

POLYNUCLEOTIDES, REAGENTS, AND METHODS FOR NUCLEIC ACID HYBRIDIZATION

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. provisional patent application number 62/673,704 filed on May 18, 2018, U.S. provisional patent application number 62/810,343 filed on February 25, 2019, U.S. provisional patent application number 62/814,749 filed on March 6, 2019, U.S. provisional patent application number 62/675,647 filed May 23, 2018, U.S. provisional patent application number 62/810,293 filed February 25, 2019, U.S. provisional patent application number 62/814,753 filed March 6, 2019, U.S. provisional patent application number 62/833,440 filed April 12, 2019, each of which is incorporated herein by reference in its entirety.

BACKGROUND

[0002] Highly efficient chemical gene synthesis with high fidelity and low cost has a central role in biotechnology and medicine, and in basic biomedical research. De novo gene synthesis is a powerful tool for basic biological research and biotechnology applications. While various methods are known for the synthesis of relatively short fragments in a small scale, these techniques often suffer from scalability, automation, speed, accuracy, and cost.

INCORPORATION BY REFERENCE

[0003] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF SUMMARY

[0004] Provided herein are methods for sequencing genomic DNA, comprising: contacting a composition comprising a first polynucleotide library comprising at least 30,000 polynucleotides, wherein each of the at least 30,000 polynucleotides is present in an amount such that, following hybridization with genomic fragments and sequencing of the hybridized genomic fragments, the polynucleotide library provides a read depth of at least 80 percent of the bases of the genomic fragments corresponding to the polynucleotides; and a total number of sequencing reads, wherein the total number of sequencing reads are capable of covering 100 percent of each of the bases of the genomic fragments corresponding to the polynucleotides at a theoretical read depth, wherein the ratio of the read depth of at least 80 percent of the bases of the genomic fragments corresponding to the polynucleotides to the theoretical read depth is at least 0.5 with a plurality of genomic

fragments; enriching at least one genomic fragment that binds to the first polynucleotide library to generate at least one enriched target polynucleotide; sequencing the at least one enriched target polynucleotide; identifying one or more positions of the at least one enriched polynucleotide having less than average read depth; repeating steps a-c, wherein a second polynucleotide library comprising at least 1500 polynucleotides is added to the composition, wherein the second polynucleotide library comprises at least one polynucleotide that binds to genomic fragments comprising the one or more positions having less than average read depth, wherein the presence of the second polynucleotide library increases the read depth at the one or more positions having less than average read depth. Provided herein are methods for sequencing genomic DNA, comprising: contacting a composition comprising a first polynucleotide library comprising at least 30,000 polynucleotides, wherein each of the at least 30,000 polynucleotides is present in an amount such that, following hybridization with genomic fragments and sequencing of the hybridized genomic fragments, the polynucleotide library provides a read depth of at least 80 percent of the bases of the genomic fragments corresponding to the polynucleotides; and a total number of sequencing reads, wherein the total number of sequencing reads are capable of covering 100 percent of each of the bases of the genomic fragments corresponding to the polynucleotides at a theoretical read depth, wherein the ratio of the read depth of at least 80 percent of the bases of the genomic fragments corresponding to the polynucleotides to the theoretical read depth is at least 0.5 with a plurality of genomic fragments; enriching at least one genomic fragment that binds to the first polynucleotide library to generate at least one enriched target polynucleotide; sequencing the at least one enriched target polynucleotide; identifying one or more positions of the at least one enriched polynucleotide having less than average read depth; repeating steps a-c, wherein a second polynucleotide library is added to the composition, wherein the second polynucleotide library comprises at least one polynucleotide that binds to genomic fragments comprising the one or more positions having less than average read depth, wherein the presence of the second polynucleotide library increases the read depth at the one or more positions having less than average read depth. Further provided herein are methods wherein the first polynucleotide library and the second polynucleotide library do not comprise any common sequences. Further provided herein are methods wherein the first polynucleotide library and the second polynucleotide library comprise at least one common sequence. Further provided herein are methods wherein the presence of the second polynucleotide library increases the read depth at the one or more positions of the least one enriched target polynucleotide having less than average read depth by at least 10 fold. Further provided herein are methods wherein the presence of the second polynucleotide library increases the read depth at the

one or more positions of the at least one enriched target polynucleotide having less than average read depth by at least 100 fold.

[0005] Provided herein are polynucleotide libraries, the polynucleotide library comprising at least 1500 polynucleotides, wherein less than all polynucleotides comprises a molecular tag, wherein each of the at least 5000 polynucleotides are present in an amount such that, following hybridization with genomic fragments and sequencing of the hybridized genomic fragments, the polynucleotide library provides a read depth of at least 90 percent of the bases of the genomic fragments corresponding to the polynucleotides; and a total number of sequencing reads, wherein the total number of sequencing reads are capable of covering 100 percent of each of the bases of the genomic fragments corresponding to the polynucleotides at a theoretical read depth, wherein the ratio of the read depth of at least 90 percent of the bases of the genomic fragments corresponding to the polynucleotides to the theoretical read depth is at least 0.5. Further provided herein are polynucleotide libraries wherein no more than 90% of the polynucleotides comprise a molecular tag. Further provided herein are polynucleotide libraries wherein no more than 80% of the polynucleotides comprise a molecular tag. Further provided herein are polynucleotide libraries wherein no more than 50% of the polynucleotides comprise a molecular tag. Further provided herein are polynucleotide libraries wherein no more than 25% of the polynucleotides comprise a molecular tag. Further provided herein are polynucleotide libraries wherein the molecular tag is biotin. Further provided herein are polynucleotide libraries wherein the at least 5000 polynucleotides encode for at least 5000 genes. Further provided herein are polynucleotide libraries wherein the polynucleotide library comprises at least 30,000 polynucleotides. Further provided herein are polynucleotide libraries wherein the polynucleotide library comprises at least 100,000 polynucleotides.

[0006] Provided herein are methods for enriching nucleic acids comprising: contacting the polynucleotide library described herein with a plurality of genomic fragments; enriching at least one genomic fragment that binds to the polynucleotide library to generate at least one enriched target polynucleotide; and sequencing the at least one enriched target polynucleotide. Further provided herein are methods wherein the polynucleotide library provides for at least 90 percent unique reads for the bases of the enriched target polynucleotide after sequencing. Further provided herein are methods wherein the polynucleotide library provides for at least 95 percent unique reads for the bases of the enriched target polynucleotide after sequencing. Further provided herein are methods wherein the polynucleotide library provides for at least 80 percent of the bases of the enriched target polynucleotide having a read depth within about 1.5 times the mean read depth. Further provided herein are methods wherein the polynucleotide library provides for at least 90

percent of the bases of the enriched target polynucleotide having a read depth within about 1.5 times the mean read depth.

[0007] Provided herein are polynucleotide libraries, the polynucleotide library comprising at least 5000 polynucleotides, wherein each of the at least 5000 polynucleotides is present in an amount such that, following hybridization with a composition comprising i) a genomic library, wherein the genomic library comprises polynucleotides each comprising genomic fragments, at least one index sequence, and at least one adapter; and ii) at least one polynucleotide blocker, wherein the polynucleotide blocker is complementary to at least a portion of the adapter sequence, but not complementary to the at least one index sequence; and sequencing of the hybridized genomic fragments, the polynucleotide library provides for at least 30 fold read depth of at least 90 percent of the bases of the genomic fragments under conditions wherein the total number of reads is no more than 55 fold higher than the total number of bases of the hybridized genomic fragments. Further provided herein are polynucleotide libraries wherein the composition comprises no more than four polynucleotide blockers. Further provided herein are polynucleotide libraries wherein the polynucleotide blocker comprises one or more nucleotide analogues. Further provided herein are polynucleotide libraries wherein the polynucleotide blocker comprises one or more locked nucleic acids (LNAs). Further provided herein are polynucleotide libraries wherein the polynucleotide blocker comprises one or more bridged nucleic acids (BNAs). Further provided herein are polynucleotide libraries wherein the polynucleotide blocker comprises at least 2 nucleotide analogues. Further provided herein are polynucleotide libraries wherein the polynucleotide blocker comprises at least 5 nucleotide analogues. Further provided herein are polynucleotide libraries wherein the polynucleotide blocker comprises at least 10 nucleotide analogues. Further provided herein are polynucleotide libraries wherein the polynucleotide blocker has a T_m of at least 70 degrees C. Further provided herein are polynucleotide libraries wherein the polynucleotide blocker has a T_m of at least 75 degrees C. Further provided herein are polynucleotide libraries wherein the polynucleotide blocker has a T_m of at least 78 degrees C. Further provided herein are polynucleotide libraries wherein the polynucleotide blocker has a T_m of at least 82 degrees C. Further provided herein are polynucleotide libraries wherein the polynucleotide blocker has a T_m of 80-90 degrees C. Further provided herein are polynucleotide libraries wherein the polynucleotide blocker has a T_m of at least 80 degrees C. Further provided herein are polynucleotide libraries wherein the genomic library comprises genomic fragments from at least 2 different samples. Further provided herein are polynucleotide libraries wherein the genomic library comprises genomic fragments from at least 10 different samples. Further provided herein are polynucleotide libraries wherein the genomic library comprises genomic fragments from at least 2

non-identical index sequences. Further provided herein are polynucleotide libraries wherein the genomic library comprises genomic fragments from at least 16 non-identical index sequences. Further provided herein are polynucleotide libraries wherein the genomic library comprises genomic fragments further comprising at least one unique molecular identifier (UMI).

[0008] Provided herein are methods for enriching nucleic acids comprising: contacting the polynucleotide libraries described herein with a plurality of genomic fragments; enriching at least one genomic fragment that binds to the polynucleotide library to generate at least one enriched target polynucleotide; and sequencing the at least one enriched target polynucleotide. Further provided herein are methods wherein the off-target rate is less than 25%. Further provided herein are methods wherein the off-target rate is less than 20%. Further provided herein are methods wherein the molar ratio between at least one polynucleotide blocker and the complementary adapter is no more than 5:1. Further provided herein are methods wherein the molar ratio between at least one polynucleotide blocker and the complementary adapter is no more than 2:1. Further provided herein are methods wherein the molar ratio between at least one polynucleotide blocker and the complementary adapter is no more than 1.5:1.

[0009] Provided herein are compositions for nucleic acid hybridization comprising: a first polynucleotide library; a second polynucleotide library, wherein at least one polynucleotide in the first library is at least partially complementary to at least one polynucleotide of the second library; and an additive, wherein the additive reduces off-target hybridization of the at least one polynucleotide of the first library with the at least one polynucleotide of the second library by decreasing a local concentration of the first polynucleotide library or the second polynucleotide library at an air-liquid interface. Further provided herein are compositions wherein the additive is mineral oil, a nucleotide triphosphate, polyether, or urea. Further provided herein are compositions wherein the additive is a hydrocarbon comprising at least six carbon atoms. Further provided herein are compositions wherein the additive is silicon oil. Further provided herein are compositions wherein the oil is derived from plant sources. Further provided herein are compositions wherein the composition further comprises dimethyl sulfoxide. Further provided herein are compositions wherein the composition does not comprise a formamide. Further provided herein are compositions wherein the size of the first polynucleotide library is less than 10 million bases. Further provided herein are compositions wherein the size of the first polynucleotide library is less than 1 million bases. Further provided herein are compositions wherein the size of the first polynucleotide library is less than 0.5 million bases. Further provided herein are compositions wherein the first polynucleotide library comprises at least one exon sequence. Further provided herein are compositions wherein first polynucleotide library comprises polynucleotides encoding for at least

10 genes. Further provided herein are compositions wherein the first polynucleotide library comprises polynucleotides encoding for at least 100 genes. Further provided herein are compositions wherein the first polynucleotide library comprises at least one genomic fragment. Further provided herein are compositions wherein the first polynucleotide library comprises RNA, DNA, cDNA, or genomic DNA. Further provided herein are compositions wherein the first polynucleotide library comprises genomic DNA.

[0010] Provided herein are compositions for nucleic acid hybridization comprising: a first polynucleotide library and a second polynucleotide library each comprising a plurality of polynucleotides, wherein at least one polynucleotide in the first library is at least partially complimentary to at least one polynucleotide of the second library; and an oil, wherein the oil reduces off-target hybridization of the at least one polynucleotide of the first library with the at least one polynucleotide of the second library by decreasing a local concentration of the first polynucleotide library or the second polynucleotide library at an air-liquid interface. Further provided herein are compositions wherein the additive is mineral oil, a nucleotide triphosphate, polyether, or urea. Further provided herein are compositions wherein the additive is a hydrocarbon comprising at least six carbon atoms. Further provided herein are compositions wherein the additive is silicon oil. Further provided herein are compositions wherein the oil is derived from plant sources. Further provided herein are compositions wherein the composition further comprises dimethyl sulfoxide. Further provided herein are compositions wherein the composition does not comprise a formamide. Further provided herein are compositions wherein the size of the first polynucleotide library is less than 10 million bases. Further provided herein are compositions wherein the size of the first polynucleotide library is less than 1 million bases. Further provided herein are compositions wherein the size of the first polynucleotide library is less than 0.5 million bases. Further provided herein are compositions wherein first polynucleotide library comprises at least one exon sequence. Further provided herein are compositions wherein first polynucleotide library comprises polynucleotides encoding for at least 10 genes. Further provided herein are compositions wherein first polynucleotide library comprises polynucleotides encoding for at least 100 genes. Further provided herein are compositions wherein the first polynucleotide library comprises at least one genomic fragment. Further provided herein are compositions wherein the first polynucleotide library comprises RNA, DNA, cDNA, or genomic DNA. Further provided herein are compositions wherein the first polynucleotide library comprises genomic DNA.

[0011] Provided herein are methods for reducing off-target nucleic acid hybridization, comprising: contacting a first polynucleotide library with a second polynucleotide library, wherein the first polynucleotide library and the second polynucleotide library each comprise a plurality of

polynucleotides, and wherein at least one polynucleotide in the first library is at least partially complimentary to at least one polynucleotide in the second library; enriching at least one genomic fragment that binds to the second polynucleotide library to generate at least one enriched target polynucleotide, wherein enriching comprises at least one aspiration step, and wherein the at least one aspiration step comprises aspirating only liquid from the area near the air/liquid interface; and sequencing the at least one enriched target polynucleotide. Further provided herein are methods wherein the additive is oil, a nucleotide triphosphate, polyether, or urea. Further provided herein are methods wherein the additive is mineral oil. Further provided herein are methods wherein the presence of the additive decreases off-target binding. Further provided herein are methods wherein the presence of the additive decreases off-target binding by at least 10%. Further provided herein are methods wherein the presence of the additive decreases off-target binding by at least 20%. Further provided herein are methods wherein the presence of the additive decreases off-target binding by at least 30%. Further provided herein are methods wherein the off-target binding is random off-target binding. Further provided herein are methods wherein the size of the first polynucleotide library is less than 10 million bases. Further provided herein are methods wherein the size of the first polynucleotide library is less than 1 million bases. Further provided herein are methods wherein the size of the first polynucleotide library is less than 0.5 million bases. Further provided herein are methods wherein first polynucleotide library comprises at least one exon sequence. Further provided herein are methods wherein first polynucleotide library comprises polynucleotides encoding for at least 10 genes. Further provided herein are methods wherein first polynucleotide library comprises polynucleotides encoding for at least 100 genes. Further provided herein are methods wherein the first polynucleotide library comprises at least one genomic fragment. Further provided herein are methods wherein the first polynucleotide library comprises RNA, DNA, cDNA, or genomic DNA. Further provided herein are methods wherein the first polynucleotide library comprises genomic DNA.

[0012] Provided herein are methods for sequencing genomic DNA, comprising: contacting a polynucleotide library with a plurality of genomic fragments and an additive to form a mixture, wherein the additive decreases a local concentration of the polynucleotide library or the genomic fragments in the mixture at an air-liquid interface; enriching at least one genomic fragment that binds to the polynucleotide library to generate at least one enriched target polynucleotide; and sequencing the at least one enriched target polynucleotide. Further provided herein are methods wherein the additive is oil, a nucleotide triphosphate, polyether, or urea. Further provided herein are methods wherein the additive is mineral oil. Further provided herein are methods wherein the presence of the additive decreases off-target binding. Further provided herein are methods wherein

the presence of the additive decreases off-target binding by at least 10%. Further provided herein are methods wherein the presence of the additive decreases off-target binding by at least 20%. Further provided herein are methods wherein the presence of the additive decreases off-target binding by at least 30%. Further provided herein are methods wherein the off-target binding is random off-target binding. Further provided herein are methods wherein the size of the first polynucleotide library is less than 10 million bases. Further provided herein are methods wherein the size of the first polynucleotide library is less than 1 million bases. Further provided herein are methods wherein the size of the first polynucleotide library is less than 0.5 million bases. Further provided herein are methods wherein the first polynucleotide library comprises at least one exon sequence. Further provided herein are methods wherein the first polynucleotide library comprises polynucleotides encoding for at least 10 genes. Further provided herein are methods wherein the first polynucleotide library comprises polynucleotides encoding for at least 100 genes. Further provided herein are methods wherein the first polynucleotide library comprises at least one genomic fragment. Further provided herein are methods wherein the first polynucleotide library comprises RNA, DNA, cDNA, or genomic DNA. Further provided herein are methods wherein the first polynucleotide library comprises genomic DNA.

BRIEF DESCRIPTION OF THE DRAWINGS

[0013] **Figure 1A** depicts a schematic workflow, including analyzing nucleic acid sequencing data, spiking in additional capture probe polynucleotide libraries that target specific areas of the analyzed nucleic acids, and obtaining new sequencing data with increased read depth at targeted regions.

[0014] **Figure 1B** depicts an exemplary a dual adapter-ligated nucleic acid with index sequences and four universal blocker polynucleotides.

[0015] **Figure 1C** depicts an exemplary workflow for enrichment and sequencing of a nucleic acid sample using partially labeled capture probes.

[0016] **Figure 2** depicts an exemplary workflow for enrichment and sequencing of a nucleic acid sample.

[0017] **Figure 3** depicts a plot of sequencing coverage vs. position at chromosome 11 after a genomic library is enriched with two different exome capture library, a smaller library panel targeting pain genes, or combinations of the exome and panel libraries.

[0018] **Figure 4A** depicts a plot of percent off bait vs. blocker type for an enrichment and sequencing analysis comparing types of blockers during probe hybridization. Conditions included no blockers (-control), specific blockers (+control), or two different designs of universal blockers.

[0019] **Figure 4B** depicts a plot of percent off bait vs. blocker mixtures of an enrichment and sequencing analysis comparing types of blockers during probe hybridization. Conditions included no blockers (-control), specific blockers (+control), or conditions wherein different combinations of universal blockers were independently tested.

[0020] **Figure 4C** depicts a plot of percent off bait vs. different designs of an enrichment and sequencing analysis comparing types of blockers during probe hybridization at different mass loadings.

[0021] **Figure 4D** depicts a plot of percent off bait vs. blocker concentration of an enrichment and sequencing analysis with universal blockers.

[0022] **Figure 4E** depicts a plot of the percent off bait vs. universal blockers comprising various amounts of locked nucleic acids for an enrichment and sequencing analysis.

[0023] **Figure 4F** depicts a plot of the percent off bait vs. universal blockers comprising various amounts of bridged nucleic acids for an enrichment and sequencing analysis.

[0024] **Figure 5A** depicts a plot of percent off bait vs. percent baits comprising biotin for an enrichment and sequencing analysis.

[0025] **Figure 5B** depicts a plot of AT or GC dropouts vs. percent baits comprising biotin for an enrichment and sequencing analysis.

[0026] **Figure 6A** depicts a plot of HS library size/target size vs. $\log_2(\text{bait mass} / \text{target size})$ for an enrichment and sequencing analysis comparing performance of an exome library and a smaller targeted pain gene exome library. The data for the exome library is fit to a linear model of dilution.

[0027] **Figure 6B** depicts a plot of HS library size/target size vs. $\log_2(\text{bait mass} / \text{target size})$ for an enrichment and sequencing analysis comparing performance of an exome library and a smaller targeted pain gene exome library. The data is fit to a logarithmic model of dilution.

[0028] **Figure 7** depicts a schematic for enriching target polynucleotides with a target binding polynucleotide library.

[0029] **Figure 8** depicts a schematic for generation of polynucleotide libraries from cluster amplification.

[0030] **Figure 9A** depicts a pair of polynucleotides for targeting and enrichment. The polynucleotides comprise complementary target binding (insert) sequences, as well as primer binding sites.

[0031] **Figure 9B** depicts a pair of polynucleotides for targeting and enrichment. The polynucleotides comprise complementary target sequence binding (insert) sequences, primer binding sites, and non-target sequences.

[0032] **Figure 10A** depicts a polynucleotide binding configuration to a target sequence of a larger polynucleotide. The target sequence is shorter than the polynucleotide binding region, and the polynucleotide binding region (or insert sequence) is offset relative to the target sequence, and also binds to a portion of adjacent sequence.

[0033] **Figure 10B** depicts a polynucleotide binding configuration to a target sequence of a larger polynucleotide. The target sequence length is less than or equal to the polynucleotide binding region, and the polynucleotide binding region is centered with the target sequence, and also binds to a portion of adjacent sequence.

[0034] **Figure 10C** depicts a polynucleotide binding configuration to a target sequence of a larger polynucleotide. The target sequence is slightly longer than the polynucleotide binding region, and the polynucleotide binding region is centered on the target sequence with a buffer region on each side.

[0035] **Figure 10D** depicts a polynucleotide binding configuration to a target sequence of a larger polynucleotide. The target sequence is longer than the polynucleotide binding region, and the binding regions of two polynucleotides are overlapped to span the target sequence.

[0036] **Figure 10E** depicts a polynucleotide binding configuration to a target sequence of a larger polynucleotide. The target sequence is longer than the polynucleotide binding region, and the binding regions of two polynucleotides are overlapped to span the target sequence.

[0037] **Figure 10F** depicts a polynucleotide binding configuration to a target sequence of a larger polynucleotide. The target sequence is longer than the polynucleotide binding region, and the binding regions of two polynucleotides are not overlapped to span the target sequence, leaving a gap 405.

[0038] **Figure 10G** depicts a polynucleotide binding configuration to a target sequence of a larger polynucleotide. The target sequence is longer than the polynucleotide binding region, and the binding regions of three polynucleotides are overlapped to span the target sequence.

[0039] **Figure 11** presents a diagram of steps demonstrating an exemplary process workflow for gene synthesis as disclosed herein.

[0040] **Figure 12** illustrates a computer system.

[0041] **Figure 13** is a block diagram illustrating an architecture of a computer system.

[0042] **Figure 14** is a diagram demonstrating a network configured to incorporate a plurality of computer systems, a plurality of cell phones and personal data assistants, and Network Attached Storage (NAS).

[0043] **Figure 15** is a block diagram of a multiprocessor computer system using a shared virtual address memory space.

- [0044] **Figure 16** is an image of a plate having 256 clusters, each cluster having 121 loci with polynucleotides extending therefrom.
- [0045] **Figure 17A** is a plot of polynucleotide representation (polynucleotide frequency versus abundance, as measured absorbance) across a plate from synthesis of 29,040 unique polynucleotides from 240 clusters, each cluster having 121 polynucleotides.
- [0046] **Figure 17B** is a plot of measurement of polynucleotide frequency versus abundance absorbance (as measured absorbance) across each individual cluster, with control clusters identified by a box.
- [0047] **Figure 18** is a plot of measurements of polynucleotide frequency versus abundance (as measured absorbance) across four individual clusters.
- [0048] **Figure 19A** is a plot of on frequency versus error rate across a plate from synthesis of 29,040 unique polynucleotides from 240 clusters, each cluster having 121 polynucleotides.
- [0049] **Figure 19B** is a plot of measurement of polynucleotide error rate versus frequency across each individual cluster, with control clusters identified by a box.
- [0050] **Figure 20** is a plot of measurements of polynucleotide frequency versus error rate across four clusters.
- [0051] **Figure 21** is a plot of GC content as a measure of the number of polynucleotides versus percent per polynucleotide.
- [0052] **Figure 22** is a plot of percent coverage verses read depth for an enrichment and sequencing analysis showing the performance of probe panels: Library 1 (757 kb) and Library 2 (803 kb).
- [0053] **Figure 23A** is a schematic of universal blockers.
- [0054] **Figure 23B** is a schematic of LNA blocker designs.
- [0055] **Figure 24** is a graph of on-target performance across for various index designs.
- [0056] **Figure 25** is a graph of on-target performance across for various panel sizes.
- [0057] **Figure 26A** is a graph of percentage of reads in each custom panel achieving 30x coverage.
- [0058] **Figure 26B** is a graph of uniformity (fold-80) of each custom panel.
- [0059] **Figure 27A** shows performance data using 810 kb panel.
- [0060] **Figure 27B** shows multiplexing performance for three panels at 1-, 8-, or 16-plex.
- [0061] **Figure 27C** shows effects of PCR cycles on uniformity.
- [0062] **Figure 27D** shows effects of library input mass on capture.
- [0063] **Figures 28A-28I** show reproducibility between custom panels. **FIG. 28A** shows quality of 800 kb panels. **FIG. 28B** shows enrichment performance of 800 kb panels. **FIG. 28C** shows

reproducibility of probe representation within same synthesis and different amplifications. **FIG. 28D** shows reproducibility of probe representation between syntheses. **FIG. 28E** shows lot to lot reproducibility capture per probe. **FIGS. 28F-28I** show reproducibility of probe target enrichment performance between syntheses. **FIG. 28F** shows lot to lot reproducibility for percent off-target capture. **FIG. 28G** shows lot to lot reproducibility for percent duplicates. **FIG. 28H** shows lot to lot reproducibility for the fraction of target bases with greater than 30X coverage. **FIG. 28I** shows lot to lot reproducibility for fold-80 base penalty.

[0064] **Figure 29A** is a schematic of adding or enhancing content to custom panels.

[0065] **Figure 29B** is a graph of uniformity (fold-80) comparing a panel with and without added content.

[0066] **Figure 29C** is a graph of duplicate rate comparing a panel with and without added content.

[0067] **Figure 29D** is a graph of percent on rate comparing a panel with and without added content.

[0068] **Figure 29E** is a graph of percent target coverage comparing a panel with and without added content, and comparator enrichment kits.

[0069] **Figure 29F** is a graph of 80-fold base penalty comparing a panel with and without added content, and comparator enrichment kits.

[0070] **Figure 30A** shows a design of control and variant panels.

[0071] **Figures 30B-30C** show distribution of mismatches on probe performance. Distribution of relative capture efficiency for probes with a single mismatch (gray) and probes with multiple mismatches (green lines; the number of mismatches is indicated in the left top corner) is shown. Solid line depicts the distribution for probes with randomly distributed mismatches (RND), and the dotted line indicates the distribution for probes with continuous mismatches (CONT). **FIG. 30B** shows a graph of probes with 3, 5, 10 or 15 mismatches (left to right). **FIG. 30C** shows a graph of probes with 20, 30, or 50 mismatches (left to right).

[0072] **Figure 30D** shows effect of temperature on capture efficiency.

[0073] **Figures 30E-30F** shows efficiency prediction for the design of 450 whole genome Zika isolates from human samples (**Figure 30E**) and all CpG islands in the human genome (**Figure 30F**).

[0074] **Figures 31A-31C** show graphs of standard vs. adaptive probe designs. **FIG. 31A** shows a comparison of standard and adaptive probe designs for percent off target rates. **FIG. 31B** shows a plot of the percent off-target reads which correlates predicted effects of selective probe removal with experimental results of selective probe removal. Various amounts of the worst performing

probes were removed from an exome capture library. **FIG. 31C** shows a graph of the percent off target as a function of selective removal of no probes (base/control), 0.4% of probes (increased), 1.7% of probes (moderate), or 3.3% of probes (strong) from an exome capture library.

[0075] **Figure 32A** shows a graph of depth coverage as percent target bases at coverage of the exome panel alone or with the RefSeq panel added.

[0076] **Figures 32B-32F** depict graphs of various enrichment/capture sequencing metrics for a standard exome panel vs. the exome panel combined with the RefSeq panel in both singleplex and 8-plex experiments. **FIG. 32B** shows a graph of specificity as percent off target for the exome panel alone or with the RefSeq panel added. **FIG. 32C** shows a graph of uniformity for the exome panel alone or with the RefSeq panel added. **FIG. 32D** shows a graph of library size for the exome panel alone or with the RefSeq panel added. **FIG. 32E** shows a graph of duplicate rate for the exome panel alone or with the RefSeq panel added. **FIG. 32F** shows a graph of coverage rate for the exome panel alone or with the RefSeq panel added.

[0077] **Figure 33A** depicts an exemplary hybridization reaction, wherein nucleic acids concentrate near a gas-liquid interface.

[0078] **Figure 33B** depicts an exemplary hybridization reaction, wherein nucleic acids are prevented from concentrating near a gas-liquid interface by an additive.

[0079] **Figure 33C** depicts a plot of the percent off target vs. binding buffer comprising various additives for an enrichment and sequencing analysis.

[0080] **Figure 34A** depicts a plot of the percent off target vs. various buffers comprising different additives for an enrichment and sequencing analysis.

[0081] **Figure 34B** depicts a plot of the percent off bait vs. number of washes and the presence of mineral oil for an enrichment and sequencing analysis.

[0082] **Figure 34C** depicts a plot of AT dropout vs. GC dropout for conditions comprising different wash numbers and the presence or absence of mineral oil for an enrichment and sequencing analysis.

[0083] **Figure 34D** depicts a plot of HS library size for conditions comprising different numbers of washes with wash buffer 1 and the presence or absence of mineral oil for an enrichment and sequencing analysis.

[0084] **Figure 34E** depicts a plot of 80 fold base penalty for conditions comprising different numbers of washes with wash buffer 1 and the presence or absence of mineral oil for an enrichment and sequencing analysis.

[0085] **Figure 35A** depicts a plot of the percent off bait vs. tube transfer and the presence of Polymer A for an enrichment and sequencing analysis.

- [0086] **Figure 35B** depicts a plot of HS library size vs. tube transfer and the presence of Polymer A for an enrichment and sequencing analysis.
- [0087] **Figure 36** depicts a plot of percent off target for conditions comprising different levels of agitation and methods of aspiration for an enrichment and sequencing analysis.
- [0088] **Figure 37A** is a plot of depth of coverage achieved (% target bases at 30x) vs. various hybridization times using either standard or fast hybridization buffers.
- [0089] **Figure 37B** is a plot of fold 80 base penalty vs. various hybridization times using either standard or fast hybridization buffers.
- [0090] **Figure 37C** is a plot of percent off bait vs. various hybridization times using either standard or fast hybridization buffers.
- [0091] **Figure 37D** is a plot of HS library size vs. various hybridization times using either standard or fast hybridization buffers.
- [0092] **Figure 37E** is a plot of percent duplicates vs. various hybridization times using either standard or fast hybridization buffers.
- [0093] **Figure 38** depicts comparison of workflows using traditional hybridization buffers vs. a streamlined target enrichment (top) workflow that can be completed in as little as 5–9 hours.
- [0094] **Figure 39A** is a series of plots for Fold-80 base penalty, On-target rate, and target bases with greater than 30X coverage obtained using a fast hybridization buffer with a 33.1 Mb exome enrichment probe library.
- [0095] **Figure 39B** is a plot of the fraction of target bases with greater than 30X coverage for 1 plex, and 8-plex experiments using either a 33.1 Mb exome probe panel or a 0.8 Mb custom cancer panel.
- [0096] **Figure 39C** is a plot of 80 fold base penalties vs. various FFPE samples.
- [0097] **Figure 39D** is a plot of duplicate rate percentage vs. various FFPE samples.
- [0098] **Figure 39E** is a plot of the percentage of target bases with greater than 20X coverage vs. various FFPE samples.
- [0099] **Figure 39F** is a plot of AT and GC dropout rates vs. various FFPE samples.
- [0100] **Figure 39G** is a plot of coverage (log vs. median) vs. position on chromosome 1 for an FFPE sample.
- [0101] **Figure 40** is a plot of exome qualitative values vs. wash buffer 1 temperature for an experiment utilizing the fast hybridization buffer.
- [0102] **Figure 41** is a plot of percent off bait for various blocker designs which target top or bottom strands of the adapters.

[0103] **Figure 42A** are plots of pre-hybridization concentration, pre-capture size, post-capture concentration, and observed size for a library generated using a tagmentation method and various configurations of universal blockers.

[0104] **Figure 42B** are plots of median insert size and HS library size for a library generated using a tagmentation method and various configurations of universal blockers.

[0105] **Figure 42C** are plots of sequencing metrics including Fold 80 base penalty and percent off bait for a library generated using a tagmentation method and various configurations of universal blockers.

[0106] **Figure 42D** are plots of sequencing metrics including percent target bases with at least 30X coverage, and duplication rate for a library generated using a tagmentation method and various configurations of universal blockers.

[0107] **Figure 42E** are plots of sequencing metrics including AT and GC dropout rates and zero coverage target percentage for a library generated using a tagmentation method and various configurations of universal blockers.

[0108] **Figure 43** is a plot of percent off bait for a library generated using a tagmentation method and various configurations of universal blockers.

[0109] **Figure 44** is a plot of melt curves in the presence or absence of blockers.

DETAILED DESCRIPTION

[0110] Provided herein are methods and compositions for designing, synthesizing and controlling hybridization events within large polynucleotide libraries. Capture probe libraries are designed and synthesized to bind to specific target sequences in a sample population of polynucleotides, which enables any number of downstream applications such as diagnostic assays, sequencing, selection assays, or other method that requires a hybridization step. Factors contributing to the overall efficiency of hybridization include capture probe stoichiometry/uniformity, capture probe labeling, dilution effects, adapter dimerization, and hybridization conditions. Another factor contributing to the overall efficiency of hybridization is the local concentration of non-target nucleic acids at an air-water interface. Such concentrations herein are controlled through the presence of additives and washing methods, leading to improved hybridization. Further provided are buffer compositions which allow reductions in hybridization times while achieving comparable sequencing depth. Further provided are blocker polynucleotides that decrease the percentage of off-target (or off-bait) reads.

[0111] Definitions

[0112] Throughout this disclosure, numerical features are presented in a range format. It should be understood that the description in range format is merely for convenience and brevity and should not be construed as an inflexible limitation on the scope of any embodiments. Accordingly, the description of a range should be considered to have specifically disclosed all the possible subranges as well as individual numerical values within that range to the tenth of the unit of the lower limit unless the context clearly dictates otherwise. For example, description of a range such as from 1 to 6 should be considered to have specifically disclosed subranges such as from 1 to 3, from 1 to 4, from 1 to 5, from 2 to 4, from 2 to 6, from 3 to 6 etc., as well as individual values within that range, for example, 1.1, 2, 2.3, 5, and 5.9. This applies regardless of the breadth of the range. The upper and lower limits of these intervening ranges may independently be included in the smaller ranges, and are also encompassed within the invention, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in the invention, unless the context clearly dictates otherwise.

[0113] The terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting of any embodiment. As used herein, the singular forms “a,” “an” and “the” are intended to include the plural forms as well, unless the context clearly indicates otherwise. It will be further understood that the terms “comprises” and/or “comprising,” when used in this specification, specify the presence of stated features, integers, steps, operations, elements, and/or components, but do not preclude the presence or addition of one or more other features, integers, steps, operations, elements, components, and/or groups thereof. As used herein, the term “and/or” includes any and all combinations of one or more of the associated listed items.

