5' region including an additional and/or remaining portion of a sequencing adapter.

An oligonucleotide component of a scaffold adapter may comprise one or more additional subcomponents including an RNA-specific tag (also referred to herein as a RNA/cDNA-specific tag, an RNA-specific barcode, or an RNA/cDNA-specific barcode) or a DNA-specific tag (also referred to herein as a DNA-specific barcode). An RNA-specific tag can mark RNA fragments (and/or cDNA generated from RNA) in a sample (e.g., a sample comprising a mixture of RNA and DNA fragments). A DNA-specific tag can mark DNA fragments in a sample (e.g., a sample comprising a mixture of RNA and DNA fragments). Typically, when an RNA-specific tag and a DNA-specific tag are used in the same library preparation, the RNA-specific tag is distinguishable from the DNA-specific tag. For example, the RNA-specific tag and the DNA-specific tag may comprise different sequences; the RNA-specific tag and the DNA-specific tag may comprise different lengths; the RNA-specific tag and the DNA-specific tag may comprise different detectable markers; or any combination of these. An RNA-specific tag or a DNA-specific tag may comprise about 5 to about 15

nucleotides. In some embodiments, an RNA-specific tag comprises 9 nucleotides. In some embodiments, a DNA-specific tag comprises 9 nucleotides. In some embodiments, an RNA-specific tag or a DNA-specific tag is located at a terminus of an oligonucleotide component of a scaffold adapter. In some embodiments, an RNA-specific tag or a DNA-specific tag is located at the 5' terminus of an oligonucleotide component of a scaffold adapter. In some embodiments, an RNA-specific tag or a DNA-specific tag is located at the 3' terminus of an oligonucleotide component of a scaffold adapter. In some embodiments, an RNA-specific tag or a DNA-specific tag is located at a terminus of an oligonucleotide component of a scaffold adapter such that the RNA-specific tag or DNA-specific tag is adjacent to an end of an ssRNA (or sscDNA) terminal region or an end of an ssDNA terminal region when the scaffold adapter is hybridized to the ssRNA (or sscDNA) or ssDNA.

An oligonucleotide component of a scaffold adapter may comprise one or more additional subcomponents including an index polynucleotide, a unique molecular identifier (UMI), one or more regions that flank a unique molecular identifier (UMI), primer binding site (e.g., P5 primer binding site, P7 primer binding site), flow cell binding region or sequencing adapter, and the like, or complementary polynucleotides thereof. An oligonucleotide may comprise a primer binding site (or a polynucleotide complementary to a primer binding site). Oligonucleotides comprising a P5 primer binding site (or complement thereof) may be referred to as P5 oligos or P5 oligonucleotides. Oligonucleotides comprising a P7 primer binding site (or complement thereof) may be referred to as P7 oligos or P7 oligonucleotides.

An oligonucleotide component of a scaffold adapter may comprise a guanine and cytosine (GC)-rich region. A GC-rich region may comprise at least about 50% guanine and cytosine nucleotides. For example, a GC-rich region may comprise about 60% guanine and cytosine nucleotides, about 70% guanine and cytosine nucleotides, about 80% guanine and cytosine nucleotides, about 90% guanine and cytosine nucleotides, or 100% guanine and cytosine nucleotides. In some embodiments, a GC-rich region comprises about 70% guanine and cytosine nucleotides. An oligonucleotide component of a scaffold adapter may comprise a guanine and cytosine (GC)-rich region at one end (e.g., at a 3' end or at a 5' end). In some embodiments, an oligonucleotide component of a scaffold adapter comprises a guanine and cytosine (GC)-rich region at the end of the oligonucleotide that is joined to an ssNA fragment (i.e., at the oligonucleotide-ssNA junction or "ligation terminus"). A scaffold polynucleotide may comprise a corresponding region that is complementary to the GC-rich region in the oligonucleotide.

An oligonucleotide component of a scaffold adapter may be phosphorylated at an end or may be dephosphorylated (i.e., may have a free OH) at an end. In some embodiments, an oligonucleotide

component of a scaffold adapter may be phosphorylated at a 5' end. In some embodiments, an oligonucleotide component of a scaffold adapter may be phosphorylated at a 3' end. In some embodiments, a first oligonucleotide component of a first scaffold adapter may be phosphorylated at a 5' end, and a second oligonucleotide component of a second scaffold adapter may be phosphorylated at a 3' end. In some embodiments, an oligonucleotide comprises a blocking modification (e.g., a blocking modification described herein) at one end and a phosphate at the other end. In some embodiments, an oligonucleotide comprises a blocking modification at a 3' end and a phosphate at a 5' end. In some embodiments, an oligonucleotide comprises a blocking modification at a 5' end and a phosphate at a 3' end. In some embodiments, a first oligonucleotide component of a first scaffold adapter comprises a blocking modification at a 3' and a phosphate at a 5' end, and a second oligonucleotide component of a second scaffold adapter comprises a blocking modification at a 5' end and a phosphate at a 3' end.

The scaffold polynucleotide may be hybridized to the oligonucleotide, forming a duplex in the scaffold adapter. Accordingly, a scaffold adapter may be referred to as a scaffold duplex, a duplex adapter, a duplex oligonucleotide, or a duplex polynucleotide. Each scaffold duplex having a unique scaffold polynucleotide (i.e., comprising a unique ssNA hybridization region sequence) may be referred to as a scaffold duplex species and a collection of multiple scaffold duplex species may be referred to as a plurality of scaffold duplex species. In some embodiments, the scaffold polynucleotide and the oligonucleotide are on separate DNA strands. In some embodiments, the scaffold polynucleotide and the oligonucleotide are on a single DNA strand (e.g., a single DNA strand capable of forming a hairpin structure).

Scaffold adapters can comprise DNA, RNA, or a combination thereof. Scaffold adapters can comprise a DNA scaffold polynucleotide and a DNA oligonucleotide, a DNA scaffold polynucleotide and an RNA oligonucleotide, an RNA scaffold polynucleotide and a DNA oligonucleotide, or an RNA scaffold polynucleotide and an RNA oligonucleotide. In one example configuration, a scaffold adapter comprises a DNA scaffold polynucleotide and a DNA oligonucleotide for combining with an RNA sample nucleic acid, and example ligases for use with such an adapter/sample configuration include T4 RNA ligase 2, T4 DNA ligase, truncated T4 RNA ligase 2, and thermostable 5' App DNA/RNA ligase. In another example adapter configuration, a scaffold adapter comprises a DNA scaffold polynucleotide and an RNA oligonucleotide for combining with an RNA sample nucleic acid, and example ligases for use with such an adapter/sample configuration include T4 RNA ligase 1, T4 RNA ligase 2, truncated T4 RNA ligase 2, and thermostable 5' App DNA/RNA ligase. In another example adapter configuration, a scaffold adapter comprises an RNA scaffold polynucleotide and an RNA oligonucleotide for combining with an RNA sample nucleic acid, and example ligases for use with such an adapter/sample configuration include T4 RNA ligase 1, T4

RNA ligase 2, truncated T4 RNA ligase 2, and thermostable 5' App DNA/RNA ligase. In some instances, the adapter nucleotide composition is selected to provide homogeneity between sample nucleic acids and scaffold adapter nucleic acids (e.g., such that at least the oligonucleotide is homogenous to the sample nucleic acids). In some instances, the adapter nucleotide composition is selected to provide homogeneity between the oligonucleotide and the sample nucleic acids and heterogeneity between the scaffold polynucleotide and the sample nucleic acids.

Unique molecular identifier (UMI)

In some embodiments, a scaffold adapter comprises a unique molecular identifier (UMI). In some embodiments, an oligonucleotide (e.g., an oligonucleotide component of a scaffold adapter) comprises a unique molecular identifier (UMI). Unique molecular identifiers (UMIs), which also may be referred to as molecular barcodes, barcodes, molecular identification domains, molecular index tags, sequence tags, and/or tags, generally are short sequences (e.g., about 3 to about 10 nucleotides in length) that may be added to nucleic acid fragments during nucleic acid library preparation to identify or mark input nucleic acid molecule(s). In certain applications, UMIs may be useful for uniquely marking molecules of interest, e.g., to determine expression levels based on the number of instances a unique tag is sequenced. UMIs typically are added prior to an amplification step (e.g., PCR amplification), and may be useful for reducing errors and quantitative bias introduced by amplification, for example. Scaffold adapters and/or oligonucleotide components of scaffold adapters comprising a UMI as described herein may be referred to as comprising an "in-line" UMI. An in-line UMI generally refers to a UMI sequence that is a component a scaffold adapter and/or an oligonucleotide described herein that becomes part of the sequence read generated by the sequencing of an ssNA fragment ligated to an oligonucleotide component of the scaffold adapter. When a scaffold adapter comprises an in-line UMI, library generation may not require certain additional processing steps (e.g., addition of a UMI to the adapter by way of an extension step using a strand displacing polymerase).

Combining scaffold adapters, or components thereof, and ssNA

A method herein may comprise combining one or more scaffold adapters, or components thereof, with a composition comprising single-stranded nucleic acid (ssNA) to form one or more complexes. The scaffold polynucleotide is designed for simultaneous hybridization to an ssNA fragment and an oligonucleotide component such that, upon complex formation, an end of the oligonucleotide component is adjacent to an end of the terminal region of the ssNA fragment. Typically, upon complex formation, a 5' end of the oligonucleotide component is adjacent to a 3' end of the terminal region of the ssNA, or a 5' end of the oligonucleotide component is adjacent to a 3' end of the terminal region of the ssNA. Upon complex formation in instances where a scaffold adapter is

attached to both ends of an ssNA fragment, a 5' end of one oligonucleotide component is adjacent to a 3' end of one terminal region of the ssNA, and a 5' end of a second oligonucleotide component is adjacent to a 3' end of a second terminal region of the ssNA.

In some embodiments, a method includes forming complexes by combining an ssNA composition, an oligonucleotide, and a plurality of heterogeneous scaffold polynucleotides having various random ssNA hybridization regions capable of acting as scaffolds for a heterogeneous population of ssNA having terminal regions of undetermined sequence. In some embodiments, a method includes forming complexes by combining an ssNA composition, a plurality of heterogeneous oligonucleotides having various UMI configurations, and a plurality of heterogeneous scaffold polynucleotides having various random ssNA hybridization regions capable of acting as scaffolds for a heterogeneous population of ssNA having terminal regions of undetermined sequence. In some embodiments, a method includes forming complexes by combining an ssNA composition, an oligonucleotide or a plurality of heterogeneous oligonucleotides having various UMI configurations, and a plurality of heterogeneous scaffold polynucleotides, where the scaffold polynucleotides are provided in an amount that exceeds the amount of oligonucleotides. In some embodiments, scaffold polynucleotides and oligonucleotides are provided at a ratio of at least 1.1 to 1 (scaffold polynucleotides to oligonucleotides). For example, scaffold polynucleotides and oligonucleotides may be provided at a ratio of at least 1.2 to 1, 1.3 to 1, 1.4 to 1, 1.5 to 1, 1.6 to 1, 1.7 to 1, 1.8 to 1, 1.9 to 1, or 2 to 1. In some embodiments, scaffold polynucleotides and oligonucleotides are provided at a ratio of 1.4 to 1 (scaffold polynucleotides to oligonucleotides). For example, a method may comprise combining an ssNA composition with 14 μ M scaffold polynucleotides and 10 μ M oligonucleotides.

In some embodiments, an ssNA hybridization region includes a known sequence designed to hybridize to an ssNA terminal region of known sequence. In some embodiments, two or more heterogeneous scaffold polynucleotides having different ssNA hybridization regions of known sequence are designed to hybridize to respective ssNA terminal regions of known sequence. Embodiments in which the ssNA hybridization regions have a known sequence may be useful, for example, for producing a nucleic acid library from a subset of ssNAs having terminal regions of known sequence. Accordingly, in certain embodiments, a method herein comprises forming complexes by combining an ssNA composition, an oligonucleotide, and one or more heterogeneous scaffold polynucleotides having one or more different ssNA hybridization regions of known sequence capable of acting as scaffolds for one or more ssNAs having one or more terminal regions of known sequence.

An ssNA fragment, an oligonucleotide, and scaffold polynucleotide may be combined in various ways. In some configurations, the combining includes combining 1) a complex comprising the scaffold polynucleotide hybridized to the oligonucleotide component via the oligonucleotide hybridization region, and 2) the ssNA fragment. In another configuration, the combining includes combining 1) a complex comprising the scaffold polynucleotide hybridized to the ssNA fragment via the ssNA hybridization region, and 2) the oligonucleotide component. In another configuration, the combining includes combining 1) the ssNA fragment, 2) the oligonucleotide, and 3) the scaffold polynucleotide, where none of the three components are pre-complexed with, or hybridized to, another component prior to the combining.

The combining may be carried out under hybridization conditions such that complexes form including a scaffold polynucleotide hybridized to a terminal region of an ssNA fragment via the ssNA hybridization region, and the scaffold polynucleotide hybridized to an oligonucleotide component via the oligonucleotide hybridization region. Whether specific hybridization occurs may be determined by factors such as the degree of complementarity between the hybridizing regions of the scaffold polynucleotide, the terminal region of the ssNA fragment, and the oligonucleotide component, as well as the length thereof, salt concentration, GC content, and the temperature at which the hybridization occurs, which may be informed by the melting temperatures (T_m) of the relevant regions.

Complexes may be formed such that an end of an oligonucleotide component is adjacent to an end of a terminal region of an ssNA fragment. Adjacent to refers the terminal nucleotide at the end of the oligonucleotide and the terminal nucleotide end of the terminal region of the ssNA fragment are sufficiently proximal to each other that the terminal nucleotides may be covalently linked, for example, by chemical ligation, enzymatic ligation, or the like. In some embodiments, the ends are adjacent to each other by virtue of the terminal nucleotide at the end of the oligonucleotide and the terminal nucleotide end of the terminal region of the ssNA being hybridized to adjacent nucleotides of the scaffold polynucleotide. The scaffold polynucleotide may be designed to ensure that an end of the oligonucleotide is adjacent to an end of the terminal region of the ssNA fragment.

In some embodiments, complexes may be formed such that an end of an RNA-specific tag in an oligonucleotide component is adjacent to an end of a terminal region of an ssRNA (or sscDNA) fragment. Adjacent to refers the terminal nucleotide at the end of the RNA-specific tag and the terminal nucleotide end of the terminal region of the ssRNA (or sscDNA) fragment are sufficiently proximal to each other that the terminal nucleotides may be covalently linked, for example, by chemical ligation, enzymatic ligation, or the like. In some embodiments, the ends are adjacent to each other by virtue of the terminal nucleotide at the end of the RNA-specific tag and the terminal

nucleotide end of the terminal region of the ssRNA (or sscDNA) being hybridized to adjacent nucleotides of the scaffold polynucleotide. The scaffold polynucleotide may be designed to ensure that an end of the RNA-specific tag is adjacent to an end of the terminal region of the ssRNA (or sscDNA) fragment.

- 5 In some embodiments, complexes may be formed such that an end of a DNA-specific tag in an oligonucleotide component is adjacent to an end of a terminal region of an ssDNA fragment. Adjacent to refers the terminal nucleotide at the end of the DNA-specific tag and the terminal nucleotide end of the terminal region of the ssDNA fragment are sufficiently proximal to each other that the terminal nucleotides may be covalently linked, for example, by chemical ligation, enzymatic
- 10 ligation, or the like. In some embodiments, the ends are adjacent to each other by virtue of the terminal nucleotide at the end of the DNA-specific tag and the terminal nucleotide end of the terminal region of the ssDNA being hybridized to adjacent nucleotides of the scaffold polynucleotide. The scaffold polynucleotide may be designed to ensure that an end of the DNA-specific tag is adjacent to an end of the terminal region of the ssDNA fragment.
- 15 A scaffold polynucleotide may be designed with one or more uracil bases in place of thymine. In some embodiments, one of the strands in a scaffold adapter duplex may be degraded by generating multiple cut sites at uracil bases, for example by using a uracil-DNA glycosylase and an endonuclease.

Scaffold adapters comprising in-line UMI designs described herein may be configured to connect to

20 one or both ends of an ssNA fragment. In some configurations, scaffold adapters are designed such that the adapter species that connects to the 5' end of an ssNA comprises an in-line UMI design described herein. In some configurations, scaffold adapters are designed such that the adapter species that connects to the 3' end of an ssNA comprises an in-line UMI design described herein. In some configurations, scaffold adapters are designed such that the adapter species that

25 connects to the 5' end of an ssNA comprises an in-line UMI design described herein and the adapter species that connects to the 3' end of the ssNA does not include an in-line UMI. In some configurations, scaffold adapters are designed such that the adapter species that connects to the 3' end of an ssNA comprises an in-line UMI design described herein and the adapter species that connects to the 5' end of the ssNA does not include an in-line UMI. In some configurations, scaffold

30 adapters are designed such that the adapter species that connects to the 5' end of an ssNA comprises an in-line UMI design described herein and the adapter species that connects to the 3' end of the ssNA also comprises an in-line UMI design described herein.

Scaffold adapters, oligonucleotide components, and scaffold polynucleotides may be referred to herein as first scaffold adapters (or first scaffold duplexes), first oligonucleotide components (or first

oligonucleotides), first unique molecular identifiers (UMIs), and first scaffold polynucleotides; or second scaffold adapters (or second scaffold duplexes), second oligonucleotide components (or second oligonucleotides), second unique molecular identifiers (UMIs), and second scaffold polynucleotides. The terms first and second generally refer to scaffold adapters, or components thereof, that hybridize to and/or are covalently linked to a first end and second end of an ssNA fragment terminus (i.e., a 5' end and a 3' end). The terms first end and second end do not always refer to a particular directionality of the ssNA fragment. Accordingly, a first end of an ssNA terminus may be a 5' end or a 3' end, and a second end of an ssNA terminus may be a 5' end or a 3' end. A first scaffold adapter, or component thereof, may refer to a P5 adapter, or component thereof, or a P7 adapter, or component thereof. A second scaffold adapter, or component thereof, may refer to a P5 adapter, or component thereof, or a P7 adapter, or component thereof.

In certain workflow configurations, ssNA, a first oligonucleotide, a plurality of first scaffold polynucleotide species, a second oligonucleotide, and a plurality of second scaffold polynucleotide species, are combined under conditions where (i) a molecule of the first scaffold polynucleotide species is hybridized to a first ssNA terminal region and a molecule of the first oligonucleotide, and (ii) a molecule of the second scaffold polynucleotide species is hybridized to a second ssNA terminal region and a molecule of the second oligonucleotide. In such configurations, a first oligonucleotide, the plurality of first scaffold polynucleotide species, the second oligonucleotide, and the plurality of second scaffold polynucleotide species are simultaneously combined with the ssNA. In such configurations, a set of hybridization products is formed. In the hybridization products, a 5' end of the molecule of the first oligonucleotide is adjacent to a 3' end of the first ssNA terminal region. Furthermore, a 3' end of the molecule of the second oligonucleotide is adjacent to a 5' end of the second ssNA terminal region. In certain configurations, the 5' end of the first oligonucleotide is phosphorylated and is adjacent to the 3' end of the first ssNA terminal region in the hybridization products. In certain configurations, the 3' end of the second oligonucleotide is phosphorylated and is adjacent to the 5' end of the second ssNA terminal region in the hybridization products.

In some instances, prior to combining scaffold adapters or components thereof with a nucleic acid sample comprising ssNA, the nucleic acid sample can be treated with a nuclease to remove unwanted nucleic acids. For example, a double-stranded specific nuclease (e.g., T7 nuclease) can be used to digest some or all double-stranded DNA, and scaffolding adapters can then be used to prepare a sequencing library of the remaining nucleic acids as disclosed herein. In an example, a double-stranded specific nuclease is used to digest double-stranded nucleic acids in a sample, leaving intact single-stranded nucleic acids such as those from single-stranded DNA viruses,

single-stranded RNA viruses, and single-stranded DNA (e.g., damaged DNA) while digesting double-stranded DNA from a host organism and/or bacteria.

Combining scaffold adapters, or components thereof, and ssRNA or sscDNA

A method herein may comprise combining one or more scaffold adapters, or components thereof, with a composition comprising single-stranded ribonucleic acid (ssRNA) or single-stranded complementary deoxyribonucleic acid (sscDNA) to form one or more complexes. The scaffold polynucleotide is designed for simultaneous hybridization to an ssRNA or sscDNA fragment and an oligonucleotide component such that, upon complex formation, an end of the oligonucleotide component is adjacent to an end of the terminal region of the ssRNA or sscDNA fragment, as described above for ssNA.

In some embodiments, a nucleic acid composition comprises sscDNA. In some embodiments, a method comprises prior to the combining, generating sscDNA from single-stranded ribonucleic acid (ssRNA). Typically, when a nucleic acid composition comprises sscDNA, a method herein uses a first-strand cDNA and does not require generating a second-strand cDNA. Thus, in some embodiments, a nucleic acid composition comprises first-strand sscDNA. In some embodiments, a nucleic acid composition consists essentially of first-strand sscDNA. A nucleic acid composition “consisting essentially of” first-strand sscDNA generally includes first-strand sscDNA and no additional protein or nucleic acid components. A nucleic acid composition consisting essentially of first-strand sscDNA generally does not comprise second-strand sscDNA. Additionally, for example, a nucleic acid composition “consisting essentially of” first-strand sscDNA may exclude double-stranded cDNA (dscDNA) or may include a low percentage of dscDNA (e.g., less than 10% dscDNA, less than 5% dscDNA, less than 1% dscDNA). A nucleic acid composition “consisting essentially of” first-strand sscDNA may exclude proteins. For example, a nucleic acid composition “consisting essentially of” first-strand sscDNA may exclude single-stranded binding proteins (SSBs) or other proteins useful for stabilizing first-strand sscDNA. A nucleic acid composition “consisting essentially of” first-strand sscDNA may include chemical components typically present in nucleic acid compositions such as buffers, salts, alcohols, crowding agents (e.g., PEG), and the like; and may include residual components (e.g., nucleic acids (e.g., residual RNA), proteins, cell membrane components) from the nucleic acid source (e.g., sample), from nucleic acid extraction, or from cDNA synthesis. A nucleic acid composition “consisting essentially of” first-strand sscDNA may include first-strand sscDNA fragments having one or more phosphates (e.g., a terminal phosphate, a 5' terminal phosphate). A nucleic acid composition “consisting essentially of” first-strand sscDNA may include first-strand sscDNA fragments comprising one or more modified nucleotides.

In some embodiments, generating the sscDNA comprises contacting the ssRNA with a primer and an agent comprising a reverse transcriptase activity, thereby generating a DNA-RNA duplex. In some embodiments, generating the sscDNA may further comprise contacting the DNA-RNA duplex with an agent comprising an RNase activity, thereby digesting the RNA and generating an sscDNA product. In some embodiments, the agent comprising a reverse transcriptase activity also comprises an RNase activity. Accordingly, in some embodiments, reverse transcription and RNase digestion are combined into one step. In some embodiments, the agent comprising a reverse transcriptase activity and an RNase activity is an M-MuLV reverse transcriptase (also referred to as M-MLV reverse transcriptase). The primer or primers may be any primer or primers suitable for use in conjunction with a reverse transcriptase. The primer or primers may be chosen from one or more of a random primer (e.g., random hexamer primer, random octamer primer), and a poly(T) primer. An sscDNA product may be purified by a suitable purification or wash method, e.g., a purification or wash method described herein.

In some embodiments, a nucleic acid composition comprises ssRNA. In such embodiments, scaffold adapters are directly hybridized to the ssRNA fragments, and the oligonucleotide component(s) is/are covalently linked to one or more ends of the ssRNA termini, thereby forming hybridization products containing one or more scaffold adapters and an ssRNA fragment. In some embodiments, a method further comprises generating single-stranded ligation products from the hybridization products (e.g., by denaturing the hybridization products). In such embodiments, single-stranded ligation products comprise an ssRNA fragment covalently linked to one or more oligonucleotide components. In some embodiments, a method further comprises contacting the single-stranded ligation products with a primer and an agent comprising a reverse transcriptase activity, thereby generating a DNA-RNA duplex. In some embodiments, a method further comprises contacting the DNA-RNA duplex with an agent comprising an RNase activity, thereby digesting the RNA and generating a single-stranded cDNA (sscDNA) product. In some embodiments, the agent comprising a reverse transcriptase activity also comprises an RNase activity. Accordingly, in some embodiments, reverse transcription and RNase digestion are combined into one step. In some embodiments, the agent comprising a reverse transcriptase activity and an RNase activity is an M-MuLV reverse transcriptase (also referred to as M-MLV reverse transcriptase). The primer may be any primer suitable for use in conjunction with a reverse transcriptase. In some embodiments, the primer comprises a nucleotide sequence complementary to a sequence in an oligonucleotide component (i.e., an oligonucleotide component covalently linked to an ssRNA fragment). An sscDNA product may be purified by a suitable purification or wash method, e.g., a purification or wash method described herein.

In some embodiments, an sscDNA product is amplified. An sscDNA product may be amplified by a suitable amplification method, e.g., an amplification method described herein. In some embodiments, amplifying an sscDNA product may be combined (e.g., combined in a single step, reaction, vessel, and/or volume) with generating a DNA-RNA duplex and/or generating an sscDNA product. Accordingly, reagents for generating a DNA-RNA duplex (e.g., one or more agents comprising a reverse transcriptase activity), reagents for generating an sscDNA product (e.g., one or more agents comprising an RNase activity), and reagents for amplifying an sscDNA product (e.g., primers, an agent comprising a polymerase activity), may be combined for use in a single step, reaction, vessel, and/or volume. In some embodiments, reagents for amplifying an sscDNA product comprise amplification primers that hybridize to a component (e.g., first oligonucleotide) of the scaffold adapters described herein. The amplification primers may be any primer suitable for use in conjunction with a polymerase. In some embodiments, each primer comprises a nucleotide sequence complementary to a sequence in an sscDNA product corresponding to an oligonucleotide component (i.e., an oligonucleotide component covalently linked to an ssRNA fragment). An amplified sscDNA product may be purified by a suitable purification or wash method, e.g., a purification or wash method described herein.

In some embodiments, a method herein comprises prior to combining the ssRNA with scaffold adapters, or components thereof, or prior to generating the sscDNA, fragmenting the ssRNA, thereby generating ssRNA fragments. Any suitable fragmentation method may be used, such as, for example, a fragmentation method described herein. In some embodiments, a method herein comprises prior to combining the ssRNA with scaffold adapters, or components thereof, or prior to generating the sscDNA, depleting ribosomal RNA (rRNA) and/or enriching messenger RNA (mRNA). In some embodiments, a method herein comprises after combining the ssRNA with scaffold adapters, or components thereof, or after generating the sscDNA and combining with scaffold adapters, or components thereof, depleting ribosomal RNA (rRNA) and/or enriching messenger RNA (mRNA). Any suitable rRNA depletion method and/or mRNA enrichment method may be used, such as, for example, an rRNA depletion method and/or mRNA enrichment method described herein.

Hybridization and ligation

Nucleic acid fragments (e.g., ssNA fragments) may be combined with scaffold adapters, or components thereof, thereby generating combined products. Combining ssNA fragments with scaffold adapters, or components thereof, may comprise hybridization and/or ligation (e.g., ligation of hybridization products). A combined product may include an ssNA fragment connected to (e.g., hybridized to and/or ligated to) a scaffold adapter, or component thereof, at one or both ends of the

ssNA fragment. A combined product may include an ssNA fragment hybridized to a scaffold adapter, or component thereof, at one or both ends of the ssNA fragment, which may be referred to as a hybridization product. A combined product may include an ssNA fragment ligated to a scaffold adapter, or component thereof, at one or both ends of the ssNA fragment, which may be referred to as a ligation product. In some embodiments, products from a cleavage step (i.e., cleaved products) may be combined with scaffold adapters, or components thereof, thereby generating combined products.

Certain methods herein comprise generating sets of combined products (e.g., a first set of combined products and a second set of combined products). In some embodiments, a first set of combined products includes ssNAs connected to (e.g., hybridized to and/or ligated to) scaffold adapters, or components thereof, from a first set of scaffold adapters, or components thereof. In some embodiments, a second set of combined products includes the first set of combined products connected to (e.g., hybridized to and/or ligated to) scaffold adapters, or components thereof, from a second set of scaffold adapters, or components thereof.

Certain methods herein comprise generating a single set of combined products. In some embodiments, a set of combined products includes ssNAs connected to (e.g., hybridized to) scaffold adapters, or components thereof, from a first set of scaffold adapters, or components thereof. The set of combined products further includes ssNAs connected to (e.g., hybridized to) scaffold adapters, or components thereof, from a second set of scaffold adapters, or components thereof.

ssNAs may be combined with scaffold adapters, or components thereof, under hybridization conditions, thereby generating hybridization products. In some embodiments, the scaffold adapters are provided as pre-hybridized products and the hybridization step includes hybridizing the scaffold adapters to the ssNA. In some embodiments, the scaffold adapter components (i.e., oligonucleotides and scaffold polynucleotides) are provided as individual components and the hybridization step includes hybridizing the scaffold adapter components 1) to each other and 2) to the ssNA. In some embodiments, the scaffold adapter components (i.e., oligonucleotides and scaffold polynucleotides) are provided sequentially as individual components and the hybridization steps includes 1) hybridizing the scaffold polynucleotides to the ssNA, and then 2) hybridizing the oligonucleotides to the oligonucleotide hybridization region of the scaffold polynucleotides. The conditions during the combining step are those conditions in which scaffold adapters, or components thereof (e.g., single-stranded scaffold regions), specifically hybridize to ssNAs having a terminal region or terminal regions that are complementary in sequence with respect to the single-stranded scaffold regions. The conditions during the combining step also may include those

conditions in which components of the scaffold adapters (e.g., oligonucleotides and oligonucleotide hybridization regions within the scaffold polynucleotides), specifically hybridize, or remain hybridized, to each other.

Specific hybridization may be affected or influenced by factors such as the degree of

5 complementarity between the single-stranded scaffold regions and the ssNA terminal region(s), or between the oligonucleotides and oligonucleotide hybridization regions, the length thereof, and the temperature at which the hybridization occurs, which may be informed by melting temperatures (T_m) of the single-stranded scaffold regions. Melting temperature generally refers to the temperature at which half of the single-stranded scaffold regions /ssNA terminal regions remain
10 hybridized and half of the single-stranded scaffold regions /ssNA terminal regions dissociate into single strands. The T_m of a duplex may be experimentally determined or predicted using the following formula $T_m = 81.5 + 16.6(\log_{10}[\text{Na}^+]) + 0.41 (\text{fraction G+C}) - (60/N)$, where N is the chain length and $[\text{Na}^+]$ is less than 1 M. Additional models that depend on various parameters also may be used to predict T_m of relevant regions depending on various hybridization conditions. Generally,
15 the on-off rate of hybridization that occurs between ssNA terminal regions and single-stranded scaffold regions is essentially constant (prior to ligation). Approaches for achieving specific nucleic acid hybridization are described, e.g., Tijssen, Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes, part I, chapter 2, "Overview of principles of hybridization and the strategy of nucleic acid probe assays," Elsevier (1993).

20 In some embodiments, a method herein comprises exposing hybridization products to conditions under which an end of an ssNA is joined to an end of a scaffold adapter to which it is hybridized. In particular, a method herein may comprise exposing hybridization products to conditions under which an end of an ssNA is joined to an end of an oligonucleotide component of a scaffold adapter to which it is hybridized. Joining may be achieved by any suitable approach that permits covalent
25 attachment of ssNA to the scaffold adapter and/or oligonucleotide component of a scaffold adapter to which it is hybridized. When one end of an ssNA is joined to an end of a scaffold adapter and/or oligonucleotide component of a scaffold adapter to which it is hybridized, typically one of two attachment events is conducted: 1) the 3' end of the ssNA to the 5' end of the oligonucleotide component of the scaffold adapter, or 2) the 5' end of the ssNA to the 3' end of the oligonucleotide
30 component of the scaffold adapter. When both ends of an ssNA are each joined to an end of a scaffold adapter and/or oligonucleotide component of a scaffold adapter to which it is hybridized, typically two attachment events are conducted: 1) the 3' end of the ssNA to the 5' end of the oligonucleotide component of a first scaffold adapter, and 2) the 5' end of the ssNA to the 3' end of the oligonucleotide component of a second scaffold adapter.

In some embodiments, a method herein comprises contacting hybridization products with an agent comprising a ligase activity under conditions in which an end of an ssNA is covalently linked to an end of a scaffold adapter and/or oligonucleotide component of a scaffold adapter to which the target nucleic acid (ssNA) is hybridized. Ligase activity may include, for example, blunt-end ligase activity, nick-sealing ligase activity, sticky end ligase activity, circularization ligase activity, cohesive end ligase activity, DNA ligase activity, RNA ligase activity, single-stranded ligase activity, and double-stranded ligase activity. Ligase activity may include ligating a 5' phosphorylated end of one polynucleotide to a 3' OH end of another polynucleotide (5'P to 3'OH). Ligase activity may include ligating a 3' phosphorylated end of one polynucleotide to a 5' OH end of another polynucleotide (3'P to 5'OH). Ligase activity may include ligating a 5' end of an ssNA to a 3' end of a scaffold adapter and/or oligonucleotide component of a scaffold adapter hybridized thereto in a ligation reaction. Ligase activity may include ligating a 3' end of an ssNA to a 5' end of a scaffold adapter and/or oligonucleotide component of a scaffold adapter hybridized thereto in a ligation reaction. Suitable reagents (e.g., ligases) and kits for performing ligation reactions are known and available. For example, Instant Sticky-end Ligase Master Mix available from New England Biolabs (Ipswich, MA) may be used. Ligases that may be used include but are not limited to, for example, T3 ligase, T4 DNA ligase (e.g., at low or high concentration), T7 DNA Ligase, E. coli DNA Ligase, Electro Ligase®, RNA ligases, T4 RNA ligase 1, T4 RNA ligase 2, truncated T4 RNA ligase 2, thermostable 5' App DNA/RNA ligase, SplintR® Ligase, RtcB ligase, Taq ligase, and the like and combinations thereof. When needed, a phosphate group may be added at the 5' end of the oligonucleotide component or ssNA fragment using a suitable kinase, for example, such as T4 polynucleotide kinase (PNK). Such kinases and guidance for using such kinases to phosphorylate 5' ends are available, for example, from New England BioLabs, Inc. (Ipswich, MA).

In some embodiments, a method comprises covalently linking the adjacent ends of an oligonucleotide component and an ssNA terminal region, thereby generating covalently linked hybridization products. In some embodiments, the covalently linking comprises contacting the hybridization products (e.g., ssNA fragments hybridized to at least one scaffold adapter herein) with an agent comprising a ligase activity under conditions in which the end of an ssNA terminal region is covalently linked to an end of the oligonucleotide component. In some embodiments, a method comprises covalently linking the adjacent ends of a first oligonucleotide component and a first ssNA terminal region, and covalently linking the adjacent ends of a second oligonucleotide component and a second ssNA terminal region, thereby generating covalently linked hybridization products. In some embodiments, the covalently linking comprises contacting hybridization products (e.g., ssNA fragments each hybridized two scaffold adapters herein) with an agent comprising a ligase activity under conditions in which an end of a first ssNA terminal region is covalently linked to an end of a

first oligonucleotide component and an end of a second ssNA terminal region is covalently linked to an end of a second oligonucleotide component. In some embodiments, the agent comprising a ligase activity is a T4 DNA ligase. In some embodiments, the T4 DNA ligase is used at an amount between about 1 unit/ μ l to about 50 units/ μ l. In some embodiments, the T4 DNA ligase is used at an amount between about 5 unit/ μ l to about 30 units/ μ l. In some embodiments, the T4 DNA ligase is used at an amount between about 5 unit/ μ l to about 15 units/ μ l. In some embodiments, the T4 DNA ligase is used at about 10 units/ μ l. In some embodiments, the T4 DNA ligase is used at an amount less than 25 units/ μ l. In some embodiments, the T4 DNA ligase is used at an amount less than 20 units/ μ l. In some embodiments, the T4 DNA ligase is used at an amount less than 15 units/ μ l. In some embodiments, the T4 DNA ligase is used at an amount less than 10 units/ μ l.

In some embodiments, a method comprises sequentially covalently linking the adjacent ends of a first oligonucleotide component to a first ssNA terminal region, and the adjacent ends of a second oligonucleotide component to a second ssNA terminal region. In some embodiments, the covalently linking comprises contacting a set of hybridization products with one or more agents comprising a ligase activity. In a sequential ligation, a first set of ligation products is generated, where (i) the adjacent ends of a first oligonucleotide and a first ssNA terminal region are ligated; and (ii) the adjacent ends of a second oligonucleotide and a second ssNA terminal region are not ligated. In certain configurations, the second oligonucleotide is blocked from ligating to the second ssNA terminal region during the initial ligation reaction. Any suitable ligation blocking mechanism may be used to prevent the second oligonucleotide from ligating to the second ssNA terminal region during the initial ligation reaction. In some embodiments, the second oligonucleotide is 3' phosphorylated, which blocks the second oligonucleotide from ligating to the second ssNA terminal region during the initial ligation reaction. In a subsequent step (i.e., after the initial ligation of the first oligonucleotide to the first ssNA terminal region is complete or substantially complete), the ligation blocking mechanism is removed, allowing the second oligonucleotide to ligate to the second ssNA terminal region. A second set of ligation products is formed where the adjacent ends of the second oligonucleotide and the second ssNA terminal region are ligated. In some embodiments, a method comprises contacting the first set of ligation products described above with one or more agents comprising a phosphatase activity. The phosphatase removes the 3' phosphate from the second oligonucleotide and allows the second oligonucleotide to ligate to the second ssNA terminal region. In some embodiments, a method comprises contacting the first set of ligation products described above with one or more agents comprising a phosphatase activity. In some embodiments, a method comprises contacting the first set of ligation products described above with one or more agents comprising a 3' phosphatase activity. In some embodiments, a method comprises contacting the first set of ligation products described above with one or more agents comprising a

3' phosphatase activity and a kinase activity. In some embodiments, a method comprises contacting the first set of ligation products described above with one or more agents comprising a kinase activity (e.g., a kinase that also includes a phosphatase activity (e.g., a 3' phosphatase activity)). In one example, the phosphatase component of the kinase removes the 3' phosphate from the second oligonucleotide and phosphorylates the 5' end of the ssNA, thus allowing the second oligonucleotide to ligate to the second ssNA terminal region. In some embodiments, a kinase is a polynucleotide kinase (PNK). In some embodiments, a kinase is a T4 polynucleotide kinase (T4 PNK). In some embodiments, additional ligase is added to the subsequent ligation reaction. In some embodiments, no additional ligase is added to the subsequent ligation reaction (e.g., when the ligase from the first ligation reaction is still present and active).

In some embodiments, hybridization products are contacted with a first agent comprising a first ligase activity and a second agent comprising a second ligase activity different than the first ligase activity. For example, the first ligase activity and the second ligase activity independently may be chosen from blunt-end ligase activity, nick-sealing ligase activity, sticky end ligase activity, circularization ligase activity, and cohesive end ligase activity, double-stranded ligase activity, single-stranded ligase activity, 5'P to 3'OH ligase activity, and 3'P to 5'OH ligase activity.

In some embodiments, a method herein comprises joining ssNAs to scaffold adapters and/or oligonucleotide components of scaffold adapters via biocompatible attachments. Methods may include, for example, click chemistry or tagging, which include biocompatible reactions useful for joining biomolecules. In some embodiments, an end of each of the oligonucleotide components comprises a first chemically reactive moiety and an end of each of the ssNAs includes a second chemically reactive moiety. In such embodiments, the first chemically reactive moiety typically is capable of reacting with the second chemically reactive moiety and forming a covalent bond between an oligonucleotide component of a scaffold adapter and an ssNA to which the scaffold adapter is hybridized. In some embodiments, a method herein includes contacting ssNA with one or more chemical agents under conditions in which the second chemically reactive moiety is incorporated at an end of each of the ssNA fragments. In some embodiments, a method herein includes exposing hybridization products to conditions in which the first chemically reactive moiety reacts with the second chemically reactive moiety forming a covalent bond between an oligonucleotide component and an ssNA to which the scaffold adapter is hybridized. In some embodiments, the first chemically reactive moiety is capable of reacting with the second chemically reactive moiety to form a 1,2,3-triazole between the oligonucleotide component and the ssNA to which the scaffold adapter is hybridized. In some embodiments, the first chemically reactive moiety is capable of reacting with the second chemically reactive moiety under conditions comprising copper. The first and second chemically reactive moieties may include any suitable pairings. For

example, the first chemically reactive moiety may be chosen from an azide-containing moiety and 5-octadiynyl deoxyuracil, and the second chemically reactive moiety may be independently chosen from an azide-containing moiety, hexynyl and 5-octadiynyl deoxyuracil. In some embodiments, the azide-containing moiety is N-hydroxysuccinimide (NHS) ester-azide.

5 Covalently linking the adjacent ends of an oligonucleotide and an ssNA fragment produces a covalently linked product, which may be referred to a ligation product. A covalently linked product that includes an ssNA fragment covalently linked to an oligonucleotide component, which remain hybridized to a scaffold polynucleotide, may be referred to as a covalently linked hybridization product. A covalently linked hybridization product may be denatured (e.g., heat-denatured) to
10 separate the ssNA fragment covalently linked to an oligonucleotide component from the scaffold polynucleotide. A covalently linked product that includes an ssNA fragment covalently linked to an oligonucleotide component, which is no longer hybridized to a scaffold polynucleotide (e.g., after denaturing), may be referred to as a single-stranded ligation product. In some instances, portions of a scaffold polynucleotide can be cleaved and/or degraded, for example by using uracil-DNA
15 glycosylase and an endonuclease at one or more uracil bases in the scaffold polynucleotide.

A covalently linked hybridization product and/or single-stranded ligation product may be purified prior to use as input in a downstream application of interest (e.g., amplification; sequencing). For example, covalently linked hybridization products and/or single-stranded ligation products may be purified from certain components present during the combining, hybridization, and/or covalently
20 linking (ligation) steps (e.g., by solid phase reversible immobilization (SPRI), column purification, and/or the like).

In some embodiments, when a method herein includes combining an ssNA composition with scaffold adapters herein, or components thereof, and covalently linking the adjacent ends of an oligonucleotide component and an ssNA fragment, the total duration of the combining and
25 covalently linking may be 4 hours or less, 3 hours or less, 2 hours or less, or 1 hour or less. In some embodiments, the total duration of the combining and covalently linking is less than 1 hour.

In some embodiments, a method herein is performed in a single vessel, a single chamber, and/or a single volume (i.e., contiguous volume), including but not limited to on a microfluidic device. In some embodiments, combining an ssNA composition with scaffold adapters herein, or components thereof, and covalently linking the adjacent ends of an oligonucleotide component and an ssNA
30 fragment are performed in a single vessel, a single chamber, and/or a single volume (i.e., contiguous volume), including but not limited to on a microfluidic device. In some embodiments, a method herein is performed in a collection of wells, droplets, emulsion, partitions, or other reaction volumes, including but not limited to on a microfluidic device. In some embodiments, combining an

ssNA composition with scaffold adapters herein, or components thereof, and covalently linking the adjacent ends of an oligonucleotide component and an ssNA fragment are performed in a collection of wells, droplets, emulsion, partitions, or other reaction volumes, including but not limited to on a microfluidic device. In some instances, the collection of reaction volumes are prepared such that a majority or all of the reaction volumes comprise at most one ssNA. In some instances, the collection of reaction volumes are prepared such that a majority or all of the reaction volumes comprise at most 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 20000, 30000, 40000, 50000, 60000, 70000, 80000, 90000, 100000, or more ssNA. Partitioning one or a limited number of ssNA into reaction volumes can provide favorable reaction kinetics, such as increasing the library conversion of rare species of sample nucleic acids.

Adapter dimers

In some embodiments, a method herein comprises one or more modifications and/or additional steps for preventing, reducing, or eliminating adapter dimers. Adapter dimers may unintentionally form during a method described herein. Adapter dimers generally refer to two or more scaffold adapters, components thereof, or parts thereof hybridizing, or hybridizing and ligating, to each other.

In certain embodiments, a scaffold adapter, or a component thereof, is modified to prevent adapter dimer formation. Described herein is a workflow that can reduce adapter dimer formation. In particular, described herein is a workflow, where one member of a pair of adapters comprises a 3' phosphorylated oligonucleotide, which blocks ligation to other adapters. Once the other member of the pair of adapters is ligated to the target nucleic acid, and the risk of adapter dimer formation is reduced, the phosphate is removed from the 3' phosphorylated oligonucleotide and the de-phosphorylated adapter is ligated to the other end of the target nucleic acid.

