

ENZYME- AND AMPLIFICATION-FREE SEQUENCING

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 62/082,883, filed November 21, 2014. The contents of the aforementioned patent application are incorporated herein by reference in their entireties.

SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing which has been submitted in ASCII format via EFS-Web and is hereby incorporated by reference in its entirety. Said ASCII copy, created on November 19, 2015, is named NATE-025_ST25.txt and is 20,860 bytes in size.

BACKGROUND OF THE INVENTION

[0003] There are currently a variety of methods for nucleic acid sequencing, *i.e.*, the process of determining the precise order of nucleotides within a nucleic acid molecule. Current methods require amplifying a nucleic acid enzymatically, *e.g.*, PCR, and/or by cloning. Further enzymatic polymerizations are required to produce a detectable signal by a light detection means. Such amplification and polymerization steps are costly and/or time-consuming. Thus, there is a need in the art for a method of nucleic acid sequencing that is amplification- and enzyme-free. The present invention addresses these needs.

SUMMARY

[0004] The present disclosure provides sequencing probes, methods, kits, and apparatuses that provide enzyme-free, amplification-free, and library-free nucleic acid sequencing that has long-read-lengths and with low error rate. Moreover, the methods, kits, and apparatuses have rapid sample-to-answer capability. These features are particularly useful for sequencing in a clinical setting.

[0005] Provided herein are sequencing probes comprising a target binding domain and a barcode domain. The target binding domain and the barcode domain may be operably linked, *e.g.*, covalently linked. A sequencing probe optionally comprises a spacer between the target binding domain and the barcode domain. The spacer can be any polymer with appropriate mechanical

complex's complementary nucleic acid or via the "extra-handle." Thus, for example, the phrase "binding to the first attachment region . . . a first complementary nucleic acid molecule of a first reporter complex" would be understood according to its plain meaning and also understood to mean "binding to the first attachment region . . . an 'extra handle' of a first reporter complex."

[0009] In examples, the terms "barcode domain" and "synthetic backbone" are synonymous.

[0010] Provided herein is a method for sequencing a nucleic acid using a sequencing probe of the present disclosure. The method comprises steps of: (1) hybridizing at least one sequencing probe, of the present disclosure, to an target nucleic acid that is immobilized (*e.g.*, at one, two, three, four, five, six, seven, eight, nine, ten or more positions) to a substrate; (2) binding to the first attachment region a first complementary nucleic acid molecule (RNA or DNA) which has a detectable label (*e.g.*, a fluorescent label) or a first complementary nucleic acid molecule of a first reporter complex comprising detectable labels (*e.g.*, fluorescent labels); (3) detecting the detectable label(s), and (4) identifying the position and identity of the first nucleotide in the immobilized target nucleic acid. Optionally, the immobilized target nucleic acid is elongated prior to being bound by the probe. The method further comprises steps of: (5) contacting the first attachment region (with or without one or two flanking single-stranded polynucleotides) with a first hybridizing nucleic acid molecule lacking a detectable label, thereby unbinding the first complementary nucleic acid molecule having a detectable label or the first complementary nucleic acid molecule of a first reporter complex comprising detectable labels and binding to, at least, the first attachment region a first hybridizing nucleic acid lacking a detectable label; (6) binding to the second attachment region a second complementary nucleic acid molecule having a detectable label or a complementary nucleic acid molecule of a second reporter complex comprising detectable labels; (7) detecting the detectable label(s); and (8) identifying the position and identity of the second nucleotide in the immobilized target nucleic acid. Steps (5) to (8) are repeated until each nucleotide in the immobilized target nucleic acid and corresponding to the target binding domain has been identified. Steps (5) and (6) may occur concurrently or sequentially. Each (*e.g.*, first, second, third, fourth, fifth, sixth, seventh, eighth, ninth, tenth, or higher) complementary nucleic acid molecule (having a detectable label or part of a reporter complex) has the same nucleic acid sequence as its corresponding (*i.e.*, first, second, third, fourth, fifth, sixth, seventh, eighth, ninth, tenth, or higher) hybridizing nucleic acid molecule lacking a detectable label. The target nucleic acid is immobilized to a substrate by binding a first

binding domain of sequencing probes in the second population of sequencing probes. Steps (16) and (17) may occur concurrently or sequentially.

