



(51) International Patent Classification:

C12Q 1/68 (2006.01) G06N 3/12 (2006.01)
G06N 3/00 (2006.01)

(21) International Application Number:

PCT/US20 17/02975 1

(22) International Filing Date:

27 April 2017 (27.04.2017)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/328,322 27 April 2016 (27.04.2016) US

(71) Applicant: PRESIDENT AND FELLOWS OF HARVARD COLLEGE [US/US]; 17 Quincy Street, Cambridge, MA 02138 (US).

(72) Inventors: LEE, Henry, Hung-yi; 127 Charles Street, Apt. 4, Boston, MA 021 14 (US). KALHOR, Reza; 40 Station Road, Apt. 2, Brookline, MA 02445 (US). CHURCH, George, M.; 218 Kent Street, Brookline, MA 02446 (US).

(74) Agent: IWANICKI, John, P.; Banner & Witcoff, Ltd., 28 State Street, Suite 1800, Boston, MA 02109 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM,

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

(84) Designated States (unless otherwise indicated, for every kind of regional protection available):

ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

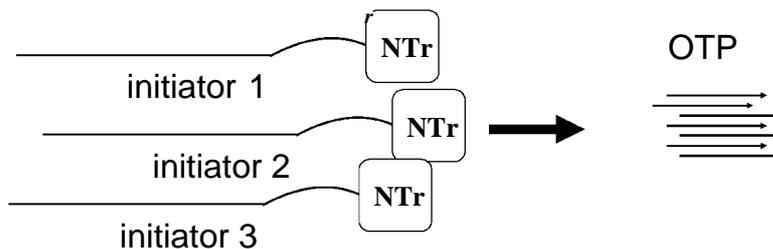
Published:

— with international search report (Art. 21(3))

(54) Title: METHOD OF SECURE COMMUNICATION VIA NUCLEOTIDE POLYMERS

Fig. 1

Generation of one-time pad of nucleotide keys



- A pool of initiator oligonucleotides is synthesized. These initiators have unique nucleotide barcodes useful for addressing each and later, it's corresponding one-time key (i.e., ATCG = key 1, CTAG = key 2, etc).
- NTrs are used to extend each initiator to add a random sequence of nucleotides. Then, a universal 3' primer is ligated to every strand.
- PCR is performed, using universal primers, to amplify a subset (or all) of the strands, such that any aliquot of this amplified mixture will have full representation of each key. Aliquots of this amplified mixture are distributed between communicating parties.
- Each aliquot becomes a one-time pad of nucleotide keys, representing a collection of initiator oligonucleotides with added random nucleotide sequences. The random sequences represents the collection of keys in the one-time pad.

(57) Abstract: Methods for making securing communications using a nucleic acid encoding a format of information are provided.



METHOD OF SECURE COMMUNICATION VIA NUCLEOTIDE POLYMERS**RELATED APPLICATION DATA**

This application claims priority to U.S. Provisional Application No. 62/328,322 filed on April 27, 2016, which is hereby incorporated herein by reference in its entirety for all purposes.

STATEMENT OF GOVERNMENT INTERESTS

This invention was made with government support under grant number 1 RO1 MH103910-01 awarded by National Institutes of Health. The government has certain rights in the invention.

FIELD

The present invention relates in general to methods of communicating a format of information using nucleotide sequences in a secure manner between a sender and a recipient.

BACKGROUND

Methods of representing a format of information using nucleotide sequences are known. See for example US-2015-0269313. Other methods of encoding digital information into DNA are described in J. Davis, *Art Journal* 55, 70-74 (1996).

SUMMARY

The disclosure provides methods of using a nucleic acid sequence or sequences, such as DNA) including nucleotides as a secure medium for information transmission between a sender and a recipient. According to one aspect, the nucleic acid sequences or oligonucleotide sequences that encode a format of information and which may be transmitted from a Sender to an Intended Recipient and are intended to be secured by the methods

described herein are in single stranded form, i.e., they are single-stranded nucleic acids. The information to be represented by a nucleic acid sequence or transmitted in the form of a nucleic acid sequence may be in text format, image format, video format or audio format. The disclosure provides for the secure representation or transmission of a format of information encoded by a nucleic acid sequence or sequences insofar as the format of information (which may be referred to herein as a "message") is encoded by a nucleic acid sequence or sequences which are then identified, selected, amplified and sequenced and the sequence is then converted into bit sequences and the bit sequences are converted to the format of information, which may then be visualized if a text, image or video format or heard, if an audio format.

The disclosure provides for the secure representation or transmission of information by using a 3' lock and key system for enabling amplification or sequencing of a nucleic acid encoding a message, such as a single-stranded nucleic acid encoding a message. It is to be understood that reference to a nucleic acid herein includes reference to a single-stranded nucleic acid. The encoded message includes a 3' lock moiety which requires optional processing to activate the 3' lock moiety and a key moiety to be combined with the 3' lock moiety to enable amplification or sequencing.

The disclosure provides for the secure representation or transmission of information by using a 3' nucleic acid sequence randomly produced, such as by using a template independent polymerase, as described herein, to create a "one-time pad", as that phrase is understood in the encryption field.

Further features and advantages of certain embodiments of the present invention will become more fully apparent in the following description of embodiments and drawings thereof, and from the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee. The foregoing and other features and advantages of the present embodiments will be more fully understood from the following detailed description of illustrative embodiments taken in conjunction with the accompanying drawings in which:

Fig. 1 is an illustration directed to a one-time pad of nucleotide keys.

Fig. 2 is an illustration directed to a one-time pad of nucleotide keys.

Fig. 3 is an illustration directed to a one-time pad of nucleotide keys.

Fig. 4 is an illustration directed to a one-time pad of nucleotide keys.

Fig. 5 is an illustration directed to a one-time pad of nucleotide keys.

Fig. 6 is an illustration directed to a one-time pad of dictionary keys.

Fig. 7 is an illustration directed to a one-time pad of dictionary keys.

Fig. 8 is an illustration directed to a one-time pad of dictionary keys.

Fig. 9 is an illustration directed to a one-time pad of dictionary keys.

Fig. 10 is an illustration directed to a one-time pad of dictionary keys.

DETAILED DESCRIPTION

The present disclosure provides methods for secure communication via nucleotide polymers with primer-based lock and key pairs as described herein and biological one-time pads as described herein.

In general, a nucleic acid encodes for a format of information as known to those of skill in the art. As an example, the format of information is converted to megabits which may form a bit stream. The megabits are then encoded into oligonucleotides. Depending on the

length of the format of information, the oligonucleotide may include a data block sequence, an address sequence (such as a barcode sequence) specifying the location of the data block in the bit stream. As an example, one bit per base is encoded. According to this aspect, a single message may be encoded in a plurality of ways, i.e., A or C for zero, G or T for the number 1. As an example, a format of information is converted to a bit stream, bit sequences are encoded into corresponding oligonucleotide sequences, the oligonucleotide sequences are synthesized, the oligonucleotide sequences are amplified and/or sequenced, the sequenced oligonucleotide sequences are decoded into bit sequences, the bit sequences may be assembled into a bit stream, and the bit stream or bit is converted into the format of information. As an example an html format of information, such as an html message or book with text and/or images, is converted to bits, i.e. zeros and ones, as commonly understood. Other formats of information that can be converted to bits are known to those of skill in the art.

In general, the disclosure provides for methods of synthesizing the nucleic acid encoding the format of information, amplifying the nucleic acid encoding the information and sequencing the nucleic acid encoding the information using methods known to those of skill in the art. As an example, a portion of an html format of information to be converted into bits may be referred to as a byte portion. The bit sequence is then converted (encoded) to a sequence of nucleotides, i.e., an oligonucleotide or DNA using a 1 bit per base encoding (ac=0; TG=1) to form a corresponding encoded oligonucleotide sequence, i.e. the oligonucleotide sequence corresponds to or encodes for the bit sequence. A single message may be encoded in a plurality of ways using one bit per base encoding, i.e., A or C for zero, G or T for the number 1. Other combinations are envisioned such as A or G for zero, C or T for the number 1 or A or T for zero, G or C for the number 1. A plurality of bit sequences

are created corresponding to a portion of or the entire html format of information. A plurality of corresponding encoded oligonucleotide sequences may be created which together may be referred to as a library. The library of encoded oligonucleotide sequences represents the html format of information. The encoded oligonucleotide sequences are then synthesized using methods known to those of skill in the art, such as using a DNA microchip. The synthesized oligonucleotides are then amplified using methods known to those of skill in the art to form a library of oligonucleotides. The library of oligonucleotides is then sequenced using methods known to those of skill in the art, such as next-generation sequencing methods. High throughput, next-generation techniques are used in both DNA synthesis and sequencing to allow for encoding and decoding of large amounts of information. The sequenced oligonucleotides are then converted into bit sequences corresponding to the html format of information. The bit sequences can be converted to the format of information using methods known to those of skill in the art. The format of information can be visualized or listened to or displayed using methods and devices known to those of skill in the art.

In general, the disclosure provides for the use of molecules, such as nucleotides, as binary bits of information. The nucleotides may be representative of a binary state, such as zero or one, and sequences of nucleotides representing sequences of binary states, such as zeros or ones, may be representative of text, an image, a video or an audio format. In this manner, a written material, a picture, a video with an audio component or an audio recording or any other medium of expression, may be encoded using nucleic acids as representative of bits. According to certain aspects, information is converted into binary bits, such as according to ASCII code, using a computer and appropriate software for example, which is a series of zeros and ones representative of the information. It is to be understood that the information may be converted to other coded bits of information, as is known in the art. A

series of nucleotides is then determined, such as by using a computer and appropriate software, which is representative of the series of coded bits of information, such as zeros and ones. The series of nucleotides are then synthesized and stored, for example, on a storage media or in a vessel. When the information is to be accessed, the series of nucleotides are determined and then translated, such as by using a computer and appropriate software, into a series of zeros and ones which is then translated into the information, for example using a computer and appropriate software. In this manner, aspects of the present disclosure are directed to the use of nucleic acids, whether fully- or partially single stranded, double-stranded, or multi-stranded, to encode information. According to one aspect, the nucleic acids are included on a support substrate whether in an ordered or random manner.

Secure Communication Using a 3' Lock and Key

The present disclosure provides for an oligonucleotide including a 5' sequence such as a common 5' sequence, for amplification and/or sequencing of the oligonucleotide. The disclosure provides for a 3' sequence or moiety as a lock moiety that requires an associated key moiety for amplification or sequencing. Accordingly, the present disclosure provides a mixture of nucleic acids including one or more true message nucleic acids having a plurality of nucleotides corresponding to a plurality of bit sequences representative of a true message, wherein the one or more nucleic acids include a lock 3' message amplification or sequencing enabling moiety unique to the one or more true message nucleic acids and having an associated key moiety for amplification or sequencing, and one or more optional dummy nucleic acids, and wherein the one or more true message nucleic acids are selectively amplifiable or sequenceable over the one or more dummy nucleic acids when the lock 3' message amplification or sequencing enabling moiety is combined with the associated key

moiety.

A "true message nucleic acid" is one that encodes for a desired or "true" message that a sender wishes to transmit to an intended recipient. A "dummy nucleic acid" encodes for a false message or one that the sender does not wish to transmit to an intended recipient. A dummy nucleic acid may be present in the mixture or it may be absent. The one or more dummy nucleic acids may include a dummy 3' message amplification or sequencing enabling moiety unique to the one or more dummy nucleic acids.

The 3' lock message amplification or sequencing enabling moiety for the true message is one that enables amplification or sequencing of the true message when paired with an associated key moiety. An exemplary lock and key pair can be a 3' priming sequence on the true message nucleic acid and the associated key can be the complementary primer sequence, which when hybridized to the 3' priming or lock sequence enables amplification or sequencing of the true message nucleic acid. The lock 3' message amplification enabling moiety may include natural nucleotides, such as adenine (A), cytosine (C), guanine (G), and thymidine (T) and uracil (U) and the associated key moiety is the corresponding complementary key sequence of nucleotides which is capable of priming nucleotide polymerization in its own 3' direction and the lock's 5' direction.