Other examples of modifications to a scaffold adapter include modified nucleotides capable of blocking covalent linkage of the scaffold adapter, oligonucleotide component, or scaffold polynucleotide, to another oligonucleotide, polynucleotide, or nucleic acid molecule (e.g., another scaffold adapter, oligonucleotide component, and/or scaffold polynucleotide). Examples of modified nucleotides are described below. Other/additional modifications to a scaffold adapter include configurations such as a Y-configuration or a hairpin configuration, which are described in further detail below. In some embodiments, scaffold adapter, oligonucleotide component, and/or scaffold polynucleotide may comprise a phosphorothioate backbone modification (e.g., a phosphorothioate bond between the last two nucleotides on a strand).

In some embodiments, a method includes a dephosphorylation step to prevent or reduce adapter dimer formation. In some embodiments, a method includes prior to combining scaffold adapters, or components thereof, with ssNA, contacting scaffold adapters, oligonucleotide components, and/or scaffold polynucleotides with an agent comprising a phosphatase activity under conditions in which the scaffold adapters, oligonucleotide components, and/or scaffold polynucleotides is/are dephosphorylated, thereby generating dephosphorylated scaffold adapters, dephosphorylated oligonucleotide components, and/or dephosphorylated scaffold polynucleotides.

In some embodiments, a method includes one or more staged ligation approaches to prevent or reduce adapter dimer formation. In some embodiments, a method includes staged ligation which comprises delaying addition of an agent comprising a phosphoryl transfer activity (e.g., until after hybridization products are formed) and/or delaying addition of a second scaffold adapter, or components thereof. For example, a method may comprise after forming hybridization products and prior to covalently linking the oligonucleotide component(s) to the ssNA terminal region(s), contacting the oligonucleotide component(s) with an agent comprising a phosphoryl transfer activity under conditions in which a 5' phosphate is added to a 5' end of an oligonucleotide component. In another example, a method may comprise combining a first set of scaffold adapters with ssNA. A first set of scaffold adapters may include an oligonucleotide component having a 3' OH. The first set of scaffold adapters are hybridized to the ssNA, and the 3' OH of the oligonucleotide component is covalently linked to the 5' end (e.g., 5' phosphorylated end) of an ssNA terminal region. The products of such first round of hybridizing and covalently linking may be referred to as intermediate covalently linked hybridization products. The intermediate covalently linked hybridization products are then combined with a second set of scaffold adapters. A second set of scaffold adapters may include an oligonucleotide component having a 5' end that may be phosphorylated as described herein. The second set of scaffold adapters are hybridized to the intermediate covalently linked hybridization products, and the 5' phosphorylated end of the oligonucleotide component is covalently linked to the 3' end of the ssNA terminal region.

In some embodiments, a method includes staged ligation which comprises use of a scaffold adapter, or component thereof, having an adenylation modification. For example, a first set of scaffold adapters may comprise an adenylation modification at the 5' end of the oligonucleotide component (5' App). The first set of scaffold adapters are hybridized to the ssNA, and the 5' App of the oligonucleotide component is covalently linked to the 3' end of an ssNA terminal region. The covalent linking may occur in the absence of ATP. The products of such first round of hybridizing and covalently linking may be referred to as intermediate covalently linked hybridization products. The intermediate covalently linked hybridization products are then combined with a second set of scaffold adapters. A second set of scaffold adapters may include an oligonucleotide component

having a 3' OH end. The second set of scaffold adapters are hybridized to the intermediate covalently linked hybridization products, and the 3' OH end of the oligonucleotide component is covalently linked to the 5' end (e.g., 5' phosphorylated end) of the ssNA terminal region (with the addition of ATP). In one variation, the first set of scaffold adapters and the second set of scaffold
5 adapters are combined with ssNA at the same time in the absence of ATP. Ligation of the first set of scaffold adapters may proceed in the absence of ATP, and ligation of the second set of scaffold adapters may proceed only until ATP is added.

In some embodiments, a method includes staged ligation which comprises use of an oligonucleotide (i.e., a single stranded oligonucleotide) having a 3' phosphorylated end. An
10 oligonucleotide having a 3' phosphorylated end may comprise any of the subcomponents described herein for oligonucleotide components of scaffold adapters (e.g., a primer binding site, an index, a UMI, a flow cell adapter, and the like). An oligonucleotide having a 3' phosphorylated end generally is single-stranded and is not hybridized to a scaffold polynucleotide. In one example, a method may comprise prior to combining scaffold adapters, or components thereof, with ssNA, combining the
15 ssNA with an oligonucleotide comprising a phosphate at the 3' end and covalently linking the 3' phosphorylated end of the oligonucleotide to the 5' end (e.g., 5' non-phosphorylated end) of an ssNA terminal region. In some embodiments, prior to the covalently linking of the oligonucleotide to the ssNA, the ssNA is contacted with an agent comprising a phosphatase activity under conditions in which the ssNA is dephosphorylated, thereby generating dephosphorylated ssNA. In some
20 embodiments, covalently linking the oligonucleotide to the ssNA comprises contacting the ssNA and the oligonucleotide with an agent comprising a single-stranded ligase activity under conditions in which the 5' end of the ssNA is covalently linked to the 3' end of the oligonucleotide. In some embodiments, the agent comprising a ligase activity is an RtcB ligase. The products of such covalently linking may be referred to as intermediate covalently linked products. The intermediate
25 covalently linked products are then combined with a set of scaffold adapters. A set of scaffold adapters may include an oligonucleotide component having a 5' phosphorylated end. The set of scaffold adapters are hybridized to the intermediate covalently linked products, and the 5' phosphorylated end of the oligonucleotide component is covalently linked to the 3' end of the ssNA terminal region.

In some embodiments, a method includes use of an oligonucleotide capable of hybridizing to an oligonucleotide dimer product to reduce or eliminate adapter dimers. An oligonucleotide dimer product may be a component of a scaffold adapter dimer, and may contain an oligonucleotide component from a first scaffold adapter covalently linked to an oligonucleotide component from a second scaffold adapter. A method herein may include a denaturing step which can release the
35 oligonucleotide dimer product from the scaffold adapter dimer. The oligonucleotide dimer product

may hybridize to an oligonucleotide having a sequence complementary to the oligonucleotide dimer product, or part thereof, thereby forming an oligonucleotide dimer hybridization product. In some embodiments, the oligonucleotide dimer hybridization product comprises a cleavage site. In some embodiments, the cleavage site is a restriction enzyme recognition site. In some embodiments, a method herein further comprises contacting the oligonucleotide dimer hybridization product with a cleavage agent (e.g., a restriction enzyme, a rare-cutter restriction enzyme).

In some embodiments, a method includes purifying or washing nucleic acid products at various stages of library preparation to reduce or eliminate adapter dimers. In some instances, purifying or washing nucleic acid products may reduce or eliminate adapter dimers. For example, covalently linked hybridization products (i.e., ssNA hybridized to scaffold adapters and covalently linked to oligonucleotide components), single-stranded ligation products (i.e., denatured covalently linked hybridization products; ssNA covalently linked to oligonucleotide components and no longer hybridized to scaffold polynucleotides), or amplification products thereof, may be purified or washed by any suitable purification or washing method. In some embodiments, purifying or washing comprises use of solid phase reversible immobilization (SPRI). SPRI beads can be resuspended in a DNA binding buffer containing, for example, about 2.5 M to about 5 M NaCl, about 0.1 mM to about 1 M EDTA, about 10 mM Tris, about 0.01% to about 0.05% TWEEN-20, and between about 8% and about 38% PEG-8000. For example, 1 ml of SPRI bead suspension can be combined with 2.5 M NaCl, 10 mM Tris, 1 mM EDTA, 0.05% Tween-20 and 20% PEG-8000. In some embodiments, SPRI includes serial SPRI (washes performed back-to-back) and or sequential SPRI (wash comprising sequential addition of SPRI beads and incubations). Serial SPRI may include a plurality of serial (back-to-back) washes, which may include 2, 3, 4, 5, 6, 7, 8, 9, 10 or more serial washes. Sequential SPRI may include a plurality of sequential addition of SPRI beads (with intervening incubations), which may include 2, 3, 4, 5, 6, 7, 8, 9, 10 or more sequential addition of SPRI beads. In some embodiments, the amount of SPRI beads used in an SPRI purification may include an amount between 0.1x to 3x SPRI beads (x is ratio of beads to nucleic acid (e.g., bead volume to reaction volume)). For example, the amount of SPRI beads used in an SPRI purification may include about 0.1x, 0.2x, 0.3x, 0.4x, 0.5x, 0.6x, 0.7x, 0.8x, 0.9x, 1.0x, 1.1x, 1.2x, 1.3x, 1.4x, 1.5x, 1.6x, 1.7x, 1.8x, 1.9x, 2.0x, 2.1x, 2.2x, 2.3x, 2.4x, 2.5x, 2.6x, 2.7x, 2.8x, 2.9x, or 3.0x SPRI beads. In some embodiments, the amount of SPRI beads used in an SPRI purification is 1.2x. In some embodiments, the amount of SPRI beads used in an SPRI purification is 1.5x. In some embodiments, purifying or washing comprises a column purification (e.g., column chromatography). In some embodiments, purifying or washing does not comprise a column purification (e.g., column chromatography). In some embodiments, covalently linked hybridization products, single-stranded ligation products, and/or amplification products thereof are not purified or washed.

An SPRI purification is typically performed in the presence of a buffer. Any suitable buffer may be used, e.g., Tris buffer, water that is of similar pH, and the like. SPRI purification beads may be added directly to a sample solution (e.g., a sample solution containing covalently linked hybridization products (ligation products), or amplified products thereof). In certain instances, buffer
5 may be added to raise the volume of the reaction so additional beads may be added. In some embodiments, an SPRI bead solution is made up of carboxylated magnetic beads added to PEG 8000 dissolved in water, NaCl, Tris, and EDTA. The amount of PEG typically determines the PEG percentage of the SPRI bead solution. For example, adding 9 g of PEG 8000 in a 50 ml SPRI bead solution may be referred to as "18% SPRI." In another example, adding 19 g of PEG 8000 in a 50
10 ml SPRI solution may be referred to as "38% SPRI." Generally, the higher proportion of PEG, the lower the size of DNA fragments retained.

In some embodiments, a purification process comprises contacting covalently linked hybridization products (ligation products) with solid phase reversible immobilization (SPRI) beads and a buffer. In some embodiments, some or all SPRI buffer is replaced with isopropanol. In some embodiments,
15 SPRI buffer comprises isopropanol. In some embodiments, SPRI buffer is completely replaced with isopropanol. In some embodiments, SPRI buffer comprises about 5% volume/volume (v/v) isopropanol to about 50% v/v isopropanol. In some embodiments, SPRI buffer comprises about 10% v/v isopropanol to about 40% v/v isopropanol. For example, SPRI buffer may comprise about 10% v/v isopropanol, 15% v/v isopropanol, 20% v/v isopropanol, 25% v/v isopropanol, 30% v/v
20 isopropanol, 35% v/v isopropanol, or 40% v/v isopropanol. In some embodiments, SPRI buffer comprises about 20% v/v isopropanol.

In some embodiments, a purifying or washing step may enrich for nucleic acid fragments, or amplification products thereof, having a particular length or range of lengths. In some
25 embodiments, an SPRI purification may enrich for nucleic acid fragments, or amplification products thereof, having a particular length or range of lengths. In some embodiments, the amount of PEG 8000 in an SPRI bead solution used in an SPRI purification may affect the length or range of lengths of fragments that are enriched. For example, an SPRI purification at 1.5x v/v ratio may recover more fragments in the < 100 base range than an SPRI purification at 1.2x because the final concentration of PEG 8000 is higher in 1.5x than in 1.2x. In some embodiments, a method herein
30 comprises adjusting an SPRI ratio to enrich for a desired fragment length or range of lengths. In some embodiments, a method herein comprises adjusting an amount of isopropanol in an SPRI purification to enrich for a desired fragment length or range of lengths. In some embodiments, a method herein comprises adjusting an amount of isopropanol in an SPRI purification to enrich for a desired fragment length or range of lengths, while minimizing the amount of unwanted artifacts
35 (e.g., adapter dimers). For example, a method herein may comprise adjusting an amount of

isopropanol in an SPRI purification to enrich for a desired fragment length or range of lengths, where the amount of adapter dimers recovered is less than about 10% of the total nucleic acid recovered. In another example, a method herein may comprise adjusting an amount of isopropanol in an SPRI purification to enrich for a desired fragment length or range of lengths, where the amount of adapter dimers recovered is less than about 5% of the total nucleic acid recovered.

In some embodiments, a method herein (e.g., combining ssNA with scaffold adapters or components thereof, hybridization, and covalently linking) may be performed in a suitable reaction volume and/or with a suitable amount of ssNA and/or suitable ratio of ssNA to scaffold adapters (or components thereof). A suitable reaction volume and/or a suitable amount of ssNA and/or a suitable ratio of ssNA to scaffold adapters (or components thereof) may include reaction volumes, amounts of ssNA, and/or ratios of ssNA and scaffold adapters that reduce or prevent adapter dimer formation. In some embodiments, a suitable amount of ssNA may range from about 250 pg to about 5 ng of ssNA. For example, a suitable amount of ssNA may be about 250 pg, 500 pg, 750 pg, 1 ng, 1.5 ng, 2 ng, 2.5 ng, 3 ng, 3.5 ng, 4 ng, 4.5 ng, or 5 ng. In some embodiments, a suitable amount of ssNA may be about 1 ng of ssNA. In some embodiments, for a 25 µl final reaction volume, 1 ng ssNA may be combined with between about 1.0 to 2.0 picomoles of each scaffold adapter (i.e., about 1.0 to 2.0 picomoles of scaffold adapters (pool of scaffold adapters that contains a plurality of scaffold adapter species) that hybridize to the 5' end of ssNA terminal regions, and about 1.0 to 2.0 picomoles of scaffold adapters (pool of scaffold adapters that contains a plurality of scaffold adapter species) that hybridize to the 3' end of ssNA terminal regions). For example, for a 25 µl final reaction volume, 1 ng ssNA may be combined with about 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, or 2.0 picomoles of each scaffold adapter. In some embodiments, for a 25 µl final reaction volume, 1 ng ssNA is combined with about 1.6 picomoles of each scaffold adapter (i.e., about 1.6 picomoles of scaffold adapters that hybridize to the 5' end of ssNA terminal regions and about 1.6 picomoles of scaffold adapters that hybridize to the 3' end of ssNA terminal regions). For larger reaction volumes, amounts of ssNA and scaffold adapters may be scaled up so long as the relative amounts are preserved. For smaller reaction volumes, amounts of ssNA and scaffold adapters may be scaled down so long as the relative amounts are preserved. In some embodiments, the scaffold adapters herein are combined with ssNA at a molar ratio between about 5:1 (scaffold adapters to ssNA) to about 50:1 (scaffold adapters to ssNA). For example, scaffold adapters may combined with ssNA at a molar ratio of about 5:1 (scaffold adapters to ssNA), about 10:1 (scaffold adapters to ssNA), about 15:1 (scaffold adapters to ssNA), about 20:1 (scaffold adapters to ssNA), about 25:1 (scaffold adapters to ssNA), about 30:1 (scaffold adapters to ssNA), about 35:1 (scaffold adapters to ssNA), about 40:1 (scaffold adapters to ssNA), about 45:1 (scaffold adapters to ssNA), or about 50:1 (scaffold adapters to ssNA). In some embodiments,

scaffold adapters are combined with ssNA at a molar ratio of about 15:1 (scaffold adapters to ssNA). In some embodiments, scaffold adapters are combined with ssNA at a molar ratio of about 30:1 (scaffold adapters to ssNA).

In some embodiments, a method herein comprises use of a crowding agent. A suitable amount of crowding agent may be used to reduce or prevent adapter dimer formation. Crowding agents may include, for example, ficoll 70, dextran 70, polyethylene glycol (PEG) 2000, and polyethylene glycol (PEG) 8000. In some embodiments, a method herein comprises use of polyethylene glycol (PEG) 8000. PEG, for example, may be used in an amount between about 15% to about 20%, which percentages refer to final concentrations of PEG in a ligation reaction. For example, PEG may be used at about 15%, 15.5%, 16%, 16.5%, 17%, 17.5%, 18%, 18.5%, 19%, 19.5%, or 20%. In some embodiments, 18.5% PEG is used. In some embodiments, 18% PEG is used.

During purification, an SPRI bead solution may be added to a sample solution, often with instructions for a v/v ratio. For example, 1.2x 18% SPRI means that, if given a 50 µl sample, add 60 µl (50 x 1.2) of 18% SPRI beads. This v/v ratio leads to a final concentration of PEG at 9.8%, assuming there is in no PEG in the sample solution. However, often after ligation, there is an existing amount of PEG present in the sample solution (i.e., ligation products). Accordingly, a user may adjust the volume of added SPRI beads to reach the desired final concentration of PEG. A desired final concentration of PEG may range from about 5% final PEG to about 15% final PEG. For example, a desired final concentration of PEG may be about 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, or 15%. In some embodiments, a desired final concentration of PEG is about 10% (e.g., for hair samples and cfDNA samples). In some embodiments, a desired final concentration of PEG is about 12% (e.g., for formalin-fixed paraffin-embedded (FFPE) samples and samples with large template fragments).

Y-adapters

In some embodiments, scaffold adapters described herein comprise two strands, with single-stranded scaffold region at a first end and two non-complementary strands at a second end. Such scaffold adapters may be referred to as Y-scaffold adapters, Y-adapters, Y-shaped scaffold adapters, Y-shaped adapters, Y-duplexes, Y-shaped duplexes, Y-scaffold duplexes, Y-shaped scaffold duplexes, and the like. A scaffold adapter having a Y-shaped structure generally comprises a double-stranded duplex region, two single stranded “arms” at one end, and single-stranded scaffold region at the other end.

Y-scaffold adapters may comprise a plurality of nucleic acid components and subcomponents. In some embodiments, Y-scaffold adapters comprise a first nucleic acid strand and a second nucleic acid strand. In some embodiments, a first nucleic acid strand is complementary to a second nucleic

acid strand. In some embodiments, a portion of a first nucleic acid strand is complementary to a portion of a second nucleic acid strand. In some embodiments, a first nucleic acid strand comprises a first region that is complementary to a first region in a second nucleic acid strand, and the first polynucleotide comprises a second region that is not complementary to a second region in the second polynucleotide. The complementary region often forms the duplex region of the Y-scaffold adapter and the non-complementary region often forms the arms, or parts thereof, of the Y-scaffold adapter. The first and second nucleic acid strands may comprise subcomponents (e.g., subcomponents of scaffold polynucleotides, subcomponents of oligonucleotides and subcomponents of sequencing adapters described herein, such as, for example, UMIs, UMI flanking regions, amplification priming sites and/or specific sequencing adapters (e.g., P5, P7 adapters)). In some embodiments, the first and second nucleic acid strands do not comprise certain subcomponents of sequencing adapters described herein, such as, for example, amplification priming sites and/or specific sequencing adapters (e.g., P5, P7 adapters).

In some embodiments, a Y-scaffold adapter comprises a single-stranded scaffold region (ssNA hybridization region). The single-stranded scaffold region of a Y-scaffold adapter typically is located adjacent to the double-stranded duplex portion and at the opposite end of the non-complementary strands (or “arms”) portion. The single-stranded scaffold region of a Y-scaffold adapter typically is complementary to a terminal region of a target nucleic acid (e.g., a terminal region of a single-stranded nucleic acid).

Hairpins

In some embodiments, a scaffold adapter comprises one strand capable of forming a hairpin structure having a single-stranded loop. In some embodiments, a scaffold adapter consists of one strand capable of forming a hairpin structure having a single-stranded loop. A scaffold adapter having a hairpin structure generally comprises a double-stranded “stem” region and a single stranded “loop” region. In some embodiments, a scaffold adapter comprises one strand (i.e., one continuous strand) capable of adopting a hairpin structure. In some embodiments, a scaffold adapter consists essentially of one strand (i.e., one continuous strand) capable of adopting a hairpin structure. Consisting essentially of one strand means that the scaffold adapter does not include any additional strands of nucleic acid (e.g., hybridized to the scaffold adapter) that are not part of the continuous strand. Thus, “consisting essentially of” here refers to the number of strands in the scaffold adapter, and the scaffold adapter can include other features not essential to the number of strands (e.g., can include a detectable label, can include other regions). A scaffold adapter comprising or consisting essentially of one strand capable of forming a hairpin structure may be referred to herein as a hairpin, hairpin scaffold adapter, or hairpin adapter.

Hairpin scaffold adapters may comprise a plurality of nucleic acid components and subcomponents within the one strand. In some embodiments, a hairpin scaffold adapter comprises an oligonucleotide and a scaffold polynucleotide. In some embodiments, the oligonucleotide is complementary to an oligonucleotide hybridization region in the scaffold polynucleotide. In some
5 embodiments, a portion of the oligonucleotide is complementary to a portion of the oligonucleotide hybridization region in the scaffold polynucleotide. In some embodiments, a hairpin scaffold adapter comprises complementary region and a non-complementary region. The complementary region often forms the stem of the hairpin adapter and the non-complementary region often forms the loop, or part thereof, of the hairpin scaffold adapter. The oligonucleotide and the scaffold
10 polynucleotide may comprise subcomponents (e.g., subcomponents of scaffold polynucleotides, subcomponents of oligonucleotides, and subcomponents of sequencing adapters described herein, such as, for example, UMIs, UMI flanking regions, amplification priming sites and/or specific sequencing adapters (e.g., P5, P7 adapters)). In some embodiments, the oligonucleotide and the scaffold polynucleotide do not comprise certain subcomponents of sequencing adapters described
15 herein, such as, for example, amplification priming sites and specific sequencing adapters (e.g., P5, P7 adapters).

Hairpin scaffold adapters may comprise one or more cleavage sites capable of being cleaved under cleavage conditions. In some embodiments, a cleavage site is located between an oligonucleotide and a scaffold polynucleotide. Cleavage at a cleavage site often generates two
20 separate strands from the hairpin scaffold adapter. In some embodiments, cleavage at a cleavage site generates a partially double stranded scaffold adapter with two unpaired strands forming a “Y” structure. Cleavage sites may include any suitable cleavage site, such as cleavage sites described herein, for example. In some embodiments, cleavage sites comprise RNA nucleotides and may be cleaved, for example, using an RNase. In some embodiments, cleavage sites comprise uracil
25 and/or deoxyuridine and may be cleaved, for example, using DNA glycosylase, endonuclease, RNase, and the like and combinations thereof. In some embodiments, cleavage sites do not comprise uracil and/or deoxyuridine. In some embodiments, a method herein comprises after combining hairpin scaffold adapters with single-stranded nucleic acids, exposing one or more cleavage sites to cleavage conditions, thereby cleaving the scaffold adapters.

In some embodiments, a hairpin scaffold adapter comprises a single-stranded scaffold region (ssNA hybridization region). The single-stranded scaffold region of a hairpin scaffold adapter typically is located adjacent to the double-stranded stem portion and at the opposite end of the loop portion. The single-stranded scaffold region of a hairpin scaffold adapter typically is complementary to a terminal region of a target nucleic acid (e.g., a terminal region of a single-stranded nucleic
35 acid).

In some embodiments, a hairpin scaffold adapter comprises in a 5' to 3' orientation: an oligonucleotide, one or more cleavage sites, and a scaffold polynucleotide comprising an oligonucleotide hybridization region and a scaffold region (ssNA hybridization region). In some embodiments, a hairpin oligonucleotide comprises in a 5' to 3' orientation: a scaffold polynucleotide comprising a scaffold region (ssNA hybridization region) and an oligonucleotide hybridization region, one or more cleavage sites, and an oligonucleotide. In some embodiments, a plurality or pool of hairpin scaffold adapter species comprises a mixture of: 1) hairpin scaffold adapters comprising in a 5' to 3' orientation: an oligonucleotide, one or more cleavage sites, and a scaffold polynucleotide comprising an oligonucleotide hybridization region and a scaffold region (ssNA hybridization region); and 2) hairpin scaffold adapters comprising in a 5' to 3' orientation: a scaffold polynucleotide comprising a scaffold region (ssNA hybridization region) and an oligonucleotide hybridization region, one or more cleavage sites, and an oligonucleotide.

Modified nucleotides

In some embodiments, a nucleic acid comprises one or more modified nucleotides. In some embodiments, a DNA molecule comprises one or more modified nucleotides. In some embodiments, an unligated DNA molecule (e.g., not ligated to one or more scaffold adapters described herein) comprises one or more modified nucleotides. In some embodiments, a scaffold adapter, or component thereof, comprises one or more modified nucleotides. In some embodiments, an unligated scaffold adapter, or component thereof, (e.g., not ligated to a DNA molecule) comprises one or more modified nucleotides. Modified nucleotides may be referred to as modified bases or non-canonical bases and may include, for example, nucleotides conjugated to a member of a binding pair, blocked nucleotides, non-natural nucleotides, nucleotide analogues, peptide nucleic acid (PNA) nucleotides, Morpholino nucleotides, locked nucleic acid (LNA) nucleotides, bridged nucleic acid (BNA) nucleotides, glycol nucleic acid (GNA) nucleotides, threose nucleic acid (TNA) nucleotides, and the like and combinations thereof. In certain configurations, a scaffold adapter, or component thereof comprises one or more nucleotides with modifications chosen from one or more of amino modifier, biotinylation, thiol, alkynes, 2'-O-methoxy-ethyl Bases (2'-MOE), RNA, fluoro bases, iso (iso-dG, iso-DC), inverted, methyl, nitro, phos, and the like.

In some embodiments, a scaffold adapter, or component thereof, comprises one or more modified nucleotides within a duplex region, within a scaffold region, at one end, or at both ends of the scaffold adapter, or component thereof. In some embodiments, a scaffold adapter, or component thereof, comprises one or more unpaired modified nucleotides. In some embodiments, a scaffold adapter, or component thereof, comprises one or more unpaired modified nucleotides at one end of the adapter. In some embodiments, a scaffold adapter, or component thereof, comprises one or

more unpaired modified nucleotides at the end of the adapter opposite to the end that hybridizes to a target nucleic acid (e.g., an end comprising a single-stranded scaffold region). A modified nucleotide may be present at the end of the strand having a 3' terminus or at the end of the strand having a 5' terminus.

- 5 In some embodiments, a nucleic acid molecule comprises one or more modified nucleotides. In some embodiments, the one or more modified nucleotides are capable of blocking covalent linkage of the nucleic acid molecule to another nucleic acid molecule, oligonucleotide, polynucleotide, or scaffold adapter. In some embodiments, a nucleic acid molecule comprises one or more modified nucleotides at one or more unligated ends. In some embodiments, an oligonucleotide component
- 10 comprises one or more modified nucleotides. In some embodiments, the one or more modified nucleotides are capable of blocking covalent linkage of the oligonucleotide component to another oligonucleotide, polynucleotide, or nucleic acid molecule. In some embodiments, an oligonucleotide component comprises one or more modified nucleotides at an end not adjacent to the ssNA. In some embodiments, a scaffold polynucleotide comprises one or more modified nucleotides. In
- 15 some embodiments, the one or more modified nucleotides are capable of blocking covalent linkage of the scaffold polynucleotide to another oligonucleotide, polynucleotide, or nucleic acid molecule. A scaffold polynucleotide may comprise the one or more modified nucleotides at one or both ends of the polynucleotide. In some embodiments, the one or more modified nucleotides comprise a ligation-blocking modification.
- 20 In some embodiments, a nucleic acid molecule comprises one or more blocked nucleotides. For example, a DNA molecule may comprise one or more blocked nucleotides at one or more unligated ends. In some embodiments, a scaffold adapter, or component thereof, comprises one or more blocked nucleotides. In one example, a scaffold adapter, or component thereof, may comprise one or more modified nucleotides that are capable of blocking hybridization to a nucleotide in another
- 25 scaffold adapter, or component thereof. In some instances, the one or more modified nucleotides are capable of blocking ligation to a nucleotide in another scaffold adapter, or component thereof. In another example, a scaffold adapter, or component thereof, may comprise one or more modified nucleotides that are capable of blocking hybridization to a nucleotide in a target nucleic acid (e.g., ssNA). In some instances, the one or more modified nucleotides are capable of blocking ligation to
- 30 a nucleotide in a target nucleic acid. In some embodiments, one or both ends of a scaffold polynucleotide include a blocking modification and/or the end of an oligonucleotide component not adjacent to an ssNA fragment may include a blocking modification.

A blocking modification refers to a modified end that cannot be linked to the end of another nucleic acid component using an approach employed to covalently link the adjacent ends of an

oligonucleotide component and an ssNA fragment. In certain embodiments, the blocking modification is a ligation-blocking modification. Examples of blocking modifications which may be included at one or both ends of a nucleic acid molecule, a scaffold polynucleotide and/or the end of an oligonucleotide component not adjacent to the ssNA, include the absence of a 3' OH, and an inaccessible 3' OH. Non-limiting examples of blocking modifications in which an end has an inaccessible 3' OH include: an amino modifier, an amino linker, a spacer, an isodeoxy-base, a dideoxy base, an inverted dideoxy base, a 3' phosphate, and the like. In some embodiments, a nucleic acid molecule, scaffold adapter, or component thereof, comprises one or more modified nucleotides that are incapable of binding to a natural nucleotide.

In some embodiments, one or more modified nucleotides comprise an isodeoxy-base. In some embodiments, one or more modified nucleotides comprise isodeoxy-guanine (iso-dG). In some embodiments, one or more modified nucleotides comprise isodeoxy-cytosine (iso-dC). Iso-dC and iso-dG are chemical variants of cytosine and guanine, respectively. Iso-dC can hydrogen bond with iso-dG but not with unmodified guanine (natural guanine). Iso-dG can base pair with Iso-dC but not with unmodified cytosine (natural cytosine). A scaffold adapter, or component thereof, containing iso-dC can be designed so that it hybridizes to a complementary oligo containing iso-dG but cannot hybridize to any naturally occurring nucleic acid sequence.

In some embodiments, one or more modified nucleotides comprise epigenetic-associated modifications, including but not limited to methylation, hydroxymethylation, and carboxylation.

Example epigenetic-associated modifications include carboxycytosine, 5-methylcytosine (5mC) and its oxidative derivatives (e.g., 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC)), N(6)-methyladenine (6mA), N4-methylcytosine (4mC), N(6)-methyladenosine (m(6)A), pseudouridine (Ψ), 5-methylcytidine (m(5)C), hydroxymethyl uracil, 2'-O-methylation at the 3' end, tRNA modifications, miRNA modifications, and snRNA modifications.

In some embodiments, one or more modified nucleotides comprise a dideoxy-base. In some embodiments, one or more modified nucleotides comprise dideoxy-adenine (ddATP). In some embodiments, one or more modified nucleotides comprise dideoxy-thymine (ddTTP). In some embodiments, one or more modified nucleotides comprise dideoxy-guanine (ddGTP). In some embodiments, one or more modified nucleotides comprise dideoxy-cytosine (ddCTP). In some embodiments, one or more modified nucleotides comprise an inverted dideoxy-base. In some embodiments, one or more modified nucleotides comprise inverted dideoxy-thymine. For example, an inverted dideoxy-thymine located at the 5' end of a sequence can prevent unwanted 5' ligations.

In some embodiments, one or more modified nucleotides comprise a spacer. In some embodiments, one or more modified nucleotides comprise a C3 spacer. A C3 spacer

phosphoramidite can be incorporated internally or at the 5'-end of an oligonucleotide. Multiple C3 spacers can be added at either end of a scaffold adapter, or component thereof, to introduce a long hydrophilic spacer arm (e.g., for the attachment of fluorophores or other pendent groups). Other spacers include, for example, photo-cleavable (PC) spacers, hexanediol, spacer 9, spacer 18, 1',2'-dideoxyribose (dSpacer), and the like.

In some embodiments, a modified nucleotide comprises an amino linker or amino blocker. In some embodiments, a modified nucleotide comprises an amino linker C6 (e.g., a 5' amino linker C6 or a 3' amino linker C6). In one example, an amino linker C6 can be used to incorporate an active primary amino group onto the 5'-end of an oligonucleotide. This can then be conjugated to a ligand.

The amino group then becomes internal to the 5' end ligand. The amino group is separated from the 5'-end nucleotide base by a 6-carbon spacer arm to reduce steric interaction between the amino group and the oligo. In some embodiments, a modified nucleotide comprises an amino linker C12 (e.g., a 5' amino linker C12 or a 3' amino linker C12). In one example, an amino linker C12 can be used to incorporate an active primary amino group onto the 5'-end of an oligonucleotide. The amino group is separated from the 5'-end nucleotide base by a 12-carbon spacer arm to minimize steric interaction between the amino group and the oligo.

In some embodiments, a modified nucleotide comprises a member of a binding pair. Binding pairs may include, for example, antibody/antigen, antibody/antibody, antibody/antibody fragment, antibody/antibody receptor, antibody/protein A or protein G, hapten/anti-hapten, biotin/avidin, biotin/streptavidin, folic acid/folate binding protein, vitamin B12/intrinsic factor, chemical reactive group/complementary chemical reactive group, digoxigenin moiety/anti-digoxigenin antibody, fluorescein moiety/anti-fluorescein antibody, steroid/steroid-binding protein, operator/ repressor, nuclease/nucleotide, lectin/polysaccharide, active compound/active compound receptor, hormone/hormone receptor, enzyme/substrate, oligonucleotide or polynucleotide/its corresponding complement, the like or combinations thereof. In some embodiments, a modified nucleotide comprises biotin.

In some embodiments, a modified nucleotide comprises a first member of a binding pair (e.g., biotin); and a second member of a binding pair (e.g., streptavidin) is conjugated to a solid support or substrate. A solid support or substrate can be any physically separable solid to which a member of a binding pair can be directly or indirectly attached including, but not limited to, surfaces provided by microarrays and wells, and particles such as beads (e.g., paramagnetic beads, magnetic beads, microbeads, nanobeads), microparticles, and nanoparticles. Solid supports also can include, for example, chips, columns, optical fibers, wipes, filters (e.g., flat surface filters), one or more capillaries, glass and modified or functionalized glass (e.g., controlled-pore glass (CPG)), quartz,

mica, diazotized membranes (paper or nylon), polyformaldehyde, cellulose, cellulose acetate, paper, ceramics, metals, metalloids, semiconductive materials, quantum dots, coated beads or particles, other chromatographic materials, magnetic particles; plastics (including acrylics, polystyrene, copolymers of styrene or other materials, polybutylene, polyurethanes, TEFLON™, polyethylene, polypropylene, polyamide, polyester, polyvinylidenedifluoride (PVDF), and the like), polysaccharides, nylon or nitrocellulose, resins, silica or silica-based materials including silicon, silica gel, and modified silicon, Sephadex®, Sepharose®, carbon, metals (e.g., steel, gold, silver, aluminum, silicon and copper), inorganic glasses, conducting polymers (including polymers such as polypyrrole and polyindole); micro or nanostructured surfaces such as nucleic acid tiling arrays, nanotube, nanowire, or nanoparticulate decorated surfaces; or porous surfaces or gels such as methacrylates, acrylamides, sugar polymers, cellulose, silicates, or other fibrous or stranded polymers. In some embodiments, a solid support or substrate may be coated using passive or chemically-derivatized coatings with any number of materials, including polymers, such as dextrans, acrylamides, gelatins or agarose. Beads and/or particles may be free or in connection with one another (e.g., sintered). In some embodiments, a solid support can be a collection of particles. In some embodiments, the particles can comprise silica, and the silica may comprise silica dioxide. In some embodiments, the silica can be porous, and in certain embodiments the silica can be non-porous. In some embodiments, the particles further comprise an agent that confers a paramagnetic property to the particles. In certain embodiments, the agent comprises a metal, and in certain embodiments the agent is a metal oxide, (e.g., iron or iron oxides, where the iron oxide contains a mixture of Fe²⁺ and Fe³⁺). A member of a binding pair may be linked to a solid support by covalent bonds or by non-covalent interactions and may be linked to a solid support directly or indirectly (e.g., via an intermediary agent such as a spacer molecule or biotin).

In some embodiments, a scaffold polynucleotide, an oligonucleotide component, or both, include one or more non-natural nucleotides, also referred to as nucleotide analogs. Non-limiting examples of non-natural nucleotides that may be included in a scaffold polynucleotide, an oligonucleotide component, or both include LNA (locked nucleic acid), PNA (peptide nucleic acid), FANA (2'-deoxy-2'-fluoroarabinonucleotide), GNA (glycol nucleic acid), TNA (threose nucleic acid), 2'-O-Me RNA, 2'-fluoro RNA, Morpholino nucleotides, and any combination thereof.

End Treatments

In some embodiments, a method herein comprises contacting a nucleic acid composition comprising single-stranded nucleic acid (ssNA) with an agent comprising an end treatment activity under conditions in which single-stranded nucleic acid (ssNA) molecules are end treated, thereby generating an end treated ssNA composition. End treatments can include but are not limited to

phosphorylation, dephosphorylation, methylation, demethylation, oxidation, de-oxidation, base modification, extension, polymerization, and combinations thereof. End treatments can be conducted with enzymes, including but not limited to ligases, polynucleotide kinases (PNK), phosphatases (e.g., recombinant shrimp alkaline phosphatase (rSAP)), kinases, terminal
5 transferases, methyltransferases, methylases (e.g., 3' methylases, 5' methylases), polymerases (e.g., poly A polymerases), oxidases, and combinations thereof.

In some embodiments, a method herein comprises contacting a nucleic acid composition comprising single-stranded nucleic acid (ssNA) with an agent comprising a phosphatase activity under conditions in which single-stranded nucleic acid (ssNA) molecules are dephosphorylated,
10 thereby generating a dephosphorylated ssNA composition. In some embodiments, single-stranded nucleic acid (ssNA) is contacted with an agent comprising a phosphatase activity prior to combining the ssNA with one or more scaffold adapters (or components thereof). In some embodiments, single-stranded nucleic acid (ssNA) is contacted with an agent comprising a phosphatase activity and simultaneously combined with single-stranded nucleic acid binding protein (SSB), as described
15 in further detail herein, to produce SSB-bound dephosphorylated ssNA. In some embodiments, ssNA is contacted with an agent comprising a phosphatase activity during a denaturation step (e.g., denaturing of dsDNA to ssDNA). In some embodiments, ssNA is contacted with an agent comprising a phosphatase activity and combined with single-stranded nucleic acid binding protein (SSB) during a denaturation step (e.g., denaturing of dsDNA to ssDNA). An agent comprising a
20 phosphatase activity may include any suitable phosphatase. In some embodiments, an agent comprising a phosphatase activity is a shrimp alkaline phosphatase (SAP). In some embodiments, an agent comprising a phosphatase activity is a recombinant shrimp alkaline phosphatase (rSAP).

In some embodiments, a method herein comprises contacting a scaffold adapter, or component thereof, with an agent comprising a phosphatase activity under conditions in which the scaffold
25 adapter, or component thereof, is dephosphorylated, thereby generating a dephosphorylated scaffold adapter, or component thereof (e.g., a dephosphorylated oligonucleotide; a dephosphorylated scaffold polynucleotide). Generally, an ssNA composition and/or scaffold adapters, or components thereof, are dephosphorylated prior to a combining step (i.e., prior to hybridization). ssNAs may be dephosphorylated and then subsequently phosphorylated prior to a
30 combining step (i.e., prior to hybridization). Scaffold adapters, or components thereof, may be dephosphorylated and then subsequently phosphorylated prior to a combining step (i.e., prior to hybridization). Scaffold adapters, or components thereof, may be dephosphorylated and then not phosphorylated prior to a combining step (i.e., prior to hybridization). Scaffold adapters, or components thereof, may be dephosphorylated, not phosphorylated prior to a combining step (i.e.,
35 prior to hybridization), and then phosphorylated after a combining step (i.e., after hybridization) and

prior to or during a ligation step. Reagents and kits for carrying out dephosphorylation of nucleic acids are known and available. For example, target nucleic acids (e.g., ssNAs) and/or scaffold adapters, or components thereof, can be treated with a phosphatase (i.e., an enzyme that uses water to cleave a phosphoric acid monoester into a phosphate ion and an alcohol).

5 In some embodiments, a method herein comprises contacting a nucleic acid composition comprising single-stranded nucleic acid (ssNA) with an agent comprising a phosphoryl transfer activity under conditions in which a 5' phosphate is added to a 5' end of ssNAs. In some embodiments, a method herein comprises contacting a dephosphorylated ssNA composition with an agent comprising a phosphoryl transfer activity under conditions in which a 5' phosphate is added to a 5' end of an ssNA. In some embodiments, a method herein comprises contacting a scaffold adapter, or component thereof, with an agent comprising a phosphoryl transfer activity under conditions in which a 5' phosphate is added to a 5' end of a scaffold adapter, or component thereof. In some embodiments, a method herein comprises contacting a dephosphorylated scaffold adapter, or component thereof, with an agent comprising a phosphoryl transfer activity under conditions in which a 5' phosphate is added to a 5' end of a scaffold adapter, or component thereof. In certain instances, an ssNA composition and/or scaffold adapters, or components thereof, are phosphorylated prior to a combining step (i.e., prior to hybridization). 5' phosphorylation of nucleic acids can be conducted by a variety of techniques. For example, an ssNA composition and/or scaffold adapters, or components thereof, can be treated with a polynucleotide kinase (PNK) (e.g., T4 PNK), which catalyzes the transfer and exchange of Pi from the γ position of ATP to the 5'-hydroxyl terminus of polynucleotides (double- and single-stranded DNA and RNA) and nucleoside 3'-monophosphates. Suitable reaction conditions include, e.g., incubation of the nucleic acids with PNK in 1X PNK reaction buffer (e.g., 70 mM Tris-HCl, 10 mM MgCl₂, 5 mM DTT, pH 7.6 @ 25°C) for 30 minutes at 37°C; and incubation of the nucleic acids with PNK in T4 DNA ligase buffer (e.g., 50 mM Tris-HCl, 10 mM MgCl₂, 1 mM ATP, 10 mM DTT, pH 7.5 @ 25°C) for 30 minutes at 37°C. Optionally, following the phosphorylation reaction, the PNK may be heat inactivated, e.g., at 65°C for 20 minutes.

In some embodiments, a method herein does not include use of an agent comprising a phosphoryl transfer activity. In some embodiments, methods do not include producing the 5' phosphorylated ssNAs by phosphorylating the 5' ends of ssNAs from a nucleic acid sample. In certain instances, a nucleic acid sample comprises ssNAs with natively phosphorylated 5' ends. In some embodiments, methods do not include producing the 5' phosphorylated scaffold adapters, or components thereof, by phosphorylating the 5' ends of scaffold adapters, or components thereof.

Cleavage

In some embodiments, ssNAs, scaffold adapters, and/or hybridization products (e.g., scaffold adapters hybridized to ssNAs) are cleaved or sheared prior to, during, or after a method described herein. In some embodiments, ssNAs, scaffold adapters, and/or hybridization products are cleaved or sheared at a cleavage site. In some embodiments, scaffold adapters and/or hybridization products are cleaved or sheared at a cleavage site within a hairpin loop. In some embodiments, scaffold adapters and/or hybridization products are cleaved or sheared at a cleavage site at an internal location in a scaffold adapter (e.g., within a duplex region of a scaffold adapter). In some embodiments, scaffold adapters are cleaved at a cleavage site (e.g., a uracil) at an internal location present only on the scaffold polynucleotide but not the complementary oligonucleotide component. Thus, in some embodiments, a scaffold polynucleotide comprises one or more uracil bases, and an oligonucleotide component comprises no uracil bases. In some embodiments, circular hybridization products are cleaved or sheared prior to, during, or after a method described herein. In some embodiments, nucleic acids, such as, for example, cellular nucleic acids and/or large fragments (e.g., greater than 500 base pairs in length) are cleaved or sheared prior to, during, or after a method described herein. Large fragments may be referred to as high molecular weight (HMW) nucleic acid, HMW DNA or HMW RNA. HMW nucleic acid fragments may include fragments greater than about 500 bp, about 600 bp, about 700 bp, about 800 bp, about 900 bp, about 1000 bp, about 2000 bp, about 3000 bp, about 4000 bp, about 5000 bp, about 10,000 bp, or more. The term “shearing” or “cleavage” generally refers to a procedure or conditions in which a nucleic acid molecule may be severed into two (or more) smaller nucleic acid molecules. Such shearing or cleavage can be sequence specific, base specific, or nonspecific, and can be accomplished by any of a variety of methods, reagents or conditions, including, for example, chemical, enzymatic, and physical (e.g., physical fragmentation). Sheared or cleaved nucleic acids may have a nominal, average or mean length of about 5 to about 10,000 base pairs, about 100 to about 1,000 base pairs, about 100 to about 500 base pairs, or about 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000 or 9000 base pairs.