[0012a] In one aspect, there is provided a method for sequencing a nucleic acid comprising steps of:

(1) hybridizing at least one first population of first sequencing probes comprising a plurality of the sequencing probes to a target nucleic acid that is immobilized to a substrate,

wherein the first sequencing probes comprise a target binding domain and a barcode domain;

wherein said target binding domain comprises at least eight nucleotides and is capable of binding a target nucleic acid, wherein at least two nucleotides in the target binding domain are universal bases capable of binding to any base within the target nucleic acid;

wherein said barcode domain comprises a synthetic backbone, said barcode domain comprising at least six attachment positions, each attachment position comprising at least one attachment region, said attachment region comprising at least one nucleic acid sequence capable of being bound by a complementary nucleic acid molecule,

wherein each attachment position of the at least six attachment positions corresponds to one nucleotide in the target binding domain and each of the at least six attachment positions have a different nucleic acid sequence, and

wherein said nucleic acid sequence of each position of the at least six attachment positions determines the position and identity of the corresponding one nucleotide in said target nucleic acid that is bound by said target binding domain, and

wherein the universal bases precede and follow the nucleotides corresponding to positions in the barcode domain, and

wherein the target nucleic acid is immobilized to the substrate at one or more positions;

(2) binding a first complementary nucleic acid molecule comprising a detectable label or a first complementary nucleic acid molecule of a first reporter complex comprising a detectable label to a first attachment position of the at least six attachment positions;

(3) detecting the detectable label of the bound first complementary nucleic acid molecule or the detectable label of the bound first complementary nucleic acid molecule of the first reporter complex;

(4) identifying the position and identity of a first nucleotide in the immobilized target nucleic acid;

(5) binding to the first attachment position a first hybridizing nucleic acid molecule lacking a detectable label, thereby unbinding the first complementary nucleic acid molecule comprising a detectable label or the first complementary nucleic acid molecule of a first reporter complex comprising a detectable label;

(6) binding a second complementary nucleic acid molecule comprising a detectable label or a second complementary nucleic acid molecule of a second reporter complex comprising a detectable label to a second attachment position of the at least six attachment positions;

(7) detecting the detectable label of the bound second complementary nucleic acid molecule or the detectable label of the bound second complementary nucleic acid molecule of the second reporter complex;

(8) identifying the position and identity of a second nucleotide in the immobilized target nucleic acid;

(9) repeating steps (5) to (8) until each attachment position of the at least six attachment positions have been bound by a complementary nucleic acid molecule comprising a detectable label or a complementary nucleic acid molecule of a reporter complex comprising a detectable label, and the detectable label of the bound complementary nucleic acid molecule or the detectable label of the bound complementary nucleic acid molecule of a reporter complex has been detected, thereby identifying the linear order of at least six nucleotides for at least a first region of the immobilized target nucleic acid that was hybridized by the target binding domain of the sequencing probe;

(10) de-hybridizing the at least one first population of first sequencing probes from the immobilized target nucleic acid;

(11) removing each de-hybridized sequencing probe of the first population;

(12) hybridizing at least one second population of second sequencing probes comprising a plurality of the sequencing probes to the immobilized target nucleic acid,

wherein the second sequencing probes comprise a target binding domain and a barcode domain,

wherein said target binding domain comprises at least eight nucleotides and is capable of binding a target nucleic acid, wherein at least two nucleotides in the target binding domain are universal bases capable of binding to any base within the target nucleic acid;

wherein said barcode domain comprises a synthetic backbone, said barcode domain comprising at least six attachment positions, each attachment position comprising at least one attachment region, said attachment region comprising at least one nucleic acid sequence capable of being bound by a complementary nucleic acid molecule,

wherein each attachment position of the at least six attachment positions corresponds to one nucleotide in the target binding domain and each of the at least six attachment positions have a different nucleic acid sequence, and

wherein said nucleic acid sequence of each position of the at least six attachment positions determines the position and identity of the corresponding one nucleotide in said target nucleic acid that is bound by said target binding domain, and

wherein the universal bases precede and follow the nucleotides corresponding to positions in the barcode domain, and

wherein the sequencing probes of the second population hybridizes to at least one region of the target nucleic acid that was hybridized to the sequencing probes of the at least first population, and

wherein each sequencing probe in the second population de-hybridizes from the immobilized target nucleic acid under about the same conditions and de-hybridizes from the immobilized target nucleic acid under different conditions from the sequencing probes in the first population;

(13) binding a first complementary nucleic acid molecule comprising a detectable label or a first complementary nucleic acid molecule of a first reporter complex comprising a detectable label to a first attachment position of the at least six attachment positions;

(14) detecting the detectable label of the bound first complementary nucleic acid molecule or the detectable label of the bound first complementary nucleic acid molecule of the first reporter complex;

(15) identifying the position and identity of a first nucleotide in the immobilized target nucleic acid;