The lock 3' message amplification enabling moiety may include one or more non-natural nucleotide analogues, or non-sequenceable nucleotide analogues, as described herein, and the associated key moiety is the corresponding complementary key sequence of natural or non-natural nucleotides which is capable of priming nucleotide polymerization in its own 3' and the lock's 5' direction. A non-sequenceable nucleotide analog is one whose complement is not one of the four natural nucleotides, and accordingly, inhibits sequencing, such as sequencing-by-synthesis or sequencing by ligation methods. Exemplary non-

sequenceable nucleotide analogues include deoxyinosine, 5-nitroindole, Iso-dC and Iso-dG. However, such non-natural or non-sequenceable nucleotide analogues may have complementary nucleotides and can be used in sequences that are intended to hybridize to one another, such as complementary primer pairs. For example Iso-dG and Iso-dC can act as cognates of each other if present in complementary primers. Accordingly, the lock 3' message amplification enabling moiety may be an amplification lock priming sequence and the associated key moiety is a corresponding complementary key primer sequence which may include natural or non-natural or non-sequenceable nucleotides.

The lock 3' message sequencing enabling moiety may include a protein or a protein binding site and the associated key moiety may be a binding protein that binds to the protein binding site and enables nanopore sequencing, as described herein, to the extent that the bound protein leads the nucleic acid sequence to and through a nanopore, as is known in the art.

The lock 3' message amplification or sequencing enabling moiety may be inactive insofar as the lock 3' message amplification or sequencing enabling moiety is required to be activated before it will enable amplification or sequencing when combined with the corresponding or associated key moiety. The lock 3' message amplification or sequencing enabling moiety may require activation before amplification or sequencing of the one or more true message nucleic acids including processing by chemical reaction, enzymatic reaction, heat, light, pH or other methods known to those of skill in the art to activate a chemical compound or to remove an inhibitor from a chemical compound, thereby activating the chemical compound. A lock 3' message amplification nor sequencing enabling moiety or its associated key moiety, such as cognate primer pairs, can be inactivated by one or more reversible chemical modifications that prevent it from being recognized by its cognate pair or

primer or any other pair or primer without reversal of said chemical modifications. An example of such a chemical modification would be attachment of chemical groups to the hydrogen-bond donor and acceptor positions of the base. Such modifications may be reversed by heat, reduction, oxidation, acid or base treatment, enzymatic treatment, etc. as is known in the art. Another example of such chemical modifications would be a branched primer in which the 3' end is covalently and reversibly attached to the base, the sugar, or the phosphate of an internal base. Such a primer would be inactive but can be linearized and activated. Reversion of the branch may be done by heat, reduction, oxidation, acid or base treatment, enzymatic treatment, etc., as is known in the art.

The lock 3' message amplification or sequencing enabling moiety may include a blocking or protecting moiety, which can be removed. A protecting group or protective group is introduced into a molecule by chemical modification of a functional group to obtain chemoselectivity in a subsequent chemical reaction. Removal of the protecting group is called deprotection, as is known to those of skill in the art. Protecting groups can be removed chemical treatment, enzymatic treatment, heat treatment, light treatment, reduction, oxidation, acid treatment, base treatment and the like. Exemplary removable protecting groups include alcohol protecting groups, amine protecting groups, carbonyl protecting groups, carboxylic acid protecting groups, and phosphate protecting groups known to those of skill in the art.

The lock 3' message amplification or sequencing enabling moiety may include a removable blocking or protecting moiety required to be removed before amplification or sequencing of the one or more true message nucleic acids.

The lock 3' message amplification or sequencing enabling moiety may be designed such that it becomes disabled upon amplification or sequencing efforts using other than

selected processing methods for activation or the associated key moiety. Such designs are known to those of skill in the art.

The lock 3' message amplification or sequencing promoting moiety may be a nucleic acid sequence including one or more non-sequenceable nucleotide analogues or one or more amplification or sequencing inhibiting bonds or non-copiable bonds between nucleic acids. Such non-sequenceable nucleotide analogues, i.e. nucleotides which are resistant to amplification or sequencing methods, are described herein and/or are known to those of skill in the art. Such amplification or sequencing inhibiting bonds or non-copiable bonds, i.e., bonds which are resistant to amplification or copying methods, between nucleic acids are described herein and/or are known to those of skill in the art. In standard primers, nucleotides are connected by a phosphodiester bond. Creating a primer in which nucleotides are connected with a different bond or a modified version of phosphodiester bond can make sequencing them impossible. For example a non-copiable primer can have amide, diphosphodiester, or phosphoramidate bonds instead of phosphodiester bonds in multiple positions. Also, as the distance between bases connected by these non-standard bonds is different from that of phosphodiester bond, they cannot be copied by a polymerase or sequenced by standard methods. However, they will serve as cognate to a primer that has the same modified bond in the complementary position.

The disclosure provides that each true message includes a unique sequence of nucleotides representing an encryption key with which the true message is decrypted, wherein the encryption key is positioned immediately 3' of the true message and 5' of the lock 3' message amplification or sequencing enabling moiety.

Accordingly, the disclosure provides a method of transmitting information from a sender to an intended recipient comprising transferring to the intended recipient a mixture

of nucleic acids including one or more true message nucleic acids having a plurality of nucleotides corresponding to a plurality of bit sequences representative of a true message, wherein the one or more nucleic acids include a lock 3' message amplification or sequencing enabling moiety unique to the one or more true message nucleic acids and having an associated key moiety for amplification or sequencing, and one or more optional dummy nucleic acids, and wherein the intended recipient possesses the associated key moiety, and wherein the one or more true message nucleic acids are selectively amplified or sequenced over the one or more dummy nucleic acids when the lock 3' message amplification or sequencing enabling moiety is combined with the associated key moiety. The disclosure provides that the mixture of nucleic acids includes one or more dummy nucleic acids. The disclosure provides that the one or more true message nucleic acids are sequenced, the nucleotides are converted to bit sequences and the bit sequences are converted to a format of information selected from the group consisting of a text format, an image format, a video format or an audio format.

The disclosure provides a method of securing a format of information encoded in one or more true message nucleic acids having a plurality of nucleotides corresponding to a plurality of bit sequences representative of a true message including adding to the one or more true message nucleic acids a lock 3' message amplification or sequencing enabling moiety unique to the one or more true message nucleic acids and having an associated key moiety for amplification or sequencing, and mixing the one or more true message nucleic acids with one or more dummy nucleic acids, and wherein the one or more true message nucleic acids are selectively amplifiable or sequenceable over the one or more dummy nucleic acids when the lock 3' message amplification or sequencing enabling moiety is combined with the associated key moiety.

The disclosure provides a method of securing a format of information encoded in one or more true message nucleic acids having a plurality of nucleotides corresponding to a plurality of bit sequences representative of a true message including combining the one or more true message nucleic acids with a mixture including a lock 3' message amplification or sequencing enabling moiety having a 5'-phosphate and/or a 3'-blocking moiety and dummy lock nucleic acid sequences, wherein the lock 3' message amplification or sequencing promoting moiety having a 5'-phosphate is ligated to the one or more true message nucleic acids at a 3'end, mixing the one or more true message nucleic acids with one or more dummy nucleic acids including dummy lock nucleic acid sequences, wherein the lock 3' message amplification or sequencing enabling moiety is unique to the one or more true message nucleic acids and having an associated key moiety for amplification or sequencing, and wherein the one or more true message nucleic acids are selectively amplifiable or sequenceable over the one or more dummy nucleic acids when the lock 3' message amplification or sequencing enabling moiety is combined with the associated key moiety. The disclosure provides that the lock 3' message amplification or sequencing enabling moiety includes one or more non-sequenceable nucleotide analogues.

Terms and symbols of nucleic acid chemistry, biochemistry, genetics, and molecular biology used herein follow those of standard treatises and texts in the field, e.g., Kornberg and Baker, DNA Replication, Second Edition (W.H. Freeman, New York, 1992); Lehninger, Biochemistry, Second Edition (Worth Publishers, New York, 1975); Strachan and Read, Human Molecular Genetics, Second Edition (Wiley-Liss, New York, 1999); Eckstein, editor, Oligonucleotides and Analogs: A Practical Approach (Oxford University Press, New York, 1991); Gait, editor, Oligonucleotide Synthesis: A Practical Approach (IRL Press, Oxford, 1984); and the like.

The following general descriptions are provided to support certain aspects of the methods described herein and are useful to those of skill in the art in practicing and understanding the present disclosure.

Bits

As used herein, the term "bit" is to be understood according to its common meaning to one of skill in the art. The term "bit" may be a contraction of "binary digit" and may refer to a basic capacity of information in computing and telecommunications. A "bit" represents either a first state or a second state, such as 1 or 0 (one or zero) only. The representation may be implemented, in a variety of systems, by means of a two state device.

Nucleic Acids and Nucleotides

As used herein, the terms "nucleic acid molecule," "nucleic acid sequence," "nucleic acid fragment" and "oligomer" are used interchangeably and are intended to include, but are not limited to, a polymeric form of nucleotides that may have various lengths, including either deoxyribonucleotides or ribonucleotides, or analogs thereof. A nucleic acid may be double-stranded or single-stranded. An exemplary nucleic acid for use in the methods of secure communication described herein is a single-stranded nucleic acid.

In general, the terms "nucleic acid molecule," "nucleic acid sequence," "nucleic acid fragment," "oligonucleotide" and "polynucleotide" are used interchangeably and are intended to include, but not limited to, a polymeric form of nucleotides that may have various lengths, either deoxyribonucleotides (DNA) or ribonucleotides (RNA), or analogs thereof. A oligonucleotide is typically composed of a specific sequence of four nucleotide bases:

adenine (A); cytosine (C); guanine (G); and thymine (T) (uracil (U) for thymine (T) when the polynucleotide is RNA). According to certain aspects, deoxynucleotides (dNTPs, such as dATP, dCTP, dGTP, dTTP) may be used. According to certain aspects, ribonucleotide triphosphates (rNTPs) may be used. According to certain aspects, ribonucleotide diphosphates (rNDPs) may be used.

The term "oligonucleotide sequence" is the alphabetical representation of a polynucleotide molecule; alternatively, the term may be applied to the polynucleotide molecule itself. This alphabetical representation can be input into databases in a computer having a central processing unit and used for bioinformatics applications such as functional genomics and homology searching. Oligonucleotides may optionally include one or more non-standard nucleotide(s), nucleotide analog(s) and/or modified nucleotides. The present disclosure contemplates any deoxyribonucleotide or ribonucleotide and chemical variants thereof, such as methylated, hydroxymethylated or glycosylated forms of the bases, and the like. According to certain aspects, natural nucleotides are used in the methods of making the nucleic acids. Natural nucleotides lack chain terminating moieties. According to another aspect, the methods of making the nucleic acids described herein do not use terminating nucleic acids or otherwise lack terminating nucleic acids, such as reversible terminators known to those of skill in the art. The methods are performed in the absence of chain terminating nucleic acids or wherein the nucleic acids are other than chain terminating nucleic acids.

Examples of modified nucleotides include, but are not limited to diaminopurine, S2T, 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-acetylcytosine, 5-(carboxyhydroxymethyl)uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine,

inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarbonylmethyluracil, 5-methoxyuracil, 2-methylthio-D46-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, 2,6-diaminopurine and the like. Nucleic acid molecules may also be modified at the base moiety (e.g., at one or more atoms that typically are available to form a hydrogen bond with a complementary nucleotide and/or at one or more atoms that are not typically capable of forming a hydrogen bond with a complementary nucleotide), sugar moiety or phosphate backbone. Nucleic acid molecules may also contain amine-modified groups, such as aminoallyl-dUTP (aa-dUTP) and aminohexylacrylamide-dCTP (aha-dCTP) to allow covalent attachment of amine reactive moieties, such as N-hydroxy succinimide esters (NHS).