[0114] Unless specifically stated or obvious from context, as used herein, the term “about” in reference to a number or range of numbers is understood to mean the stated number and numbers +/- 10% thereof, or 10% below the lower listed limit and 10% above the higher listed limit for the values listed for a range.

[0115] As used herein, the terms “preselected sequence”, “predefined sequence” or “predetermined sequence” are used interchangeably. The terms mean that the sequence of the polymer is known and chosen before synthesis or assembly of the polymer. In particular, various aspects of the invention are described herein primarily with regard to the preparation of nucleic acids molecules, the sequence of the oligonucleotide or polynucleotide being known and chosen before the synthesis or assembly of the nucleic acid molecules.

[0116] The term nucleic acid encompasses double- or triple-stranded nucleic acids, as well as single-stranded molecules. In double- or triple-stranded nucleic acids, the nucleic acid strands need not be coextensive (i.e., a double-stranded nucleic acid need not be double-stranded along the entire length of both strands). Nucleic acid sequences, when provided, are listed in the 5' to 3' direction, unless stated otherwise. Methods described herein provide for the generation of isolated nucleic acids. Methods described herein additionally provide for the generation of isolated and purified nucleic acids. The length of polynucleotides, when provided, are described as the number of bases and abbreviated, such as nt (nucleotides), bp (bases), kb (kilobases), or Gb (gigabases).

[0117] Provided herein are methods and compositions for production of synthetic (i.e. de novo synthesized or chemically synthesizes) polynucleotides. The term oligonucleic acid, oligonucleotide, oligo, and polynucleotide are defined to be synonymous throughout. Libraries of synthesized polynucleotides described herein may comprise a plurality of polynucleotides collectively encoding for one or more genes or gene fragments. In some instances, the polynucleotide library comprises coding or non-coding sequences. In some instances, the polynucleotide library encodes for a plurality of cDNA sequences. Reference gene sequences from which the cDNA sequences are based may contain introns, whereas cDNA sequences exclude introns. Polynucleotides described herein may encode for genes or gene fragments from an organism. Exemplary organisms include, without limitation, prokaryotes (e.g., bacteria) and eukaryotes (e.g., mice, rabbits, humans, and non-human primates). In some instances, the polynucleotide library comprises one or more polynucleotides, each of the one or more polynucleotides encoding sequences for multiple exons. Each polynucleotide within a library described herein may encode a different sequence, i.e., non-identical sequence. In some instances, each polynucleotide within a library described herein comprises at least one portion that is complementary to sequence of another polynucleotide within the library. Polynucleotide sequences described herein may be, unless stated otherwise, comprise DNA or RNA. A polynucleotide library described herein may comprise at least 10, 20, 50, 100, 200, 500, 1,000, 2,000, 5,000, 10,000, 20,000, 30,000, 50,000, 100,000, 200,000, 500,000, 1,000,000, or more than 1,000,000 polynucleotides. A polynucleotide library described herein may have no more than 10, 20, 50, 100, 200, 500, 1,000, 2,000, 5,000, 10,000, 20,000, 30,000, 50,000, 100,000, 200,000, 500,000, or no more than 1,000,000 polynucleotides. A polynucleotide library described herein may comprise 10 to 500, 20 to 1000, 50 to 2000, 100 to 5000, 500 to 10,000, 1,000 to 5,000, 10,000 to 50,000, 100,000 to 500,000, or to 50,000 to 1,000,000 polynucleotides. A polynucleotide library described herein may comprise about 370,000; 400,000; 500,000 or more different polynucleotides.

[0118] Methods for Hybridization

[0119] Described herein are methods of hybridization designed to improve the efficiency and accuracy of capture probes binding to target nucleic acids (**FIG. 33A-33C**). Such methods comprise changing the stoichiometry of individual or groups of capture probes in a capture probe library, supplementing a capture probe library with capture probes targeting alternative sequences, preventing off-target binding interactions by use of blocking polynucleotides comprising nucleobase analogues, and partial labeling of capture probe libraries. Also provided are methods to reduce off-target (or off-bait) sequencing metrics (**FIG. 33A-33B**). Without being bound by theory, factors which contribute to off-target rates include the ability of probes to freely interact/hybridize with the target nucleic acids, as well as the efficiency of washing away non-hybridized, non-target nucleic acids. These factors may be influenced by a non-uniform concentration of nucleic acids in a solution, such as at a gas-liquid interface. Such hybridization reactions may be improved by addition of additives that prevent such non-uniform concentrations, and/or by controlled manipulation of such solutions. Additives or buffers in some instances also result in decreased hybridization times (**FIG. 38**). Such improvements often lead to significant decreases for off-target rates with smaller polynucleotide libraries (e.g., less than 1Mb), but are also used with larger libraries, such as exome libraries. Also provided herein are de novo synthesized polynucleotides for use in hybridization to genomic DNA, for example in the context of a sequencing process. In a first step of an exemplary sequencing workflow (**FIG. 2**), a nucleic acid sample **208** comprising target polynucleotides is fragmented by mechanical or enzymatic shearing to form a library of fragments **209**. Adapters **215** optionally comprising primer sequences and/or barcodes are ligated to form an adapter-tagged library **210**. This library is then optionally amplified, and hybridized with target binding polynucleotides **217** which hybridize to target polynucleotides, along with blocking polynucleotides **216** that prevent hybridization between target binding polynucleotides **217** and adapters **215**. Capture of target polynucleotide-target binding polynucleotide hybridization pairs **212**, and removal of target binding polynucleotides **217** allows isolation/enrichment of target polynucleotides **213**, which are then optionally amplified and sequenced **214**. In some instances the addition of blockers to the hybridization reaction reduces off-target rates by preventing adapter-adapter interactions (**FIG. 1B**).

[0120] A first method described herein comprises changing the stoichiometry of individual or groups of capture probes in a capture probe library. For example, an enrichment and sequencing analysis is run on a nucleic acid sample, and one or more regions of the targeted sequences comprise less than desired read depth (**FIG. 1A**, black bar, left). Addition of a second “spike in,” targeted, or (targeted) panel library increases the read depth at these less than average read depth

regions (**FIG. 1A**, black bars, right). Such regions are in some instances regions that are already targeted by a larger capture probe library, for example an exome probe or other library.

Alternatively or in combination, such regions are not already targeted by the larger probe library, and the targeted panel library adds additional sequencing information to new regions of the nucleic acid sample. Exemplary panels in some instances target genes with specific function (development, disease state, pain, physical trait, or other function), or non-coding regions such as introns. In some embodiments, the panels comprise target genes involved in disease including but not limited to cancer, neurodegenerative disease, and mitochondrial disorders.

[0121] A second method described herein comprises the use of universal blockers to prevent off-target binding of capture probes to adapters ligated to genomic fragments **101**, or adapter-adaptor hybridization (**FIG. 1B**). Adapter blockers used for preventing off-target hybridization may target a portion or the entire adapter **102**. In some instances, specific blockers are used that are complementary to a portion of the adapter **102** that includes the unique index sequence **103**. In cases where the adapter-tagged genomic library **100** comprises a large number of different indices **103**, it can be beneficial to design blockers which either do not target the index sequence **103**, or do not hybridize strongly to it. For example, a “universal” blocker **104** targets a portion of the adapter **102** that does not comprise an index sequence (index independent), which allows a minimum number of blockers to be used regardless of the number of different index sequences employed (**FIG. 1B**). In some instances, no more than 8 universal blockers are used. In some instances, 4 universal blockers are used. In some instances, 3 universal blockers are used. In some instances, 2 universal blockers are used. In some instances, 1 universal blocker is used. In an exemplary arrangement, 4 universal blockers are used with adapters comprising at least 4, 8, 16, 32, 64, 96, or at least 128 different index sequences. In some instances, the different index sequences comprises at least or about 4, 6, 8, 10, 12, 14, 16, 18, 20, or more than 20 base pairs (bp). In some instances, a universal blocker is not configured to bind to a barcode sequence. In some instances, a universal blocker partially binds to a barcode sequence. In some instances, a universal blocker which partially binds to a barcode sequence further comprises nucleotide analogs, such as those that increase the T_m of binding to the adapter (e.g., LNAs or BNAs).

[0122] The universal blockers may be used with panel libraries of varying size. In some embodiments, the panel libraries comprises at least or about 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 1.0, 2.0, 4.0, 8.0, 10.0, 12.0, 14.0, 16.0, 18.0, 20.0, 22.0, 24.0, 26.0, 28.0, 30.0, 40.0, 50.0, 60.0, or more than 60.0 megabases (Mb).

[0123] Blockers as described herein may improve on-target performance. In some embodiments, on-target performance is improved by at least or about 5%, 10%, 15%, 20%, 25%,

30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or more than 95%. In some embodiments, the on-target performance is improved by at least or about 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or more than 95% for various index designs. In some embodiments, the on-target performance is improved by at least or about 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or more than 95% is improved for various panel sizes.

[0124] Blockers may contain any number of different nucleobases (DNA, RNA, etc.), nucleobase analogues (non-canonical), or non-nucleobase linkers or spacers. For example, a blocker comprises one or more nucleobase analogues or other groups that enhance hybridization (T_m) between the blocker and the adapter. Nucleobase analogues and other groups include but are not limited to locked nucleic acids (LNAs), bicyclic nucleic acids (BNAs), C5-modified pyrimidine bases, 2'-O-methyl substituted RNA, peptide nucleic acids (PNAs), glycol nucleic acid (GNAs), threose nucleic acid (TNAs), xenonucleic acids (XNAs) morpholino backbone-modified bases, minor groove binders (MGBs), spermine, G-clamps, or a anthraquinone (Uaq) caps. In instances, blockers comprise spacer elements that connect two polynucleotide chains. In some instances, blockers comprise one or more nucleobase analogues selected from **Table 1**. In some instances, such nucleobase analogues are added to control the T_m of a blocker.

Table 1

Base	A	T	G	C	U
Locked Nucleic Acid (LNA)					
Bridged Nucleic Acid* (BNA)					

*R is H or Me.

[0125] A third method described herein comprises addition of one or more additives to a hybridization reaction to decrease off-target rates. Additives are added at any step in the hybridization workflow, such as during hybridization, or during washing steps. In an exemplary arrangement, additives are added to buffers such as hybridization buffers, binding buffers, wash buffers, or any combination thereof. In some instances, additives are added to two or more buffers,

such as a hybridization buffer and a binding buffer. An exemplary hybridization reaction **3000** in a container **3001** is shown in **FIG. 33A**, wherein a solution **3002** comprising nucleic acid targets and polynucleotide probes is in contact with a gas **3004**, forming a gas-liquid interface **3005** (such as an air-water interface). Such hybridization reactions are often hindered by a higher concentration of nucleic acids at the area **3003** adjacent to the gas-liquid interface **3005**, which in some instances limits the uniform hybridization of probes to target nucleic acids, or prevents non-target nucleic acids from being removed in a wash step. Addition of additives, such as additive **3006**, in some instances reduces the concentration of non-target nucleic acids at the area **3003** adjacent to the gas-liquid interface **3005**, which results in decreased off-target binding. In some instances, addition of at least one additive results in a decrease in random off-target binding.

[0126] Methods described herein may comprise one or more washing steps or tube transfer steps. In some instances, washing or tube transfers are combined with the use of additives. In some instances, 1, 2, 3, 4, or more than 4 washes are performed after capture of target sequences on a solid support. In some instances, one or more wash steps is substituted with a tube transfer, wherein the captured target sequences are transferred to an unused tube or other container. In some instances, tube transfers are used in combination with wash steps. In some instances, 1, 2, 3, 4, or more than 4 tube transfers are performed during the methods described herein.

[0127] Additives for hybridization may include any number of chemical agents, or mixtures thereof that influence the structure or solubility of polynucleotides. Additives for hybridization include salts, oils, waxes, nucleotides (or nucleotide analogues), polymers, kosmotropes, chaotropes, or other additive that influences local concentrations of polynucleotides. Oils include but are not limited to petroleum-based agents (e.g., light oil, jet fuel, gasoline, kerosene, naphtha, petroleum ether, petroleum spirits, mineral oil, light mineral oil, white mineral oil), plant-based oils (olive oil, vegetable oil, soybean oil, or other plant-based oil). Polymers in some instances are hydrophobic (e.g., polysilanes) or hydrophilic (polyethers such as polyethylene glycol). In some instances, oils comprise alkanes, cycloalkanes, or silanes (silicon oils). In some instances, additives comprise liquid polymers, such as high-molecular weight, low vapor pressure, and/or low water solubility polymers. In some instances, chaotropes include alcohols (e.g., n-butanol, ethanol), guanidinium chloride, lithium perchlorate, lithium acetate, magnesium chloride, phenol, 2-propanol, sodium dodecyl sulfate, thiourea, urea, thiocyanate, or other agent that disrupts hydrogen bonding networks. In some instances kosmotropes include carbonate, sulfate, hydrogen phosphate, magnesium, lithium, zinc, aluminum, or other agent that stabilizes hydrogen bonding networks.

[0128] Additives described herein may be present at any concentration suitable for reducing off-target binding. Such concentrations are often represented as a percent by weight, percent by

volume, or percent weight per volume. For example, an additive is present at about 0.0001%, 0.0002%, 0.0005%, 0.0008%, 0.001%, 0.002%, 0.005%, 0.008%, 0.01%, 0.02%, 0.05%, 0.08%, 0.1%, 0.2%, 0.5%, 0.8%, 1%, 1.2%, 1.5%, 1.8%, 2%, 5%, 10%, 20%, or about 30%. In some instances, an additive is present at no more than 0.0001%, 0.0002%, 0.0005%, 0.0008%, 0.001%, 0.002%, 0.005%, 0.008%, 0.01%, 0.02%, 0.05%, 0.08%, 0.1%, 0.2%, 0.5%, 0.8%, 1%, 1.2%, 1.5%, 1.8%, 2%, 5%, 10%, 20%, or no more than 30%. In some instances, an additive is present in at least 0.0001%, 0.0002%, 0.0005%, 0.0008%, 0.001%, 0.002%, 0.005%, 0.008%, 0.01%, 0.02%, 0.05%, 0.08%, 0.1%, 0.2%, 0.5%, 0.8%, 1%, 1.2%, 1.5%, 1.8%, 2%, 5%, 10%, 20%, or at least 30%. In some instances, an additive is present at 0.0001%-10%, 0.0002%-5%, 0.0005%-1.5%, 0.0008%-1%, 0.001%-0.2%, 0.002%-0.08%, 0.005%-0.02%, or 0.008%-0.05%. In some instances, an additive is present at 0.005%-0.1%. In some instances, an additive is present at 0.05%-0.1%. In some instances, an additive is present at 0.005%-0.6%. In some instances, an additive is present at 1%-30%, 5%-25%, 10%-30%, 15%-30%, or 1%-15%. Liquid additives may be present as a percentage of the total reaction volume. In some instances, an additive is about 10%, 20%, 30%, 40%, 50%, 60%, 75%, or about 90% of the total volume. In some instances, an additive is at least 10%, 20%, 30%, 40%, 50%, 60%, 75%, or at least 90% of the total volume. In some instances, an additive is no more than 10%, 20%, 30%, 40%, 50%, 60%, 75%, or no more than 90% of the total volume. In some instances, an additive is 5%-75%, 5%-65%, 5%-55%, 10%-50%, 15%-40%, 20%-50%, 20%-30%, 25%-35%, 5%-35%, 10%-35%, or 20%-40% of the total volume. In some instances, an additive is 25%-45% of the total volume.

[0129] A fourth method provided herein comprises controlled fluid transfer that results in a decrease of off-target rates. Without being bound by theory, such controlled transfer minimizes contamination of non-hybridized (non-target) nucleic acids with target nucleic acids. In some instances, a controlled transfer decreases local non-uniform concentration of nucleic acids in a solution, such as at a gas-liquid interface. In some instances, non-target nucleic acids are present at a higher concentration near a gas-liquid interface **3005**. In some instances, the interface is an air-water interface. In this method, controlled fluid transfer of liquid near or in the local area **3003** adjacent to the gas-liquid interface provides for selective removal of off-target nucleic acids during hybridization and/or capture steps. For example, liquid is removed only from this local area in a continuous fashion, until all liquid **3002** is removed. The local area is in some instances defined as a volume of liquid near the gas-liquid interface, and related to the total volume of the liquid. For example, the local area volume is about the upper 10% of the total volume. In some instances, the local area volume is about the upper 1%, 2%, 5%, 8%, 10%, 15%, 20%, or about 25% of the total volume. In some instances, the local area volume is about the upper 1%-25%, 2%-20%, 5%-15%,

8%-12%, 10%-25%, 1%-10%, 20%, or about 25% of the total volume. The location of liquid removal in some instances depends on the surface area of the gas-liquid interface. In some instances, a higher interface surface area decreases the local area volume from which liquid is removed.

[0130] Various temperatures and times are used for hybridization of probes to target nucleic acids. In some instances, the hybridization temperature is at least 50, 60, 70, 80, 90, or at least 95 C. In some instances, the hybridization temperature is about 50, 55, 60, 65, 70, 75, 80, 85, or 90 C. In some instances, the hybridization temperature is 40-50 C, 40-80 C, 50-70 C, 50-80 C, 60-90 C, 55-70 C, or 60-80 C. In some instances, probes are hybridized for no more than 5, 10, 15, 20, 30, 45, 60, or no more than 60 minutes. In some instances, probes are hybridized for about 0.1, 0.2, 0.3, 0.5, 0.75, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or about 12 hours. In some instances, probes are hybridized for about 10 min to 8 hours, 15 min to 6 hours, 20 min to 4 hrs, 15 min to 2 hrs, 10 min to 6 hrs, 30 min to 5 hrs, 1 hr to 8 hrs, or 2 hrs to 10 hrs.

[0131] Various temperatures and times are used for wash buffers used with the methods and compositions described herein. Washes in some instances are performed when hybridized nucleic acids are bound to a solid support. In some instances a wash buffer is pre-heated to about 50, 55, 57, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, or 80 C prior to use. In some instances a wash buffer is pre-heated to 50-80, 50-75, 50-70, 60-75, 60-70, 65-75, 70-80, 67-74, or 55-75 C prior to use. In some instances, more than one wash is performed, and each wash buffer used is the same or a different temperature. In some instances a first wash buffer (or wash buffer 1) is pre-heated to about 50, 55, 57, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, or 80 C prior to use. In some instances a first wash buffer is pre-heated to 50-80, 50-75, 50-70, 60-75, 60-70, 65-75, 70-80, 67-74, or 55-75 C prior to use.

[0132] Hybridization Blockers

[0133] Blockers may comprise any number of nucleobase analogues (such as LNAs or BNAs), depending on the desired hybridization T_m . For example, a blocker comprises 20 to 40 nucleobase analogues. In some instances, a blocker comprises 8 to 16 nucleobase analogues. In some instances, a blocker comprises at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or at least 12 nucleobase analogues. In some instances, a blocker comprises about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, or about 16 nucleobase analogues. In some instances, the number of nucleobase analogous is expressed as a percent of the total bases in the blocker. For example, a blocker comprises at least 1%, 2%, 5%, 10%, 12%, 18%, 24%, 30%, or more than 30% nucleobase analogues. In some instances, the blocker comprising a nucleobase analogue raises the T_m in a range of about 2 °C to about 8 °C for each nucleobase analogue. In some instances, the T_m is raised by at least or about 1 °C, 2 °C, 3 °C,

4 °C, 5 °C, 6 °C, 7 °C, 8 °C, 9 °C, 10 °C, 12 °C, 14 °C, or 16 °C for each nucleobase analogue. Such blockers in some instances are configured to bind to the top or “sense” strand of an adapter. Blockers in some instances are configured to bind to the bottom or “anti-sense” strand of an adapter. In some instances a set of blockers includes sequences which are configured to bind to both top and bottom strands of an adapter. Additional blockers in some instances are configured to the complement, reverse, forward, or reverse complement of an adapter sequence. In some instances, a set of blockers targeting a top (binding to the top) or bottom strand (or both) is designed and tested, followed by optimization, such as replacing a top blocker with a bottom blocker, or a bottom blocker with a top blocker.

[0134] Blockers may be any length, depending on the size of the adapter or hybridization T_m . For example, blockers are 20 to 50 bases in length. In some instances, blockers are 25 to 45 bases, 30 to 40 bases, 20 to 40 bases, or 30 to 50 bases in length. In some instances, blockers are 25 to 35 bases in length. In some instances blockers are at least 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, or at least 35 bases in length. In some instances, blockers are no more than 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, or no more than 35 bases in length. In some instances, blockers are about 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, or about 35 bases in length. In some instances, blockers are about 50 bases in length. A set of blockers targeting an adapter-tagged genomic library fragment in some instances comprises blockers of more than one length. Two blockers are in some instances tethered together with a linker. Various linkers are well known in the art, and in some instances comprise alkyl groups, polyether groups, amine groups, amide groups, or other chemical group. In some instances, linkers comprise individual linker units, which are connected together (or attached to blocker polynucleotides) through a backbone such as phosphate, thiophosphate, amide, or other backbone. In an exemplary arrangement, a linker spans the index region between a first blocker that each targets the 5' end of the adapter sequence and a second blocker that targets the 3' end of the adapter sequence. In some instances, capping groups are added to the 5' or 3' end of the blocker to prevent downstream amplification. Capping groups variously comprise polyethers, polyalcohols, alkanes, or other non-hybridizable group that prevents amplification. Such groups are in some instances connected through phosphate, thiophosphate, amide, or other backbone. In some instances, one or more blockers are used. In some instances, at least 4 non-identical blockers are used. In some instances, a first blocker spans a first 3' end of an adaptor sequence, a second blocker spans a first 5' end of an adaptor sequence, a third blocker spans a second 3' end of an adaptor sequence, and a fourth blockers spans a second 5' end of an adaptor sequence. In some instances a first blocker is at least 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, or at least 35 bases in length. In some instances a second blocker is at least 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34,

or at least 35 bases in length. In some instances a third blocker is at least 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, or at least 35 bases in length. In some instances a fourth blocker is at least 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, or at least 35 bases in length. In some instances, a first blocker, second blocker, third blocker, or fourth blocker comprises a nucleobase analogue. In some instances, the nucleobase analogue is LNA.

[0135] The design of blockers may be influenced by the desired hybridization T_m to the adapter sequence. In some instances, non-canonical nucleic acids (for example locked nucleic acids, bridged nucleic acids, or other non-canonical nucleic acid or analog) are inserted into blockers to increase or decrease the blocker's T_m . In some instances, the T_m of a blocker is calculated using a tool specific to calculating T_m for polynucleotides comprising a non-canonical amino acid. In some instances, a T_m is calculated using the Exiqon TM online prediction tool. In some instances, blocker T_m described herein are calculated in-silico. In some instances, the blocker T_m is calculated in-silico, and is correlated to experimental in-vitro conditions. Without being bound by theory, an experimentally determined T_m may be further influenced by experimental parameters such as salt concentration, temperature, presence of additives, or other factor. In some instances, T_m described herein are in-silico determined T_m that are used to design or optimize blocker performance. In some instances, T_m values are predicted, estimated, or determined from melting curve analysis experiments. In some instances, blockers have a T_m of 70 degrees C to 99 degrees C. In some instances, blockers have a T_m of 75 degrees C to 90 degrees C. In some instances, blockers have a T_m of at least 85 degrees C. In some instances, blockers have a T_m of at least 70, 72, 75, 77, 80, 82, 85, 88, 90, or at least 92 degrees C. In some instances, blockers have a T_m of about 70, 72, 75, 77, 80, 82, 85, 88, 90, 92, or about 95 degrees C. In some instances, blockers have a T_m of 78 degrees C to 90 degrees C. In some instances, blockers have a T_m of 79 degrees C to 90 degrees C. In some instances, blockers have a T_m of 80 degrees C to 90 degrees C. In some instances, blockers have a T_m of 81 degrees C to 90 degrees C. In some instances, blockers have a T_m of 82 degrees C to 90 degrees C. In some instances, blockers have a T_m of 83 degrees C to 90 degrees C. In some instances, blockers have a T_m of 84 degrees C to 90 degrees C. In some instances, a set of blockers have an average T_m of 78 degrees C to 90 degrees C. In some instances, a set of blockers have an average T_m of 80 degrees C to 90 degrees C. In some instances, a set of blockers have an average T_m of at least 80 degrees C. In some instances, a set of blockers have an average T_m of at least 81 degrees C. In some instances, a set of blockers have an average T_m of at least 82 degrees C. In some instances, a set of blockers have an average T_m of at least 83 degrees C. In some instances, a set of blockers have an average T_m of at least 84 degrees C. In some instances, a set of blockers have an average T_m of at least 86 degrees C. Blocker T_m are in some instances modified as a result of other

components described herein, such as use of a fast hybridization buffer and/or hybridization enhancer.

[0136] The molar ratio of blockers to adapter targets may influence the off-bait (and subsequently off-target) rates during hybridization. The more efficient a blocker is at binding to the target adapter, the less blocker is required. Blockers described herein in some instances achieve sequencing outcomes of no more than 20% off-target reads with a molar ratio of less than 20:1 (blocker:target). In some instances, no more than 20% off-target reads are achieved with a molar ratio of less than 10:1 (blocker:target). In some instances, no more than 20% off-target reads are achieved with a molar ratio of less than 5:1 (blocker:target). In some instances, no more than 20% off-target reads are achieved with a molar ratio of less than 2:1 (blocker:target). In some instances, no more than 20% off-target reads are achieved with a molar ratio of less than 1.5:1 (blocker:target). In some instances, no more than 20% off-target reads are achieved with a molar ratio of less than 1.2:1 (blocker:target). In some instances, no more than 20% off-target reads are achieved with a molar ratio of less than 1.05:1 (blocker:target).

[0137] A third method described herein comprises improving the efficiency of polynucleotide probe libraries by selectively labeling only a portion of the probes (**FIG. 1C**). If a library of polynucleotide probes that is fully labeled is diluted, the result is often an increase in off bait, and a decrease in HS library size. By keeping the total ratio of polynucleotides to target genomic sequences constant, all target genomic sequences are still bound to a complementary probe and inter or intramolecular hybridization of such sequences is reduced. In an exemplary workflow, a library of sample polynucleotides **109** is hybridized with a plurality of probes, some of which are labeled probes **118** or unlabeled probes **117**. The hybridization mixture **119** can then be subjected to further purification to isolate target polynucleotides binding to labeled probes **118**. The percentage of labeled probes may vary depending on the application, library size, and genomic targets. For example, about 1%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% of all probes are labeled. In some instances at least 1%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or at least 90% of all probes are labeled. In some instances no more than 1%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or no more than 90% of all probes are labeled. In some instances, 10-90%, 20-80%, 30-70%, 40-50%, 1-40%, 20-60%, 40-70%, 50-90%, 60-99%, 70-99%, or 80-99% of all probes are labeled. In some instances, the label is a molecular tag, such as biotin or other molecular tag. In some instances, polynucleotide probe libraries comprising less than 15 % labeled probes results in less than 40% off bait. In some instances, polynucleotide probe libraries comprising less than 15 % labeled probes results in less than 40% off bait. Partial labeling of probes may also result in a decrease in AT and GC dropouts. For example, polynucleotide probe

libraries comprising 1-50% labeled probes results in less than 1.9% AT dropout. In some instances, polynucleotide probe libraries comprising 1-50% labeled probes results in less than 1.3% GC dropout. In some instances, polynucleotide probe libraries comprising 12.5-50% labeled probes results in less than 1.3% GC dropout. In some instances, polynucleotide probe libraries comprising 12.5-50% labeled probes results in less than 1.9% AT dropout.

[0138] Hybridization Buffers

[0139] Any number of buffers may be used with the hybridization methods described herein. For example, a buffer comprises numerous chemical components, such as polymers, solvents, salts, surfactants, or other component. In some instances, hybridization buffers decrease the hybridization times (e.g., “fast” hybridization buffers) required to achieve a given sequencing result or level of quality. Such components in some instances lead to improved hybridization outcomes, such as increased on-target rate, improved sequencing outcomes (e.g., sequencing depth or other metric), or decreased off-target rates. Such components may be introduced at any concentration to achieve such outcomes. In some instances, buffer components are added in specific order. For example, water is added first. In some instances, salts are added after water. In some instances, salts are added after thickening agents and surfactants. In some instances, hybridization buffers such as “fast” hybridization buffers described herein are used in conjunction with universal blockers and liquid polymer additives.

[0140] Hybridization buffers described herein may comprise solvents, or mixtures of two or more solvents. In some instances, a hybridization buffer comprises a mixture of two solvents, three solvents or more than three solvents. In some instances, a hybridization buffer comprises a mixture of an alcohol and water. In some instances, a hybridization buffer comprises a mixture of a ketone containing solvent and water. In some instances, a hybridization buffer comprises a mixture of an ethereal solvent and water. In some instances, a hybridization buffer comprises a mixture of a sulfoxide-containing solvent and water. In some instances, a hybridization buffer comprises a mixture of an amide-containing solvent and water. In some instances, a hybridization buffer comprises a mixture of an ester-containing solvent and water. In some instances, hybridization buffers comprise solvents such as water, ethanol, methanol, propanol, butanol, other alcohol solvent, or a mixture thereof. In some instances, hybridization buffers comprise solvents such as acetone, methyl ethyl ketone, 2-butanone, ethyl acetate, methyl acetate, tetrahydrofuran, diethyl ether, or a mixture thereof. In some instances, hybridization buffers comprise solvents such as DMSO, DMF, DMA, HMPA, or a mixture thereof. In some instances, hybridization buffers comprise a mixture of water, HMPA, and an alcohol. In some instances, two solvents are present at a 1:1, 1:2, 1:3, 1:4, 1:5, 1:8, 1:9, 1:10, 1:20, 1:50, 1:100, or 1:500 ratio.

[0141] Hybridization buffers described herein may comprise polymers. Polymers include but are not limited to thickening agents, polymeric solvents, dielectric materials, or other polymer. Polymers are in some instances hydrophobic or hydrophilic. In some instances, polymers are silicon polymers. In some instances, polymers comprise repeating polyethylene or polypropylene units, or a mixture thereof. In some instances, polymers comprise polyvinylpyrrolidone or polyvinylpyridine. In some instances, polymers comprise amino acids. For example, in some instances polymers comprise proteins. In some instances, polymers comprise casein, milk proteins, bovine serum albumin, or other protein. In some instances, polymers comprise nucleotides, for example, DNA or RNA. In some instances, polymers comprise polyA, polyT, Cot-1 DNA, or other nucleic acid. In some instances, polymers comprise sugars. For example, in some instances a polymer comprises glucose, arabinose, galactose, mannose, or other sugar. In some instances, a polymer comprises cellulose or starch. In some instances, a polymer comprises agar, carboxyalkyl cellulose, xanthan, guar gum, locust bean gum, gum karaya, gum tragacanth, gum Arabic. In some instances, a polymer comprises a derivative of cellulose or starch, or nitrocellulose, dextran, hydroxyethyl starch, ficoll, or a combination thereof. In some instances, mixtures of polymers are used in hybridization buffers described herein. In some instances, hybridization buffers comprise Denhardt's solution. Polymers described herein may be present at any concentration suitable for reducing off-target binding. Such concentrations are often represented as a percent by weight, percent by volume, or percent weight per volume. For example, a polymer is present at about 0.0001%, 0.0002%, 0.0005%, 0.0008%, 0.001%, 0.002%, 0.005%, 0.008%, 0.01%, 0.02%, 0.05%, 0.08%, 0.1%, 0.2%, 0.5%, 0.8%, 1%, 1.2%, 1.5%, 1.8%, 2%, 5%, 10%, 20%, or about 30%. In some instances, a polymer is present at no more than 0.0001%, 0.0002%, 0.0005%, 0.0008%, 0.001%, 0.002%, 0.005%, 0.008%, 0.01%, 0.02%, 0.05%, 0.08%, 0.1%, 0.2%, 0.5%, 0.8%, 1%, 1.2%, 1.5%, 1.8%, 2%, 5%, 10%, 20%, or no more than 30%. In some instances, a polymer is present in at least 0.0001%, 0.0002%, 0.0005%, 0.0008%, 0.001%, 0.002%, 0.005%, 0.008%, 0.01%, 0.02%, 0.05%, 0.08%, 0.1%, 0.2%, 0.5%, 0.8%, 1%, 1.2%, 1.5%, 1.8%, 2%, 5%, 10%, 20%, or at least 30%. In some instances, a polymer is present at 0.0001%-10%, 0.0002%-5%, 0.0005%-1.5%, 0.0008%-1%, 0.001%-0.2%, 0.002%-0.08%, 0.005%-0.02%, or 0.008%-0.05%. In some instances, a polymer is present at 0.005%-0.1%. In some instances, a polymer is present at 0.05%-0.1%. In some instances, a polymer is present at 0.005%-0.6%. In some instances, a polymer is present at 1%-30%, 5%-25%, 10%-30%, 15%-30%, or 1%-15%. Liquid polymers may be present as a percentage of the total reaction volume. In some instances, a polymer is about 10%, 20%, 30%, 40%, 50%, 60%, 75%, or about 90% of the total volume. In some instances, a polymer is at least 10%, 20%, 30%, 40%, 50%, 60%, 75%, or at least 90% of the total volume. In some

instances, a polymer is no more than 10%, 20%, 30%, 40%, 50%, 60%, 75%, or no more than 90% of the total volume. In some instances, a polymer is 5%-75%, 5%-65%, 5%-55%, 10%-50%, 15%-40%, 20%-50%, 20%-30%, 25%-35%, 5%-35%, 10%-35%, or 20%-40% of the total volume. In some instances, a polymer is 25%-45% of the total volume. In some instances, hybridization buffers described herein are used in conjunction with universal blockers and liquid polymer additives.