(16) binding to the first attachment position a first hybridizing nucleic acid molecule lacking a detectable label, thereby unbinding the first complementary nucleic acid molecule comprising a detectable label or the first complementary nucleic acid molecule of a first reporter complex comprising a detectable label;

(17) binding a second complementary nucleic acid molecule comprising a detectable label or a second complementary nucleic acid molecule of a second reporter complex comprising a detectable label to a second attachment position of the at least six attachment positions;

(18) detecting the detectable label of the bound second complementary nucleic acid molecule or the detectable label of the bound second complementary nucleic acid molecule of the second reporter complex;

(19) identifying the position and identity of a second nucleotide in the immobilized target nucleic acid;

(20) repeating steps (16) to (19) until each attachment position of the at least six attachment positions have been bound by a complementary nucleic acid molecule comprising a detectable label or a complementary nucleic acid molecule of a reporter complex comprising a detectable label, and the detectable label of the bound complementary nucleic acid molecule or the detectable label of the bound complementary nucleic acid molecule of a reporter complex has been detected, thereby identifying the linear order of at least six nucleotides for at least a second region of the immobilized target nucleic acid that was hybridized by the target binding domain of the sequencing probe; and

(21) de-hybridizing the at least one second population of second sequencing probes from the immobilized target nucleic acid.

[0013] Each sequencing probe in the second population may de-hybridize from the immobilized target nucleic acid at a different condition (*e.g.*, a higher temperature, higher level of chaotropic agent, higher salt concentration, higher flow rate, and different pH) than the average condition for which the sequencing probes in the first population de-hybridize from the target nucleic acid.

[0014] However, when more than two populations of probes are used, then probes in two sequential populations may de-hybridize at different conditions and probes in non-sequential populations may de-hybridize at similar conditions. As an example, probes in a first population and third population may de-hybridize under similar conditions. In examples, sequential populations of probes de-hybridized at increasingly more stringent conditions (*e.g.*, higher levels of chaotropic agent, salt concentration, and temperature). For a microfluidic device, using temperature as an example, a first population of probes may remain hybridized at a first temperature but de-hybridize at a second temperature, which is higher than the first. A second population of probes may remain hybridized at the second temperature but de-hybridize at a third temperature, which is higher than the second. In this example, solutions (comprising reagents required by the present method) flowing over a target nucleic acid for initial probe populations are at a lower temperature than solutions flowing over the target nucleic acid for later probe populations.

[0015] In some examples, after a population of probes has been used, the population of probes is de-hybridized from the target nucleic acid and a new aliquot of the same population of probes is used. For example, after a first population of probes has been hybridized, detected, and de-hybridized, a subsequent aliquot of the first population of probes is hybridized. Alternately, as an example, a first population of probes may be de-hybridized and replaced with a second population of probes; once the second population has been detected and de-hybridized, a subsequent aliquot of the first population of probes is hybridized to the target nucleic acid. Thus, a probe in the subsequent population may hybridize to a region of the target nucleic acid that had been previously sequenced (thereby gaining duplicative and/or confirmatory sequence information) or a probe in the subsequent population may hybridize to a region of the target nucleic acid that had not previously been sequenced (thereby gaining new sequence information). Accordingly, a population of probes may be re-aliquoted when a prior read was

[0021] Alternately or additionally, the concentration of sequencing probes in a population may be increased to increase coverage of probes in a specific region of a target nucleic acid. The concentration of sequencing probes may be reduced to decrease coverage of probes in a specific region of a target nucleic acid, *e.g.*, to above the resolution limit of the sequencing apparatus.

[0022] The methods for sequencing a nucleic acid further comprises steps of assembling each identified linear order of nucleotides for each region of the immobilized target nucleic acid, thereby identifying a sequence for the immobilized target nucleic acid. Steps of assembling use a non-transitory computer-readable storage medium with an executable program stored thereon which instructs a microprocessor to arrange each identified linear order of nucleotides, thereby obtaining the sequence of the nucleic acid. Assembling can occur in “real time”, *i.e.*, while data is being collected from sequencing probes rather than after all data has been collected.

[0023] The target nucleic acid, *i.e.*, that is sequenced, may be between about 4 and 1,000,000 nucleotides in length. The target may include a whole, intact chromosome or a fragment thereof either of which is greater than 1,000,000 nucleotides in length.

[0024] Provided herein is an apparatus for performing a method as described herein.