Alternatives to standard DNA base pairs or RNA base pairs in the oligonucleotides of the present disclosure can provide higher density in bits per cubic mm, higher safety (resistant to accidental or purposeful synthesis of natural toxins), easier discrimination in photo-programmed polymerases, or lower secondary structure. Such alternative base pairs compatible with natural and mutant polymerases for de novo and/or amplification synthesis are described in Betz K, Malyshev DA, Lavergne T, Welte W, Diederichs K, Dwyer TJ, Ordoukhanian P, Romesberg FE, Marx A (2012) KlenTaq polymerase replicates unnatural base pairs by inducing a Watson-Crick geometry, *Nature Chem. Biol.* 8:612-614; See YJ, Malyshev DA, Lavergne T, Ordoukhanian P, Romesberg FE. *J Am Chem Soc.* 2011 Dec

14;133(49): 19878-88, Site-specific labeling of DNA and RNA using an efficiently replicated and transcribed class of unnatural base pairs; Switzer CY, Moroney SE, Benner SA. (1993) *Biochemistry*. 32(39): 10489-96. Enzymatic recognition of the base pair between isocytidine and isoguanosine; Yamashige R, Kimoto M, Takezawa Y, Sato A, Mitsui T, Yokoyama S, Hirao I. *Nucleic Acids Res.* 2012 Mar;40(6):2793-806. Highly specific unnatural base pair systems as a third base pair for PCR amplification; and Yang Z, Chen F, Alvarado JB, Benner SA. *J Am Chem Soc.* 2011 Sep 28; 133(38): 15 105-12, Amplification, mutation, and sequencing of a six-letter synthetic genetic system. Other non-standard nucleotides may be used such as described in Malyshev, D.A., et al., *Nature*, vol. 509, pp. 385-388 (15 May 2014) hereby incorporated by reference in its entirety.

Methods of Making Nucleic Acids

In certain exemplary embodiments, oligonucleotide sequences may be prepared using one or more of the phosphoramidite linkers and/or sequencing by ligation methods known to those of skill in the art. Oligonucleotide sequences may also be prepared by any suitable method, e.g., standard phosphoramidite methods such as those described herein below as well as those described by Beaucage and Carruthers ((1981) *Tetrahedron Lett.* 22: 1859) or the triester method according to Matteucci et al. (1981) *J. Am. Chem. Soc.* 103:3185), or by other chemical methods using either a commercial automated oligonucleotide synthesizer or high-throughput, high-density array methods known in the art (see U.S. Patent Nos. 5,602,244, 5,574,146, 5,554,744, 5,428,148, 5,264,566, 5,141,813, 5,959,463, 4,861,571 and 4,659,774, incorporated herein by reference in its entirety for all purposes). Pre-synthesized oligonucleotides may also be obtained commercially from a variety of vendors.

In certain exemplary embodiments, oligonucleotide sequences may be prepared using a variety of microarray technologies known in the art. Pre-synthesized oligonucleotide and/or polynucleotide sequences may be attached to a support or synthesized in situ using light-directed methods, flow channel and spotting methods, inkjet methods, pin-based methods and bead-based methods set forth in the following references: McGall et al. (1996) Proc. Natl. Acad. Sci. U.S.A. 93:13555; Synthetic DNA Arrays In Genetic Engineering, Vol. 20:111, Plenum Press (1998); Duggan et al. (1999) Nat. Genet. S21:10; Microarrays: Making Them and Using Them In Microarray Bioinformatics, Cambridge University Press, 2003; U.S. Patent Application Publication Nos. 2003/0068633 and 2002/0081582; U.S. Patent Nos. 6,833,450, 6,830,890, 6,824,866, 6,800,439, 6,375,903 and 5,700,637; and PCT Application Nos. WO 04/031399, WO 04/031351, WO 04/029586, WO 03/100012, WO 03/066212, WO 03/065038, WO 03/064699, WO 03/064027, WO 03/064026, WO 03/046223, WO 03/040410 and WO 02/24597.

According to certain aspects, oligonucleotide sequences may be prepared using inkjet techniques known to those of skill in the art, electrochemical techniques known to those of skill in the art, microfluidic techniques known to those of skill in the art, photogenerated acids known to those of skill in the art, or photodeprotected monomers known to those of skill in the art. Such techniques have the advantage of making oligonucleotides at high speed, low cost, fewer toxic chemicals, enhanced portability and ability to interleave DNA biochemistry (e.g. modifications, polymerases, hybridization etc.) with de novo (digital or analog) synthesis. For example, spatially patterned light, either directly from camera optics or from Digital Micromirror Display devices (DMD), can be used with aqueous chemistry. See US2003/0228611. For example, a template-independent polymerase like Terminal deoxynucleotidyl Transferase (TdT) or poly(A) polymerase ~ alternatively, a template-

dependent polymerase like Taq or Phi29 derivatives, can have their basic polymerase function, base-specificity or fidelity programmable by light by incorporating an azobenzene amino acid (see Hoppmann C, Schmieder P, Heinrich N, Beyermann M. (2011) *Chembiochem*.12(17):2555-9. doi: 10.1002/cbic.201100578. Epub 2011 Oct 13, Photoswitchable click amino acids: light control of conformation and bioactivity) into the active site of the polymerase or 5' to 3' exonuclease domains (if present).

Light sensitive neurons (optogenetics) can trigger ion-sensitive polymerases (see Zamft B, Marblestone A, Kording K, Schmidt D, Martin-Alarcon D, Tyo K, Boyden E, Church GM(2012) Measuring Cation Dependent DNA Polymerase Fidelity Landscapes by Deep Sequencing. *PLoS One*, in press) or, for some applications, the ion flux patterns themselves can constitute the stored datasets.

According to certain aspects, nucleic acids can be manufactured on substrates using electrode arrays, conventional camera optics, microscopy optics, flat optics (fresnel or bead microlens), curved imaging planes. Square, trigonal, hexagonal, or other repeating motifs (as in digital photography) arrays or analog imaging (as in conventional silver halide photography). If light is used, the spatial patterning can be via DMD (digital micro mirror device), other digital project methods or natural (analog) light fields.

According to certain aspects, nucleic acids can be made by electrochemical solid phase synthesis as disclosed in US 6,093,302 hereby incorporated by reference in its entirety. According to this aspect, diverse sequences of separate polymers or nucleic acids sequences are prepared using electrochemical placement of monomers or nucleotides at a specific location on a substrate containing at least one electrode that is preferentially in contact with a buffering or scavenging solution to prevent chemical crosstalk between electrodes due to diffusion of electrochemically generated reagents.

According to certain aspects, photogenerated acids may be used to synthesize nucleic acids as described in Church et al, Nature, Vol. 432, 23/30 December 2004 hereby incorporated by reference in its entirety.

According to certain aspects, methods of providing or delivering dNTP, rNTP or rNDP are useful in making nucleic acids. Release of a lipase or other membrane-lytic enzyme from pH-sensitive viral particoles inside dNTP filled-liposomes is described in *J Clin Microbiol.* May 1988; 26(5): 804-807. Photo-caged rNTPs or dNTPs from which NTPs can be released, typically nitrobenzyl derivatives sensitive to 350nm light, are commercially available from Lifetechnologies. Rhoposin or bacterio-opsin triggered signal transduction resulting in vesicular or other secretion of nucleotides is known in the art. With these methods for delivering dNTPs, the nucleotides should be removed or sequestered between the first primer-polymerase encountered and any downstream.

According to certain aspects, ligases are useful in making nucleic acids. Such ligases include DNA ligases known to those of skill in the art and RNA ligases known to those of skill in the art. DNA ligases include bacterial and mammalian DNA ligases. Exemplary ligases include T3 ligase, T4 ligase, T7 ligase, E. coli DNA ligase, Taq DNA ligase, circ-ligase and the like.

According to certain aspects, methods of using pH or light to modulate polymerase activity is useful in making nucleic acids. Polymerases having an optimal pH range for nucleotide incorporation and a pH range in which reversible activity occurs are known in the art. Azobenzene amino acids can be incorporated into the DNA or RNA polymerases via synthetic peptides or unique genetic codes with altered tRNAs as described in ACS Nano. 2014 May 27;8(5):4 157-65. Further useful methods are described in Nature, 500(7463) August 22, 2013.

Polymerases may be used to build nucleic acid molecules representing information which is referred to herein as being recorded in the nucleic acid sequence. Polymerases are enzymes that produce a nucleic acid sequence, for example, using DNA or RNA as a template. Polymerases that produce RNA polymers are known as RNA polymerases, while polymerases that produce DNA polymers are known as DNA polymerases. Template-independent polymerases such as terminal deoxynucleotidyl transferase (TdT), also known as DNA nucleotidylexotransferase (DNNT) or terminal transferase create nucleic acid strands by catalyzing the addition of nucleotides to the 3' terminus of a DNA molecule without a template. The preferred substrate of TdT is a 3'-overhang, but it can also add nucleotides to blunt or recessed 3' ends. Cobalt is a cofactor, however the enzyme catalyzes reaction upon Mg and Mn administration in vitro. Nucleic acid initiators may be 4 or 5 nucleotides or longer and may be single stranded or double stranded. Double stranded initiators may have a 3' overhang or they may be blunt ended or they may have a 3' recessed end.

TdT, like all DNA polymerases, also requires divalent metal ions for catalysis. Further description of TdT is provided in *Biochim Biophys Acta.*, May 2010; 1804(5): 115 1-1166 hereby incorporated by reference in its entirety. Another useful polymerase is the eta-polymerase described in Matsuda et al. (2000) *Nature* 404(6781): 101 1-10 13 hereby incorporated by reference in its entirety.

Methods of Amplifying Nucleic Acids

In general, "amplifying" includes the production of copies of a nucleic acid molecule via repeated rounds of primed enzymatic synthesis. "In situ" amplification indicates that the amplification takes place with the template nucleic acid molecule positioned on a support or a bead, rather than in solution. In situ amplification methods are described in U.S. Patent No.

6,432,360. Varied choices of polymerases exist with different properties, such as temperature, strand displacement, and proof-reading. Amplification can be isothermal and in similar adaptation such as multiple displacement amplification (MDA) described by Dean et al., Comprehensive human genome amplification using multiple displacement amplification, *Proc. Natl. Acad. Sci. U.S.A.*, vol. 99, p. 5261-5266. 2002; also Dean et al, Rapid amplification of plasmid and phage DNA using phi29 DNA polymerase and multiply-primed rolling circle amplification, *Genome Res.*, vol. 11, p. 1095-1099. 2001; also Aviel-Ronen et al., Large fragment Bst DNA polymerase for whole genome amplification of DNA formalin-fixed paraffin-embedded tissues, *BMC Genomics*, vol. 7, p. 312. 2006. Amplification can also cycle through different temperature regiments, such as the traditional polymerase chain reaction (PCR) popularized by Mullis et al., Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction. *Cold Spring Harbor Symp. Quant. Biol.*, vol. 51, p. 263-273. 1986. Variations more applicable to genome amplification are described by Zhang et al, Whole genome amplification from a single cell: implications for genetic analysis, *Proc. Natl. Acad. Sci. U.S.A.*, vol. 89, p. 5847-5851. 1992; and Telenius et al., Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer, *Genomics*, vol. 13, p. 718-725. 1992. Other methods include Polony PCR described by Mitra and Church, In situ localized amplification and contact replication of many individual DNA molecules, *Nuc. Acid. Res.*, vol. 27, pages e34. 1999; emulsion PCR (ePCR) described by Shendure et al., Accurate multiplex polony sequencing of an evolved bacterial genome, *Science*, vol. 309, p. 1728-32. 2005; and Williams et al., Amplification of complex gene libraries by emulsion PCR, *Nat. Methods*, vol. 3, p. 545-550. 2006. Any amplification method can be combined with a reverse transcription step, a priori, to allow amplification of RNA. According to certain aspects, amplification is not absolutely required since probes,

reporters and detection systems with sufficient sensitivity can be used to allow detection of a single molecule.

Amplification methods useful in the present disclosure may comprise contacting a nucleic acid with one or more primers that specifically hybridize to the nucleic acid under conditions that facilitate hybridization and chain extension. Exemplary methods for amplifying nucleic acids include the polymerase chain reaction (PCR) (see, e.g., Mullis et al. (1986) *Cold Spring Harb. Symp. Quant. Biol.* 51 Pt 1:263 and Cleary et al. (2004) *Nature Methods* 1:241; and U.S. Patent Nos. 4,683,195 and 4,683,202), anchor PCR, RACE PCR, ligation chain reaction (LCR) (see, e.g., Landegran et al. (1988) *Science* 241:1077-1080; and Nakazawa et al. (1994) *Proc. Natl. Acad. Sci. U.S.A.* 91:360-364), self sustained sequence replication (Guatelli et al. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87:1874), transcriptional amplification system (Kwoh et al. (1989) *Proc. Natl. Acad. Sci. U.S.A.* 86:1173), Q-Beta Replicase (Lizardi et al. (1988) *BioTechnology* 6:1197), recursive PCR (Jaffe et al. (2000) *J. Biol. Chem.* 275:2619; and Williams et al. (2002) *J. Biol. Chem.* 277:7790), the amplification methods described in U.S. Patent Nos. 6,391,544, 6,365,375, 6,294,323, 6,261,797, 6,124,090 and 5,612,199, isothermal amplification (e.g., rolling circle amplification (RCA), hyperbranched rolling circle amplification (HRCa), strand displacement amplification (SDA), helicase-dependent amplification (HDA), PWGA) or any other nucleic acid amplification method using techniques well known to those of skill in the art.