[0142] Hybridization buffers described herein may comprise salts such as cations or anions. For example, hybridization buffer comprises a monovalent or divalent cation. In some instances, a hybridization buffer comprises a monovalent or divalent anion. Cations in some instances comprise sodium, potassium, magnesium, lithium, tris, or other salt. Anions in some instances comprise sulfate, bisulfite, hydrogensulfate, nitrate, chloride, bromide, citrate, ethylenediaminetetraacetate, dihydrogenphosphate, hydrogenphosphate, or phosphate. In some instances, hybridization buffers comprise salts comprising any combination of anions and cations (e.g. sodium chloride, sodium sulfate, potassium phosphate, or other salt). In some instance, a hybridization buffer comprises an ionic liquid. Salts described herein may be present at any concentration suitable for reducing off-target binding. Such concentrations are often represented as a percent by weight, percent by volume, or percent weight per volume. For example, a salt is present at about 0.0001%, 0.0002%, 0.0005%, 0.0008%, 0.001%, 0.002%, 0.005%, 0.008%, 0.01%, 0.02%, 0.05%, 0.08%, 0.1%, 0.2%, 0.5%, 0.8%, 1%, 1.2%, 1.5%, 1.8%, 2%, 5%, 10%, 20%, or about 30%. In some instances, a salt is present at no more than 0.0001%, 0.0002%, 0.0005%, 0.0008%, 0.001%, 0.002%, 0.005%, 0.008%, 0.01%, 0.02%, 0.05%, 0.08%, 0.1%, 0.2%, 0.5%, 0.8%, 1%, 1.2%, 1.5%, 1.8%, 2%, 5%, 10%, 20%, or no more than 30%. In some instances, a salt is present in at least 0.0001%, 0.0002%, 0.0005%, 0.0008%, 0.001%, 0.002%, 0.005%, 0.008%, 0.01%, 0.02%, 0.05%, 0.08%, 0.1%, 0.2%, 0.5%, 0.8%, 1%, 1.2%, 1.5%, 1.8%, 2%, 5%, 10%, 20%, or at least 30%. In some instances, a salt is present at 0.0001%-10%, 0.0002%-5%, 0.0005%-1.5%, 0.0008%-1%, 0.001%-0.2%, 0.002%-0.08%, 0.005%-0.02%, or 0.008%-0.05%. In some instances, a salt is present at 0.005%-0.1%. In some instances, a salt is present at 0.05%-0.1%. In some instances, a salt is present at 0.005%-0.6%. In some instances, a salt is present at 1%-30%, 5%-25%, 10%-30%, 15%-30%, or 1%-15%. Liquid polymers may be present as a percentage of the total reaction volume. In some instances, a salt is about 10%, 20%, 30%, 40%, 50%, 60%, 75%, or about 90% of the total volume. In some instances, a salt is at least 10%, 20%, 30%, 40%, 50%, 60%, 75%, or at least 90% of the total volume. In some instances, a salt is no more than 10%, 20%, 30%, 40%, 50%, 60%, 75%, or no more than 90% of the total volume. In some instances, a salt is 5%-75%, 5%-65%, 5%-55%, 10%-

50%, 15%-40%, 20%-50%, 20%-30%, 25%-35%, 5%-35%, 10%-35%, or 20%-40% of the total volume. In some instances, a salt is 25%-45% of the total volume.

[0143] Hybridization buffers described herein may comprise surfactants (or emulsifiers). For example, a hybridization buffer comprises SDS (sodium dodecyl sulfate), CTAB, cetylpyridinium, benzalkonium tergitol, fatty acid sulfonates (e.g., sodium lauryl sulfate), ethyloxylated propylene glycol, lignin sulfonates, benzene sulfonate, lecithin, phospholipids, dialkyl sulfosuccinates (e.g., dioctyl sodium sulfosuccinate), glycerol diester, polyethoxylated octyl phenol, abietic acid, sorbitan monoester, perfluoro alkanols, sulfonated polystyrene, betaines, dimethyl polysiloxanes, or other surfactant. In some instances, a hybridization buffer comprises a sulfate, phosphate, or tetralkyl ammonium group. Surfactants described herein may be present at any concentration suitable for reducing off-target binding. Such concentrations are often represented as a percent by weight, percent by volume, or percent weight per volume. For example, a surfactant is present at about 0.0001%, 0.0002%, 0.0005%, 0.0008%, 0.001%, 0.002%, 0.005%, 0.008%, 0.01%, 0.02%, 0.05%, 0.08%, 0.1%, 0.2%, 0.5%, 0.8%, 1%, 1.2%, 1.5%, 1.8%, 2%, 5%, 10%, 20%, or about 30%. In some instances, a surfactant is present at no more than 0.0001%, 0.0002%, 0.0005%, 0.0008%, 0.001%, 0.002%, 0.005%, 0.008%, 0.01%, 0.02%, 0.05%, 0.08%, 0.1%, 0.2%, 0.5%, 0.8%, 1%, 1.2%, 1.5%, 1.8%, 2%, 5%, 10%, 20%, or no more than 30%. In some instances, a surfactant is present in at least 0.0001%, 0.0002%, 0.0005%, 0.0008%, 0.001%, 0.002%, 0.005%, 0.008%, 0.01%, 0.02%, 0.05%, 0.08%, 0.1%, 0.2%, 0.5%, 0.8%, 1%, 1.2%, 1.5%, 1.8%, 2%, 5%, 10%, 20%, or at least 30%. In some instances, a surfactant is present at 0.0001%-10%, 0.0002%-5%, 0.0005%-1.5%, 0.0008%-1%, 0.001%-0.2%, 0.002%-0.08%, 0.005%-0.02%, or 0.008%-0.05%. In some instances, a surfactant is present at 0.005%-0.1%. In some instances, a surfactant is present at 0.05%-0.1%. In some instances, a surfactant is present at 0.005%-0.6%. In some instances, a surfactant is present at 1%-30%, 5%-25%, 10%-30%, 15%-30%, or 1%-15%. Liquid polymers may be present as a percentage of the total reaction volume. In some instances, a surfactant is about 10%, 20%, 30%, 40%, 50%, 60%, 75%, or about 90% of the total volume. In some instances, a surfactant is at least 10%, 20%, 30%, 40%, 50%, 60%, 75%, or at least 90% of the total volume. In some instances, a surfactant is no more than 10%, 20%, 30%, 40%, 50%, 60%, 75%, or no more than 90% of the total volume. In some instances, a surfactant is 5%-75%, 5%-65%, 5%-55%, 10%-50%, 15%-40%, 20%-50%, 20%-30%, 25%-35%, 5%-35%, 10%-35%, or 20%-40% of the total volume. In some instances, a surfactant is 25%-45% of the total volume.

[0144] Buffers used in the methods described herein may comprise any combination of components. In some instances, a buffer described herein is a hybridization buffer. In some instances, a hybridization buffer described herein is a fast hybridization buffer. Such fast

hybridization buffers allow for lower hybridization times such as less than 8 hours, 6 hours, 4 hours, 2 hours, 1 hour, 45 minutes, 30 minutes, or less than 15 minutes. Hybridization buffers described herein in some instances comprise a buffer described in **Tables 2A-2G**. In some instances, the buffers described in Tables 1A-1I may be used as fast hybridization buffers. In some instances, the buffers described in Tables 1B, 1C, and 1D may be used as fast hybridization buffers. In some instances, a fast hybridization buffer as described herein is described in Table 1B. In some instances, a fast hybridization buffer as described herein is described in Table 1C. In some instances, a fast hybridization buffer as described herein is described in Table 1D.

[0145] Table 2A. Buffers A

Buffer Component	Volume (mL)	Buffer Component	Volume (mL)
Water	5-300	Water	100-300
DMF	0-3	DMSO	0-3
NaCl (5M)	0.01-0.5	NaCl (5M)	0.01-0.5
20% SDS	0.05-0.5	20% SDS	0.05-0.5
Tergitol (1% by weight)	0.2-3	EDTA (1M)	0-2
Denhardt's Solution (50X)	1-10	Denhardt's Solution (50X)	1-10
NaH ₂ PO ₄ (5M)	0.01-1.5	NaH ₂ PO ₄ (5M)	0.01-1.5

[0146] Table 2B. Buffers B

Buffer Component	Volume (mL)	Buffer Component	Volume (mL)
Water	5-30	Water	5-30
DMSO	0.5-3	DMSO	0.5-3
NaCl (5M)	0.01-0.5	NaCl (5M)	0.01-0.5
20% SDS	0.05-0.5	20% CTAB	0.05-0.5
EDTA (1M)	0.05-2	EDTA (1M)	0.05-2
Denhardt's Solution (50X)	1-10	Denhardt's Solution (50X)	1-10
NaH ₂ PO ₄ (5M)	0.01-1.5	NaH ₂ PO ₄ (5M)	0.01-1.5

[0147] Table 2C. Buffers C

Buffer Component	Volume (mL)	Buffer Component	Volume (mL)
Water	5-30	Water	5-30
DMSO	0.5-3	DMSO	0.5-3

NaCl (1M)	0.01-0.5	NaCl (5M)	0.01-0.5
20% SDS	0.05-0.5	20% SDS	0.05-0.5
TrisHCl (1M)	0.01-2.5	Dextran Sulfate (50%)	0.05-2
Denhardt's Solution (50X)	1-10	Denhardt's Solution (50X)	1-10
NaH ₂ PO ₄ (5M)	0.01-1.5	NaH ₂ PO ₄ (5M)	0.01-1.5
EDTA (0.5 M)	0.05-1.5	EDTA (0.5 M)	0.05-1.5

[0148] Table 2D. Buffers D

Buffer Component	Volume (mL)	Buffer Component	Volume (mL)
Water	5-30	Water	5-30
Methanol	0.1-3	DMSO	0.5-3
NaCl (1M)	0.01-0.5	NaCl (5M)	0.01-0.5
20% Dextran Sulfate	0.05-0.5	20% SDS	0.05-0.5
TrisHCl (1M)	0.01-2.5	hydroxyethyl starch (20%)	0.05-2
Denhardt's Solution (50X)	1-10	Denhardt's Solution (50X)	1-10
NaH ₂ PO ₄ (1M)	0.01-1.5	NaH ₂ PO ₄ (5M)	0.01-1.5
EDTA (0.5 M)	0.05-1.5	EDTA (0.5 M)	0.05-1.5

[0149] Table 2E. Buffers E

Buffer Component	Volume (mL)	Buffer Component	Volume (mL)
Water	5-300	Water	5-300
DMF	0.1-30	DMSO	0.5-30
NaCl (1M)	0.01-0.5	NaCl (5M)	0.01-1.0
hydroxyethyl starch (20%)	0.01-2.5	hydroxyethyl starch (20%)	0.01-2.5
Denhardt's Solution (50X)	1-10	Denhardt's Solution (50X)	0.05-2
NaH ₂ PO ₄ (1M)	0.01-1.5	NaH ₂ PO ₄ (5M)	1-10

[0150] Table 2F. Buffers F

Buffer Component	Volume (mL)	Buffer Component	Volume (mL)
Water	50-300	Water	50-300
DMF	15-300	DMSO	15-300
NaCl (5M)	2-100	NaCl (5M)	2-100

Denhardt's Solution (50X)	1-10	saline-sodium citrate 20X	1-50
Tergitol (1% by weight)	0.2-2.0	20% SDS	0-2

[0151] Table 2G. Buffers G

Buffer Component	Volume (mL)	Buffer Component	Volume (mL)
Water	5-30	Water	5-30
Ethanol	0-3	Methanol	0-3
NaCl (1M)	0.01-0.5	NaCl (5M)	0.01-0.5
NaH ₂ PO ₄ (5M)	0.01-1.5	NaH ₂ PO ₄ (5M)	0-2
EDTA (0.5 M)	0-1.5	EDTA (0.5 M)	1-10

[0152] Table 2H. Buffers H

Buffer Component	Volume (mL)	Buffer Component	Volume (mL)
Water	50-300	Water	10-300
EDTA (0.5 M)	0-1.5	NaCl (5M)	0.01-0.5
NaCl (5M)	5-70	10% Triton X-100	0.05-0.5
Tergitol (1% by weight)	0.2-2.0	EDTA (1M)	0-2
TrisHCl (1M)	0.01-2.5	TrisHCl (1M)	0.1-5

[0153] Table 2I. Buffers I

Buffer Component	Volume (mL)	Buffer Component	Volume (mL)
Water	5-200	Water	10-200
EDTA (0.5 M)	0-1.5	NaCl (5M)	0.01-0.5
NaCl (5M)	5-100	Sodium Lauryl sulfate (10%)	0.05-0.5
CTAB (0.2M)	0.05-0.5	EDTA (1M)	0-2

[0154] Buffers such as binding buffers and wash buffers are described herein. Binding buffers in some instances are used to prepare mixtures of sample polynucleotides and probes after hybridization. In some instances, binding buffers facilitate capture of sample polynucleotides on a column or other solid support. In some instances, the buffers described in **Tables 2A-2I** may be used as binding buffers. Binding buffers in some instances comprise a buffer described in **Tables 2A, 2H, and 2I**. In some instances, a binding buffer as described herein is described in **Table 2A**. In some instances, a binding buffer as described herein is described in **Table 2H**. In some

instances, a binding buffer as described herein is described in **Table 2I**. In some instances, the buffers described herein may be used as wash buffers. Wash buffers in some instances are used to remove non-binding polynucleotides from a column or solid support. In some instances, the buffers described in **Tables 2A-2I** may be used as wash buffers. In some instances, a wash buffer comprises a buffer as described in **Tables 2E, 2F, and 2G**. In some instances, a wash buffer as described herein is described in **Table 2E**. In some instances, a wash buffer as described herein is described in **Table 2F**. In some instances, a wash buffer as described herein is described in **Table 2G**. Wash buffers used with the compositions and methods described herein are in some instances described as a first wash buffer (wash buffer 1), second wash buffer (wash buffer 2), etc.

[0155] De Novo Synthesis of Small Polynucleotide Populations for Amplification Reactions

[0156] Described herein are methods of synthesis of polynucleotides from a surface, e.g., a plate. In some instances, the polynucleotides are synthesized on a cluster of loci for polynucleotide extension, released and then subsequently subjected to an amplification reaction, e.g., PCR. An exemplary workflow of synthesis of polynucleotides from a cluster is depicted in **FIG. 8**. A silicon plate **801** includes multiple clusters **803**. Within each cluster are multiple loci **821**. Polynucleotides are synthesized **807** de novo on a plate **801** from the cluster **803**. Polynucleotides are cleaved **811** and removed **813** from the plate to form a population of released polynucleotides **815**. The population of released polynucleotides **815** is then amplified **817** to form a library of amplified polynucleotides **219**.

[0157] Provided herein are methods where amplification of polynucleotides synthesized on a cluster provide for enhanced control over polynucleotide representation compared to amplification of polynucleotides across an entire surface of a structure without such a clustered arrangement. In some instances, amplification of polynucleotides synthesized from a surface having a clustered arrangement of loci for polynucleotides extension provides for overcoming the negative effects on representation due to repeated synthesis of large polynucleotide populations. Exemplary negative effects on representation due to repeated synthesis of large polynucleotide populations include, without limitation, amplification bias resulting from high/low GC content, repeating sequences, trailing adenines, secondary structure, affinity for target sequence binding, or modified nucleotides in the polynucleotide sequence.

[0158] Cluster amplification as opposed to amplification of polynucleotides across an entire plate without a clustered arrangement can result in a tighter distribution around the mean. For example, if 100,000 reads are randomly sampled, an average of 8 reads per sequence would yield a library with a distribution of about 1.5X from the mean. In some cases, single cluster amplification

results in at most about 1.5X, 1.6X, 1.7X, 1.8X, 1.9X, or 2.0X from the mean. In some cases, single cluster amplification results in at least about 1.0X, 1.2X, 1.3X, 1.5X, 1.6X, 1.7X, 1.8X, 1.9X, or 2.0X from the mean.

[0159] Cluster amplification methods described herein when compared to amplification across a plate can result in a polynucleotide library that requires less sequencing for equivalent sequence representation. In some instances at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 95% less sequencing is required. In some instances up to 10%, up to 20%, up to 30%, up to 40%, up to 50%, up to 60%, up to 70%, up to 80%, up to 90%, or up to 95% less sequencing is required. Sometimes 30% less sequencing is required following cluster amplification compared to amplification across a plate. Sequencing of polynucleotides in some instances is verified by high-throughput sequencing such as by next generation sequencing. Sequencing of the sequencing library can be performed with any appropriate sequencing technology, including but not limited to single-molecule real-time (SMRT) sequencing, Polony sequencing, sequencing by ligation, reversible terminator sequencing, proton detection sequencing, ion semiconductor sequencing, nanopore sequencing, electronic sequencing, pyrosequencing, Maxam-Gilbert sequencing, chain termination (e.g., Sanger) sequencing, +S sequencing, or sequencing by synthesis. The number of times a single nucleotide or polynucleotide is identified or “read” is defined as the sequencing depth or read depth. In some cases, the read depth is referred to as a fold coverage, for example, 55 fold (or 55X) coverage, optionally describing a percentage of bases.

[0160] Libraries described herein may have a reduced number of dropouts after amplification. In some instances, amplification from a clustered arrangement compared to amplification across a plate results in less dropouts, or sequences which are not detected after sequencing of amplification product. Dropouts can be of AT and/or GC. In some instances, a number of dropouts is at most about 1%, 2%, 3%, 4%, or 5% of a polynucleotide population. In some cases, the number of dropouts is zero.

[0161] A cluster as described herein comprises a collection of discrete, non-overlapping loci for polynucleotide synthesis. A cluster can comprise about 50-1000, 75-900, 100-800, 125-700, 150-600, 200-500, or 300-400 loci. In some instances, each cluster includes 121 loci. In some instances, each cluster includes about 50-500, 50-200, 100-150 loci. In some instances, each cluster includes at least about 50, 100, 150, 200, 500, 1000 or more loci. In some instances, a single plate includes 100, 500, 10000, 20000, 30000, 50000, 100000, 500000, 700000, 1000000 or more loci. A locus can be a spot, well, microwell, channel, or post. In some instances, each cluster has at least 1X, 2X,

3X, 4X, 5X, 6X, 7X, 8X, 9X, 10X, or more redundancy of separate features supporting extension of polynucleotides having identical sequence.

[0162] Generation of Polynucleotide Libraries with Controlled Stoichiometry of Sequence Content

[0163] Provided herein are polynucleotide libraries synthesized with a specified distribution of desired polynucleotide sequences. Adjusting polynucleotide libraries for enrichment of specific desired sequences may provide for improved downstream application outcomes. For example, one or more specific sequences can be selected based on their evaluation in a downstream application. In some instances, the evaluation is binding affinity to target sequences for amplification, enrichment, or detection, stability, melting temperature, biological activity, ability to assemble into larger fragments, or other property of polynucleotides. In some instances, the evaluation is empirical or predicted from prior experiments and/or computer algorithms. An exemplary application includes increasing sequences in a probe library which correspond to areas of a genomic target having less than average read depth. The selected sequences for adjustment in a polynucleotide library can be at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or more than 95% of the sequences. In some instances, selected sequences for adjustment in a polynucleotide library are at most 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or at most 100% of the sequences. In some cases, selected sequences are in a range of about 5-95%, 10-90%, 30-80%, 40-75%, or 50-70% of the sequences. Polynucleotide libraries can be adjusted for the frequency of each selected sequence for adjustment. In some instances, polynucleotide libraries favor a higher number of selected sequences. For example, a library is designed where increased polynucleotide frequency of selected sequences is in a range of about 40% to about 90%. In some instances, polynucleotide libraries contain a low number of selected sequences. For example, a library is designed where increased polynucleotide frequency of the selected sequences is in a range of about 10% to about 60%. A library can be designed to favor a higher and lower frequency of selected sequences. In some instances, a library favors uniform sequence representation. For example, polynucleotide frequency is uniform with regard to selected sequence frequency, in a range of about 10% to about 90%. In some instances, a library comprises polynucleotides with a selected sequence frequency of about 10% to about 95% of the sequences.

[0164] Generation of polynucleotide libraries with a specified selected sequence for adjustment frequency may occur by combining at least 2 polynucleotide libraries with different selected sequence for adjustment frequency content. In some instances, at least 2, 3, 4, 5, 6, 7, 10, or more than 10 polynucleotide libraries are combined to generate a population of polynucleotides with a specified selected sequence frequency. In some cases, no more than 2, 3, 4, 5, 6, 7, or 10

polynucleotide libraries are combined to generate a population of non-identical polynucleotides with a specified selected sequence frequency.

[0165] As described herein, selected sequence for adjustment frequency is adjusted by synthesizing fewer or more polynucleotides per cluster. For example, at least 25, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, or more than 1000 non-identical polynucleotides are synthesized on a single cluster. In some cases, no more than about 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 non-identical polynucleotides are synthesized on a single cluster. In some instances, 50 to 500 non-identical polynucleotides are synthesized on a single cluster. In some instances, 100 to 200 non-identical polynucleotides are synthesized on a single cluster. In some instances, about 100, about 120, about 125, about 130, about 150, about 175, or about 200 non-identical polynucleotides are synthesized on a single cluster.

[0166] In some cases, selected sequence for adjustment frequency is adjusted by synthesizing non-identical polynucleotides of varying length. For example, the length of each of the non-identical polynucleotides synthesized may be at least or about at least 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 150, 200, 300, 400, 500, 2000 nucleotides, or more. The length of the non-identical polynucleotides synthesized may be at most or about at most 2000, 500, 400, 300, 200, 150, 100, 50, 45, 35, 30, 25, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10 nucleotides, or less. The length of each of the non-identical polynucleotides synthesized may fall from 10-2000, 10-500, 9-400, 11-300, 12-200, 13-150, 14-100, 15-50, 16-45, 17-40, 18-35, and 19-25.

[0167] Polynucleotide Probe Structures

[0168] Libraries of polynucleotide probes can be used to enrich particular target sequences in a larger population of sample polynucleotides. In some instances, polynucleotide probes each comprise a target binding sequence complementary to one or more target sequences, one or more non-target binding sequences, and one or more primer binding sites, such as universal primer binding sites. Target binding sequences that are complementary or at least partially complementary in some instances bind (hybridize) to target sequences. Primer binding sites, such as universal primer binding sites facilitate simultaneous amplification of all members of the probe library, or a subpopulation of members. In some instances, the probes or adapters further comprise a barcode or index sequence. Barcodes are nucleic acid sequences that allow some feature of a polynucleotide with which the barcode is associated to be identified. After sequencing, the barcode region provides an indicator for identifying a characteristic associated with the coding region or sample source. Barcodes can be designed at suitable lengths to allow sufficient degree of identification, e.g., at least about 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54,

55, or more bases in length. Multiple barcodes, such as about 2, 3, 4, 5, 6, 7, 8, 9, 10, or more barcodes, may be used on the same molecule, optionally separated by non-barcode sequences. In some embodiments, each barcode in a plurality of barcodes differ from every other barcode in the plurality at least three base positions, such as at least about 3, 4, 5, 6, 7, 8, 9, 10, or more positions. Use of barcodes allows for the pooling and simultaneous processing of multiple libraries for downstream applications, such as sequencing (multiplex). In some instances, at least 4, 8, 16, 32, 48, 64, 128, or more 512 barcoded libraries are used. In some instances, the polynucleotides are ligated to one or more molecular (or affinity) tags such as a small molecule, peptide, antigen, metal, or protein to form a probe for subsequent capture of the target sequences of interest. In some instances, only a portion of the polynucleotides are ligated to a molecular tag. In some instances, two probes that possess complementary target binding sequences which are capable of hybridization form a double stranded probe pair. Polynucleotide probes or adapters may comprise unique molecular identifiers (UMI). UMIs allow for internal measurement of initial sample concentrations or stoichiometry prior to downstream sample processing (e.g., PCR or enrichment steps) which can introduce bias. In some instances, UMIs comprise one or more barcode sequences.

[0169] Probes described here may be complementary to target sequences which are sequences in a genome. Probes described here may be complementary to target sequences which are exome sequences in a genome. Probes described here may be complementary to target sequences which are intron sequences in a genome. In some instances, probes comprise a target binding sequence complementary to a target sequence, and at least one non-target binding sequence that is not complementary to the target. In some instances, the target binding sequence of the probe is about 120 nucleotides in length, or at least 10, 15, 20, 25, 50, 75, 100, 110, 120, 125, 140, 150, 160, 175, 200, 300, 400, 500, or more than 500 nucleotides in length. The target binding sequence is in some instances no more than 10, 15, 20, 25, 50, 75, 100, 125, 150, 175, 200, or no more than 500 nucleotides in length. The target binding sequence of the probe is in some instances about 120 nucleotides in length, or about 10, 15, 20, 25, 40, 50, 60, 70, 80, 85, 87, 90, 95, 97, 100, 105, 110, 115, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 135, 140, 145, 150, 155, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 175, 180, 190, 200, 210, 220, 230, 240, 250, 300, 400, or about 500 nucleotides in length. The target binding sequence is in some instances about 20 to about 400 nucleotides in length, or about 30 to about 175, about 40 to about 160, about 50 to about 150, about 75 to about 130, about 90 to about 120, or about 100 to about 140 nucleotides in length. The non-target binding sequence(s) of the probe is in some instances at least about 20 nucleotides in length, or at least about 1, 5, 10, 15, 17, 20, 23, 25, 50, 75, 100, 110, 120, 125, 140, 150, 160, 175, or more than about 175 nucleotides in length. The non-target binding

sequence often is no more than about 5, 10, 15, 20, 25, 50, 75, 100, 125, 150, 175, or no more than about 200 nucleotides in length. The non-target binding sequence of the probe often is about 20 nucleotides in length, or about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 25, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, or about 200 nucleotides in length. The non-target binding sequence in some instances is about 1 to about 250 nucleotides in length, or about 20 to about 200, about 10 to about 100, about 10 to about 50, about 30 to about 100, about 5 to about 40, or about 15 to about 35 nucleotides in length. The non-target binding sequence often comprises sequences that are not complementary to the target sequence, and/or comprise sequences that are not used to bind primers. In some instances, the non-target binding sequence comprises a repeat of a single nucleotide, for example polyadenine or polythymidine. A probe often comprises none or at least one non-target binding sequence. In some instances, a probe comprises one or two non-target binding sequences. The non-target binding sequence may be adjacent to one or more target binding sequences in a probe. For example, a non-target binding sequence is located on the 5' or 3' end of the probe. In some instances, the non-target binding sequence is attached to a molecular tag or spacer.

[0170] As described herein, non-target binding sequence(s) may be a primer binding site. The primer binding sites often are each at least about 20 nucleotides in length, or at least about 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, or at least about 40 nucleotides in length. Each primer binding site in some instances is no more than about 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, or no more than about 40 nucleotides in length. Each primer binding site in some instances is about 10 to about 50 nucleotides in length, or about 15 to about 40, about 20 to about 30, about 10 to about 40, about 10 to about 30, about 30 to about 50, or about 20 to about 60 nucleotides in length. In some instances the polynucleotide probes comprise at least two primer binding sites. In some instances, primer binding sites may be universal primer binding sites, wherein all probes comprise identical primer binding sequences at these sites. In some instances, a pair of polynucleotide probes targeting a particular sequence and its reverse complement (e.g., a region of genomic DNA) are represented by **900** in **FIG. 9A**, comprising a first target binding sequence **901**, a second target binding sequence **902**, a first non-target binding sequence **903**, and a second non-target binding sequence **904**. For example, a pair of polynucleotide probes complementary to a particular sequence (e.g., a region of genomic DNA).

[0171] In some instances, the first target binding sequence **901** is the reverse complement of the second target binding sequence **902**. In some instances, both target binding sequences are chemically synthesized prior to amplification. In an alternative arrangement, a pair of polynucleotide probes targeting a particular sequence and its reverse complement (e.g., a region of

genomic DNA) are represented by **905** in **FIG. 9B**, comprising a first target binding sequence **901**, a second target binding sequence **902**, a first non-target binding sequence **903**, a second non-target binding sequence **904**, a third non-target binding sequence **906**, and a fourth non-target binding sequence **907**. In some instances, the first target binding sequence **901** is the reverse complement of the second target binding sequence **902**. In some instances, one or more non-target binding sequences comprise polyadenine or polythymidine.

[0172] Probes described herein may comprise molecular tags. In some instances, both probes in the pair are labeled with at least one molecular tag. In some instances, PCR is used to introduce molecular tags (via primers comprising the molecular tag) onto the probes during amplification. In some instances, the molecular tag comprises one or more biotin, folate, a polyhistidine, a FLAG tag, glutathione, or other molecular tag consistent with the specification. In some instances probes are labeled at the 5' terminus. In some instances, the probes are labeled at the 3' terminus. In some instances, both the 5' and 3' termini are labeled with a molecular tag. In some instances, the 5' terminus of a first probe in a pair is labeled with at least one molecular tag, and the 3' terminus of a second probe in the pair is labeled with at least one molecular tag. In some instances, a spacer is present between one or more molecular tags and the nucleic acids of the probe. In some instances, the spacer may comprise an alkyl, polyol, or polyamino chain, a peptide, or a polynucleotide. The solid support used to capture probe-target nucleic acid complexes in some instances, is a bead or a surface. The solid support in some instances comprises glass, plastic, or other material capable of comprising a capture moiety that will bind the molecular tag. In some instances, a bead is a magnetic bead. For example, probes labeled with biotin are captured with a magnetic bead comprising streptavidin. The probes are contacted with a library of nucleic acids to allow binding of the probes to target sequences. In some instances, blocking polynucleic acids are added to prevent binding of the probes to one or more adapter sequences attached to the target nucleic acids. In some instances, blocking polynucleic acids comprise one or more nucleic acid analogues. In some instances, blocking polynucleic acids have a uracil substituted for thymine at one or more positions.

[0173] Probes described herein may comprise complementary target binding sequences which bind to one or more target nucleic acid sequences. In some instances, the target sequences are any DNA or RNA nucleic acid sequence. In some instances, target sequences may be longer than the probe insert. In some instance, target sequences may be shorter than the probe insert. In some instance, target sequences may be the same length as the probe insert. For example, the length of the target sequence may be at least or about at least 2, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 150, 200, 300, 400, 500, 1000, 2000, 5,000, 12,000, 20,000 nucleotides, or more. The length of the

target sequence may be at most or about at most 20,000, 12,000, 5,000, 2,000, 1,000, 500, 400, 300, 200, 150, 100, 50, 45, 35, 30, 25, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 2 nucleotides, or less. The length of the target sequence may fall from 2-20,000, 3-12,000, 5-5,000, 10-2,000, 10-1,000, 10-500, 9-400, 11-300, 12-200, 13-150, 14-100, 15-50, 16-45, 17-40, 18-35, and 19-25. The probe sequences may target sequences associated with specific genes, diseases, regulatory pathways, or other biological functions consistent with the specification.

[0174] A probe described herein may bind to a target sequences in any number of suitable arrangements. In some instances, a single probe insert **1003** is complementary to one or more target sequences **1002** (**FIGS. 10A-10G**) in a larger polynucleic acid **1000**. An exemplary target sequence is an exon. In some instances, one or more probes target a single target sequence (**FIGS. 10A-10G**). In some instances, a single probe may target more than one target sequence. In some instances, the target binding sequence of the probe targets both a target sequence **1002** and an adjacent sequence **1001** (**FIG. 10A and 10B**). In some instances, a first probe targets a first region and a second region of a target sequence, and a second probe targets the second region and a third region of the target sequence (**FIG. 10D and FIG. 10E**). In some instances, a plurality of probes targets a single target sequence, wherein the target binding sequences of the plurality of probes contain one or more sequences which overlap with regard to complementarity to a region of the target sequence (**FIG. 10G**). In some instances, probe inserts do not overlap with regard to complementarity to a region of the target sequence. In some instances, at least at least 2, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 150, 200, 300, 400, 500, 1000, 2000, 5,000, 12,000, 20,000, or more than 20,000 probes target a single target sequence. In some instances no more than 4 probes directed to a single target sequence overlap, or no more than 3, 2, 1, or no probes targeting a single target sequence overlap. In some instances, one or more probes do not target all bases in a target sequence, leaving one or more gaps (**FIG. 10C and FIG. 10F**). In some instances, the gaps are near the middle of the target sequence **1005** (**FIG. 10F**). In some instances, the gaps **1004** are at the 5' or 3' ends of the target sequence (**FIG. 10C**). In some instances, the gaps are 6 nucleotides in length. In some instances, the gaps are no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, or no more than 50 nucleotides in length. In some instances, the gaps are at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, or at least 50 nucleotides in length. In some instances, the gaps length falls within 1-50, 1-40, 1-30, 1-20, 1-10, 2-30, 2-20, 2-10, 3-50, 3-25, 3-10, or 3-8 nucleotides in length. In some instances, a set of probes targeting a sequence do not comprise overlapping regions amongst probes in the set when hybridized to complementary sequence. In some instances, a set of probes targeting a sequence do not have any gaps amongst probes in the set when hybridized to complementary sequence. Probes may be designed to maximize uniform binding to target sequences. In some

instances, probes are designed to minimize target binding sequences of high or low GC content, secondary structure, repetitive/palindromic sequences, or other sequence feature that may interfere with probe binding to a target. In some instances, a single probe may target a plurality of target sequences.

[0175] A probe library described herein may comprise at least 10, 20, 50, 100, 200, 500, 1,000, 2,000, 5,000, 10,000, 20,000, 50,000, 100,000, 200,000, 500,000, 1,000,000 or more than 1,000,000 probes. A probe library may have no more than 10, 20, 50, 100, 200, 500, 1,000, 2,000, 5,000, 10,000, 20,000, 50,000, 100,000, 200,000, 500,000, or no more than 1,000,000 probes. A probe library may comprise 10 to 500, 20 to 1000, 50 to 2000, 100 to 5000, 500 to 10,000, 1,000 to 5,000, 10,000 to 50,000, 100,000 to 500,000, or to 50,000 to 1,000,000 probes. A probe library may comprise about 370,000; 400,000; 500,000 or more different probes.

[0176] Next Generation Sequencing Applications

[0177] Downstream applications of polynucleotide libraries may include next generation sequencing. For example, enrichment of target sequences with a controlled stoichiometry polynucleotide probe library results in more efficient sequencing. The performance of a polynucleotide library for capturing or hybridizing to targets may be defined by a number of different metrics describing efficiency, accuracy, and precision. For example, Picard metrics comprise variables such as HS library size (the number of unique molecules in the library that correspond to target regions, calculated from read pairs), mean target coverage (the percentage of bases reaching a specific coverage level), depth of coverage (number of reads including a given nucleotide) fold enrichment (sequence reads mapping uniquely to the target/reads mapping to the total sample, multiplied by the total sample length/target length), percent off-bait bases (percent of bases not corresponding to bases of the probes/baits), percent off-target (percent of bases not corresponding to bases of interest), usable bases on target, AT or GC dropout rate, fold 80 base penalty (fold over-coverage needed to raise 80 percent of non-zero targets to the mean coverage level), percent zero coverage targets, PF reads (the number of reads passing a quality filter), percent selected bases (the sum of on-bait bases and near-bait bases divided by the total aligned bases), percent duplication, or other variable consistent with the specification.

[0178] Read depth (sequencing depth, or sampling) represents the total number of times a sequenced nucleic acid fragment (a “read”) is obtained for a sequence. Theoretical read depth is defined as the expected number of times the same nucleotide is read, assuming reads are perfectly distributed throughout an idealized genome. Read depth is expressed as function of % coverage (or coverage breadth). For example, 10 million reads of a 1 million base genome, perfectly distributed, theoretically results in 10X read depth of 100% of the sequences. In practice, a greater number of

reads (higher theoretical read depth, or oversampling) may be needed to obtain the desired read depth for a percentage of the target sequences. In some instances, the efficiency in sequencing is defined as a ratio of reads for a population of bases in a sample vs. the total reads obtained for the sample. In some instances, a population of bases is selected using probes described herein. In some instances, the ratio of reads for a population of bases in a sample vs. the total reads is at least 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, or at least 0.95. In some instances, the ratio of reads for a population of bases in a sample vs. the total reads is about 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, or about 0.95. In some instances, the ratio of reads for a population of bases in a sample vs. the total reads is 0.1 to 0.9, 0.2 to 0.8, 0.3 to 0.7, 0.2 to 0.8, 0.3 to 0.7, 0.5-0.7, or 0.4-0.7. In some instances, the ratio of reads for a population of bases in a sample vs. the total reads is at least 0.5. In some instances, the ratio of reads for a population of bases in a sample vs. the total reads is at least 0.6. In some instances, the ratio of reads for a population of bases in a sample vs. the total reads is at least 0.7. In some instances, the ratio of reads for a population of bases in a sample vs. the total reads is at least 0.8. Enrichment of target sequences with a controlled stoichiometry probe library increases the efficiency of downstream sequencing, as fewer total reads will be required to obtain an outcome with an acceptable number of reads over a desired % of target sequences. For example, in some instances 55x theoretical read depth of target sequences results in at least 30x coverage of at least 90% of the sequences. In some instances no more than 55x theoretical read depth of target sequences results in at least 30x read depth of at least 80% of the sequences. In some instances no more than 55x theoretical read depth of target sequences results in at least 30x read depth of at least 95% of the sequences. In some instances no more than 55x theoretical read depth of target sequences results in at least 10x read depth of at least 98% of the sequences. In some instances, 55x theoretical read depth of target sequences results in at least 20x read depth of at least 98% of the sequences. In some instances no more than 55x theoretical read depth of target sequences results in at least 5x read depth of at least 98% of the sequences. Increasing the concentration of probes during hybridization with targets can lead to an increase in read depth. In some instances, the concentration of probes is increased by at least 1.5x, 2.0x, 2.5x, 3x, 3.5x, 4x, 5x, or more than 5x. In some instances, increasing the probe concentration results in at least a 1000% increase, or a 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 200%, 300%, 500%, 750%, 1000%, or more than a 1000% increase in read depth. In some instances, increasing the probe concentration by 3x results in a 1000% increase in read depth.