[0025] Provided herein are kits including sequencing probes of the present disclosure and for performing methods of the present disclosure. In examples, the kits include a substrate capable of immobilizing a nucleic acid via a capture probe, a plurality of sequencing probes of the present disclosure, at least one capture probe, at least one complementary nucleic acid molecule having a detectable label, at least one complementary nucleic acid molecule which lacks a detectable label, and instructions for use. In examples, the kit comprises about or at least 4096 unique sequencing probes. 4096 is the minimum number of unique probes necessary to include each possible hexameric combination (*i.e.*, for probes each having six attachment regions in the barcode domains). Here, “4096” is achieved since there are four nucleotides options for six positions: 4^6 . For a set of probes having four attachment regions in the barcode domains, only 256 (*i.e.*, 4^4) unique probes will be needed. For a set of probes having eight nucleotides in their target binding domains, 4^8 (*i.e.*, 65,536) unique probes will be needed. For a set of probes having ten nucleotides in their target binding domains, 4^{10} (*i.e.*, 1,048,576) unique probes will be needed.

[0026] In examples, the kit comprises about or at least twenty four distinct complementary nucleic acid molecule having a detectable label and about or at least twenty four distinct

What is claimed is:

1. A method for sequencing a nucleic acid comprising steps of:

(1) hybridizing at least one first population of first sequencing probes comprising a plurality of the sequencing probes to a target nucleic acid that is immobilized to a substrate,

wherein the first sequencing probes comprise a target binding domain and a barcode domain;

wherein said target binding domain comprises at least eight nucleotides and is capable of binding a target nucleic acid, wherein at least two nucleotides in the target binding domain are universal bases capable of binding to any base within the target nucleic acid;

wherein said barcode domain comprises a synthetic backbone, said barcode domain comprising at least six attachment positions, each attachment position comprising at least one attachment region, said attachment region comprising at least one nucleic acid sequence capable of being bound by a complementary nucleic acid molecule,

wherein each attachment position of the at least six attachment positions corresponds to one nucleotide in the target binding domain and each of the at least six attachment positions have a different nucleic acid sequence, and

wherein said nucleic acid sequence of each position of the at least six attachment positions determines the position and identity of the corresponding one nucleotide in said target nucleic acid that is bound by said target binding domain, and

wherein the universal bases precede and follow the nucleotides corresponding to positions in the barcode domain, and

wherein the target nucleic acid is immobilized to the substrate at one or more positions;

(2) binding a first complementary nucleic acid molecule comprising a detectable label or a first complementary nucleic acid molecule of a first reporter complex comprising a detectable label to a first attachment position of the at least six attachment positions;

(3) detecting the detectable label of the bound first complementary nucleic acid molecule or the detectable label of the bound first complementary nucleic acid molecule of the first reporter complex;

(4) identifying the position and identity of a first nucleotide in the immobilized target nucleic acid;

(5) binding to the first attachment position a first hybridizing nucleic acid molecule lacking a detectable label, thereby unbinding the first complementary nucleic acid molecule comprising a detectable label or the first complementary nucleic acid molecule of a first reporter complex comprising a detectable label;

(6) binding a second complementary nucleic acid molecule comprising a detectable label or a second complementary nucleic acid molecule of a second reporter complex comprising a detectable label to a second attachment position of the at least six attachment positions;

(7) detecting the detectable label of the bound second complementary nucleic acid molecule or the detectable label of the bound second complementary nucleic acid molecule of the second reporter complex;

(8) identifying the position and identity of a second nucleotide in the immobilized target nucleic acid;

(9) repeating steps (5) to (8) until each attachment position of the at least six attachment positions have been bound by a complementary nucleic acid molecule comprising a detectable label or a complementary nucleic acid molecule of a reporter complex comprising a detectable label, and the detectable label of the bound complementary nucleic acid molecule or the detectable label of the bound complementary nucleic acid molecule of a reporter complex has been detected, thereby identifying the linear order of at least six nucleotides for at least a first region of the immobilized target nucleic acid that was hybridized by the target binding domain of the sequencing probe;

(10) de-hybridizing the at least one first population of first sequencing probes from the immobilized target nucleic acid;

(11) removing each de-hybridized sequencing probe of the first population;

(12) hybridizing at least one second population of second sequencing probes comprising a plurality of the sequencing probes to the immobilized target nucleic acid,

wherein the second sequencing probes comprise a target binding domain and a barcode domain,

wherein said target binding domain comprises at least eight nucleotides and is capable of binding a target nucleic acid, wherein at least two nucleotides in the target binding domain are universal bases capable of binding to any base within the target nucleic acid;

wherein said barcode domain comprises a synthetic backbone, said barcode domain

comprising at least six attachment positions, each attachment position comprising at least one attachment region, said attachment region comprising at least one nucleic acid sequence capable of being bound by a complementary nucleic acid molecule,

wherein each attachment position of the at least six attachment positions corresponds to one nucleotide in the target binding domain and each of the at least six attachment positions have a different nucleic acid sequence, and

wherein said nucleic acid sequence of each position of the at least six attachment positions determines the position and identity of the corresponding one nucleotide in said target nucleic acid that is bound by said target binding domain, and

wherein the universal bases precede and follow the nucleotides corresponding to positions in the barcode domain, and

wherein the sequencing probes of the second population hybridizes to at least one region of the target nucleic acid that was hybridized to the sequencing probes of the at least first population, and

wherein each sequencing probe in the second population de-hybridizes from the immobilized target nucleic acid under about the same conditions and de-hybridizes from the immobilized target nucleic acid under different conditions from the sequencing probes in the first population;