In exemplary embodiments, the methods disclosed herein utilize PCR amplification. "Polymerase chain reaction," or "PCR," refers to a reaction for the in vitro amplification of specific DNA sequences by the simultaneous primer extension of complementary strands of DNA. In other words, PCR is a reaction for making multiple copies or replicates of a target nucleic acid flanked by primer binding sites, such reaction comprising one or more

repetitions of the following steps: (i) denaturing the target nucleic acid, (ii) annealing primers to the primer binding sites, and (iii) extending the primers by a nucleic acid polymerase in the presence of nucleoside triphosphates. Usually, the reaction is cycled through different temperatures optimized for each step in a thermal cycler instrument. Particular temperatures, durations at each step, and rates of change between steps depend on many factors well-known to those of ordinary skill in the art, e.g., exemplified by the references: McPherson et al., editors, PCR: A Practical Approach and PCR2: A Practical Approach (IRL Press, Oxford, 1991 and 1995, respectively). For example, in a conventional PCR using Taq DNA polymerase, a double stranded target nucleic acid may be denatured at a temperature greater than 90 °C, primers annealed at a temperature in the range 50-75 °C, and primers extended at a temperature in the range 68-78 °C.

The term "PCR" encompasses derivative forms of the reaction, including but not limited to, RT-PCR, real-time PCR, nested PCR, quantitative PCR, multiplexed PCR, assembly PCR and the like. Reaction volumes range from a few hundred nanoliters, e.g., 200 nL, to a few hundred microliters, e.g., 200 µL. "Reverse transcription PCR," or "RT-PCR," means a PCR that is preceded by a reverse transcription reaction that converts a target RNA to a complementary single stranded DNA, which is then amplified, e.g., Tecott et al, U.S. Patent No. 5,168,038. "Real-time PCR" means a PCR for which the amount of reaction product, i.e., amplicon, is monitored as the reaction proceeds. There are many forms of real-time PCR that differ mainly in the detection chemistries used for monitoring the reaction product, e.g., Gelfand et al., U.S. Patent No. 5,210,015 ("Taqman"); Wittwer et al., U.S. Patent Nos. 6,174,670 and 6,569,627 (intercalating dyes); Tyagi et al., U.S. Patent No. 5,925,517 (molecular beacons). Detection chemistries for real-time PCR are reviewed in Mackay et al, Nucleic Acids Research, 30:1292-1305 (2002). "Nested PCR" means a two-stage PCR

wherein the amplicon of a first PCR becomes the sample for a second PCR using a new set of primers, at least one of which binds to an interior location of the first amplicon. As used herein, "initial primers" in reference to a nested amplification reaction mean the primers used to generate a first amplicon, and "secondary primers" mean the one or more primers used to generate a second, or nested, amplicon. "Multiplexed PCR" means a PCR wherein multiple target sequences (or a single target sequence and one or more reference sequences) are simultaneously carried out in the same reaction mixture, e.g. Bernard et al. (1999) *Anal. Biochem.*, 273:221-228 (two-color real-time PCR). Usually, distinct sets of primers are employed for each sequence being amplified. "Quantitative PCR" means a PCR designed to measure the abundance of one or more specific target sequences in a sample or specimen. Techniques for quantitative PCR are well-known to those of ordinary skill in the art, as exemplified in the following references: Freeman et al., *Biotechniques*, 26: 112-126 (1999); Becker-Andre et al., *Nucleic Acids Research*, 17:9437-9447 (1989); Zimmerman et al., *Biotechniques*, 21:268-279 (1996); Diviacco et al., *Gene*, 122:3013-3020 (1992); Becker-Andre et al., *Nucleic Acids Research*, 17:9437-9446 (1989); and the like.

Rolling Circle Amplification (RCA) (Zhong (2001) *Proc. Natl. Acad. Sci. USA* 98(7):3940-3945) represents an alternative to polony amplification since it is continuous replication and does not require thermal cycling. With only one primer (or nick), it grows one long tail from the original circle at a rate linear with time. Isothermal amplification of a circular or linear nucleic acid template also can be performed according to Tabor and Richardson (WO 00/41524) using methods in which enzymatic synthesis of nucleic acid molecules occurs in the absence of oligonucleotide primers. When a second primer from the opposite strand is also included, highly branched structures are produced, with mass growing initially exponentially with respect to time ($m=k*\exp(t)$, or at least $m=kt^2$).

The nucleic acids can be read with or without polymerase amplification. Amplification can be via thermal cycling or isothermal. The amplicons can be short, i.e. 100 to 200 mers as is convenient for current chemical synthesis or up to 1 Mbp as might be achievable with polymerases.

Methods of Sequencing Nucleic Acids

The present disclosure provides methods of sequencing nucleic acids known to those of skill in the art such as high throughput disclosed in Mitra (1999) *Nucleic Acids Res.* 27(24):e34; pp. 1-6. Sequencing methods useful in the present disclosure include Shendure et al., Accurate multiplex polony sequencing of an evolved bacterial genome, *Science*, vol. 309, p. 1728-32. 2005; Drmanac et al., Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays, *Science*, vol. 327, p. 78-81. 2009; McKernan et al., Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding, *Genome Res.*, vol. 19, p. 1527-41. 2009; Rodrigue et al, Unlocking short read sequencing for metagenomics, *PLoS One*, vol. 28, e11840. 2010; Rothberg et al, An integrated semiconductor device enabling non-optical genome sequencing, *Nature*, vol. 475, p. 348-352. 2011; Margulies et al, Genome sequencing in microfabricated high-density picolitre reactors, *Nature*, vol. 437, p. 376-380. 2005; Rasko et al. Origins of the E. coli strain causing an outbreak of hemolytic-uremic syndrome in Germany, *N. Engl. J. Med.*, Epub. 2011; Hutter et al., Labeled nucleoside triphosphates with reversibly terminating aminoalkoxyl groups, *Nucleos. Nucleot. Nucl*, vol. 92, p. 879-895. 2010; Seo et al., Four-color DNA sequencing by synthesis on a chip using photocleavable fluorescent nucleotides, *Proc. Natl. Acad. Sci. USA.*, Vol. 102, P. 5926-5931 (2005); Olejnik et al.; Photocleavable biotin derivatives: a versatile approach for the isolation

of biomolecules, *Proc. Natl. Acad. Sci. U.S.A.*, vol. 92, p. 7590-7594. 1995; US 2009/0062 129 and US 2009/0 191553.

Sequencing primers according to the present disclosure are those that are capable of binding to a known binding region of the target polynucleotide and facilitating ligation of an oligonucleotide probe of the present disclosure. Sequencing primers may be designed with the aid of a computer program such as, for example, DNAWorks, or Gene201igo. The binding region can vary in length but it should be long enough to hybridize the sequencing primer. Target polynucleotides may have multiple different binding regions thereby allowing different sections of the target polynucleotide to be sequenced. Sequencing primers are selected to form highly stable duplexes so that they remain hybridized during successive cycles of ligation. Sequencing primers can be selected such that ligation can proceed in either the 5' to 3' direction or the 3' to 5' direction or both. Sequencing primers may contain modified nucleotides or bonds to enhance their hybridization efficiency, or improve their stability, or prevent extension from a one terminus or the other.

According to certain aspects, polymers, including non-nucleotide based polymers, identified herein may be sequenced by passing the polymer through nanopores or nanogaps or nanochannels to determine the individual monomers in the polymer. Briefly, the polymer is in an electrically conductive medium and is passed through a nanopore under the influence of a voltage differential. Interface dependent changes in ionic current are used to differentiate between individual monomers.

Nanopore and Nanopore Sequencing

"Nanopore" means a hole or passage having a nanometer scale width. Exemplary nanopores include a hole or passage through a membrane formed by a multimeric protein

ring. Typically, the passage is 0.2-25 nm wide. Nanopores, as used herein, may include transmembrane structures that may permit the passage of molecules through a membrane. Examples of nanopores include α -hemolysin (Staphylococcus aureus) and MspA (Mycobacterium smegmatis). Other examples of nanopores may be found in the art describing nanopore sequencing or described in the art as pore-forming toxins, such as the β -PFTs Pantone-Valentine leukocidin S, aerolysin, and Clostridial Epsilon-toxin, the α -PFTs cytolysin A, the binary PFT anthrax toxin, or others such as pneumolysin or gramicidin. Nanopores have become technologically and economically significant with the advent of nanopore sequencing technology. Methods for nanopore sequencing are known in the art, for example, as described in U.S.P.N. 5,795,782, which is incorporated by reference. Briefly, nanopore detection involves a nanopore-perforated membrane immersed in a voltage-conducting fluid, such as an ionic solution including, for example, KCl, NaCl, NiCl, LiCl or other ion forming inorganic compounds known to those of skill in the art. A voltage is applied across the membrane, and an electric current results from the conduction of ions through the nanopore. When the nanopore interacts with polymers, such as DNA or other non-DNA polymers, flow through the nanopore is modulated in a monomer-specific manner, resulting in a change in the current that permits identification of the monomer(s). Nanopores within the scope of the present disclosure include solid state nonprotein nanopores known to those of skill in the art and DNA origami nanopores known to those of skill in the art. Such nanopores provide a nanopore width larger than known protein nanopores which allow the passage of larger molecules for detection while still being sensitive enough to detect a change in ionic current when the complex passes through the nanopore.

"Nanopore sequencing" means a method of determining the components of a polymer based upon interaction of the polymer with the nanopore. Nanopore sequencing may be

achieved by measuring a change in the conductance of ions through a nanopore that occurs when the size of the opening is altered by interaction with the polymer.

In addition to a nanopore, the present disclosure envisions the use of a nanogap which is known in the art as being a gap between two electrodes where the gap is about a few nanometers in width such as between about 0.2 nm to about 25 nm or between about 2 and about 5 nm. The gap mimics the opening in a nanopore and allows polymers to pass through the gap and between the electrodes. Aspects of the present disclosure also envision use of a nanochannel electrodes are placed adjacent to a nanochannel through which the polymer passes. It is to be understood that one of skill will readily envision different embodiments of molecule or moiety identification and sequencing based on movement of a molecule or moiety through an electric field and creating a distortion of the electric field representative of the structure passing through the electric field.

Supports and Attachment

While nucleic acids or oligonucleotide sequences as described herein may be in solution in a vessel or container, one or more oligonucleotide sequences described herein may be immobilized on a support (e.g., a solid and/or semi-solid support). In certain aspects, an oligonucleotide sequence can be attached to a support using one or more of the phosphoramidite linkers described herein. Suitable supports include, but are not limited to, slides, beads, chips, particles, strands, gels, sheets, tubing, spheres, containers, capillaries, pads, slices, films, plates and the like. In various embodiments, a solid support may be biological, nonbiological, organic, inorganic, or any combination thereof. Supports of the

present invention can be any shape, size, or geometry as desired. For example, the support may be square, rectangular, round, flat, planar, circular, tubular, spherical, and the like. When using a support that is substantially planar, the support may be physically separated into regions, for example, with trenches, grooves, wells, or chemical barriers (e.g., hydrophobic coatings, etc.). Supports may be made from glass (silicon dioxide), metal, ceramic, polymer or other materials known to those of skill in the art. Supports may be a solid, semi-solid, elastomer or gel. In certain exemplary embodiments, a support is a microarray. As used herein, the term "microarray" refers in one embodiment to a type of array that comprises a solid phase support having a substantially planar surface on which there is an array of spatially defined non-overlapping regions or sites that each contain an immobilized hybridization probe. "Substantially planar" means that features or objects of interest, such as probe sites, on a surface may occupy a volume that extends above or below a surface and whose dimensions are small relative to the dimensions of the surface. For example, beads disposed on the face of a fiber optic bundle create a substantially planar surface of probe sites, or oligonucleotides disposed or synthesized on a porous planar substrate create a substantially planar surface. Spatially defined sites may additionally be "addressable" in that its location and the identity of the immobilized probe at that location are known or determinable.