[0179] On-target rate represents the percentage of sequencing reads that correspond with the desired target sequences. In some instances, a controlled stoichiometry polynucleotide probe library

results in an on-target rate of at least 30%, or at least 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, or at least 90%. Increasing the concentration of polynucleotide probes during contact with target nucleic acids leads to an increase in the on-target rate. In some instances, the concentration of probes is increased by at least 1.5x, 2.0x, 2.5x, 3x, 3.5x, 4x, 5x, or more than 5x. In some instances, increasing the probe concentration results in at least a 20% increase, or a 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 200%, 300%, or at least a 500% increase in on-target binding. In some instances, increasing the probe concentration by 3x results in a 20% increase in on-target rate.

[0180] Coverage uniformity is in some cases calculated as the read depth as a function of the target sequence identity. Higher coverage uniformity results in a lower number of sequencing reads needed to obtain the desired read depth. For example, a property of the target sequence may affect the read depth, for example, high or low GC or AT content, repeating sequences, trailing adenines, secondary structure, affinity for target sequence binding (for amplification, enrichment, or detection), stability, melting temperature, biological activity, ability to assemble into larger fragments, sequences containing modified nucleotides or nucleotide analogues, or any other property of polynucleotides. Enrichment of target sequences with controlled stoichiometry polynucleotide probe libraries results in higher coverage uniformity after sequencing. In some instances, 95% of the sequences have a read depth that is within 1x of the mean library read depth, or about 0.05, 0.1, 0.2, 0.5, 0.7, 1, 1.2, 1.5, 1.7 or about within 2x the mean library read depth. In some instances, 80%, 85%, 90%, 95%, 97%, or 99% of the sequences have a read depth that is within 1x of the mean.

[0181] The methods and compositions described herein may be used for specific sample types, including but not limited to DNA, RNA, mRNA, cfDNA, fetal cfDNA, siRNA, rRNA, miRNA, FFPE or other nucleic acid sample. In some instances, mechanical shearing is used to prepare nucleic acid samples for ligation of adapters, capture, enrichment, and sequencing. In some instances, enzymatic cleavage is used to prepare nucleic acid samples for ligation of adapters, capture, enrichment, and sequencing. In some instances, FFPE samples are analyzed, such as FFPE samples from different tissues. Tissues include but are not limited to brain, neck, lymph node, lung, liver, spleen, heart, kidney, skin, uterus, testis, pancreas, intestine, colon, stomach, prostate, or other tissue. In some instances, the tissue is a cancer, such as a solid tumor. In some instances, the solid tumor is a carcinoma. In some instances, use of probes described herein result in increased uniformity and sensitivity of sequencing data obtained using the methods described herein.

[0182] Enrichment of Target Nucleic Acids with a Polynucleotide Probe Library

[0183] A probe library described herein may be used to enrich target polynucleotides present in a population of sample polynucleotides, for a variety of downstream applications. In some instances, a sample is obtained from one or more sources, and the population of sample polynucleotides is isolated. Samples are obtained (by way of non-limiting example) from biological sources such as saliva, blood, tissue, skin, or completely synthetic sources. The plurality of polynucleotides obtained from the sample are fragmented, end-repaired, and adenylated to form a double stranded sample nucleic acid fragment. In some instances, end repair is accomplished by treatment with one or more enzymes, such as T4 DNA polymerase, klenow enzyme, and T4 polynucleotide kinase in an appropriate buffer. A nucleotide overhang to facilitate ligation to adapters is added, in some instances with 3' to 5' exonuclease klenow fragment and dATP.

[0184] Adapters may be ligated to both ends of the sample polynucleotide fragments with a ligase, such as T4 ligase, to produce a library of adapter-tagged polynucleotide strands, and the adapter-tagged polynucleotide library is amplified with primers, such as universal primers. In some instances, the adapters are Y-shaped adapters comprising one or more primer binding sites, one or more grafting regions, and one or more index (or barcode) regions. In some instances, the one or more index region is present on each strand of the adapter. In some instances, grafting regions are complementary to a flowcell surface, and facilitate next generation sequencing of sample libraries. In some instances, Y-shaped adapters comprise partially complementary sequences. In some instances, Y-shaped adapters comprise a single thymidine overhang which hybridizes to the overhanging adenine of the double stranded adapter-tagged polynucleotide strands. Y-shaped adapters may comprise modified nucleic acids, that are resistant to cleavage. For example, a phosphorothioate backbone is used to attach an overhanging thymidine to the 3' end of the adapters. The library of double stranded sample nucleic acid fragments is then denatured in the presence of adapter blockers. Adapter blockers minimize off-target hybridization of probes to the adapter sequences (instead of target sequences) present on the adapter-tagged polynucleotide strands, and/or prevent intermolecular hybridization of adapters (i.e., "daisy chaining").

Denaturation is carried out in some instances at 96°C, or at about 85, 87, 90, 92, 95, 97, 98 or about 99°C. A polynucleotide targeting library (probe library) is denatured in a hybridization solution, in some instances at 96°C, at about 85, 87, 90, 92, 95, 97, 98 or 99°C. The denatured adapter-tagged polynucleotide library and the hybridization solution are incubated for a suitable amount of time and at a suitable temperature to allow the probes to hybridize with their complementary target sequences. In some instances, a suitable hybridization temperature is about 45 to 80°C, or at least 45, 50, 55, 60, 65, 70, 75, 80, 85, or 90°C. In some instances, the hybridization temperature is 70°C. In some instances, a suitable hybridization time is 16 hours, or at least 4, 6, 8, 10, 12, 14, 16,

18, 20, 22, or more than 22 hours, or about 12 to 20 hours. Binding buffer is then added to the hybridized adapter-tagged-polynucleotide probes, and a solid support comprising a capture moiety are used to selectively bind the hybridized adapter-tagged polynucleotide-probes. The solid support is washed with buffer to remove unbound polynucleotides before an elution buffer is added to release the enriched, tagged polynucleotide fragments from the solid support. In some instances, the solid support is washed 2 times, or 1, 2, 3, 4, 5, or 6 times. The enriched library of adapter-tagged polynucleotide fragments is amplified and the enriched library is sequenced.

[0185] A plurality of nucleic acids (i.e. genomic sequence) may obtained from a sample, and fragmented, optionally end-repaired, and adenylated. Adapters are ligated to both ends of the polynucleotide fragments to produce a library of adapter-tagged polynucleotide strands, and the adapter-tagged polynucleotide library is amplified. The adapter-tagged polynucleotide library is then denatured at high temperature, preferably 96°C, in the presence of adapter blockers. A polynucleotide targeting library (probe library) is denatured in a hybridization solution at high temperature, preferably about 90 to 99°C, and combined with the denatured, tagged polynucleotide library in hybridization solution for about 10 to 24 hours at about 45 to 80°C. Binding buffer is then added to the hybridized tagged polynucleotide probes, and a solid support comprising a capture moiety are used to selectively bind the hybridized adapter-tagged polynucleotide-probes. The solid support is washed one or more times with buffer, preferably about 2 and 5 times to remove unbound polynucleotides before an elution buffer is added to release the enriched, adapter-tagged polynucleotide fragments from the solid support. The enriched library of adapter-tagged polynucleotide fragments is amplified and then the library is sequenced. Alternative variables such as incubation times, temperatures, reaction volumes/concentrations, number of washes, or other variables consistent with the specification are also employed in the method.

[0186] A population of polynucleotides may be enriched prior to adapter ligation. In one example, a plurality of polynucleotides is obtained from a sample, fragmented, optionally end-repaired, and denatured at high temperature, preferably 90-99°C. A polynucleotide targeting library (probe library) is denatured in a hybridization solution at high temperature, preferably about 90 to 99°C, and combined with the denatured, tagged polynucleotide library in hybridization solution for about 10 to 24 hours at about 45 to 80°C. Binding buffer is then added to the hybridized tagged polynucleotide probes, and a solid support comprising a capture moiety are used to selectively bind the hybridized adapter-tagged polynucleotide-probes. The solid support is washed one or more times with buffer, preferably about 2 and 5 times to remove unbound polynucleotides before an elution buffer is added to release the enriched, adapter-tagged polynucleotide fragments from the solid support. The enriched polynucleotide fragments are then polyadenylated, adapters are ligated

to both ends of the polynucleotide fragments to produce a library of adapter-tagged polynucleotide strands, and the adapter-tagged polynucleotide library is amplified. The adapter-tagged polynucleotide library is then sequenced.

[0187] A polynucleotide targeting library may also be used to filter undesired sequences from a plurality of polynucleotides, by hybridizing to undesired fragments. For example, a plurality of polynucleotides is obtained from a sample, and fragmented, optionally end-repaired, and adenylated. Adapters are ligated to both ends of the polynucleotide fragments to produce a library of adapter-tagged polynucleotide strands, and the adapter-tagged polynucleotide library is amplified. Alternatively, adenylation and adapter ligation steps are instead performed after enrichment of the sample polynucleotides. The adapter-tagged polynucleotide library is then denatured at high temperature, preferably 90-99°C, in the presence of adapter blockers. A polynucleotide filtering library (probe library) designed to remove undesired, non-target sequences is denatured in a hybridization solution at high temperature, preferably about 90 to 99°C, and combined with the denatured, tagged polynucleotide library in hybridization solution for about 10 to 24 hours at about 45 to 80°C. Binding buffer is then added to the hybridized tagged polynucleotide probes, and a solid support comprising a capture moiety are used to selectively bind the hybridized adapter-tagged polynucleotide-probes. The solid support is washed one or more times with buffer, preferably about 1 and 5 times to elute unbound adapter-tagged polynucleotide fragments. The enriched library of unbound adapter-tagged polynucleotide fragments is amplified and then the amplified library is sequenced.

[0188] A polynucleotide targeting library may be designed to target genes with specific functions. For example, the target genes are mitochondrial genes. In some instances, the target genes are involved in a disease such as cancer or a neurodegenerative disease.

[0189] A polynucleotide targeting library may be designed to target a number of genes. In some instances, the number of genes comprises at least or about 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200, 300, 400, 500, 600, 700, 800, 900, 1000, or more than 1000 genes. In some instances, a size of the target gene is at least or about 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 1.0, 2.0, 4.0, 8.0, 10.0, 12.0, 14.0, 16.0, 18.0, 20.0, 22.0, 24.0, 26.0, 28.0, 30.0, 40.0, 50.0, 60.0, or more than 60.0 megabases (Mb). A number of probes in the polynucleotide targeting library, in some instances, comprises at least or about 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 20000, 30000, 40000, 50000, 60000, 70000, 80000, 90000, 100000, 200000, 300000, 400000, 500000, 600000, 700000, 800000, 900000, 1000000, or more than 1000000 probes.

[0190] Described herein are polynucleotide targeting libraries with improved performance. In some instances, the polynucleotide targeting library comprises sequences that are highly uniform. In some instances, polynucleotide sequences are within at least or about 0.05, 0.1, 0.2, 0.5, 0.7, 1, 1.2, 1.5, 1.7, or 2x the mean. In some instances, 80%, 85%, 90%, 95%, 97%, or 99% of the sequences are within 1x of the mean. In some instances, the polynucleotide targeting libraries result in an on-target rate of at least 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, or at least 90%. In some instances, the polynucleotide targeting libraries result in a duplication rate of at most or about 0.5%, 1.0%, 1.5%, 2.0%, 2.5%, 3.0%, 3.5%, 4.0%, 4.5%, or 5.0%. In some instances, the polynucleotide targeting libraries result in at least 30x coverage of at least 80%, 85%, 90%, 95%, or 99% of the sequences. In some instances, the polynucleotide targeting libraries result in at least 30x coverage of at least 95% of the sequences. In some instances, the polynucleotide targeting libraries result in at least 30x coverage of at least 99% of the sequences.

[0191] A polynucleotide targeting library as described herein may be used for multiplexed reactions. In some instances, the polynucleotide targeting library is used for a 1-, 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9-, 10-, 11-, 12-, 13-, 14-, 15-, 16-, 17-, 18-, 19-, or a 20-plex enrichment reaction. In some instances, the polynucleotide targeting library used for multiplexed reactions result in improved performance. In some instances, the polynucleotide targeting library used for multiplexed reactions comprises sequences that are highly uniform. In some instances, polynucleotide sequences are within at least or about 0.05, 0.1, 0.2, 0.5, 0.7, 1, 1.2, 1.5, 1.7, or 2x the mean. In some instances, 80%, 85%, 90%, 95%, 97%, or 99% of the sequences are within 1x of the mean. In some instances, the polynucleotide targeting library used for multiplexed reactions result in an on-target rate of at least 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, or at least 90%. In some instances, the polynucleotide targeting library used for multiplexed reactions result in a duplication rate of at most or about 0.5%, 1.0%, 1.5%, 2.0%, 2.5%, 3.0%, 3.5%, 4.0%, 4.5%, or 5.0%. In some instances, the polynucleotide targeting library used for multiplexed reactions result in a duplication rate of at most or about 2.0%. In some instances, the polynucleotide targeting library used for multiplexed reactions result in a duplication rate of at most or about 3.0%. In some instances, the improved performance is regardless of panel size. In some instances, the polynucleotide library results in improved performance for panels comprising at least or about 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 1.0, 2.0, 4.0, 8.0, 10.0, 12.0, 14.0, 16.0, 18.0, 20.0, 22.0, 24.0, 26.0, 28.0, 30.0, 40.0, 50.0, 60.0, or more than 60.0 megabases (Mb). In some instances, the improved performance is regardless of sample mass. In some instances, the polynucleotide library results in improved performance for panels comprising at least or about 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 300, 400, 500,

600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, or more than 500 nanograms (ng).

[0192] Polynucleotide targeting libraries as described herein are highly accurate. In some instances, a first polynucleotide targeting library and a second polynucleotide targeting library comprise similar target enrichment. In some instances, a first polynucleotide targeting library and a second polynucleotide targeting library comprise similar probe abundance.

[0193] Polynucleotide targeting libraries as described herein are highly flexible and modular. For example, content of the polynucleotide targeting libraries may be added or enhanced. Adding content can increase a number of targets covered or enhancing content can augment the coverage of specific regions. In some instances, at least or about 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 1.0, 2.0, 4.0, 8.0, 10.0, 12.0, 14.0, 16.0, 18.0, 20.0 megabases (Mb) of content is added or enhanced. In some instances, addition or enhancement of content results in increased coverage. In some instances, coverage is improved to at least 80%, 85%, 90%, 95%, 99%, or more than 99%. In some instances, polynucleotide targeting libraries comprising added or enhanced content have high uniformity, high on-target rate, low duplicate rate, or a combination thereof. In some instances, the polynucleotide targeting library comprising added or enhanced content comprises sequences that are highly uniform. In some instances, polynucleotide sequences are within at least or about 0.05, 0.1, 0.2, 0.5, 0.7, 1, 1.2, 1.5, 1.7, or 2x the mean. In some instances, 80%, 85%, 90%, 95%, 97%, or 99% of the sequences are within 1x of the mean. In some instances, the polynucleotide targeting libraries comprising added or enhanced content result in an on-target rate of at least 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, or at least 90%. In some instances, the polynucleotide targeting libraries comprising added or enhanced content result in a duplication rate of at most or about 0.5%, 1.0%, 1.5%, 2.0%, 2.5%, 3.0%, 3.5%, 4.0%, 4.5%, or 5.0%.

[0194] Polynucleotide targeting libraries as described herein may be designed to improve capture uniformity. For example, polynucleotide targeting libraries are designed to result in less than 1.0%, 1.5%, 2.0%, 2.5%, 3.0%, 3.5%, 4.0%, 4.5%, 5.0%, 6.0%, 6.5%, 7.0%, 7.5%, 8.0%, 8.5%, 9.0%, 9.5%, or 10% AT dropout. In some instances, polynucleotide targeting libraries are designed to result in less than 1.0%, 1.5%, 2.0%, 2.5%, 3.0%, 3.5%, 4.0%, 4.5%, or 5.0% AT dropout. In some instances, polynucleotide targeting libraries are designed to result in less than 1.0%, 1.5%, 2.0%, 2.5%, 3.0%, 3.5%, 4.0%, 4.5%, 5.0%, 6.0%, 6.5%, 7.0%, 7.5%, 8.0%, 8.5%, 9.0%, 9.5%, or 10% GC dropout. In some instances, polynucleotide targeting libraries are designed to result in less than 1.0%, 1.5%, 2.0%, 2.5%, 3.0%, 3.5%, 4.0%, 4.5%, or 5.0% GC dropout. In some instances, the polynucleotide targeting libraries designed for improved capture uniformity result in polynucleotide sequences are within at least or about 0.05, 0.1, 0.2, 0.5, 0.7, 1, 1.2, 1.5,

1.7, or 2x the mean. In some instances, 80%, 85%, 90%, 95%, 97%, or 99% of the sequences are within 1x of the mean. In some instances, the polynucleotide targeting libraries designed for improved capture uniformity result in an on-target rate of at least 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, or at least 90%. In some instances, the polynucleotide targeting libraries designed for improved capture uniformity result in a duplication rate of at most or about 0.5%, 1.0%, 1.5%, 2.0%, 2.5%, 3.0%, 3.5%, 4.0%, 4.5%, or 5.0%. In some instances, the polynucleotide targeting libraries designed for improved capture uniformity result in at least 30x coverage of at least 80%, 85%, 90%, 95%, or 99% of the sequences. In some instances, the polynucleotide targeting libraries designed for improved capture uniformity result in at least 30x coverage of at least 95% of the sequences. In some instances, the polynucleotide targeting libraries designed for improved capture uniformity result in at least 30x coverage of at least 99% of the sequences. In some instances, the polynucleotide targeting libraries designed for improved capture uniformity result in at least 20x coverage of at least 80%, 85%, 90%, 95%, or 99% of the sequences. In some instances, the polynucleotide targeting libraries designed for improved capture uniformity result in at least 20x coverage of at least 95% of the sequences. In some instances, the polynucleotide targeting libraries result in at least 30x coverage of at least 99% of the sequences.

[0195] Polynucleotide targeting libraries may iteratively optimized based on performance of the library. In some instances, polynucleotides are removed from a library. In some instances, removal of a portion of the polynucleotides results in increased on-target rates or a decrease in off-target rates. In some instances, about 0.1%, 0.2%, 0.5%, 1%, 2%, 3%, 4%, or about 5% of the polynucleotides are removed. In some instances, no more than 0.1%, 0.2%, 0.5%, 1%, 2%, 3%, 4%, or no more than 5% of the polynucleotides are removed. In some instances, 0.01%-1%, 0.02-0.4%, 0.3-0.5%, 0.2-1.5%, 0.5-2%, 1-2%, 1-5%, 2-4% or 0.7-3% of the polynucleotides are removed. In some instances, removal of one or more probes from a polynucleotide library used in a method described herein results in enhanced enrichment performance of the library (e.g., on target rate, off target rate, 80-fold base penalty, off-bait rate, % bases >30X coverage, or other sequencing metric).

[0196] Highly Parallel De Novo Nucleic Acid Synthesis

[0197] Described herein is a platform approach utilizing miniaturization, parallelization, and vertical integration of the end-to-end process from polynucleotide synthesis to gene assembly within Nano wells on silicon to create a revolutionary synthesis platform. Devices described herein provide, with the same footprint as a 96-well plate, a silicon synthesis platform is capable of increasing throughput by a factor of 100 to 1,000 compared to traditional synthesis methods, with production of up to approximately 1,000,000 polynucleotides in a single highly-parallelized run. In some instances, a single silicon plate described herein provides for synthesis of about 6,100 non-

identical polynucleotides. In some instances, each of the non-identical polynucleotides is located within a cluster. A cluster may comprise 50 to 500 non-identical polynucleotides.

[0198] Methods described herein provide for synthesis of a library of polynucleotides each encoding for a predetermined variant of at least one predetermined reference nucleic acid sequence. In some cases, the predetermined reference sequence is nucleic acid sequence encoding for a protein, and the variant library comprises sequences encoding for variation of at least a single codon such that a plurality of different variants of a single residue in the subsequent protein encoded by the synthesized nucleic acid are generated by standard translation processes. The synthesized specific alterations in the nucleic acid sequence can be introduced by incorporating nucleotide changes into overlapping or blunt ended polynucleotide primers. Alternatively, a population of polynucleotides may collectively encode for a long nucleic acid (e.g., a gene) and variants thereof. In this arrangement, the population of polynucleotides can be hybridized and subject to standard molecular biology techniques to form the long nucleic acid (e.g., a gene) and variants thereof. When the long nucleic acid (e.g., a gene) and variants thereof are expressed in cells, a variant protein library is generated. Similarly, provided here are methods for synthesis of variant libraries encoding for RNA sequences (e.g., miRNA, shRNA, and mRNA) or DNA sequences (e.g., enhancer, promoter, UTR, and terminator regions). Also provided here are downstream applications for variants selected out of the libraries synthesized using methods described here. Downstream applications include identification of variant nucleic acid or protein sequences with enhanced biologically relevant functions, e.g., biochemical affinity, enzymatic activity, changes in cellular activity, and for the treatment or prevention of a disease state.

[0199] Substrates

[0200] Provided herein are substrates comprising a plurality of clusters, wherein each cluster comprises a plurality of loci that support the attachment and synthesis of polynucleotides. The term “locus” as used herein refers to a discrete region on a structure which provides support for polynucleotides encoding for a single predetermined sequence to extend from the surface. In some instances, a locus is on a two dimensional surface, e.g., a substantially planar surface. In some instances, a locus refers to a discrete raised or lowered site on a surface e.g., a well, micro well, channel, or post. In some instances, a surface of a locus comprises a material that is actively functionalized to attach to at least one nucleotide for polynucleotide synthesis, or preferably, a population of identical nucleotides for synthesis of a population of polynucleotides. In some instances, polynucleotide refers to a population of polynucleotides encoding for the same nucleic acid sequence. In some instances, a surface of a device is inclusive of one or a plurality of surfaces of a substrate.

[0201] Provided herein are structures that may comprise a surface that supports the synthesis of a plurality of polynucleotides having different predetermined sequences at addressable locations on a common support. In some instances, a device provides support for the synthesis of more than 2,000; 5,000; 10,000; 20,000; 30,000; 50,000; 75,000; 100,000; 200,000; 300,000; 400,000; 500,000; 600,000; 700,000; 800,000; 900,000; 1,000,000; 1,200,000; 1,400,000; 1,600,000; 1,800,000; 2,000,000; 2,500,000; 3,000,000; 3,500,000; 4,000,000; 4,500,000; 5,000,000; 10,000,000 or more non-identical polynucleotides. In some instances, the device provides support for the synthesis of more than 2,000; 5,000; 10,000; 20,000; 30,000; 50,000; 75,000; 100,000; 200,000; 300,000; 400,000; 500,000; 600,000; 700,000; 800,000; 900,000; 1,000,000; 1,200,000; 1,400,000; 1,600,000; 1,800,000; 2,000,000; 2,500,000; 3,000,000; 3,500,000; 4,000,000; 4,500,000; 5,000,000; 10,000,000 or more polynucleotides encoding for distinct sequences. In some instances, at least a portion of the polynucleotides have an identical sequence or are configured to be synthesized with an identical sequence.

[0202] Provided herein are methods and devices for manufacture and growth of polynucleotides about 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, or 2000 bases in length. In some instances, the length of the polynucleotide formed is about 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, or 225 bases in length. A polynucleotide may be at least 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 bases in length. A polynucleotide may be from 10 to 225 bases in length, from 12 to 100 bases in length, from 20 to 150 bases in length, from 20 to 130 bases in length, or from 30 to 100 bases in length.

[0203] In some instances, polynucleotides are synthesized on distinct loci of a substrate, wherein each locus supports the synthesis of a population of polynucleotides. In some instances, each locus supports the synthesis of a population of polynucleotides having a different sequence than a population of polynucleotides grown on another locus. In some instances, the loci of a device are located within a plurality of clusters. In some instances, a device comprises at least 10, 500, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 11000, 12000, 13000, 14000, 15000, 20000, 30000, 40000, 50000 or more clusters. In some instances, a device comprises more than 2,000; 5,000; 10,000; 100,000; 200,000; 300,000; 400,000; 500,000; 600,000; 700,000; 800,000; 900,000; 1,000,000; 1,100,000; 1,200,000; 1,300,000; 1,400,000; 1,500,000; 1,600,000; 1,700,000; 1,800,000; 1,900,000; 2,000,000; 300,000; 400,000; 500,000; 600,000; 700,000; 800,000; 900,000; 1,000,000; 1,200,000; 1,400,000; 1,600,000; 1,800,000; 2,000,000; 2,500,000; 3,000,000; 3,500,000; 4,000,000; 4,500,000; 5,000,000; or 10,000,000 or more distinct loci. In some instances, a device comprises about 10,000 distinct loci. The amount of loci within a single

cluster is varied in different instances. In some instances, each cluster includes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 130, 150, 200, 300, 400, 500, 1000 or more loci. In some instances, each cluster includes about 50-500 loci. In some instances, each cluster includes about 100-200 loci. In some instances, each cluster includes about 100-150 loci. In some instances, each cluster includes about 109, 121, 130 or 137 loci. In some instances, each cluster includes about 19, 20, 61, 64 or more loci.

[0204] The number of distinct polynucleotides synthesized on a device may be dependent on the number of distinct loci available in the substrate. In some instances, the density of loci within a cluster of a device is at least or about 1 locus per mm^2 , 10 loci per mm^2 , 25 loci per mm^2 , 50 loci per mm^2 , 65 loci per mm^2 , 75 loci per mm^2 , 100 loci per mm^2 , 130 loci per mm^2 , 150 loci per mm^2 , 175 loci per mm^2 , 200 loci per mm^2 , 300 loci per mm^2 , 400 loci per mm^2 , 500 loci per mm^2 , 1,000 loci per mm^2 or more. In some instances, a device comprises from about 10 loci per mm^2 to about 500 mm^2 , from about 25 loci per mm^2 to about 400 mm^2 , from about 50 loci per mm^2 to about 500 mm^2 , from about 100 loci per mm^2 to about 500 mm^2 , from about 150 loci per mm^2 to about 500 mm^2 , from about 10 loci per mm^2 to about 250 mm^2 , from about 50 loci per mm^2 to about 250 mm^2 , from about 10 loci per mm^2 to about 200 mm^2 , or from about 50 loci per mm^2 to about 200 mm^2 . In some instances, the distance from the centers of two adjacent loci within a cluster is from about 10 μm to about 500 μm , from about 10 μm to about 200 μm , or from about 10 μm to about 100 μm . In some instances, the distance from two centers of adjacent loci is greater than about 10 μm , 20 μm , 30 μm , 40 μm , 50 μm , 60 μm , 70 μm , 80 μm , 90 μm or 100 μm . In some instances, the distance from the centers of two adjacent loci is less than about 200 μm , 150 μm , 100 μm , 80 μm , 70 μm , 60 μm , 50 μm , 40 μm , 30 μm , 20 μm or 10 μm . In some instances, each locus has a width of about 0.5 μm , 1 μm , 2 μm , 3 μm , 4 μm , 5 μm , 6 μm , 7 μm , 8 μm , 9 μm , 10 μm , 20 μm , 30 μm , 40 μm , 50 μm , 60 μm , 70 μm , 80 μm , 90 μm or 100 μm . In some instances, the each locus is has a width of about 0.5 μm to 100 μm , about 0.5 μm to 50 μm , about 10 μm to 75 μm , or about 0.5 μm to 50 μm .

[0205] In some instances, the density of clusters within a device is at least or about 1 cluster per 100 mm^2 , 1 cluster per 10 mm^2 , 1 cluster per 5 mm^2 , 1 cluster per 4 mm^2 , 1 cluster per 3 mm^2 , 1 cluster per 2 mm^2 , 1 cluster per 1 mm^2 , 2 clusters per 1 mm^2 , 3 clusters per 1 mm^2 , 4 clusters per 1 mm^2 , 5 clusters per 1 mm^2 , 10 clusters per 1 mm^2 , 50 clusters per 1 mm^2 or more. In some instances, a device comprises from about 1 cluster per 10 mm^2 to about 10 clusters per 1 mm^2 . In some instances, the distance from the centers of two adjacent clusters is less than about 50 μm , 100 μm , 200 μm , 500 μm , 1000 μm , or 2000 μm or 5000 μm . In some instances, the distance from the centers of two adjacent clusters is from about 50 μm and about 100 μm , from about 50 μm and

about 200 μm , from about 50 μm and about 300 μm , from about 50 μm and about 500 μm , and from about 100 μm to about 2000 μm . In some instances, the distance from the centers of two adjacent clusters is from about 0.05 mm to about 50 mm, from about 0.05 mm to about 10 mm, from about 0.05 mm and about 5 mm, from about 0.05 mm and about 4 mm, from about 0.05 mm and about 3 mm, from about 0.05 mm and about 2 mm, from about 0.1 mm and 10 mm, from about 0.2 mm and 10 mm, from about 0.3 mm and about 10 mm, from about 0.4 mm and about 10 mm, from about 0.5 mm and 10 mm, from about 0.5 mm and about 5 mm, or from about 0.5 mm and about 2 mm. In some instances, each cluster has a diameter or width along one dimension of about 0.5 to 2 mm, about 0.5 to 1 mm, or about 1 to 2 mm. In some instances, each cluster has a diameter or width along one dimension of about 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9 or 2 mm. In some instances, each cluster has an interior diameter or width along one dimension of about 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.1, 1.15, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9 or 2 mm.

[0206] A device may be about the size of a standard 96 well plate, for example from about 100 and 200 mm by from about 50 and 150 mm. In some instances, a device has a diameter less than or equal to about 1000 mm, 500 mm, 450 mm, 400 mm, 300 mm, 250 mm, 200 mm, 150 mm, 100 mm or 50 mm. In some instances, the diameter of a device is from about 25 mm and 1000 mm, from about 25 mm and about 800 mm, from about 25 mm and about 600 mm, from about 25 mm and about 500 mm, from about 25 mm and about 400 mm, from about 25 mm and about 300 mm, or from about 25 mm and about 200. Non-limiting examples of device size include about 300 mm, 200 mm, 150 mm, 130 mm, 100 mm, 76 mm, 51 mm and 25 mm. In some instances, a device has a planar surface area of at least about 100 mm^2 ; 200 mm^2 ; 500 mm^2 ; 1,000 mm^2 ; 2,000 mm^2 ; 5,000 mm^2 ; 10,000 mm^2 ; 12,000 mm^2 ; 15,000 mm^2 ; 20,000 mm^2 ; 30,000 mm^2 ; 40,000 mm^2 ; 50,000 mm^2 or more. In some instances, the thickness of a device is from about 50 mm and about 2000 mm, from about 50 mm and about 1000 mm, from about 100 mm and about 1000 mm, from about 200 mm and about 1000 mm, or from about 250 mm and about 1000 mm. Non-limiting examples of device thickness include 275 mm, 375 mm, 525 mm, 625 mm, 675 mm, 725 mm, 775 mm and 925 mm. In some instances, the thickness of a device varies with diameter and depends on the composition of the substrate. For example, a device comprising materials other than silicon has a different thickness than a silicon device of the same diameter. Device thickness may be determined by the mechanical strength of the material used and the device must be thick enough to support its own weight without cracking during handling. In some instances, a structure comprises a plurality of devices described herein.

[0207] Surface Materials

[0208] Provided herein is a device comprising a surface, wherein the surface is modified to support polynucleotide synthesis at predetermined locations and with a resulting low error rate, a low dropout rate, a high yield, and a high oligo representation. In some embodiments, surfaces of a device for polynucleotide synthesis provided herein are fabricated from a variety of materials capable of modification to support a de novo polynucleotide synthesis reaction. In some cases, the devices are sufficiently conductive, *e.g.*, are able to form uniform electric fields across all or a portion of the device. A device described herein may comprise a flexible material. Exemplary flexible materials include, without limitation, modified nylon, unmodified nylon, nitrocellulose, and polypropylene. A device described herein may comprise a rigid material. Exemplary rigid materials include, without limitation, glass, fused silica, silicon, silicon dioxide, silicon nitride, plastics (for example, polytetrafluoroethylene, polypropylene, polystyrene, polycarbonate, and blends thereof, and metals (for example, gold, platinum). Device disclosed herein may be fabricated from a material comprising silicon, polystyrene, agarose, dextran, cellulosic polymers, polyacrylamides, polydimethylsiloxane (PDMS), glass, or any combination thereof. In some cases, a device disclosed herein is manufactured with a combination of materials listed herein or any other suitable material known in the art.

[0209] A listing of tensile strengths for exemplary materials described herein is provided as follows: nylon (70 MPa), nitrocellulose (1.5 MPa), polypropylene (40 MPa), silicon (268 MPa), polystyrene (40 MPa), agarose (1-10 MPa), polyacrylamide (1-10 MPa), polydimethylsiloxane (PDMS) (3.9-10.8 MPa). Solid supports described herein can have a tensile strength from 1 to 300, 1 to 40, 1 to 10, 1 to 5, or 3 to 11 MPa. Solid supports described herein can have a tensile strength of about 1, 1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 20, 25, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 270, or more MPa. In some instances, a device described herein comprises a solid support for polynucleotide synthesis that is in the form of a flexible material capable of being stored in a continuous loop or reel, such as a tape or flexible sheet.

[0210] Young's modulus measures the resistance of a material to elastic (recoverable) deformation under load. A listing of Young's modulus for stiffness of exemplary materials described herein is provided as follows: nylon (3 GPa), nitrocellulose (1.5 GPa), polypropylene (2 GPa), silicon (150 GPa), polystyrene (3 GPa), agarose (1-10 GPa), polyacrylamide (1-10 GPa), polydimethylsiloxane (PDMS) (1-10 GPa). Solid supports described herein can have a Young's moduli from 1 to 500, 1 to 40, 1 to 10, 1 to 5, or 3 to 11 GPa. Solid supports described herein can have a Young's moduli of about 1, 1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 20, 25, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 400, 500 GPa, or more. As the relationship between flexibility and stiffness are

inverse to each other, a flexible material has a low Young's modulus and changes its shape considerably under load.