(13) binding a first complementary nucleic acid molecule comprising a detectable label or a first complementary nucleic acid molecule of a first reporter complex comprising a detectable label to a first attachment position of the at least six attachment positions;

(14) detecting the detectable label of the bound first complementary nucleic acid molecule or the detectable label of the bound first complementary nucleic acid molecule of the first reporter complex;

(15) identifying the position and identity of a first nucleotide in the immobilized target nucleic acid;

(16) binding to the first attachment position a first hybridizing nucleic acid molecule lacking a detectable label, thereby unbinding the first complementary nucleic acid molecule comprising a detectable label or the first complementary nucleic acid molecule of a first reporter complex comprising a detectable label;

(17) binding a second complementary nucleic acid molecule comprising a detectable

label or a second complementary nucleic acid molecule of a second reporter complex comprising a detectable label to a second attachment position of the at least six attachment positions;

(18) detecting the detectable label of the bound second complementary nucleic acid molecule or the detectable label of the bound second complementary nucleic acid molecule of the second reporter complex;

(19) identifying the position and identity of a second nucleotide in the immobilized target nucleic acid;

(20) repeating steps (16) to (19) until each attachment position of the at least six attachment positions have been bound by a complementary nucleic acid molecule comprising a detectable label or a complementary nucleic acid molecule of a reporter complex comprising a detectable label, and the detectable label of the bound complementary nucleic acid molecule or the detectable label of the bound complementary nucleic acid molecule of a reporter complex has been detected, thereby identifying the linear order of at least six nucleotides for at least a second region of the immobilized target nucleic acid that was hybridized by the target binding domain of the sequencing probe; and

(21) de-hybridizing the at least one second population of second sequencing probes from the immobilized target nucleic acid.

2. The method of claim 1, wherein steps (5) and (6) occur sequentially or concurrently.
3. The method of claim 1, wherein steps (16) and (17) occur sequentially or concurrently.
4. The method of claim 1, wherein the first complementary nucleic acid molecule and the first hybridizing nucleic acid molecule lacking a detectable label comprise the same nucleic acid sequence.
5. The method of claim 1, wherein said synthetic backbone comprises single-stranded DNA.
6. The method of claim 1, wherein said sequencing probes comprises a double-stranded DNA spacer between the target binding domain and the barcode domain.

7. The method of claim 1, wherein the number of nucleotides in a target binding domain is at least two more than the number of attachment regions in the corresponding barcode domain.
8. The method of claim 1, wherein each position in a barcode domain has: (a) the same number of attachment regions; (b) one attachment region; or (c) more than one attachment region.
9. The method of claim 1, wherein each complementary nucleic acid molecule is directly linked to a primary nucleic acid molecule or is indirectly linked to a primary nucleic acid molecule via a nucleic acid spacer.
10. The method of claim 1, wherein each complementary nucleic acid molecule for each attachment position comprises between 8 nucleotides and 20 nucleotides.
11. The method of claim 10, wherein each complementary nucleic acid molecule comprises 12 nucleotides.
12. The method of claim 9, wherein each primary nucleic acid molecule is hybridized to at least one secondary nucleic acid molecule.
13. The method of claim 12, wherein the secondary nucleic acid molecule or molecules comprise at least one detectable label.
14. The method of claim 12, wherein each secondary nucleic acid molecule is hybridized to at least one tertiary nucleic acid molecule comprising at least one detectable label.
15. The method of claim 1, wherein one or more attachment regions in the barcode domain are adjacent to at least one flanking single-stranded polynucleotide.
16. The method of claim 1, wherein the first hybridizing nucleic acid molecule lacking a detectable label comprises a nucleic acid sequence complementary to a flanking single-stranded

polynucleotide adjacent to the first attachment region.

17. The method of claim 1, further comprising steps of assembling each identified linear order of nucleotides in the at least first region and at least second region of the immobilized target nucleic acid, thereby identifying a sequence for the immobilized target nucleic acid.