The solid supports can also include a semi-solid support such as a compressible matrix with both a solid and a liquid component, wherein the liquid occupies pores, spaces or other interstices between the solid matrix elements. Preferably, the semi-solid support materials include polyacrylamide, cellulose, poly dimethyl siloxane, polyamide (nylon) and cross-linked agarose, -dextran and -polyethylene glycol. Solid supports and semi-solid supports can be used together or independent of each other.

Supports can also include immobilizing media. Such immobilizing media that are of use according to the invention are physically stable and chemically inert under the conditions required for nucleic acid molecule deposition and amplification. A useful support matrix withstands the rapid changes in, and extremes of, temperature required for PCR. The support material permits enzymatic nucleic acid synthesis. Useful support materials include both organic and inorganic substances, and include, but are not limited to, polyacrylamide, cellulose and polyamide (nylon), as well as cross-linked agarose, dextran or polyethylene glycol.

Methods of immobilizing oligonucleotides to a support are known in the art (beads: Dressman et al. (2003) Proc. Natl. Acad. Sci. USA 100:8817, Brenner et al. (2000) Nat. Biotech. 18:630, Albretsen et al. (1990) Anal. Biochem. 189:40, and Lang et al. Nucleic Acids Res. (1988) 16:10861; nitrocellulose: Ranki et al. (1983) Gene 21:77; cellulose: Goldkorn (1986) Nucleic Acids Res. 14:9171; polystyrene: Ruth et al. (1987) Conference of Therapeutic and Diagnostic Applications of Synthetic Nucleic Acids, Cambridge U.K.; teflon-aery lamide: Duncan et al. (1988) Anal. Biochem. 169:104; polypropylene: Polsky-Cynkin et al. (1985) Clin. Chem. 31:1438; nylon: Van Ness et al. (1991) Nucleic Acids Res. 19:3345; agarose: Polsky-Cynkin et al., Clin. Chem. (1985) 31:1438; and sephacryl: Langdale et al. (1985) Gene 36:201; latex: Wolf et al. (1987) Nucleic Acids Res. 15:2911). Supports may be coated with attachment chemistry or polymers, such as amino-silane, NHS-esters, click chemistry, polylysine, etc., to bind a nucleic acid to the support. Nucleic acids that have been synthesized on the surface of a support may be removed, such as by a cleavable linker or linkers known to those of skill in the art.

As used herein, the term "attach" refers to both covalent interactions and noncovalent interactions. A covalent interaction is a chemical linkage between two atoms or radicals

formed by the sharing of a pair of electrons (i.e., a single bond), two pairs of electrons (i.e., a double bond) or three pairs of electrons (i.e., a triple bond). Covalent interactions are also known in the art as electron pair interactions or electron pair bonds. Noncovalent interactions include, but are not limited to, van der Waals interactions, hydrogen bonds, weak chemical bonds (i.e., via short-range noncovalent forces), hydrophobic interactions, ionic bonds and the like. A review of noncovalent interactions can be found in Alberts et al., in *Molecular Biology of the Cell*, 3d edition, Garland Publishing, 1994.

Computer-based Methods

The practice of the methods disclosed herein may employ conventional biology methods, software, computers and computer systems. Accordingly, the methods described herein may be computer implemented methods in whole or in part. Computer software utilized in the methods of the present disclosure include computer readable medium having computer-executable instructions for performing logic steps of the method of the invention. Suitable computer readable medium include, but are not limited to, a floppy disk, CD-ROM/DVD/DVD-ROM, hard-disk drive, flash memory, ROM/RAM, magnetic tapes, and others that may be developed. The computer executable instructions may be written in a suitable computer language or combination of several computer languages. The methods described herein may also make use of various commercially available computers and computer program products and software for a variety of purposes including translating text or images into binary code, designing nucleic acids sequences representative of the binary code, analyzing sequencing data from the nucleic acid sequences, translating the nucleic acid sequence data into binary code, and translating the binary code into text or images.

EXAMPLE I**Secure Communication Using Primer-based Lock and Keys**

Sender wants to send information encoded in DNA ("true message DNA") to intended recipient or receiver while making it difficult to a third party to intercept and decode the DNA. Sender can mix the information encoded in single-stranded nucleic acid, such as single-stranded DNA with random single-stranded DNA sequences ("dummy message DNA"). The true message DNA can be selectively amplified from a mixture of single-stranded DNA sequences using PCR with a primer pair specific to the true message DNA. The priming sequence functions as a lock while the complementary primer sequence functions as a key to amplify and sequence the true message. Oligonucleotide primer pairs function as designed lock and key systems where the oligonucleotide encoding the message can be unlocked by the recipient having knowledge of the associated key, such as the primer key sequence.

As an example, Sender has the capacity to synthesize DNA. Intended Recipient has the capacity to sequence DNA. Sender sets up communication protocol with intended recipient as follows. Sender synthesizes a "lock'V'key" primer set. Sender gives Intended Recipient the "key" primer to be used in amplification and/or sequencing while keeping the "lock" primer. Sender synthesizes single-stranded DNA corresponding to and encoding the true message and ligates this true message from its 3' end to the 5' end of the "lock" primer using a single-stranded DNA ligase. Sender also synthesizes dummy or false single-stranded DNA messages of similar chemical and entropic properties and ligates them to a set of dummy "lock" primers of similar chemical or entropic properties as the true "lock" primer. The DNA containing the true message and the DNA containing dummy or false messages are mixed in one tube and sent to intended recipient. To identify the DNA which contains the

true message from the mixed sample, Intended Recipient selectively amplifies the true message by PCR with "key" primers that are complementary to the "lock" sequence that was ligated to the true message. Intended Recipient possesses the 'key' primers, so when the Intended Recipient receives the DNA mix sent from Sender, only the Intended Recipient can selectively amplify the true message from the mixture of dummy DNAs. As an example, Intended recipient possessed a lab-on-a-chip device which holds generic primer and the key primer. The Intended Recipient performs PCR to selectively amplify the true strand to quantities sufficient to perform DNA sequencing to read out the true message (which can be plaintext or ciphertext). If an Interceptor were to intercept the sample in any way, Interceptor will be unable to identify the true message since Interceptor doesn't have access to the 'key' primer and does not know its sequence.

Accordingly, a single-stranded nucleotide polymer is generated that encodes a message into DNA bases (i.e., A, C, G, and T). The Sender can also synthesize false messages ligated to dummy primers that have chemical and entropic properties similar to the true strand. By combining all DNA strands, Sender can obfuscate the true message and key primers. The corresponding DNA sequence is then synthesized along a 'generic' or common primer sequence that does not need to be a secret between Sender and Intended Recipient. The 3' end of true message DNA is ligated to a 'lock' primer. The reverse complement of the 'lock' primer that is ligated to the 3' of a DNA message is the 'key' primer as only this primer can be used to amplify the full message DNA. The 'key' primer should only be accessible to the party that has to read the message (i.e., Intended Recipient). The sequence of the 'key' primer and or its corresponding 'lock' should not be known to anyone other than Sender and Intended Recipient. For deciphering a true message, true message DNA is

amplified in a PCR reaction using the 'key' and the 'generic' primer. The PCR product is sequenced and the message is decoded.

EXAMPLE II

Secure Communication Using Primer-based Lock and Keys

Sender wants to be able to send messages encoded in single-stranded DNA to Intended Recipient, but they do not have the opportunity for a pre-deployment set up and exchange of primers. As a result, Intended Recipient sends Sender the lock primer to which Intended Recipient already holds the key. Both Sender and Intended Recipient have the capacity to synthesize and sequence DNA. They set up a communication protocol with the following steps. Intended Recipient synthesizes a "lock'V'key" primer set. Intended Recipient transmits the "lock" primer to Sender to be used for communicating Sender's messages to Intended Recipient. Sender synthesizes single-stranded DNA corresponding to Sender's message and ligates this message to Intended Recipient's "lock" primer. Sender also synthesizes dummy DNA messages of similar chemical and entropic properties and ligates them to a set of dummy "lock" primers of Sender's own making. The DNA containing the true message with Intended Recipient's lock and the DNA containing dummy messages with dummy locks that Sender made are mixed in one tube and sent to Intended Recipient who uses the key primer to amplify and/or sequence the true message.

Accordingly, sender encodes a message into single-stranded DNA and synthesizes it on the 3' of a 'generic' primer. Sender ligates Intended Recipient's 'lock' primer, which Sender received, to the 3' end of true message DNA. Intended Recipient amplifies the true message DNA by PCR using the 'generic' primer and Intended Recipient's own 'key' primer. The true message DNA is sequenced to read the message.

Alternatively, Intended Recipient has a primer pair, one key (K) and one lock (L). They are complementary. The lock has a 5' phosphate and a 3' block. Intended Recipient sends his lock primer in a tube or vessel to Sender so that Sender can use it to lock the true message to be sent to Intended Recipient. The tube with Intended Recipient's lock primer contains other random similar DNA for additional obfuscation - only the lock has a 5'-phosphate and thus can be ligated, while the other sequences lack a 5' phosphate and may be not blocked on their 3'end. Sender synthesizes the true message as single-stranded DNA on a 'generic' primer and "locks" the message by ligating Intended Recipient's lock on the 3' end of the message using single stranded DNA ligase. Sender separately generates dummy messages and dummy locks (LD). Sender mixes the true message and dummy messages in one tube or vessel and sends this tube or vessel to Intended Recipient. Intended Recipient can "unlock" the true message by selectively amplifying the true message by PCR primers, 'generic', and Intended Recipient's own key primer. Should Interceptor intercept the message from Sender to Intended Recipient, Interceptor cannot identify the true message because the message is obfuscated by a very large number of dummy messages which Interceptor cannot distinguish from the true message. Should Inteceptor intercept the locks sent from Intended Recipient, Inceptor may try to identify the sequence of the lock and synthesize the corresponding key, which is the reverse-complement of the lock. However, the lock can be composed largely of non-sequenceable nucleotide analogues and may be inactive. The primer may be designed such that deciphering its sequence is not possible. This can be accomplished by using non-sequenceable nucleotide analogues or non-copiable bonds between different nucleotides. Furthermore, the primer is inactive and requires specific treatments to be activated. The nature and sequence of such treatments are only known to

Intended Recipient. Applying the wrong treatment or the wrong sequence of treatments to the primer can cause destruction of the primer altogether.

EXAMPLE III

Secure Communication Using a Biological One-Time Pad

The present disclosure provides methods for making and using a biological one-time pad including a plurality of first nucleic acid strands which may include a barcode sequence, a true message sequence, an initiator sequence or other useful sequence and having attached to each of the plurality of first nucleic acid strands a different unique nucleotide sequence key. The concept of a "one-time pad" is found in cryptography. The one-time pad (OTP) is an encryption technique where a plaintext is paired with a random secret key (also referred to as a one-time pad). Then, each bit or character of the plaintext is encrypted by combining it with the corresponding bit or character from the pad using modular addition. The key should be truly random, should be at least as long as the plaintext, should never be reused in whole or in part, and should be kept completely secret in order to prevent the ciphertext from being decrypted or broken. The "pad" part of the name comes from early implementations where the key material was distributed as a pad of paper, so that the top sheet could be easily torn off and destroyed after use. The key is any random string that a Sender and Intended Recipient have agreed to in advance of transmission of an encoded message. Examples of one-time pad applications can be found at world wide website cs.utsa.edu/~wagner/laws/pad.html and Shannon, C.E., "A Mathematical Theory of Communication." Bell System Technical Journal, July 1948, P. 623.

The present disclosure utilizes properties and characteristics of a "one-time pad" with nucleic acid sequences encoding a true message and a having an associated different and

unique nucleotide sequence, referred to herein as a key. A plurality of truly random key nucleic acid sequences can be produced by using TdT as described herein to extend a nucleic acid key sequence from a nucleic acid strand. The nucleic acid strand may be the true message strand or it may be an initiator strand and the random sequence may be removed to the initiator strand and added to a true message strand. The random nucleic acid key should be at least as long as the true message, should never be reused in whole or in part, and should be kept completely secret in order to prevent the ciphertext from being decrypted or broken. The key is used to encrypt a message with an encryption algorithm known to those of skill in the art. Each bit of the true message is encrypted by a modular arithmetic or an XOR operation with each bit of the key. Once these operations have been performed on every bit of the message, the string represents the encrypted message. It is to be understood that many software tools are known to those of skill in the art to perform such modular arithmetic or XOR operations or more complex operations. The key is used to decrypt a message with a decryption algorithm known to those of skill in the art. Each bit of the true message is decrypted by a modular arithmetic or an XOR operation with each bit of the key. Once these operations have been performed on every bit of the message, the string represents the decrypted message. It is to be understood that many software tools are known to those of skill in the art to perform such modular arithmetic or XOR operations or more complex operations.