[0211] In some cases, a device disclosed herein comprises a silicon dioxide base and a surface layer of silicon oxide. Alternatively, the device may have a base of silicon oxide. Surface of the device provided here may be textured, resulting in an increase overall surface area for polynucleotide synthesis. Device disclosed herein may comprise at least 5 %, 10%, 25%, 50%, 80%, 90%, 95%, or 99% silicon. A device disclosed herein may be fabricated from a silicon on insulator (SOI) wafer.

[0212] Surface Architecture

[0213] Provided herein are devices comprising raised and/or lowered features. One benefit of having such features is an increase in surface area to support polynucleotide synthesis. In some instances, a device having raised and/or lowered features is referred to as a three-dimensional substrate. In some instances, a three-dimensional device comprises one or more channels. In some instances, one or more loci comprise a channel. In some instances, the channels are accessible to reagent deposition via a deposition device such as a polynucleotide synthesizer. In some instances, reagents and/or fluids collect in a larger well in fluid communication one or more channels. For example, a device comprises a plurality of channels corresponding to a plurality of loci with a cluster, and the plurality of channels are in fluid communication with one well of the cluster. In some methods, a library of polynucleotides is synthesized in a plurality of loci of a cluster.

[0214] In some instances, the structure is configured to allow for controlled flow and mass transfer paths for polynucleotide synthesis on a surface. In some instances, the configuration of a device allows for the controlled and even distribution of mass transfer paths, chemical exposure times, and/or wash efficacy during polynucleotide synthesis. In some instances, the configuration of a device allows for increased sweep efficiency, for example by providing sufficient volume for a growing a polynucleotide such that the excluded volume by the growing polynucleotide does not take up more than 50, 45, 40, 35, 30, 25, 20, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1%, or less of the initially available volume that is available or suitable for growing the polynucleotide. In some instances, a three-dimensional structure allows for managed flow of fluid to allow for the rapid exchange of chemical exposure.

[0215] Provided herein are methods to synthesize an amount of DNA of 1 fM, 5 fM, 10 fM, 25 fM, 50 fM, 75 fM, 100 fM, 200 fM, 300 fM, 400 fM, 500 fM, 600 fM, 700 fM, 800 fM, 900 fM, 1 pM, 5 pM, 10 pM, 25 pM, 50 pM, 75 pM, 100 pM, 200 pM, 300 pM, 400 pM, 500 pM, 600 pM, 700 pM, 800 pM, 900 pM, or more. In some instances, a polynucleotide library may span the length of about 1 %, 2 %, 3 %, 4 %, 5 %, 10 %, 15 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 100 %, or more.

95 %, or 100 % of a gene. A gene may be varied up to about 1 %, 2 %, 3 %, 4 %, 5 %, 10 %, 15 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 85 %, 90 %, 95 %, or 100 %.

[0216] Non-identical polynucleotides may collectively encode a sequence for at least 1 %, 2 %, 3 %, 4 %, 5 %, 10 %, 15 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 85 %, 90 %, 95 %, or 100 % of a gene. In some instances, a polynucleotide may encode a sequence of 50 %, 60 %, 70 %, 80 %, 85 %, 90 %, 95 %, or more of a gene. In some instances, a polynucleotide may encode a sequence of 80 %, 85 %, 90 %, 95 %, or more of a gene.

[0217] In some instances, segregation is achieved by physical structure. In some instances, segregation is achieved by differential functionalization of the surface generating active and passive regions for polynucleotide synthesis. Differential functionalization is also achieved by alternating the hydrophobicity across the device surface, thereby creating water contact angle effects that cause beading or wetting of the deposited reagents. Employing larger structures can decrease splashing and cross-contamination of distinct polynucleotide synthesis locations with reagents of the neighboring spots. In some instances, a device, such as a polynucleotide synthesizer, is used to deposit reagents to distinct polynucleotide synthesis locations. Substrates having three-dimensional features are configured in a manner that allows for the synthesis of a large number of polynucleotides (*e.g.*, more than about 10,000) with a low error rate (*e.g.*, less than about 1:500, 1:1000, 1:1500, 1:2,000; 1:3,000; 1:5,000; or 1:10,000). In some instances, a device comprises features with a density of about or greater than about 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 300, 400 or 500 features per mm².

[0218] A well of a device may have the same or different width, height, and/or volume as another well of the substrate. A channel of a device may have the same or different width, height, and/or volume as another channel of the substrate. In some instances, the width of a cluster is from about 0.05 mm to about 50 mm, from about 0.05 mm to about 10 mm, from about 0.05 mm and about 5 mm, from about 0.05 mm and about 4 mm, from about 0.05 mm and about 3 mm, from about 0.05 mm and about 2 mm, from about 0.05 mm and about 1 mm, from about 0.05 mm and about 0.5 mm, from about 0.05 mm and about 0.1 mm, from about 0.1 mm and 10 mm, from about 0.2 mm and 10 mm, from about 0.3 mm and about 10 mm, from about 0.4 mm and about 10 mm, from about 0.5 mm and 10 mm, from about 0.5 mm and about 5 mm, or from about 0.5 mm and about 2 mm. In some instances, the width of a well comprising a cluster is from about 0.05 mm to about 50 mm, from about 0.05 mm to about 10 mm, from about 0.05 mm and about 5 mm, from about 0.05 mm and about 4 mm, from about 0.05 mm and about 3 mm, from about 0.05 mm and about 2 mm, from about 0.05 mm and about 1 mm, from about 0.05 mm and about 0.5 mm, from about 0.05 mm and about 0.1 mm, from about 0.1 mm and 10 mm, from about 0.2 mm and 10 mm,

from about 0.3 mm and about 10 mm, from about 0.4 mm and about 10 mm, from about 0.5 mm and 10 mm, from about 0.5 mm and about 5 mm, or from about 0.5 mm and about 2 mm. In some instances, the width of a cluster is less than or about 5 mm, 4 mm, 3 mm, 2 mm, 1 mm, 0.5 mm, 0.1 mm, 0.09 mm, 0.08 mm, 0.07 mm, 0.06 mm or 0.05 mm. In some instances, the width of a cluster is from about 1.0 and 1.3 mm. In some instances, the width of a cluster is about 1.150 mm. In some instances, the width of a well is less than or about 5 mm, 4 mm, 3 mm, 2 mm, 1 mm, 0.5 mm, 0.1 mm, 0.09 mm, 0.08 mm, 0.07 mm, 0.06 mm or 0.05 mm. In some instances, the width of a well is from about 1.0 and 1.3 mm. In some instances, the width of a well is about 1.150 mm. In some instances, the width of a cluster is about 0.08 mm. In some instances, the width of a well is about 0.08 mm. The width of a cluster may refer to clusters within a two-dimensional or three-dimensional substrate.

[0219] In some instances, the height of a well is from about 20 um to about 1000 um, from about 50 um to about 1000 um, from about 100 um to about 1000 um, from about 200 um to about 1000 um, from about 300 um to about 1000 um, from about 400 um to about 1000 um, or from about 500 um to about 1000 um. In some instances, the height of a well is less than about 1000 um, less than about 900 um, less than about 800 um, less than about 700 um, or less than about 600 um.

[0220] In some instances, a device comprises a plurality of channels corresponding to a plurality of loci within a cluster, wherein the height or depth of a channel is from about 5 um to about 500 um, from about 5 um to about 400 um, from about 5 um to about 300 um, from about 5 um to about 200 um, from about 5 um to about 100 um, from about 5 um to about 50 um, or from about 10 um to about 50 um. In some instances, the height of a channel is less than 100 um, less than 80 um, less than 60 um, less than 40 um or less than 20 um.

[0221] In some instances, the diameter of a channel, locus (*e.g.*, in a substantially planar substrate) or both channel and locus (*e.g.*, in a three-dimensional device wherein a locus corresponds to a channel) is from about 1 um to about 1000 um, from about 1 um to about 500 um, from about 1 um to about 200 um, from about 1 um to about 100 um, from about 5 um to about 100 um, or from about 10 um to about 100 um, for example, about 90 um, 80 um, 70 um, 60 um, 50 um, 40 um, 30 um, 20 um or 10 um. In some instances, the diameter of a channel, locus, or both channel and locus is less than about 100 um, 90 um, 80 um, 70 um, 60 um, 50 um, 40 um, 30 um, 20 um or 10 um. In some instances, the distance from the center of two adjacent channels, loci, or channels and loci is from about 1 um to about 500 um, from about 1 um to about 200 um, from about 1 um to about 100 um, from about 5 um to about 200 um, from about 5 um to about 100 um, from about 5 um to about 50 um, or from about 5 um to about 30 um, for example, about 20 um.

[0222] Surface Modifications

[0223] In various instances, surface modifications are employed for the chemical and/or physical alteration of a surface by an additive or subtractive process to change one or more chemical and/or physical properties of a device surface or a selected site or region of a device surface. For example, surface modifications include, without limitation, (1) changing the wetting properties of a surface, (2) functionalizing a surface, i.e., providing, modifying or substituting surface functional groups, (3) defunctionalizing a surface, i.e., removing surface functional groups, (4) otherwise altering the chemical composition of a surface, *e.g.*, through etching, (5) increasing or decreasing surface roughness, (6) providing a coating on a surface, *e.g.*, a coating that exhibits wetting properties that are different from the wetting properties of the surface, and/or (7) depositing particulates on a surface.

[0224] In some instances, the addition of a chemical layer on top of a surface (referred to as adhesion promoter) facilitates structured patterning of loci on a surface of a substrate. Exemplary surfaces for application of adhesion promotion include, without limitation, glass, silicon, silicon dioxide and silicon nitride. In some instances, the adhesion promoter is a chemical with a high surface energy. In some instances, a second chemical layer is deposited on a surface of a substrate. In some instances, the second chemical layer has a low surface energy. In some instances, surface energy of a chemical layer coated on a surface supports localization of droplets on the surface. Depending on the patterning arrangement selected, the proximity of loci and/or area of fluid contact at the loci are alterable.

[0225] In some instances, a device surface, or resolved loci, onto which nucleic acids or other moieties are deposited, *e.g.*, for polynucleotide synthesis, are smooth or substantially planar (*e.g.*, two-dimensional) or have irregularities, such as raised or lowered features (*e.g.*, three-dimensional features). In some instances, a device surface is modified with one or more different layers of compounds. Such modification layers of interest include, without limitation, inorganic and organic layers such as metals, metal oxides, polymers, small organic molecules and the like. Non-limiting polymeric layers include peptides, proteins, nucleic acids or mimetics thereof (*e.g.*, peptide nucleic acids and the like), polysaccharides, phospholipids, polyurethanes, polyesters, polycarbonates, polyureas, polyamides, polyethyleneamines, polyarylene sulfides, polysiloxanes, polyimides, polyacetates, and any other suitable compounds described herein or otherwise known in the art. In some instances, polymers are heteropolymeric. In some instances, polymers are homopolymeric. In some instances, polymers comprise functional moieties or are conjugated.

[0226] In some instances, resolved loci of a device are functionalized with one or more moieties that increase and/or decrease surface energy. In some instances, a moiety is chemically inert. In some instances, a moiety is configured to support a desired chemical reaction, for example,

one or more processes in a polynucleotide synthesis reaction. The surface energy, or hydrophobicity, of a surface is a factor for determining the affinity of a nucleotide to attach onto the surface. In some instances, a method for device functionalization may comprise: (a) providing a device having a surface that comprises silicon dioxide; and (b) silanizing the surface using, a suitable silanizing agent described herein or otherwise known in the art, for example, an organofunctional alkoxysilane molecule.

[0227] In some instances, the organofunctional alkoxysilane molecule comprises dimethylchloro-octodecyl-silane, methylchloro-octodecyl-silane, trichloro-octodecyl-silane, trimethyl-octodecyl-silane, triethyl-octodecyl-silane, or any combination thereof. In some instances, a device surface comprises functionalized with polyethylene/polypropylene (functionalized by gamma irradiation or chromic acid oxidation, and reduction to hydroxyalkyl surface), highly crosslinked polystyrene-divinylbenzene (derivatized by chloromethylation, and aminated to benzylamine functional surface), nylon (the terminal aminohexyl groups are directly reactive), or etched with reduced polytetrafluoroethylene. Other methods and functionalizing agents are described in U.S. Patent No. 5474796, which is herein incorporated by reference in its entirety.

[0228] In some instances, a device surface is functionalized by contact with a derivatizing composition that contains a mixture of silanes, under reaction conditions effective to couple the silanes to the device surface, typically via reactive hydrophilic moieties present on the device surface. Silanization generally covers a surface through self-assembly with organofunctional alkoxysilane molecules.

[0229] A variety of siloxane functionalizing reagents can further be used as currently known in the art, *e.g.*, for lowering or increasing surface energy. The organofunctional alkoxysilanes can be classified according to their organic functions.

[0230] Provided herein are devices that may contain patterning of agents capable of coupling to a nucleoside. In some instances, a device may be coated with an active agent. In some instances, a device may be coated with a passive agent. Exemplary active agents for inclusion in coating materials described herein includes, without limitation, N-(3-triethoxysilylpropyl)-4-hydroxybutyramide (HAPS), 11-acetoxyundecyltriethoxysilane, n-decyltriethoxysilane, (3-aminopropyl)trimethoxysilane, (3-aminopropyl)triethoxysilane, 3-glycidoxypropyltrimethoxysilane (GOPS), 3-iodo-propyltrimethoxysilane, butyl-aldehyde-trimethoxysilane, dimeric secondary aminoalkyl siloxanes, (3-aminopropyl)-diethoxy-methylsilane, (3-aminopropyl)-dimethyl-ethoxysilane, and (3-aminopropyl)-trimethoxysilane, (3-glycidoxypropyl)-dimethyl-ethoxysilane, glycidoxy-trimethoxysilane, (3-mercaptopropyl)-trimethoxysilane, 3-4 epoxycyclohexyl-

ethyltrimethoxysilane, and (3-mercaptopropyl)-methyl-dimethoxysilane, allyl trichlorochlorosilane, 7-oct-1-enyl trichlorochlorosilane, or bis (3-trimethoxysilylpropyl) amine.

[0231] Exemplary passive agents for inclusion in a coating material described herein includes, without limitation, perfluorooctyltrichlorosilane; tridecafluoro-1,1,2,2-tetrahydrooctyl)trichlorosilane; 1H, 1H, 2H, 2H-fluorooctyltriethoxysilane (FOS); trichloro(1H, 1H, 2H, 2H - perfluorooctyl)silane; tert-butyl-[5-fluoro-4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)indol-1-yl]-dimethyl-silane; CYTOP™; Fluorinert™; perfluorooctyltrichlorosilane (PFOTCS); perfluorooctyldimethylchlorosilane (PFODCS); perfluorodecyltriethoxysilane (PFDTES); pentafluorophenyl-dimethylpropylchloro-silane (PFPTES); perfluorooctyltriethoxysilane; perfluorooctyltrimethoxysilane; octylchlorosilane; dimethylchloro-octadecyl-silane; methylchloro-octadecyl-silane; trichloro-octadecyl-silane; trimethyl-octadecyl-silane; triethyl-octadecyl-silane; or octadecyltrichlorosilane.

[0232] In some instances, a functionalization agent comprises a hydrocarbon silane such as octadecyltrichlorosilane. In some instances, the functionalizing agent comprises 11-acetoxyundecyltriethoxysilane, n-decyltriethoxysilane, (3-aminopropyl)trimethoxysilane, (3-aminopropyl)triethoxysilane, glycidyloxypropyl/trimethoxysilane and N-(3-triethoxysilylpropyl)-4-hydroxybutyramide.

[0233] Polynucleotide Synthesis

[0234] Methods of the current disclosure for polynucleotide synthesis may include processes involving phosphoramidite chemistry. In some instances, polynucleotide synthesis comprises coupling a base with phosphoramidite. Polynucleotide synthesis may comprise coupling a base by deposition of phosphoramidite under coupling conditions, wherein the same base is optionally deposited with phosphoramidite more than once, i.e., double coupling. Polynucleotide synthesis may comprise capping of unreacted sites. In some instances, capping is optional. Polynucleotide synthesis may also comprise oxidation or an oxidation step or oxidation steps. Polynucleotide synthesis may comprise deblocking, detritylation, and sulfurization. In some instances, polynucleotide synthesis comprises either oxidation or sulfurization. In some instances, between one or each step during a polynucleotide synthesis reaction, the device is washed, for example, using tetrazole or acetonitrile. Time frames for any one step in a phosphoramidite synthesis method may be less than about 2 minutes, 1 minute, 50 seconds, 40 seconds, 30 seconds, 20 seconds and 10 seconds.

[0235] Polynucleotide synthesis using a phosphoramidite method may comprise a subsequent addition of a phosphoramidite building block (e.g., nucleoside phosphoramidite) to a growing polynucleotide chain for the formation of a phosphite triester linkage. Phosphoramidite

polynucleotide synthesis proceeds in the 3' to 5' direction. Phosphoramidite polynucleotide synthesis allows for the controlled addition of one nucleotide to a growing nucleic acid chain per synthesis cycle. In some instances, each synthesis cycle comprises a coupling step. Phosphoramidite coupling involves the formation of a phosphite triester linkage between an activated nucleoside phosphoramidite and a nucleoside bound to the substrate, for example, via a linker. In some instances, the nucleoside phosphoramidite is provided to the device activated. In some instances, the nucleoside phosphoramidite is provided to the device with an activator. In some instances, nucleoside phosphoramidites are provided to the device in a 1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100-fold excess or more over the substrate-bound nucleosides. In some instances, the addition of nucleoside phosphoramidite is performed in an anhydrous environment, for example, in anhydrous acetonitrile. Following addition of a nucleoside phosphoramidite, the device is optionally washed. In some instances, the coupling step is repeated one or more additional times, optionally with a wash step between nucleoside phosphoramidite additions to the substrate. In some instances, a polynucleotide synthesis method used herein comprises 1, 2, 3 or more sequential coupling steps. Prior to coupling, in many cases, the nucleoside bound to the device is de-protected by removal of a protecting group, where the protecting group functions to prevent polymerization. A common protecting group is 4,4'-dimethoxytrityl (DMT).

[0236] Following coupling, phosphoramidite polynucleotide synthesis methods optionally comprise a capping step. In a capping step, the growing polynucleotide is treated with a capping agent. A capping step is useful to block unreacted substrate-bound 5'-OH groups after coupling from further chain elongation, preventing the formation of polynucleotides with internal base deletions. Further, phosphoramidites activated with 1H-tetrazole may react, to a small extent, with the O6 position of guanosine. Without being bound by theory, upon oxidation with I₂/water, this side product, possibly via O6-N7 migration, may undergo depurination. The apurinic sites may end up being cleaved in the course of the final deprotection of the polynucleotide thus reducing the yield of the full-length product. The O6 modifications may be removed by treatment with the capping reagent prior to oxidation with I₂/water. In some instances, inclusion of a capping step during polynucleotide synthesis decreases the error rate as compared to synthesis without capping. As an example, the capping step comprises treating the substrate-bound polynucleotide with a mixture of acetic anhydride and 1-methylimidazole. Following a capping step, the device is optionally washed.

[0237] In some instances, following addition of a nucleoside phosphoramidite, and optionally after capping and one or more wash steps, the device bound growing nucleic acid is oxidized. The

oxidation step comprises the phosphite triester is oxidized into a tetracoordinated phosphate triester, a protected precursor of the naturally occurring phosphate diester internucleoside linkage. In some instances, oxidation of the growing polynucleotide is achieved by treatment with iodine and water, optionally in the presence of a weak base (*e.g.*, pyridine, lutidine, collidine). Oxidation may be carried out under anhydrous conditions using, *e.g.* *tert*-Butyl hydroperoxide or (1*S*)-(+)-(10-camphorsulfonyl)-oxaziridine (CSO). In some methods, a capping step is performed following oxidation. A second capping step allows for device drying, as residual water from oxidation that may persist can inhibit subsequent coupling. Following oxidation, the device and growing polynucleotide is optionally washed. In some instances, the step of oxidation is substituted with a sulfurization step to obtain polynucleotide phosphorothioates, wherein any capping steps can be performed after the sulfurization. Many reagents are capable of the efficient sulfur transfer, including but not limited to 3-(Dimethylaminomethylidene)amino)-3H-1,2,4-dithiazole-3-thione, DDTT, 3H-1,2-benzodithiol-3-one 1,1-dioxide, also known as Beaucage reagent, and N,N,N'N'-Tetraethylthiuram disulfide (TETD).

[0238] In order for a subsequent cycle of nucleoside incorporation to occur through coupling, the protected 5' end of the device bound growing polynucleotide is removed so that the primary hydroxyl group is reactive with a next nucleoside phosphoramidite. In some instances, the protecting group is DMT and deblocking occurs with trichloroacetic acid in dichloromethane. Conducting detritylation for an extended time or with stronger than recommended solutions of acids may lead to increased depurination of solid support-bound polynucleotide and thus reduces the yield of the desired full-length product. Methods and compositions of the disclosure described herein provide for controlled deblocking conditions limiting undesired depurination reactions. In some instances, the device bound polynucleotide is washed after deblocking. In some instances, efficient washing after deblocking contributes to synthesized polynucleotides having a low error rate.

[0239] Methods for the synthesis of polynucleotides typically involve an iterating sequence of the following steps: application of a protected monomer to an actively functionalized surface (*e.g.*, locus) to link with either the activated surface, a linker or with a previously deprotected monomer; deprotection of the applied monomer so that it is reactive with a subsequently applied protected monomer; and application of another protected monomer for linking. One or more intermediate steps include oxidation or sulfurization. In some instances, one or more wash steps precede or follow one or all of the steps.

[0240] Methods for phosphoramidite-based polynucleotide synthesis comprise a series of chemical steps. In some instances, one or more steps of a synthesis method involve reagent cycling,

where one or more steps of the method comprise application to the device of a reagent useful for the step. For example, reagents are cycled by a series of liquid deposition and vacuum drying steps. For substrates comprising three-dimensional features such as wells, microwells, channels and the like, reagents are optionally passed through one or more regions of the device via the wells and/or channels.

[0241] Methods and systems described herein relate to polynucleotide synthesis devices for the synthesis of polynucleotides. The synthesis may be in parallel. For example at least or about at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 1000, 10000, 50000, 75000, 100000 or more polynucleotides can be synthesized in parallel. The total number polynucleotides that may be synthesized in parallel may be from 2-100000, 3-50000, 4-10000, 5-1000, 6-900, 7-850, 8-800, 9-750, 10-700, 11-650, 12-600, 13-550, 14-500, 15-450, 16-400, 17-350, 18-300, 19-250, 20-200, 21-150, 22-100, 23-50, 24-45, 25-40, 30-35. Those of skill in the art appreciate that the total number of polynucleotides synthesized in parallel may fall within any range bound by any of these values, for example 25-100. The total number of polynucleotides synthesized in parallel may fall within any range defined by any of the values serving as endpoints of the range. Total molar mass of polynucleotides synthesized within the device or the molar mass of each of the polynucleotides may be at least or at least about 10, 20, 30, 40, 50, 100, 250, 500, 750, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 25000, 50000, 75000, 100000 picomoles, or more. The length of each of the polynucleotides or average length of the polynucleotides within the device may be at least or about at least 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 150, 200, 300, 400, 500 nucleotides, or more. The length of each of the polynucleotides or average length of the polynucleotides within the device may be at most or about at most 500, 400, 300, 200, 150, 100, 50, 45, 35, 30, 25, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10 nucleotides, or less. The length of each of the polynucleotides or average length of the polynucleotides within the device may fall from 10-500, 9-400, 11-300, 12-200, 13-150, 14-100, 15-50, 16-45, 17-40, 18-35, 19-25. Those of skill in the art appreciate that the length of each of the polynucleotides or average length of the polynucleotides within the device may fall within any range bound by any of these values, for example 100-300. The length of each of the polynucleotides or average length of the polynucleotides within the device may fall within any range defined by any of the values serving as endpoints of the range.

[0242] Methods for polynucleotide synthesis on a surface provided herein allow for synthesis at a fast rate. As an example, at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 55, 60, 70, 80, 90, 100, 125, 150, 175, 200

nucleotides per hour, or more are synthesized. Nucleotides include adenine, guanine, thymine, cytosine, uridine building blocks, or analogs/modified versions thereof. In some instances, libraries of polynucleotides are synthesized in parallel on substrate. For example, a device comprising about or at least about 100; 1,000; 10,000; 30,000; 75,000; 100,000; 1,000,000; 2,000,000; 3,000,000; 4,000,000; or 5,000,000 resolved loci is able to support the synthesis of at least the same number of distinct polynucleotides, wherein polynucleotide encoding a distinct sequence is synthesized on a resolved locus. In some instances, a library of polynucleotides are synthesized on a device with low error rates described herein in less than about three months, two months, one month, three weeks, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 days, 24 hours or less. In some instances, larger nucleic acids assembled from a polynucleotide library synthesized with low error rate using the substrates and methods described herein are prepared in less than about three months, two months, one month, three weeks, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 days, 24 hours or less.

[0243] In some instances, methods described herein provide for generation of a library of polynucleotides comprising variant polynucleotides differing at a plurality of codon sites. In some instances, a polynucleotide may have 1 site, 2 sites, 3 sites, 4 sites, 5 sites, 6 sites, 7 sites, 8 sites, 9 sites, 10 sites, 11 sites, 12 sites, 13 sites, 14 sites, 15 sites, 16 sites, 17 sites, 18 sites, 19 sites, 20 sites, 30 sites, 40 sites, 50 sites, or more of variant codon sites.

[0244] In some instances, the one or more sites of variant codon sites may be adjacent. In some instances, the one or more sites of variant codon sites may be not be adjacent and separated by 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more codons.

[0245] In some instances, a polynucleotide may comprise multiple sites of variant codon sites, wherein all the variant codon sites are adjacent to one another, forming a stretch of variant codon sites. In some instances, a polynucleotide may comprise multiple sites of variant codon sites, wherein none the variant codon sites are adjacent to one another. In some instances, a polynucleotide may comprise multiple sites of variant codon sites, wherein some the variant codon sites are adjacent to one another, forming a stretch of variant codon sites, and some of the variant codon sites are not adjacent to one another.

[0246] Referring to the Figures, **FIG. 11** illustrates an exemplary process workflow for synthesis of nucleic acids (e.g., genes) from shorter polynucleotides. The workflow is divided generally into phases: (1) de novo synthesis of a single stranded polynucleotide library, (2) joining polynucleotides to form larger fragments, (3) error correction, (4) quality control, and (5) shipment. Prior to de novo synthesis, an intended nucleic acid sequence or group of nucleic acid sequences is preselected. For example, a group of genes is preselected for generation.

[0247] Once large polynucleotides for generation are selected, a predetermined library of polynucleotides is designed for de novo synthesis. Various suitable methods are known for generating high density polynucleotide arrays. In the workflow example, a device surface layer **1101** is provided. In the example, chemistry of the surface is altered in order to improve the polynucleotide synthesis process. Areas of low surface energy are generated to repel liquid while areas of high surface energy are generated to attract liquids. The surface itself may be in the form of a planar surface or contain variations in shape, such as protrusions or microwells which increase surface area. In the workflow example, high surface energy molecules selected serve a dual function of supporting DNA chemistry, as disclosed in International Patent Application Publication WO/2015/021080, which is herein incorporated by reference in its entirety.

[0248] In situ preparation of polynucleotide arrays is generated on a solid support and utilizes single nucleotide extension process to extend multiple oligomers in parallel. A material deposition device, such as a polynucleotide synthesizer, is designed to release reagents in a step wise fashion such that multiple polynucleotides extend, in parallel, one residue at a time to generate oligomers with a predetermined nucleic acid sequence **1102**. In some instances, polynucleotides are cleaved from the surface at this stage. Cleavage includes gas cleavage, *e.g.*, with ammonia or methylamine.

[0249] The generated polynucleotide libraries are placed in a reaction chamber. In this exemplary workflow, the reaction chamber (also referred to as “nanoreactor”) is a silicon coated well, containing PCR reagents and lowered onto the polynucleotide library **1103**. Prior to or after the sealing **1104** of the polynucleotides, a reagent is added to release the polynucleotides from the substrate. In the exemplary workflow, the polynucleotides are released subsequent to sealing of the nanoreactor **1105**. Once released, fragments of single stranded polynucleotides hybridize in order to span an entire long range sequence of DNA. Partial hybridization **1105** is possible because each synthesized polynucleotide is designed to have a small portion overlapping with at least one other polynucleotide in the population.

[0250] After hybridization, a PCR reaction is commenced. During the polymerase cycles, the polynucleotides anneal to complementary fragments and gaps are filled in by a polymerase. Each cycle increases the length of various fragments randomly depending on which polynucleotides find each other. Complementarity amongst the fragments allows for forming a complete large span of double stranded DNA **1106**.

[0251] After PCR is complete, the nanoreactor is separated from the device **1107** and positioned for interaction with a device having primers for PCR **1108**. After sealing, the nanoreactor is subject to PCR **1109** and the larger nucleic acids are amplified. After PCR **1110**, the nanochamber is opened **1111**, error correction reagents are added **1112**, the chamber is sealed **1113**

and an error correction reaction occurs to remove mismatched base pairs and/or strands with poor complementarity from the double stranded PCR amplification products **1114**. The nanoreactor is opened and separated **1115**. Error corrected product is next subject to additional processing steps, such as PCR and molecular bar coding, and then packaged **1122** for shipment **1123**.

[0252] In some instances, quality control measures are taken. After error correction, quality control steps include for example interaction with a wafer having sequencing primers for amplification of the error corrected product **1116**, sealing the wafer to a chamber containing error corrected amplification product **1117**, and performing an additional round of amplification **1118**. The nanoreactor is opened **1119** and the products are pooled **1120** and sequenced **1121**. After an acceptable quality control determination is made, the packaged product **1122** is approved for shipment **1123**.

[0253] In some instances, a nucleic acid generate by a workflow such as that in **FIG. 11** is subject to mutagenesis using overlapping primers disclosed herein. In some instances, a library of primers are generated by in situ preparation on a solid support and utilize single nucleotide extension process to extend multiple oligomers in parallel. A deposition device, such as a polynucleotide synthesizer, is designed to release reagents in a step wise fashion such that multiple polynucleotides extend, in parallel, one residue at a time to generate oligomers with a predetermined nucleic acid sequence **1102**.

[0254] Large Polynucleotide Libraries Having Low Error Rates

[0255] Average error rates for polynucleotides synthesized within a library using the systems and methods provided may be less than 1 in 1000, less than 1 in 1250, less than 1 in 1500, less than 1 in 2000, less than 1 in 3000 or less often. In some instances, average error rates for polynucleotides synthesized within a library using the systems and methods provided are less than 1/500, 1/600, 1/700, 1/800, 1/900, 1/1000, 1/1100, 1/1200, 1/1250, 1/1300, 1/1400, 1/1500, 1/1600, 1/1700, 1/1800, 1/1900, 1/2000, 1/3000, or less. In some instances, average error rates for polynucleotides synthesized within a library using the systems and methods provided are less than 1/1000.

[0256] In some instances, aggregate error rates for polynucleotides synthesized within a library using the systems and methods provided are less than 1/500, 1/600, 1/700, 1/800, 1/900, 1/1000, 1/1100, 1/1200, 1/1250, 1/1300, 1/1400, 1/1500, 1/1600, 1/1700, 1/1800, 1/1900, 1/2000, 1/3000, or less compared to the predetermined sequences. In some instances, aggregate error rates for polynucleotides synthesized within a library using the systems and methods provided are less than 1/500, 1/600, 1/700, 1/800, 1/900, or 1/1000. In some instances, aggregate error rates for

polynucleotides synthesized within a library using the systems and methods provided are less than 1/1000.

[0257] In some instances, an error correction enzyme may be used for polynucleotides synthesized within a library using the systems and methods provided can use. In some instances, aggregate error rates for polynucleotides with error correction can be less than 1/500, 1/600, 1/700, 1/800, 1/900, 1/1000, 1/1100, 1/1200, 1/1300, 1/1400, 1/1500, 1/1600, 1/1700, 1/1800, 1/1900, 1/2000, 1/3000, or less compared to the predetermined sequences. In some instances, aggregate error rates with error correction for polynucleotides synthesized within a library using the systems and methods provided can be less than 1/500, 1/600, 1/700, 1/800, 1/900, or 1/1000. In some instances, aggregate error rates with error correction for polynucleotides synthesized within a library using the systems and methods provided can be less than 1/1000.

[0258] Error rate may limit the value of gene synthesis for the production of libraries of gene variants. With an error rate of 1/300, about 0.7% of the clones in a 1500 base pair gene will be correct. As most of the errors from polynucleotide synthesis result in frame-shift mutations, over 99% of the clones in such a library will not produce a full-length protein. Reducing the error rate by 75% would increase the fraction of clones that are correct by a factor of 40. The methods and compositions of the disclosure allow for fast de novo synthesis of large polynucleotide and gene libraries with error rates that are lower than commonly observed gene synthesis methods both due to the improved quality of synthesis and the applicability of error correction methods that are enabled in a massively parallel and time-efficient manner. Accordingly, libraries may be synthesized with base insertion, deletion, substitution, or total error rates that are under 1/300, 1/400, 1/500, 1/600, 1/700, 1/800, 1/900, 1/1000, 1/1250, 1/1500, 1/2000, 1/2500, 1/3000, 1/4000, 1/5000, 1/6000, 1/7000, 1/8000, 1/9000, 1/10000, 1/12000, 1/15000, 1/20000, 1/25000, 1/30000, 1/40000, 1/50000, 1/60000, 1/70000, 1/80000, 1/90000, 1/100000, 1/125000, 1/150000, 1/200000, 1/300000, 1/400000, 1/500000, 1/600000, 1/700000, 1/800000, 1/900000, 1/1000000, 1/1250000, 1/1500000, 1/2000000, 1/3000000, 1/4000000, 1/5000000, 1/6000000, 1/7000000, 1/8000000, 1/9000000, 1/10000000, or less, across the library, or across more than 80%, 85%, 90%, 93%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.8%, 99.9%, 99.95%, 99.98%, 99.99%, or more of the library. The methods and compositions of the disclosure further relate to large synthetic polynucleotide and gene libraries with low error rates associated with at least 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 93%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.8%, 99.9%, 99.95%, 99.98%, 99.99%, or more of the polynucleotides or genes in at least a subset of the library to relate to error free sequences in comparison to a predetermined/preselected sequence. In some instances, at least 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 93%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.8%, 99.9%, 99.95%, 99.98%, 99.99%, or more of the polynucleotides or genes in an isolated volume within the library have the

same sequence. In some instances, at least 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 93%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.8%, 99.9%, 99.95%, 99.98%, 99.99%, or more of any polynucleotides or genes related with more than 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9% or more similarity or identity have the same sequence. In some instances, the error rate related to a specified locus on a polynucleotide or gene is optimized. Thus, a given locus or a plurality of selected loci of one or more polynucleotides or genes as part of a large library may each have an error rate that is less than 1/300, 1/400, 1/500, 1/600, 1/700, 1/800, 1/900, 1/1000, 1/1250, 1/1500, 1/2000, 1/2500, 1/3000, 1/4000, 1/5000, 1/6000, 1/7000, 1/8000, 1/9000, 1/10000, 1/12000, 1/15000, 1/20000, 1/25000, 1/30000, 1/40000, 1/50000, 1/60000, 1/70000, 1/80000, 1/90000, 1/100000, 1/125000, 1/150000, 1/200000, 1/300000, 1/400000, 1/500000, 1/600000, 1/700000, 1/800000, 1/900000, 1/1000000, or less. In various instances, such error optimized loci may comprise at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 30000, 50000, 75000, 100000, 500000, 1000000, 2000000, 3000000 or more loci. The error optimized loci may be distributed to at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 30000, 75000, 100000, 500000, 1000000, 2000000, 3000000 or more polynucleotides or genes.