Sheared or cleaved nucleic acids can be generated by a suitable method, non-limiting examples of which include physical methods (e.g., shearing, e.g., sonication, ultrasonication, French press, heat, UV irradiation, the like), enzymatic processes (e.g., enzymatic cleavage agents (e.g., a suitable nuclease, a suitable restriction enzyme), chemical methods (e.g., alkylation, DMS, piperidine, acid hydrolysis, base hydrolysis, heat, the like, or combinations thereof), ultraviolet (UV) light (e.g., at a photo-cleavable site (e.g., comprising a photo-cleavable spacer), the like or

combinations thereof. The average, mean or nominal length of the resulting nucleic acid fragments can be controlled by selecting an appropriate fragment-generating method.

The term “cleavage agent” generally refers to an agent, sometimes a chemical or an enzyme that can cleave a nucleic acid at one or more specific or non-specific sites. Specific cleavage agents often cleave specifically according to a particular nucleotide sequence at a particular site, which may be referred to as a cleavage site. Cleavage agents may include enzymatic cleavage agents, chemical cleavage agents, and light (e.g., ultraviolet (UV) light).

Examples of enzymatic cleavage agents include without limitation endonucleases;

deoxyribonucleases (DNase; e.g., DNase I, II); ribonucleases (RNase; e.g., RNase A, RNase E, RNase F, RNase H, RNase III, RNase L, RNase P, RNase PhyM, RNase T1, RNase T2, RNase U2, and RNase V); endonuclease VIII; CLEAVASE enzyme; TAQ DNA polymerase; E. coli DNA polymerase I; eukaryotic structure-specific endonucleases; murine FEN-1 endonucleases; nicking enzymes; type I, II or III restriction endonucleases (i.e., restriction enzymes) such as Acc I, Acil, Afl III, Alu I, Alw44 I, Apa I, Asn I, Ava I, Ava II, BamH I, Ban II, Bcl I, Bgl I, Bgl II, Bln I, Bsm I, BssH II, BstE II, BstUI, Cfo I, Cla I, Dde I, Dpn I, Dra I, EclX I, EcoR I, EcoR I, EcoR II, EcoR V, Hae II, Hae II, HhaI, Hind II, Hind III, Hpa I, Hpa II, Kpn I, Ksp I, MaeII, McrBC, Mlu I, MluN I, Msp I, Nci I, Nco I, Nde I, Nde II, Nhe I, Not I, Nru I, Nsi I, Pst I, Pvu I, Pvu II, Rsa I, Sac I, Sal I, Sau3A I, Sca I, ScrF I, Sfi I, Sma I, Spe I, Sph I, Ssp I, Stu I, Sty I, Swa I, Taq I, Xba I, Xho I; glycosylases (e.g., uracil-DNA glycosylase (UDG), 3-methyladenine DNA glycosylase, 3-methyladenine DNA glycosylase II, pyrimidine hydrate-DNA glycosylase, FaPy-DNA glycosylase, thymine mismatch-DNA glycosylase (e.g., hypoxanthine-DNA glycosylase, uracil DNA glycosylase (UDG), 5-Hydroxymethyluracil DNA glycosylase (HmUDG), 5-Hydroxymethylcytosine DNA glycosylase, or 1,N6-etheno-adenine DNA glycosylase); exonucleases (e.g., exonuclease I, exonuclease II, exonuclease III, exonuclease IV, exonuclease V, exonuclease VI, exonuclease VII, exonuclease VIII); 5' to 3' exonucleases (e.g. exonuclease II); 3' to 5' exonucleases (e.g. exonuclease I); poly(A)-specific 3' to 5' exonucleases; ribozymes; DNazymes; and the like and combinations thereof.

In some embodiments, a cleavage site comprises a restriction enzyme recognition site. In some embodiments, a cleavage agent comprises a restriction enzyme. In some embodiments, a cleavage site comprises a rare-cutter restriction enzyme recognition site (e.g., a NotI recognition sequence). In some embodiments, a cleavage agent comprises a rare-cutter enzyme (e.g., a rare-cutter restriction enzyme). A rare-cutter enzyme generally refers to a restriction enzyme with a recognition sequence which occurs only rarely in a genome (e.g., a human genome). An example is NotI, which cuts after the first GC of a 5'-GCGGCCGC-3' sequence. Restriction enzymes with seven and eight base pair recognition sequences often are considered as rare-cutter enzymes.

Cleavage methods and procedures for selecting restriction enzymes for cutting DNA at specific sites are well known to the skilled artisan. For example, many suppliers of restriction enzymes provide information on conditions and types of DNA sequences cut by specific restriction enzymes, including New England BioLabs, Pro-Mega Biochems, Boehringer-Mannheim, and the like.

- 5 Enzymes often are used under conditions that will enable cleavage of the DNA with about 95%-100% efficiency, preferably with about 98%-100% efficiency.

In some embodiments, a cleavage site comprises one or more ribonucleic acid (RNA) nucleotides. In some embodiments, a cleavage site comprises a single stranded portion comprising one or more RNA nucleotides. In some embodiments, the single stranded portion is flanked by duplex portions.

- 10 In some embodiments, the single stranded portion is a hairpin loop. In some embodiments, a cleavage site comprises one RNA nucleotide. In some embodiments, a cleavage site comprises two RNA nucleotides. In some embodiments, a cleavage site comprises three RNA nucleotides. In some embodiments, a cleavage site comprises four RNA nucleotides. In some embodiments, a cleavage site comprises five RNA nucleotides. In some embodiments, a cleavage site comprises more than five RNA nucleotides. In some embodiments, a cleavage site comprises one or more RNA nucleotides chosen from adenine (A), cytosine (C), guanine (G), and uracil (U). In some
15 embodiments, a cleavage site comprises one or more RNA nucleotides chosen from adenine (A), cytosine (C), and guanine (G). In some embodiments, a cleavage site comprises no uracil (U). In some embodiments, a cleavage site comprises one or more RNA nucleotides comprising guanine (G). In some embodiments, a cleavage site comprises one or more RNA nucleotides consisting of guanine (G). In some embodiments, a cleavage site comprises one or more RNA nucleotides comprising cytosine (C). In some embodiments, a cleavage site comprises one or more RNA nucleotides consisting of cytosine (C). In some embodiments, a cleavage site comprises one or more RNA nucleotides comprising adenine (A). In some embodiments, a cleavage site comprises one or more RNA nucleotides consisting of adenine (A). In some embodiments, a cleavage site
20 comprises one or more RNA nucleotides consisting of adenine (A), cytosine (C), and guanine (G). In some embodiments, a cleavage site comprises one or more RNA nucleotides consisting of adenine (A) and cytosine (C). In some embodiments, a cleavage site comprises one or more RNA nucleotides consisting of adenine (A) and guanine (G). In some embodiments, a cleavage site comprises one or more RNA nucleotides consisting of cytosine (C) and guanine (G). In some
25 embodiments, a cleavage agent comprises a ribonuclease (RNase). In some embodiments, an RNase is an endoribonuclease. An RNase may be chosen from one or more of RNase A, RNase E, RNase F, RNase H, RNase III, RNase L, RNase P, RNase PhyM, RNase T1, RNase T2, RNase U2, and RNase V.

In some embodiments, a cleavage site comprises a photo-cleavable spacer or photo-cleavable modification. Photo-cleavable modifications may contain, for example, a photolabile functional group that is cleavable by ultraviolet (UV) light of specific wavelength (e.g., 300–350 nm). An example photo-cleavable spacer (available from Integrated DNA Technologies; product no. 1707) is a 10-atom linker arm that can only be cleaved when exposed to UV light within the appropriate spectral range. An oligonucleotide comprising a photo-cleavable spacer can have a 5' phosphate group that is available for subsequent ligase reactions. Photo-cleavable spacers can be placed between DNA bases or between an oligo and a terminal modification (e.g., a fluorophore). In such embodiments, ultraviolet (UV) light may be considered as a cleavage agent.

In some embodiments, a cleavage site comprises a diol. For example, a cleavage site may comprise vicinal diol incorporated in a 5' to 5' linkage. Cleavage sites comprising a diol may be chemically cleaved, for example, using a periodate. In some embodiments, a cleavage site comprises a blunt end restriction enzyme recognition site. Cleavage sites comprising a blunt end restriction enzyme recognition site may be cleaved by a blunt end restriction enzyme.

Nick seal and fill-in

In some embodiments, a method herein comprises performing a nick seal reaction (e.g., using a DNA ligase or other suitable enzyme, and, in certain instances, a kinase adapted to 5' phosphorylate nucleic acids (e.g., a polynucleotide kinase (PNK))). In some embodiments, a method herein comprises performing a fill-in reaction. For example, when scaffold adapters are present as duplexes, some or all of the duplexes may include an overhang at the end of the duplex opposite the end that hybridizes to the ssNAs. When such duplex overhangs exist, subsequent to the combining, a method herein may further include filling in the overhangs formed by the duplexes. In some embodiments, a fill-in reaction is performed to generate a blunt-ended hybridization product. Any suitable reagent for carrying out a fill-in reaction may be used. Polymerases suitable for performing fill-in reactions include, e.g., DNA polymerase I, large (Klenow) fragment of DNA polymerase I, T4 DNA polymerase, *Bacillus stearothermophilus* (Bst) DNA polymerase, thermostable DNA polymerases (e.g., from hyperthermophilic marine Archaea), 9°N™ DNA Polymerase (GENBANK accession no. AAA88769.1), THERMINATOR polymerase (9°N™ DNA Polymerase with mutations: D141A, E143A, A485L), and the like. In some embodiments, a strand displacing polymerase is used (e.g., Bst DNA polymerase).

Exonuclease treatment

In some embodiments, nucleic acid (e.g., RNA-DNA duplexes, hybridization products; circularized hybridization products) is treated with an exonuclease. In some embodiments, RNA in an RNA-DNA duplex (e.g., an RNA-DNA duplex generated by first strand cDNA synthesis) is treated with an

exonuclease. Exonucleases are enzymes that work by cleaving nucleotides one at a time from the end of a polynucleotide chain through a hydrolyzing reaction that breaks phosphodiester bonds at either the 3' or the 5' end. Exonucleases include, for example, DNAses, RNAses (e.g., RNaseH), 5' to 3' exonucleases (e.g. exonuclease II), 3' to 5' exonucleases (e.g. exonuclease I), and poly(A)-specific 3' to 5' exonucleases. In some embodiments, exonuclease activity is provided by a reverse transcriptase (e.g., RNase activity provided by M-MLV reverse transcriptase having a fully functional RNaseH domain). In some embodiments, hybridization products are treated with an exonuclease to remove contaminating nucleic acids such as, for example, single stranded oligonucleotides, nucleic acid fragments, or RNA from an RNA-DNA duplex. In some embodiments, circularized hybridization products are treated with an exonuclease to remove any non-circularized hybridization products, non-hybridized oligonucleotides, non-hybridized target nucleic acids, oligonucleotide dimers, and the like and combinations thereof.

Samples

Provided herein are methods and compositions for processing and/or analyzing nucleic acid.

Nucleic acid or a nucleic acid mixture utilized in methods and compositions described herein may be isolated from a sample obtained from a subject (e.g., a test subject). A subject can be any living or non-living organism, including but not limited to a human, a non-human animal, a plant, a bacterium, a fungus, a protist or a pathogen. Any human or non-human animal can be selected, and may include, for example, mammal, reptile, avian, amphibian, fish, ungulate, ruminant, bovine (e.g., cattle), equine (e.g., horse), caprine and ovine (e.g., sheep, goat), swine (e.g., pig), camelid (e.g., camel, llama, alpaca), monkey, ape (e.g., gorilla, chimpanzee), ursid (e.g., bear), poultry, dog, cat, mouse, rat, fish, dolphin, whale and shark. A subject may be a male or female (e.g., woman, a pregnant woman). A subject may be any age (e.g., an embryo, a fetus, an infant, a child, an adult). A subject may be a cancer patient, a patient suspected of having cancer, a patient in remission, a patient with a family history of cancer, and/or a subject obtaining a cancer screen. A subject may be a patient having an infection or infectious disease or infected with a pathogen (e.g., bacteria, virus, fungus, protozoa, and the like), a patient suspected of having an infection or infectious disease or being infected with a pathogen, a patient recovering from an infection, infectious disease, or pathogenic infection, a patient with a history of infections, infectious disease, pathogenic infections, and/or a subject obtaining an infectious disease or pathogen screen. A subject may be a transplant recipient. A subject may be a patient undergoing a microbiome analysis. In some embodiments, a test subject is a female. In some embodiments, a test subject is a human female. In some embodiments, a test subject is a male. In some embodiments, a test subject is a human male.

A nucleic acid sample may be isolated or obtained from any type of suitable biological specimen or sample (e.g., a test sample). A nucleic acid sample may be isolated or obtained from a single cell, a plurality of cells (e.g., cultured cells), cell culture media, conditioned media, a tissue, an organ, or an organism (e.g., bacteria, yeast, or the like). In some embodiments, a nucleic acid sample is isolated or obtained from a cell(s), tissue, organ, and/or the like of an animal (e.g., an animal subject). In some embodiments, a nucleic acid sample is isolated or obtained from a source such as bacteria, yeast, insects (e.g., drosophila), mammals, amphibians (e.g., frogs (e.g., *Xenopus*)), viruses, plants, or any other mammalian or non-mammalian nucleic acid sample source.

A nucleic acid sample may be isolated or obtained from an extant organism or animal. In some instances, a nucleic acid sample may be isolated or obtained from an extinct (or "ancient") organism or animal (e.g., an extinct mammal; an extinct mammal from the genus *Homo*). In some instances, a nucleic acid sample may be obtained as part of a diagnostic analysis.

In some instances, a nucleic acid sample may be obtained as part of a forensics analysis. In some embodiments, a single-stranded nucleic acid library preparation (ssPrep) method described herein is applied to a forensic sample or specimen. A forensic sample or specimen may include any biological substance that contains nucleic acid. For example, a forensic sample or specimen may include blood, semen, hair, skin, sweat, saliva, decomposed tissue, bone, fingernail scrapings, licked stamps/envelopes, sluff, touch DNA, razor residue, and the like.

A sample or test sample may be any specimen that is isolated or obtained from a subject or part thereof (e.g., a human subject, a pregnant female, a cancer patient, a patient having an infection or infectious disease, a transplant recipient, a fetus, a tumor, an infected organ or tissue, a transplanted organ or tissue, a microbiome). A sample sometimes is from a pregnant female subject bearing a fetus at any stage of gestation (e.g., first, second or third trimester for a human subject), and sometimes is from a post-natal subject. A sample sometimes is from a pregnant subject bearing a fetus that is euploid for all chromosomes, and sometimes is from a pregnant subject bearing a fetus having a chromosome aneuploidy (e.g., one, three (i.e., trisomy (e.g., T21, T18, T13)), or four copies of a chromosome) or other genetic variation. Non-limiting examples of specimens include fluid or tissue from a subject, including, without limitation, blood or a blood product (e.g., serum, plasma, or the like), umbilical cord blood, chorionic villi, amniotic fluid, cerebrospinal fluid, spinal fluid, lavage fluid (e.g., bronchoalveolar, gastric, peritoneal, ductal, ear, arthroscopic), biopsy sample (e.g., from pre-implantation embryo; cancer biopsy), celocentesis sample, cells (blood cells, placental cells, embryo or fetal cells, fetal nucleated cells or fetal cellular remnants, normal cells, abnormal cells (e.g., cancer cells)) or parts thereof (e.g., mitochondrial, nucleus, extracts, or the like), washings of female reproductive tract, urine, feces, sputum, saliva,

nasal mucous, prostate fluid, lavage, semen, lymphatic fluid, bile, tears, sweat, breast milk, breast fluid, the like or combinations thereof. In some embodiments, a biological sample is a cervical swab from a subject. A fluid or tissue sample from which nucleic acid is extracted may be acellular (e.g., cell-free). In some embodiments, a fluid or tissue sample may contain cellular elements or cellular remnants. In some embodiments, fetal cells or cancer cells may be included in the sample.

A sample can be a liquid sample. A liquid sample can comprise extracellular nucleic acid (e.g., circulating cell-free DNA). Examples of liquid samples include, but are not limited to, blood or a blood product (e.g., serum, plasma, or the like), urine, cerebral spinal fluid, saliva, sputum, biopsy sample (e.g., liquid biopsy for the detection of cancer), a liquid sample described above, the like or combinations thereof. In certain embodiments, a sample is a liquid biopsy, which generally refers to an assessment of a liquid sample from a subject for the presence, absence, progression or remission of a disease (e.g., cancer). A liquid biopsy can be used in conjunction with, or as an alternative to, a solid biopsy (e.g., tumor biopsy). In certain instances, extracellular nucleic acid is analyzed in a liquid biopsy.

In some embodiments, a biological sample may be blood, plasma or serum. The term "blood" encompasses whole blood, blood product or any fraction of blood, such as serum, plasma, buffy coat, or the like as conventionally defined. Blood or fractions thereof often comprise nucleosomes. Nucleosomes comprise nucleic acids and are sometimes cell-free or intracellular. Blood also comprises buffy coats. Buffy coats are sometimes isolated by utilizing a ficoll gradient. Buffy coats can comprise white blood cells (e.g., leukocytes, T-cells, B-cells, platelets, and the like). Blood plasma refers to the fraction of whole blood resulting from centrifugation of blood treated with anticoagulants. Blood serum refers to the watery portion of fluid remaining after a blood sample has coagulated. Fluid or tissue samples often are collected in accordance with standard protocols hospitals or clinics generally follow. For blood, an appropriate amount of peripheral blood (e.g., between 3 to 40 milliliters, between 5 to 50 milliliters) often is collected and can be stored according to standard procedures prior to or after preparation.

An analysis of nucleic acid found in a subject's blood may be performed using, e.g., whole blood, serum, or plasma. An analysis of fetal DNA found in maternal blood, for example, may be performed using, e.g., whole blood, serum, or plasma. An analysis of tumor or cancer DNA found in a patient's blood, for example, may be performed using, e.g., whole blood, serum, or plasma. An analysis of pathogen DNA found in a patient's blood, for example, may be performed using, e.g., whole blood, serum, or plasma. An analysis of transplant DNA found in a transplant recipient's blood, for example, may be performed using, e.g., whole blood, serum, or plasma. Methods for preparing serum or plasma from blood obtained from a subject (e.g., a maternal subject; patient;

cancer patient) are known. For example, a subject's blood (e.g., a pregnant woman's blood; patient's blood; cancer patient's blood) can be placed in a tube containing EDTA or a specialized commercial product such as Cell-Free DNA BCT (Streck, Omaha, NE) or Vacutainer SST (Becton Dickinson, Franklin Lakes, N.J.) to prevent blood clotting, and plasma can then be obtained from whole blood through centrifugation. Serum may be obtained with or without centrifugation-following blood clotting. If centrifugation is used then it is typically, though not exclusively, conducted at an appropriate speed, e.g., 1,500-3,000 times g. Plasma or serum may be subjected to additional centrifugation steps before being transferred to a fresh tube for nucleic acid extraction. In addition to the acellular portion of the whole blood, nucleic acid may also be recovered from the cellular fraction, enriched in the buffy coat portion, which can be obtained following centrifugation of a whole blood sample from the subject and removal of the plasma.

A sample may be a tumor nucleic acid sample (i.e., a nucleic acid sample isolated from a tumor). The term "tumor" generally refers to neoplastic cell growth and proliferation, whether malignant or benign, and may include pre-cancerous and cancerous cells and tissues. The terms "cancer" and "cancerous" generally refer to the physiological condition in mammals that is typically characterized by unregulated cell growth/proliferation. Examples of cancer include, but are not limited to, carcinoma, lymphoma, blastoma, sarcoma, leukemia, squamous cell cancer, small-cell lung cancer, non-small cell lung cancer, adenocarcinoma of the lung, squamous carcinoma of the lung, cancer of the peritoneum, hepatocellular cancer, gastrointestinal cancer, pancreatic cancer, glioblastoma, cervical cancer, ovarian cancer, liver cancer, bladder cancer, hepatoma, breast cancer, colon cancer, colorectal cancer, endometrial or uterine carcinoma, salivary gland carcinoma, kidney cancer, liver cancer, prostate cancer, vulval cancer, thyroid cancer, hepatic carcinoma, various types of head and neck cancer, and the like.

A sample may be heterogeneous. For example, a sample may include more than one cell type and/or one or more nucleic acid species. In some instances, a sample may include (i) fetal cells and maternal cells, (ii) cancer cells and non-cancer cells, and/or (iii) pathogenic cells and host cells. In some instances, a sample may include (i) cancer and non-cancer nucleic acid, (ii) pathogen and host nucleic acid, (iii) fetal derived and maternal derived nucleic acid, and/or more generally, (iv) mutated and wild-type nucleic acid. In some instances, a sample may include a minority nucleic acid species and a majority nucleic acid species, as described in further detail below. In some instances, a sample may include cells and/or nucleic acid from a single subject or may include cells and/or nucleic acid from multiple subjects.

Nucleic acid

Provided herein are methods and compositions for processing and/or analyzing nucleic acid. The terms nucleic acid(s), nucleic acid molecule(s), nucleic acid fragment(s), target nucleic acid(s), nucleic acid template(s), template nucleic acid(s), nucleic acid target(s), target nucleic acid(s), polynucleotide(s), polynucleotide fragment(s), target polynucleotide(s), polynucleotide target(s), and the like may be used interchangeably throughout the disclosure. The terms refer to nucleic acids of any composition from, such as DNA (e.g., complementary DNA (cDNA; synthesized from any RNA or DNA of interest), genomic DNA (gDNA), genomic DNA fragments, mitochondrial DNA (mtDNA), recombinant DNA (e.g., plasmid DNA), and the like), RNA (e.g., message RNA (mRNA), short inhibitory RNA (siRNA), ribosomal RNA (rRNA), transfer RNA (tRNA), microRNA, transacting small interfering RNA (ta-siRNA), natural small interfering RNA (nat-siRNA), small nucleolar RNA (snoRNA), small nuclear RNA (snRNA), long non-coding RNA (lncRNA), non-coding RNA (ncRNA), transfer-messenger RNA (tmRNA), precursor messenger RNA (pre-mRNA), small Cajal body-specific RNA (scaRNA), piwi-interacting RNA (piRNA), endoribonuclease-prepared siRNA (esiRNA), small temporal RNA (stRNA), signal recognition RNA, telomere RNA, RNA highly expressed by a fetus or placenta, and the like), and/or DNA or RNA analogs (e.g., containing base analogs, sugar analogs and/or a non-native backbone and the like), RNA/DNA hybrids and polyamide nucleic acids (PNAs), all of which can be in single- or double-stranded form, and unless otherwise limited, can encompass known analogs of natural nucleotides that can function in a similar manner as naturally occurring nucleotides. A nucleic acid may be, or may be from, a plasmid, phage, virus, bacterium, autonomously replicating sequence (ARS), mitochondria, centromere, artificial chromosome, chromosome, or other nucleic acid able to replicate or be replicated in vitro or in a host cell, a cell, a cell nucleus or cytoplasm of a cell in certain embodiments. A template nucleic acid in some embodiments can be from a single chromosome (e.g., a nucleic acid sample may be from one chromosome of a sample obtained from a diploid organism). Unless specifically limited, the term encompasses nucleic acids containing known analogs of natural nucleotides that have similar binding properties as the reference nucleic acid and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (e.g., degenerate codon substitutions), alleles, orthologs, single nucleotide polymorphisms (SNPs), and complementary sequences as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues. The term nucleic acid is used interchangeably with locus, gene, cDNA, and mRNA encoded by a gene. The term also may include, as equivalents, derivatives,

variants and analogs of RNA or DNA synthesized from nucleotide analogs, single-stranded ("sense" or "antisense," "plus" strand or "minus" strand, "forward" reading frame or "reverse" reading frame) and double-stranded polynucleotides. The term "gene" refers to a section of DNA involved in producing a polypeptide chain; and generally includes regions preceding and following the coding region (leader and trailer) involved in the transcription/translation of the gene product and the regulation of the transcription/translation, as well as intervening sequences (introns) between individual coding regions (exons). A nucleotide or base generally refers to the purine and pyrimidine molecular units of nucleic acid (e.g., adenine (A), thymine (T), guanine (G), and cytosine (C)). For RNA, the base thymine is replaced with uracil. Nucleic acid length or size may be expressed as a number of bases.

Target nucleic acids may be any nucleic acids of interest. Nucleic acids may be polymers of any length composed of deoxyribonucleotides (i.e., DNA bases), ribonucleotides (i.e., RNA bases), or combinations thereof, e.g., 10 bases or longer, 20 bases or longer, 50 bases or longer, 100 bases or longer, 200 bases or longer, 300 bases or longer, 400 bases or longer, 500 bases or longer, 1000 bases or longer, 2000 bases or longer, 3000 bases or longer, 4000 bases or longer, 5000 bases or longer. In certain aspects, nucleic acids are polymers composed of deoxyribonucleotides (i.e., DNA bases), ribonucleotides (i.e., RNA bases), or combinations thereof, e.g., 10 bases or less, 20 bases or less, 50 bases or less, 100 bases or less, 200 bases or less, 300 bases or less, 400 bases or less, 500 bases or less, 1000 bases or less, 2000 bases or less, 3000 bases or less, 4000 bases or less, or 5000 bases or less.

Nucleic acid may be single or double stranded. Single stranded DNA (ssDNA), for example, can be generated by denaturing double stranded DNA by heating or by treatment with alkali, for example. Accordingly, in some embodiments, ssDNA is derived from double-stranded DNA (dsDNA). In some embodiments, a method herein comprises prior to combining a nucleic acid composition comprising dsDNA with the scaffold adapters herein, or components thereof, denaturing the dsDNA, thereby generating ssDNA.

In certain embodiments, nucleic acid is in a D-loop structure, formed by strand invasion of a duplex DNA molecule by an oligonucleotide or a DNA-like molecule such as peptide nucleic acid (PNA). D loop formation can be facilitated by addition of E. Coli RecA protein and/or by alteration of salt concentration, for example, using methods known in the art.

Nucleic acid (e.g., nucleic acid targets, single-stranded nucleic acid (ssNA), oligonucleotides, overhangs, scaffold polynucleotides and hybridization regions thereof (e.g., ssNA hybridization region, oligonucleotide hybridization region)) may be described herein as being complementary to another nucleic acid, having a complementarity region, being capable of hybridizing to another

nucleic acid, or having a hybridization region. The terms “complementary” or “complementarity” or “hybridization” generally refer to a nucleotide sequence that base-pairs by non-covalent bonds to a region of a nucleic acid (e.g., the nucleotide sequence of an ssNA hybridization region that hybridizes to the terminal region of an ssNA fragment, and the nucleotide sequence of an oligonucleotide hybridization region that hybridizes to an oligonucleotide component of a scaffold adapter). In the canonical Watson-Crick base pairing, adenine (A) forms a base pair with thymine (T), and guanine (G) pairs with cytosine (C) in DNA. In RNA, thymine (T) is replaced by uracil (U). As such, A is complementary to T and G is complementary to C. In RNA, A is complementary to U and vice versa. In a DNA-RNA duplex, A (in a DNA strand) is complementary to U (in an RNA strand). In some embodiments, one or more thymine (T) bases are replaced by uracil (U) in a scaffold adapter, or a component thereof, and is/are complementary to adenine (A). Typically, “complementary” or “complementarity” or “capable of hybridizing” refer to a nucleotide sequence that is at least partially complementary. These terms may also encompass duplexes that are fully complementary such that every nucleotide in one strand is complementary or hybridizes to every nucleotide in the other strand in corresponding positions.

In certain instances, a nucleotide sequence may be partially complementary to a target, in which not all nucleotides are complementary to every nucleotide in the target nucleic acid in all the corresponding positions. For example, an ssNA hybridization region may be perfectly (i.e., 100%) complementary to a target ssNA terminal region, or an ssNA hybridization region may share some degree of complementarity which is less than perfect (e.g., 70%, 75%, 85%, 90%, 95%, 99%). In another example, an oligonucleotide hybridization region may be perfectly (i.e., 100%) complementary to an oligonucleotide, or an oligonucleotide hybridization region may share some degree of complementarity which is less than perfect (e.g., 70%, 75%, 85%, 90%, 95%, 99%).

The percent identity of two nucleotide sequences can be determined by aligning the sequences for optimal comparison purposes (e.g., gaps can be introduced in the sequence of a first sequence for optimal alignment). The nucleotides at corresponding positions are then compared, and the percent identity between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % identity = # of identical positions / total # of positions × 100). When a position in one sequence is occupied by the same nucleotide as the corresponding position in the other sequence, then the molecules are identical at that position.

In some embodiments, nucleic acids in a mixture of nucleic acids are analyzed. A mixture of nucleic acids can comprise two or more nucleic acid species having the same or different nucleotide sequences, different lengths, different origins (e.g., genomic origins, fetal vs. maternal origins, cell or tissue origins, cancer vs. non-cancer origin, tumor vs. non-tumor origin, host vs. pathogen, host

vs. transplant, host vs. microbiome, sample origins, subject origins, and the like), different overhang lengths, different overhang types (e.g., 5' overhangs, 3' overhangs, no overhangs), or combinations thereof. In some embodiments, a mixture of nucleic acids comprises single-stranded nucleic acid and double-stranded nucleic acid. In some embodiment, a mixture of nucleic acids comprises DNA and RNA. In some embodiment, a mixture of nucleic acids comprises ribosomal RNA (rRNA) and messenger RNA (mRNA). Nucleic acid provided for processes described herein may contain nucleic acid from one sample or from two or more samples (e.g., from 1 or more, 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, 10 or more, 11 or more, 12 or more, 13 or more, 14 or more, 15 or more, 16 or more, 17 or more, 18 or more, 19 or more, or 20 or more samples).

In some embodiments, target nucleic acids (e.g., ssNAs) comprise degraded DNA. Degraded DNA may be referred to as low-quality DNA or highly degraded DNA. Degraded DNA may be highly fragmented, and may include damage such as base analogs and abasic sites subject to miscoding lesions and/or intermolecular crosslinking. For example, sequencing errors resulting from deamination of cytosine residues may be present in certain sequences obtained from degraded DNA (e.g., miscoding of C to T and G to A). In some embodiments, target nucleic acids (e.g., ssNAs) are derived from nicked double-stranded nucleic acid fragments. Nicked double-stranded nucleic acid fragments may be denatured (e.g., heat denatured) to generate ssNA fragments.

Nucleic acid may be derived from one or more sources (e.g., biological sample, blood, cells, serum, plasma, buffy coat, urine, lymphatic fluid, skin, hair, soil, and the like) by methods known in the art. Any suitable method can be used for isolating, extracting and/or purifying DNA from a biological sample (e.g., from blood or a blood product), non-limiting examples of which include methods of DNA preparation (e.g., described by Sambrook and Russell, *Molecular Cloning: A Laboratory Manual* 3d ed., 2001), various commercially available reagents or kits, such as DNeasy®, RNeasy®, QIAprep®, QIAquick®, and QIAamp® (e.g., QIAamp® Circulating Nucleic Acid Kit, QIAamp® DNA Mini Kit or QIAamp® DNA Blood Mini Kit) nucleic acid isolation/purification kits by Qiagen, Inc. (Germantown, Md); GenomicPrep™ Blood DNA Isolation Kit (Promega, Madison, Wis.); GFX™ Genomic Blood DNA Purification Kit (Amersham, Piscataway, N.J.); DNAzol®, ChargeSwitch®, Purelink®, GeneCatcher® nucleic acid isolation/purification kits by Life Technologies, Inc. (Carlsbad, CA); NucleoMag®, NucleoSpin®, and NucleoBond® nucleic acid isolation/purification kits by Clontech Laboratories, Inc. (Mountain View, CA); the like or combinations thereof. In certain aspects, the nucleic acid is isolated from a fixed biological sample, e.g., formalin-fixed, paraffin-embedded (FFPE) tissue. Genomic DNA from FFPE tissue may be isolated using commercially available kits – such as the AllPrep® DNA/RNA FFPE kit by Qiagen, Inc. (Germantown, Md), the

RecoverAll® Total Nucleic Acid Isolation kit for FFPE by Life Technologies, Inc. (Carlsbad, CA), and the NucleoSpin® FFPE kits by Clontech Laboratories, Inc. (Mountain View, CA).

In some embodiments, nucleic acid is extracted from cells using a cell lysis procedure. Cell lysis procedures and reagents are known in the art and may generally be performed by chemical (e.g., detergent, hypotonic solutions, enzymatic procedures, and the like, or combination thereof), physical (e.g., French press, sonication, and the like), or electrolytic lysis methods. Any suitable lysis procedure can be utilized. For example, chemical methods generally employ lysing agents to disrupt cells and extract the nucleic acids from the cells, followed by treatment with chaotropic salts. Physical methods such as freeze/thaw followed by grinding, the use of cell presses and the like also are useful. In some instances, a high salt and/or an alkaline lysis procedure may be utilized. In some instances, a lysis procedure may include a lysis step with EDTA/Proteinase K, a binding buffer step with high amount of salts (e.g., guanidinium chloride (GuHCl), sodium acetate) and isopropanol, and binding DNA in this solution to silica-based column. In some instances, a lysis protocol includes certain procedures described in Dabney et al., Proceedings of the National Academy of Sciences 110, no. 39 (2013): 15758-15763.

Nucleic acids can include extracellular nucleic acid in certain embodiments. The term "extracellular nucleic acid" as used herein can refer to nucleic acid isolated from a source having substantially no cells and also is referred to as "cell-free" nucleic acid (cell-free DNA, cell-free RNA, or both), "circulating cell-free nucleic acid" (e.g., CCF fragments, ccfDNA) and/or "cell-free circulating nucleic acid." Extracellular nucleic acid can be present in and obtained from blood (e.g., from the blood of a human subject). Extracellular nucleic acid often includes no detectable cells and may contain cellular elements or cellular remnants. Non-limiting examples of acellular sources for extracellular nucleic acid are blood, blood plasma, blood serum and urine. In certain aspects, cell-free nucleic acid is obtained from a body fluid sample chosen from whole blood, blood plasma, blood serum, amniotic fluid, saliva, urine, pleural effusion, bronchial lavage, bronchial aspirates, breast milk, colostrum, tears, seminal fluid, peritoneal fluid, pleural effusion, and stool. As used herein, the term "obtain cell-free circulating sample nucleic acid" includes obtaining a sample directly (e.g., collecting a sample, e.g., a test sample) or obtaining a sample from another who has collected a sample. Extracellular nucleic acid may be a product of cellular secretion and/or nucleic acid release (e.g., DNA release). Extracellular nucleic acid may be a product of any form of cell death, for example. In some instances, extracellular nucleic acid is a product of any form of type I or type II cell death, including mitotic, oncotic, toxic, ischemic, and the like and combinations thereof. Without being limited by theory, extracellular nucleic acid may be a product of cell apoptosis and cell breakdown, which provides basis for extracellular nucleic acid often having a series of lengths across a spectrum (e.g., a "ladder"). In some instances, extracellular nucleic acid is a product of

cell necrosis, necroptosis, oncosis, entosis, pyroptosis, and the like and combinations thereof. In some embodiments, sample nucleic acid from a test subject is circulating cell-free nucleic acid. In some embodiments, circulating cell free nucleic acid is from blood plasma or blood serum from a test subject. In some aspects, cell-free nucleic acid is degraded. In some embodiments, cell-free

5 nucleic acid comprises cell-free fetal nucleic acid (e.g., cell-free fetal DNA). In certain aspects, cell-free nucleic acid comprises circulating cancer nucleic acid (e.g., cancer DNA). In certain aspects, cell-free nucleic acid comprises circulating tumor nucleic acid (e.g., tumor DNA). In some

embodiments, cell-free nucleic acid comprises infectious agent nucleic acid (e.g., pathogen DNA). In some embodiments, cell-free nucleic acid comprises nucleic acid (e.g., DNA) from a transplant.

10 In some embodiments, cell-free nucleic acid comprises nucleic acid (e.g., DNA) from a microbiome (e.g., microbiome of gut, microbiome of blood, microbiome of mouth, microbiome of spinal fluid, microbiome of feces).

Cell-free DNA (cfDNA) may originate from degraded sources and often provides limiting amounts of DNA when extracted. Methods described herein for generating single-stranded DNA (ssDNA)

15 libraries are able to capture a larger amount of short DNA fragments from cfDNA. cfDNA from cancer samples, for example, tends to have a higher population of short fragments. In certain instances, short fragments in cfDNA may be enriched for fragments originating from transcription factors rather than nucleosomes.

Extracellular nucleic acid can include different nucleic acid species, and therefore is referred to

20 herein as "heterogeneous" in certain embodiments. For example, blood serum or plasma from a person having a tumor or cancer can include nucleic acid from tumor cells or cancer cells (e.g., neoplasia) and nucleic acid from non-tumor cells or non-cancer cells. In another example, blood serum or plasma from a pregnant female can include maternal nucleic acid and fetal nucleic acid. In another example, blood serum or plasma from a patient having an infection or infectious disease

25 can include host nucleic acid and infectious agent or pathogen nucleic acid. In another example, a sample from a subject having received a transplant can include host nucleic acid and nucleic acid from the donor organ or tissue. In some instances, cancer nucleic acid, tumor nucleic acid, fetal nucleic acid, pathogen nucleic acid, or transplant nucleic acid sometimes is about 5% to about 50%

30 of the overall nucleic acid (e.g., about 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, or 49% of the total nucleic acid is cancer, tumor, fetal, pathogen, transplant, or microbiome nucleic acid). In another example, heterogeneous nucleic acid may include nucleic acid from two or more subjects (e.g., a sample from a crime scene).

At least two different nucleic acid species can exist in different amounts in extracellular nucleic acid and sometimes are referred to as minority species and majority species. In certain instances, a minority species of nucleic acid is from an affected cell type (e.g., cancer cell, wasting cell, cell attacked by immune system). In certain embodiments, a genetic variation or genetic alteration (e.g., copy number alteration, copy number variation, single nucleotide alteration, single nucleotide variation, chromosome alteration, and/or translocation) is determined for a minority nucleic acid species. In certain embodiments, a genetic variation or genetic alteration is determined for a majority nucleic acid species. Generally, it is not intended that the terms "minority" or "majority" be rigidly defined in any respect. In one aspect, a nucleic acid that is considered "minority," for example, can have an abundance of at least about 0.1% of the total nucleic acid in a sample to less than 50% of the total nucleic acid in a sample. In some embodiments, a minority nucleic acid can have an abundance of at least about 1% of the total nucleic acid in a sample to about 40% of the total nucleic acid in a sample. In some embodiments, a minority nucleic acid can have an abundance of at least about 2% of the total nucleic acid in a sample to about 30% of the total nucleic acid in a sample. In some embodiments, a minority nucleic acid can have an abundance of at least about 3% of the total nucleic acid in a sample to about 25% of the total nucleic acid in a sample. For example, a minority nucleic acid can have an abundance of about 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29% or 30% of the total nucleic acid in a sample. In some instances, a minority species of extracellular nucleic acid sometimes is about 1% to about 40% of the overall nucleic acid (e.g., about 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39% or 40% of the nucleic acid is minority species nucleic acid). In some embodiments, the minority nucleic acid is extracellular DNA. In some embodiments, the minority nucleic acid is extracellular DNA from apoptotic tissue. In some embodiments, the minority nucleic acid is extracellular DNA from tissue where some cells therein underwent apoptosis. In some embodiments, the minority nucleic acid is extracellular DNA from necrotic tissue. In some embodiments, the minority nucleic acid is extracellular DNA from tissue where some cells therein underwent necrosis. Necrosis may refer to a post-mortem process following cell death, in certain instances. In some embodiments, the minority nucleic acid is extracellular DNA from tissue affected by a cell proliferative disorder (e.g., cancer). In some embodiments, the minority nucleic acid is extracellular DNA from a tumor cell. In some embodiments, the minority nucleic acid is extracellular fetal DNA. In some embodiments, the minority nucleic acid is extracellular DNA from a pathogen. In some embodiments, the minority

nucleic acid is extracellular DNA from a transplant. In some embodiments, the minority nucleic acid is extracellular DNA from a microbiome.

In another aspect, a nucleic acid that is considered "majority," for example, can have an abundance greater than 50% of the total nucleic acid in a sample to about 99.9% of the total nucleic acid in a sample. In some embodiments, a majority nucleic acid can have an abundance of at least about 60% of the total nucleic acid in a sample to about 99% of the total nucleic acid in a sample. In some embodiments, a majority nucleic acid can have an abundance of at least about 70% of the total nucleic acid in a sample to about 98% of the total nucleic acid in a sample. In some embodiments, a majority nucleic acid can have an abundance of at least about 75% of the total nucleic acid in a sample to about 97% of the total nucleic acid in a sample. For example, a majority nucleic acid can have an abundance of at least about 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% of the total nucleic acid in a sample. In some embodiments, the majority nucleic acid is extracellular DNA. In some embodiments, the majority nucleic acid is extracellular maternal DNA. In some embodiments, the majority nucleic acid is DNA from healthy tissue. In some embodiments, the majority nucleic acid is DNA from non-tumor cells. In some embodiments, the majority nucleic acid is DNA from host cells.

In some embodiments, a minority species of extracellular nucleic acid is of a length of about 500 base pairs or less (e.g., about 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% of minority species nucleic acid is of a length of about 500 base pairs or less). In some embodiments, a minority species of extracellular nucleic acid is of a length of about 300 base pairs or less (e.g., about 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% of minority species nucleic acid is of a length of about 300 base pairs or less). In some embodiments, a minority species of extracellular nucleic acid is of a length of about 250 base pairs or less (e.g., about 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% of minority species nucleic acid is of a length of about 250 base pairs or less). In some embodiments, a minority species of extracellular nucleic acid is of a length of about 200 base pairs or less (e.g., about 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% of minority species nucleic acid is of a length of about 200 base pairs or less). In some embodiments, a minority species of extracellular nucleic acid is of a length of about 150 base pairs or less (e.g., about 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% of minority species nucleic acid is of a length of about 150 base pairs or less). In some embodiments, a minority species of extracellular nucleic acid is of a length of about 100 base pairs or less (e.g., about 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% of minority species nucleic acid is of a length of about 100 base pairs or less). In some embodiments, a minority species of extracellular nucleic acid is of a length of about

50 base pairs or less (e.g., about 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% of minority species nucleic acid is of a length of about 50 base pairs or less).

Nucleic acid may be provided for conducting methods described herein with or without processing of the sample(s) containing the nucleic acid. In some embodiments, nucleic acid is provided for conducting methods described herein after processing of the sample(s) containing the nucleic acid. For example, a nucleic acid can be extracted, isolated, purified, partially purified or amplified from the sample(s). The term "isolated" as used herein refers to nucleic acid removed from its original environment (e.g., the natural environment if it is naturally occurring, or a host cell if expressed exogenously), and thus is altered by human intervention (e.g., "by the hand of man") from its original environment. The term "isolated nucleic acid" as used herein can refer to a nucleic acid removed from a subject (e.g., a human subject). An isolated nucleic acid can be provided with fewer non-nucleic acid components (e.g., protein, lipid) than the amount of components present in a source sample. A composition comprising isolated nucleic acid can be about 50% to greater than 99% free of non-nucleic acid components. A composition comprising isolated nucleic acid can be about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or greater than 99% free of non-nucleic acid components. The term "purified" as used herein can refer to a nucleic acid provided that contains fewer non-nucleic acid components (e.g., protein, lipid, carbohydrate) than the amount of non-nucleic acid components present prior to subjecting the nucleic acid to a purification procedure. A composition comprising purified nucleic acid may be about 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or greater than 99% free of other non-nucleic acid components. The term "purified" as used herein can refer to a nucleic acid provided that contains fewer nucleic acid species than in the sample source from which the nucleic acid is derived. A composition comprising purified nucleic acid may be about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or greater than 99% free of other nucleic acid species. For example, fetal nucleic acid can be purified from a mixture comprising maternal and fetal nucleic acid. In certain examples, small fragments of nucleic acid (e.g., 30 to 500 bp fragments) can be purified, or partially purified, from a mixture comprising nucleic acid fragments of different lengths. In certain examples, nucleosomes comprising smaller fragments of nucleic acid can be purified from a mixture of larger nucleosome complexes comprising larger fragments of nucleic acid. In certain examples, larger nucleosome complexes comprising larger fragments of nucleic acid can be purified from nucleosomes comprising smaller fragments of nucleic acid. In certain examples, small fragments of fetal nucleic acid (e.g., 30 to 500 bp fragments) can be purified, or partially purified, from a mixture comprising both fetal and maternal nucleic acid fragments. In certain examples, nucleosomes comprising smaller fragments of fetal nucleic acid can be purified from a mixture of larger nucleosome complexes comprising larger

fragments of maternal nucleic acid. In certain examples, cancer cell nucleic acid can be purified from a mixture comprising cancer cell and non-cancer cell nucleic acid. In certain examples, nucleosomes comprising small fragments of cancer cell nucleic acid can be purified from a mixture of larger nucleosome complexes comprising larger fragments of non-cancer nucleic acid. In some
5 embodiments, nucleic acid is provided for conducting methods described herein without prior processing of the sample(s) containing the nucleic acid. For example, nucleic acid may be analyzed directly from a sample without prior extraction, purification, partial purification, and/or amplification.