Accordingly, the disclosure provides a method of associating a different unique nucleotide sequence key with each of a plurality of barcoded initiator sequence strands including extending each of the plurality of barcoded initiator sequence strands with a random sequence of nucleotides forming the unique nucleotide key sequence, ligating a universal 3' sequence to the 5' end of each of the unique nucleotide key sequences to form

amplifiable strands, and amplifying the amplifiable strands to form a collection of amplified barcoded initiator sequence strands having unique nucleotide key sequences. The disclosure provides that each of the plurality of barcoded initiator sequence strands is extended with a random sequence of nucleotides forming the unique nucleotide key sequence using a template independent polymerase. The disclosure provides that each of the plurality of barcoded initiator sequence strands is ligated to a pre-synthesized random sequence of nucleotides forming the unique nucleotide key sequence. The disclosure provides that only a substantially small fraction of barcoded initiators with universal 3' sequence are amplified. The disclosure provides that the reverse complementary strands resulting from amplification of each amplified barcoded initiator sequence strands having unique nucleotide sequences is removed. According to one aspect, in order to perform the probe:barcode hybridization, the nucleic acids of the one-time pad or one-time pad dictionary may be single stranded. The complementary strand created, due to PCR amplification, may be removed before aliquoting the keys between communicating parties. There are many techniques to remove the complementary strand which are known to those skilled in the art. An example of such a technique includes PCR amplification with a reverse primer which has a biotin on the 5' end. To remove the reverse complementary strand, the PCR products are hybridized to beads with streptavidin and sodium hydroxide is added to denature the double stranded product. The reverse complementary strand will stay bound to the beads and the supernatant will contain the desired sense strand. The disclosure provides that the collection of amplified barcoded initiator sequence strands having unique nucleotide key sequences is distributed in aliquots to communicating parties. In this manner, the communicating parties, i.e. Sender and Intended Recipient, each receive and possess a collection of nucleic acids representing the one-time pad to be used with encrypting and decrypting nucleic acids with encoded messages.

The disclosure provides a method of encrypting a message including selecting a barcode from a collection of amplified barcoded initiator sequence strands having unique nucleotide key sequences, wherein each barcode indicates a unique nucleotide key sequence, isolating the amplified barcoded initiator sequence strands having the selected barcode from the collection, identifying the unique nucleotide key sequence from the isolated amplified barcoded initiator sequence strands, using the unique nucleotide key sequence to encrypt a message using an encryption algorithm. The disclosure provides that the isolated amplified barcoded initiator sequence strands are destroyed. The disclosure provides that the amplified barcoded initiator sequence strands having the selected barcode are isolated from the collection by hybridizing the selected barcode with a probe and removing or extracting the probe:barcode hybridization product from the collection. The disclosure provides that the amplified barcoded initiator sequence strands having the selected barcode are isolated from the collection by hybridizing the selected barcode to an immobilized probe and removing remaining amplified barcoded initiator sequence strands and recovering the amplified barcoded initiator sequence strands immobilized to the probe.

The disclosure provides a method of decrypting a message encrypted using a unique nucleotide key sequence associated with a selected barcode including isolating amplified barcoded initiator sequence strands having the selected barcode from a collection of amplified barcoded initiator sequence strands having unique nucleotide key sequences, wherein each barcode indicates a unique nucleotide key sequence, identifying the unique nucleotide key sequence from the isolated amplified barcoded initiator sequence strands, and using the unique nucleotide key sequence to decrypt the message using a decryption algorithm. The disclosure provides that the isolated amplified barcoded initiator sequence strands are destroyed. The disclosure provides that the amplified barcoded initiator sequence

strands having the selected barcode are isolated from the collection by hybridizing the selected barcode with a probe and removing or extracting the probe:barcode hybridization product from the collection. The disclosure provides that the amplified barcoded initiator sequence strands having the selected barcode are isolated from the collection by hybridizing the selected barcode to an immobilized probe and removing remaining amplified barcoded initiator sequence strands and recovering the amplified barcoded initiator sequence strands immobilized to the probe.

The disclosure provides for methods of making and using a one-time pad of dictionary keys, such as a method of making a library of nucleic acid sequences with each nucleic acid sequence representing a word, symbol, understandable message or format of information (all of which may be used interchangeably herein) including preparing a plurality of nucleic acid sequences each representing a word, extending each of the plurality of nucleic acid sequences with a random sequence of nucleotides forming a unique nucleotide key sequence, with the unique nucleotide key sequence representing the word, ligating a universal 3' nucleotide sequence to each of the unique nucleotide key sequences to form amplifiable strands, and amplifying the amplifiable strands to form a collection of amplified barcoded nucleic acid sequences having unique nucleotide key sequences.

The disclosure provides a method of encrypting one or more words including selecting the one or more words, sequencing a library of nucleic acid sequences, each sequence having a first sequence with a known association with a given word and having a second sequence of random nucleotides forming a unique nucleotide key sequence, identifying the unique nucleotide key sequence for each of the one or more words, and associating the unique nucleotide key sequence for each of the one or more words. The disclosure provides that the collection of amplified barcoded nucleic acid sequences having

unique nucleotide key sequences is ligated to the 3' end of an additional barcoded initiator. The disclosure provides that the collection of amplified barcoded nucleic acid sequences having unique nucleotide key sequences is independently ligated to the 3' end of a plurality of additionally barcoded initiators. The disclosure provides that the reverse complementary strands resulting from amplification of collection of amplified barcoded nucleic acid sequences having unique nucleotide key sequences is removed. The disclosure provides that the collection of amplified barcoded nucleic acid sequences having unique nucleotide key sequences is distributed in aliquots to communicating parties. The disclosure provides that the library of nucleic acid sequences is destroyed.

The disclosure provides a method of encrypting one or more words including selecting the one or more words, isolating a library of nucleic acid sequences having a target common barcode sequence from a plurality of libraries of nucleic acids each having a different common barcode sequence, sequencing the isolated library of nucleic acid sequences, each sequence having a first sequence with a known association with a given word and having a second sequence of random nucleotides forming a unique nucleotide key sequence, identifying the unique nucleotide key sequence for each of the one or more words, and associating the unique nucleotide key sequence for each of the one or more words. The disclosure provides that the library of nucleic acid sequences is destroyed.

The disclosure provides a method of decrypting one or more words including receiving one or more random nucleic acid sequences representative of one or more words, isolating a library of nucleic acid sequences having a target common barcode sequence from a plurality of libraries of nucleic acids each having a different common barcode sequence, sequencing the isolated library of nucleic acid sequences, each sequence having a first sequence with a known association with a given word and having a second

sequence of random nucleotides forming a unique nucleotide key sequence, identifying the first sequence with a known association with a given word from the unique nucleotide key sequence, and identifying the one or more words. The disclosure provides that the library of nucleic acid sequences is destroyed.

EXAMPLE IV

Generation of One-Time Pad of Nucleotide Keys

With reference to Fig. 1, the disclosure provides that a pool of initiator oligonucleotides is synthesized. These initiators have unique nucleotide barcodes useful for addressing each initiator oligonucleotide and a corresponding one-time key (i.e., ATCG = key 1, CTAG = key 2, etc). Nucleotidyl transferases ("NTrs") are used to extend each initiator to add a random sequence of nucleotides. Then, a universal 3' primer is ligated to every strand. PCR is performed, using universal primers, to amplify a subset (or all) of the strands, such that any aliquot of this amplified mixture will have full representation of each key. Aliquots of this amplified mixture are distributed between communicating parties. Each aliquot is used as a one-time pad of random nucleotide keys, representing a collection of initiator oligonucleotides with added random nucleotide sequences. The random sequences represent the collection of keys in the one-time pad.

With reference to Fig. 2, Sender ("Alice") and Intended Recipient ("Bob") both have DNA sequencing and synthesis capabilities. Sender and Intended Recipient have an aliquot of the amplified mixture and so have the same one-time pad of random nucleic acid keys. Both agree on an order with which to use the keys. i.e., use key prefixed with "ACTG" first, "CATG" second, etc. Sender uses a one-time key from the one-time pad to encrypt a message Sender wishes to send to Intended Recipient.

With reference to Fig. 3, as an example, Sender and Intended Recipient agree to use the key prefixed (barcoded) with "ACTG". In order to encrypt the message, Sender identifies, or isolates or reveals the key connected with the "ACTG" prefix, use the key to encrypt the message, and send the encrypted message to the Intended Recipient. In order to reveal or identify or isolate the key, Sender synthesizes the complementary strand "TGAC" and uses it to pull out, isolate or identify (such as by hybridization) all instances of the specific key from the pool of keys in the shared one-time pad. Sender sequences the key and performs an XOR between the key and the message, for example using an encryption algorithm, producing a ciphertext. Sender sends the ciphertext (digitally, or in encoded in DNA), to Intended Recipient. The used key is now depleted from the one-time pad, and is destroyed (by an active process such as DNase or chemical treatment or by throwing the DNA key into the environment).

With reference to Figs. 4 and 5, Sender sends the ciphertext to Intended recipient. To decrypt the ciphertext from Sender, Intended Recipient synthesizes a complementary oligo "TGAC" and uses it to pull out (by hybridization) all instances of the specific key from the pool of keys in the shared one-time pad. Intended Recipient sequences the key and performs an XOR operation, for example using an encryption algorithm, between the key and the message, decrypting, i.e, recovering, the plaintext communication from Sender. The used key is now depleted from the one-time pad, and is destroyed (by an active process such as DNase or chemical treatment or by throwing the DNA key into the environment).

EXAMPLE V

Generation of One-Time Pad of Nucleotide Dictionary Keys

The disclosure provides a one-time pad of a library of one-time use dictionary keys. Sender and Intended Recipient have the same one-time pad and both agree on an order with which to use the keys. Sender and Recipient both have DNA sequencing and synthesis capabilities. Sender uses a set of one-time dictionary keys from the one-time pad to encrypt a message Sender wishes to send to Intended Recipient.

With reference to Fig. 6, a one-time pad of dictionary keys is a one-time pad of nucleotide keys in which each initiator or nucleic acid sequence represents a word from the English dictionary. A pool of initiators is synthesized such that each oligo encodes a word (i.e. from the English dictionary) in DNA. NTRs are used to extend each initiator to produce a nucleotide polymer of random sequence. Then, a universal 3' primer is ligated to every strand.

PCR is performed, using universal primers, to amplify a subset (or all) of the strands, such that any aliquot of this amplified mixture will have full representation of each key. Aliquots of this amplified mixture are distributed between communicating parties. This becomes a one-time pad of nucleotide keys as a dictionary, which is aliquotted between communicating parties. Each set of dictionary keys can be prefixed with a barcode (i.e., ATCG = dictionary 1, CTAG = dictionary 2, etc.) such that multiple pads can be generated and pooled.

With reference to Fig. 7, Sender and Intended Recipient have the same one-time pad and both agree on an order with which to use the keys. Sender and Recipient both have DNA sequencing and synthesis capabilities. Sender uses a set of one-time dictionary keys from the one-time pad to encrypt a message Sender wishes to send to Intended recipient.

With reference to Fig. 8, Sender and Intended Recipient agree to use the key prefixed (barcoded) with "ACTG", marked in red. Sender synthesizes the complementary strand "TGAC" and uses it to pull out (by hybridization) all instances of the appropriate one-time

key from the shared one-time pad. Sender sequences the dictionary of keys and performs a substitution of each word with the corresponding key, i.e. ATTACKDAWN for GATTGCGTTGAGTC. With reference to Fig. 9, Sender sends the cipher text (digitally, or in encoded in DNA), to Intended Recipient. The used dictionary keys are now depleted from the one-time pad, and are destroyed (by an active process such as DNase or chemical treatment or by throwing the DNA key into the environment).

With reference to Fig. 10, Intended Recipient sequences the key dictionary and performs local alignments between the key and the message, recovering the plaintext communication from Sender. The used dictionary keys are now depleted from the one-time pad, and are destroyed (by an active process such as DNase or chemical treatment or by throwing the DNA key into the environment).