[0259] The error rates can be achieved with or without error correction. The error rates can be achieved across the library, or across more than 80%, 85%, 90%, 93%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.8%, 99.9%, 99.95%, 99.98%, 99.99%, or more of the library.

[0260] Computer systems

[0261] Any of the systems described herein, may be operably linked to a computer and may be automated through a computer either locally or remotely. In various instances, the methods and systems of the disclosure may further comprise software programs on computer systems and use thereof. Accordingly, computerized control for the synchronization of the dispense/vacuum/refill functions such as orchestrating and synchronizing the material deposition device movement, dispense action and vacuum actuation are within the bounds of the disclosure. The computer systems may be programmed to interface between the user specified base sequence and the position of a material deposition device to deliver the correct reagents to specified regions of the substrate.

[0262] The computer system **1200** illustrated in **FIG. 12** may be understood as a logical apparatus that can read instructions from media **1211** and/or a network port **1205**, which can

optionally be connected to server **1209** having fixed media **1212**. The system, such as shown in **FIG. 12** can include a CPU **1201**, disk drives **1203**, optional input devices such as keyboard **1215** and/or mouse **1216** and optional monitor **1207**. Data communication can be achieved through the indicated communication medium to a server at a local or a remote location. The communication medium can include any means of transmitting and/or receiving data. For example, the communication medium can be a network connection, a wireless connection or an internet connection. Such a connection can provide for communication over the World Wide Web. It is envisioned that data relating to the present disclosure can be transmitted over such networks or connections for reception and/or review by a party **1222** as illustrated in **FIG. 12**.

[0263] **FIG. 13** is a block diagram illustrating a first example architecture of a computer system **1300** that can be used in connection with example instances of the present disclosure. As depicted in **FIG. 13**, the example computer system can include a processor **1302** for processing instructions. Non-limiting examples of processors include: Intel Xeon™ processor, AMD Opteron™ processor, Samsung 32-bit RISC ARM 1176JZ(F)-S v1.0™ processor, ARM Cortex-A8 Samsung S5PC100™ processor, ARM Cortex-A8 Apple A4™ processor, Marvell PXA 930™ processor, or a functionally-equivalent processor. Multiple threads of execution can be used for parallel processing. In some instances, multiple processors or processors with multiple cores can also be used, whether in a single computer system, in a cluster, or distributed across systems over a network comprising a plurality of computers, cell phones, and/or personal data assistant devices.

[0264] As illustrated in **FIG. 13**, a high speed cache **1304** can be connected to, or incorporated in, the processor **1302** to provide a high speed memory for instructions or data that have been recently, or are frequently, used by processor **1302**. The processor **1302** is connected to a north bridge **1306** by a processor bus **1308**. The north bridge **1306** is connected to random access memory (RAM) **1310** by a memory bus **1312** and manages access to the RAM **1310** by the processor **1302**. The north bridge **1306** is also connected to a south bridge **1314** by a chipset bus **1316**. The south bridge **1314** is, in turn, connected to a peripheral bus **1318**. The peripheral bus can be, for example, PCI, PCI-X, PCI Express, or other peripheral bus. The north bridge and south bridge are often referred to as a processor chipset and manage data transfer between the processor, RAM, and peripheral components on the peripheral bus **1318**. In some alternative architectures, the functionality of the north bridge can be incorporated into the processor instead of using a separate north bridge chip. In some instances, system **1300** can include an accelerator card **1322** attached to the peripheral bus **1318**. The accelerator can include field programmable gate arrays (FPGAs) or other hardware for accelerating certain processing. For example, an accelerator can be used for adaptive data restructuring or to evaluate algebraic expressions used in extended set processing.

[0265] Software and data are stored in external storage **1324** and can be loaded into RAM **1310** and/or cache **1304** for use by the processor. The system **1300** includes an operating system for managing system resources; non-limiting examples of operating systems include: Linux, WindowsTM, MACOSTM, BlackBerry OSTM, iOSTM, and other functionally-equivalent operating systems, as well as application software running on top of the operating system for managing data storage and optimization in accordance with example instances of the present disclosure. In this example, system **1300** also includes network interface cards (NICs) **1320** and **1321** connected to the peripheral bus for providing network interfaces to external storage, such as Network Attached Storage (NAS) and other computer systems that can be used for distributed parallel processing.

[0266] **FIG. 14** is a diagram showing a network **1400** with a plurality of computer systems **1402a**, and **1402b**, a plurality of cell phones and personal data assistants **1402c**, and Network Attached Storage (NAS) **1404a**, and **1404b**. In example instances, systems **1402a**, **1402b**, and **1402c** can manage data storage and optimize data access for data stored in Network Attached Storage (NAS) **1404a** and **1404b**. A mathematical model can be used for the data and be evaluated using distributed parallel processing across computer systems **1402a**, and **1402b**, and cell phone and personal data assistant systems **1402c**. Computer systems **1402a**, and **1402b**, and cell phone and personal data assistant systems **1402c** can also provide parallel processing for adaptive data restructuring of the data stored in Network Attached Storage (NAS) **1404a** and **1404b**. **FIG. 14** illustrates an example only, and a wide variety of other computer architectures and systems can be used in conjunction with the various instances of the present disclosure. For example, a blade server can be used to provide parallel processing. Processor blades can be connected through a back plane to provide parallel processing. Storage can also be connected to the back plane or as Network Attached Storage (NAS) through a separate network interface. In some example instances, processors can maintain separate memory spaces and transmit data through network interfaces, back plane or other connectors for parallel processing by other processors. In other instances, some or all of the processors can use a shared virtual address memory space.

[0267] **FIG. 15** is a block diagram of a multiprocessor computer system **1500** using a shared virtual address memory space in accordance with an example instance. The system includes a plurality of processors **1502a-f** that can access a shared memory subsystem **1504**. The system incorporates a plurality of programmable hardware memory algorithm processors (MAPs) **1506a-f** in the memory subsystem **1504**. Each MAP **1506a-f** can comprise a memory **1508a-f** and one or more field programmable gate arrays (FPGAs) **1510a-f**. The MAP provides a configurable functional unit and particular algorithms or portions of algorithms can be provided to the FPGAs

1510a-f for processing in close coordination with a respective processor. For example, the MAPs can be used to evaluate algebraic expressions regarding the data model and to perform adaptive data restructuring in example instances. In this example, each MAP is globally accessible by all of the processors for these purposes. In one configuration, each MAP can use Direct Memory Access (DMA) to access an associated memory **1508a-f**, allowing it to execute tasks independently of, and asynchronously from the respective microprocessor **1502a-f**. In this configuration, a MAP can feed results directly to another MAP for pipelining and parallel execution of algorithms.

[0268] The above computer architectures and systems are examples only, and a wide variety of other computer, cell phone, and personal data assistant architectures and systems can be used in connection with example instances, including systems using any combination of general processors, co-processors, FPGAs and other programmable logic devices, system on chips (SOCs), application specific integrated circuits (ASICs), and other processing and logic elements. In some instances, all or part of the computer system can be implemented in software or hardware. Any variety of data storage media can be used in connection with example instances, including random access memory, hard drives, flash memory, tape drives, disk arrays, Network Attached Storage (NAS) and other local or distributed data storage devices and systems.

[0269] In example instances, the computer system can be implemented using software modules executing on any of the above or other computer architectures and systems. In other instances, the functions of the system can be implemented partially or completely in firmware, programmable logic devices such as field programmable gate arrays (FPGAs) as referenced in **FIG. 15**, system on chips (SOCs), application specific integrated circuits (ASICs), or other processing and logic elements. For example, the Set Processor and Optimizer can be implemented with hardware acceleration through the use of a hardware accelerator card, such as accelerator card **1322** illustrated in **FIG. 13**.

[0270] Embodiments

[0271] Provided herein are polynucleotide libraries comprising: a first polynucleotide library comprising at least 30,000 polynucleotides, wherein each of the at least 30,000 polynucleotides is present in an amount such that, following hybridization with genomic fragments and sequencing of the hybridized genomic fragments, the polynucleotide library provides for at least 25 fold read depth of at least 80 percent of the bases of a first set of hybridized genomic fragments and at least 40 fold average read depth; and a second polynucleotide library comprising at least 1500 polynucleotides, wherein each of the at least 1500 polynucleotides is present in an amount such that, following hybridization with genomic fragments and sequencing of the hybridized genomic fragments, the polynucleotide library provides for at least 15 fold read depth of at least 80 percent

of the bases of a second set of hybridized genomic fragments and at least 24 fold average read depth. Further provided herein are polynucleotide libraries, wherein the first polynucleotide library comprises at least 100,000 polynucleotides. Further provided herein are polynucleotide libraries wherein the second polynucleotide library comprises at least 5,000 polynucleotides. Further provided herein are polynucleotide libraries wherein the first polynucleotide library comprises at least 100,000 polynucleotides and the second polynucleotide library comprises at least 5,000 polynucleotides. Further provided herein are polynucleotide libraries wherein the first polynucleotide library provides for at least 25 fold read depth of at least 90 percent of the bases of the first set of hybridized genomic fragments and at least 40 fold average read depth. Further provided herein are polynucleotide libraries wherein the first polynucleotide library provides for at least 40 fold read depth of at least 80 percent of the bases of the first set of hybridized genomic fragments and at least 50 fold average read depth. Further provided herein are polynucleotide libraries wherein the second polynucleotide library provides for at least 15 fold read depth of at least 90 percent of the bases of the second set of hybridized genomic fragments and at least 24 fold average read depth. Further provided herein are polynucleotide libraries wherein the second polynucleotide library provides for at least 20 fold read depth of at least 80 percent of the bases of the second set of hybridized genomic fragments and at least 30 fold average read depth. Further provided herein are polynucleotide libraries wherein at least 90% of the bases sequenced are at least 99.5% correct. Further provided herein are polynucleotide libraries wherein at least 90% of the bases sequenced are at least 99.9% correct. Further provided herein are polynucleotide libraries wherein at least 90% of the bases sequenced are at least 99.95% correct. Further provided herein are polynucleotide libraries wherein each of the genomic fragments is about 100 bases to about 500 bases in length. Further provided herein are polynucleotide libraries wherein the at least 30,000 polynucleotides encode for at least 1000 genes. Further provided herein are polynucleotide libraries wherein the at least 30,000 polynucleotides encode for at least one exon sequence. Further provided herein are polynucleotide libraries wherein the at least 1500 polynucleotides encode for at least one exon sequence. Further provided herein are polynucleotide libraries wherein the at least 1500 polynucleotides encode for at least 10 genes. Further provided herein are polynucleotide libraries wherein the at least 1500 polynucleotides encode for at least 100 genes. Further provided herein are polynucleotide libraries wherein the at least 1500 polynucleotides encode for at least one intron. Further provided herein are polynucleotide libraries wherein the at least 1500 polynucleotides encode for at least one single nucleotide polymorphism (SNP). Further provided herein are polynucleotide libraries wherein the single nucleotide polymorphism (SNP) is heterozygous.

[0272] Provided herein are methods for sequencing genomic DNA, comprising: contacting the first library and the second library of the polynucleotide libraries described herein with a plurality of genomic fragments; enriching at least one genomic fragment that binds to the first library or the second library to generate at least one enriched target polynucleotide; and sequencing the at least one enriched target polynucleotide.

[0273] Provided herein are methods for sequencing genomic DNA, comprising: contacting a composition comprising a first polynucleotide library of the polynucleotide libraries described herein with a plurality of genomic fragments; enriching at least one genomic fragment that binds to the first polynucleotide library to generate at least one enriched target polynucleotide; sequencing the at least one enriched target polynucleotide; identifying one or more positions of the at least one enriched polynucleotide having less than average read depth; repeating steps a-c, wherein the second polynucleotide library of the polynucleotide libraries described herein is added to the composition, wherein the second polynucleotide library comprises at least one polynucleotide that binds to genomic fragments comprising the one or more positions having less than average read depth, wherein the presence of the second polynucleotide library increases the read depth at the one or more positions having less than average read depth. Further provided herein are methods wherein the first polynucleotide library and the second polynucleotide library do not comprise any common sequences. Further provided herein are methods wherein the first polynucleotide library and the second polynucleotide library comprise at least one common sequence. Further provided herein are methods wherein the presence of the second polynucleotide library increases the read depth at the one or more positions of the least one enriched target polynucleotide having less than average read depth by at least 10 fold. Further provided herein are methods wherein the presence of the second polynucleotide library increases the read depth at the one or more positions of the at least one enriched target polynucleotide having less than average read depth by at least 100 fold.

[0274] Provided herein are polynucleotide libraries, the polynucleotide library comprising at least 1500 polynucleotides, wherein less than all polynucleotides comprises a molecular tag, wherein each of the at least 5000 polynucleotides are present in an amount such that, following hybridization with genomic fragments and sequencing of the hybridized genomic fragments, the polynucleotide library provides for at least 30 fold read depth of at least 90 percent of the bases of the hybridized genomic fragments under conditions wherein the total number of reads is no more than 55 fold higher than the total number of bases of the hybridized genomic fragments. Further provided herein are polynucleotide libraries wherein no more than 90% of the polynucleotides comprise a molecular tag. Further provided herein are polynucleotide libraries wherein no more than 80% of the polynucleotides comprise a molecular tag. Further provided herein are

polynucleotide libraries wherein no more than 50% of the polynucleotides comprise a molecular tag. Further provided herein are polynucleotide libraries wherein no more than 25% of the polynucleotides comprise a molecular tag. Further provided herein are polynucleotide libraries wherein the molecular tag is biotin. Further provided herein are polynucleotide libraries wherein the at least 5000 polynucleotides encode for at least 5000 genes. Further provided herein are polynucleotide libraries wherein the polynucleotide library comprises at least 30,000 polynucleotides. Further provided herein are polynucleotide libraries wherein the polynucleotide library comprises at least 100,000 polynucleotides.

[0275] Provided herein are methods for enriching nucleic acids comprising: contacting the polynucleotide library described herein with a plurality of genomic fragments; enriching at least one genomic fragment that binds to the polynucleotide library to generate at least one enriched target polynucleotide; and sequencing the at least one enriched target polynucleotide. Further provided herein are methods wherein the polynucleotide library provides for at least 90 percent unique reads for the bases of the enriched target polynucleotide after sequencing. Further provided herein are methods wherein the polynucleotide library provides for at least 95 percent unique reads for the bases of the enriched target polynucleotide after sequencing. Further provided herein are methods wherein the polynucleotide library provides for at least 80 percent of the bases of the enriched target polynucleotide having a read depth within about 1.5 times the mean read depth. Further provided herein are methods wherein the polynucleotide library provides for at least 90 percent of the bases of the enriched target polynucleotide having a read depth within about 1.5 times the mean read depth.

[0276] Provided herein are polynucleotide libraries, the polynucleotide library comprising at least 5000 polynucleotides, wherein each of the at least 5000 polynucleotides is present in an amount such that, following hybridization with a composition comprising i) a genomic library, wherein the genomic library comprises polynucleotides each comprising genomic fragments, at least one index sequence, and at least one adapter; and ii) at least one polynucleotide blocker, wherein the polynucleotide blocker is complementary to at least a portion of the adapter sequence, but not complementary to the at least one index sequence; and sequencing of the hybridized genomic fragments, the polynucleotide library provides for at least 30 fold read depth of at least 90 percent of the bases of the genomic fragments under conditions wherein the total number of reads is no more than 55 fold higher than the total number of bases of the hybridized genomic fragments. Further provided herein are polynucleotide libraries wherein the composition comprises no more than four polynucleotide blockers. Further provided herein are polynucleotide libraries wherein the polynucleotide blocker comprises one or more nucleotide analogues.

[0277] Further provided herein are polynucleotide libraries wherein the polynucleotide blocker comprises one or more locked nucleic acids (LNAs). Further provided herein are polynucleotide libraries wherein the polynucleotide blocker comprises one or more bridged nucleic acids (BNAs). Further provided herein are polynucleotide libraries wherein the polynucleotide blocker comprises at least 2 nucleotide analogues. Further provided herein are polynucleotide libraries wherein the polynucleotide blocker comprises at least 5 nucleotide analogues. Further provided herein are polynucleotide libraries wherein the polynucleotide blocker comprises at least 10 nucleotide analogues. Further provided herein are polynucleotide libraries wherein the polynucleotide blocker has a T_m of at least 70 degrees C. Further provided herein are polynucleotide libraries wherein the polynucleotide blocker has a T_m of at least 75 degrees C. Further provided herein are polynucleotide libraries wherein the polynucleotide blocker has a T_m of at least 80 degrees C. Further provided herein are polynucleotide libraries wherein the genomic library comprises genomic fragments from at least 2 different samples. Further provided herein are polynucleotide libraries wherein the genomic library comprises genomic fragments from at least 10 different samples. Further provided herein are polynucleotide libraries wherein the genomic library comprises genomic fragments from at least 2 non-identical index sequences. Further provided herein are polynucleotide libraries wherein the genomic library comprises genomic fragments from at least 16 non-identical index sequences. Further provided herein are polynucleotide libraries wherein the genomic library comprises genomic fragments further comprising at least one unique molecular identifier (UMI).

[0278] Provided herein are methods for enriching nucleic acids comprising: contacting the polynucleotide libraries described herein with a plurality of genomic fragments; enriching at least one genomic fragment that binds to the polynucleotide library to generate at least one enriched target polynucleotide; and sequencing the at least one enriched target polynucleotide. Further provided herein are methods wherein the off-target rate is less than 25%. Further provided herein are methods wherein the off-target rate is less than 20%. Further provided herein are methods wherein the molar ratio between at least one polynucleotide blocker and the complementary adapter is no more than 5:1. Further provided herein are methods wherein the molar ratio between at least one polynucleotide blocker and the complementary adapter is no more than 2:1. Further provided herein are methods wherein the molar ratio between at least one polynucleotide blocker and the complementary adapter is no more than 1.5:1.

[0279] Provided herein are compositions for nucleic acid hybridization comprising: a first polynucleotide library; a second polynucleotide library, wherein at least one polynucleotide in the first library is at least partially complimentary to at least one polynucleotide of the second library;

and an additive, wherein the additive reduces off-target hybridization of the at least one polynucleotide of the first library with the at least one polynucleotide of the second library by decreasing a local concentration of the first polynucleotide library or the second polynucleotide library at an air-liquid interface. Further provided herein are compositions wherein the additive is mineral oil, a nucleotide triphosphate, polyether, or urea. Further provided herein are compositions wherein the additive is a hydrocarbon comprising at least six carbon atoms. Further provided herein are compositions wherein the additive is silicon oil. Further provided herein are compositions wherein the oil is derived from plant sources. Further provided herein are compositions wherein the composition further comprises dimethyl sulfoxide. Further provided herein are compositions wherein the composition does not comprise a formamide. Further provided herein are compositions wherein the size of the first polynucleotide library is less than 10 million bases. Further provided herein are compositions wherein the size of the first polynucleotide library is less than 1 million bases. Further provided herein are compositions wherein the size of the first polynucleotide library is less than 0.5 million bases. Further provided herein are compositions wherein the first polynucleotide library comprises at least one exon sequence. Further provided herein are compositions wherein first polynucleotide library comprises polynucleotides encoding for at least 10 genes. Further provided herein are compositions wherein the first polynucleotide library comprises polynucleotides encoding for at least 100 genes. Further provided herein are compositions wherein the first polynucleotide library comprises at least one genomic fragment. Further provided herein are compositions wherein the first polynucleotide library comprises RNA, DNA, cDNA, or genomic DNA. Further provided herein are compositions wherein the first polynucleotide library comprises genomic DNA.

[0280] Provided herein are compositions for nucleic acid hybridization comprising: a first polynucleotide library and a second polynucleotide library each comprising a plurality of polynucleotides, wherein at least one polynucleotide in the first library is at least partially complimentary to at least one polynucleotide of the second library; and an oil, wherein the oil reduces off-target hybridization of the at least one polynucleotide of the first library with the at least one polynucleotide of the second library by decreasing a local concentration of the first polynucleotide library or the second polynucleotide library at an air-liquid interface. Further provided herein are compositions wherein the additive is mineral oil, a nucleotide triphosphate, polyether, or urea. Further provided herein are compositions wherein the additive is a hydrocarbon comprising at least six carbon atoms. Further provided herein are compositions wherein the additive is silicon oil. Further provided herein are compositions wherein the oil is derived from plant sources. Further provided herein are compositions wherein the composition further comprises

dimethyl sulfoxide. Further provided herein are compositions wherein the composition does not comprise a formamide. Further provided herein are compositions wherein the size of the first polynucleotide library is less than 10 million bases. Further provided herein are compositions wherein the size of the first polynucleotide library is less than 1 million bases. Further provided herein are compositions wherein the size of the first polynucleotide library is less than 0.5 million bases. Further provided herein are compositions wherein first polynucleotide library comprises at least one exon sequence. Further provided herein are compositions wherein first polynucleotide library comprises polynucleotides encoding for at least 10 genes. Further provided herein are compositions wherein first polynucleotide library comprises polynucleotides encoding for at least 100 genes. Further provided herein are compositions wherein the first polynucleotide library comprises at least one genomic fragment. Further provided herein are compositions wherein the first polynucleotide library comprises RNA, DNA, cDNA, or genomic DNA. Further provided herein are compositions wherein the first polynucleotide library comprises genomic DNA.

[0281] Provided herein are methods for reducing off-target nucleic acid hybridization, comprising: contacting a first polynucleotide library with a second polynucleotide library, wherein the first polynucleotide library and the second polynucleotide library each comprise a plurality of polynucleotides, and wherein at least one polynucleotide in the first library is at least partially complimentary to at least one polynucleotide in the second library; enriching at least one genomic fragment that binds to the second polynucleotide library to generate at least one enriched target polynucleotide, wherein enriching comprises at least one aspiration step, and wherein the at least one aspiration step comprises aspirating only liquid from the area near the air/liquid interface; and sequencing the at least one enriched target polynucleotide. Further provided herein are methods wherein the additive is oil, a nucleotide triphosphate, polyether, or urea. Further provided herein are methods wherein the additive is mineral oil. Further provided herein are methods wherein the presence of the additive decreases off-target binding. Further provided herein are methods wherein the presence of the additive decreases off-target binding by at least 10%. Further provided herein are methods wherein the presence of the additive decreases off-target binding by at least 20%. Further provided herein are methods wherein the presence of the additive decreases off-target binding by at least 30%. Further provided herein are methods wherein the off-target binding is random off-target binding. Further provided herein are methods wherein the size of the first polynucleotide library is less than 10 million bases. Further provided herein are methods wherein the size of the first polynucleotide library is less than 1 million bases. Further provided herein are methods wherein the size of the first polynucleotide library is less than 0.5 million bases. Further provided herein are methods wherein first polynucleotide library comprises at least one exon

sequence. Further provided herein are methods wherein first polynucleotide library comprises polynucleotides encoding for at least 10 genes. Further provided herein are methods wherein first polynucleotide library comprises polynucleotides encoding for at least 100 genes. Further provided herein are methods wherein the first polynucleotide library comprises at least one genomic fragment. Further provided herein are methods wherein the first polynucleotide library comprises RNA, DNA, cDNA, or genomic DNA. Further provided herein are methods wherein the first polynucleotide library comprises genomic DNA.

[0282] Provided herein are methods for sequencing genomic DNA, comprising: contacting a polynucleotide library with a plurality of genomic fragments and an additive to form a mixture, wherein the additive decreases a local concentration of the polynucleotide library or the genomic fragments in the mixture at an air-liquid interface; enriching at least one genomic fragment that binds to the polynucleotide library to generate at least one enriched target polynucleotide; and sequencing the at least one enriched target polynucleotide. Further provided herein are methods wherein the additive is oil, a nucleotide triphosphate, polyether, or urea. Further provided herein are methods wherein the additive is mineral oil. Further provided herein are methods wherein the presence of the additive decreases off-target binding. Further provided herein are methods wherein the presence of the additive decreases off-target binding by at least 10%. Further provided herein are methods wherein the presence of the additive decreases off-target binding by at least 20%. Further provided herein are methods wherein the presence of the additive decreases off-target binding by at least 30%. Further provided herein are methods wherein the off-target binding is random off-target binding. Further provided herein are methods wherein the size of the first polynucleotide library is less than 10 million bases. Further provided herein are methods wherein the size of the first polynucleotide library is less than 1 million bases. Further provided herein are methods wherein the size of the first polynucleotide library is less than 0.5 million bases. Further provided herein are methods wherein the first polynucleotide library comprises at least one exon sequence. Further provided herein are methods wherein the first polynucleotide library comprises polynucleotides encoding for at least 10 genes. Further provided herein are methods wherein the first polynucleotide library comprises polynucleotides encoding for at least 100 genes. Further provided herein are methods wherein the first polynucleotide library comprises at least one genomic fragment. Further provided herein are methods wherein the first polynucleotide library comprises RNA, DNA, cDNA, or genomic DNA. Further provided herein are methods wherein the first polynucleotide library comprises genomic DNA.

EXAMPLES

[0283] The following examples are given for the purpose of illustrating various embodiments of the invention and are not meant to limit the present invention in any fashion. The present examples, along with the methods described herein are presently representative of preferred embodiments, are exemplary, and are not intended as limitations on the scope of the invention. Changes therein and other uses which are encompassed within the spirit of the invention as defined by the scope of the claims will occur to those skilled in the art.

[0284] Example 1: Functionalization of a substrate surface

[0285] A substrate was functionalized to support the attachment and synthesis of a library of polynucleotides. The substrate surface was first wet cleaned using a piranha solution comprising 90% H₂SO₄ and 10% H₂O₂ for 20 minutes. The substrate was rinsed in several beakers with DI water, held under a DI water gooseneck faucet for 5 minutes, and dried with N₂. The substrate was subsequently soaked in NH₄OH (1:100; 3 mL:300 mL) for 5 minutes, rinsed with DI water using a handgun, soaked in three successive beakers with DI water for 1 minute each, and then rinsed again with DI water using the handgun. The substrate was then plasma cleaned by exposing the substrate surface to O₂. A SAMCO PC-300 instrument was used to plasma etch O₂ at 250 watts for 1 minute in downstream mode.

[0286] The cleaned substrate surface was actively functionalized with a solution comprising N-(3-triethoxysilylpropyl)-4-hydroxybutyramide using a YES-1224P vapor deposition oven system with the following parameters: 0.5 to 1 torr, 60 minutes, 70 °C, 135 °C vaporizer. The substrate surface was resist coated using a Brewer Science 200X spin coater. SPR™ 3612 photoresist was spin coated on the substrate at 2500 rpm for 40 seconds. The substrate was pre-baked for 30 minutes at 90 °C on a Brewer hot plate. The substrate was subjected to photolithography using a Karl Suss MA6 mask aligner instrument. The substrate was exposed for 2.2 seconds and developed for 1 minute in MSF 26A. Remaining developer was rinsed with the handgun and the substrate soaked in water for 5 minutes. The substrate was baked for 30 minutes at 100 °C in the oven, followed by visual inspection for lithography defects using a Nikon L200. A descum process was used to remove residual resist using the SAMCO PC-300 instrument to O₂ plasma etch at 250 watts for 1 minute.

[0287] The substrate surface was passively functionalized with a 100 µL solution of perfluorooctyltrichlorosilane mixed with 10 µL light mineral oil. The substrate was placed in a chamber, pumped for 10 minutes, and then the valve was closed to the pump and left to stand for 10 minutes. The chamber was vented to air. The substrate was resist stripped by performing two soaks for 5 minutes in 500 mL NMP at 70 °C with ultrasonication at maximum power (9 on Crest

system). The substrate was then soaked for 5 minutes in 500 mL isopropanol at room temperature with ultrasonication at maximum power. The substrate was dipped in 300 mL of 200 proof ethanol and blown dry with N₂. The functionalized surface was activated to serve as a support for polynucleotide synthesis.

[0288] Example 2: Synthesis of a 50-mer sequence on a polynucleotide synthesis device

[0289] A two dimensional polynucleotide synthesis device was assembled into a flowcell, which was connected to a flowcell (Applied Biosystems (ABI394 DNA Synthesizer"). The polynucleotide synthesis device was uniformly functionalized with N-(3-TRIETHOXYSILYLPROPYL)-4-HYDROXYBUTYRAMIDE (Gelest) was used to synthesize an exemplary polynucleotide of 50 bp ("50-mer polynucleotide") using polynucleotide synthesis methods described herein.

[0290] The sequence of the 50-mer was as described in SEQ ID NO.: 1.

5'AGACAATCAACCATTGTTGGGGTGGACAGCCTTGACCTCTAGACTTCGGCAT##TTTTTTT TTT3' (SEQ ID NO.: 1), where # denotes Thymidine-succinyl hexamide CED phosphoramidite (CLP-2244 from ChemGenes), which is a cleavable linker enabling the release of polynucleotides from the surface during deprotection.

[0291] The synthesis was done using standard DNA synthesis chemistry (coupling, capping, oxidation, and deblocking) according to the protocol in **Table 3** and an ABI synthesizer.

Table 3

General DNA Synthesis Process Name	Process Step	Time (seconds)
WASH (Acetonitrile Wash Flow)	Acetonitrile System Flush	4
	Acetonitrile to Flowcell	23
	N ₂ System Flush	4
	Acetonitrile System Flush	4
DNA BASE ADDITION (Phosphoramidite + Activator Flow)	Activator Manifold Flush	2
	Activator to Flowcell	6
	Activator + Phosphoramidite to Flowcell	6
	Activator to Flowcell	0.5
	Activator + Phosphoramidite to Flowcell	5
	Activator to Flowcell	0.5
	Activator + Phosphoramidite to Flowcell	5
	Activator to Flowcell	0.5
	Activator +	5

General DNA Synthesis Process Name	Process Step	Time (seconds)
	Phosphoramidite to Flowcell	
	Incubate for 25sec	25
WASH (Acetonitrile Wash Flow)	Acetonitrile System Flush	4
	Acetonitrile to Flowcell	15
	N2 System Flush	4
	Acetonitrile System Flush	4
DNA BASE ADDITION (Phosphoramidite + Activator Flow)	Activator Manifold Flush	2
	Activator to Flowcell	5
	Activator + Phosphoramidite to Flowcell	18
	Incubate for 25sec	25
WASH (Acetonitrile Wash Flow)	Acetonitrile System Flush	4
	Acetonitrile to Flowcell	15
	N2 System Flush	4
	Acetonitrile System Flush	4
CAPPING (CapA+B, 1:1, Flow)	CapA+B to Flowcell	15
WASH (Acetonitrile Wash Flow)	Acetonitrile System Flush	4
	Acetonitrile to Flowcell	15
	Acetonitrile System Flush	4
OXIDATION (Oxidizer Flow)	Oxidizer to Flowcell	18
WASH (Acetonitrile Wash Flow)	Acetonitrile System Flush	4
	N2 System Flush	4
	Acetonitrile System Flush	4
	Acetonitrile to Flowcell	15
	Acetonitrile System Flush	4
	Acetonitrile to Flowcell	15
	N2 System Flush	4
	Acetonitrile System Flush	4
	Acetonitrile to Flowcell	23
	N2 System Flush	4
	Acetonitrile System Flush	4
DEBLOCKING (Deblock Flow)	Deblock to Flowcell	36
WASH (Acetonitrile Wash Flow)	Acetonitrile System Flush	4
	N2 System Flush	4
	Acetonitrile System Flush	4
	Acetonitrile to Flowcell	18
	N2 System Flush	4.13
	Acetonitrile System Flush	4.13
	Acetonitrile to Flowcell	15

[0292] The phosphoramidite/activator combination was delivered similar to the delivery of bulk reagents through the flowcell. No drying steps were performed as the environment stays "wet" with reagent the entire time.

[0293] The flow restrictor was removed from the ABI 394 synthesizer to enable faster flow. Without flow restrictor, flow rates for amidites (0.1M in ACN), Activator, (0.25M Benzoylthiotetrazole ("BTT"; 30-3070-xx from GlenResearch) in ACN), and Ox (0.02M I₂ in 20% pyridine, 10% water, and 70% THF) were roughly ~100uL/second, for acetonitrile ("ACN") and capping reagents (1:1 mix of CapA and CapB, wherein CapA is acetic anhydride in THF/Pyridine and CapB is 16% 1-methylimidazole in THF), roughly ~200uL/second, and for Deblock (3% dichloroacetic acid in toluene), roughly ~300uL/second (compared to ~50uL/second for all reagents with flow restrictor). The time to completely push out Oxidizer was observed, the timing for chemical flow times was adjusted accordingly and an extra ACN wash was introduced between different chemicals. After polynucleotide synthesis, the chip was deprotected in gaseous ammonia overnight at 75 psi. Five drops of water were applied to the surface to recover polynucleotides. The recovered polynucleotides were then analyzed on a BioAnalyzer small RNA chip (data not shown).

[0294] **Example 3: Synthesis of a 100-mer sequence on a polynucleotide synthesis device**

[0295] The same process as described in **Example 2** for the synthesis of the 50-mer sequence was used for the synthesis of a 100-mer polynucleotide ("100-mer polynucleotide"; 5' CGGGATCCTTATCGTCATCGTCGTACAGATCCCGACCCATTTGCTGTCCACCAGTCATG CTAGCCATACCATGATGATGATGATGATGAGAACCCCGCAT##TTTTTTTTTT3', where # denotes Thymidine-succinyl hexamide CED phosphoramidite (CLP-2244 from ChemGenes); SEQ ID NO.: 2) on two different silicon chips, the first one uniformly functionalized with N-(3-TRIETHOXSILYL)PROPYL)-4-HYDROXYBUTYRAMIDE and the second one functionalized with 5/95 mix of 11-acetoxyundecyltriethoxysilane and n-decyltriethoxysilane, and the polynucleotides extracted from the surface were analyzed on a BioAnalyzer instrument (data not shown).

[0296] All ten samples from the two chips were further PCR amplified using a forward (5'ATGCGGGGTTCTCATCATC3'; SEQ ID NO.: 3) and a reverse (5'CGGGATCCTTATCGTCATCG3'; SEQ ID NO.: 4) primer in a 50uL PCR mix (25uL NEB Q5 master mix, 2.5uL 10uM Forward primer, 2.5uL 10uM Reverse primer, 1uL polynucleotide extracted from the surface, and water up to 50uL) using the following thermal cycling program:

98 C, 30 seconds

98 C, 10 seconds; 63C, 10 seconds; 72C, 10 seconds; repeat 12 cycles

72C, 2 minutes

[0297] The PCR products were also run on a BioAnalyzer (data not shown), demonstrating sharp peaks at the 100-mer position. Next, the PCR amplified samples were cloned, and Sanger sequenced. **Table 4** summarizes the results from the Sanger sequencing for samples taken from spots 1-5 from chip 1 and for samples taken from spots 6-10 from chip 2.

Table 4

Spot	Error rate	Cycle efficiency
1	1/763 bp	99.87%
2	1/824 bp	99.88%
3	1/780 bp	99.87%
4	1/429 bp	99.77%
5	1/1525 bp	99.93%
6	1/1615 bp	99.94%
7	1/531 bp	99.81%
8	1/1769 bp	99.94%
9	1/854 bp	99.88%
10	1/1451 bp	99.93%

[0298] Thus, the high quality and uniformity of the synthesized polynucleotides were repeated on two chips with different surface chemistries. Overall, 89%, corresponding to 233 out of 262 of the 100-mers that were sequenced were perfect sequences with no errors.