Nucleic acids may be amplified under amplification conditions. The term “amplified” or
10 “amplification” or “amplification conditions” as used herein refers to subjecting a target nucleic acid (e.g., ssNA) in a sample or a nucleic acid product generated by a method herein to a process that linearly or exponentially generates amplicon nucleic acids having the same or substantially the same nucleotide sequence as the target nucleic acid (e.g., ssNA), or part thereof. In certain
15 embodiments, the term “amplified” or “amplification” or “amplification conditions” refers to a method that comprises a polymerase chain reaction (PCR). In certain instances, an amplified product can contain one or more nucleotides more than the amplified nucleotide region of a nucleic acid
20 template sequence (e.g., a primer can contain “extra” nucleotides such as a transcriptional initiation sequence, in addition to nucleotides complementary to a nucleic acid template gene molecule, resulting in an amplified product containing “extra” nucleotides or nucleotides not corresponding to
the amplified nucleotide region of the nucleic acid template gene molecule).

Nucleic acid also may be exposed to a process that modifies certain nucleotides in the nucleic acid before providing nucleic acid for a method described herein. A process that selectively modifies
nucleic acid based upon the methylation state of nucleotides therein can be applied to nucleic acid,
for example. In addition, conditions such as high temperature, ultraviolet radiation, x-radiation, can
25 induce changes in the sequence of a nucleic acid molecule. Nucleic acid may be provided in any suitable form useful for conducting a sequence analysis.

In some embodiments, target nucleic acids (e.g., ssNAs) are not modified in prior to combining with the scaffold adapters herein, or components thereof. In some embodiments, target nucleic acids
(e.g., ssNAs) are not modified in length prior to combining with the scaffold adapters herein, or
30 components thereof. In this context, “not modified” means that target nucleic acids are isolated from a sample and then combined with scaffold adapters, or components thereof, without modifying the length or the composition of the target nucleic acids. For example, target nucleic acids (e.g., ssNAs) may not be shortened (e.g., they are not contacted with a restriction enzyme or nuclease or physical condition that reduces length (e.g., shearing condition, cleavage condition)) and may not

be increased in length by one or more nucleotides (e.g., ends are not filled in at overhangs; no nucleotides are added to the ends). Adding a phosphate or chemically reactive group to one or both ends of a target nucleic acid (e.g., ssNA) generally is not considered modifying the nucleic acid or modifying the length of the nucleic acid. Denaturing a double-stranded nucleic acid (dsNA) fragment to generate an ssNA fragment generally is not considered modifying the nucleic acid or modifying the length of the nucleic acid.

In some embodiments, one or both native ends of target nucleic acids (e.g., ssNAs) are present when the ssNA is combined with the scaffold adapters herein, or components thereof. Native ends generally refer to unmodified ends of a nucleic acid fragment. In some embodiments, native ends of target nucleic acids (e.g., ssNAs) are not modified in length prior to combining with the scaffold adapters herein, or components thereof. In this context, "not modified" means that target nucleic acids are isolated from a sample and then combined with scaffold adapters, or components thereof, without modifying the length of the native ends of target nucleic acids. For example, target nucleic acids (e.g., ssNAs) are not shortened (e.g., they are not contacted with a restriction enzyme or nuclease or physical condition that reduces length (e.g., shearing condition, cleavage condition) to generate non-native ends) and are not increased in length by one or more nucleotides (e.g., native ends are not filled in at overhangs; no nucleotides are added to the native ends). Adding a phosphate or chemically reactive group to one or both native ends of a target nucleic acid generally is not considered modifying the length of the nucleic acid.

In some embodiments, target nucleic acids (e.g., ssNAs) are not contacting with a cleavage agent (e.g., endonuclease, exonuclease, restriction enzyme) and/or a polymerase prior to combining with the scaffold adapters herein, or components thereof. In some embodiments, target nucleic acids are not subjected to mechanical shearing (e.g., ultrasonication (e.g., Adaptive Focused Acoustics™ (AFA) process by Covaris)) prior to combining with the scaffold adapters herein, or components thereof. In some embodiments, target nucleic acids are not contacting with an exonuclease (e.g., DNase) prior to combining with the scaffold adapters herein, or components thereof. In some embodiments, target nucleic acids are not amplified prior to combining with the scaffold adapters herein, or components thereof. In some embodiments, target nucleic acids are not attached to a solid support prior to combining with the scaffold adapters herein, or components thereof. In some embodiments, target nucleic acids are not conjugated to another molecule prior to combining with the scaffold adapters herein, or components thereof. In some embodiments, target nucleic acids are not cloned into a vector prior to combining with the scaffold adapters herein, or components thereof. In some embodiments, target nucleic acids may be subjected to dephosphorylation prior to combining with the scaffold adapters herein, or components thereof. In some embodiments, target

nucleic acids may be subjected to phosphorylation prior to combining with the scaffold adapters herein, or components thereof.

In some embodiments, combining target nucleic acids (e.g., ssNAs) with the scaffold adapters herein, or components thereof, comprises isolating the target nucleic acids, and combining the isolated target nucleic acids with the scaffold adapters herein, or components thereof. In some
5 embodiments, combining target nucleic acids with the scaffold adapters herein, or components thereof, comprises isolating the target nucleic acids, phosphorylating the isolated target nucleic acids, and combining the phosphorylated target nucleic acids with the scaffold adapters herein, or components thereof. In some embodiments, combining target nucleic acids with the scaffold
10 adapters herein, or components thereof, comprises isolating the target nucleic acids, dephosphorylating the scaffold adapters herein, or components thereof, and combining the isolated target nucleic acids with the dephosphorylated scaffold adapters herein, or dephosphorylated components thereof. In some embodiments, combining target nucleic acids with the scaffold
adapters herein, or components thereof, comprises isolating the target nucleic acids,
15 dephosphorylating the isolated target nucleic acids, phosphorylating the dephosphorylated target nucleic acids, and combining the phosphorylated target nucleic acids with the scaffold adapters herein, or components thereof. In some embodiments, combining target nucleic acids with the scaffold adapters herein, or components thereof, comprises isolating the target nucleic acids,
dephosphorylating the isolated target nucleic acids, phosphorylating the dephosphorylated target
20 nucleic acids, dephosphorylating the scaffold adapters, or components thereof, and combining the phosphorylated target nucleic acids with the dephosphorylated scaffold adapters herein, or dephosphorylated components thereof.

In some embodiments, combining target nucleic acids (e.g., ssNAs) with the scaffold adapters herein, or components thereof, consists of isolating the target nucleic acids, and combining the
25 isolated target nucleic acids with the scaffold adapters herein, or components thereof. In some embodiments, combining target nucleic acids with the scaffold adapters herein, or components thereof, consists of isolating the target nucleic acids, phosphorylating the isolated target nucleic acids, and combining the phosphorylated target nucleic acids with the scaffold adapters herein, or components thereof. In some embodiments, combining target nucleic acids with the scaffold
30 adapters herein, or components thereof, consists of isolating the target nucleic acids, dephosphorylating the scaffold adapters, or components thereof, and combining the isolated target nucleic acids with the dephosphorylated scaffold adapters herein, or dephosphorylated components thereof. In some embodiments, combining target nucleic acids with the scaffold
adapters herein, or components thereof, consists of isolating the target nucleic acids,
35 dephosphorylating the isolated target nucleic acids, phosphorylating the dephosphorylated target

nucleic acids, and combining the phosphorylated target nucleic acids with the scaffold adapters herein, or components thereof. In some embodiments, combining target nucleic acids with the scaffold adapters herein, or components thereof, consists of isolating the target nucleic acids, dephosphorylating the isolated target nucleic acids, phosphorylating the dephosphorylated target nucleic acids, dephosphorylating the scaffold adapters, or components thereof, and combining the phosphorylated target nucleic acids with the dephosphorylated scaffold adapters herein, or dephosphorylated components thereof.

Single-stranded nucleic acid

Provided herein are methods and compositions for capturing single-stranded nucleic acid (ssNA) using specialized adapters (e.g., for generating a sequencing library). Single-stranded nucleic acid or ssNA generally refers to a collection of polynucleotides which are single-stranded (i.e., not hybridized intermolecularly or intramolecularly) over 70% or more of their length. In some embodiments, ssNA is single-stranded over 75% or more, 80% or more, 85% or more, 90% or more, 95% or more, or 99% or more, of the length of the polynucleotides. In certain aspects, the ssNA is single-stranded over the entire length of the polynucleotides. Single-stranded nucleic acid may be referred to herein as target nucleic acid.

ssNA may include single-stranded deoxyribonucleic acid (ssDNA). In some embodiments, ssDNA includes, but is not limited to, ssDNA derived from double-stranded DNA (dsDNA). For example, ssDNA may be derived from double-stranded DNA which is denatured (e.g., heat denatured and/or chemically denatured) to produce ssDNA. In some embodiments, a method herein comprises, prior to combining ssDNA with scaffold adapters described herein, or components thereof, generating the ssDNA by denaturing dsDNA.

In some embodiments, ssNA includes single-stranded ribonucleic acid (ssRNA). RNA may include, for example, messenger RNA (mRNA), microRNA (miRNA), small interfering RNA (siRNA), transacting small interfering RNA (ta-siRNA), natural small interfering RNA (nat-siRNA), ribosomal RNA (rRNA), transfer RNA (tRNA), small nucleolar RNA (snoRNA), small nuclear RNA (snRNA), long non-coding RNA (lncRNA), non-coding RNA (ncRNA), transfer-messenger RNA (tmRNA), precursor messenger RNA (pre-mRNA), small Cajal body-specific RNA (scaRNA), piwi-interacting RNA (piRNA), endoribonuclease-prepared siRNA (esiRNA), small temporal RNA (stRNA), signal recognition RNA, telomere RNA, ribozyme, or a combination thereof. In some embodiments, when the ssNA is ssRNA, the ssRNA is mRNA. In some embodiments, ssNA includes single stranded complementary DNA (cDNA).

In some embodiments, a method herein comprises contacting ssNA with a single-stranded nucleic acid binding agent. In some embodiments, a method herein comprises contacting ssNA with single-

stranded nucleic acid binding protein (SSB) to produce SSB-bound ssNA. In some embodiments, a method herein comprises contacting sscDNA with single-stranded nucleic acid binding protein (SSB) to produce SSB-bound sscDNA. In some embodiments, a method herein comprises contacting ssDNA with single-stranded nucleic acid binding protein (SSB) to produce SSB-bound ssDNA. In some embodiments, a method herein comprises contacting ssRNA with single-stranded nucleic acid binding protein (SSB) to produce SSB-bound ssRNA. SSB generally binds in a cooperative manner to ssNA and typically does not bind well to double-stranded nucleic acid (dsNA). Upon binding ssDNA, SSB destabilizes helical duplexes. SSBs may be prokaryotic SSB (e.g., bacterial or archaeal SSB) or eukaryotic SSB. Examples of SSBs may include E. coli SSB, E. coli RecA, Extreme Thermostable Single-Stranded DNA Binding Protein (ET SSB), Thermus thermophilus (Tth) RecA, T4 Gene 32 Protein, replication protein A (RPA – a eukaryotic SSB), and the like. ET SSB, Tth RecA, E. coli RecA, T4 Gene 32 Protein, as well buffers and detailed protocols for preparing SSB-bound ssNA using such SSBs are commercially available (e.g., New England Biolabs, Inc. (Ipswich, MA)).

In some embodiments, a method herein does not comprise contacting ssNA with single-stranded nucleic acid binding protein (SSB) to produce SSB-bound ssNA. Accordingly, a method herein may omit the step of producing SSB-bound ssNA. For example, a method herein may comprise combining ssNA with scaffold adapters described herein, or components thereof, without contacting the ssNA with SSB. In such instances, a method herein may be referred to an “SSB-free” method for producing a nucleic acid library. Certain SSB-free methods described herein may produce libraries having parameters similar to parameters for libraries prepared using SSB, as shown in the Drawings and discussed in the Examples. In some embodiments, a method herein comprises contacting ssNA with a single-stranded nucleic acid binding agent other than SSB. Such single-stranded nucleic acid binding agents can stably bind single stranded nucleic acids, can prevent or reduce formation of nucleic acid duplexes, can still allow the bound nucleic acids to be ligated or otherwise terminally modified, and can be thermostable. Example single-stranded nucleic acid binding agents include but are not limited to topoisomerases, helicases, domains thereof, and fusion proteins comprising domains thereof.

In some embodiments, a method herein comprises combining a nucleic acid composition comprising single-stranded nucleic acid (ssNA) with scaffold adapters described herein, or components thereof. In some embodiments, a method herein comprises combining a nucleic acid composition consisting of single-stranded nucleic acid (ssNA) with scaffold adapters described herein, or components thereof. In some embodiments, a method herein comprises combining a nucleic acid composition consisting essentially of single-stranded nucleic acid (ssNA) with scaffold adapters described herein, or components thereof. A nucleic acid composition “consisting

essentially of" single-stranded nucleic acid (ssNA) generally includes ssNA and no additional protein or nucleic acid components. For example, a nucleic acid composition "consisting essentially of" single-stranded nucleic acid (ssNA) may exclude double-stranded nucleic acid (dsNA) or may include a low percentage of dsNA (e.g., less than 10% dsNA, less than 5% dsNA, less than 1% dsNA). A nucleic acid composition "consisting essentially of" single-stranded nucleic acid (ssNA) may exclude proteins. For example, a nucleic acid composition "consisting essentially of" single-stranded nucleic acid (ssNA) may exclude single-stranded binding proteins (SSBs) or other proteins useful for stabilizing ssNA. A nucleic acid composition "consisting essentially of" single-stranded nucleic acid (ssNA) may include chemical components typically present in nucleic acid compositions such as buffers, salts, alcohols, crowding agents (e.g., PEG), and the like; and may include residual components (e.g., nucleic acids, proteins, cell membrane components) from the nucleic acid source (e.g., sample) or nucleic acid extraction. A nucleic acid composition "consisting essentially of" single-stranded nucleic acid (ssNA) may include ssNA fragments having one or more phosphates (e.g., a terminal phosphate, a 5' terminal phosphate). A nucleic acid composition "consisting essentially of" single-stranded nucleic acid (ssNA) may include ssNA fragments comprising one or more modified nucleotides.

Enriching nucleic acids

In some embodiments, nucleic acid (e.g., extracellular nucleic acid) is enriched or relatively enriched for a subpopulation or species of nucleic acid. Nucleic acid subpopulations can include, for example, fetal nucleic acid, maternal nucleic acid, cancer nucleic acid, tumor nucleic acid, patient nucleic acid, host nucleic acid, pathogen nucleic acid, transplant nucleic acid, microbiome nucleic acid, nucleic acid comprising fragments of a particular length or range of lengths, or nucleic acid from a particular genome region (e.g., single chromosome, set of chromosomes, and/or certain chromosome regions). Such enriched samples can be used in conjunction with a method provided herein. Thus, in certain embodiments, methods of the technology comprise an additional step of enriching for a subpopulation of nucleic acid in a sample. In certain embodiments, nucleic acid from normal tissue (e.g., non-cancer cells, host cells) is selectively removed (partially, substantially, almost completely or completely) from the sample. In certain embodiments, maternal nucleic acid is selectively removed (partially, substantially, almost completely or completely) from the sample. In certain embodiments, enriching for a particular low copy number species nucleic acid (e.g., cancer, tumor, fetal, pathogen, transplant, microbiome nucleic acid) may improve quantitative sensitivity. Methods for enriching a sample for a particular species of nucleic acid are described, for example, in U.S. Patent No. 6,927,028, International Patent Application Publication No. WO2007/140417, International Patent Application Publication No. WO2007/147063, International Patent Application Publication No. WO2009/032779, International Patent Application

Publication No. WO2009/032781, International Patent Application Publication No. WO2010/033639, International Patent Application Publication No. WO2011/034631, International Patent Application Publication No. WO2006/056480, and International Patent Application Publication No. WO2011/143659, the entire content of each is incorporated herein by reference, including all text, tables, equations and drawings.

In some embodiments, nucleic acid is enriched for certain target fragment species and/or reference fragment species. In certain embodiments, nucleic acid is enriched for a specific nucleic acid fragment length or range of fragment lengths using one or more length-based separation methods described below. In certain embodiments, nucleic acid is enriched for fragments from a select genomic region (e.g., chromosome) using one or more sequence-based separation methods described herein and/or known in the art.

Non-limiting examples of methods for enriching for a nucleic acid subpopulation in a sample include methods that exploit epigenetic differences between nucleic acid species (e.g., methylation-based fetal nucleic acid enrichment methods described in U.S. Patent Application Publication No. 2010/0105049, which is incorporated by reference herein); restriction endonuclease enhanced polymorphic sequence approaches (e.g., such as a method described in U.S. Patent Application Publication No. 2009/0317818, which is incorporated by reference herein); selective enzymatic degradation approaches; massively parallel signature sequencing (MPSS) approaches; amplification (e.g., PCR)-based approaches (e.g., loci-specific amplification methods, multiplex SNP allele PCR approaches; universal amplification methods); pull-down approaches (e.g., biotinylated ultramer pull-down methods); extension and ligation-based methods (e.g., molecular inversion probe (MIP) extension and ligation); and combinations thereof.

In some embodiments, modified nucleic acids can be enriched for. Nucleic acid modifications include but are not limited to carboxycytosine, 5-methylcytosine (5mC) and its oxidative derivatives (e.g., 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC)), N(6)-methyladenine (6mA), N4-methylcytosine (4mC), N(6)-methyladenosine (m(6)A), pseudouridine (Ψ), 5-methylcytidine (m(5)C), hydroxymethyl uracil, 2'-O-methylation at the 3' end, tRNA modifications, miRNA modifications, and snRNA modifications. Nucleic acids comprising one or more modifications can be enriched for by a variety of methods, including but not limited to antibody-based pulldown. Modified nucleic acid enrichment can be conducted before or after denaturation of dsDNA. Enrichment prior to denaturation can result in also enriching for the complementary strand which may lack the modification, while enrichment after denaturation does not enrich for complementary strands lacking modification.

In some embodiments, nucleic acid is enriched for fragments from a select genomic region (e.g., chromosome) using one or more sequence-based separation methods described herein.

Sequence-based separation generally is based on nucleotide sequences present in the fragments of interest (e.g., target and/or reference fragments) and substantially not present in other fragments of the sample or present in an insubstantial amount of the other fragments (e.g., 5% or less). In some embodiments, sequence-based separation can generate separated target fragments and/or separated reference fragments. Separated target fragments and/or separated reference fragments often are isolated away from the remaining fragments in the nucleic acid sample. In certain embodiments, the separated target fragments and the separated reference fragments also are isolated away from each other (e.g., isolated in separate assay compartments). In certain embodiments, the separated target fragments and the separated reference fragments are isolated together (e.g., isolated in the same assay compartment). In some embodiments, unbound fragments can be differentially removed or degraded or digested.

In some embodiments, scaffold adapters are used to enrich for target nucleic acids. For example, scaffold adapters can be designed such that some or all of the bases in the ssNA hybridization region are defined or known bases. These scaffold adapters can hybridize preferentially to target nucleic acids with sequences complementary to the defined or known bases of the scaffold adapter ssNA hybridization region, thereby enriching for the target nucleic acids in the resulting library. For example, including a GC dinucleotide in the ssNA hybridization region can be used to enrich for target nucleic acids that have terminal CG (also called CpG) dinucleotides. Any other defined sequence can be targeted in a similar manner, using some or all of the length of the scaffold adapter ssNA hybridization region, including but not limited to nuclease cleavage sites, gene promoter regions, pathogen sequences, tumor-related sequences, and other motifs. In an example, libraries were prepared using non-enriching scaffold adapters and CG dinucleotide enriching scaffold adapters. For libraries prepared without enrichment, 1.7% of reads started with CG and 1.1% of reads ended with CG. For libraries prepared with enrichment, 5.2% of reads started with CG and 19.6% of reads ended with CG. In another example, a sample comprising RNA (e.g., host and pathogen RNA) is reverse transcribed with primers specific to pathogen RNA of interest to generate cDNA; the cDNA is then purified and prepared with single-stranded library preparation methods as discussed herein, either with standard scaffold adapters or with scaffold adapters with ssNA hybridization regions targeted to the regions enriched by the reverse transcription primers. Pathogenic DNA can be similarly enriched.

In some instances, the target nucleic acid sequence at the 5' or 3' nucleic acid termini is defined or known. In other instances, scaffold adapters can be used to identify novel targets of interest at 5' or 3' nucleic acid termini. Nucleic acid sequences or patterns of interest may be characterized from

the scaffold adapter library output with or without enrichment. In some instances, a specific sequence or sequence pattern at 5', 3', or both nucleic acid termini may be associated with a particular state. Such states include but are not limited to disease state, methylation state, and gene expression state. The scaffold adapters can be used to quantify the presence or relative abundance of a known or novel target sequence(s) at nucleic acid termini between samples and controls, for example, cell-free DNA from cancer patients and healthy controls. These data can be used to learn the relationship between the sequence information at DNA termini and a given state. By training on a well-characterized dataset of patient and healthy samples, in one example, an analytical method or algorithm can be used to predict the state or transitions through the state. For example, an increase of AT dinucleotides and reduction of CpG dinucleotides at 5' and 3' DNA termini in cfDNA from patients with Acute Myeloid leukemia (AML) is observed when compared to non-AML patient samples. In this example, an analytical tool may be used cfDNA termini sequence information to predict a person's risk for developing AML.

In some embodiments, a selective nucleic acid capture process is used to separate target and/or reference fragments away from a nucleic acid sample. Commercially available nucleic acid capture systems include, for example, Nimblegen sequence capture system (Roche NimbleGen, Madison, WI); ILLUMINA BEADARRAY platform (Illumina, San Diego, CA); Affymetrix GENECHIP platform (Affymetrix, Santa Clara, CA); Agilent SureSelect Target Enrichment System (Agilent Technologies, Santa Clara, CA); and related platforms. Such methods typically involve hybridization of a capture oligonucleotide to a part or all of the nucleotide sequence of a target or reference fragment and can include use of a solid phase (e.g., solid phase array) and/or a solution-based platform. Capture oligonucleotides (sometimes referred to as "bait") can be selected or designed such that they preferentially hybridize to nucleic acid fragments from selected genomic regions or loci, or a particular sequence in a nucleic acid target. In certain embodiments, a hybridization-based method (e.g., using oligonucleotide arrays) can be used to enrich for fragments containing certain nucleic acid sequences. Thus, in some embodiments, a nucleic acid sample is optionally enriched by capturing a subset of fragments using capture oligonucleotides complementary to, for example, selected sequences in sample nucleic acid. In certain instances, captured fragments are amplified. For example, captured fragments containing adapters may be amplified using primers complementary to the adapter sequences to form collections of amplified fragments, indexed according to adapter sequence. In some embodiments, nucleic acid is enriched for fragments from a select genomic region (e.g., chromosome, a gene) by amplification of one or more regions of interest using oligonucleotides (e.g., PCR primers) complementary to sequences in fragments containing the region(s) of interest, or part(s) thereof.

In some embodiments, nucleic acid is enriched for a particular nucleic acid fragment length, range of lengths, or lengths under or over a particular threshold or cutoff using one or more length-based separation methods. Nucleic acid fragment length typically refers to the number of nucleotides in the fragment. Nucleic acid fragment length also is sometimes referred to as nucleic acid fragment size. In some embodiments, a length-based separation method is performed without measuring lengths of individual fragments. In some embodiments, a length-based separation method is performed in conjunction with a method for determining length of individual fragments. In some embodiments, length-based separation refers to a size fractionation procedure where all or part of the fractionated pool can be isolated (e.g., retained) and/or analyzed. Size fractionation procedures are known in the art (e.g., separation on an array, separation by a molecular sieve, separation by gel electrophoresis, separation by column chromatography (e.g., size-exclusion columns), and microfluidics-based approaches). In certain instances, length-based separation approaches can include selective sequence tagging approaches, fragment circularization, chemical treatment (e.g., formaldehyde, polyethylene glycol (PEG) precipitation), mass spectrometry and/or size-specific nucleic acid amplification, for example.

In some embodiments, nucleic acid is enriched for fragments associated with one or more nucleic acid binding proteins. Example enrichment methods include but are not limited to chromatin immunoprecipitation (ChIP), cross-linked ChIP (XChIP), native ChIP (NChIP), bead-free ChIP, carrier ChIP (CChIP), fast ChIP (qChIP), quick and quantitative ChIP (Q²ChIP), microchip (μ ChIP), matrix ChIP, pathology-ChIP (PAT-ChIP), ChIP-exo, ChIP-on-chip, RIP-ChIP, HiChIP, ChIA-PET, and HiChIRP.

In some embodiments, a method herein includes enriching an RNA species in a mixture of RNA species. In some embodiments, a method herein includes depleting an RNA species in a mixture of RNA species. For example, a method herein may include enriching and/or depleting an RNA species chosen from message RNA (mRNA), short inhibitory RNA (siRNA), ribosomal RNA (rRNA), transfer RNA (tRNA), microRNA, transacting small interfering RNA (ta-siRNA), natural small interfering RNA (nat-siRNA), small nucleolar RNA (snoRNA), small nuclear RNA (snRNA), long non-coding RNA (lncRNA), non-coding RNA (ncRNA), transfer-messenger RNA (tmRNA), precursor messenger RNA (pre-mRNA), small Cajal body-specific RNA (scaRNA), piwi-interacting RNA (piRNA), endoribonuclease-prepared siRNA (esiRNA), small temporal RNA (stRNA), signal recognition RNA, telomere RNA, RNA highly expressed by a fetus or placenta, and the like. In some embodiments, a method herein includes enriching an sscDNA species in a mixture of sscDNA species. For example, a method herein may comprise enriching messenger RNA (mRNA) present in a mixture of mRNA and ribosomal RNA (rRNA), or enriching sscDNA corresponding to mRNA present in a mixture of sscDNA corresponding to mRNA and rRNA. Any suitable mRNA

enrichment method may be used, which includes rRNA depletion and/or mRNA enrichment methods such as rRNA depletion with magnetic beads (e.g., Ribo-zero™, Ribominus™, and MICROBExpress™, which use rRNA depletion probes in combination with magnetic beads to deplete rRNAs from a sample, thus enriching mRNAs), oligo(dT)-based poly(A) enrichment (e.g.,
5 BioMag® Oligo (dT)20), nuclease-based rRNA depletion (e.g., digestion of rRNA with Terminator™ 5'-Phosphate Dependent Exonuclease), and combinations thereof.

Enrichment strategies can increase the relative abundance (e.g., as assessed by percent of sequencing reads) of the targeted nucleic acids by at least 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 200%, 300%, 400%, 500%, 600%, 700%, 800%, 900%, 1000%, 1100%,
10 1200%, 1300%, 1400%, 1500%, 1600%, 1700%, 1800%, 1900%, 2000%, 3000%, 4000%, 5000%, 6000%, 7000%, 8000%, 9000%, 10000%, or more.

In some embodiments, a method herein comprises contacting under hybridization conditions a set of ligation products (e.g., a set of ssNAs ligated to one or more scaffold adapters herein) with one or more probe species. In some embodiments, each probe species comprises a sequence
15 complementary to an RNA sequence. As such, a probe sequence may be complementary to an RNA sequence, or complement, or reverse complement thereof, and a probe sequence may be complementary to an sscDNA sequence generated from an RNA fragment, or complement, or reverse complement thereof. A probe sequence may be complementary to an entire RNA sequence or an entire sscDNA sequence, or may be complementary to a portion of an RNA
20 sequence or a portion of an sscDNA sequence. In some embodiments, each probe species comprises a sequence complementary to a ribosomal RNA (rRNA) sequence. As such, a probe sequence may be complementary to an rRNA sequence, or complement, or reverse complement thereof, and a probe sequence may be complementary to an sscDNA sequence generated from an rRNA fragment, or complement, or reverse complement thereof. A probe sequence may be
25 complementary to an entire rRNA sequence or an entire sscDNA sequence, or may be complementary to a portion of an rRNA sequence or a portion of an sscDNA sequence.

In some embodiments, a set of probe species is used to hybridize to ssNA in a workflow described herein. A set of probe species may include, for example, probes designed to hybridize to one or more sequences in an RNA molecule. A set of probe species may include, for example, probes
30 designed to hybridize to one or more ribosomal RNA (rRNA) sequences chosen from sequences encoding 18S subunit, 28S subunit, 5S subunit, 5.8S subunit, 12S subunit, 16S subunit, and RN7SL (e.g., RN7SL1 and RN7SL2), or to the complements thereof. rRNA sequences may be cytoplasmic and/or mitochondrial (mtrRNA). In some embodiments, a set of probes comprises one or more probes targeting adapter dimers, such as one or more probes targeting -P5 and -P7

regions. In some embodiments, a set of probes comprises one or more probes targeting transcripts from genes that are commonly highly expressed across many or all samples of a given type, such as ribonucleoprotein complex genes like RN7SL1 and RN7SL2. In some embodiments, a set of probes comprises one or more probes comprising nucleic acid sequences set forth in SEQ ID NOs: 1-174.

In some embodiments, a probe species can specifically hybridize to a ssNA, or portion thereof, comprising an RNA sequence, or a corresponding sscDNA sequence, complement thereof, or reverse complement thereof. In some embodiments, a probe species can specifically hybridize to a ssNA, or portion thereof, comprising a ribosomal RNA (rRNA) sequence, or a corresponding sscDNA sequence, complement thereof, or reverse complement thereof. Specific hybridization may be affected or influenced by factors such as the degree of complementarity between the ssNA and the probe, the length thereof, and the temperature at which the hybridization occurs, which may be informed by melting temperatures (T_m) of the probe and/or ssNA. Melting temperature generally refers to the temperature at which half of the probes /ssNAs remain hybridized and half of the probes /ssNAs dissociate into single strands. The T_m of a duplex may be experimentally determined or predicted using the following formula $T_m = 81.5 + 16.6(\log_{10}[\text{Na}^+]) + 0.41 (\text{fraction G+C}) - (60/N)$, where N is the chain length and $[\text{Na}^+]$ is less than 1 M. Additional models that depend on various parameters also may be used to predict T_m of relevant regions depending on various hybridization conditions. Generally, the on-off rate of hybridization that occurs between probes and ssNAs is essentially constant (prior to ligation). Approaches for achieving specific nucleic acid hybridization are described, e.g., Tijssen, Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes, part I, chapter 2, "Overview of principles of hybridization and the strategy of nucleic acid probe assays," Elsevier (1993).

In some embodiments, a probe species comprises (e.g., is conjugated to; comprises one or more nucleotides conjugated to) a member of a binding pair. Binding pairs may include, for example, antibody/antigen, antibody/antibody, antibody/antibody fragment, antibody/antibody receptor, antibody/protein A or protein G, hapten/anti-hapten, biotin/avidin, biotin/streptavidin, folic acid/folate binding protein, vitamin B12/intrinsic factor, chemical reactive group/complementary chemical reactive group, digoxigenin moiety/anti-digoxigenin antibody, fluorescein moiety/anti-fluorescein antibody, steroid/steroid-binding protein, operator/ repressor, nuclease/nucleotide, lectin/polysaccharide, active compound/active compound receptor, hormone/hormone receptor, enzyme/substrate, oligonucleotide or polynucleotide/its corresponding complement, the like or combinations thereof. In some embodiments, the binding pair is biotin and streptavidin.

In some embodiments, a probe species comprises a first member of a binding pair (e.g., biotin); and a second member of a binding pair (e.g., streptavidin) is conjugated to a solid support or substrate. A solid support or substrate can be any physically separable solid to which a member of a binding pair can be directly or indirectly attached including, but not limited to, surfaces provided
5 by microarrays and wells, and particles such as beads (e.g., paramagnetic beads, magnetic beads, microbeads, nanobeads), microparticles, and nanoparticles. Solid supports also can include, for example, chips, columns, optical fibers, wipes, filters (e.g., flat surface filters), one or more capillaries, glass and modified or functionalized glass (e.g., controlled-pore glass (CPG)), quartz, mica, diazotized membranes (paper or nylon), polyformaldehyde, cellulose, cellulose acetate,
10 paper, ceramics, metals, metalloids, semiconductive materials, quantum dots, coated beads or particles, other chromatographic materials, magnetic particles; plastics (including acrylics, polystyrene, copolymers of styrene or other materials, polybutylene, polyurethanes, TEFLON™, polyethylene, polypropylene, polyamide, polyester, polyvinylidenedifluoride (PVDF), and the like), polysaccharides, nylon or nitrocellulose, resins, silica or silica-based materials including silicon, silica gel, and modified silicon, Sephadex®, Sepharose®, carbon, metals (e.g., steel, gold, silver,
15 aluminum, silicon and copper), inorganic glasses, conducting polymers (including polymers such as polypyrrole and polyindole); micro or nanostructured surfaces such as nucleic acid tiling arrays, nanotube, nanowire, or nanoparticulate decorated surfaces; or porous surfaces or gels such as methacrylates, acrylamides, sugar polymers, cellulose, silicates, or other fibrous or stranded
20 polymers. In some embodiments, a solid support or substrate may be coated using passive or chemically-derivatized coatings with any number of materials, including polymers, such as dextrans, acrylamides, gelatins or agarose. Beads and/or particles may be free or in connection with one another (e.g., sintered). In some embodiments, a solid support can be a collection of particles. In some embodiments, the particles can comprise silica, and the silica may comprise
25 silica dioxide. In some embodiments, the silica can be porous, and in certain embodiments the silica can be non-porous. In some embodiments, the particles further comprise an agent that confers a paramagnetic property to the particles. In certain embodiments, the agent comprises a metal, and in certain embodiments the agent is a metal oxide, (e.g., iron or iron oxides, where the iron oxide contains a mixture of Fe²⁺ and Fe³⁺). A member of a binding pair may be linked to a
30 solid support by covalent bonds or by non-covalent interactions and may be linked to a solid support directly or indirectly (e.g., via an intermediary agent such as a spacer molecule or biotin).

After probe hybridization, a mixture of hybridized ligation products and unhybridized ligation products is generated, in some embodiments. Hybridized ligation products and unhybridized ligation products may be separated using any suitable separation method. In some embodiments,
35 hybridized ligation products and unhybridized ligation products may be separated using a suitable a

probe pull-down method. In some embodiments, hybridized ligation products comprising a first member of a binding pair may be separated using a solid support conjugated to a second member of a binding pair. For example, hybridized ligation products comprising a biotinylated probe may be separated using a solid support (e.g., magnetic bead) conjugated to a streptavidin. In some
5 embodiments, unhybridized ligation products are retained for further analysis (e.g., analysis of mRNA sequences). In some embodiments, hybridized ligation products are retained for further analysis (e.g., analysis of RNA sequences or a specific subset thereof; analysis of rRNA sequences or a specific subset thereof).

Length-based separation

10 In some embodiments, a method herein comprises separating target nucleic acids (e.g., ssNAs) according to fragment length. For example, target nucleic acids (e.g., ssNAs) may be enriched for a particular nucleic acid fragment length, range of lengths, or lengths under or over a particular threshold or cutoff using one or more length-based separation methods. Nucleic acid fragment length typically refers to the number of nucleotides in the fragment. Nucleic acid fragment length
15 also may be referred to as nucleic acid fragment size. In some embodiments, a length-based separation method is performed without measuring lengths of individual fragments. In some embodiments, a length-based separation method is performed in conjunction with a method for determining length of individual fragments. In some embodiments, length-based separation refers to a size fractionation procedure where all or part of the fractionated pool can be isolated (e.g.,
20 retained) and/or analyzed. Size fractionation procedures are known in the art (e.g., separation on an array, separation by a molecular sieve, separation by gel electrophoresis, separation by column chromatography (e.g., size-exclusion columns), and microfluidics-based approaches). In some embodiments, length-based separation approaches can include fragment circularization, chemical treatment (e.g., formaldehyde, polyethylene glycol (PEG)), mass spectrometry and/or size-specific
25 nucleic acid amplification, for example. In some embodiments, length-based separation is performed using Solid Phase Reversible Immobilization (SPRI) beads.

In some embodiments, nucleic acid fragments of a certain length, range of lengths, or lengths under or over a particular threshold or cutoff are separated from the sample. In some
embodiments, fragments having a length under a particular threshold or cutoff (e.g., 500 bp, 400
30 bp, 300 bp, 200 bp, 150 bp, 100 bp) are referred to as “short” fragments and fragments having a length over a particular threshold or cutoff (e.g., 500 bp, 600 bp, 700 bp, 800 bp, 900 bp, 1000 bp) are referred to as “long” fragments, large fragments, and/or high molecular weight (HMW) fragments. In some embodiments, fragments of a certain length, range of lengths, or lengths under or over a particular threshold or cutoff are retained for analysis while fragments of a different length

or range of lengths, or lengths over or under the threshold or cutoff are not retained for analysis. In some embodiments, fragments that are less than about 500 bp are retained. In some embodiments, fragments that are less than about 400 bp are retained. In some embodiments, fragments that are less than about 300 bp are retained. In some embodiments, fragments that are less than about 200 bp are retained. In some embodiments, fragments that are less than about 150 bp are retained. For example, fragments that are less than about 190 bp, 180 bp, 170 bp, 160 bp, 150 bp, 140 bp, 130 bp, 120 bp, 110 bp or 100 bp are retained. In some embodiments, fragments that are about 100 bp to about 200 bp are retained. For example, fragments that are about 190 bp, 180 bp, 170 bp, 160 bp, 150 bp, 140 bp, 130 bp, 120 bp or 110 bp are retained. In some embodiments, fragments that are in the range of about 100 bp to about 200 bp are retained. For example, fragments that are in the range of about 110 bp to about 190 bp, 130 bp to about 180 bp, 140 bp to about 170 bp, 140 bp to about 150 bp, 150 bp to about 160 bp, or 145 bp to about 155 bp are retained.

In some embodiments, target nucleic acids (e.g., ssNAs) having fragment lengths of less than about 1000 bp are combined with a plurality or pool of scaffold adapter species, or components of scaffold adapter species, described herein. In some embodiments, target nucleic acids (e.g., ssNAs) having fragment lengths of less than about 500 bp are combined with a plurality or pool of scaffold adapter species, or components of scaffold adapter species, described herein. In some embodiments, target nucleic acids (e.g., ssNAs) having fragment lengths of less than about 400 bp are combined with a plurality or pool of scaffold adapter species, or components of scaffold adapter species, described herein. In some embodiments, target nucleic acids (e.g., ssNAs) having fragment lengths of less than about 300 bp are combined with a plurality or pool of scaffold adapter species, or components of scaffold adapter species, described herein. In some embodiments, target nucleic acids (e.g., ssNAs) having fragment lengths of less than about 200 bp are combined with a plurality or pool of scaffold adapter species, or components of scaffold adapter species, described herein. In some embodiments, target nucleic acids (e.g., ssNAs) having fragment lengths of less than about 100 bp are combined with a plurality or pool of scaffold adapter species, or components of scaffold adapter species, described herein.

In some embodiments, target nucleic acids (e.g., ssNAs) having fragment lengths of about 100 bp or more are combined with a plurality or pool of scaffold adapter species, or components of scaffold adapter species, described herein. In some embodiments, target nucleic acids (e.g., ssNAs) having fragment lengths of about 200 bp or more are combined with a plurality or pool of scaffold adapter species, or components of scaffold adapter species, described herein. In some embodiments, target nucleic acids (e.g., ssNAs) having fragment lengths of about 300 bp or more are combined with a plurality or pool of scaffold adapter species, or components of scaffold adapter species,

described herein. In some embodiments, target nucleic acids (e.g., ssNAs) having fragment lengths of about 400 bp or more are combined with a plurality or pool of scaffold adapter species, or components of scaffold adapter species, described herein. In some embodiments, target nucleic acids (e.g., ssNAs) having fragment lengths of about 500 bp or more are combined with a plurality or pool of scaffold adapter species, or components of scaffold adapter species, described herein. In some embodiments, target nucleic acids (e.g., ssNAs) having fragment lengths of about 1000 bp or more are combined with a plurality or pool of scaffold adapter species, or components of scaffold adapter species, described herein.

In some embodiments, target nucleic acids (e.g., ssNAs) having any fragment length or any combination of fragment lengths are combined with a plurality or pool of scaffold adapter species, or components of scaffold adapter species, described herein. For example, target nucleic acids (e.g., ssNAs) having fragment lengths of less than 500 bp and fragments lengths of 500 bp or more may be combined with a plurality or pool of scaffold adapter species, or components of scaffold adapter species, described herein.

Certain length-based separation methods that can be used with methods described herein employ a selective sequence tagging approach, for example. In such methods, a fragment size species (e.g., short fragments) nucleic acids are selectively tagged in a sample that includes long and short nucleic acids. Such methods typically involve performing a nucleic acid amplification reaction using a set of nested primers which include inner primers and outer primers. In some embodiments, one or both of the inner can be tagged to thereby introduce a tag onto the target amplification product. The outer primers generally do not anneal to the short fragments that carry the (inner) target sequence. The inner primers can anneal to the short fragments and generate an amplification product that carries a tag and the target sequence. Typically, tagging of the long fragments is inhibited through a combination of mechanisms which include, for example, blocked extension of the inner primers by the prior annealing and extension of the outer primers. Enrichment for tagged fragments can be accomplished by any of a variety of methods, including for example, exonuclease digestion of single stranded nucleic acid and amplification of the tagged fragments using amplification primers specific for at least one tag.

Another length-based separation method that can be used with methods described herein involves subjecting a nucleic acid sample to polyethylene glycol (PEG) precipitation. Examples of methods include those described in International Patent Application Publication Nos. WO2007/140417 and WO2010/115016. This method in general entails contacting a nucleic acid sample with PEG in the presence of one or more monovalent salts under conditions sufficient to substantially precipitate

large nucleic acids without substantially precipitating small (e.g., less than 300 nucleotides) nucleic acids.

Another length-based enrichment method that can be used with methods described herein involves circularization by ligation, for example, using circ ligase. Short nucleic acid fragments typically can be circularized with higher efficiency than long fragments. Non-circularized sequences can be separated from circularized sequences, and the enriched short fragments can be used for further analysis.

Nucleic acid library

Methods herein may include preparing a nucleic acid library and/or modifying nucleic acids for a nucleic acid library. In some embodiments, ends of nucleic acid fragments are modified such that the fragments, or amplified products thereof, may be incorporated into a nucleic acid library.

Generally, a nucleic acid library refers to a plurality of polynucleotide molecules (e.g., a sample of nucleic acids) that are prepared, assembled and/or modified for a specific process, non-limiting examples of which include immobilization on a solid phase (e.g., a solid support, a flow cell, a bead), enrichment, amplification, cloning, detection and/or for nucleic acid sequencing. In certain embodiments, a nucleic acid library is prepared prior to or during a sequencing process. A nucleic acid library (e.g., sequencing library) can be prepared by a suitable method as known in the art. A nucleic acid library can be prepared by a targeted or a non-targeted preparation process.

In some embodiments, a library of nucleic acids is modified to comprise a chemical moiety (e.g., a functional group) configured for immobilization of nucleic acids to a solid support. In some embodiments a library of nucleic acids is modified to comprise a biomolecule (e.g., a functional group) and/or member of a binding pair configured for immobilization of the library to a solid support, non-limiting examples of which include thyroxine-binding globulin, steroid-binding proteins, antibodies, antigens, haptens, enzymes, lectins, nucleic acids, repressors, protein A, protein G, avidin, streptavidin, biotin, complement component C1q, nucleic acid-binding proteins, receptors, carbohydrates, oligonucleotides, polynucleotides, complementary nucleic acid sequences, the like and combinations thereof. Some examples of specific binding pairs include, without limitation: an avidin moiety and a biotin moiety; an antigenic epitope and an antibody or immunologically reactive fragment thereof; an antibody and a hapten; a digoxigenin moiety and an anti-digoxigenin antibody; a fluorescein moiety and an anti-fluorescein antibody; an operator and a repressor; a nuclease and a nucleotide; a lectin and a polysaccharide; a steroid and a steroid-binding protein; an active compound and an active compound receptor; a hormone and a hormone receptor; an enzyme and a substrate; an immunoglobulin and protein A; an oligonucleotide or polynucleotide and its corresponding complement; the like or combinations thereof.