OTHER EMBODIMENTS

Other embodiments will be evident to those of skill in the art. It should be understood that the foregoing description is provided for clarity only and is merely exemplary. The spirit and scope of the present invention are not limited to the above examples, but are encompassed by the following claims. All publications and patent applications cited above are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication or patent application were specifically and individually indicated to be so incorporated by reference.

Claims:

1. A mixture of nucleic acids comprising
one or more true message single-stranded nucleic acids having a plurality of nucleotides corresponding to a plurality of bit sequences representative of a true message, wherein the one or more nucleic acids include a lock 3' message amplification or sequencing enabling moiety unique to the one or more true message nucleic acids and having an associated key moiety for amplification or sequencing; and
one or more optional dummy single-stranded nucleic acids, and
wherein the one or more true message nucleic acids are selectively amplifiable or sequenceable over the one or more dummy nucleic acids when the lock 3' message amplification or sequencing enabling moiety is combined with the associated key moiety.
2. The mixture of nucleic acids of claim 1 including one or more dummy nucleic acids.
3. The mixture of nucleic acids of claim 1 wherein the lock 3' message amplification enabling moiety is comprised of natural nucleotides and the associated key moiety is the corresponding complementary key sequence of nucleotides which is capable of priming nucleotide polymerization in its own 3' direction and the lock's 5' direction.
4. The mixture of nucleic acids of claim 1 wherein the lock 3' message amplification enabling moiety is comprised of non-natural nucleotide analogues and the associated key moiety is the corresponding complementary key sequence of natural or non-

natural nucleotides which is capable of priming nucleotide polymerization in its own 3' and the lock's 5' direction.

5. The mixture of nucleic acids of claim 1 wherein the lock 3' message amplification enabling moiety is an amplification lock priming sequence and the associated key moiety is a corresponding complementary key primer sequence.

6. The mixture of nucleic acids of claim 1 wherein the lock 3' message sequencing enabling moiety includes a protein or a protein binding site and the associated key moiety is a binding protein that enables nanopore sequencing.

7. The mixture of nucleic acids of claim 1 wherein the one or more dummy nucleic acids includes a dummy 3' message amplification or sequencing enabling moiety unique to the one or more dummy nucleic acids.

8. The mixture of nucleic acids of claim 1 wherein the lock 3' message amplification or sequencing enabling moiety is inactive.

9. The mixture of nucleic acids of claim 1 wherein the lock 3' message amplification or sequencing enabling moiety includes a blocking moiety.

10. The mixture of nucleic acids of claim 1 wherein the lock 3' message amplification or sequencing enabling moiety includes a removable blocking moiety required

to be removed before amplification or sequencing of the one or more true message nucleic acids.

11. The mixture of nucleic acids of claim 1 wherein the lock 3' message amplification or sequencing enabling moiety requires activation before amplification or sequencing of the one or more true message nucleic acids.

12. The mixture of nucleic acids of claim 1 wherein the lock 3' message amplification or sequencing enabling moiety requires activation before amplification or sequencing of the one or more true message nucleic acids including processing by chemical reaction, enzymatic reaction, heat, light, or pH.

13. The mixture of nucleic acids of claim 1 wherein the lock 3' message amplification or sequencing enabling moiety becomes disabled upon amplification or sequencing efforts using other than the associated key moiety.

14. The mixture of nucleic acids of claim 1 wherein the lock 3' message amplification or sequencing promoting moiety is a nucleic acid sequence including one or more non-sequenceable nucleotide analogues or one or more amplification or sequencing inhibiting bonds between nucleic acids.

15. The mixture of nucleic acids of claim 1 wherein each true message includes a unique sequence of nucleotides representing an encryption key with which the true message

is decrypted, wherein the encryption key is positioned immediately 3' of the true message and 5' of the lock 3' message amplification or sequencing enabling moiety.

16. A method of transmitting information from a sender to an intended recipient comprising

transferring to the intended recipient a mixture of nucleic acids comprising

one or more true message single stranded nucleic acids having a plurality of nucleotides corresponding to a plurality of bit sequences representative of a true message, wherein the one or more nucleic acids include a lock 3' message amplification or sequencing enabling moiety unique to the one or more true message nucleic acids and having an associated key moiety for amplification or sequencing, and

one or more optional dummy single-stranded nucleic acids, and

wherein the intended recipient possesses the associated key moiety, and

wherein the one or more true message nucleic acids are selectively amplified or sequenced over the one or more dummy nucleic acids when the lock 3' message amplification or sequencing enabling moiety is combined with the associated key moiety.

17. The method of claim 16 including one or more dummy nucleic acids.

18. The method of claim 16 wherein the one or more true message nucleic acids are sequenced, the nucleotides are converted to bit sequences and the bit sequences are converted to a format of information selected from the group consisting of a text format, an image format, a video format or an audio format.

19. A method of securing a format of information encoded in one or more true message single-stranded nucleic acids having a plurality of nucleotides corresponding to a plurality of bit sequences representative of a true message comprising

adding to the one or more true message single-stranded nucleic acids a lock 3' message amplification or sequencing enabling moiety unique to the one or more true message nucleic acids and having an associated key moiety for amplification or sequencing; and

mixing the one or more true message nucleic acids with one or more dummy single-stranded nucleic acids, and

wherein the one or more true message nucleic acids are selectively amplifiable or sequenceable over the one or more dummy nucleic acids when the lock 3' message amplification or sequencing enabling moiety is combined with the associated key moiety.

20. A method of securing a format of information encoded in one or more true message single-stranded nucleic acids having a plurality of nucleotides corresponding to a plurality of bit sequences representative of a true message comprising

combining the one or more true message single-stranded nucleic acids with a mixture including a lock 3' message amplification or sequencing enabling moiety having a 5'-phosphate and/or a 3'-blocking moiety and dummy lock nucleic acid sequences,

wherein the lock 3' message amplification or sequencing promoting moiety having a 5'-phosphate is ligated to the one or more true message nucleic acids at a 3' end using single-stranded ligation methods,

mixing the one or more true message nucleic acids with one or more dummy nucleic acids including dummy lock nucleic acid sequences,

wherein the lock 3' message amplification or sequencing enabling moiety is unique to the one or more true message nucleic acids and having an associated key moiety for amplification or sequencing; and

wherein the one or more true message nucleic acids are selectively amplifiable or sequenceable over the one or more dummy nucleic acids when the lock 3' message amplification or sequencing enabling moiety is combined with the associated key moiety.

21. The method of claim 20 wherein the lock 3' message amplification or sequencing enabling moiety includes one or more non-sequenceable nucleotide analogues.

22. The method of claim 20 wherein the lock 3' message amplification or sequencing enabling moiety includes one or more non-copiable bonds between nucleotides.

23. A method of associating a different unique nucleotide sequence key with each of a plurality of barcoded initiator sequence strands comprising

extending each of the plurality of barcoded initiator sequence strands with a random sequence of nucleotides forming the unique nucleotide key sequence,

ligating a universal 3' sequence to the 5' end of each of the unique nucleotide key sequences to form amplifiable strands, and

amplifying the amplifiable strands to form a collection of amplified barcoded initiator sequence strands having unique nucleotide key sequences.

24. The method of claim 23 wherein each of the plurality of barcoded initiator sequence strands is extended with a random sequence of nucleotides forming the unique nucleotide key sequence using a template independent polymerase.

25. The method of claim 23 wherein each of the plurality of barcoded initiator sequence strands is ligated to a pre-synthesized random sequence of nucleotides forming the unique nucleotide key sequence.

26. The method of claim 23 wherein only a substantially small fraction of barcoded initiators with universal 3' sequence are amplified.

27. The method of claim 23 wherein the reverse complementary strands resulting from amplification of each amplified barcoded initiator sequence strands having unique nucleotide sequences is removed.

28. The method of claim 23 wherein the collection of amplified barcoded initiator sequence strands having unique nucleotide key sequences is distributed in aliquots to communicating parties.

29. A method of encrypting a message comprising
selecting a barcode from a collection of amplified barcoded initiator sequence strands having unique nucleotide key sequences, wherein each barcode indicates a unique nucleotide key sequence,

isolating the amplified barcoded initiator sequence strands having the selected barcode from the collection,

identifying the unique nucleotide key sequence from the isolated amplified barcoded initiator sequence strands,

using the unique nucleotide key sequence to encrypt a message using an encryption algorithm.

30. The method of claim 29 wherein the isolated amplified barcoded initiator sequence strands are destroyed.

31. The method of claim 29 where the amplified barcoded initiator sequence strands having the selected barcode are isolated from the collection by hybridizing the selected barcode with a probe and removing or extracting the probe:barcode hybridization product from the collection.

32. The method of claim 29 where the amplified barcoded initiator sequence strands having the selected barcode are isolated from the collection by hybridizing the selected barcode to an immobilized probe and removing remaining amplified barcoded initiator sequence strands and recovering the amplified barcoded initiator sequence strands immobilized to the probe.

33. A method of decrypting a message encrypted using a unique nucleotide key sequence associated with a selected barcode comprising

isolating amplified barcoded initiator sequence strands having the selected barcode from a collection of amplified barcoded initiator sequence strands having unique nucleotide key sequences, wherein each barcode indicates a unique nucleotide key sequence,

identifying the unique nucleotide key sequence from the isolated amplified barcoded initiator sequence strands, and

using the unique nucleotide key sequence to decrypt the message using a decryption algorithm.

34. The method of claim 33 wherein the isolated amplified barcoded initiator sequence strands are destroyed.

35. The method of claim 33 where the amplified barcoded initiator sequence strands having the selected barcode are isolated from the collection by hybridizing the selected barcode with a probe and removing or extracting the probe:barcode hybridization product from the collection.

36. The method of claim 33 where the amplified barcoded initiator sequence strands having the selected barcode are isolated from the collection by hybridizing the selected barcode to an immobilized probe and removing remaining amplified barcoded initiator sequence strands and recovering the amplified barcoded initiator sequence strands immobilized to the probe.

37. A method of making a library of nucleic acid sequences with each nucleic acid sequence representing a word comprising

preparing a plurality of nucleic acid sequences each representing a word,

extending each of the plurality of nucleic acid sequences with a random sequence of nucleotides forming a unique nucleotide key sequence, with the unique nucleotide key sequence representing the word,

ligating a universal 3' nucleotide sequence to each of the unique nucleotide key sequences to form amplifiable strands, and

amplifying the amplifiable strands to form a collection of amplified barcoded nucleic acid sequences having unique nucleotide key sequences.

38. A method of encrypting one or more words comprising

selecting the one or more words,

sequencing a library of nucleic acid sequences, each sequence having a first sequence with a known association with a given word and having a second sequence of random nucleotides forming a unique nucleotide key sequence,

identifying the unique nucleotide key sequence for each of the one or more words,

and

associating the unique nucleotide key sequence for each of the one or more words.

39. The method of claim 37 wherein the collection of amplified barcoded nucleic acid sequences having unique nucleotide key sequences is ligated to the 3' end of an additional barcoded initiator.

40. The method of claim 37 wherein the collection of amplified barcoded nucleic acid sequences having unique nucleotide key sequences is independently ligated to the 3' end of a plurality of additionally barcoded initiators.

41. The method of claim 37 wherein the reverse complementary strands resulting from amplification of collection of amplified barcoded nucleic acid sequences having unique nucleotide key sequences is removed.

42. The method of claim 37 wherein the collection of amplified barcoded nucleic acid sequences having unique nucleotide key sequences is distributed in aliquots to communicating parties.

43. The method of claim 38 wherein the library of nucleic acid sequences is destroyed.

44. A method of encrypting one or more words comprising
selecting the one or more words,
isolating a library of nucleic acid sequences having a target common barcode sequence from a plurality of libraries of nucleic acids each having a different common barcode sequence,
sequencing the isolated library of nucleic acid sequences, each sequence having a first sequence with a known association with a given word and having a second sequence of random nucleotides forming a unique nucleotide key sequence,
identifying the unique nucleotide key sequence for each of the one or more words,
and
associating the unique nucleotide key sequence for each of the one or more words.