[0299] Finally, **Table 5** summarizes error characteristics for the sequences obtained from the polynucleotides samples from spots 1-10.

Table 5

Sample ID/Spot no.	OSA_00 46/1	OSA_00 47/2	OSA_00 48/3	OSA_00 49/4	OSA_00 50/5	OSA_00 51/6	OSA_00 52/7	OSA_00 53/8	OSA_00 54/9	OSA_005 5/10
Total Sequences	32	32	32	32	32	32	32	32	32	32
Sequencing Quality	25 of 28	27 of 27	26 of 30	21 of 23	25 of 26	29 of 30	27 of 31	29 of 31	28 of 29	25 of 28
Oligo Quality	23 of 25	25 of 27	22 of 26	18 of 21	24 of 25	25 of 29	22 of 27	28 of 29	26 of 28	20 of 25
ROI Match Count	2500	2698	2561	2122	2499	2666	2625	2899	2798	2348
ROI Mutation	2	2	1	3	1	0	2	1	2	1
ROI Multi Base Deletion	0	0	0	0	0	0	0	0	0	0
ROI Small Insertion	1	0	0	0	0	0	0	0	0	0

ROI Single Base Deletion	0	0	0	0	0	0	0	0	0	0
Large Deletion Count	0	0	1	0	0	1	1	0	0	0
Mutation: G>A	2	2	1	2	1	0	2	1	2	1
Mutation: T>C	0	0	0	1	0	0	0	0	0	0
ROI Error Count	3	2	2	3	1	1	3	1	2	1
ROI Error Rate	Err: ~1 in 834	Err: ~1 in 1350	Err: ~1 in 1282	Err: ~1 in 708	Err: ~1 in 2500	Err: ~1 in 2667	Err: ~1 in 876	Err: ~1 in 2900	Err: ~1 in 1400	Err: ~1 in 2349
ROI Minus Primer Error Rate	MP Err: ~1 in 763	MP Err: ~1 in 824	MP Err: ~1 in 780	MP Err: ~1 in 429	MP Err: ~1 in 1525	MP Err: ~1 in 1615	MP Err: ~1 in 531	MP Err: ~1 in 1769	MP Err: ~1 in 854	MP Err: ~1 in 1451

[0300] Example 4: Parallel assembly of 29,040 unique polynucleotides

[0301] A structure comprising 256 clusters **1605** each comprising 121 loci on a flat silicon plate **1601** was manufactured as shown in **FIG. 16**. An expanded view of a cluster is shown in **1610** with 121 loci. Loci from 240 of the 256 clusters provided an attachment and support for the synthesis of polynucleotides having distinct sequences. Polynucleotide synthesis was performed by phosphoramidite chemistry using general methods from **Example 3**. Loci from 16 of the 256 clusters were control clusters. The global distribution of the 29,040 unique polynucleotides synthesized (240 x 121) is shown in **FIG. 17A**. Polynucleotide libraries were synthesized at high uniformity. 90% of sequences were present at signals within 4x of the mean, allowing for 100% representation. Distribution was measured for each cluster, as shown in **FIG. 17B**. The distribution of unique polynucleotides synthesized in 4 representative clusters is shown in **FIG. 18**. On a global level, all polynucleotides in the run were present and 99% of the polynucleotides had abundance that was within 2x of the mean indicating synthesis uniformity. This same observation was consistent on a per-cluster level.

[0302] The error rate for each polynucleotide was determined using an Illumina MiSeq gene sequencer. The error rate distribution for the 29,040 unique polynucleotides is shown in **FIG. 19A** and averages around 1 in 500 bases, with some error rates as low as 1 in 800 bases. Distribution was measured for each cluster, as shown in **FIG. 19B**. The error rate distribution for unique polynucleotides in four representative clusters is shown in **FIG. 20**. The library of 29,040 unique polynucleotides was synthesized in less than 20 hours.

[0303] Analysis of GC percentage versus polynucleotide representation across all of the 29,040 unique polynucleotides showed that synthesis was uniform despite GC content, **FIG. 21**.

[0304] Example 5. Use of a controlled stoichiometry polynucleotide library for exome targeting with Next Generation Sequencing (NGS)

[0305] A first polynucleotide cDNA targeting library (probe library), comprising up to 370,000 or more non-identical polynucleotides which overlap with one or more gene exons is designed and synthesized on a structure by phosphoramidite chemistry using the general methods from Example 3. The polynucleotides are ligated to a molecular tag such as biotin using PCR (or directly during solid-phase synthesis) to form a probe for subsequent capture of the target exons of interest. The probes are hybridized to sequences in a library of genomic nucleic acids, and separated from non-binding sequences. Unbound probes are washed away, leaving the target library enriched in cDNA sequences. The enriched library is then sequenced using NGS, and reads for each expected gene are measured as a function of the cDNA probe(s) used to target the gene.

[0306] A target sequence's frequency of reads is affected by target sequence abundance, probe binding, secondary structure, or other factors which decrease representation after sequencing of the target sequence despite enrichment. Polynucleotide library stoichiometric control is performed by modifying the stoichiometry of the first polynucleotide cDNA targeting library to obtain a second polynucleotide cDNA targeting library, with increased stoichiometry for polynucleotide probe sequences that lead to fewer reads. This second cDNA targeting library is designed and synthesized on a structure by phosphoramidite chemistry using the general methods from Example 3, and used to enrich sequence exons of the target genomic DNA library as described previously.

[0307] Example 6: Genomic DNA capture with an exome-targeting polynucleotide probe library

[0308] A polynucleotide targeting library comprising at least 500,000 non-identical polynucleotides targeting the human exome was designed and synthesized on a structure by phosphoramidite chemistry using the general methods from Example 3, and the stoichiometry controlled using the general methods of Example 5 to generate Library 4. The polynucleotides were then labeled with biotin, and then dissolved to form an exome probe library solution. A dried indexed library pool was obtained from a genomic DNA (gDNA) sample using the general methods of Example 16.

[0309] The exome probe library solution, a hybridization solution, a blocker mix A, and a blocker mix B were mixed by pulse vortexing for 2 seconds. The hybridization solution was heated at 65°C for 10 minutes, or until all precipitate was dissolved, and then brought to room temperature on the benchtop for 5 additional minutes. 20 µL of hybridization solution and 4 µL of the exome probe library solution were added to a thin-walled PCR 0.2 mL strip-tube and mixed gently by pipetting. The combined hybridization solution/exome probe solution was heated to 95°C for 2 minutes in a thermal cycler with a 105°C lid and immediately cooled on ice for at least 10 minutes. The solution was then allowed to cool to room temperature on the benchtop for 5 minutes. While

the hybridization solution/exome probe library solution was cooling, water was added to 9 μ L for each genomic DNA sample, and 5 μ L of blocker mix A, and 2 μ L of blocker mix B were added to the dried indexed library pool in the thin-walled PCR 0.2 mL strip-tube. The solution was then mixed by gentle pipetting. The pooled library/blocker tube was heated at 95°C for 5 minutes in a thermal cycler with a 105°C lid, then brought to room temperature on the benchtop for no more than 5 minutes before proceeding onto the next step. The hybridization mix/probe solution was mixed by pipetting and added to the entire 24 μ L of the pooled library/blocker tube. The entire capture reaction well was mixed by gentle pipetting, to avoid generating bubbles. The sample tube was pulse-spun to make sure the tube was sealed tightly. The capture/hybridization reaction was heated at 70 °C for 16 hours in a PCR thermocycler, with a lid temperature of 85 °C.

[0310] Binding buffer, wash Buffer 1 and wash Buffer 2 were heated at 48°C until all precipitate was dissolved into solution. 700 μ L of wash buffer 2 was aliquoted per capture and preheated to 48°C. Streptavidin binding beads and DNA purification beads were equilibrated at room temperature for at least 30 minutes. A polymerase, such as KAPA HiFi HotStart ReadyMix and amplification primers were thawed on ice. Once the reagents were thawed, they were mixed by pulse vortexing for 2 seconds. 500 μ L of 80 percent ethanol per capture reaction was prepared. Streptavidin binding beads were pre-equilibrated at room temperature and vortexed until homogenized. 100 μ L of streptavidin binding beads were added to a clean 1.5 mL microcentrifuge tube per capture reaction. 200 μ L of binding buffer was added to each tube and each tube was mixed by pipetting until homogenized. The tube was placed on magnetic stand. Streptavidin binding beads were pelleted within 1 minute. The tube was removed and the clear supernatant was discarded, making sure not to disturb the bead pellet. The tube was removed from the magnetic stand, and the washes were repeated two additional times. After the third wash, the tube was removed and the clear supernatant was discarded. A final 200 μ L of binding buffer was added, and beads were resuspended by vortexing until homogeneous.

[0311] After completing the hybridization reaction, the thermal cycler lid was opened and the full volume of capture reaction was quickly transferred (36-40 μ L) into the washed streptavidin binding beads. The mixture was mixed for 30 minutes at room temperature on a shaker, rocker, or rotator at a speed sufficient to keep capture reaction/streptavidin binding bead solution homogenized. The capture reaction/streptavidin binding bead solution was removed from mixer and pulse-spun to ensure all solution was at the bottom of the tube. The sample was placed on a magnetic stand, and streptavidin binding beads pelleted, leaving a clear supernatant within 1 minute. The clear supernatant was removed and discarded. The tube was removed from the magnetic stand and 200 μ L of wash buffer was added at room temperature, followed by mixing by

pipetting until homogenized. The tube was pulse-spun to ensure all solution was at the bottom of the tube. A thermal cycler was programmed with the following conditions (**Table 6**).

[0312] The temperature of the heated lid was set to 105°C.

Table 6

Step	Temperature	Time	Cycle Number
1	98°C	45 seconds	1
2	98°C	15 seconds	9
	60°C	30 seconds	
	72°C	30 seconds	
3	72°C	1 minute	1
4	4°C	HOLD	

[0313] Amplification primers (2.5 µL) and a polymerase, such as KAPA HiFi HotStart ReadyMix (25 µL) were added to a tube containing the water/streptavidin binding bead slurry, and the tube mixed by pipetting. The tube was then split into two reactions. The tube was pulse-spun and transferred to the thermal cycler and the cycling program in **Table 6** was started. When thermal cycler program was complete, samples were removed from the block and immediately subjected to purification. DNA purification beads pre-equilibrated at room temperature were vortexed until homogenized. 90 µL (1.8x) homogenized DNA purification beads were added to the tube, and mixed well by vortexing. The tube was incubated for 5 minutes at room temperature, and placed on a magnetic stand. DNA purification beads pelleted, leaving a clear supernatant within 1 minute. The clear supernatant was discarded, and the tube was left on the magnetic stand. The DNA purification bead pellet was washed with 200 µL of freshly prepared 80 percent ethanol, incubated for 1 minute, then removed and the ethanol discarded. The wash was repeated once, for a total of two washes, while keeping the tube on the magnetic stand. All remaining ethanol was removed and discarded with a 10 µL pipette, making sure to not disturb the DNA purification bead pellet. The DNA purification bead pellet was air-dried on a magnetic stand for 5-10 minutes or until the pellet was dry. The tube was removed from the magnetic stand and 32 µL of water was added, mixed by pipetting until homogenized, and incubated at room temperature for 2 minutes. The tube was placed on a magnetic stand for 3 minutes or until beads were fully pelleted. 30 µL of clear supernatant was recovered and transferred to a clean thin-walled PCR 0.2 mL strip-tube, making sure not to disturb DNA purification bead pellet. Average fragment length was between about 375 bp to about 425 bp using a range setting of 150 bp to 1000 bp on an analysis instrument. Ideally, the final

concentration values is at least about 15 ng/μL. Each capture was quantified and validated using Next Generation Sequencing (NGS).

[0314] A summary of NGS metrics is shown in **Table 7**, **Table 8** as compared to a comparator exome capture kit (Comparator Kit D). Library 4 has probes (baits) that correspond to a higher percentage of exon targets than Comparator Kit D. This results in less sequencing to obtain comparable quality and coverage of target sequences using Library 4.

Table 7

NGS Metric	Comparator Kit D	Library 4
Target Territory	38.8 Mb	33.2 Mb
Bait Territory	50.8 Mb	36.7 Mb
Bait Design Efficiency	76.5%	90.3%
Capture Plex	8-plex	8-plex
PF Reads	57.7M	49.3M
Normalized Coverage	150X	150X
HS Library Size	30.3 M	404.0 M
Percent Duplication	32.5%	2.5%
Fold Enrichment	43.2	48.6
Fold 80 Base Penalty	1.84	1.40

Table 8

NGS Metric	Comparator Kit D	Library 4
Percent Pass Filtered Unique Reads (PCT_PF_UQ_READS)	67.6%	97.5%
Percent Target Bases at 1X	99.8%	99.8%
Percent Target Bases at 20X	90.3%	99.3%
Percent Target Bases at 30X	72.4%	96.2%

[0315] A comparison of overlapping target regions for both Kit D and Library 4 (total reads normalized to 96X coverage) is shown in **Table 9**. Library 4 was processed as 8 samples per hybridization, and Kit D was processed at 2 samples per hybridization. Additionally, for both libraries, single nucleotide polymorphism and in-frame deletion calls from overlapping regions were compared against high-confidence regions identified from “Genome in a Bottle” NA12878 reference data (**Table 10**). Library 4 performed similarly or better (higher indel precision) than Kit D in identifying SNPs and indels.

Table 9

NGS Metric	Comparator Kit D	Library 4
Percent Pass Filtered Reads (PCT_PF_UQ_READS)	94.60%	97.7%
Percent Selected Bases	79%	80%
Percent Target Bases at 1X	100%	100%
Percent Target Bases at 20X	90%	96%
Percent Target Bases at 30X	71%	77%
Fold Enrichment	44.9	49.9
Fold 80 Base Penalty	1.76	1.4
HS Library Size	122 M	267 M

Table 10

Variants	Comparator Kit D		Library 4	
	Precision	Sensitivity	Precision	Sensitivity
Single Nucleotide Polymorphisms (SNPs)	98.59%	99.23%	99.05%	99.27%
In-Frame Deletions (Indels)	76.42%	94.12%	87.76%	94.85%
Total	98.14%	99.15%	98.85%	99.20%

[0316] Precision represents the ratio of true positive calls to total (true and false) positive calls. Sensitivity represents the ratio of true positive calls to total true values (true positive and false negative).

[0317] Example 7: Exome probes with a pain gene panel

[0318] Sequencing data was acquired using the general method of **Example 6**, with modification: different combinations of probe sets were evaluated. Two different exome probe libraries were used (Exome 1 and Exome 2) as well as a second polynucleotide probe library (panel) which targeted genes associated with pain. Both exome panels were evaluated individually, as well as with the pain gene panels mixed. This resulted in additional sequencing coverage of these genomic regions; one such exemplary region of chromosome 11 is shown in **FIG. 3**. This result was compared with separate analyses in which various exome panel and pain gene panels were individually evaluated, or combined for areas spanning multiple chromosomes, such as chromosome 1, 2, 6, and 22 (data not shown).

[0319] Example 8: Universal Blockers with Locked Nucleic Acids

[0320] Sequencing data was acquired using the general method of Example 6, with modification: four polynucleotide blockers were evaluated in separate analyses for their ability to reduce off-target binding (**FIG. 4A**). Universal blockers comprising LNAs performed comparably to positive control conditions with specific blockers, achieving less than 20% off bait across two different index sequences.

[0321] **Example 9: Combinations of Universal Blockers with Locked Nucleic Acids**

[0322] Sequencing data was acquired using the general method of Example 6, with modification: different combinations of four polynucleotide blockers were evaluated in separate conditions for their ability to reduce off-target binding (**FIG. 4B**). Universal blockers comprising LNAs performed comparably to positive control conditions with specific blockers when all four blockers (C, D, E, F) were present, achieving less than 20% off bait.

[0323] **Example 10: Universal Blockers with Locked Nucleic Acids**

[0324] Sequencing data was acquired using the general method of Example 6, with modification: four polynucleotide blockers were evaluated in separate conditions for their ability to reduce off-target binding (**FIG. 4C**) in conditions comprising 1 or 8 different index sequences (1- or 8-plex). Universal blockers comprising LNAs at 0.125 nmol each performed comparably to positive control conditions with 1 nmole specific blockers, achieving less than 20% off bait across both 1-plex and 8-plex conditions. Universal blockers comprising LNA performed better (less than 20% off bait) than specific blockers (more than 20% off bait) when they were each present in comparable amounts by mass (**FIG. 4C**).

[0325] **Example 11: Titration of Universal Blockers with Locked Nucleic Acids**

[0326] Sequencing data was acquired using the general method of Example 6, with modification: four polynucleotide blockers were evaluated in separate conditions for their ability to reduce off-target binding (**FIG. 4D**). Universal blockers comprising LNAs present in amounts less than 0.01 nmole each achieved less than 20% off bait.

[0327] **Example 12: Universal Blockers with Varying Amounts of Locked Nucleic Acids**

[0328] Sequencing data was acquired using the general method of Example 6, with modification: four polynucleotide blockers comprising varying amounts of LNAs were evaluated in separate conditions for their ability to reduce off-target binding (**FIG. 4E**). Universal blockers comprising at least 8 LNAs performed comparably to positive control conditions with specific blockers, achieving less than 20% off bait.

[0329] **Example 13: Universal Blockers with Bridged Nucleic Acids**

[0330] Sequencing data was acquired using the general method of Example 6, with modification: four different polynucleotide blockers sets were evaluated in separate conditions for

their ability to reduce off-target binding (**FIG. 4F**). Universal blockers comprising BNAs performed better than positive control conditions with specific blockers, achieving less than 27% off bait.

[0331] Example 14: Use of Partially Biotinylated Probes

[0332] Sequencing data is acquired using the general method of Example 6, with modification: separate conditions were run varying ratios of biotinylated to non-biotinylated exome probes, and percent off bait and dropout rates were measured. Probe libraries comprising only 50% biotinylated baits achieved a percent off bait rate of less than 25% (**FIG. 5A**), and A/T and G/C dropout rates of less than 2% (**FIG. 5B**).

[0333] Example 15: Dilution Calibration

[0334] Sequencing data is acquired using the general method of Example 6, with modification: separate conditions were run by varying the dilution of probes (probe mass:target size), and the HS library size:target size was analyzed. Although the exome library targets roughly followed a linear distribution, the smaller panel did not vary linearly (**FIG. 6A**). When the data was refit to a kinetic model, both the exome and gene panel are fit on the same curve for various dilutions (**FIG. 6B**). This allowed the accurate prediction of an optimal ratio of exome:gene panel probes to achieve a desired capture amount. For example, to capture 45% of the targets for both the exome and gene library, the gene panel probes were spiked in at 22% per bait mass relative to the exome library.

[0335] Example 16: Performance of a Custom Panel Library

[0336] Sequencing data was acquired using the general method of Example 6, with modification: two different custom probe panels Library 1 (757 kb) and Library 2 (803 kb) were used to target different areas of the genome (**FIG. 22**). The two panels resulted in a high percentage of on-target reads, as well as a high percentage of targets with >20% read depth (**Table 10**). Library 1 demonstrated an off-target rate of 9%.

Table 11

Condition	HS Library Size	Fold 80 base penalty	% on-target	% targets with read depth >20%	% targets with read depth >30%
Library 1, sample 1	25.2	1.25	91	98	96
Library 1, sample 2	16.9	1.24	91	98	96
Library 2, sample 1	31.2	1.22	69	99	98
Library 2, sample 2	24.2	1.22	70	99	98

[0337] Example 17: Evaluation of Probe Performance and Tuning

[0338] A subset of polynucleotide probes is selectively removed from the capture library of Example 6, and the capture/sequencing method is repeated on the same sample using the general

method of Example 6. Outcome metrics such as on-bait coverage, off-target, and fold 80 base penalty are measured. The process is iterated with different probe subsets, and the sequencing results correlated. The best performing probe subsets are then combined and evaluated in a similar manner.

[0339] Example 18: Exome probes with additional SNP panel

[0340] A subset of polynucleotide probes (panel) is selectively added to the capture library of **Example 6**, and the capture/sequencing method is repeated on the same sample using the general method of **Example 6**. The subset of polynucleotides targets areas of the genome comprising single nucleotide polymorphisms (SNPs). The panel allows for the identity of bases at each of the SNPs to be determined by increasing the read depth at these sites, including sites which are heterozygous.

[0341] Example 19: Exome probes with an intron panel

[0342] Sequencing data is acquired using the general method of Example 6, with modification: a second polynucleotide probe library which targets introns is mixed with the exome library. This results in additional sequencing coverage of these genomic regions. Data not shown.

[0343] Example 20: Universal Blockers with Bridged Nucleic Acids (8-Plex)

[0344] Sequencing data is acquired using the general method of Example 6, with modification: adapter-tagged genomic fragments comprising 8 different barcode sequences are used, and four different polynucleotide blockers are evaluated for their ability to reduce off-target binding.

[0345] Example 21: Exome probes with a custom panel

[0346] Sequencing data is acquired using the general method of Example 6, with modification: different combinations of probe sets are evaluated. Two different exome probe libraries are used (Exome 1 and Exome 2) as well as either Library 1 or Library 2 which target additional regions of the genome. Both exome panels are evaluated individually, as well as with Library 1 or Library 2 panels mixed in with each. Sequencing metrics are obtained and evaluated for both the exome, as well as areas targeted by Library 1 or Library 2.

[0347] Example 22. Universal Blockers with improved on-target performance

[0348] Universal blockers were used with adapter-tagged genomic fragments. *See as an example FIG. 23A.* Individual libraries were generated from a single genomic source (NA12878; Coriell) and compatible adapters. Each prepared library was then captured either in the absence or presence of universal blockers. Following sequencing, reads were downsampled to 150X of targeted bases and evaluated using Picard metric tools with a MapQuality filter = 20. Error bars denote one standard deviation; $N \geq 2$. As seen in **FIG. 24**, there was improved on-target performance across a wide range of index designs. Cot DNA was present in all samples. As seen

in **FIG. 25**, there was improved on-target performance across a range of panel sizes using the universal blockers.

[0349] Example 23. Custom panel designs across a range of panel sizes and target regions

[0350] Sequencing data was acquired using the general method of **Example 6**. Details of the library are seen in **Table 12**. Briefly, hybrid capture was performed using several target enrichment panels designed herein using 500 ng of gDNA (NA12878; Coriell) per single-plex pool following manufacturer's recommendations. Sequencing was performed with a NextSeq 500/550 High Output v2 kit to generate 2x76 paired end reads. Data was downsampled to 150x of target size and analyzed using Picard Metrics with a mapping quality of 20; N = 2. The panels resulted in a high percentage of on-target reads, as well improved uniformity and low duplication rate (**Table 12**). **FIG. 26A** shows percentage of reads in each panel achieving 30x coverage and **FIG. 26B** shows uniformity (fold-80).

Table 12.

Panel Description				Performance (Picard Metrics)		
Name	Target Size (Mb)	Probes	Genes	Uniformity (Fold-80)	On-Target Rate	Duplication Rate
mtDNA Library	0.017	139	37	1.22	82%	0.8%
Cancer Library	0.037	384	50	1.36	68%	1.9%
Neurodegenerative Library	0.6	6,024	118	1.23	61%	1.0%
Cancer Library 2	0.81	7,446	127	1.25	70%	2.2%
Cancer Library 3	1.69	19,661	522	1.27	78%	1.4%
Pan-Cancer Library	3.4	31,002	578	1.27	62%	1.9%
Exploratory Cancer Library	13.2	135,937	5,442	1.30	80%	3.0%

[0351] Example 24. Custom panel performance during multiplex target enrichment

[0352] Sequencing data was acquired using the general method of **Example 6**. Data from multiplex target enrichment is seen in **Table 13** below and **FIG. 27A**. **FIG. 27A** shows coverage distribution and cumulative coverage with the x-axis for both charts of **FIG. 27A** is coverage, and

the y-axis for the left chart is % bases with coverage, and for the right chart is bases with % bases with coverage > X. All data were subsampled to 150x coverage. MapQuality filter = 20.

Sequencing was performed on an Illumina® NextSeq® instrument using 2 x 76 reads. The data show high uniformity for all levels of multiplexing, high on-target rates that do not vary with higher levels of multiplexing, and low duplication rates across all levels of multiplexing.

Table 13.

	Performance (Picard Metrics)		
Multiplexing	Uniformity (Fold-80)	On-Target Rate	Duplication Rate
Singleplex	1.25	70%	1.8%
8-Plex	1.27	69%	2.2%
16-Plex	1.30	69%	2.7%

[0353] Probes were designed to maximize the capture of unique molecules and minimize sequencing duplicates to delivery high multiplex performance. High capture performance was determined on three panels of 800 kb, 3.3 Mb and a fixed Exome of 33.1 Mb. Consistent capture coverage at 30x is observed across all samples and multiplexing conditions (**FIG. 27B**). The magnitude of duplicate rate increase was minimal. For an 800kb panel duplication rate increased from 1.8% to 2.7% between 1-plex and 16-plex captures, respectively, and similar observations were made with larger panels. The impact to performance was confirmed with consistent 30x coverage.

[0354] Probes were also designed to support multiplexing without increasing sample mass. Hybrid capture was performed using an exome target enrichment panel described herein (33.1 Mb) using 500 ng of library (NA12878; Coriell) as a single-plex capture following manufacturer's recommendations. N = 2. **FIG. 27C** shows effect on number of PCR cycles on uniformity.

[0355] Hybrid capture was performed using an exome target enrichment panel described herein (33.1 Mb) using 150 ng (18.75 ng per library) or 1500 ng (187.5 ng per library) of library (NA12878; Coriell) per 8-plex pool following manufacturer's recommendations. Data was down-sampled to 100x of target size; N = 2. Consistent 30x coverage clearly demonstrates the capacity of this system to multiplex with reduced mass input without degradation to performance (**FIG. 27D**).

[0356] **Example 25. Custom panel reproducibility**

[0357] Sequencing data was acquired using the general method of Example 6 to assess the reproducibility of custom panels from lot to lot. As seen in **FIGS. 28A-28I**, the custom panels

demonstrate a low lot-to-lot variation. Lots A and B were independent lots produced using two synthesis runs. Each dot represents probe abundance (**FIG. 28A**) or probe coverage following NGS target enrichment at 1500x coverage (**FIG 28B**). **FIG. 28A** shows consistent quality of 800 kb panels as assessed by NGS.

[0358] A panel containing roughly 7,400 probes (800 kb) was re-synthesized ~1 month apart (Lot1 and Lot2), with two amplification replicates in each Lot (Replicate 1 and 2). **FIG. 28C** shows reproducibility of probe representation within same synthesis and different amplifications. **FIG. 28D** shows reproducibility of probe representation between syntheses.

[0359] **FIGS. 28E-28I** show data that was downsampled to 150x of target size and analyzed using Picard Metrics with a mapping quality of 20; N = 2. **FIG. 28E** show lot to lot reproducibility capture per probe. **FIGS. 28F-28I** show reproducibility of probe target enrichment performance between syntheses.

[0360] **Example 26. Flexible and modular custom panels**

[0361] Content can be added to or enhanced. *See FIG. 29A*. Adding content to the panel increases the number of targets covered. Enhancing content to the panel refers to the coverage of specific regions.

[0362] 3 Mb of additional target regions was added derived from the RefSeq database. The production of this panel increased coverage and did not decrease performance. Coverage improved to >99% of the RefSeq, CCDS, and GENCODE databases. Further, the custom panel displayed high uniformity and on-target rate, as well as a low duplicate rate (all results based on 150x sequencing).

[0363] The database coverage as seen in **Table 14** was increased using the custom panels as described herein. The data compared the overlap between panel content to the protein-coding regions in the databases annotated on the primary human genome assembly (alternative chromosomes were excluded) as of May 2018 (UCSC genome browser). A1, A2, and I-1 are commercially available comparator panels from different vendors. Comparisons were performed using the BEDtools suite and genome version indicated in parentheses. The addition of 3 Mb of content improved the coverage of RefSeq and GENCODE databases to >99%.

Table 14.

	Database Coverage		
	RefSeq (35.9Mb)	CCDS21 (33.2Mb)	GENCODE v28 (34.8Mb)
Panel 1	92.3%	99.5%	95.1%
Panel 1 + Added Content	99.2%	99.5%	99.1%

A-1 (hg19)*	88.3 %	91.9%	90.8%
A-2 (hg38)*	91.0%	94.6%	94.0%
I-1 (hg19)	94.1%	98.3%	95.7%

[0364] FIGS. 29B-29D show data from Panel 1 and Panel 1 + Added Content on Fold (FIG. 29B), duplicate rate (FIG. 29C), and percent on target (FIG. 29D). FIG. 29E and FIG. 29F show comparative data for target coverage (FIG. 29E) and fold-80 base penalty (FIG. 29F).

[0365] **Example 27. Effect of 30,000 Probes on Capture**

[0366] The effect of mismatches on capture was determined for optimizing probe design. To examine the effects of number and distribution of mismatches on capture efficiency, two panels, Control and Variant were designed and synthesized. Each panel (Variant and Control) contained 28,794 probes. The Control panel contained probes selected from the human exome panel designed and synthesized using methods as described herein that perfectly match the human genome reference. The Variant panel contained the same probes but with 1-50 mismatches distributed at random, or as one continuous stretch (FIG. 30A). In the control panel, the probes were designed to be complementary to their targets. In the variant panel 1-50 mismatches (yellow) were introduced either randomly along the probe (RND) or all together in a single continuous stretch (CONT). Also, 382 control probes without mismatches were added to both panels for normalization (in grey), thus the Control and Variant pools contained a total of 29,176 probes.

[0367] FIGS. 30B-30C shows probes with varying numbers of mismatches on capture efficiency. Distribution of relative capture efficiency for probes with a single mismatch (gray) and probes with multiple mismatches (green lines; the number of mismatches is indicated in the left top corner) is shown. Solid line depicts the distribution for probes with randomly distributed mismatches (RND), and the dotted line indicates the distribution for probes with continuous mismatches (CONT). Probes with 50 mismatches arranged in one continuous stretch capture as well as probes with 10-15 mismatches distributed randomly, while probes with 50 mismatches distributed randomly were completely ineffective.

[0368] Other factors such as GC, length of perfect match and hybridization temperature can modulate capture efficiency in the presence of mismatches. FIG. 30D shows the effect on temperature on capture efficiency in the presence of mismatches.

[0369] FIGS. 30E-30F show metagenomic and bisulfide capture efficiency prediction for the design of 450 whole genome Zika isolates from human samples (FIG. 30E) and all CpG islands in the human genome (FIG. 30F). CpG islands were downloaded from the UCSC annotation track for human genome hg38 and designed using design methods as described herein.

[0370] **Example 28. Probe Specificity for Downstream Applications**

[0371] Adaptive designs where experimental results from a first pass design were used to determine sequences that should be removed. **FIG. 31A** shows improvements after a single pass adaptive design for moderate and aggressive off target reduction in panels with challenging target regions (respectively 37Kb and 800Kb, 3 probes and ~4% of probes removed). **FIG. 31B** shows the level off target predicted by our model compared to that measured by experimentation (axes) and the fraction out of the total number of baits required in each case to achieve it. **FIG. 31C** shows results for a custom design against a particularly hard set of target regions, various levels of stringency, and the effectiveness of bait removal based on methods described herein.

[0372] **Example 29. RefSeq Design**

[0373] A RefSeq panel design was designed in hg38 and included the union of CCDS21, RefSeq all coding sequence, and GENCODE v28 basic coding sequences. The size of RefSeq alone (Exome) was 3.5Mb and the combined Core Exome+RefSeq (Exome+RefSeq) was 36.5Mb. Experiments were run using 50 ng of gDNA (NA12878) as 1-plex and 8-plex run in triplicate, and evaluated at 150x sequencing with 76bp reads. The target file was 36.5 Mb.

[0374] The RefSeq panel design was assessed for depth of coverage, specificity, uniformity, library complexity, duplicate rate, and coverage rate. **FIG. 32A** shows depth of coverage. More than 95% of target bases at 20X were observed. More than 90% of target bases at 30X were observed. **FIG. 32B** shows specificity of the RefSeq panel. The percent off target was less than 0.2. **FIG. 32C** shows uniformity of the RefSeq panel. The fold 80 was less than 1.5. **FIG. 32D** shows the complexity of the library. The library size was greater than 320 million. **FIG. 32E** shows the duplicate rate of the RefSeq panel. The duplicate rate was less than 4%. **FIG. 32F** shows the coverage ratio of the RefSeq panel. The coverage ratio was between 0.9 and 1.1. As seen in **FIG. 32F**, the coverage ratio was less than 1.1.

[0375] **Example 30: Genomic DNA capture with an exome-targeting polynucleotide probe library, using various additives in the binding buffer**

[0376] Sequencing data is acquired using the general method of **Example 6**, with modification: various binding buffers comprising different additives were used in separate sequencing runs, and a 0.8 Mb custom probe panel library was used instead of the 36.7 Mb probe library. The results of the sequencing analysis are found in **FIG. 33C**. Addition of mineral oil to the binding buffer led to a significant decrease in the percent off target rates. Addition of 5% PEG to the binding buffer also led to a decrease in off target rates relative to the control run (water added).

[0377] **Example 31: Genomic DNA capture with an exome-targeting polynucleotide probe library, using mineral oil in buffers**

[0378] Sequencing data is acquired using the general method of **Example 6**, with modification: various buffers comprising mineral oil were used in separate sequencing runs, the number of washes was varied, and a 0.8 Mb custom probe panel library was used instead of the 36.7 Mb probe library. Conditions were run in duplicate. The results of the sequencing analysis for off target rates are found in **FIG. 34A**. Addition of mineral oil to wash buffer 1, first wash with wash buffer 2, or last wash with wash buffer 2 gave off-target rates that were comparable to no mineral oil conditions. Addition of mineral oil to hybridization buffer, first binding buffer, or last binding buffer led to a significant decrease in the percent off target rates.

[0379] **Example 32: Genomic DNA capture with an exome-targeting polynucleotide probe library, using mineral oil and washes**

[0380] Sequencing data is acquired using the general method of **Example 6**, with modification: hybridization and binding buffers comprising mineral oil were used in, the number of washes was varied, and a 0.8 Mb custom probe panel library was used instead of the 36.7 Mb probe library. Conditions were run in 2-7 replicates. The results of the sequencing analysis are found in **FIG. 34B-34E**. Four washes with wash buffer 1 generally led to a decrease in percent off bait (4 washes: 38.31% vs. 1 wash: 56.86%, without mineral oil), unless mineral oil was used (1 wash: 34.89% vs. 4 washes: 38.31% **FIG. 34B**); mineral oil in conjunction with a single wash with wash buffer 1 led to an average off bait percentage of 34.89%. Addition of mineral oil in general lowered GC dropout rates (**FIG. 34C**, intersections of dashed lines indicate average values). Additional washes led to less run to run variance in HS library size and fold 80 base penalty, independent of mineral oil addition (**FIG. 34D** and **FIG. 34E**).