In some embodiments, a library of nucleic acids is modified to comprise one or more polynucleotides of known composition, non-limiting examples of which include an identifier (e.g., a tag, an indexing tag), a capture sequence, a label, an adapter, a restriction enzyme site, a promoter, an enhancer, an origin of replication, a stem loop, a complimentary sequence (e.g., a primer binding site, an annealing site), a suitable integration site (e.g., a transposon, a viral integration site), a modified nucleotide, a unique molecular identifier (UMI) described herein, a palindromic sequence described herein, the like or combinations thereof. Polynucleotides of known sequence can be added at a suitable position, for example on the 5' end, 3' end or within a nucleic acid sequence. Polynucleotides of known sequence can be the same or different sequences. In some embodiments, a polynucleotide of known sequence is configured to hybridize to one or more oligonucleotides immobilized on a surface (e.g., a surface in flow cell). For example, a nucleic acid molecule comprising a 5' known sequence may hybridize to a first plurality of oligonucleotides while the 3' known sequence may hybridize to a second plurality of oligonucleotides. In some embodiments, a library of nucleic acid can comprise chromosome-specific tags, capture sequences, labels and/or adapters (e.g., oligonucleotide adapters described herein). In some embodiments, a library of nucleic acids comprises one or more detectable labels. In some embodiments one or more detectable labels may be incorporated into a nucleic acid library at a 5' end, at a 3' end, and/or at any nucleotide position within a nucleic acid in the library. In some embodiments, a library of nucleic acids comprises hybridized oligonucleotides. In certain embodiments hybridized oligonucleotides are labeled probes. In some embodiments, a library of nucleic acids comprises hybridized oligonucleotide probes prior to immobilization on a solid phase.

In some embodiments, a polynucleotide of known sequence comprises a universal sequence. A universal sequence is a specific nucleotide sequence that is integrated into two or more nucleic acid molecules or two or more subsets of nucleic acid molecules where the universal sequence is the same for all molecules or subsets of molecules that it is integrated into. A universal sequence is often designed to hybridize to and/or amplify a plurality of different sequences using a single universal primer that is complementary to a universal sequence. In some embodiments two (e.g., a pair) or more universal sequences and/or universal primers are used. A universal primer often comprises a universal sequence. In some embodiments adapters (e.g., universal adapters) comprise universal sequences. In some embodiments one or more universal sequences are used to capture, identify and/or detect multiple species or subsets of nucleic acids.

In certain embodiments of preparing a nucleic acid library, (e.g., in certain sequencing by synthesis procedures), nucleic acids are size selected and/or fragmented into lengths of several hundred base pairs, or less (e.g., in preparation for library generation). In some embodiments, library preparation is performed without fragmentation (e.g., when using cell-free DNA).

In certain embodiments, a ligation-based library preparation method is used (e.g., ILLUMINA TRUSEQ, Illumina, San Diego CA). Ligation-based library preparation methods often make use of an adapter (e.g., a methylated adapter) design which can incorporate an index sequence (e.g., a sample index sequence to identify sample origin for a nucleic acid sequence) at the initial ligation
5 step and often can be used to prepare samples for single-read sequencing, paired-end sequencing and multiplexed sequencing. For example, nucleic acids (e.g., fragmented nucleic acids or cell-free DNA) may be end repaired by a fill-in reaction, an exonuclease reaction or a combination thereof. In some embodiments, the resulting blunt end repaired nucleic acid can then be extended by a single nucleotide, which is complementary to a single nucleotide overhang on the 3' end of an
10 adapter/primer. Any nucleotide can be used for the extension/overhang nucleotides. In some embodiments, end repair is omitted and scaffold adapters (e.g., scaffold adapters described herein) are ligated directly to the native ends of nucleic acids (e.g., single-stranded nucleic acids, fragmented nucleic acids, and/or cell-free DNA).

In some embodiments, nucleic acid library preparation comprises ligating a scaffold adapter, or
15 component thereof, (e.g., to a sample nucleic acid, to a sample nucleic acid fragment, to a template nucleic acid, to a target nucleic acid, to an ssNA), such as a scaffold adapter described herein. Scaffold adapters, or components thereof, may comprise sequences complementary to flow-cell anchors, and sometimes are utilized to immobilize a nucleic acid library to a solid support, such as the inside surface of a flow cell, for example. In some embodiments, a scaffold adapter, or
20 component thereof, comprises an identifier, one or more sequencing primer hybridization sites (e.g., sequences complementary to universal sequencing primers, single end sequencing primers, paired end sequencing primers, multiplexed sequencing primers, and the like), or combinations thereof (e.g., adapter/sequencing, adapter/identifier, adapter/identifier/sequencing). In some
25 embodiments, a scaffold adapter, or component thereof, comprises one or more of primer annealing polynucleotide, also referred to herein as priming sequence or primer binding domain, (e.g., for annealing to flow cell attached oligonucleotides and/or to free amplification primers), an index polynucleotide (e.g., sample index sequence for tracking nucleic acid from different samples; also referred to as a sample ID), a barcode polynucleotide (e.g., single molecule barcode (SMB) for
30 tracking individual molecules of sample nucleic acid that are amplified prior to sequencing; also referred to as a molecular barcode or a unique molecular identifier (UMI)). In some embodiments, a primer annealing component (or priming sequence or primer binding domain) of a scaffold adapter, or component thereof, comprises one or more universal sequences (e.g., sequences complementary to one or more universal amplification primers). In some embodiments, an index polynucleotide (e.g., sample index; sample ID) is a component of a scaffold adapter, or component

thereof. In some embodiments, an index polynucleotide (e.g., sample index; sample ID) is a component of a universal amplification primer sequence.

In some embodiments, scaffold adapters, or components thereof, when used in combination with amplification primers (e.g., universal amplification primers) are designed generate library constructs comprising one or more of: universal sequences, molecular barcodes (UMIs), UMI flanking sequence, sample ID sequences, spacer sequences, and a sample nucleic acid sequence (e.g., ssNA sequence). In some embodiments, scaffold adapters, or components thereof, when used in combination with universal amplification primers are designed to generate library constructs comprising an ordered combination of one or more of: universal sequences, molecular barcodes (UMIs), sample ID sequences, spacer sequences, and a sample nucleic acid sequence (e.g., ssNA sequence). For example, a library construct may comprise a first universal sequence, followed by a second universal sequence, followed by first molecular barcode (UMI), followed by a spacer sequence, followed by a template sequence (e.g., sample nucleic acid sequence; ssNA sequence), followed by a spacer sequence, followed by a second molecular barcode (UMI), followed by a third universal sequence, followed by a sample ID, followed by a fourth universal sequence. In some embodiments, scaffold adapters, or components thereof, when used in combination with amplification primers (e.g., universal amplification primers) are designed generate library constructs for each strand of a template molecule (e.g., sample nucleic acid molecule; ssNA molecule). In some embodiments, scaffold adapters are duplex adapters.

An identifier can be a suitable detectable label incorporated into or attached to a nucleic acid (e.g., a polynucleotide) that allows detection and/or identification of nucleic acids that comprise the identifier. In some embodiments, an identifier is incorporated into or attached to a nucleic acid during a sequencing method (e.g., by a polymerase). In some embodiments, an identifier is incorporated into or attached to a nucleic acid prior to a sequencing method (e.g., by an extension reaction, by an amplification reaction, by a ligation reaction). Non-limiting examples of identifiers include nucleic acid tags, nucleic acid indexes or barcodes, a radiolabel (e.g., an isotope), metallic label, a fluorescent label, a chemiluminescent label, a phosphorescent label, a fluorophore quencher, a dye, a protein (e.g., an enzyme, an antibody or part thereof, a linker, a member of a binding pair), the like or combinations thereof. In some embodiments, an identifier (e.g., a nucleic acid index or barcode) is a unique, known and/or identifiable sequence of nucleotides or nucleotide analogues. In some embodiments, identifiers are six or more contiguous nucleotides. A multitude of fluorophores are available with a variety of different excitation and emission spectra. Any suitable type and/or number of fluorophores can be used as an identifier. In some embodiments 1 or more, 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, 10 or more, 20 or more, 30 or more or 50 or more different identifiers are utilized in a method described herein

(e.g., a nucleic acid detection and/or sequencing method). In some embodiments, one or two types of identifiers (e.g., fluorescent labels) are linked to each nucleic acid in a library. Detection and/or quantification of an identifier can be performed by a suitable method, apparatus or machine, non-limiting examples of which include flow cytometry, quantitative polymerase chain reaction (qPCR),
5 gel electrophoresis, a luminometer, a fluorometer, a spectrophotometer, a suitable gene-chip or microarray analysis, Western blot, mass spectrometry, chromatography, cytofluorimetric analysis, fluorescence microscopy, a suitable fluorescence or digital imaging method, confocal laser scanning microscopy, laser scanning cytometry, affinity chromatography, manual batch mode separation, electric field suspension, a suitable nucleic acid sequencing method and/or nucleic acid
10 sequencing apparatus, the like and combinations thereof.

In some embodiments, an identifier, a sequencing-specific index/barcode, and a sequencer-specific flow-cell binding primer sites are incorporated into a nucleic acid library by single-primer extension (e.g., by a strand displacing polymerase).

In some embodiments, a nucleic acid library or parts thereof are amplified (e.g., amplified by a
15 PCR-based method) under amplification conditions. In some embodiments, a sequencing method comprises amplification of a nucleic acid library. A nucleic acid library can be amplified prior to or after immobilization on a solid support (e.g., a solid support in a flow cell). Nucleic acid amplification includes the process of amplifying or increasing the numbers of a nucleic acid template and/or of a complement thereof that are present (e.g., in a nucleic acid library), by producing one or more
20 copies of the template and/or its complement. Amplification can be carried out by a suitable method. A nucleic acid library can be amplified by a thermocycling method or by an isothermal amplification method. In some embodiments, a rolling circle amplification method is used. In some embodiments, amplification takes place on a solid support (e.g., within a flow cell) where a nucleic acid library or portion thereof is immobilized. In certain sequencing methods, a nucleic acid library
25 is added to a flow cell and immobilized by hybridization to anchors under suitable conditions. This type of nucleic acid amplification is often referred to as solid phase amplification. In some embodiments of solid phase amplification, all or a portion of the amplified products are synthesized by an extension initiating from an immobilized primer. Solid phase amplification reactions are analogous to standard solution phase amplifications except that at least one of the amplification
30 oligonucleotides (e.g., primers) is immobilized on a solid support. In some embodiments, modified nucleic acid (e.g., nucleic acid modified by addition of adapters) is amplified.

In some embodiments, solid phase amplification comprises a nucleic acid amplification reaction comprising only one species of oligonucleotide primer immobilized to a surface. In certain embodiments, solid phase amplification comprises a plurality of different immobilized

oligonucleotide primer species. In some embodiments, solid phase amplification may comprise a nucleic acid amplification reaction comprising one species of oligonucleotide primer immobilized on a solid surface and a second different oligonucleotide primer species in solution. Multiple different species of immobilized or solution-based primers can be used. Non-limiting examples of solid
5 phase nucleic acid amplification reactions include interfacial amplification, bridge amplification, emulsion PCR, WildFire amplification (e.g., U.S. Patent Application Publication No. 2013/0012399), the like or combinations thereof.

In some embodiments, nucleic acids are differentially amplified. Differentially amplified generally refers to amplifying a first nucleic acid species to a greater degree than a second nucleic acid
10 species. For example, a first nucleic acid species may be amplified at least about 2x, 3x, 4x, 5x, 6x, 7x, 8x, 9x, 10x, or more compared to the amplification of a second nucleic acid. In some embodiments, a first nucleic acid species is exponentially amplified and a second nucleic acid species is linearly amplified. A nucleic acid species may refer to a source or origin of the nucleic acid. For example, a source may be RNA (e.g., single-stranded RNA) or DNA (e.g., double-
15 stranded DNA). In some embodiments, a first species or source is RNA. In some embodiments, a first species or source is DNA. In some embodiments, a second species or source is RNA. In some embodiments, a second species or source is DNA. In some embodiments, a first species or source is RNA, and a second species or source is DNA. Accordingly, in some embodiments, a method herein can differentially amplify nucleic acid originating from an RNA source and nucleic acid
20 originating from a DNA source, where the nucleic acid originating from the RNA source is amplified to a greater degree compared to the nucleic acid originating from the DNA source. In some embodiments, nucleic acids in a library produced by a method described herein are differentially amplified. In some embodiments, nucleic acids in a library produced by a method shown in Fig. 1, Fig. 2, or Fig.3 are differentially amplified. In some embodiments, a nucleic acid library comprises
25 nucleic acid originating from an RNA source and nucleic acid originating from a DNA source, where both types of nucleic acid molecules comprise a common priming site at one end and a different priming site at the other end. For example, both types of nucleic acid molecules may have priming site A at one end, nucleic acid originating from the RNA source may have priming site B at the opposite end, and nucleic acid originating from the DNA source may have priming site C at the
30 opposite end. An amplification reaction that includes primers binding to A and B, and excludes a primer binding to C, will result in exponential amplification of nucleic acid originating from the RNA source and linear amplification of nucleic acid originating from the DNA source.

Nucleic acid sequencing

In some embodiments, nucleic acid (e.g., nucleic acid fragments, sample nucleic acid, cell-free nucleic acid, single-stranded nucleic acid, single-stranded DNA, single-stranded RNA) is sequenced. In some embodiments, ssNA hybridized to scaffold adapters provided herein
5 (“hybridization products”) are sequenced by a sequencing process. In some embodiments, ssNA ligated to oligonucleotide components provided herein (“single-stranded ligation products”) are sequenced by a sequencing process. In some embodiments, hybridization products and/or single-stranded ligation products are amplified by an amplification process, and the amplification products are sequenced by a sequencing process. In some embodiments, hybridization products and/or
10 single-stranded ligation products are not amplified by an amplification process, and the hybridization products and/or single-stranded ligation products are sequenced without prior amplification by a sequencing process. In some embodiments, the sequencing process generates sequence reads (or sequencing reads). In some embodiments, a method herein comprises determining the sequence of a single-stranded nucleic acid molecule based on the sequence
15 reads.

For certain sequencing platforms (e.g., paired-end sequencing), generating sequence reads may include generating forward sequence reads and generating reverse sequence reads. For example, sequencing using certain paired-end sequencing platforms sequence each nucleic acid fragment from both directions, generally resulting in two reads per nucleic acid fragment, with the first read in
20 a forward orientation (forward read) and the second read in reverse-complement orientation (reverse read). For certain platforms, a forward read is generated off a particular primer within a sequencing adapter (e.g., ILLUMINA adapter, P5 primer), and a reverse read is generated off a different primer within a sequencing adapter (e.g., ILLUMINA adapter, P7 primer).

Nucleic acid may be sequenced using any suitable sequencing platform including a Sanger
25 sequencing platform, a high throughput or massively parallel sequencing (next generation sequencing (NGS)) platform, or the like, such as, for example, a sequencing platform provided by Illumina® (e.g., HiSeq™, MiSeq™ and/or Genome Analyzer™ sequencing systems); Oxford Nanopore™ Technologies (e.g., MinION sequencing system), Ion Torrent™ (e.g., Ion PGM™ and/or Ion Proton™ sequencing systems); Pacific Biosciences (e.g., PACBIO RS II sequencing
30 system); Life Technologies™ (e.g., SOLiD sequencing system); Roche (e.g., 454 GS FLX+ and/or GS Junior sequencing systems); or any other suitable sequencing platform. In some embodiments, the sequencing process is a highly multiplexed sequencing process. In certain instances, a full or substantially full sequence is obtained and sometimes a partial sequence is obtained. Nucleic acid sequencing generally produces a collection of sequence reads. As used herein, “reads” (e.g., “a

read,” “a sequence read”) are short sequences of nucleotides produced by any sequencing process described herein or known in the art. Reads can be generated from one end of nucleic acid fragments (single-end reads), and sometimes are generated from both ends of nucleic acid fragments (e.g., paired-end reads, double-end reads). In some embodiments, a sequencing
5 process generates short sequencing reads or “short reads.” In some embodiments, the nominal, average, mean or absolute length of short reads sometimes is about 10 continuous nucleotides to about 250 or more contiguous nucleotides. In some embodiments, the nominal, average, mean or absolute length of short reads sometimes is about 50 continuous nucleotides to about 150 or more contiguous nucleotides.

10 The length of a sequence read is often associated with the particular sequencing technology utilized. High-throughput methods, for example, provide sequence reads that can vary in size from tens to hundreds of base pairs (bp). Nanopore sequencing, for example, can provide sequence reads that can vary in size from tens to hundreds to thousands of base pairs. In some
15 embodiments, sequence reads are of a mean, median, average or absolute length of about 15 bp to about 900 bp long. In certain embodiments sequence reads are of a mean, median, average or absolute length of about 1000 bp or more. In some embodiments sequence reads are of a mean, median, average or absolute length of about 1500, 2000, 2500, 3000, 3500, 4000, 4500, or 5000 bp or more. In some embodiments, sequence reads are of a mean, median, average or absolute length of about 100 bp to about 200 bp.

20 In some embodiments, the nominal, average, mean or absolute length of single-end reads sometimes is about 10 continuous nucleotides to about 250 or more contiguous nucleotides, about 15 contiguous nucleotides to about 200 or more contiguous nucleotides, about 15 contiguous nucleotides to about 150 or more contiguous nucleotides, about 15 contiguous nucleotides to about 125 or more contiguous nucleotides, about 15 contiguous nucleotides to about 100 or more
25 contiguous nucleotides, about 15 contiguous nucleotides to about 75 or more contiguous nucleotides, about 15 contiguous nucleotides to about 60 or more contiguous nucleotides, 15 contiguous nucleotides to about 50 or more contiguous nucleotides, about 15 contiguous nucleotides to about 40 or more contiguous nucleotides, and sometimes about 15 contiguous nucleotides or about 36 or more contiguous nucleotides. In certain embodiments the nominal,
30 average, mean or absolute length of single-end reads is about 20 to about 30 bases, or about 24 to about 28 bases in length. In certain embodiments the nominal, average, mean or absolute length of single-end reads is about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 21, 22, 23, 24, 25, 26, 27, 28 or about 29 bases or more in length. In certain embodiments the nominal,
35 average, mean or absolute length of single-end reads is about 20 to about 200 bases, about 100 to about 200 bases, or about 140 to about 160 bases in length. In certain embodiments the nominal,

average, mean or absolute length of single-end reads is about 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, or about 200 bases or more in length. In certain embodiments, the nominal, average, mean or absolute length of paired-end reads sometimes is about 10 contiguous nucleotides to about 25 contiguous nucleotides or more (e.g., about 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 or 25 nucleotides in length or more), about 15 contiguous nucleotides to about 20 contiguous nucleotides or more, and sometimes is about 17 contiguous nucleotides or about 18 contiguous nucleotides. In certain embodiments, the nominal, average, mean or absolute length of paired-end reads sometimes is about 25 contiguous nucleotides to about 400 contiguous nucleotides or more (e.g., about 25, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, or 400 nucleotides in length or more), about 50 contiguous nucleotides to about 350 contiguous nucleotides or more, about 100 contiguous nucleotides to about 325 contiguous nucleotides, about 150 contiguous nucleotides to about 325 contiguous nucleotides, about 200 contiguous nucleotides to about 325 contiguous nucleotides, about 275 contiguous nucleotides to about 310 contiguous nucleotides, about 100 contiguous nucleotides to about 200 contiguous nucleotides, about 100 contiguous nucleotides to about 175 contiguous nucleotides, about 125 contiguous nucleotides to about 175 contiguous nucleotides, and sometimes is about 140 contiguous nucleotides to about 160 contiguous nucleotides. In certain embodiments, the nominal, average, mean, or absolute length of paired-end reads is about 150 contiguous nucleotides, and sometimes is 150 contiguous nucleotides.

Reads generally are representations of nucleotide sequences in a physical nucleic acid. For example, in a read containing an ATGC depiction of a sequence, "A" represents an adenine nucleotide, "T" represents a thymine nucleotide, "G" represents a guanine nucleotide and "C" represents a cytosine nucleotide, in a physical nucleic acid. Sequence reads obtained from a sample from a subject can be reads from a mixture of a minority nucleic acid and a majority nucleic acid. For example, sequence reads obtained from the blood of a cancer patient can be reads from a mixture of cancer nucleic acid and non-cancer nucleic acid. In another example, sequence reads obtained from the blood of a pregnant female can be reads from a mixture of fetal nucleic acid and maternal nucleic acid. In another example, sequence reads obtained from the blood of a patient having an infection or infectious disease can be reads from a mixture of host nucleic acid and pathogen nucleic acid. In another example, sequence reads obtained from the blood of a transplant recipient can be reads from a mixture of host nucleic acid and transplant nucleic acid. In another example, sequence reads obtained from a sample can be reads from a mixture of nucleic acid from microorganisms collectively comprising a microbiome (e.g., microbiome of gut, microbiome of blood, microbiome of mouth, microbiome of spinal fluid, microbiome of feces) in a subject. In

another example, sequence reads obtained from a sample can be reads from a mixture of nucleic acid from microorganisms collectively comprising a microbiome (e.g., microbiome of gut, microbiome of blood, microbiome of mouth, microbiome of spinal fluid, microbiome of feces), and nucleic acid from the host subject. A mixture of relatively short reads can be transformed by processes described herein into a representation of genomic nucleic acid present in the subject, and/or a representation of genomic nucleic acid present in a tumor, a fetus, a pathogen, a transplant, or a microbiome.

In certain embodiments, “obtaining” nucleic acid sequence reads of a sample from a subject and/or “obtaining” nucleic acid sequence reads of a biological specimen from one or more reference persons can involve directly sequencing nucleic acid to obtain the sequence information. In some embodiments, “obtaining” can involve receiving sequence information obtained directly from a nucleic acid by another.

In some embodiments, some or all nucleic acids in a sample are enriched and/or amplified (e.g., non-specifically, e.g., by a PCR based method) prior to or during sequencing. In certain embodiments, specific nucleic acid species or subsets in a sample are enriched and/or amplified prior to or during sequencing. In some embodiments, a species or subset of a pre-selected pool of nucleic acids is sequenced randomly. In some embodiments, nucleic acids in a sample are not enriched and/or amplified prior to or during sequencing.

In some embodiments, a representative fraction of a genome is sequenced and is sometimes referred to as “coverage” or “fold coverage.” For example, a 1-fold coverage indicates that roughly 100% of the nucleotide sequences of the genome are represented by reads. In some instances, fold coverage is referred to as (and is directly proportional to) “sequencing depth.” In some embodiments, “fold coverage” is a relative term referring to a prior sequencing run as a reference. For example, a second sequencing run may have 2-fold less coverage than a first sequencing run.

In some embodiments, a genome is sequenced with redundancy, where a given region of the genome can be covered by two or more reads or overlapping reads (e.g., a “fold coverage” greater than 1, e.g., a 2-fold coverage). In some embodiments, a genome (e.g., a whole genome) is sequenced with about 0.01-fold to about 100-fold coverage, about 0.1-fold to 20-fold coverage, or about 0.1-fold to about 1-fold coverage (e.g., about 0.015-, 0.02-, 0.03-, 0.04-, 0.05-, 0.06-, 0.07-, 0.08-, 0.09-, 0.1-, 0.2-, 0.3-, 0.4-, 0.5-, 0.6-, 0.7-, 0.8-, 0.9-, 1-, 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9-, 10-, 15-, 20-, 30-, 40-, 50-, 60-, 70-, 80-, 90-fold or greater coverage). In some embodiments, specific parts of a genome (e.g., genomic parts from targeted methods) are sequenced and fold coverage values generally refer to the fraction of the specific genomic parts sequenced (i.e., fold coverage values do not refer to the whole genome). In some instances, specific genomic parts are sequenced at 1000-

fold coverage or more. For example, specific genomic parts may be sequenced at 2000-fold, 5,000-fold, 10,000-fold, 20,000-fold, 30,000-fold, 40,000-fold or 50,000-fold coverage. In some embodiments, sequencing is at about 1,000-fold to about 100,000-fold coverage. In some
5 embodiments, sequencing is at about 10,000-fold to about 70,000-fold coverage. In some
embodiments, sequencing is at about 20,000-fold to about 60,000-fold coverage. In some
embodiments, sequencing is at about 30,000-fold to about 50,000-fold coverage.

In some embodiments, one nucleic acid sample from one individual is sequenced. In certain
embodiments, nucleic acids from each of two or more samples are sequenced, where samples are
from one individual or from different individuals. In certain embodiments, nucleic acid samples from
10 two or more biological samples are pooled, where each biological sample is from one individual or
two or more individuals, and the pool is sequenced. In the latter embodiments, a nucleic acid
sample from each biological sample often is identified by one or more unique identifiers.

In some embodiments, a sequencing method utilizes identifiers that allow multiplexing of sequence
reactions in a sequencing process. The greater the number of unique identifiers, the greater the
15 number of samples and/or chromosomes for detection, for example, that can be multiplexed in a
sequencing process. A sequencing process can be performed using any suitable number of unique
identifiers (e.g., 4, 8, 12, 24, 48, 96, or more).

A sequencing process sometimes makes use of a solid phase, and sometimes the solid phase
comprises a flow cell on which nucleic acid from a library can be attached and reagents can be
20 flowed and contacted with the attached nucleic acid. A flow cell sometimes includes flow cell lanes,
and use of identifiers can facilitate analyzing a number of samples in each lane. A flow cell often is
a solid support that can be configured to retain and/or allow the orderly passage of reagent
solutions over bound analytes. Flow cells frequently are planar in shape, optically transparent,
generally in the millimeter or sub-millimeter scale, and often have channels or lanes in which the
25 analyte/reagent interaction occurs. In some embodiments, the number of samples analyzed in a
given flow cell lane is dependent on the number of unique identifiers utilized during library
preparation and/or probe design. Multiplexing using 12 identifiers, for example, allows simultaneous
analysis of 96 samples (e.g., equal to the number of wells in a 96 well microwell plate) in an 8-lane
flow cell. Similarly, multiplexing using 48 identifiers, for example, allows simultaneous analysis of
30 384 samples (e.g., equal to the number of wells in a 384 well microwell plate) in an 8-lane flow cell.
Non-limiting examples of commercially available multiplex sequencing kits include Illumina's
multiplexing sample preparation oligonucleotide kit and multiplexing sequencing primers and PhiX
control kit (e.g., Illumina's catalog numbers PE-400-1001 and PE-400-1002, respectively).

Any suitable method of sequencing nucleic acids can be used, non-limiting examples of which include Maxim & Gilbert, chain-termination methods, sequencing by synthesis, sequencing by ligation, sequencing by mass spectrometry, microscopy-based techniques, the like or combinations thereof. In some embodiments, a first-generation technology, such as, for example, Sanger sequencing methods including automated Sanger sequencing methods, including microfluidic Sanger sequencing, can be used in a method provided herein. In some embodiments, sequencing technologies that include the use of nucleic acid imaging technologies (e.g., transmission electron microscopy (TEM) and atomic force microscopy (AFM)), can be used. In some embodiments, a high-throughput sequencing method is used. High-throughput sequencing methods generally involve clonally amplified DNA templates or single DNA molecules that are sequenced in a massively parallel fashion, sometimes within a flow cell. Next generation (e.g., 2nd and 3rd generation) sequencing techniques capable of sequencing DNA in a massively parallel fashion can be used for methods described herein and are collectively referred to herein as “massively parallel sequencing” (MPS). In some embodiments, MPS sequencing methods utilize a targeted approach, where specific chromosomes, genes or regions of interest are sequenced. In certain embodiments, a non-targeted approach is used where most or all nucleic acids in a sample are sequenced, amplified and/or captured randomly.

In some embodiments a targeted enrichment, amplification and/or sequencing approach is used. A targeted approach often isolates, selects and/or enriches a subset of nucleic acids in a sample for further processing by use of sequence-specific oligonucleotides. In some embodiments, a library of sequence-specific oligonucleotides is utilized to target (e.g., hybridize to) one or more sets of nucleic acids in a sample. Sequence-specific oligonucleotides and/or primers are often selective for particular sequences (e.g., unique nucleic acid sequences) present in one or more chromosomes, genes, exons, introns, and/or regulatory regions of interest. Any suitable method or combination of methods can be used for enrichment, amplification and/or sequencing of one or more subsets of targeted nucleic acids. In some embodiments targeted sequences are isolated and/or enriched by capture to a solid phase (e.g., a flow cell, a bead) using one or more sequence-specific anchors. In some embodiments targeted sequences are enriched and/or amplified by a polymerase-based method (e.g., a PCR-based method, by any suitable polymerase-based extension) using sequence-specific primers and/or primer sets. Sequence specific anchors often can be used as sequence-specific primers.

MPS sequencing sometimes makes use of sequencing by synthesis and certain imaging processes. A nucleic acid sequencing technology that may be used in a method described herein is sequencing-by-synthesis and reversible terminator-based sequencing (e.g., Illumina’s Genome Analyzer; Genome Analyzer II; HISEQ 2000; HISEQ 2500 (Illumina, San Diego CA)). With this

technology, millions of nucleic acid (e.g., DNA) fragments can be sequenced in parallel. In one example of this type of sequencing technology, a flow cell is used which contains an optically transparent slide with 8 individual lanes on the surfaces of which are bound oligonucleotide anchors (e.g., adapter primers).

5 Sequencing by synthesis generally is performed by iteratively adding (e.g., by covalent addition) a nucleotide to a primer or preexisting nucleic acid strand in a template directed manner. Each iterative addition of a nucleotide is detected and the process is repeated multiple times until a sequence of a nucleic acid strand is obtained. The length of a sequence obtained depends, in part, on the number of addition and detection steps that are performed. In some embodiments of
10 sequencing by synthesis, one, two, three or more nucleotides of the same type (e.g., A, G, C or T) are added and detected in a round of nucleotide addition. Nucleotides can be added by any suitable method (e.g., enzymatically or chemically). For example, in some embodiments a polymerase or a ligase adds a nucleotide to a primer or to a preexisting nucleic acid strand in a template directed manner. In some embodiments of sequencing by synthesis, different types of
15 nucleotides, nucleotide analogues and/or identifiers are used. In some embodiments, reversible terminators and/or removable (e.g., cleavable) identifiers are used. In some embodiments, fluorescent labeled nucleotides and/or nucleotide analogues are used. In certain embodiments sequencing by synthesis comprises a cleavage (e.g., cleavage and removal of an identifier) and/or a washing step. In some embodiments the addition of one or more nucleotides is detected by a
20 suitable method described herein or known in the art, non-limiting examples of which include any suitable imaging apparatus, a suitable camera, a digital camera, a CCD (Charge Couple Device) based imaging apparatus (e.g., a CCD camera), a CMOS (Complementary Metal Oxide Silicon) based imaging apparatus (e.g., a CMOS camera), a photo diode (e.g., a photomultiplier tube), electron microscopy, a field-effect transistor (e.g., a DNA field-effect transistor), an ISFET ion
25 sensor (e.g., a CHEMFET sensor), the like or combinations thereof.

Any suitable MPS method, system or technology platform for conducting methods described herein can be used to obtain nucleic acid sequence reads. Non-limiting examples of MPS platforms include ILLUMINA/SOLEX/HISEQ (e.g., Illumina's Genome Analyzer; Genome Analyzer II; HISEQ 2000; HISEQ), SOLiD, Roche/454, PACBIO and/or SMRT, Helicos True Single Molecule
30 Sequencing, Ion Torrent and Ion semiconductor-based sequencing (e.g., as developed by Life Technologies), WildFire, 5500, 5500xl W and/or 5500xl W Genetic Analyzer based technologies (e.g., as developed and sold by Life Technologies, U.S. Patent Application Publication No. 2013/0012399); Polony sequencing, Pyrosequencing, Massively Parallel Signature Sequencing (MPSS), RNA polymerase (RNAP) sequencing, LaserGen systems and methods, Nanopore-based
35 platforms, chemical-sensitive field effect transistor (CHEMFET) array, electron microscopy-based

sequencing (e.g., as developed by ZS Genetics, Halcyon Molecular), nanoball sequencing, the like or combinations thereof. Other sequencing methods that may be used to conduct methods herein include digital PCR, sequencing by hybridization, nanopore sequencing, chromosome-specific sequencing (e.g., using DANSR (digital analysis of selected regions) technology).

5 In some embodiments, nucleic acid is sequenced and the sequencing product (e.g., a collection of sequence reads) is processed prior to, or in conjunction with, an analysis of the sequenced nucleic acid. For example, sequence reads may be processed according to one or more of the following: aligning, mapping, filtering, counting, normalizing, weighting, generating a profile, and the like, and combinations thereof. Certain processing steps may be performed in any order and certain
10 processing steps may be repeated.

Methods of the present disclosure can be used to reduce sequencing error rates. In some embodiments, prior to an initial denaturing, double-stranded molecules can be labeled with a barcode such that, after subsequent denaturing, single-stranded library preparation, and sequencing, sequences from nucleic acid molecules that were originally paired together can be
15 associated. In some embodiments, after initial ligation of scaffold adapters, a pool of index primers is used to conduct index PCR such that copies are generated of both original sample nucleic acid molecules and nucleic acids from initial PCR first strand synthesis that both comprise the same barcode or UMI (or the complement thereof). By these or other means of associating strands that were originally hybridized (and therefore have complementary sequences), sequencing read
20 information for both strands can be compared and used to reduce the sequencing error rate.

Mapping reads

Sequence reads can be mapped and the number of reads mapping to a specified nucleic acid region (e.g., a chromosome or portion thereof) are referred to as counts. Any suitable mapping method (e.g., process, algorithm, program, software, module, the like or combination thereof) can
25 be used. Certain aspects of mapping processes are described hereafter.

Mapping nucleotide sequence reads (i.e., sequence information from a fragment whose physical genomic position is unknown) can be performed in a number of ways, and often comprises alignment of the obtained sequence reads with a matching sequence in a reference genome. In such alignments, sequence reads generally are aligned to a reference sequence and those that
30 align are designated as being "mapped," as "a mapped sequence read" or as "a mapped read." In certain embodiments, a mapped sequence read is referred to as a "hit" or "count." In some embodiments, mapped sequence reads are grouped together according to various parameters and assigned to particular genomic portions, which are discussed in further detail below.

The terms “aligned,” “alignment,” or “aligning” generally refer to two or more nucleic acid sequences that can be identified as a match (e.g., 100% identity) or partial match. Alignments can be done manually or by a computer (e.g., a software, program, module, or algorithm), non-limiting examples of which include the Efficient Local Alignment of Nucleotide Data (ELAND) computer program distributed as part of the ILLUMINA Genomics Analysis pipeline. Alignment of a sequence read can be a 100% sequence match. In some instances, an alignment is less than a 100% sequence match (i.e., non-perfect match, partial match, partial alignment). In some embodiments an alignment is about a 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, 89%, 88%, 87%, 86%, 85%, 84%, 83%, 82%, 81%, 80%, 79%, 78%, 77%, 76% or 75% match. In some 10 embodiments, an alignment comprises a mismatch. In some embodiments, an alignment comprises 1, 2, 3, 4 or 5 mismatches. Two or more sequences can be aligned using either strand (e.g., sense or antisense strand). In certain embodiments a nucleic acid sequence is aligned with the reverse complement of another nucleic acid sequence.

Various computational methods can be used to map each sequence read to a portion. Non-limiting 15 examples of computer algorithms that can be used to align sequences include, without limitation, BLAST, BLITZ, FASTA, BOWTIE 1, BOWTIE 2, ELAND, MAQ, PROBEMATCH, SOAP, BWA or SEQMAP, or variations thereof or combinations thereof. In some embodiments, sequence reads can be aligned with sequences in a reference genome. In some embodiments, sequence reads can be found and/or aligned with sequences in nucleic acid databases known in the art including, for 20 example, GenBank, dbEST, dbSTS, EMBL (European Molecular Biology Laboratory) and DDBJ (DNA Databank of Japan). BLAST or similar tools can be used to search identified sequences against a sequence database. Search hits can then be used to sort the identified sequences into appropriate portions (described hereafter), for example.

In some embodiments, a read may uniquely or non-uniquely map to portions in a reference 25 genome. A read is considered as “uniquely mapped” if it aligns with a single sequence in the reference genome. A read is considered as “non-uniquely mapped” if it aligns with two or more sequences in the reference genome. In some embodiments, non-uniquely mapped reads are eliminated from further analysis (e.g. quantification). A certain, small degree of mismatch (0-1) may be allowed to account for single nucleotide polymorphisms that may exist between the reference 30 genome and the reads from individual samples being mapped, in certain embodiments. In some embodiments, no degree of mismatch is allowed for a read mapped to a reference sequence.

As used herein, the term “reference genome” can refer to any particular known, sequenced or characterized genome, whether partial or complete, of any organism or virus which may be used to reference identified sequences from a subject. For example, a reference genome used for human

subjects as well as many other organisms can be found at the National Center for Biotechnology Information at World Wide Web URL ncbi.nlm.nih.gov. A “genome” refers to the complete genetic information of an organism or virus, expressed in nucleic acid sequences. As used herein, a reference sequence or reference genome often is an assembled or partially assembled genomic sequence from an individual or multiple individuals. In some embodiments, a reference genome is an assembled or partially assembled genomic sequence from one or more human individuals. In some embodiments, a reference genome comprises sequences assigned to chromosomes.

In certain embodiments, mappability is assessed for a genomic region (e.g., portion, genomic portion). Mappability is the ability to unambiguously align a nucleotide sequence read to a portion of a reference genome, typically up to a specified number of mismatches, including, for example, 0, 1, 2 or more mismatches. For a given genomic region, the expected mappability can be estimated using a sliding-window approach of a preset read length and averaging the resulting read-level mappability values. Genomic regions comprising stretches of unique nucleotide sequence sometimes have a high mappability value.

For paired-end sequencing, reads may be mapped to a reference genome by use of a suitable mapping and/or alignment program or algorithm, non-limiting examples of which include BWA (Li H. and Durbin R. (2009) *Bioinformatics* 25, 1754–60), Novoalign [Novocraft (2010)], Bowtie (Langmead B, et al., (2009) *Genome Biol.* 10:R25), SOAP2 (Li R, et al., (2009) *Bioinformatics* 25, 1966–67), BFAST (Homer N, et al., (2009) *PLoS ONE* 4, e7767), GASSST (Rizk, G. and Lavenier, D. (2010) *Bioinformatics* 26, 2534–2540), and MPscan (Rivals E., et al. (2009) *Lecture Notes in Computer Science* 5724, 246–260), and the like. Reads can be trimmed and/or merged by use of a suitable trimming and/or merging program or algorithm, non-limiting examples of which include Cutadapt, trimmomatic, SeqPrep, and usearch. Some paired-end reads, such as those from nucleic acid templates that are shorter than the sequencing read length, can have portions sequenced by both the forward read and the reverse read; in such instances, the forward and reverse reads can be merged into a single read using the overlap between the forward and reverse reads. Reads that do not overlap or that do not overlap sufficiently can remain unmerged and be mapped as paired reads. Paired-end reads may be mapped and/or aligned using a suitable short read alignment program or algorithm. Non-limiting examples of short read alignment programs include BarraCUDA, BFAST, BLASTN, BLAT, Bowtie, BWA, CASHX, CUDA-EC, CUSHAW, CUSHAW2, drFAST, ELAND, ERNE, GNUMAP, GEM, GensearchNGS, GMAP, Geneious Assembler, iSAAC, LAST, MAQ, mrFAST, mrsFAST, MOSAIK, MPscan, Novoalign, NovoalignCS, Novocraft, NextGENe, Omixon, PALMapper, Partek, PASS, PerM, QPalma, RazerS, REAL, cREAL, RMAP, rNA, RTG, Segemehl, SeqMap, Shrec, SHRiMP, SLIDER, SOAP, SOAP2, SOAP3, SOCS, SSAHA, SSAHA2, Stampy, STORM, Subread, Subjunc, Taipan, UGENE, VelociMapper, TimeLogic, XpressAlign,

ZOOM, the like or combinations thereof. Paired-end reads are often mapped to opposing ends of the same polynucleotide fragment, according to a reference genome. In some embodiments, read mates are mapped independently. In some embodiments, information from both sequence reads (i.e., from each end) is factored in the mapping process. A reference genome is often used to determine and/or infer the sequence of nucleic acids located between paired-end read mates. The term “discordant read pairs” as used herein refers to a paired-end read comprising a pair of read mates, where one or both read mates fail to unambiguously map to the same region of a reference genome defined, in part, by a segment of contiguous nucleotides. In some embodiments discordant read pairs are paired-end read mates that map to unexpected locations of a reference genome.

Non-limiting examples of unexpected locations of a reference genome include (i) two different chromosomes, (ii) locations separated by more than a predetermined fragment size (e.g., more than 300 bp, more than 500 bp, more than 1000 bp, more than 5000 bp, or more than 10,000 bp), (iii) an orientation inconsistent with a reference sequence (e.g., opposite orientations), the like or a combination thereof. In some embodiments discordant read mates are identified according to a length (e.g., an average length, a predetermined fragment size) or expected length of template polynucleotide fragments in a sample. For example, read mates that map to a location that is separated by more than the average length or expected length of polynucleotide fragments in a sample are sometimes identified as discordant read pairs. Read pairs that map in opposite orientation are sometimes determined by taking the reverse complement of one of the reads and comparing the alignment of both reads using the same strand of a reference sequence. Discordant read pairs can be identified by any suitable method and/or algorithm known in the art or described herein (e.g., SVDetect, Lumpy, BreakDancer, BreakDancerMax, CREST, DELLY, the like or combinations thereof).

Sequence read quantification

Sequence reads that are mapped or partitioned based on a selected feature or variable can be quantified to determine the amount or number of reads that are mapped to one or more portions (e.g., portion of a reference genome). In certain embodiments, the quantity of sequence reads that are mapped to a portion or segment is referred to as a count or read density.

A count often is associated with a genomic portion. In some embodiments a count is determined from some or all of the sequence reads mapped to (i.e., associated with) a portion. In certain embodiments, a count is determined from some or all of the sequence reads mapped to a group of portions (e.g., portions in a segment or region).

A count can be determined by a suitable method, operation or mathematical process. A count sometimes is the direct sum of all sequence reads mapped to a genomic portion or a group of

genomic portions corresponding to a segment, a group of portions corresponding to a sub-region of a genome (e.g., copy number variation region, copy number alteration region, copy number duplication region, copy number deletion region, microduplication region, microdeletion region, chromosome region, autosome region, sex chromosome region) and/or sometimes is a group of portions corresponding to a genome. A read quantification sometimes is a ratio, and sometimes is a ratio of a quantification for portion(s) in region *a* to a quantification for portion(s) in region *b*. Region *a* sometimes is one portion, segment region, copy number variation region, copy number alteration region, copy number duplication region, copy number deletion region, microduplication region, microdeletion region, chromosome region, autosome region and/or sex chromosome region. Region *b* independently sometimes is one portion, segment region, copy number variation region, copy number alteration region, copy number duplication region, copy number deletion region, microduplication region, microdeletion region, chromosome region, autosome region, sex chromosome region, a region including all autosomes, a region including sex chromosomes and/or a region including all chromosomes.

In some embodiments, a count is derived from raw sequence reads and/or filtered sequence reads. In certain embodiments a count is an average, mean or sum of sequence reads mapped to a genomic portion or group of genomic portions (e.g., genomic portions in a region). In some embodiments, a count is associated with an uncertainty value. A count sometimes is adjusted. A count may be adjusted according to sequence reads associated with a genomic portion or group of portions that have been weighted, removed, filtered, normalized, adjusted, averaged, derived as a mean, derived as a median, added, or combination thereof.

A sequence read quantification sometimes is a read density. A read density may be determined and/or generated for one or more segments of a genome. In certain instances, a read density may be determined and/or generated for one or more chromosomes. In some embodiments a read density comprises a quantitative measure of counts of sequence reads mapped to a segment or portion of a reference genome. A read density can be determined by a suitable process. In some embodiments a read density is determined by a suitable distribution and/or a suitable distribution function. Non-limiting examples of a distribution function include a probability function, probability distribution function, probability density function (PDF), a kernel density function (kernel density estimation), a cumulative distribution function, probability mass function, discrete probability distribution, an absolutely continuous univariate distribution, the like, any suitable distribution, or combinations thereof. A read density may be a density estimation derived from a suitable probability density function. A density estimation is the construction of an estimate, based on observed data, of an underlying probability density function. In some embodiments a read density comprises a density estimation (e.g., a probability density estimation, a kernel density estimation).

A read density may be generated according to a process comprising generating a density estimation for each of the one or more portions of a genome where each portion comprises counts of sequence reads. A read density may be generated for normalized and/or weighted counts mapped to a portion or segment. In some instances, each read mapped to a portion or segment may contribute to a read density, a value (e.g., a count) equal to its weight obtained from a normalization process described herein. In some embodiments read densities for one or more portions or segments are adjusted. Read densities can be adjusted by a suitable method. For example, read densities for one or more portions can be weighted and/or normalized.

Reads quantified for a given portion or segment can be from one source or different sources. In one example, reads may be obtained from nucleic acid from a subject having cancer or suspected of having cancer. In such circumstances, reads mapped to one or more portions often are reads representative of both healthy cells (i.e., non-cancer cells) and cancer cells (e.g., tumor cells). In certain embodiments, some of the reads mapped to a portion are from cancer cell nucleic acid and some of the reads mapped to the same portion are from non-cancer cell nucleic acid. In another example, reads may be obtained from a nucleic acid sample from a pregnant female bearing a fetus. In such circumstances, reads mapped to one or more portions often are reads representative of both the fetus and the mother of the fetus (e.g., a pregnant female subject). In certain embodiments some of the reads mapped to a portion are from a fetal genome and some of the reads mapped to the same portion are from a maternal genome.