[0381] **Example 33: Genomic DNA capture with an exome-targeting polynucleotide probe library, using a liquid polymer and tube transfers**

[0382] Sequencing data is acquired using the general method of **Example 6**, with modification: hybridization and binding buffers comprising a liquid polymer (Polymer A) additive were used in, a tube transfer was optionally performed during washes, and 800 kb and 40 kb custom probe panel libraries were used in independent runs instead of the 36.7 Mb probe library. Polymer A is a high molecular weight liquid polymer, that has a vapor pressure of < 1 mm Hg, and a water solubility of <100 ppb. Conditions were generally run in duplicate. Transferring tubes between washes and/or use of liquid polymer generally led to a decrease in percent off bait (**FIG. 35A**), as well as an increase in HS Library size for both 40 kb and 800 kb libraries (**FIG. 35B**). Other variables such as fold 80 base penalty and GC dropouts were relatively unaffected, although use of either tube changes, liquid polymer additive, or a combination of both resulted in fewer AT dropouts (data not shown).

[0383] Example 34: Genomic DNA capture with an exome-targeting polynucleotide probe library, using agitation and controlled aspiration

[0384] Sequencing data is acquired using the general method of **Example 6**, with modification: different levels of agitation/mixing and aspiration methods were used in separate sequencing runs, and a 0.8 Mb custom probe panel library was used instead of the 36.7 Mb probe library. High agitation comprised a short vortexing of the hybridization and binding buffer during mixing, while low agitation comprised flicking the tube during mixing. Top aspirate comprised collecting only liquid near the air-water interface, and slowly lowering a pipette tip as the liquid level dropped. Higher levels of agitation increased the off target rates relative to low levels of agitation (**FIG. 36**). The lowest off target rates were achieved with a combination of low agitation and aspirating from the top of the tube.

[0385] Example 35: Fast hybridization buffers

[0386] Sequencing data is acquired using the general method of **Example 6**, with modification: genomic DNA (NA12878, Coriell) is hybridized and captured using either the a 33.1 Mb exome probe library or an 800 kb targeted library. Two different workflows are compared (**FIG. 38**). A standard buffer or “fast” hybridization buffer is used during hybridization of two different probe libraries (exome probes or an 800kb custom panel) to the nucleic acid sample, and the capture/hybridization reaction is heated to 50-75 °C for various periods of time (15 minutes to 8 hours) in a PCR thermocycler, with a lid temperature of 80-95 °C. Following sequencing, Picard HS_Metric tools (Pct_Target_Bases_30X) with default values are used for sequence analysis. Data are downsampled to 150x raw coverage of targeted bases for evaluation. Use of fast hybridization buffers results in a workflow that is completed in 5-9 hours.

[0387] Example 36: Fast hybridization buffers with liquid polymer

[0388] Sequencing data was acquired using the general method of **Examples 6 and 10**, with modification: genomic DNA (NA12878, Coriell) was hybridized and captured using either a 33.1 Mb exome probe library or an 800 kb targeted library. A “fast” hybridization buffer was used with liquid polymer during hybridization of two different probe libraries (exome probes or an 800kb custom panel) to the nucleic acid sample, and the capture/hybridization reaction was heated at 65 °C for various periods of time in a PCR thermocycler, with a lid temperature of 85 °C. Following sequencing, Picard HS_Metric tools (Pct_Target_Bases_30X) with default values were used for sequence analysis. Data were downsampled to 150x raw coverage of targeted bases for evaluation. For either panels a 15-min hybridization in Fast Hybridization Solution produced an equivalent performance to the 16-hr standard hybridization, and increasing hybridization times improved performance over the standard protocol using conventional hybridization buffers (**FIG. 37A-37E**).

Overall, fast hybridization conditions led to low 80 fold base penalty, high on target rate, and >90% of target bass sequenced at greater than 30X. (FIG. 39A, 39B) The protocol also performed similarly in a multiplex experiment (FIG. 39B).

[0389] Example 37: Capture of genomic DNA from an FFPE Sample

[0390] Sequencing data was acquired using the general method of Example 8, with modification: six different formalin-fixed paraffin-embedded (FFPE) samples covering four different tissues were used as samples. Genomic DNA samples from FFPE samples were sheared mechanically with Covaris AFA equipment with compatible oneTUBE AFA vessels. The instrument settings were adjusted for gDNA fragmentation to target a size distribution with a mode of 200–250 bp. Following end repair, A-tailing, and ligation of adapters, indexed libraries were subjected to capture in multiplexed reactions (8 libraries; 187.5 ng each; 1500 ng total) with a 33.1 Mb target region exome capture probe set in 16-hour hybridization reactions. Samples were sequenced on a NextSeq system (Illumina) with a NextSeq 500/550 High Output v2 kit to generate 2 x 76 paired-end reads and downsampled to 150x of targeted bases for evaluation. Picard HS_metrics tools with a mapping quality of 20 were utilized for sequence analysis. Average values presented with $N \geq 3$ for all observations. Positive controls were sheared with AFA but not subject to FFPE extraction (Table 15). Sequencing metrics for the FFPE samples are shown in FIGS. 39C-39G.

Table 15. Sequencing results using mechanical shearing of FFPE samples.

Metric	Q-ratio		Fold-80 base penalty	30x depth of coverage	Percent On-Target	Percent duplication rate
Variable →	Q129/Q41	Q305/Q41	FOLD_80_BASE_PENALTY	PCT_TARGET_BASES_30X	1 - (PCT_OFF_BAIT)	PCT_EXC_DUPE
High-Quality gDNA (positive control; NA12878)	---	---	1.31	93.3%	83.4%	3.1%
Uterus	1.24	0.34	1.43	93.9%	83.6%	5.7%
Uterus Carcinoma	0.99	0.19	1.4	94.9%	84.2%	7.2%
Lung	0.78	0.12	1.55	92.1%	83.7%	5.8%

Liver	0.91	0.09	1.69	88.4%	84.0%	10.6%
Kidney	0.62	0.06	1.69	89.7%	84.9%	8.9%

[0391] Example 38: Fast hybridization buffers with variable wash buffer 1 temperature

[0392] Sequencing data was acquired using the general method of **Examples 6 and 10**, with modification: the temperature of wash buffer 1 was varied to modify sequencing results, and the protocol was carried out as described below.

[0393] Step 1. Eight samples, each approximately 187.5 ng (1500 ng total) were transferred to a 0.2-ml thin-walled PCR strip-tube or 96-well plate. 4 uL comprising the exome capture probe panel, optionally 4 uL of a second panel, 8 uL of universal blockers, and 5 uL of blocker solution/buffer were added, the mixture pulse-spun, and the mixture evaporated using low or no heat.

[0394] Step 2. A 96-well thermal cycler was programmed with the following conditions and the heated lid set to 85°C, as shown in **Table 16**.

Table 16.

Step	Temperature	Time
1	95°C	HOLD
2	95°C	5 minutes
3	Hybridization temperature (e.g., 60°C)	HOLD

[0395] The dried hybridization reactions were each resuspended in 20 µl fast hybridization buffer, and mixed by flicking. The tubes were pulse spun to minimize bubbles. 30 µl of liquid polymer was then added to the top of the hybridization reaction, and the tube pulse-spun. Tubes were transferred to the preheated thermal cycler and moved to Step 2 of the thermocycler program (incubate at 95°C for 5 minutes). The tubes were then incubated at 60°C for a time of 15 minutes to 4 hours in a thermal cycler with the lid at 85°C. 450 µl wash buffer 1 was heated the desired temperature (e.g., 70°C, or other temperature depending on desired sequencing metrics) and 700 µl wash buffer 2 was heated to 48°C. Streptavidin Binding Beads were equilibrated to room temperature for at least 30 minutes and then vortexed until mixed. 100 µl Streptavidin Binding Beads were added to a 1.5-ml microcentrifuge tube. One tube was prepared for each hybridization reaction. 200 µl fast binding buffer was added to the tubes and mixed by pipetting. The tubes were placed on a magnetic stand for 1 minute, then removed and the clear supernatant discarded, without disturbing the bead pellet. The tube was then removed from the magnetic stand. The pellet was washed two more times for a total of three washes with the fast binding buffer. After removing the

clear supernatant from the third wash, a final 200 µl fast binding buffer was added and the beads resuspended by vortexing until homogenized. The tubes of the hybridization reaction were mixed with the Streptavidin Binding Beads for 30 minutes at room temperature on a shaker, rocker, or rotator at a speed sufficient to keep the solution mixed.

[0396] Step 3. Tubes containing the hybridization reaction with Streptavidin Binding Beads were removed from the mixer and pulse-spun to ensure solution was at the bottom of the tubes, and the tubes were placed on a magnetic stand for 1 minute. The clear supernatant including the liquid polymer was removed and discarded with disturbing the pellet. The tubes were removed from the magnetic stand and 200 µl preheated fast wash buffer 1 was added, then mixed by pipetting. The tubes were incubated for 5 minutes at 70°C, and placed on a magnetic stand for 1 minute. The clear supernatant was removed and discarded without disturbing the bead pellet. The tubes were then removed from the magnetic stand and an additional 200 µl of preheated fast wash buffer 1 was added, followed by mixing and incubation 5 minutes at 70°C. The tubes were pulse-spun to ensure solution was at the bottom of the tubes. After the hybridization is complete, the thermal cycler lid was opened and the volume of each hybridization reaction including liquid polymer quickly transferred into a corresponding tube of washed Streptavidin Binding Beads, then mixed. The entire volume (~200 µl) was transferred into a new 1.5-ml microcentrifuge tube, one per hybridization reaction. The tubes were placed on a magnetic stand for 1 minute, followed by removal and discard of the clear supernatant. The tubes were removed from the magnetic stand and 200 µl of 48°C wash buffer 2 was added, mixed by pipetting, and then pulse-spun to ensure the solution was at the bottom of the tubes. The tubes were then incubated for 5 minutes at 48°C, placed on a magnetic stand for 1 minute, and the clear supernatant removed and discarded with disturbing the pellet. The wash step was repeated two more times, for a total of three washes. After the final wash, a 10 µl pipette was used to remove traces of supernatant. Without allowing the pellet to dry, the tubes were removed from the magnetic stand and 45 µl of water added, mixed, and then incubated on ice (hereafter referred to as the Streptavidin Binding Bead slurry).

[0397] Step 4. A thermal cycler was programmed with the following conditions in **Table 17**, and the heated lid set to 105°C. 22.5 µl of the Streptavidin Binding Bead slurry was transferred to a 0.2-ml thin-walled PCR strip- tubes and kept on ice until ready for use in the next step. A PCR mixture was prepared by adding a PCR polymerase mastermix and adapter-specific primers to the tubes containing the Streptavidin Binding Bead slurry and mixed by pipetting. The tubes were pulse-spun, and transferred to the thermal cycler and start the cycling program.

Table 17. Thermocycler program for PCR library amplification.

Step		Temperature	Time	Number of Cycles																				
1	Initialization	98 C	45 seconds	1	<table><tr><th>Custom Panel Size</th><th>Number of Cycles</th></tr><tr><td>>100 Mb</td><td>5</td></tr><tr><td>50-100 Mb</td><td>7</td></tr><tr><td>10-500 Mb</td><td>8</td></tr><tr><td>1-10 Mb</td><td>9</td></tr><tr><td>500-1,000 kb</td><td>11</td></tr><tr><td>100-500 kb</td><td>13</td></tr><tr><td>50-100 kb</td><td>14</td></tr><tr><td><50 kb</td><td>15</td></tr></table>	Custom Panel Size	Number of Cycles	>100 Mb	5	50-100 Mb	7	10-500 Mb	8	1-10 Mb	9	500-1,000 kb	11	100-500 kb	13	50-100 kb	14	<50 kb	15	
Custom Panel Size	Number of Cycles																							
>100 Mb	5																							
50-100 Mb	7																							
10-500 Mb	8																							
1-10 Mb	9																							
500-1,000 kb	11																							
100-500 kb	13																							
50-100 kb	14																							
<50 kb	15																							
2	Denaturation	98 C	15 seconds	Varies																				
	Annealing	60 C	30 seconds																					
	Extension	72 C	30 seconds																					
3	Final Extension	72 C	1 minute	1																				
4	Final Hold	4 C	HOLD	-																				

[0398] 50 µl (1.0x) homogenized DNA Purification Beads were added to the tubes, mixed by vortexing, and incubated for 5 minutes at room temperature. The tubes were then placed on a magnetic plate for 1 minute. The clear supernatant was removed from the tubes. The DNA Purification Bead pellet was washed with 200 µl freshly prepared 80% ethanol for 1 minute, then the ethanol was removed and discarded. This wash was repeated once, for a total of two washes, while keeping the tube on the magnetic plate. A 10 µl pipet was used to remove residual ethanol, making sure to not disturb the bead pellet. The bead pellet was air-dried on a magnetic plate for 5–10 minutes or until the bead pellet was dry. The tubes were removed from the magnetic plate and 32 µl water was added. The resulting solution was mixed by pipetting until homogenized and incubated at room temperature for 2 minutes. The tubes were then placed on a magnetic plate and let stand for 3 minutes or until the beads fully pelleted. 30 µl of the clear supernatant containing the enriched library was transferred to a clean thin-walled PCR 0.2-ml strip-tube.

[0399] Step 5. Each enriched library was validated and quantified for size and quality using an appropriate assay, such as the Agilent BioAnalyzer High Sensitivity DNA Kit and a Thermo Fisher scientific Qubit dsDNA High Sensitivity Quantitation Assay. Samples were then loaded onto an Illumina sequencing instrument for analysis. Sampling was conducted at 150X (theoretical read depth), and mapping quality was >20. The effects on various NGS sequencing metrics for various fast hybridization wash buffer 1 temperatures are shown in **FIG. 40**.

[0400] **Example 39: Blockers targeting strands of the adapter**

[0401] The general procedures of Example 8 were executed with modification: additional blockers were added that target the top strand, bottom strand, or both strands of the adapter

sequence. The results are shown in **Table 18**. “Outside” refers to the portion of the adapter between the terminus and the barcode. “Inside” refers to the portion of the adapter between the barcode and genomic insert. The percent off bait is shown in **FIG. 41**.

[0402] Table 18. Strand-specific blockers

Blockers used during hybridization (to target strands of the adapter)	Percent off-bait observed post-sequencing of target enriched genomic material
Standard four blockers targeting the top strand	17.4%
Two “outside” blockers targeting the top strand and two “inside” blockers targeting the bottom strand	20.7%
Two “inside” blockers targeting the top strand and two “outside” blockers targeting the bottom strand	23.6%
Four blockers targeting the top strand and four blockers targeting the bottom strand	64.1%
Four blockers targeting the bottom strand	25.7%

[0403] Example 40: Blockers with tagmentation-based library generation

[0404] Following the general procedures of Example 8, a genomic library was treated with an engineered transposon to fragment the DNA and tag the fragments with an adapter sequencing in a single step to generate fragments of approximately 300 bases in length. The resulting library of fragments were then amplified with a limited PCR-cycle procedure using primers that add additional adapter sequences to both ends of the DNA fragments. Prior to sequencing, the adapter-ligated genomic library was enriched using an exome panel in the presence of either four universal blockers designed specifically for the tagmentation adapters (DEJL-1 or DEJL-2); four non-tagmentation universal blockers (CDEF), two universal blockers targeting the adapter region adjacent to the genomic insert (JL), or a control experiment without blockers (NB). Blockers targeting the tagmentation adapters comprised 11-13 locked nucleic acids (32-45% of the bases), a T_m of 84-90 degrees C, and a length of 29-34 bases. The addition of blockers led to significant decreases in off-bait capture. Off-bait percentage was approximately 25%, AT dropout was approximately 7%, percent 30X base coverage was approximately 30%, and fold 80 base penalty was 1.6. The results after sequencing for various NGS metrics are shown in **42A-42E** and **FIGS. 43**. Without being bound by theory, gDNA library size

[0405] Example 41: Location of modified bases in blockers

[0406] The general procedures of Example 8 were followed with modification: three of four universal blockers were held constant, and the fourth blocker designed was manipulated by changing the location of the positions comprising locked nucleic acids. All blocker designs maintained an overall T_m of at least 82 degrees C, regardless of locked nucleic acid placement. All

designs tested gave comparable results that were independent of locked nucleic acid placement, provided the overall T_m was at least 82 degrees C (data not shown).

[0407] Example 42: Blockers and alternative adapter designs

[0408] The general procedures of Example 8 are followed with modification: Y-adapters are replaced with “bubble” adapters or “clamp” adapters. After capture using blockers, sequencing metrics such as percent bases at 30X, off-bait percentage, AT/GC dropout, 80 fold base penalty, and on-target percent are measured.

[0409] Example 43: Multiplex Fast hybridization buffers with liquid polymer

[0410] Sequencing data is acquired using the general method of Example 38, with modification: samples from 16 different sources are individually, uniquely barcoded by sample and processed using the fast hybridization buffer protocol. Sequencing metrics for the 16 samples are comparable to experiments using only a single sample.

[0411] Example 44: Multiplex Fast hybridization buffers with liquid polymer

[0412] Sequencing data is acquired using the general method of Example 38 with modification: samples from 96 different sources are individually, uniquely barcoded by sample and processed using the fast hybridization buffer protocol. Sequencing metrics for the 96 samples are comparable to experiments using only a single sample.

[0413] Example 45: Fast hybridization buffers with tagmentation blockers

[0414] Sequencing data is acquired using the general method of Example 38 with modification: the library was prepared using the tagmentation procedure of Example 40.

[0415] Example 46: Fast hybridization buffers with Blockers and alternative adapter designs

[0416] Sequencing data is acquired using the general method of Example 38 with modification: the Y-adapters are replaced with “bubble” adapters or “clamp” adapters. After capture using blockers with the fast hybridization buffer, sequencing metrics such as percent bases at 30X, off-bait percentage, AT/GC dropout, 80 fold base penalty, and on-target percent are measured.

[0417] Example 47: Melting curve analysis for universal blockers

[0418] An experiment was conducted to empirically measure T_m between universal blockers and adapter-ligated genomic DNA (gDNA). gDNA libraries with adapters at 27 ng/ul, non-modified full length specific blockers at 1 nmol/ul total, and LNA-containing blockers at 0.5 nmol/ul total were used. Appropriate components were mixed with 10 μ M SYTO9 and 50nM ROX fluorescence dyes, denatured at 95°C, and heated from 40° to 95°C over 16 hours, holding at each 0.1° for 1 minute and 44 seconds. During the heat curve, fluorescence was recorded in a qPCR system and graphed as a normalized derivative. (**FIG. 44**). {gDNA} provided a maximum value at

~45°C, {gDNA + non-modified full length specific blockers} provided a maximum value at ~55°C, and {gDNA + LNA blockers} provided a maximum value at ~65°C in this experiment.

[0419] While preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

CLAIMS**WHAT IS CLAIMED IS:**

1. A method for sequencing genomic DNA, comprising:
 - (a) contacting a composition comprising a first polynucleotide library comprising at least 30,000 polynucleotides, wherein each of the at least 30,000 polynucleotides is present in an amount such that, following hybridization with genomic fragments and sequencing of the hybridized genomic fragments, the polynucleotide library provides
 - a read depth of at least 80 percent of the bases of the genomic fragments corresponding to the polynucleotides; and
 - a total number of sequencing reads, wherein the total number of sequencing reads are capable of covering 100 percent of each of the bases of the genomic fragments corresponding to the polynucleotides at a theoretical read depth, wherein the ratio of the read depth of at least 80 percent of the bases of the genomic fragments corresponding to the polynucleotides to the theoretical read depth is at least 0.5 with a plurality of genomic fragments;
 - (b) enriching at least one genomic fragment that binds to the first polynucleotide library to generate at least one enriched target polynucleotide;
 - (c) sequencing the at least one enriched target polynucleotide;
 - (d) identifying one or more positions of the at least one enriched polynucleotide having less than average read depth;
 - (e) repeating steps a-c, wherein a second polynucleotide library comprising at least 1500 polynucleotides is added to the composition, wherein the second polynucleotide library comprises at least one polynucleotide that binds to genomic fragments comprising the one or more positions having less than average read depth, wherein the presence of the second polynucleotide library increases the read depth at the one or more positions having less than average read depth.
2. The method of claim 1, wherein the first polynucleotide library and the second polynucleotide library do not comprise any common sequences.
3. The method of claim 1, wherein the first polynucleotide library and the second polynucleotide library comprise at least one common sequence.
4. The method of any one of claims 1-3, wherein the presence of the second polynucleotide library increases the read depth at the one or more positions of the least one enriched target polynucleotide having less than average read depth by at least 10 fold.

5. The method of claims 1-3, wherein the presence of the second polynucleotide library increases the read depth at the one or more positions of the at least one enriched target polynucleotide having less than average read depth by at least 100 fold.
6. The method of any one of claims 1-5, wherein the first method comprises at least 100,000 polynucleotides.
7. The method of claim 1-6, wherein the second method comprises at least 5,000 polynucleotides.
8. The method of any one of claims 1-5, wherein the first method comprises at least 100,000 polynucleotides and the second method comprises at least 5,000 polynucleotides.
9. The method of any one of claims 1-8 wherein the ratio of the read depth of at least 80 percent of the bases of the genomic fragments corresponding to the polynucleotides to the theoretical read depth is at least 0.5 with a plurality of genomic fragments;
10. The method of any one of claims 1-9, wherein at least 90% of the bases sequenced are at least 99.5% correct.
11. The method of any one of claims 1-9, wherein at least 90% of the bases sequenced are at least 99.9% correct.
12. The method of any one of claims 1-9, wherein at least 90% of the bases sequenced are at least 99.95% correct.
13. The method of any one of claims 1-12, wherein each of the genomic fragments is about 100 bases to about 500 bases in length.
14. The method of any one of claims 1-13, wherein the at least 30,000 polynucleotides encode for at least 1000 genes.
15. The method of any one of claims 1-14, wherein the at least 30,000 polynucleotides encode for at least one exon sequence.
16. The method of any one of claims 1-15, wherein the at least 1500 polynucleotides encode for at least one exon sequence.
17. The method of any one of claims 1-16 wherein the at least 1500 polynucleotides encode for at least 10 genes.
18. The method of any one of claims 1-16, wherein the at least 1500 polynucleotides encode for at least 100 genes.
19. The method of claim any one of claims 1-18, wherein the at least 1500 polynucleotides encode for at least one intron.
20. The method of claim any one of claims 1-19, wherein the at least 1500 polynucleotides encode for at least one single nucleotide polymorphism (SNP).

21. The method of claim 20, wherein the single nucleotide polymorphism (SNP) is heterozygous.
22. The method of any one of claims 1-21, wherein the fragments are sequenced by a method comprising sequencing by synthesis.
23. The method of claim 22, wherein the fragments are sequenced on an Illumina sequencing instrument.
24. A polynucleotide library, the polynucleotide library comprising at least 1500 polynucleotides, wherein less than all polynucleotides comprises a molecular tag, wherein each of the at least 5000 polynucleotides are present in an amount such that, following hybridization with genomic fragments and sequencing of the hybridized genomic fragments, the polynucleotide library provides
 - a read depth of at least 90 percent of the bases of the genomic fragments corresponding to the polynucleotides; and
 - a total number of sequencing reads, wherein the total number of sequencing reads are capable of covering 100 percent of each of the bases of the genomic fragments corresponding to the polynucleotides at a theoretical read depth, wherein the ratio of the read depth of at least 90 percent of the bases of the genomic fragments corresponding to the polynucleotides to the theoretical read depth is at least 0.5
25. The polynucleotide library of claim 24, wherein no more than 90% of the polynucleotides comprise a molecular tag.
26. The polynucleotide library of claims 24 or 25, wherein no more than 80% of the polynucleotides comprise a molecular tag.
27. The polynucleotide library of any one of claims 24-26, wherein no more than 50% of the polynucleotides comprise a molecular tag.
28. The polynucleotide library of any one of claims 24-26, wherein no more than 25% of the polynucleotides comprise a molecular tag.
29. The polynucleotide library of any one of claims 24-28, wherein the molecular tag is biotin.
30. The polynucleotide library of any one of claims 24-29, wherein the at least 5000 polynucleotides encode for at least 5000 genes.
31. The polynucleotide library of any one of claims 24-30, wherein the polynucleotide library comprises at least 30,000 polynucleotides.
32. The polynucleotide library of any one of claims 24-30, wherein the polynucleotide library comprises at least 100,000 polynucleotides.
33. A method for enriching nucleic acids comprising:

- (a) contacting the polynucleotide library of any one of claims 24-32 with a plurality of genomic fragments;
 - (b) enriching at least one genomic fragment that binds to the polynucleotide library to generate at least one enriched target polynucleotide; and
 - (c) sequencing the at least one enriched target polynucleotide.
34. The method of claim 33, wherein the polynucleotide library provides for at least 90 percent unique reads for the bases of the enriched target polynucleotide after sequencing.
35. The method of claim 33, wherein the polynucleotide library provides for at least 95 percent unique reads for the bases of the enriched target polynucleotide after sequencing.
36. The method of any one of claims 33-35, wherein the polynucleotide library provides for at least 80 percent of the bases of the enriched target polynucleotide having a read depth within about 1.5 times the mean read depth.
37. The method of claim 33-35, wherein the polynucleotide library provides for at least 90 percent of the bases of the enriched target polynucleotide having a read depth within about 1.5 times the mean read depth.
38. A polynucleotide library, the polynucleotide library comprising at least 5000 polynucleotides, wherein each of the at least 5000 polynucleotides is present in an amount such that, following hybridization with a composition comprising
- i) a genomic library, wherein the genomic library comprises polynucleotides each comprising genomic fragments, at least one index sequence, and at least one adapter; and
 - ii) at least one polynucleotide blocker, wherein the polynucleotide blocker is complementary to at least a portion of the adapter sequence, but not complementary to the at least one index sequence;
- and sequencing of the hybridized genomic fragments, the polynucleotide library provides for
- a read depth of at least 90 percent of the bases of the genomic fragments corresponding to the polynucleotides; and
 - a total number of sequencing reads, wherein the total number of sequencing reads are capable of covering 100 percent of each of the bases of the genomic fragments corresponding to the polynucleotides at a theoretical read depth, wherein the ratio of the read depth of at least 90 percent of the bases of the genomic fragments corresponding to the polynucleotides to the theoretical read depth is at least 0.5.

39. The composition of claim 38, wherein the composition comprises no more than four polynucleotide blockers.
40. The composition of claim 38 or 39, wherein the polynucleotide blocker comprises one or more nucleotide analogues.
41. The composition of any one of claims 38-40, wherein the polynucleotide blocker comprises one or more locked nucleic acids (LNAs) or one or more bridged nucleic acids (BNAs).
42. the composition of any one of claims 38-41, wherein the polynucleotide blocker comprises at least 5 nucleotide analogues.
43. The composition of any one of claims 38-41, wherein the polynucleotide blocker comprises at least 10 nucleotide analogues.
44. The composition of any one of claims 38-43, wherein the polynucleotide blocker has a T_m of at least 78 degrees C.
45. The composition of any one of claims 38-43, wherein the polynucleotide blocker has a T_m of at least 80 degrees C.
46. The composition of any one of claims 38-43, wherein the polynucleotide blocker has a T_m of at least 82 degrees C.
47. The composition of any one of claims 38-43, wherein the polynucleotide blocker has a T_m of 80-90 degrees C.
48. The composition of any one of claims 38-47, wherein the genomic library comprises genomic fragments from at least 2 different samples.
49. The composition of any one of claims 38-47, wherein the genomic library comprises genomic fragments from at least 10 different samples.
50. The composition of any one of claims 38-49, wherein the genomic library comprises genomic fragments from at least 2 non-identical index sequences.
51. The composition of any one of claims 38-49, wherein the genomic library comprises genomic fragments from at least 16 non-identical index sequences.
52. The composition of any one of claims 38-51, wherein the genomic library comprises genomic fragments further comprising at least one unique molecular identifier (UMI).
53. A method for enriching nucleic acids comprising:
 - (a) contacting the polynucleotide library of any one of claims 38-52 with a plurality of genomic fragments;
 - (b) enriching at least one genomic fragment that binds to the polynucleotide library to generate at least one enriched target polynucleotide; and
 - (c) sequencing the at least one enriched target polynucleotide.

54. The method of claim 53, wherein the off-target rate is less than 25%.
55. The method of claim 53, wherein the off-target rate is less than 20%.
56. The method of any one of claims 53-55, wherein the molar ratio between at least one polynucleotide blocker and the complementary adapter is no more than 5:1.
57. The method of any one of claims 53-55, wherein the molar ratio between at least one polynucleotide blocker and the complementary adapter is no more than 2:1.
58. The method of any one of claims 53-55, wherein the molar ratio between at least one polynucleotide blocker and the complementary adapter is no more than 1.5:1.
59. A composition for nucleic acid hybridization comprising:
 - a first polynucleotide library;
 - a second polynucleotide library, wherein at least one polynucleotide in the first library is at least partially complementary to at least one polynucleotide of the second library; and
 - an additive, wherein the additive reduces off-target hybridization of the at least one polynucleotide of the first library with the at least one polynucleotide of the second library by decreasing a local concentration of the first polynucleotide library or the second polynucleotide library at an air-liquid interface.
60. A composition for nucleic acid hybridization comprising:
 - a first polynucleotide library and a second polynucleotide library each comprising a plurality of polynucleotides, wherein at least one polynucleotide in the first library is at least partially complementary to at least one polynucleotide of the second library; and
 - an oil, wherein the oil reduces off-target hybridization of the at least one polynucleotide of the first library with the at least one polynucleotide of the second library by decreasing a local concentration of the first polynucleotide library or the second polynucleotide library at an air-liquid interface.
61. The composition of claim 59, wherein the additive is mineral oil, a nucleotide triphosphate, polyether, or urea.
62. The composition of any one of claims 59-61, wherein the additive is a hydrocarbon comprising at least six carbon atoms.
63. The composition of any one of claims 59-62, wherein the additive is silicon oil.
64. The composition of claim 59, wherein the oil is derived from plant sources.
65. The composition of any one of claims 59-63, wherein the composition further comprises dimethyl sulfoxide.

66. The composition of any one of claims 59-64, wherein the composition does not comprise a formamide.
67. The composition of claim 59 or 66, wherein the size of the first polynucleotide library is less than 10 million bases.
68. The composition of claim 59 or 66, wherein the size of the first polynucleotide library is less than 1 million bases.
69. The composition of claim 59 or 66, wherein the size of the first polynucleotide library is less than 0.5 million bases.
70. The composition of any one of claims 59-69, wherein the first polynucleotide library comprises at least one exon sequence.
71. The composition of any one of claims 59-70, wherein the first polynucleotide library comprises polynucleotides encoding for at least 10 genes.
72. The composition of any one of claims 59-70, wherein the first polynucleotide library comprises polynucleotides encoding for at least 100 genes.
73. The composition of any one of claims 59-72, wherein the first polynucleotide library comprises at least one genomic fragment.
74. The composition of any one of claims 59-72, wherein the first polynucleotide library comprises RNA, DNA, cDNA, or genomic DNA.
75. The composition of any one of claims 59-74, wherein the first polynucleotide library comprises genomic DNA.
76. A method for reducing off-target nucleic acid hybridization, comprising:
 - contacting a first polynucleotide library with a second polynucleotide library, wherein the first polynucleotide library and the second polynucleotide library each comprising a plurality of polynucleotides, and wherein at least one polynucleotide in the first library is at least partially complementary to at least one polynucleotide in the second library;
 - enriching at least one genomic fragment that binds to the second polynucleotide library to generate at least one enriched target polynucleotide, wherein enriching comprises at least one aspiration step, and wherein the at least one aspiration step comprises aspirating only liquid from the area near the air/liquid interface; and
 - sequencing the at least one enriched target polynucleotide.
77. A method for sequencing genomic DNA, comprising:

contacting a polynucleotide library with a plurality of genomic fragments and an additive to form a mixture, wherein the additive decreases a local concentration of the polynucleotide library or the genomic fragments in the mixture at an air-liquid interface; enriching at least one genomic fragment that binds to the polynucleotide library to generate at least one enriched target polynucleotide; and sequencing the at least one enriched target polynucleotide.

78. The method of claim 77, wherein the additive is oil, a nucleotide triphosphate, polyether, or urea.
79. The method of claim 78, wherein the additive is mineral oil.
80. The method of any one of claims 77-79, wherein the presence of the additive decreases off-target binding.
81. The method of any one of claims 77-80, wherein the presence of the additive decreases off-target binding by at least 10%.
82. The method of any one of claims 77-80, wherein the presence of the additive decreases off-target binding by at least 20%.
83. The method of any one of claims 77-80, wherein the presence of the additive decreases off-target binding by at least 30%.
84. The method of any one of claims 80-83, wherein the off-target binding is random off-target binding.
85. The method of any one of claims 77-84, wherein the size of the first polynucleotide library is less than 10 million bases.
86. The method of any one of claims 77-84, wherein the size of the first polynucleotide library is less than 1 million bases.
87. The method of any one of claims 77-84, wherein the size of the first polynucleotide library is less than 0.5 million bases.
88. The method of any one of claims 77-87, wherein the first polynucleotide library comprises at least one exon sequence.
89. The method of any one of claims 77-88, wherein the first polynucleotide library comprises polynucleotides encoding for at least 10 genes.
90. The method of any one of claims 77-88, wherein the first polynucleotide library comprises polynucleotides encoding for at least 100 genes.
91. The method of any one of claims 77-90, wherein the first polynucleotide library comprises at least one genomic fragment.

92. The method of any one of claims 77-91, wherein the first polynucleotide library comprises RNA, DNA, cDNA, or genomic DNA.
93. The method of any one of claims 77-92, wherein the first polynucleotide library comprises genomic DNA.
94. An aqueous composition for polynucleotide hybridization comprising
 - at least one solvent;
 - at least one surfactant; and
 - at least one thickening agent.
95. The composition of claim 94, wherein the at least one solvent comprises DMSO, DMF, DMAc, or HMPA.
96. The composition of claim 94 or 95, wherein the at least one solvent comprises acetone, methyl ethyl ketone, THF, or diethyl ether.
97. The composition of any one of claims 94-96, wherein the at least one surfactant comprises SDS, CTAB, tergitol, or sodium lauryl sulfate.
98. The composition of any one of claims 94-97, wherein the at least one surfactant comprises a trialkylammonium salt.
99. The composition of any one of claims 94-98, wherein the at least one thickening agent comprises cellulose, starch, or a derivative thereof.
100. The composition of any one of claims 94-99, wherein the at least one thickening agent comprises dextran sulfate, hydroxymethylcellulose, hydroxyethylcellulose, polyethylene glycol, or ficoll.
101. The composition of any one of claims 94-100, wherein the at least one thickening agent is 10-40% (w/v).
102. The composition of any one of claims 94-101, wherein the at least one surfactant is 0.001-0.1% (w/v).
103. The composition of any one of claims 94-102, wherein the at least one solvent is 10-30% (v/v) in water.
104. A method for sequencing genomic DNA, comprising:
 - contacting a polynucleotide library with a plurality of genomic fragments and a composition of any one of claims 94-103, wherein contacting occurs for less than 8 hours;
 - enriching at least one genomic fragment that binds to the polynucleotide library to generate at least one enriched target polynucleotide; and
 - sequencing the at least one enriched target polynucleotide.
105. The method of claim 104, wherein contacting occurs for less than 5 hours.

- 106. The method of claim 104, wherein contacting occurs for less than 1 hour.
- 107. The method of claim 104, wherein contacting occurs for less than 15 minutes.