Classifications and uses thereof

Methods described herein can provide an outcome indicative of one or more characteristics of a sample or source described above. Methods described herein sometimes provide an outcome indicative of a phenotype and/or presence or absence of a medical condition for a test sample (e.g., providing an outcome determinative of the presence or absence of a medical condition and/or phenotype). An outcome often is part of a classification process, and a classification (e.g., classification of one or more characteristics of a sample or source; and/or presence or absence of a genotype, phenotype, genetic variation and/or medical condition for a test sample) sometimes is based on and/or includes an outcome. An outcome and/or classification sometimes is based on and/or includes a result of data processing for a test sample that facilitates determining one or more characteristics of a sample or source and/or presence or absence of a genotype, phenotype, genetic variation, genetic alteration, and/or medical condition in a classification process (e.g., a statistic value). An outcome and/or classification sometimes includes or is based on a score determinative of, or a call of, one or more characteristics of a sample or source and/or presence or absence of a genotype, phenotype, genetic variation, genetic alteration, and/or medical condition.

In certain embodiments, an outcome and/or classification includes a conclusion that predicts and/or determines one or more characteristics of a sample or source and/or presence or absence of a genotype, phenotype, genetic variation, genetic alteration, and/or medical condition in a classification process.

5 Any suitable expression of an outcome and/or classification can be provided. An outcome and/or classification sometimes is based on and/or includes one or more numerical values generated using a processing method described herein in the context of one or more considerations of probability. Non-limiting examples of values that can be utilized include a sensitivity, specificity, standard deviation, median absolute deviation (MAD), measure of certainty, measure of
10 confidence, measure of certainty or confidence that a value obtained for a test sample is inside or outside a particular range of values, measure of uncertainty, measure of uncertainty that a value obtained for a test sample is inside or outside a particular range of values, coefficient of variation (CV), confidence level, confidence interval (e.g., about 95% confidence interval), standard score (e.g., z-score), chi value, phi value, result of a t-test, p-value, ploidy value, fitted minority species
15 fraction, area ratio, median level, the like or combination thereof. In some embodiments, an outcome and/or classification comprises a read density, a read density profile and/or a plot (e.g., a profile plot). In certain embodiments, multiple values are analyzed together, sometimes in a profile for such values (e.g., z-score profile, p-value profile, chi value profile, phi value profile, result of a t-test, value profile, the like, or combination thereof). A consideration of probability can facilitate
20 determining one or more characteristics of a sample or source and/or whether a subject is at risk of having, or has, a genotype, phenotype, genetic variation and/or medical condition, and an outcome and/or classification determinative of the foregoing sometimes includes such a consideration.

In certain embodiments, an outcome and/or classification is based on and/or includes a conclusion that predicts and/or determines a risk or probability of the presence or absence of a genotype,
25 phenotype, genetic variation and/or medical condition for a test sample. A conclusion sometimes is based on a value determined from a data analysis method described herein (e.g., a statistics value indicative of probability, certainty and/or uncertainty (e.g., standard deviation, median absolute deviation (MAD), measure of certainty, measure of confidence, measure of certainty or confidence that a value obtained for a test sample is inside or outside a particular range of values, measure of
30 uncertainty, measure of uncertainty that a value obtained for a test sample is inside or outside a particular range of values, coefficient of variation (CV), confidence level, confidence interval (e.g., about 95% confidence interval), standard score (e.g., z-score), chi value, phi value, result of a t-test, p-value, sensitivity, specificity, the like or combination thereof). An outcome and/or classification sometimes is expressed in a laboratory test report for particular test sample as a
35 probability (e.g., odds ratio, p-value), likelihood, or risk factor, associated with the presence or

absence of a genotype, phenotype, genetic variation and/or medical condition. An outcome and/or classification for a test sample sometimes is provided as “positive” or “negative” with respect a particular genotype, phenotype, genetic variation and/or medical condition. For example, an outcome and/or classification sometimes is designated as “positive” in a laboratory test report for a particular test sample where presence of a genotype, phenotype, genetic variation and/or medical condition is determined, and sometimes an outcome and/or classification is designated as “negative” in a laboratory test report for a particular test sample where absence of a genotype, phenotype, genetic variation and/or medical condition is determined. An outcome and/or classification sometimes is determined and sometimes includes an assumption used in data processing.

There typically are four types of classifications generated in a classification process: true positive, false positive, true negative and false negative. The term “true positive” as used herein refers to presence of a genotype, phenotype, genetic variation, or medical condition correctly determined for a test sample. The term “false positive” as used herein refers to presence of a genotype, phenotype, genetic variation, or medical condition incorrectly determined for a test sample. The term “true negative” as used herein refers to absence of a genotype, phenotype, genetic variation, or medical condition correctly determined for a test sample. The term “false negative” as used herein refers to absence of a genotype, phenotype, genetic variation, or medical condition incorrectly determined for a test sample. Two measures of performance for a classification process can be calculated based on the ratios of these occurrences: (i) a sensitivity value, which generally is the fraction of predicted positives that are correctly identified as being positives; and (ii) a specificity value, which generally is the fraction of predicted negatives correctly identified as being negative.

In certain embodiments, a laboratory test report generated for a classification process includes a measure of test performance (e.g., sensitivity and/or specificity) and/or a measure of confidence (e.g., a confidence level, confidence interval). A measure of test performance and/or confidence sometimes is obtained from a clinical validation study performed prior to performing a laboratory test for a test sample. In certain embodiments, one or more of sensitivity, specificity and/or confidence are expressed as a percentage. In some embodiments, a percentage expressed independently for each of sensitivity, specificity or confidence level, is greater than about 90% (e.g., about 90, 91, 92, 93, 94, 95, 96, 97, 98 or 99%, or greater than 99% (e.g., about 99.5%, or greater, about 99.9% or greater, about 99.95% or greater, about 99.99% or greater)). A confidence interval expressed for a particular confidence level (e.g., a confidence level of about 90% to about 99.9% (e.g., about 95%)) can be expressed as a range of values, and sometimes is expressed as a range or sensitivities and/or specificities for a particular confidence level. Coefficient of variation (CV) in

some embodiments is expressed as a percentage, and sometimes the percentage is about 10% or less (e.g., about 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1%, or less than 1% (e.g., about 0.5% or less, about 0.1% or less, about 0.05% or less, about 0.01% or less)). A probability (e.g., that a particular outcome and/or classification is not due to chance) in certain embodiments is expressed as a
5 standard score (e.g., z-score), a p-value, or result of a t-test. In some embodiments, a measured variance, confidence level, confidence interval, sensitivity, specificity and the like (e.g., referred to collectively as confidence parameters) for an outcome and/or classification can be generated using one or more data processing manipulations described herein.

An outcome and/or classification for a test sample often is ordered by, and often is provided to, a
10 health care professional or other qualified individual (e.g., physician or assistant) who transmits an outcome and/or classification to a subject from whom the test sample is obtained. In certain embodiments, an outcome and/or classification is provided using a suitable visual medium (e.g., a peripheral or component of a machine, e.g., a printer or display). A classification and/or outcome often is provided to a healthcare professional or qualified individual in the form of a report. A report
15 typically comprises a display of an outcome and/or classification (e.g., a value, one or more characteristics of a sample or source, or an assessment or probability of presence or absence of a genotype, phenotype, genetic variation and/or medical condition), sometimes includes an associated confidence parameter, and sometimes includes a measure of performance for a test used to generate the outcome and/or classification. A report sometimes includes a
20 recommendation for a follow-up procedure (e.g., a procedure that confirms the outcome or classification). A report sometimes includes a visual representation of a chromosome or portion thereof (e.g., a chromosome ideogram or karyogram), and sometimes shows a visualization of a duplication and/or deletion region for a chromosome (e.g., a visualization of a whole chromosome for a chromosome deletion or duplication; a visualization of a whole chromosome with a deleted
25 region or duplicated region shown; a visualization of a portion of chromosome duplicated or deleted; a visualization of a portion of a chromosome remaining in the event of a deletion of a portion of a chromosome) identified for a test sample.

A report can be displayed in a suitable format that facilitates determination of presence or absence of a genotype, phenotype, genetic variation and/or medical condition by a health professional or
30 other qualified individual. Non-limiting examples of formats suitable for use for generating a report include digital data, a graph, a 2D graph, a 3D graph, and 4D graph, a picture (e.g., a jpg, bitmap (e.g., bmp), pdf, tiff, gif, raw, png, the like or suitable format), a pictograph, a chart, a table, a bar graph, a pie graph, a diagram, a flow chart, a scatter plot, a map, a histogram, a density chart, a function graph, a circuit diagram, a block diagram, a bubble map, a constellation diagram, a

contour diagram, a cartogram, spider chart, Venn diagram, nomogram, and the like, or combination of the foregoing.

A report may be generated by a computer and/or by human data entry, and can be transmitted and communicated using a suitable electronic medium (e.g., via the internet, via computer, via

5 facsimile, from one network location to another location at the same or different physical sites), or by another method of sending or receiving data (e.g., mail service, courier service and the like).

Non-limiting examples of communication media for transmitting a report include auditory file, computer readable file (e.g., pdf file), paper file, laboratory file, medical record file, or any other medium described in the previous paragraph. A laboratory file or medical record file may be in

10 tangible form or electronic form (e.g., computer readable form), in certain embodiments. After a report is generated and transmitted, a report can be received by obtaining, via a suitable communication medium, a written and/or graphical representation comprising an outcome and/or classification, which upon review allows a healthcare professional or other qualified individual to make a determination as to one or more characteristics of a sample or source, or presence or
15 absence of a genotype, phenotype, genetic variation and/or or medical condition for a test sample.

An outcome and/or classification may be provided by and obtained from a laboratory (e.g., obtained from a laboratory file). A laboratory file can be generated by a laboratory that carries out one or more tests for determining one or more characteristics of a sample or source and/or presence or absence of a genotype, phenotype, genetic variation and/or medical condition for a test sample.

20 Laboratory personnel (e.g., a laboratory manager) can analyze information associated with test samples (e.g., test profiles, reference profiles, test values, reference values, level of deviation, patient information) underlying an outcome and/or classification. For calls pertaining to presence or absence of a genotype, phenotype, genetic variation and/or medical condition that are close or questionable, laboratory personnel can re-run the same procedure using the same (e.g., aliquot of
25 the same sample) or different test sample from a test subject. A laboratory may be in the same location or different location (e.g., in another country) as personnel assessing the presence or absence of a genotype, phenotype, genetic variation and/or a medical condition from the laboratory file. For example, a laboratory file can be generated in one location and transmitted to another location in which the information for a test sample therein is assessed by a healthcare professional
30 or other qualified individual, and optionally, transmitted to the subject from which the test sample was obtained. A laboratory sometimes generates and/or transmits a laboratory report containing a classification of presence or absence of genomic instability, a genotype, phenotype, a genetic variation and/or a medical condition for a test sample. A laboratory generating a laboratory test report sometimes is a certified laboratory, and sometimes is a laboratory certified under the Clinical
35 Laboratory Improvement Amendments (CLIA).

An outcome and/or classification sometimes is a component of a diagnosis for a subject, and sometimes an outcome and/or classification is utilized and/or assessed as part of providing a diagnosis for a test sample. For example, a healthcare professional or other qualified individual may analyze an outcome and/or classification and provide a diagnosis based on, or based in part on, the outcome and/or classification. In some embodiments, determination, detection or diagnosis of a medical condition, disease, syndrome or abnormality comprises use of an outcome and/or classification determinative of presence or absence of a genotype, phenotype, genetic variation and/or medical condition. Thus, provided herein are methods for diagnosing presence or absence of a genotype, phenotype, a genetic variation and/or a medical condition for a test sample according to an outcome or classification generated by methods described herein, and optionally according to generating and transmitting a laboratory report that includes a classification for presence or absence of the genotype, phenotype, a genetic variation and/or a medical condition for the test sample.

Machines, software and interfaces

Certain processes and methods described herein (e.g., selecting a subset of sequence reads, generating a sequence read profile, processing sequence read data, processing sequence read quantifications, determining one or more characteristics of a sample based on sequence read data or a sequence read profile) often are too complex for performing in the mind and cannot be performed without a computer, microprocessor, software, module or other machine. Methods described herein may be computer-implemented methods, and one or more portions of a method sometimes are performed by one or more processors (e.g., microprocessors), computers, systems, apparatuses, or machines (e.g., microprocessor-controlled machine).

Computers, systems, apparatuses, machines and computer program products suitable for use often include, or are utilized in conjunction with, computer readable storage media. Non-limiting examples of computer readable storage media include memory, hard disk, CD-ROM, flash memory device and the like. Computer readable storage media generally are computer hardware, and often are non-transitory computer-readable storage media. Computer readable storage media are not computer readable transmission media, the latter of which are transmission signals per se.

Provided herein are computer readable storage media with an executable program stored thereon, where the program instructs a microprocessor to perform a method described herein. Provided also are computer readable storage media with an executable program module stored thereon, where the program module instructs a microprocessor to perform part of a method described herein. Also provided herein are systems, machines, apparatuses and computer program products that include computer readable storage media with an executable program stored thereon, where the program

instructs a microprocessor to perform a method described herein. Provided also are systems, machines and apparatuses that include computer readable storage media with an executable program module stored thereon, where the program module instructs a microprocessor to perform part of a method described herein.

- 5 Also provided are computer program products. A computer program product often includes a computer usable medium that includes a computer readable program code embodied therein, the computer readable program code adapted for being executed to implement a method or part of a method described herein. Computer usable media and readable program code are not transmission media (i.e., transmission signals per se). Computer readable program code often is adapted for being executed by a processor, computer, system, apparatus, or machine.

In some embodiments, methods described herein (e.g., selecting a subset of sequence reads, generating a sequence read profile, processing sequence read data, processing sequence read quantifications, determining one or more characteristics of a sample based on sequence read data or a sequence read profile) are performed by automated methods. In some embodiments, one or more steps of a method described herein are carried out by a microprocessor and/or computer, and/or carried out in conjunction with memory. In some embodiments, an automated method is embodied in software, modules, microprocessors, peripherals and/or a machine comprising the like, that perform methods described herein. As used herein, software refers to computer readable program instructions that, when executed by a microprocessor, perform computer operations, as described herein.

Machines, software and interfaces may be used to conduct methods described herein. Using machines, software and interfaces, a user may enter, request, query or determine options for using particular information, programs or processes (e.g., processing sequence read data, processing sequence read quantifications, and/or providing an outcome), which can involve implementing statistical analysis algorithms, statistical significance algorithms, statistical algorithms, iterative steps, validation algorithms, and graphical representations, for example. In some embodiments, a data set may be entered by a user as input information, a user may download one or more data sets by suitable hardware media (e.g., flash drive), and/or a user may send a data set from one system to another for subsequent processing and/or providing an outcome (e.g., send sequence read data from a sequencer to a computer system for sequence read processing; send processed sequence read data to a computer system for further processing and/or yielding an outcome and/or report).

A system typically comprises one or more machines. Each machine comprises one or more of memory, one or more microprocessors, and instructions. Where a system includes two or more

machines, some or all of the machines may be located at the same location, some or all of the machines may be located at different locations, all of the machines may be located at one location and/or all of the machines may be located at different locations. Where a system includes two or more machines, some or all of the machines may be located at the same location as a user, some
5 or all of the machines may be located at a location different than a user, all of the machines may be located at the same location as the user, and/or all of the machine may be located at one or more locations different than the user.

A system sometimes comprises a computing machine and a sequencing apparatus or machine, where the sequencing apparatus or machine is configured to receive physical nucleic acid and
10 generate sequence reads, and the computing apparatus is configured to process the reads from the sequencing apparatus or machine. The computing machine sometimes is configured to determine an outcome from the sequence reads (e.g., a characteristic of a sample).

A user may, for example, place a query to software which then may acquire a data set via internet access, and in certain embodiments, a programmable microprocessor may be prompted to acquire
15 a suitable data set based on given parameters. A programmable microprocessor also may prompt a user to select one or more data set options selected by the microprocessor based on given parameters. A programmable microprocessor may prompt a user to select one or more data set options selected by the microprocessor based on information found via the internet, other internal or external information, or the like. Options may be chosen for selecting one or more data feature
20 selections, one or more statistical algorithms, one or more statistical analysis algorithms, one or more statistical significance algorithms, iterative steps, one or more validation algorithms, and one or more graphical representations of methods, machines, apparatuses, computer programs or a non-transitory computer-readable storage medium with an executable program stored thereon.

Systems addressed herein may comprise general components of computer systems, such as, for
25 example, network servers, laptop systems, desktop systems, handheld systems, personal digital assistants, computing kiosks, and the like. A computer system may comprise one or more input means such as a keyboard, touch screen, mouse, voice recognition or other means to allow the user to enter data into the system. A system may further comprise one or more outputs, including, but not limited to, a display screen (e.g., CRT or LCD), speaker, FAX machine, printer (e.g., laser,
30 ink jet, impact, black and white or color printer), or other output useful for providing visual, auditory and/or hardcopy output of information (e.g., outcome and/or report).

In a system, input and output components may be connected to a central processing unit which may comprise among other components, a microprocessor for executing program instructions and memory for storing program code and data. In some embodiments, processes may be

implemented as a single user system located in a single geographical site. In certain embodiments, processes may be implemented as a multi-user system. In the case of a multi-user implementation, multiple central processing units may be connected by means of a network. The network may be local, encompassing a single department in one portion of a building, an entire building, span
5 multiple buildings, span a region, span an entire country or be worldwide. The network may be private, being owned and controlled by a provider, or it may be implemented as an internet-based service where the user accesses a web page to enter and retrieve information. Accordingly, in certain embodiments, a system includes one or more machines, which may be local or remote with respect to a user. More than one machine in one location or multiple locations may be accessed by
10 a user, and data may be mapped and/or processed in series and/or in parallel. Thus, a suitable configuration and control may be utilized for mapping and/or processing data using multiple machines, such as in local network, remote network and/or "cloud" computing platforms.

A system can include a communications interface in some embodiments. A communications interface allows for transfer of software and data between a computer system and one or more
15 external devices. Non-limiting examples of communications interfaces include a modem, a network interface (such as an Ethernet card), a communications port, a PCMCIA slot and card, and the like. Software and data transferred via a communications interface generally are in the form of signals, which can be electronic, electromagnetic, optical and/or other signals capable of being received by a communications interface. Signals often are provided to a communications interface via a
20 channel. A channel often carries signals and can be implemented using wire or cable, fiber optics, a phone line, a cellular phone link, an RF link and/or other communications channels. Thus, in an example, a communications interface may be used to receive signal information that can be detected by a signal detection module.

Data may be input by a suitable device and/or method, including, but not limited to, manual input
25 devices or direct data entry devices (DDEs). Non-limiting examples of manual devices include keyboards, concept keyboards, touch sensitive screens, light pens, mouse, tracker balls, joysticks, graphic tablets, scanners, digital cameras, video digitizers and voice recognition devices. Non-limiting examples of DDEs include bar code readers, magnetic strip codes, smart cards, magnetic ink character recognition, optical character recognition, optical mark recognition, and turnaround
30 documents.

In some embodiments, output from a sequencing apparatus or machine may serve as data that can be input via an input device. In certain embodiments, sequence read information may serve as data that can be input via an input device. In certain embodiments, mapped sequence reads may serve as data that can be input via an input device. In certain embodiments, nucleic acid fragment size

(e.g., length) may serve as data that can be input via an input device. In certain embodiments, output from a nucleic acid capture process (e.g., genomic region origin data) may serve as data that can be input via an input device. In certain embodiments, a combination of nucleic acid fragment size (e.g., length) and output from a nucleic acid capture process (e.g., genomic region origin data) may serve as data that can be input via an input device. In certain embodiments, simulated data is generated by an in silico process and the simulated data serves as data that can be input via an input device. The term "in silico" refers to research and experiments performed using a computer. In silico processes include, but are not limited to, mapping sequence reads and processing mapped sequence reads according to processes described herein.

A system may include software useful for performing a process or part of a process described herein, and software can include one or more modules for performing such processes (e.g., sequencing module, logic processing module, data display organization module). The term "software" refers to computer readable program instructions that, when executed by a computer, perform computer operations. Instructions executable by the one or more microprocessors sometimes are provided as executable code, that when executed, can cause one or more microprocessors to implement a method described herein. A module described herein can exist as software, and instructions (e.g., processes, routines, subroutines) embodied in the software can be implemented or performed by a microprocessor. For example, a module (e.g., a software module) can be a part of a program that performs a particular process or task. The term "module" refers to a self-contained functional unit that can be used in a larger machine or software system. A module can comprise a set of instructions for carrying out a function of the module. A module can transform data and/or information. Data and/or information can be in a suitable form. For example, data and/or information can be digital or analogue. In certain embodiments, data and/or information sometimes can be packets, bytes, characters, or bits. In some embodiments, data and/or information can be any gathered, assembled or usable data or information. Non-limiting examples of data and/or information include a suitable media, pictures, video, sound (e.g., frequencies, audible or non-audible), numbers, constants, a value, objects, time, functions, instructions, maps, references, sequences, reads, mapped reads, levels, ranges, thresholds, signals, displays, representations, or transformations thereof. A module can accept or receive data and/or information, transform the data and/or information into a second form, and provide or transfer the second form to a machine, peripheral, component or another module. A microprocessor can, in certain embodiments, carry out the instructions in a module. In some embodiments, one or more microprocessors are required to carry out instructions in a module or group of modules. A module can provide data and/or information to another module, machine or source and can receive data and/or information from another module, machine or source.

A computer program product sometimes is embodied on a tangible computer-readable medium, and sometimes is tangibly embodied on a non-transitory computer-readable medium. A module sometimes is stored on a computer readable medium (e.g., disk, drive) or in memory (e.g., random access memory). A module and microprocessor capable of implementing instructions from a
5 module can be located in a machine or in a different machine. A module and/or microprocessor capable of implementing an instruction for a module can be located in the same location as a user (e.g., local network) or in a different location from a user (e.g., remote network, cloud system). In embodiments in which a method is carried out in conjunction with two or more modules, the modules can be located in the same machine, one or more modules can be located in different
10 machine in the same physical location, and one or more modules may be located in different machines in different physical locations.

A machine, in some embodiments, comprises at least one microprocessor for carrying out the instructions in a module. Sequence read quantifications (e.g., counts) sometimes are accessed by a microprocessor that executes instructions configured to carry out a method described herein.
15 Sequence read quantifications that are accessed by a microprocessor can be within memory of a system, and the sequence read counts can be accessed and placed into the memory of the system after they are obtained. In some embodiments, a machine includes a microprocessor (e.g., one or more microprocessors) which microprocessor can perform and/or implement one or more instructions (e.g., processes, routines and/or subroutines) from a module. In some embodiments, a
20 machine includes multiple microprocessors, such as microprocessors coordinated and working in parallel. In some embodiments, a machine operates with one or more external microprocessors (e.g., an internal or external network, server, storage device and/or storage network (e.g., a cloud)). In some embodiments, a machine comprises a module (e.g., one or more modules). A machine comprising a module often is capable of receiving and transferring one or more of data and/or
25 information to and from other modules.

In certain embodiments, a machine comprises peripherals and/or components. In certain embodiments, a machine can comprise one or more peripherals or components that can transfer data and/or information to and from other modules, peripherals and/or components. In certain
30 embodiments, a machine interacts with a peripheral and/or component that provides data and/or information. In certain embodiments, peripherals and components assist a machine in carrying out a function or interact directly with a module. Non-limiting examples of peripherals and/or components include a suitable computer peripheral, I/O or storage method or device including but not limited to scanners, printers, displays (e.g., monitors, LED, LCT or CRTs), cameras, microphones, pads (e.g., ipads, tablets), touch screens, smart phones, mobile phones, USB I/O
35 devices, USB mass storage devices, keyboards, a computer mouse, digital pens, modems, hard

drives, jump drives, flash drives, a microprocessor, a server, CDs, DVDs, graphic cards, specialized I/O devices (e.g., sequencers, photo cells, photo multiplier tubes, optical readers, sensors, etc.), one or more flow cells, fluid handling components, network interface controllers, ROM, RAM, wireless transfer methods and devices (Bluetooth, WiFi, and the like,), the world wide web (www), the internet, a computer and/or another module.

Software often is provided on a program product containing program instructions recorded on a computer readable medium, including, but not limited to, magnetic media including floppy disks, hard disks, and magnetic tape; and optical media including CD-ROM discs, DVD discs, magneto-optical discs, flash memory devices (e.g., flash drives), RAM, floppy discs, the like, and other such media on which the program instructions can be recorded. In online implementation, a server and web site maintained by an organization can be configured to provide software downloads to remote users, or remote users may access a remote system maintained by an organization to remotely access software. Software may obtain or receive input information. Software may include a module that specifically obtains or receives data (e.g., a data receiving module that receives sequence read data and/or mapped read data) and may include a module that specifically processes the data (e.g., a processing module that processes received data (e.g., filters, normalizes, provides an outcome and/or report). The terms "obtaining" and "receiving" input information refers to receiving data (e.g., sequence reads, mapped reads) by computer communication means from a local, or remote site, human data entry, or any other method of receiving data. The input information may be generated in the same location at which it is received, or it may be generated in a different location and transmitted to the receiving location. In some embodiments, input information is modified before it is processed (e.g., placed into a format amenable to processing (e.g., tabulated)).

Software can include one or more algorithms in certain embodiments. An algorithm may be used for processing data and/or providing an outcome or report according to a finite sequence of instructions. An algorithm often is a list of defined instructions for completing a task. Starting from an initial state, the instructions may describe a computation that proceeds through a defined series of successive states, eventually terminating in a final ending state. The transition from one state to the next is not necessarily deterministic (e.g., some algorithms incorporate randomness). By way of example, and without limitation, an algorithm can be a search algorithm, sorting algorithm, merge algorithm, numerical algorithm, graph algorithm, string algorithm, modeling algorithm, computational genometric algorithm, combinatorial algorithm, machine learning algorithm, cryptography algorithm, data compression algorithm, parsing algorithm and the like. An algorithm can include one algorithm or two or more algorithms working in combination. An algorithm can be of any suitable complexity class and/or parameterized complexity. An algorithm can be used for calculation and/or data processing, and in some embodiments, can be used in a deterministic or

probabilistic/predictive approach. An algorithm can be implemented in a computing environment by use of a suitable programming language, non-limiting examples of which are C, C++, Java, Perl, Python, Fortran, and the like. In some embodiments, an algorithm can be configured or modified to include margin of errors, statistical analysis, statistical significance, and/or comparison to other information or data sets (e.g., applicable when using a neural net or clustering algorithm).

In certain embodiments, several algorithms may be implemented for use in software. These algorithms can be trained with raw data in some embodiments. For each new raw data sample, the trained algorithms may produce a representative processed data set or outcome. A processed data set sometimes is of reduced complexity compared to the parent data set that was processed.

Based on a processed set, the performance of a trained algorithm may be assessed based on sensitivity and specificity, in some embodiments. An algorithm with the highest sensitivity and/or specificity may be identified and utilized, in certain embodiments.

In certain embodiments, simulated (or simulation) data can aid data processing, for example, by training an algorithm or testing an algorithm. In some embodiments, simulated data includes hypothetical various samplings of different groupings of sequence reads. Simulated data may be based on what might be expected from a real population or may be skewed to test an algorithm and/or to assign a correct classification. Simulated data also is referred to herein as “virtual” data. Simulations can be performed by a computer program in certain embodiments. One possible step in using a simulated data set is to evaluate the confidence of identified results, e.g., how well a random sampling matches or best represents the original data. One approach is to calculate a probability value (p-value), which estimates the probability of a random sample having better score than the selected samples. In some embodiments, an empirical model may be assessed, in which it is assumed that at least one sample matches a reference sample (with or without resolved variations). In some embodiments, another distribution, such as a Poisson distribution for example, can be used to define the probability distribution.

A system may include one or more microprocessors in certain embodiments. A microprocessor can be connected to a communication bus. A computer system may include a main memory, often random-access memory (RAM), and can also include a secondary memory. Memory in some embodiments comprises a non-transitory computer-readable storage medium. Secondary memory can include, for example, a hard disk drive and/or a removable storage drive, representing a floppy disk drive, a magnetic tape drive, an optical disk drive, memory card and the like. A removable storage drive often reads from and/or writes to a removable storage unit. Non-limiting examples of removable storage units include a floppy disk, magnetic tape, optical disk, and the like, which can

be read by and written to by, for example, a removable storage drive. A removable storage unit can include a computer-usable storage medium having stored therein computer software and/or data.

A microprocessor may implement software in a system. In some embodiments, a microprocessor may be programmed to automatically perform a task described herein that a user could perform.

5 Accordingly, a microprocessor, or algorithm conducted by such a microprocessor, can require little to no supervision or input from a user (e.g., software may be programmed to implement a function automatically). In some embodiments, the complexity of a process is so large that a single person or group of persons could not perform the process in a timeframe short enough for determining one or more characteristics of a sample.

10 In some embodiments, secondary memory may include other similar means for allowing computer programs or other instructions to be loaded into a computer system. For example, a system can include a removable storage unit and an interface device. Non-limiting examples of such systems include a program cartridge and cartridge interface (such as that found in video game devices), a removable memory chip (such as an EPROM, or PROM) and associated socket, and other
15 removable storage units and interfaces that allow software and data to be transferred from the removable storage unit to a computer system.

Kits

Provided in certain embodiments are kits. The kits may include any components and compositions described herein (e.g., scaffold adapters and components/subcomponents thereof,
20 oligonucleotides, oligonucleotide components/regions, scaffold polynucleotides, scaffold polynucleotide components/regions, nucleic acids, single-stranded nucleic acids, primers, single-stranded binding proteins, enzymes) useful for performing any of the methods described herein, in any suitable combination. Kits may further include any reagents, buffers, or other components useful for carrying out any of the methods described herein. For example, a kit may include one or
25 more of a plurality of scaffold adapter species, a plurality of oligonucleotide species, and/or a plurality of scaffold polynucleotide species, a kinase adapted to 5' phosphorylate nucleic acids (e.g., a polynucleotide kinase (PNK)), a kinase adapted to 3' dephosphorylate nucleic acids (e.g., a polynucleotide kinase (PNK)), a phosphatase (e.g., rSAP), a DNA ligase (e.g., T4 DNA ligase), and any combination thereof. In some embodiments, a kit includes a first 5' phosphorylated
30 oligonucleotide species and a second 3' phosphorylated oligonucleotide species. In some embodiments, a kit includes one or more agents comprising a ligase activity. In some embodiments, a kit includes one or more agents comprising a kinase activity. In some embodiments, a kit includes one or more agents comprising a phosphatase activity. In some embodiments, a kit includes one or more agents comprising a 3' phosphatase activity. In some

embodiments, a kit includes one or more agents comprising a 3' phosphatase activity and a kinase activity. In some embodiments, a kit further comprises one or more of a reverse transcriptase, a polymerase, single stranded binding proteins (SSBs), one or more distinctive nucleotides, a hairpin adapter, and the like. In some embodiments, a kit further comprises one or more of an RNA probe set (e.g., a probe set comprising one or more probe species, where each probe species comprises a sequence complementary to an RNA sequence), as described herein. In some embodiments, a kit further comprises one or more of an rRNA probe set (e.g., a probe set comprising one or more probe species, where each probe species comprises a sequence complementary to a ribosomal RNA (rRNA) sequence), as described herein.

Kits may include components for capturing single-stranded DNA, single-stranded RNA, and/or single-stranded cDNA. Kits for capturing single-stranded DNA may be configured such that a user provides double-stranded or single-stranded DNA. Kits for capturing single-stranded RNA may be configured such that a user provides cDNA (either single or double stranded), or provides RNA (e.g., total RNA or rRNA-depleted RNA). A kit for capturing single-stranded RNA may include rRNA depletion reagents, mRNA enrichment reagents, fragmentation reagents, cDNA synthesis reagents, and/or RNA digestion reagents.

Components of a kit may be present in separate containers, or multiple components may be present in a single container. Suitable containers include a single tube (e.g., vial), one or more wells of a plate (e.g., a 96-well plate, a 384-well plate, and the like), and the like.

Kits may also comprise instructions for performing one or more methods described herein and/or a description of one or more components described herein. For example, a kit may include instructions for using scaffold adapters described herein, or components thereof, to capture single-stranded nucleic acid fragments and/or to produce a nucleic acid library. Instructions and/or descriptions may be in printed form and may be included in a kit insert. In some embodiments, instructions and/or descriptions are provided as an electronic storage data file present on a suitable computer readable storage medium, e.g., portable flash drive, DVD, CD-ROM, diskette, and the like. A kit also may include a written description of an internet location that provides such instructions or descriptions.

Certain Implementations

Following are non-limiting examples of certain implementations of the technology.

A1. A method of producing a nucleic acid library, comprising:

(a) combining (i) a nucleic acid composition comprising single-stranded nucleic acid (ssNA),
 5 (ii) a first oligonucleotide, (iii) a plurality of first scaffold polynucleotide species, (iv) a second oligonucleotide, and (v) a plurality of second scaffold polynucleotide species, wherein:

(1) the first oligonucleotide comprises a 5' phosphate;

(2) the second oligonucleotide comprises a 3' phosphate;

(3) each polynucleotide in the plurality of first scaffold polynucleotide species
 10 comprises an ssNA hybridization region and a first oligonucleotide hybridization region;

(4) each polynucleotide in the plurality of second scaffold polynucleotide species
 comprises an ssNA hybridization region and a second oligonucleotide hybridization region;
 and

(5) the nucleic acid composition, the first oligonucleotide, the plurality of first scaffold
 15 polynucleotide species, the second oligonucleotide, and the plurality of second scaffold polynucleotide species, are combined under conditions wherein:

(i) a molecule of the first scaffold polynucleotide species is hybridized to a
 first ssNA terminal region and a molecule of the first oligonucleotide, and

(ii) a molecule of the second scaffold polynucleotide species is hybridized to
 20 a second ssNA terminal region and a molecule of the second oligonucleotide,
 thereby forming a set of hybridization products in which a 5' end of the molecule of
 the first oligonucleotide is adjacent to a 3' end of the first ssNA terminal region and a
 3' end of the molecule of the second oligonucleotide is adjacent to a 5' end of the
 second ssNA terminal region;

(b) contacting the set of hybridization products with one or more agents comprising a ligase
 25 activity, thereby generating a first set of ligation products, wherein:

(i) the adjacent ends of the first oligonucleotide and the first ssNA terminal
 region are ligated; and

(ii) the adjacent ends of the second oligonucleotide and the second ssNA
 30 terminal region are not ligated; and

(c) contacting the first set of ligation products with one or more agents comprising a 3'
 phosphatase activity, thereby generating a second set of ligation products wherein the adjacent
 ends of the second oligonucleotide and the second ssNA terminal region are ligated.

A2. The method of embodiment A1, further comprising prior to the combining in (a), contacting the ssNA with one or more agents comprising a phosphatase activity under conditions in which the ssNA is dephosphorylated, thereby generating dephosphorylated ssNA.

A3. The method of embodiment A1 or A2, further comprising prior to the combining in (a),
5 contacting the ssNA with single-stranded nucleic acid binding protein (SSB) to produce SSB-bound ssNA.

A4. The method of embodiment A1, further comprising prior to the combining in (a), contacting the ssNA with an agent comprising a phosphatase activity under conditions in which the ssNA is dephosphorylated, and contacting the ssNA with single-stranded nucleic acid binding protein (SSB)
10 thereby generating SSB-bound dephosphorylated ssNA.

A4.1 The method of embodiment A4, wherein the ssNA is simultaneously contacted with (i) SSB and (ii) an agent comprising a phosphatase activity.

A5. The method of any one of embodiments A1-A4.1, wherein prior to the combining in (a), each of the first scaffold polynucleotide species is hybridized to a first oligonucleotide to form a plurality of
15 first scaffold duplex species, and/or each of the second scaffold polynucleotide species is hybridized to a second oligonucleotide to form a plurality of second scaffold duplex species.

A6. The method of any one of embodiments A1-A5, wherein the one or more agents comprising a ligase activity comprise a T4 DNA ligase.

A7. The method of any one of embodiments A1-A6, wherein the one or more agents comprising a
20 3' phosphatase activity further comprise a kinase activity.

A7.1 The method of any one of embodiments A1-A7, wherein the one or more agents comprising a 3' phosphatase activity comprise a polynucleotide kinase (PNK).

A8. The method of any one of embodiments A1-A7.1, wherein the ssNA hybridization region of each of the first polynucleotide species is different than the ssNA hybridization region in other first
25 polynucleotide species in the plurality of first polynucleotide species.

A9. The method of any one of embodiments A1-A8, wherein the ssNA hybridization region of each of the second polynucleotide species is different than the ssNA hybridization region in other second polynucleotide species in the plurality of second polynucleotide species.

A10. The method of any one of embodiments A1-A9, wherein the ssNA hybridization region in the
30 first scaffold polynucleotide species and/or second scaffold polynucleotide species comprises a random sequence.

A11. The method of any one of embodiments A1-A10, wherein the nucleic acid composition comprises single-stranded DNA (ssDNA).

A12. The method of embodiment A11, wherein the ssDNA is derived from double-stranded DNA (dsDNA).

- 5 A13. The method of embodiment A12, comprising prior to the combining in (a), denaturing the dsDNA, thereby generating the ssDNA.

A13.1 The method of embodiment A13, wherein the dsDNA is simultaneously denatured and contacted with (i) SSB and (ii) an agent comprising a phosphatase activity.

- 10 A14. The method of any one of embodiments A1-A13.1, wherein the ssDNA is from cell-free nucleic acid.

A15. The method of any one of embodiments A1-A14, further comprising denaturing the second set of ligation products, thereby generating single-stranded ligation products.

A16. The method of embodiment A15, further comprising amplifying the single-stranded ligation products, thereby generating amplified ligation products.

- 15 A17. The method of embodiment A16, further comprising sequencing the amplified ligation products.

A18. The method of any one of embodiments A1-A17, wherein the nucleic acid composition, the first oligonucleotide, the plurality of first scaffold polynucleotide species, the second oligonucleotide, and the plurality of second scaffold polynucleotide species, are combined simultaneously in (a).

- 20 A19. The method of any one of embodiments A1-A18, wherein (a) and (b) are performed simultaneously.

A20. A method of producing a nucleic acid library, comprising:

- 25 (a) combining (i) a nucleic acid composition comprising cell-free single-stranded nucleic acid (ssDNA), (ii) a first oligonucleotide, (iii) a plurality of first scaffold polynucleotide species, (iv) a second oligonucleotide, and (v) a plurality of second scaffold polynucleotide species, wherein:

- 30 (1) the first oligonucleotide comprises a 5' phosphate;
 (2) the second oligonucleotide comprises a 3' phosphate;
 (3) each polynucleotide in the plurality of first scaffold polynucleotide species comprises an ssDNA hybridization region and a first oligonucleotide hybridization region;
 (4) each polynucleotide in the plurality of second scaffold polynucleotide species comprises an ssDNA hybridization region and a second oligonucleotide hybridization region;

(5) the nucleic acid composition, the first oligonucleotide, the plurality of first scaffold polynucleotide species, the second oligonucleotide, and the plurality of second scaffold polynucleotide species, are combined under conditions wherein:

(i) a molecule of the first scaffold polynucleotide species is hybridized to a first ssNA terminal region and a molecule of the first oligonucleotide, and

(ii) a molecule of the second scaffold polynucleotide species is hybridized to a second ssNA terminal region and a molecule of the second oligonucleotide, thereby forming a set of hybridization products in which a 5' end of the molecule of the first oligonucleotide is adjacent to a 3' end of the first ssNA terminal region and a 3' end of the molecule of the second oligonucleotide is adjacent to a 5' end of the second ssNA terminal region; and

(6) the ssNA comprises single-stranded DNA;

(b) contacting the set of hybridization products with a T4 DNA ligase, thereby generating a first set of ligation products, wherein:

(i) the adjacent ends of the first oligonucleotide and the first ssNA terminal region are ligated; and

(ii) the adjacent ends of the second oligonucleotide and the second ssNA terminal region are not ligated; and

(c) contacting the first set of ligation products with a polynucleotide kinase (PNK), thereby generating a second set of ligation products wherein the adjacent ends of the second oligonucleotide and the second ssNA terminal region are ligated.

B1. A kit comprising:

(a) a first oligonucleotide, a plurality of first scaffold polynucleotide species, a second oligonucleotide, and a plurality of second scaffold polynucleotide species, wherein:

(i) the first oligonucleotide comprises a 5' phosphate;

(ii) the second oligonucleotide comprises a 3' phosphate;

(iii) each polynucleotide in the plurality of first scaffold polynucleotide species comprises an ssNA hybridization region and a first oligonucleotide hybridization region; and

(iv) each polynucleotide in the plurality of second scaffold polynucleotide species comprises an ssNA hybridization region and a second oligonucleotide hybridization region;

(b) one or more agents comprising a ligase activity; and

(c) one or more agents comprising a 3' phosphatase activity.

B2. The kit of embodiment B1, wherein the one or more agents in (c) further comprise a kinase activity.

B3. The kit of embodiment B1 or B2, wherein the one or more agents in (c) comprise a polynucleotide kinase (PNK).

B4. The kit of any one of embodiments B1-B3, further comprising single-stranded nucleic acid binding protein (SSB).

5 B5. The kit of any one of embodiments B1-B4, wherein the one or more agents comprising a ligase activity comprise a T4 DNA ligase.

B6. The kit of any one of embodiments B1-B5, wherein the ssNA hybridization region of each of the first polynucleotide species is different than the ssNA hybridization region in other first polynucleotide species in the plurality of first polynucleotide species.

10 B7. The kit of any one of embodiments B1-B6, wherein the ssNA hybridization region of each of the second polynucleotide species is different than the ssNA hybridization region in other second polynucleotide species in the plurality of second polynucleotide species.

15 B8. The kit of any one of embodiments B1-B7, wherein the ssNA hybridization region in the first scaffold polynucleotide species and/or second scaffold polynucleotide species comprises a random sequence.

C1. A method of producing a nucleic acid library, comprising:

(a) combining (i) a nucleic acid composition comprising single-stranded nucleic acid (ssNA), (ii) a first oligonucleotide, (iii) a plurality of first scaffold polynucleotide species, (iv) a second oligonucleotide, and (v) a plurality of second scaffold polynucleotide species, wherein:

20 (1) each polynucleotide in the plurality of first scaffold polynucleotide species comprises an ssNA hybridization region and a first oligonucleotide hybridization region;

(2) each polynucleotide in the plurality of second scaffold polynucleotide species comprises an ssNA hybridization region and a second oligonucleotide hybridization region; and

25 (3) the nucleic acid composition, the first oligonucleotide, the plurality of first scaffold polynucleotide species, the second oligonucleotide, and the plurality of second scaffold polynucleotide species, are combined under conditions wherein:

(i) a molecule of the first scaffold polynucleotide species is hybridized to a first ssNA terminal region and a molecule of the first oligonucleotide, and

30 (ii) a molecule of the second scaffold polynucleotide species is hybridized to a second ssNA terminal region and a molecule of the second oligonucleotide, thereby forming a set of hybridization products;

(b) contacting the set of hybridization products with one or more agents comprising a ligase activity, thereby generating a set of ligation products;

(c) contacting under hybridization conditions the set of ligation products with one or more probe species, wherein each probe species comprises a sequence complementary to an RNA sequence, thereby generating a mixture of hybridized ligation products and unhybridized ligation products; and

(d) separating the hybridized ligation products from the unhybridized ligation products.

C1.1 A method of producing a nucleic acid library, comprising:

(a) combining (i) a nucleic acid composition comprising single-stranded nucleic acid (ssNA), (ii) a first oligonucleotide, (iii) a plurality of first scaffold polynucleotide species, (iv) a second oligonucleotide, and (v) a plurality of second scaffold polynucleotide species, wherein:

(1) each polynucleotide in the plurality of first scaffold polynucleotide species comprises an ssNA hybridization region and a first oligonucleotide hybridization region;

(2) each polynucleotide in the plurality of second scaffold polynucleotide species comprises an ssNA hybridization region and a second oligonucleotide hybridization region; and

(3) the nucleic acid composition, the first oligonucleotide, the plurality of first scaffold polynucleotide species, the second oligonucleotide, and the plurality of second scaffold polynucleotide species, are combined under conditions wherein:

(i) a molecule of the first scaffold polynucleotide species is hybridized to a first ssNA terminal region and a molecule of the first oligonucleotide, and

(ii) a molecule of the second scaffold polynucleotide species is hybridized to a second ssNA terminal region and a molecule of the second oligonucleotide, thereby forming a set of hybridization products;

(b) contacting the set of hybridization products with one or more agents comprising a ligase activity, thereby generating a set of ligation products;

(c) contacting under hybridization conditions the set of ligation products with one or more probe species, wherein each probe species comprises a sequence complementary to a ribosomal RNA (rRNA) sequence, thereby generating a mixture of hybridized ligation products and unhybridized ligation products; and

(d) separating the hybridized ligation products from the unhybridized ligation products.

C1.2 The method of embodiment C1 or C1.1, wherein, in the set of hybridization products in

(a)(3)(ii), a 5' end of the molecule of the first oligonucleotide is adjacent to a 3' end of the first ssNA terminal region and a 3' end of the molecule of the second oligonucleotide is adjacent to a 5' end of the second ssNA terminal region.

C2. The method of any one of embodiments C1-C1.1, wherein the nucleic acid composition comprises single-stranded complementary DNA (ssDNA).

C3. The method of embodiment C2, comprising prior to the combining in (a), generating the ssDNA from single-stranded ribonucleic acid (ssRNA).

5 C4. The method of embodiment C3, wherein generating the ssDNA comprises contacting the ssRNA with a primer and an agent comprising a reverse transcriptase activity, thereby generating a DNA-RNA duplex.

C5. The method of embodiment C4, wherein generating the ssDNA comprises contacting the DNA-RNA duplex with an agent comprising an RNase activity, thereby digesting the RNA and
10 generating an ssDNA product.

C6. The method of embodiment C5, further comprising purifying the ssDNA product.

C7. The method of embodiment C5 or C6, wherein the agent comprising a reverse transcriptase activity also comprises an RNase activity.

C8. The method of embodiment C7, wherein the agent is an M-MuLV reverse transcriptase.

15 C9. The method of any one of embodiments C4-C8, wherein the primer is chosen from one or more of a random hexamer primer, a random octamer primer, and a poly(T) primer.

C10. The method of any one of embodiments C1-C9, wherein the one or more probe species are chosen from one or more probes comprising nucleic acid sequences of SEQ ID NOs: 1-174.

C10.1 The method of any one of embodiments C1-C10, wherein each probe species further
20 comprises a member of a binding pair.

C11. The method of embodiment C10.1, wherein the binding pair is biotin and streptavidin.

C12. The method of any one of embodiments C1-C11, wherein the separating in (d) comprises a probe pull-down method.

C13. The method of any one of embodiments C1-C12, wherein the first oligonucleotide comprises
25 a 5' phosphate, and the second oligonucleotide comprises a 3' phosphate.

C14. The method of embodiment C13, wherein (d) comprises:

contacting the set of hybridization products with one or more agents comprising a ligase activity, thereby generating a first subset of ligation products, wherein:

(i) the adjacent ends of the first oligonucleotide and the first ssNA terminal region
30 are ligated; and

(ii) the adjacent ends of the second oligonucleotide and the second ssNA terminal region are not ligated; and

contacting the first subset of ligation products with one or more agents comprising a 3' phosphatase activity, thereby generating a second subset of ligation products wherein the adjacent ends of the second oligonucleotide and the second ssNA terminal region are ligated.

C15. The method of any one of embodiments C1-C14, further comprising prior to the combining in (a), contacting the ssNA with one or more agents comprising a phosphatase activity under conditions in which the ssNA is dephosphorylated, thereby generating dephosphorylated ssNA.

C16. The method of any one of embodiments C1-C15, further comprising prior to the combining in (a), contacting the ssNA with single-stranded nucleic acid binding protein (SSB) to produce SSB-bound ssNA.

C17. The method of any one of embodiments C1-C14, further comprising prior to the combining in (a), contacting the ssNA with an agent comprising a phosphatase activity under conditions in which the ssNA is dephosphorylated, and contacting the ssNA with single-stranded nucleic acid binding protein (SSB) thereby generating SSB-bound dephosphorylated ssNA.

C18. The method of embodiment C17, wherein the ssNA is simultaneously contacted with (i) SSB and (ii) an agent comprising a phosphatase activity.

C19. The method of any one of embodiments C1-C18, wherein prior to the combining in (a), each of the first scaffold polynucleotide species is hybridized to a first oligonucleotide to form a plurality of first scaffold duplex species, and/or each of the second scaffold polynucleotide species is hybridized to a second oligonucleotide to form a plurality of second scaffold duplex species.

C20. The method of any one of embodiments C1-C19, wherein the one or more agents comprising a ligase activity comprise a T4 DNA ligase.

C21. The method of any one of embodiments C14-C20, wherein the one or more agents comprising a 3' phosphatase activity further comprise a kinase activity.

C22. The method of any one of embodiments C14-C21, wherein the one or more agents comprising a 3' phosphatase activity comprise a polynucleotide kinase (PNK).

C23. The method of any one of embodiments C1-C22, wherein the ssNA hybridization region of each of the first polynucleotide species is different than the ssNA hybridization region in other first polynucleotide species in the plurality of first polynucleotide species.

C24. The method of any one of embodiments C1-C23, wherein the ssNA hybridization region of each of the second polynucleotide species is different than the ssNA hybridization region in other second polynucleotide species in the plurality of second polynucleotide species.

5 C25. The method of any one of embodiments C1-C24, wherein the ssNA hybridization region in the first scaffold polynucleotide species and/or second scaffold polynucleotide species comprises a random sequence.

C26. The method of any one of embodiments C1-C25, further comprising denaturing the set of ligation products, thereby generating single-stranded ligation products.

10 C27. The method of any one of embodiments C1-C26, further comprising after (d) amplifying the unhybridized ligation products, thereby generating amplified unhybridized ligation products.

C28. The method of embodiment C27, further comprising sequencing the amplified unhybridized ligation products.

C29. The method of any one of embodiments C1-C26, further comprising after (d) denaturing the hybridized ligation products, thereby generating denatured hybridized ligation products.

15 C30. The method of embodiment C29, further comprising amplifying the denatured hybridized ligation products, thereby generating amplified denatured hybridized ligation products.

C31. The method of embodiment C30, further comprising sequencing the amplified denatured hybridized ligation products.

20 C32. The method of any one of embodiments C1-C31, wherein the nucleic acid composition, the first oligonucleotide, the plurality of first scaffold polynucleotide species, the second oligonucleotide, and the plurality of second scaffold polynucleotide species, are combined simultaneously in (a).

C33. The method of any one of embodiments C1-C32, wherein (a) and (b) are performed simultaneously.

D1. A kit comprising:

25 (a) a first oligonucleotide, a plurality of first scaffold polynucleotide species, a second oligonucleotide, and a plurality of second scaffold polynucleotide species, wherein:

(ii) each polynucleotide in the plurality of first scaffold polynucleotide species comprises an ssNA hybridization region and a first oligonucleotide hybridization region; and

30 (ii) each polynucleotide in the plurality of second scaffold polynucleotide species comprises an ssNA hybridization region and a second oligonucleotide hybridization region;

(b) one or more agents comprising a ligase activity; and

(c) one or more probe species, wherein each probe species comprises a sequence complementary to an RNA sequence.

D1.1 A kit comprising:

- 5 (a) a first oligonucleotide, a plurality of first scaffold polynucleotide species, a second oligonucleotide, and a plurality of second scaffold polynucleotide species, wherein:
 - (ii) each polynucleotide in the plurality of first scaffold polynucleotide species comprises an ssNA hybridization region and a first oligonucleotide hybridization region; and
 - (ii) each polynucleotide in the plurality of second scaffold polynucleotide species comprises an ssNA hybridization region and a second oligonucleotide hybridization region;
- 10 (b) one or more agents comprising a ligase activity; and
- (c) one or more probe species, wherein each probe species comprises a sequence complementary to a ribosomal RNA (rRNA) sequence.

D2. The kit of embodiment D1 or D1.1, further comprising a primer and an agent comprising a reverse transcriptase activity.

- 15 D3. The kit of any one of embodiments D1-D2, further comprising an agent comprising an RNase activity.

D4. The kit of embodiment D2 or D3, wherein the agent comprising a reverse transcriptase activity also comprises an RNase activity.

D5. The kit of embodiment D4, wherein the agent is an M-MuLV reverse transcriptase.

- 20 D6. The kit of any one of embodiments D2-D5, wherein the primer is chosen from one or more of a random hexamer primer, a random octamer primer, and a poly(T) primer.

D7. The kit of any one of embodiments D1-D6, wherein the one or more probe species are chosen from one or more probes comprising nucleic acid sequences of SEQ ID NOs: 1-174.

- 25 D7.1 The kit of any one of embodiments D1-D7, wherein each probe species further comprises a member of a binding pair.

D8. The kit of embodiment D7.1, wherein the binding pair is biotin and streptavidin.

D9. The kit of any one of embodiments D1-D8, wherein the first oligonucleotide comprises a 5' phosphate, and the second oligonucleotide comprises a 3' phosphate.

- 30 D10. The kit of embodiment D9, further comprising one or more agents comprising a 3' phosphatase activity.

D11. The kit of embodiment D10, wherein the one or more agents comprising a 3' phosphatase activity further comprise a kinase activity.

D12. The kit of embodiment D10 or D11, wherein the one or more agents comprising a 3' phosphatase activity comprise a polynucleotide kinase (PNK).

5 D13. The kit of any one of embodiments D1-D12, further comprising single-stranded nucleic acid binding protein (SSB).

D14. The kit of any one of embodiments D1-D13, wherein the one or more agents comprising a ligase activity comprise a T4 DNA ligase.

10 D15. The kit of any one of embodiments D1-D14, wherein the ssNA hybridization region of each of the first polynucleotide species is different than the ssNA hybridization region in other first polynucleotide species in the plurality of first polynucleotide species.

D16. The kit of any one of embodiments D1-D15, wherein the ssNA hybridization region of each of the second polynucleotide species is different than the ssNA hybridization region in other second polynucleotide species in the plurality of second polynucleotide species.

15 D17. The kit of any one of embodiments D1-D16, wherein the ssNA hybridization region in the first scaffold polynucleotide species and/or second scaffold polynucleotide species comprises a random sequence.

Examples

The examples set forth below illustrate certain implementations and do not limit the technology.

20 *Example 1: Modified single-stranded library prep (SRSLY 2.0)*

In previous single-stranded library prep (SRSLY) workflows, ligation takes place with simultaneous addition of P5 and P7 adapters in the presence of T4 DNA ligase and T4 PNK in order to perform 5' phosphorylation of the template DNA, 3' dephosphorylation of the template DNA, and adapter ligation all at once. Contrary to this, the modified single-stranded library prep (SRSLY 2.0) is two-
25 fold:

1. The P5 SRSLY scaffold adapter now contains a 3' phosphate on the ligatable 3' end of the adapter. The 3' phosphate acts as a reversible 3' ligation blocking modification. In previous single-stranded library prep (SRSLY) configurations, the P5 adapter just has a free 3' OH.

30 2. The SRSLY ligation reagents and conditions are almost the same as in previous single-stranded library prep (SRSLY) workflows with one significant exception – the absence of PNK in the initial reaction.

The P5 adapter and the P7 adapter are still added simultaneously during the start of the ligation. However, at this point only the P7 adapter can ligate because it has a 5' phosphate. The P5 SRSLY adapter is essentially inert in the reaction at this step because it has both a 5' blocking modification (as per previous single-stranded library prep (SRSLY) workflows) and the added 3' phosphate reversible blocking modification. This setup drives the SRSLY P7 adapter to only ligate to template DNA and not to create dimers. After x amount of ligation time, PNK is spiked into the reaction mixture. This liberates the 3' phosphate from the SRSLY P5 adapter and allows it to ligate to the template DNA. Since the P7 adapter quantity is now vastly reduced in the reaction, less dimers and more product forms overall.

Another modification of the single-stranded library prep (SRSLY) workflow is use of a phosphatase upstream of SRSLY adapter ligation in order to "end prep" or "pre-treat" the template DNA for SRSLY ligation. This phosphatase step strips all 5' and 3' phosphates from the template DNA. Different template DNA fragments have varying 5' and 3' phosphorylation status. The phosphatase "pre-treatment" works to make all, or substantially all, template DNA fragments more uniform and behave more predictably in a single-stranded library prep. Initial observations indicate phosphatase pre-treatment of the template DNA does indeed increase yield and library complexity. Theoretically stripping template DNA of all its phosphates and then performing the SRSLY ligation as defined above should decrease chimeras since the template DNA should not be able to self-ligate without phosphates until the PNK is added. Unlike previous workflows that included phosphatase pre-treatment, the phosphatase pre-treatment here is incorporated into the single-stranded binding protein (SSB) step.

When the phosphatase pre-treatment is combined with the above adapters the overall protocol drives only SRSLY P7 adapter ligation during the first part of the ligation reaction. After x amount of time when the PNK is added, it functions to strip the 3' phosphate from the SRSLY P5 adapter as well as create a 5' phosphate on the template DNA such that the P5 adapter and the T4 ligase can now work in concert to ligate the SRSLY P5 adapter to the template DNA. The end result of all of this is fewer dimers, higher yield, and overall a more robust and better SRSLY protocol.

Initial tests using the above modified single-stranded library prep (SRSLY 2.0) workflow were performed with 250 pg of cfDNA as input and the following was observed:

- Greater than 50% dimer reduction
- 50-75% improvement in yield

The modified single-stranded library prep (SRSLY 2.0) workflow also allows the following:

- Combining the adapters into one tube for customers without fear of creating dimers since the P5 SRSly adapter is essentially inert

- Fewer pipetting steps and more pipetting volume for customers

- Increased total SRSly ligation volume to 75µl (currently 50µl). This is advantageous

5 because increased volume provides a less viscous reaction master mix to customers (this was not optimal in previous SRSly workflows because increasing the SRSly ligation volume decreased yield in previous SRSly workflows, but the modified workflow described herein overcomes such yield deficits).

10 Example 2: *Single-stranded library prep (SRSly or SRSly 2.0) and RNA-Seq with rRNA sequence depletion*

In this Example, a single-stranded library prep method for RNA fragments was performed as described herein and as previously described (e.g., WO 2020/206143 and/or WO 2021/262805) using a modified workflow that included a ribosomal RNA (rRNA) sequence depletion step after the SRSly adapter ligation step. The modified workflow is shown in Fig. 3 and includes:

- 15 1) RNA fragmentation and cDNA synthesis;
- 2) RNase H treatment;
- 3) SRSly adapter ligation;
- 4) rRNA probe hybridization;
- 5) probe pull-down; and
- 20 6) index PCR.

In some workflows, a second round of rRNA probe hybridization and probe pull-down is performed. The workflow described in this Example is useful for about 10 ng to about 250 ng total RNA input. In this workflow, ribosomal sequence depletion (i.e., depletion of sscDNA corresponding to rRNA sequences) takes place after SRSly ligation and uses a biotin/streptavidin interaction for

25 subtractive hybridization. One benefit to this approach is that as SRSly is a single-stranded method there is no need for 2nd strand cDNA synthesis, which reduces cost and process time. Furthermore, by performing the depletion after ligation, there are some additional benefits. A larger amount of input nucleic acid into the adapter ligation step can reduce the formation of adapter dimers as there is more template for the adapters to ligate to, reducing the chance of adapter-

30 adapter ligation events. Therefore, performing depletion after ligation improves the quality of the resultant library as there is more template to ligate with adapters and thereby reduce dimer formation.

The rRNA probes target cytoplasmic ribosomes (28S, 18s, 5s, 5.3s), mitochondrial ribosomes (12s, 16s), and RN7SL1 and RN7SL2, which are highly expressed transcripts for ribosomal associated proteins. The probe set used in this Example included:

- about 30 probes to target 18S subunit
- 5 about 80 probes to target 28S subunit
- 2 probes to target 5S subunit
- 2 probes to target 5.8S subunit
- 14 additional probes to target 28S and 18S overrepresented sequences
- 16 probes to target 12S mtrRNA
- 10 26 probes to target 16S mtrRNA
- 4 probes to target RN7SL
- 7 Probes (28S/18S) removed

Probe Name	Sequence	SEQ ID NO
18S_1	/5Biosg/GTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGA AGGATC	1
18S_2	/5Biosg/CGGGGTCTGGCCCCACGGCCCTGGCGGAGCGCTGA GAAGACG	2
18S_3	/5Biosg/GCCCTTTGTACACACCGCCCGTCGCTACTACCGA TTGGAT	3
18S_4	/5Biosg/TATTCCTCATGAACGAGGAATTCCCAGTAAGTGCG GGTCA	4
18S_5	/5Biosg/GCCTACCCTACGCCGGCAGGCGCGGGTAACCCG TTGAACC	5
18S_6	/5Biosg/CAATAACAGGTCTGTGATGCCCTTAGATGTCCGG GGCTGC	6
18S_7	/5Biosg/CGACCCCCGAGCGGTCTGGCGTCCCCCAACTTCTT AGAGGG	7
18S_8	/5Biosg/GTCTGGTTAATTCCGATAACGAACGAGACTCTGGC ATGCT	8
18S_9	/5Biosg/ACAGATTGATAGCTCTTTCTCGATTCCGTGGGTGG TGGTG	9
18S_10	/5Biosg/TTTGACTCAACACGGGAAACCTCACCCGGCCCCGG ACACGG	10
18S_11	/5Biosg/AAGGAATTGACGGAAGGGCACCACCAGGAGTGGA GCCTG	11

Probe Name	Sequence	SEQ ID NO
18S_12	/5Biosg/CGCCGGGCAGCTTCCGGGAAACCAAAGTCTTTGG GTTCCG	12
18S_13	/5Biosg/GTAGTTCCGACCATAAACGATGCCGACCGGCGAT GCGGC	13
18S_14	/5Biosg/AGAACGAAAGTCGGAGGTTCTGAAGACGATCAGAT ACCGTC	14
18S_15	/5Biosg/TGGACCGGCGCAAGACGGACCAGAGCGAAAGCA TTTGCCA	15
18S_16	/5Biosg/TCGGAACTGAGGCCATGATTAAGAGGGACGGCCG GGGGCA	16
18S_17	/5Biosg/CAGGCCCGAGCCGCCTGGATACCGCAGCTAGGA ATAATGG	17
18S_18	/5Biosg/CTCGATGCTCTTAGCTGAGTGTCCCGCGGGGCCC GAAGCG	18
18S_19	/5Biosg/AGCGGGCGGGCGGTCCGCCGCGAGGCGAGCCA CCGCCCCG	19
18S_20	/5Biosg/ATTCCAGCTCCAATAGCGTATATTAAAGTTGCTGC AGTT	20
18S_21	/5Biosg/ATCCTTTAACGAGGATCCATTGGAGGGCAAGTCTG GTGC	21
18S_22	/5Biosg/ACAGGACTCTTTCGAGGCCCTGTAATTGGAATGAG TCCA	22
18S_23	/5Biosg/TCCAAGGAAGGCAGCAGGCGCGCAAATTACCCAC TCCCGA	23
18S_24	/5Biosg/CATGGTGACCACGGGTGACGGGGAATCAGGGTTC GATTCC	24
18S_25	/5Biosg/CCCCCGTGGCGGCGACGACCCATTGAAACGTCT GCCCTA	25
18S_26	/5Biosg/CCGGGGGGCGGGCGCCGGCGGCTTTGGTGACTC TAGATAA	26
18S_27	/5Biosg/TGCGTGCAATTTATCAGATCAAACCAACCCGGTCA GCCCC	27
18S_28	/5Biosg/GGTAATTCTAGAGCTAATACATGCCGACGGGCGC TGACCC	28
18S_29	/5Biosg/ATGGCTCATTAAATCAGTTATGGTTCCTTTGGTCG CTCGC	29
18S_30	/5Biosg/GCTTGTCTCAAAGATTAAGCCATGCATGTCTGAGT ACGCA	30
Anti_P5	/5Biosg/ACACTCTTTCCTACACGACGCTCTTCCGATCT	31
Anti_P7	/5Biosg/AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC	32

Probe Name	Sequence	SEQ ID NO
hum_28s_depl_001	/5Biosg/TTTCGTACGTAGCAGAGCAGCTCCCTCGCTGCGA TCTATT	33
hum_28s_depl_002	/5Biosg/GTCACGCACCGCACGTTTCGTGGGGAACCTGGCG CTAAACC	34
hum_28s_depl_003	/5Biosg/GTCCGGTGCGGAGTGCCCTTCGTCCTGGGAAACG GGGCG	35
hum_28s_depl_004	/5Biosg/CGCGCGGGAGGGCGCGTGCCCCGCCGCGCGCC GGGACCG	36
hum_28s_depl_005	/5Biosg/CGGATAGCCGGTCCCCCGCCTGTCCCCGCCGGC GGGCCG	37
hum_28s_depl_006	/5Biosg/TGAACGCCTCTAAGTCAGAATCCCGCCCAGGCGG AACGAT	38
hum_28s_depl_007	/5Biosg/TTCAGACATTTGGTGTATGTGCTTGGCTGAGGAGC CAATG	39
hum_28s_depl_008	/5Biosg/GGTTAGTTTTACCCTACTGATGATGTGTTGTTGCC ATGGT	40
hum_28s_depl_009	/5Biosg/GTTGGATTGTTCACCCACTAATAGGGAACGTGAGC TGGGT	41
hum_28s_depl_010	/5Biosg/CATAGCGACGTCGCTTTTTGATCCTTCGATGTCGG CTCTT	42
hum_28s_depl_011	/5Biosg/AAGTTACCACAGGGATAACTGGCTTGTGGCGGCC AAGCG	43
hum_28s_depl_012	/5Biosg/ATACAGACCGTGAAAGCGGGGCCTCACGATCCTT CTGAC	44
hum_28s_depl_013	/5Biosg/CTCAGGGAGGACAGAAACCTCCCGTGAGCAGAA GGGCAA	45
hum_28s_depl_014	/5Biosg/TGCCAGGTGGGGAGTTTGAAGTGGGGCGGTACACC TGTCAA	46
hum_28s_depl_015	/5Biosg/CGCCAAGCGCCCGGCCGCGCGCCGGCCGGGCG CGACCCG	47
hum_28s_depl_017	/5Biosg/GGCGGGGTCCGCCGGCCCTGCGGGCCGCGCGGT GAAATAC	48
hum_28s_depl_018	/5Biosg/AGAGACATGAGAGGTGTAGAATAAGTGGGAGGCC CCCGGC	49
hum_28s_depl_019	/5Biosg/GGGAACGGGCTTGGCGGAATCAGCGGGGAAAGA AGACCCT	50
hum_28s_depl_020	/5Biosg/GCGCATGAATGGATGAACGAGATTCCCACTGTCC CTACCT	51
hum_28s_depl_021	/5Biosg/GGCGGGAGTAACTATGACTCTCTTAAGGTAGCCA AATGCC	52
hum_28s_depl_022	/5Biosg/GCGATGTGATTTCTGCCAGTGCTCTGAATGTCAA AGTGA	53

Probe Name	Sequence	SEQ ID NO
hum_28s_depl_023	/5Biosg/TTAAAACAAAGCATCGCGAAGGCCCGCGGCGGGT GTTGAC	54
hum_28s_depl_024	/5Biosg/CGCCTCGCCTCGGCCGGCGCCTAGCAGCCGACT TAGAACT	55
hum_28s_depl_025	/5Biosg/CCCCGCCGGGTCCGCCCCCGGGGCCGCGGTTCC GCGCGG	56
hum_28s_depl_026	/5Biosg/CTCCTCGGCGCGCGGCCGGCGGCGGCGGCAGGC GGCGGAG	57
hum_28s_depl_027	/5Biosg/CGGGGCGGGGCGGTTCGTCCCCCGCCCTACCC CCCCGG	58
hum_28s_depl_028	/5Biosg/CACGTCTCGTCGCGCGCGCTCCGCTGGGGGCG GGGAGCG	59
hum_28s_depl_029	/5Biosg/CAGCTGCGGCGGGCGTCGCGGCCGCCCCCGGG GAGCCCG	60
hum_28s_depl_030	/5Biosg/CGGCGACTCTGGACGCGAGCCGGGCCCTTCCCG TGGATCG	61
hum_28s_depl_031	/5Biosg/CGGGGGCACGGTCCCCCGCGAGGGGGGCCCGG GCACCCG	62
hum_28s_depl_032	/5Biosg/CGGCGGCGGGGGGAGAAGGGTCGGGGCGGCAG GGGCCGG	63
hum_28s_depl_033	/5Biosg/CGCTCCCCGTCTCCCCCTCCCCGGGGGAGCG CCGCGTG	64
hum_28s_depl_034	/5Biosg/CCACCCGCGCGCGCGCTCGCTCCCTCCCCGC CCCGCG	65
hum_28s_depl_035	/5Biosg/CGCGGCTGGACGAGGCGCCGCCGCCCCCCCCAC GCCCGG	66
hum_28s_depl_036	/5Biosg/TTCGGGATAAGGATTGGCTCTAAGGGCTGGGTCTG GTCGGG	67
hum_28s_depl_037	/5Biosg/CCAAGGTGAACAGCCTCTGGCATGTTGGAACAAT GTAGGT	68
hum_28s_depl_038	/5Biosg/GGCCCTTGAAAATCCGGGGGAGAGGGTGTAATC TCGCGC	69
hum_28s_depl_039	/5Biosg/TTCGCCCCGAGAGAGGGGCCCGTGCTTGAAA GCGTCGC	70
hum_28s_depl_040	/5Biosg/CGGAGAAGCCGGCGGGAGCCCCGGGGAGAGTTC TCTTTTC	71
hum_28s_depl_041	/5Biosg/TCAGATCCCCGAATCCGGAGTGGCGGAGATGGGC GCCGCG	72
hum_28s_depl_042	/5Biosg/GGGCGAGCGCCGTTCCGAAGGGACGGGCGATGG CCTCCGT	73
hum_28s_depl_043	/5Biosg/AGGCCGAAGTGGAGAAGGGTTCCATGTGAACAGC AGTTG	74

Probe Name	Sequence	SEQ ID NO
hum_28s_depl_044	/5Biosg/GTGGAGCCGCGCAGGTGCAGATCTTGGTGGTAGTAGCAA	75
hum_28s_depl_045	/5Biosg/CGCGGACGCTACGCCGCGACGAGTAGGAGGGCCGCTGCGG	76
hum_28s_depl_046	/5Biosg/CACGCCTCCTCCCCTCCTCCCGCCCACGCCCCGC TCCCCG	77
hum_28s_depl_047	/5Biosg/TGGTGTGCGTCCGAGGGCGGCGGCGGCGGCGGCGGCGGG	78
hum_28s_depl_048	/5Biosg/CCATACCCGGCCGTCGCCGGCAGTCGAGAGTGGACGGGAG	79
hum_28s_depl_049	/5Biosg/TCAACTAGCCCTGAAAATGGATGGCGCTGGAGCGTCGGGC	80
hum_28s_depl_050	/5Biosg/TATAGACAGCAGGACGGTGGCCATGGAAGTCGGAATCCGC	81
hum_28s_depl_051	/5Biosg/GATGAACCGAACGCCGGGTAAAGGCGCCCGATGCGACGC	82
hum_28s_depl_052	/5Biosg/CGTGGAGCCGGGCGTGGAATGCGAGTGCCTAGTGGGCCA	83
hum_28s_depl_053	/5Biosg/GGTCTTGGGGCCGAAACGATCTCAACCTATTCTCAAACTT	84
hum_28s_depl_054	/5Biosg/CGCTCTCGCAGACCCGACGCACCCCCGCCACGCAGTTTT	85
hum_28s_depl_055	/5Biosg/TAATCGAACCATCTAGTAGCTGGTTCCCTCCGAAGTTTCC	86
hum_28s_depl_056	/5Biosg/CTCTGGTGGAGGTCCGTAGCGGTCCTGACGTGCA AATCGG	87
hum_28s_depl_057	/5Biosg/GTGTTAGGACCCGAAAGATGGTGA ACTATGCCTGGCAGG	88
hum_28s_depl_058	/5Biosg/CCGTCTCGCCCGCCGCGCCGGGGAGGTGGAGCACGAGCG	89
hum_28s_depl_059	/5Biosg/CGAGGTGGGATCCCGAGGCCTCTCCAGTCCGCCGAGGGCG	90
hum_28s_depl_060	/5Biosg/TGCGCGAGTCGGGGGCTCGCACGAAAGCCGCCGTGGCGCA	91
hum_28s_depl_061	/5Biosg/GTCGGCGGCGACGTCGGCTACCCACCCGACCCGTCTTGA	92
hum_28s_depl_062	/5Biosg/CTCTCGGGGCCACGCGCGCGTCCCCCGAAGAGGGGGACG	93
hum_28s_depl_063	/5Biosg/CGGACTGTCCCCAGTGCGCCCGGGCGGGTTCGGGCCGTCTG	94
hum_28s_depl_064	/5Biosg/CGGCGACGGGGGGGGTGCCGCGCGCGGGTTCGGGGGCGG	95

Probe Name	Sequence	SEQ ID NO
hum_28s_depl_065	/5Biosg/TCCCACGGCGCGACCGCTCTCCCACCCCTCCTCC CCGCG	96
hum_28s_depl_066	/5Biosg/CGCGGGGGCGCGCCGGCGTCTCCTCGTGGGGG GGCCGGG	97
hum_28s_depl_067	/5Biosg/CTCCCGGCGCCCACCCCCGCGGGGAATCCCCCG CGAGGG	98
hum_28s_depl_068	/5Biosg/CGGCAGCAGCACTCGCCGAATCCCGGGGCCGAG GGAGCG	99
hum_28s_depl_069	/5Biosg/GCGGCGGTGGCGGCGGCGGCGGCGGCGGG ACCGAAA	100
hum_28s_depl_070	/5Biosg/CGCGTCCTCCCTCGGGAGGGCGCGCGGGTGGG GCGGCG	101
hum_28s_depl_071	/5Biosg/AAGGTGGCTCGGGGGGCCCGTCCGTCCGTCCG TCCGTCC	102
hum_28s_depl_072	/5Biosg/CGGCGGTGCGCCGCGACCGGCTCCGGGACGGCT GGGAAG	103
hum_28s_depl_073	/5Biosg/CCGACCGGCGACCGGCCCGCCGGGCGCATTT CCACCG	104
hum_28s_depl_074	/5Biosg/TGGGCGGGCGGGGCCGGGGTGGGGTCGGCGG GGGACCG	105
hum_28s_depl_075	/5Biosg/CTCCTCCTCCCCGGAGGGGGCGGGCTCCGGCGG GTGCGG	106
hum_28s_depl_076	/5Biosg/CGGATCTTTCCCGCCCCCGTTCCTCCCGACCCC TCCAC	107
hum_28s_depl_077	/5Biosg/CGCAGTCCGCCCGGAGGATTCAACCCGGCGGCG GGTCCG	108
hum_28s_depl_078	/5Biosg/AGAGGGCGTGAAACCGTTAAGAGGTAAACGGGTG GGGTCC	109
hum_28s_depl_081	/5Biosg/CTTCTGATCGAGGCCAGCCCGTGGACGGTGTGA GGCCG	110
hum_28s_depl_082	/5Biosg/GCGGCGGGGCGCGGGACATGTGGCGTACGGAAG ACCCGCT	111
hum_28s_depl_083	/5Biosg/CCAGGATTCCCTCAGTAACGGCGAGTGAACAGGG AAGAGC	112
hum_28s_depl_084	/5Biosg/GCGACCTCAGATCAGACGTGGCGACCCGCTGAAT TTAAGC	113
hum_5s_depl_001	/5Biosg/CTAAGCAGGGTCGGGCCTGGTTAGTACTTGGATG GGAGAC	114
hum_5s_depl_002	/5Biosg/GTCTACGGCCATACCACCCTGAACGCGCCCGATC TCGTC	115
hum_5-8s_depl_001	/5Biosg/ATCATCGACACTTCGAACGCACTTGCGGCCCGG GTTCT	116

Probe Name	Sequence	SEQ ID NO
hum_5-8s_depl_002	/5Biosg/TCACTCGGCTCGTGCGTCGATGAAGAACGCAGCTAGCTGC	117
18S_28S_seqs_002	/5Biosg/GACCAAGGAGTCTAACACGTGCGCGAGTCGGGGGCTCGCA	118
18S_28S_seqs_003	/5Biosg/GTGGCGACCCGCTGAATTTAAGCATATTAGTCAGCGGAGG	119
18S_28S_seqs_004	/5Biosg/ATACCACTACTCTGATCGTTTTTTCACTGACCCGGTGAGG	120
18S_28S_seqs_005	/5Biosg/CGGAGTCGGGTGCTTGGAATGCAGCCCAAAGCGGGTGG	121
18S_28S_seqs_006	/5Biosg/CGAGACCGATAGTCAACAAGTACCGTAAGGGAAAGTTGAA	122
18S_28S_seqs_007	/5Biosg/CGCGACCCGCTCCGGGGACAGTGCCAGGTGGGGAGTTTG	123
18S_28S_seqs_009	/5Biosg/GTGACTCTAGATAACCTCGGGCCGATCGCACGCCCCCCGT	124
18S_28S_seqs_010	/5Biosg/CTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAA TAGCGT	125
18S_28S_seqs_011	/5Biosg/TCCTGCCAGTAGCATATGCTTGTCTCAAAGATTAA GCCA	126
18S_28S_seqs_013	/5Biosg/ACCCCCCTCGCGGGGGGGATGCGTGCATTTATCA GATCAA	127
rn7sl1_2_depl_oligos_001	/5Biosg/CAGTAGTGGGATCGCGCCTGTGAATAGCCACTGC ACTCCA	128
rn7sl1_2_depl_oligos_002	/5Biosg/ACCAGGTTGCCTAAGGAGGGGTGAACCGGCCCA GGTCCGA	129
rn7sl1_2_depl_oligos_003	/5Biosg/TGCGCTATGCCGATCGGGTGTCCGCACTAAGTTC GGCATC	130
rn7sl1_2_depl_oligos_004	/5Biosg/CTGTAGTCCCAGCTACTCGGGAGGCTGAGGCTGG AGGAT	131
rn7sl1_2_depl_oligos_005	/5Biosg/CTGTAGTCCCAGCTACTCGGGAGGCTGAGGTGGG AGGAT	132
12s_oligo_001	/5Biosg/AGTCGTAACATGGTAAGTGTAAGTGGAAAGTGCACT TGGAC	133
12s_oligo_002	/5Biosg/GTCACCCTCCTCAAGTATACTTCAAAGGACATTTA ACTA	134
12s_oligo_003	/5Biosg/CTAAGAGTAGAGTGCTTAGTTGAACAGGGCCCTG AAGCGC	135
12s_oligo_004	/5Biosg/GCCCTTATGAACTTAAGGGTCGAAGGTGGATTTA GCAGT	136
12s_oligo_005	/5Biosg/GTTAGGTCAAGGTGTAGCCCATGAGGTGGCAAGA AATGGG	137

Probe Name	Sequence	SEQ ID NO
12s_oligo_006	/5Biosg/CTATATACCGCCATCTTCAGCAAACCCTGATGAAG GCTAC	138
12s_oligo_007	/5Biosg/ATCGATAAACCCCGATCAACCTCACCACCTCTTGC TCAGC	139
12s_oligo_008	/5Biosg/AACTCAAAGGACCTGGCGGTGCTTCATATCCCTCT AGAG	140
12s_oligo_009	/5Biosg/TCAACAAAAGTCTCGCCAGAACACTACGAGCCA CAGCTT	141
12s_oligo_010	/5Biosg/AGACCCAAAGTGGGATTAGATACCCCACTATGCTT AGCCC	142
12s_oligo_011	/5Biosg/ACTCCAGTTGACACAAAATAGACTACGAAAGTGGC TTTA	143
12s_oligo_012	/5Biosg/AAGCCGGCGTAAAGAGTGTTTTAGATCACCCCTC CCCAA	144
12s_oligo_013	/5Biosg/CTATACTAACCCAGGGTTGGTCAATTCGTGCCA GCCAC	145
12s_oligo_014	/5Biosg/TTAGCCTAGCCACACCCCCACGGGAAACAGCAGT GATTAA	146
12s_oligo_015	/5Biosg/GGAACAAGCATCAAGCACGCAGCAATGCAGCTCA AAACGC	147
12s_oligo_016	/5Biosg/TCTTAGTAAGATTACACATGCAAGCATCCCCGTT CAGTG	148
16s_oligo_001	/5Biosg/ACTTAGTATTATACCCACACCCACCCAAGAACAGG GTTT	149
16s_oligo_002	/5Biosg/AGAGAAATAAGGCCTACTTCACAAAGCGCCTTCCC CCGT	150
16s_oligo_003	/5Biosg/CTGAGTTCAGACCGGAGTAATCCAGGTCGGTTTCT ATCTA	151
16s_oligo_004	/5Biosg/GATCAGGACATCCCGATGGTGCAGCCGCTATTAA AGGTTC	152
16s_oligo_005	/5Biosg/ACCCTAGGGATAACAGCGCAATCCTATTCTAGAGT CCATA	153
16s_oligo_006	/5Biosg/CTTCACCAGTCAAAGCGAACTACTATACTCAATTG ATCCA	154
16s_oligo_007	/5Biosg/AAATTCGGTTGGGGCGACCTCGGAGCAGAACCC AACCT	155
16s_oligo_008	/5Biosg/GTACCTAACAAACCCACAGGTCCTAACTACCAAA CCTGC	156
16s_oligo_009	/5Biosg/GGCGGGCATAACACAGCAAGACGAGAAGACCCTA TGGAGC	157
16s_oligo_010	/5Biosg/GTATGAATGGCTCCACGAGGGTTCAGCTGTCTCTT ACTTT	158

Probe Name	Sequence	SEQ ID NO
16s_oligo_011	/5Biosg/ATGTTTAACGGCCGCGGTACCCTAACCGTGCAAA GGTAGC	159
16s_oligo_012	/5Biosg/AAACATCACCTCTAGCATCACCAGTATTAGAGGCA CCGC	160
16s_oligo_013	/5Biosg/GTAAAAGGAACTCGGCAAATCTTACCCCGCCTGTT TACCA	161
16s_oligo_014	/5Biosg/AGTCATTATTACCCTCACTGTCAACCCAACACAGG CATG	162
16s_oligo_015	/5Biosg/GCCTGCGTCAGATTAACAACTGAACTGACAATTA ACAGC	163
16s_oligo_016	/5Biosg/ATGTTAGTATAAGTAACATGAAAACATTCTCCTCCG CAT	164
16s_oligo_017	/5Biosg/ACTGAACTCCTCACACCCAATTGGACCAATCTATC ACCCT	165
16s_oligo_018	/5Biosg/GCAGCCACCAATTAAGAAAGCGTTCAAGCTCAACA CCCAC	166
16s_oligo_019	/5Biosg/AGCTCTTTGGACACTAGGAAAAAACCTTGTAGAGA GAGTA	167
16s_oligo_020	/5Biosg/ACTTTAAATTTGCCACAGAACCCCTCTAAATCCCC TTGT	168
16s_oligo_021	/5Biosg/TATAGGTAGAGGCGACAAACCTACCGAGCCTGGT GATAGC	169
16s_oligo_022	/5Biosg/GACGAGCTACCTAAGAACAGCTAAAAGAGCACAC CCGTCT	170
16s_oligo_023	/5Biosg/ACTTTGCAAGGAGAGCCAAAGCTAAGACCCCCGA AACCA	171
16s_oligo_024	/5Biosg/AGCATAATATAGCAAGGACTAACCCCTATACCTTC TGCA	172
16s_oligo_025	/5Biosg/ACCTGGCGCAATAGATATAGTACCGCAAGGGAAA GATGA	173
16s_oligo_026	/5Biosg/CCACCTTACTACCAGACAACCTTAGCCAAACCATT TACCC	174

The above probe set was specifically optimized in a few ways to capture adapter ligated sscDNA fragments. For example, this probe set hybridizes to cDNA, not RNA. Most, if not all, of the existing probe sets and protocols hybridize to total RNA prior to cDNA synthesis. Since this workflow is applied to cDNA, the probes themselves are the reverse complement of standard probe sets that target RNA, and therefore many of the sequences in this probe set are opposite to sequences in an RNA-targeting probe set. This probe set also includes anti SRSLY -P5 and -P7 probes to deplete adapter dimers.

The probe set also contains four probes that target the RN7SL1 and RN7SL2 genes. These genes are associated with the ribonucleoprotein complex and therefore are commonly highly expressed gene transcripts, making them not especially informative to a gene expression study. By depleting these as well, valuable space on the sequencer may go to more informative reads.

5 SRSLY rRNA depletion and RNA-Seq results are shown in Fig. 4A to Fig. 8.

* * *

The entirety of each patent, patent application, publication and document referenced herein is incorporated by reference. Citation of patents, patent applications, publications and documents is not an admission that any of the foregoing is pertinent prior art, nor does it constitute any admission
10 as to the contents or date of these publications or documents. Their citation is not an indication of a search for relevant disclosures. All statements regarding the date(s) or contents of the documents is based on available information and is not an admission as to their accuracy or correctness.

The technology has been described with reference to specific implementations. The terms and expressions that have been utilized herein to describe the technology are descriptive and not
15 necessarily limiting. Certain modifications made to the disclosed implementations can be considered within the scope of the technology. Certain aspects of the disclosed implementations suitably may be practiced in the presence or absence of certain elements not specifically disclosed herein.

Each of the terms "comprising," "consisting essentially of," and "consisting of" may be replaced with
20 either of the other two terms. The term "a" or "an" can refer to one of or a plurality of the elements it modifies (e.g., "a reagent" can mean one or more reagents) unless it is contextually clear either one of the elements or more than one of the elements is described. The term "about" as used herein refers to a value within 10% of the underlying parameter (i.e., plus or minus 10%; e.g., a weight of "about 100 grams" can include a weight between 90 grams and 110 grams). Use of the term
25 "about" at the beginning of a listing of values modifies each of the values (e.g., "about 1, 2 and 3" refers to "about 1, about 2 and about 3"). When a listing of values is described the listing includes all intermediate values and all fractional values thereof (e.g., the listing of values "80%, 85% or 90%" includes the intermediate value 86% and the fractional value 86.4%). When a listing of values is followed by the term "or more," the term "or more" applies to each of the values listed (e.g., the
30 listing of "80%, 90%, 95%, or more" or "80%, 90%, 95% or more" or "80%, 90%, or 95% or more" refers to "80% or more, 90% or more, or 95% or more"). When a listing of values is described, the listing includes all ranges between any two of the values listed (e.g., the listing of "80%, 90% or 95%" includes ranges of "80% to 90%," "80% to 95%" and "90% to 95%").

Certain implementations of the technology are set forth in the claim(s) that follow(s).

What is claimed is:

1. A method of producing a nucleic acid library, comprising:

(a) combining (i) a nucleic acid composition comprising single-stranded nucleic acid (ssNA), (ii) a first oligonucleotide, (iii) a plurality of first scaffold polynucleotide species, (iv) a second oligonucleotide, and (v) a plurality of second scaffold polynucleotide species, wherein:

(1) the first oligonucleotide comprises a 5' phosphate;

(2) the second oligonucleotide comprises a 3' phosphate;

(3) each polynucleotide in the plurality of first scaffold polynucleotide species comprises an ssNA hybridization region and a first oligonucleotide hybridization region;

(4) each polynucleotide in the plurality of second scaffold polynucleotide species comprises an ssNA hybridization region and a second oligonucleotide hybridization region; and

(5) the nucleic acid composition, the first oligonucleotide, the plurality of first scaffold polynucleotide species, the second oligonucleotide, and the plurality of second scaffold polynucleotide species, are combined under conditions wherein:

(i) a molecule of the first scaffold polynucleotide species is hybridized to a first ssNA terminal region and a molecule of the first oligonucleotide, and

(ii) a molecule of the second scaffold polynucleotide species is hybridized to a second ssNA terminal region and a molecule of the second oligonucleotide, thereby forming a set of hybridization products in which a 5' end of the molecule of the first oligonucleotide is adjacent to a 3' end of the first ssNA terminal region and a 3' end of the molecule of the second oligonucleotide is adjacent to a 5' end of the second ssNA terminal region;

(b) contacting the set of hybridization products with one or more agents comprising a ligase activity, thereby generating a first set of ligation products, wherein:

(i) the adjacent ends of the first oligonucleotide and the first ssNA terminal region are ligated; and

(ii) the adjacent ends of the second oligonucleotide and the second ssNA terminal region are not ligated; and

(c) contacting the first set of ligation products with one or more agents comprising a 3' phosphatase activity, thereby generating a second set of ligation products wherein the adjacent ends of the second oligonucleotide and the second ssNA terminal region are ligated.

2. The method of claim 1, further comprising prior to the combining in (a), contacting the ssNA with one or more agents comprising a phosphatase activity under conditions in which the ssNA is dephosphorylated, thereby generating dephosphorylated ssNA.

3. The method of claim 1 or 2, further comprising prior to the combining in (a), contacting the ssNA with single-stranded nucleic acid binding protein (SSB) to produce SSB-bound ssNA.
4. The method of any one of claims 1-3, wherein the one or more agents comprising a 3' phosphatase activity further comprise a kinase activity.
5. The method of any one of claims 1-4, wherein the one or more agents comprising a 3' phosphatase activity comprise a polynucleotide kinase (PNK).
6. The method of any one of claims 1-5, wherein the nucleic acid composition comprises single-stranded DNA (ssDNA), wherein the ssDNA is derived from double-stranded DNA (dsDNA).
7. The method of claim 6, further comprising prior to the combining in (a), denaturing the dsDNA, thereby generating the ssDNA, wherein the dsDNA is simultaneously denatured and contacted with (i) SSB and (ii) an agent comprising a phosphatase activity.
8. The method of any one of claims 1-7, wherein the nucleic acid composition, the first oligonucleotide, the plurality of first scaffold polynucleotide species, the second oligonucleotide, and the plurality of second scaffold polynucleotide species, are combined simultaneously in (a).
9. The method of any one of claims 1-8, wherein (a) and (b) are performed simultaneously.
10. A kit comprising:
 - (a) a first oligonucleotide, a plurality of first scaffold polynucleotide species, a second oligonucleotide, and a plurality of second scaffold polynucleotide species, wherein:
 - (i) the first oligonucleotide comprises a 5' phosphate;
 - (ii) the second oligonucleotide comprises a 3' phosphate;
 - (iii) each polynucleotide in the plurality of first scaffold polynucleotide species comprises an ssNA hybridization region and a first oligonucleotide hybridization region; and
 - (iv) each polynucleotide in the plurality of second scaffold polynucleotide species comprises an ssNA hybridization region and a second oligonucleotide hybridization region;
 - (b) one or more agents comprising a ligase activity; and
 - (c) one or more agents comprising a 3' phosphatase activity.
11. The kit of claim 10, wherein the one or more agents in (c) further comprise a kinase activity.
12. The kit of claim 10 or 11, wherein the one or more agents in (c) comprise a polynucleotide kinase (PNK).

13. The kit of any one of claims 10-12, further comprising single-stranded nucleic acid binding protein (SSB).

14. A method of producing a nucleic acid library, comprising:

(a) combining (i) a nucleic acid composition comprising single-stranded nucleic acid (ssNA), (ii) a first oligonucleotide, (iii) a plurality of first scaffold polynucleotide species, (iv) a second oligonucleotide, and (v) a plurality of second scaffold polynucleotide species, wherein:

(1) each polynucleotide in the plurality of first scaffold polynucleotide species comprises an ssNA hybridization region and a first oligonucleotide hybridization region;

(2) each polynucleotide in the plurality of second scaffold polynucleotide species comprises an ssNA hybridization region and a second oligonucleotide hybridization region; and

(3) the nucleic acid composition, the first oligonucleotide, the plurality of first scaffold polynucleotide species, the second oligonucleotide, and the plurality of second scaffold polynucleotide species, are combined under conditions wherein:

(i) a molecule of the first scaffold polynucleotide species is hybridized to a first ssNA terminal region and a molecule of the first oligonucleotide, and

(ii) a molecule of the second scaffold polynucleotide species is hybridized to a second ssNA terminal region and a molecule of the second oligonucleotide, thereby forming a set of hybridization products;

(b) contacting the set of hybridization products with one or more agents comprising a ligase activity, thereby generating a set of ligation products;

(c) contacting under hybridization conditions the set of ligation products with one or more probe species, wherein each probe species comprises a sequence complementary to an RNA sequence, thereby generating a mixture of hybridized ligation products and unhybridized ligation products; and

(d) separating the hybridized ligation products from the unhybridized ligation products.

15. The method of claim 14, wherein the RNA sequence comprises a ribosomal RNA (rRNA) sequence.

16. The method of claim 14 or 15, wherein, in the set of hybridization products in (a)(3)(ii), a 5' end of the molecule of the first oligonucleotide is adjacent to a 3' end of the first ssNA terminal region and a 3' end of the molecule of the second oligonucleotide is adjacent to a 5' end of the second ssNA terminal region.

17. The method of any one of claims 14-16, wherein the one or more probe species are chosen from one or more probes comprising nucleic acid sequences of SEQ ID NOs: 1-174.

18. The method of any one of claims 14-17, wherein each probe species further comprises a member of a binding pair.

19. The method of claim 18, wherein the binding pair is biotin and streptavidin.

20. The method of any one of claims 14-19, wherein the separating in (d) comprises a probe pull-down method.

21. A kit comprising:

(a) a first oligonucleotide, a plurality of first scaffold polynucleotide species, a second oligonucleotide, and a plurality of second scaffold polynucleotide species, wherein:

(i) each polynucleotide in the plurality of first scaffold polynucleotide species comprises an ssNA hybridization region and a first oligonucleotide hybridization region; and

(ii) each polynucleotide in the plurality of second scaffold polynucleotide species comprises an ssNA hybridization region and a second oligonucleotide hybridization region;

(b) one or more agents comprising a ligase activity; and

(c) one or more probe species, wherein each probe species comprises a sequence complementary to an RNA sequence.

22. The kit of claim 21, wherein the RNA sequence comprises a ribosomal RNA (rRNA) sequence.

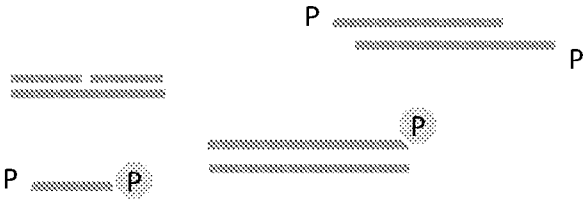
23. The kit of claim 21 or 22, wherein the one or more probe species are chosen from one or more probes comprising nucleic acid sequences of SEQ ID NOs: 1-174.

24. The kit of any one of claims 21-23, wherein each probe species further comprises a member of a binding pair.

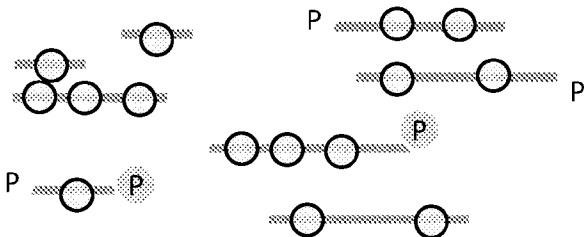
25. The kit of claim 24, wherein the binding pair is biotin and streptavidin.

Current Single-Stranded Prep (SRSLY)

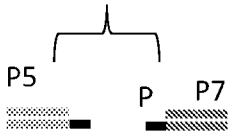
Template DNA



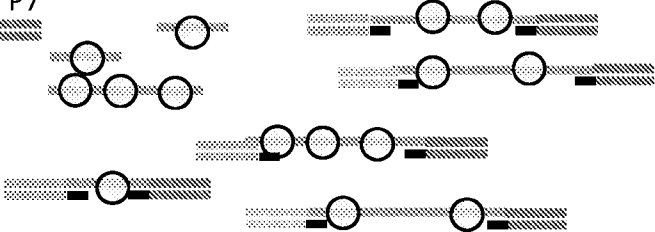
SSB & heat denature



Simultaneous addition



Kinase, Phosphatase, Ligation



P

5' Phos

P

3' Phos

O

SSB

SRSLY P5 splint-adapter

SRSLY P7 splint-adapter

50µl Reaction

Fig. 1

Modified Single-Stranded Prep (SRSLY 2.0)

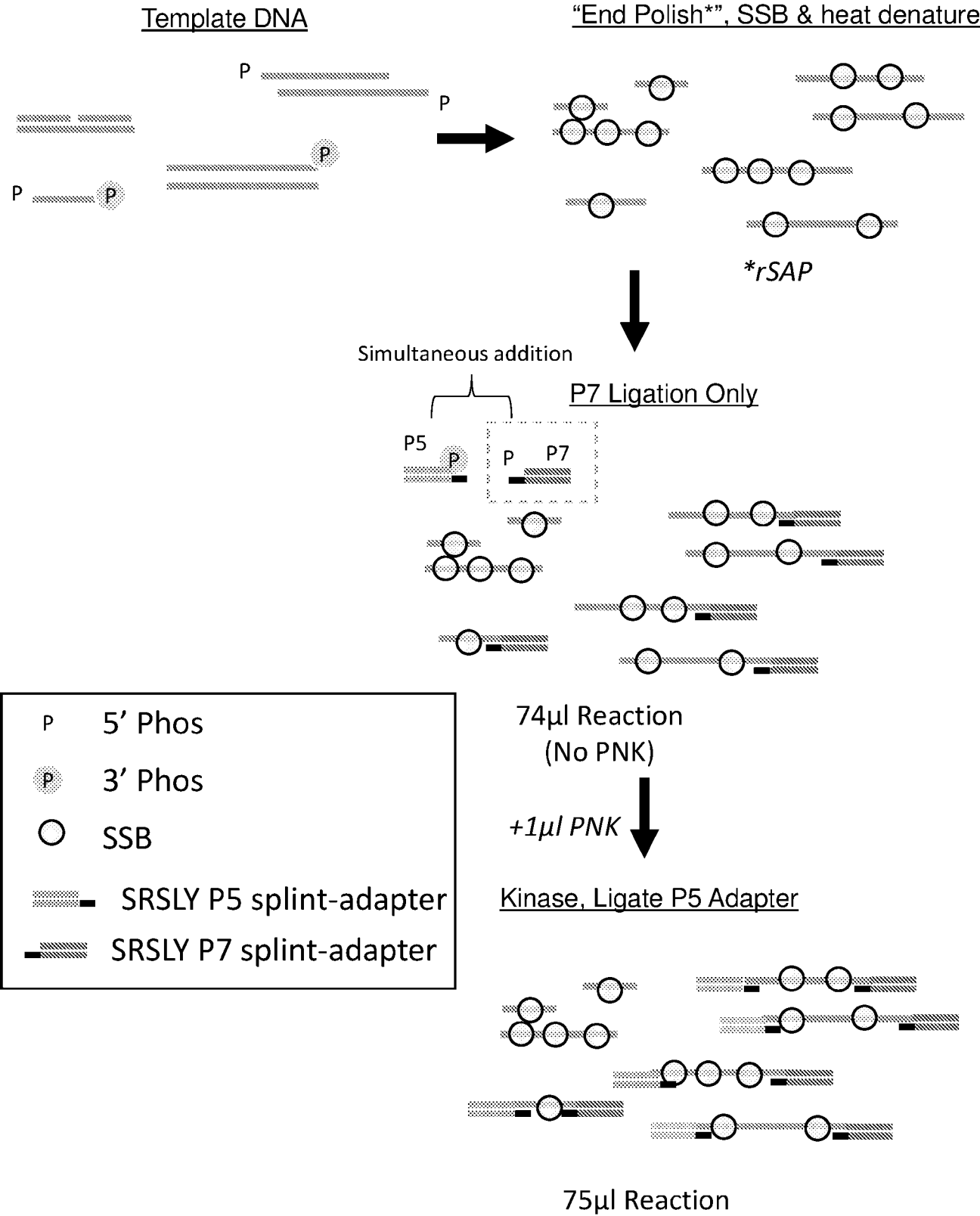


Fig. 2

RNA Depletion- Protocol Overview

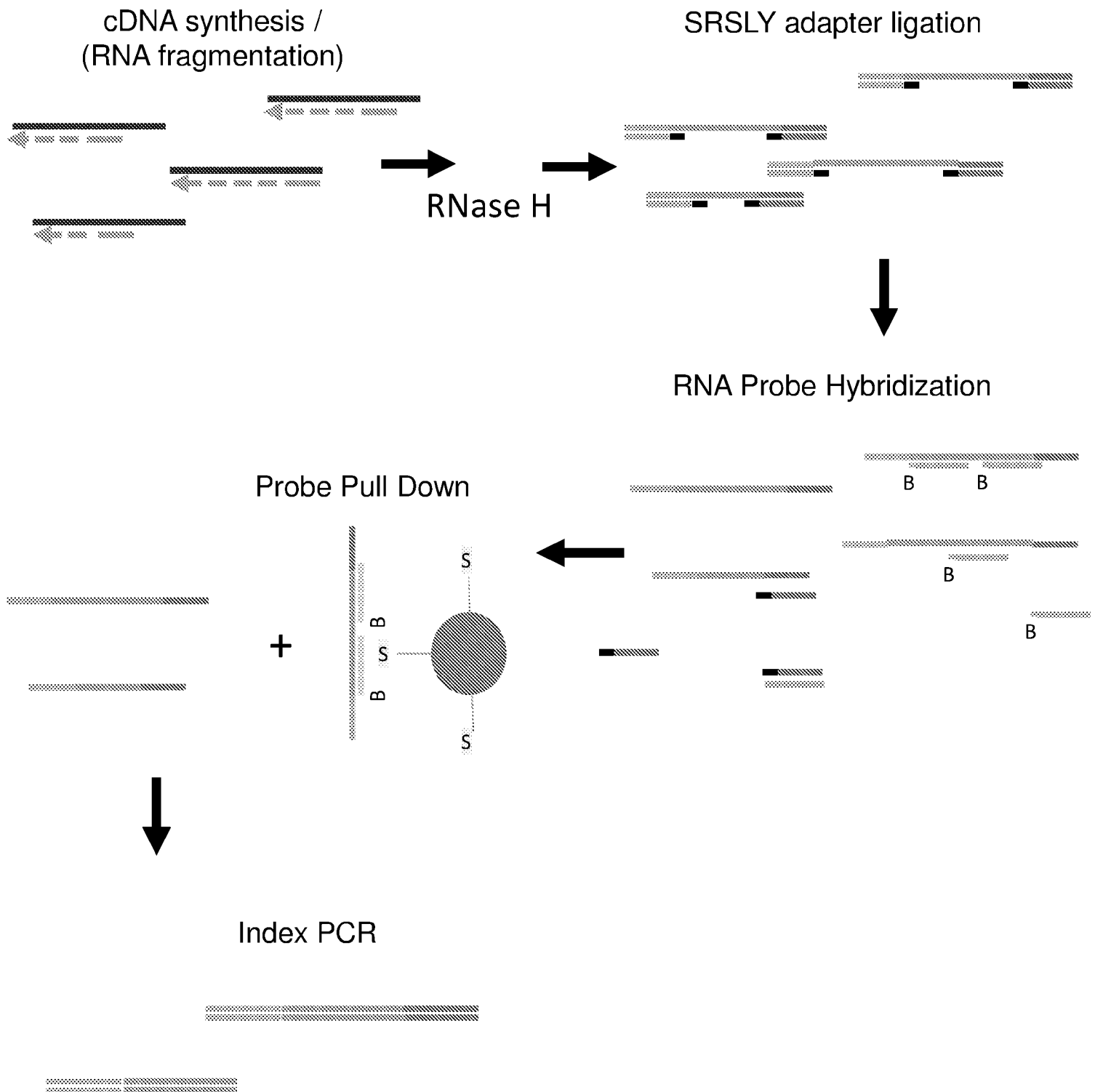


Fig. 3A

rRNA Depletion- Protocol Overview

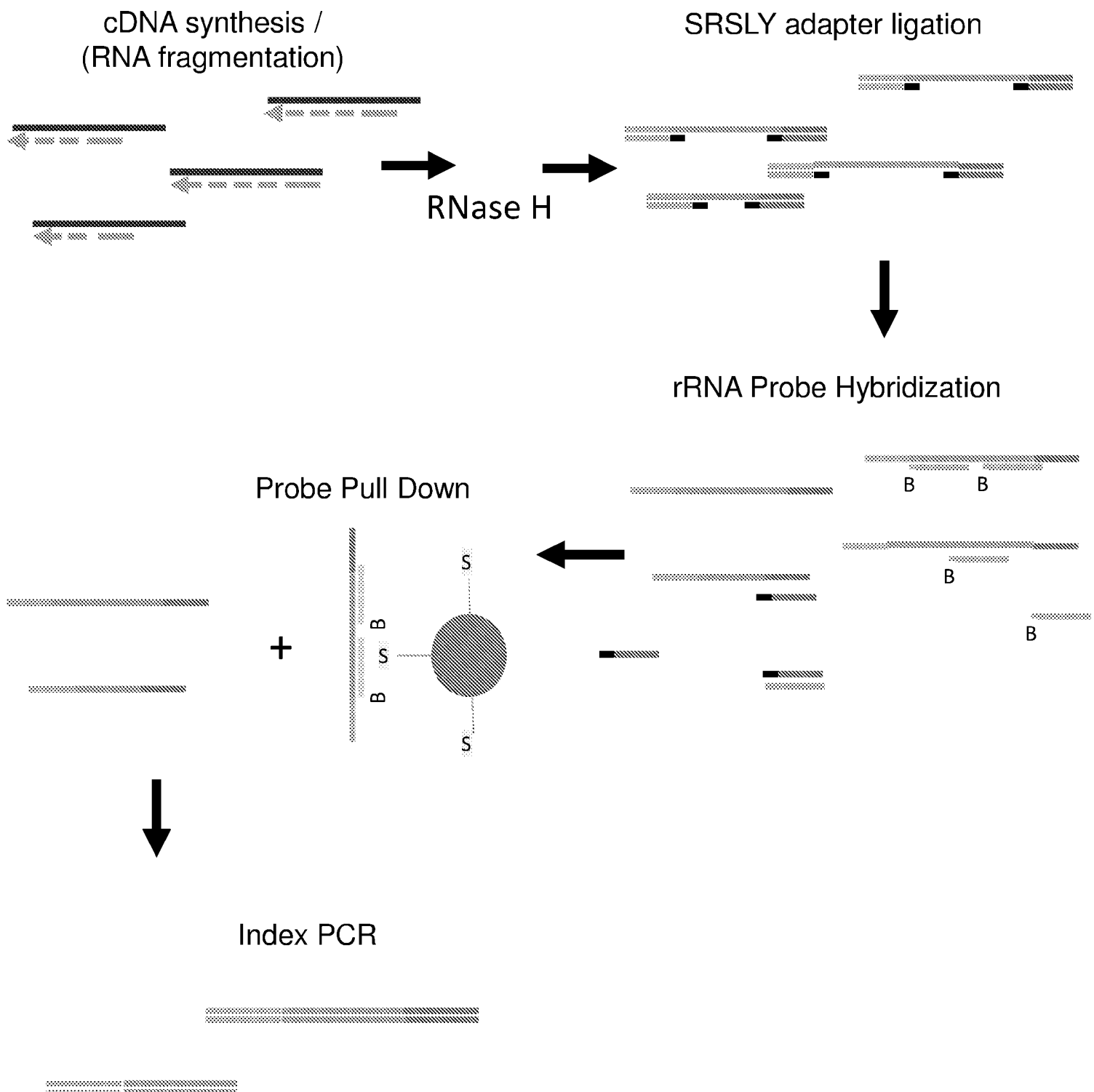


Fig. 3B

RNA-Seq Results

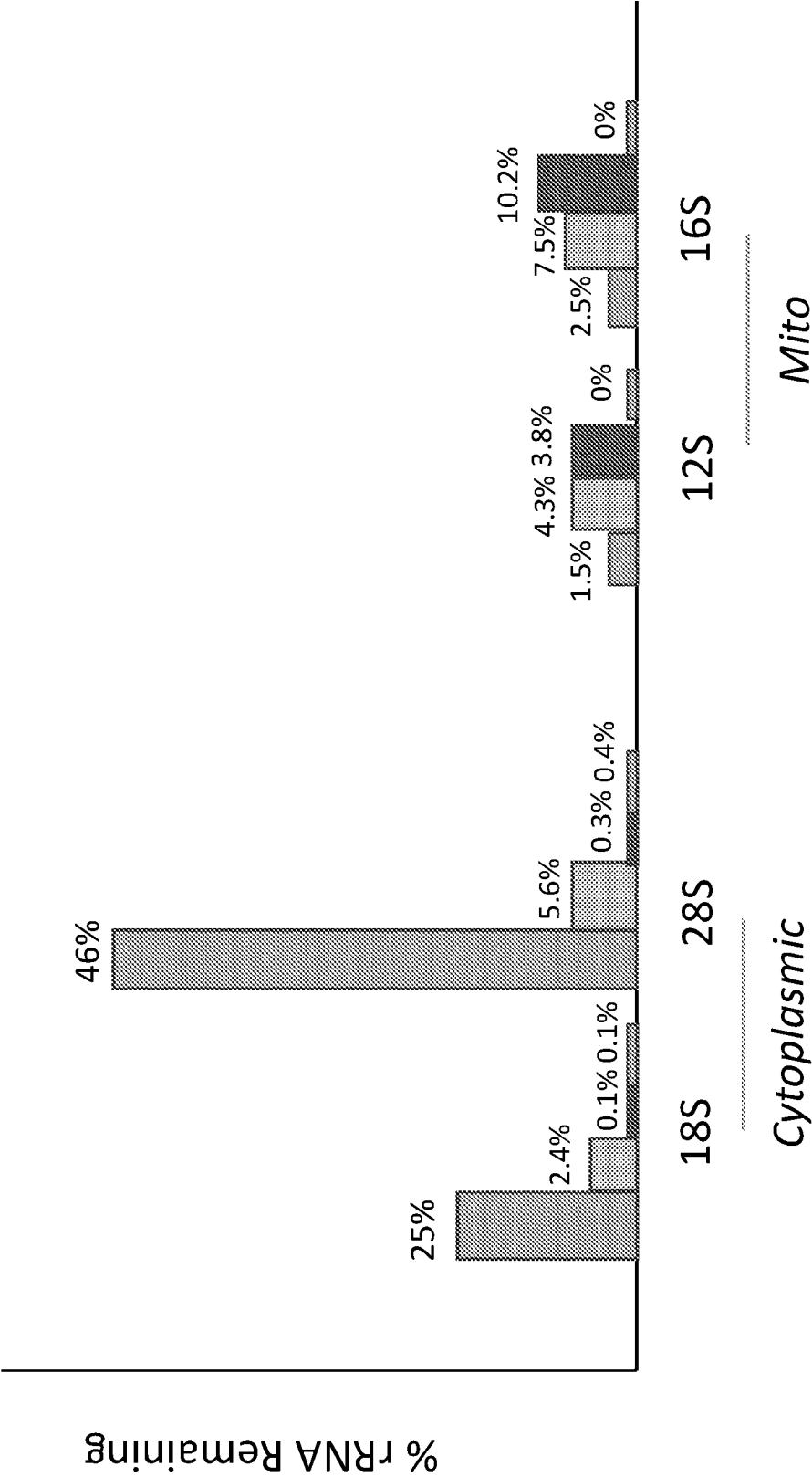


Fig. 4A

RNA-Seq Results

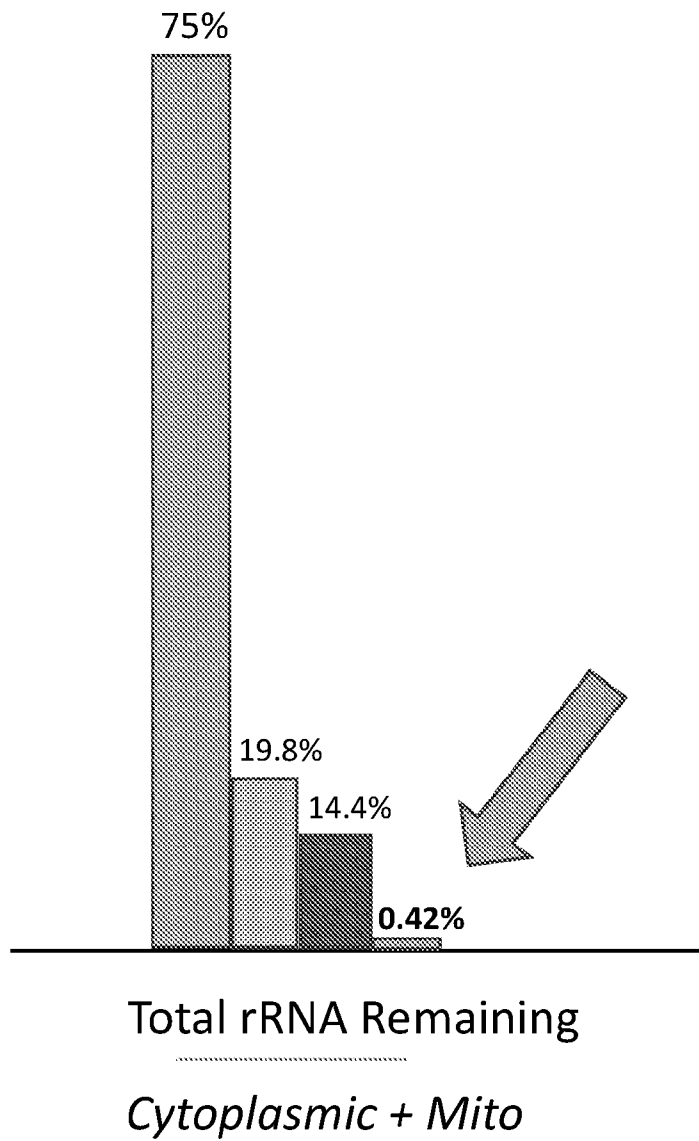


Fig. 4B

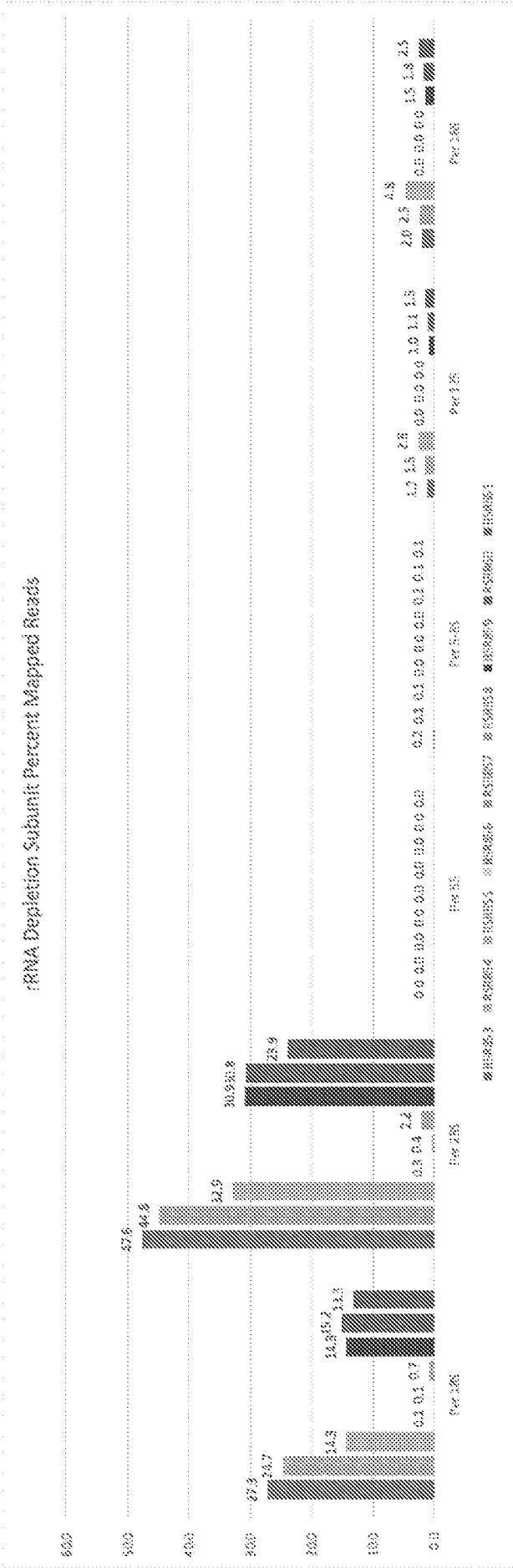


Fig. 5

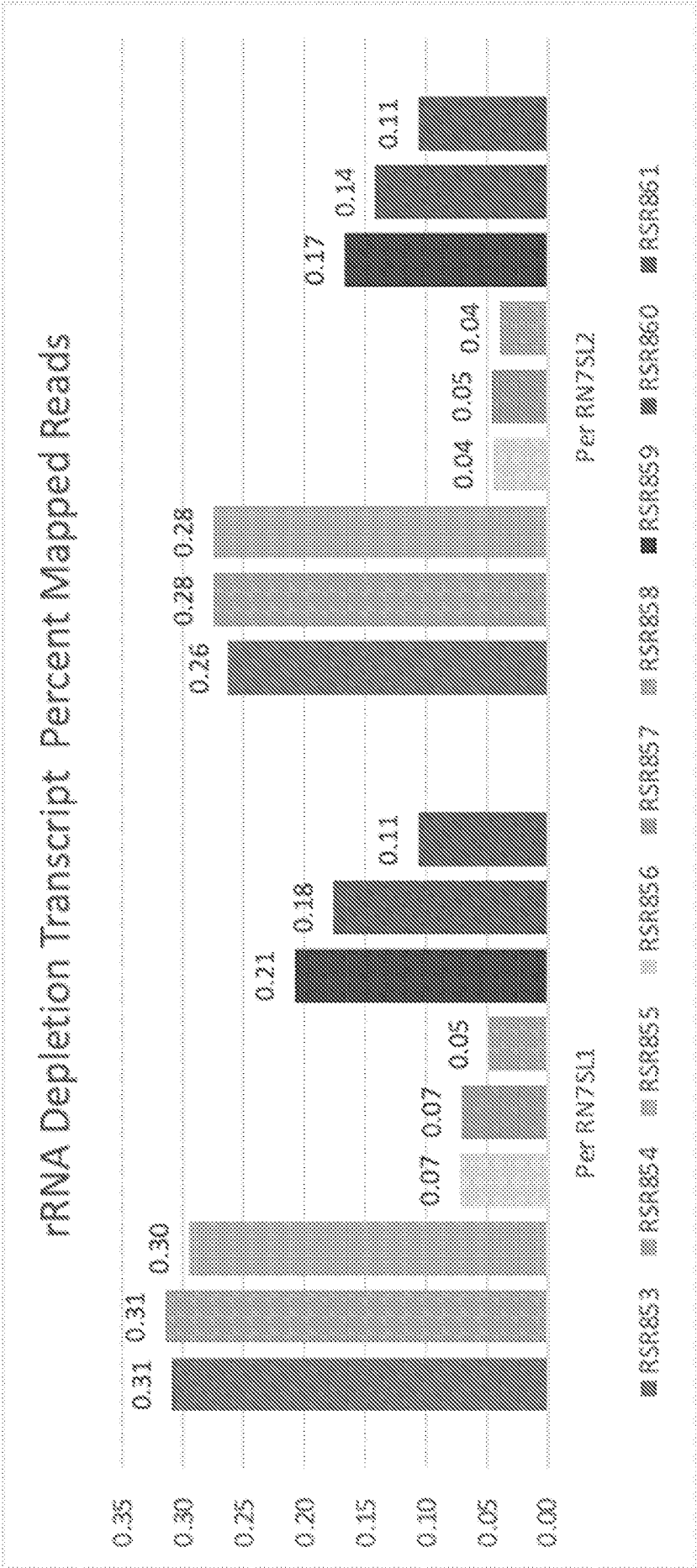


Fig. 6

id	Description	pairs	percent_kept	percent_mapped_unique	percent_mapped_mappers	percent_unmapped_bushart	percent_discarded	percent_dup	percent_size	percent_ribosomal_bases	percent_correct_strand	percent_mrna_bases	percent_coding_bases	percent_intronic_bases	percent_intergenic_bases	percent usable_bases
RS083-230711	50mg No Depletion REALLYY Control	586069	98.1	22.5	74.6	2.1	0.12	30.8	709718	74.15	98.44	9.98	5.13	4.85	7.62	9.93
RS084-230711	100mg No Depletion REALLYY Control	653755	98.2	27.6	69.1	2.5	0.20	30.9	785910	68.46	98.96	12.25	6.43	5.82	9.05	12.05
RS085-230711	500mg No Depletion REALLYY Control	585582	97.9	48.0	46.2	3.0	0.46	22.3	1044661	45.65	98.87	20.23	11.29	8.94	18.08	19.92
RS086-230711	500mg Depletion Protocol with 8ul Probe set stock 'New Probe Set 6 230812'	565942	99.5	92.4	4.8	2.4	0.29	0.8	3374359	0.33	98.45	16.36	22.07	17.29	16.09	38.82
RS087-230711	100mg Depletion Protocol with 8ul Probe set stock 'New Probe Set 6 230812'	586308	99.5	92.6	4.7	2.2	0.35	0.8	3538812	0.42	98.38	16.55	21.65	17.03	16.62	38.16
RS088-230711	500mg Depletion Protocol with 8ul Probe set stock 'New Probe Set 6 230812'	608558	99.6	90.7	6.3	2.4	0.47	1.0	2962997	2.82	97.98	33.00	18.08	14.52	19.66	32.60
RS089-230711	500mg Depletion Protocol with 8ul Probe set stock 'New Probe Set 6 230812' 1 round of step beads only	718705	98.7	50.4	47.1	2.0	0.18	19.3	1546320	45.18	98.42	20.36	11.10	9.27	12.39	20.09
RS090-230711	100mg Depletion Protocol with 8ul Probe set stock 'New Probe Set 6 230812' 1 round of step beads only	614790	99.1	49.8	47.8	2.0	0.19	20.3	1257182	46.14	98.41	19.28	10.63	8.66	11.87	19.03
RS091-230711	500mg Depletion Protocol with 8ul Probe set stock 'New Probe Set 6 230812' 1 round of step beads only	638306	99.1	53.1	49.1	2.3	0.17	17.7	1820155	37.42	97.96	19.92	10.94	8.98	12.21	19.68

Fig. 7

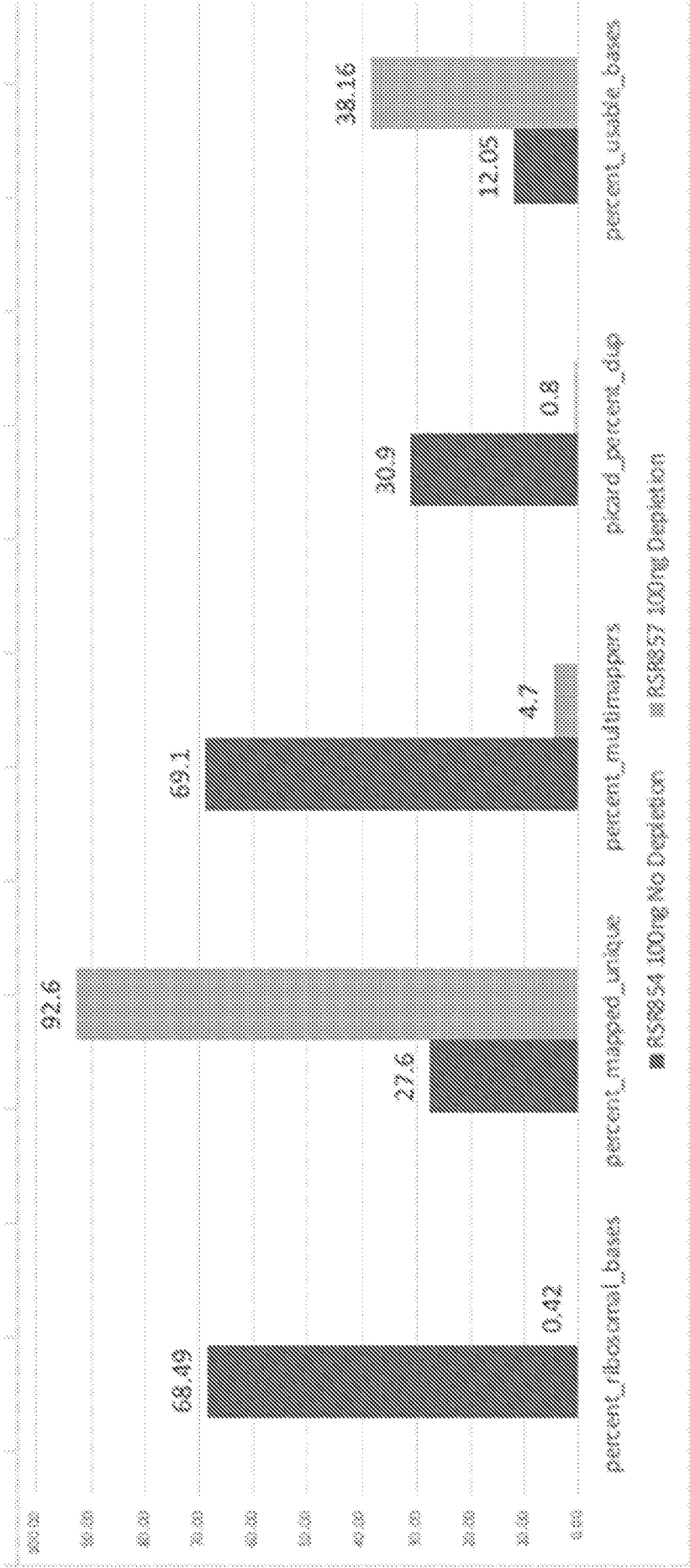


Fig. 8

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2024/032417

A. CLASSIFICATION OF SUBJECT MATTERIPC: **C40B 50/06** (2024.01); **C12N 15/10** (2024.01); **C12P 19/34** (2024.01); **C12Q 1/6813** (2024.01); **C12Q 1/6869** (2024.01); **C12Q 1/6876** (2024.01)CPC: **C40B 50/06**; **C12Q 1/6869**; **C12P 19/34**; **C12N 15/10**; **C12Q 1/6813**; **C12Q 1/6876**; **C12N 2330/31**; **C12N 9/93**

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History Document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

See Search History Document

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

See Search History Document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2022/0348906 A1 (CLARET BIOSCIENCE LLC) 03 November 2022 (03.11.2022) entire document	1-3, 14-16
A	US 7,427,479 B2 (KARGER et al.) 23 September 2008 (23.09.2008) entire document	1-3, 14-16
A	HEAD et al. Library construction for next-generation sequencing: Overviews and challenges. 1 February 2014. Biotechniques. Vol. 56(2), Pgs. 61-68, passim. [retrieved on 05.08.2024] . Retrieved from the internet: < https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4351865/ >. entire document	1-3, 14-16
A	NIKROO et al. Switchable Control of Scaffold Protein Activity via Engineered Phosphoregulated Autoinhibition. ACS Synthetic Biology. 9 June 2022. Vol. 11, Issue. 7. Pgs. 2464-2472. [retrieved on 05.08.2024]. Retrieved from the internet: < https://pubs.acs.org/doi/full/10.1021/acssynbio.2c00122 >. entire document	1-3, 14-16



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“D” document cited by the applicant in the international application

“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

05 August 2024 (05.08.2024)

Date of mailing of the international search report

27 September 2024 (27.09.2024)

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US
Commissioner for Patents
P.O. Box 1450, Alexandria, VA 22313-1450

Facsimile No. 571-273-8300

Authorized officer

MATOS
TAINA

Telephone No. 571-272-4300

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2024/032417

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
 - a. ☐ forming part of the international application as filed.
 - b. ☒ furnished subsequent to the international filing date for the purposes of international search (Rule 13ter.1(a)),
☒ accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.
2. ☐ With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.
3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2024/032417

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

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

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☒ Claims Nos.: **4-9, 13, 17-20, 24, 25**
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

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

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

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees need to be paid.

Group I: claims 1-3 and 14-16 are drawn to methods of producing a nucleic acid library.

Group II+: claims 10-12 and 21-23 are drawn to kits.

An exemplary election of Group II+ is a probe species comprising a nucleic acid sequence selected to be SEQ ID NO: 1, and kits comprising the same. The first named invention has been selected based on the guidance set forth in section 10.54 of the PCT International Search and Preliminary Examination Guidelines. Specifically, the first named invention was selected based on the first listed compound species presented in the claims (see claim 23).

Applicant is invited to elect additional nucleic acid sequence(s) for each additional probe to be searched in a specific combination by paying an additional fee for each set of election. Another exemplary election would be a probe species comprising a nucleic acid sequence selected to be SEQ ID NO: 2, and kits comprising the same. Additional nucleic acid sequence(s) will be searched upon the payment of additional fees. Applicants must specify the claims that read on any additional elected inventions. Applicants must further indicate, if applicable, the claims which read on the first named invention if different than what was indicated above for this group. Failure to clearly identify how any paid additional invention fees are to be applied to the "+" group(s) will result in only the first claimed invention to be searched/examined.

The inventions listed in Groups I and II+ do not relate to a single general inventive concept under PCT Rule 13.1, because under PCT Rule 13.2 they lack the same or corresponding special technical features for the following reasons:

The special technical features of Group I, methods of producing a nucleic acid library, are not present in Group II+; and the special technical features of Group II+, kits, are not present in Group I+.

The Groups II+ formulas do not share a significant structural element, requiring the selection of alternative nucleic acid sequences where "the one or more probe species are chosen from one or more probes comprising nucleic acid sequences of SEQ ID NOs: 1-174."

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

Additionally, even if Groups I and II+ were considered to share the technical features of a kit comprising: (a) a first oligonucleotide, a plurality of first scaffold polynucleotide species, a second oligonucleotide, and a plurality of second scaffold polynucleotide species, wherein: (i) the first oligonucleotide comprises a 5' phosphate; (ii) the second oligonucleotide comprises a 3' phosphate; (iii) each polynucleotide in the plurality of first scaffold polynucleotide species comprises an ssNA hybridization region and a first oligonucleotide hybridization region; and (iv) each polynucleotide in the plurality of second scaffold polynucleotide species comprises an ssNA hybridization region and a second oligonucleotide hybridization region; (b) one or more agents comprising a ligase activity; and (c) one or more agents comprising a 3' phosphatase activity; and kit comprising: (a) a first oligonucleotide, a plurality of first scaffold polynucleotide species, a second oligonucleotide, and a plurality of second scaffold polynucleotide species, wherein: (ii) each polynucleotide in the plurality of first scaffold polynucleotide species comprises an ssNA hybridization region and a first oligonucleotide hybridization region; and (ii) each polynucleotide in the plurality of second scaffold polynucleotide species comprises an ssNA hybridization region and a second oligonucleotide hybridization region; (b) one or more agents comprising a ligase activity; and (c) one or more probe species, wherein each probe species comprises a sequence complementary to an RNA sequence, these shared technical features do not represent a contribution over the prior art as disclosed by US 2022/0348906 to Claret Bioscience, LLC (hereinafter, "Claret").

Claret teaches a kit comprising: a) a first oligonucleotide, a plurality of first scaffold polynucleotide species, a second oligonucleotide, and a plurality of second scaffold polynucleotide species (a kit comprising a first oligonucleotide; Paras. [0718]-[0719]; the kit further comprising a second oligonucleotide; Paras. [0722]-[0723]), wherein: i) the first oligonucleotide comprises a 5' phosphate (a 5' phosphate is added to a 5' end of the first oligonucleotide; Para. [0884]); ii) the second oligonucleotide comprises a 3' phosphate (the second oligonucleotide comprises a 3' phosphate; Paras. [0500], [0739]); iii) each polynucleotide in the plurality of first scaffold polynucleotide species comprises an ssNA hybridization region and a first oligonucleotide hybridization region (each polynucleotide in the plurality of first scaffold polynucleotide species comprises an ssNA hybridization region and a first oligonucleotide hybridization region; Para. [0487], [0720]); and iv) each polynucleotide in the plurality of second scaffold polynucleotide species comprises an ssNA hybridization region and a second oligonucleotide hybridization region (each polynucleotide in the plurality of second scaffold polynucleotide species comprises an ssNA hybridization region and a second oligonucleotide hybridization region; Paras. [0504], [0724]); b) one or more agents comprising a ligase activity (a ligase; Para. [0734]); and c) one or more agents comprising a 3' phosphatase activity (a phosphatase; Para. [0738]; the second oligonucleotide comprises a phosphate at the 3' end; Para. [0500]; and contacting the second oligonucleotide with an agent comprising a phosphatase activity (thereby comprising a 3' phosphatase activity); Para. [0506]); and a kit comprising: a) a first oligonucleotide, a plurality of first scaffold polynucleotide species, a second oligonucleotide, and a plurality of second scaffold polynucleotide species (a kit comprising a first oligonucleotide; Paras. [0718]-[0719]; the kit further comprising a second oligonucleotide; Paras. [0722]-[0723]), wherein: ii) each polynucleotide in the plurality of first scaffold polynucleotide species comprises an ssNA hybridization region and a first oligonucleotide hybridization region (each polynucleotide in the plurality of first scaffold polynucleotide species comprises an ssNA hybridization region and a first oligonucleotide hybridization region; Para. [0487], [0720]); and ii) each polynucleotide in the plurality of second scaffold polynucleotide species comprises an ssNA hybridization region and a second oligonucleotide hybridization region (each polynucleotide in the plurality of second scaffold polynucleotide species comprises an ssNA hybridization region and a second oligonucleotide hybridization region; Paras. [0504], [0724]); b) one or more agents comprising a ligase activity (a ligase; Para. [0734]); and c) one or more probe species, wherein each probe species comprises a sequence complementary to an RNA sequence (rRNA depletion probes in combination with magnetic beads to deplete rRNA from a sample, thus enriching mRNA; Para. [0216]).

The inventions listed in Groups I and II+ therefore lack unity under Rule 13 because they do not share a same or corresponding special technical feature.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2024/032417

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

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: **1-3, 14-16**

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.