45. The method of claim 44 wherein the library of nucleic acid sequences is destroyed.

46. A method of decrypting one or more words comprising receiving one or more random nucleic acid sequences representative of one or more words,

isolating a library of nucleic acid sequences having a target common barcode sequence from a plurality of libraries of nucleic acids each having a different common barcode sequence,

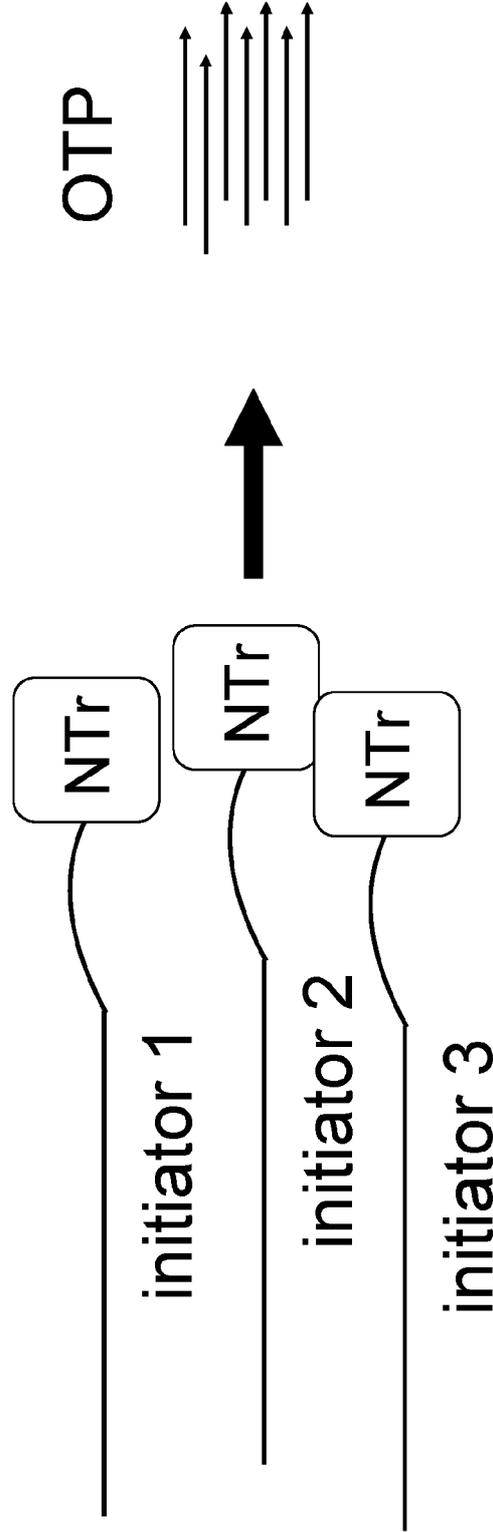
sequencing the isolated library of nucleic acid sequences, each sequence having a first sequence with a known association with a given word and having a second sequence of random nucleotides forming a unique nucleotide key sequence,

identifying the first sequence with a known association with a given word from the unique nucleotide key sequence, and

identifying the one or more words.

47. The method of claim 46 wherein the wherein the library of nucleic acid sequences is destroyed.

Fig. 1 Generation of one-time pad of nucleotide keys



- A pool of initiator oligonucleotides is synthesized. These initiators have unique nucleotide barcodes useful for addressing each and later, it's corresponding one-time key (i.e., ATCG = key 1, CTAG = key 2, etc).
- NTrs are used to extend each initiator to add a random sequence of nucleotides. Then, a universal 3' primer is ligated to every strand.
- PCR is performed, using universal primers, to amplify a subset (or all) of the strands, such that any aliquot of this amplified mixture will have full representation of each key. Aliquots of this amplified mixture are distributed between communicating parties.
- Each aliquot becomes a one-time pad of nucleotide keys, representing a collection of initiator oligonucleotides with added random nucleotide sequences. The random sequences represents the collection of keys in the one-time pad.

Fig. 2
**Example of using biological one-time pad of
 nucleotide keys for communication**



- This one-time pad consists of a library of one-time use keys.
- Alice and Bob both have DNA sequencing and synthesis capabilities.
- Alice and Bob have the same one-time pad and both agree on an order with which to use the keys. i.e., use key prefixed with “ACTG” first, “CATG” second, etc.
- Alice uses a one-time key from the one-time pad to encrypt a message she wishes to send to Bob.

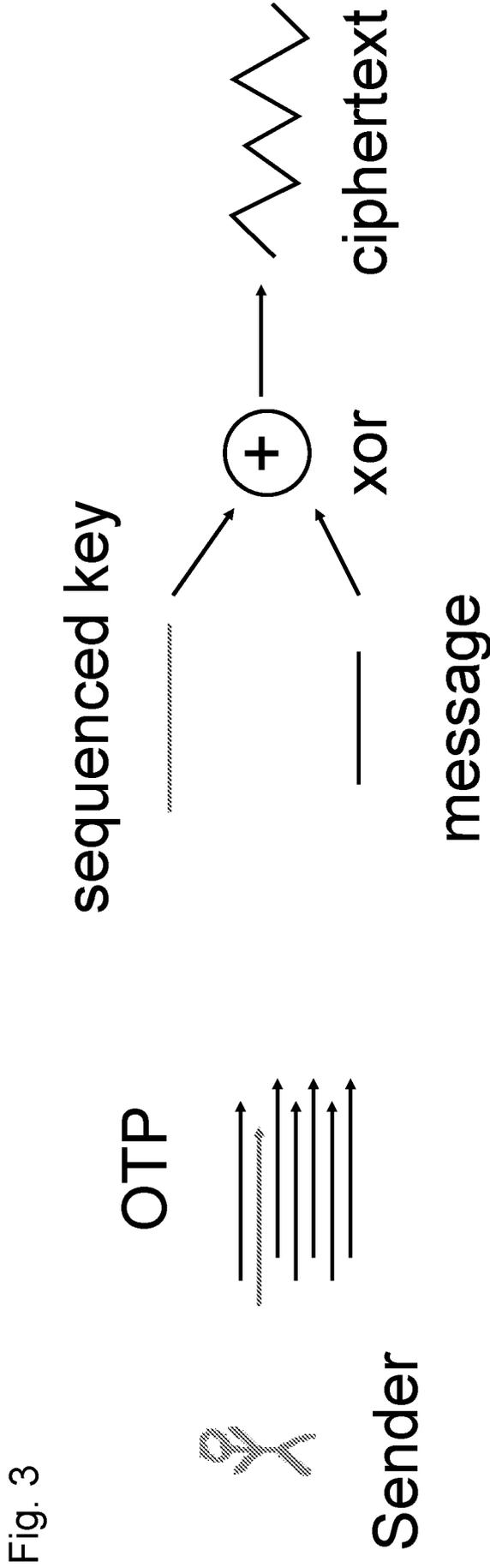
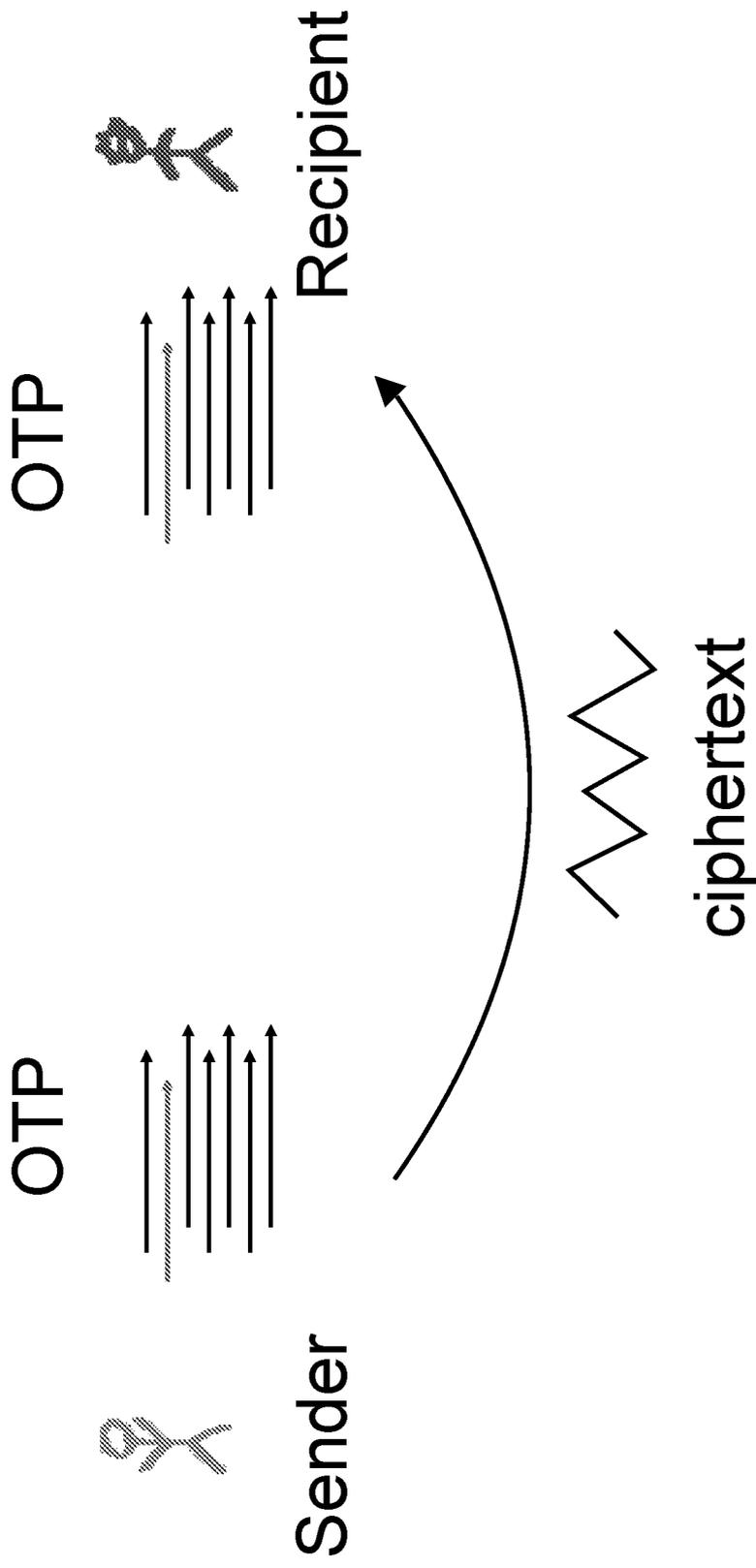


Fig. 3

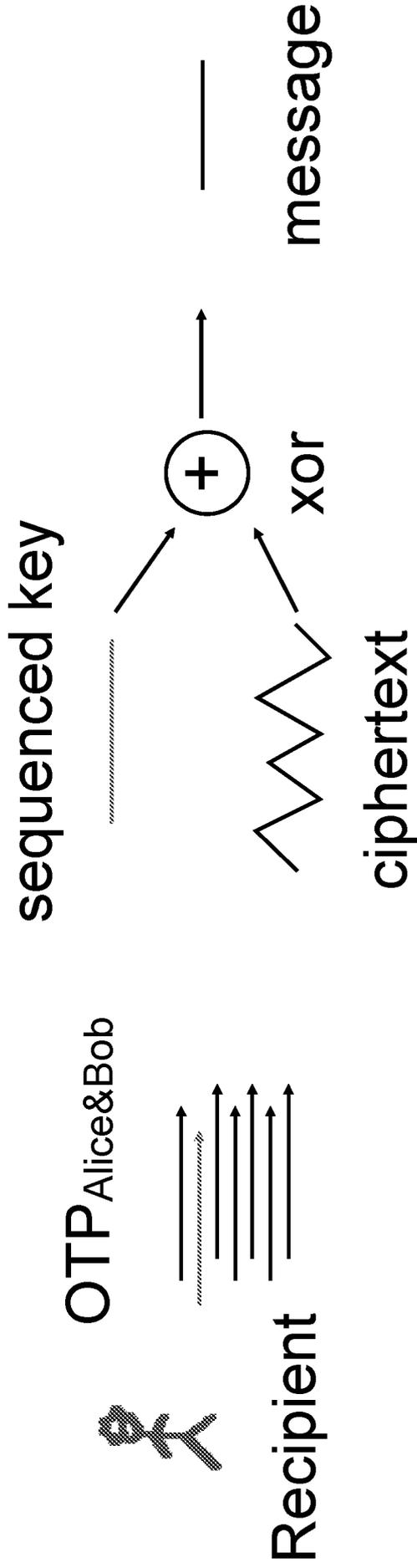
- Alice and Bob agree to use the key prefixed (barcoded) with “ACTG”.
- In order to encrypt her message – Alice needs to (1) reveal the key connected with “ACTG” prefix (2) use the key to encrypt her message (3) send to Bob.
 1. In order to reveal the key, Alice synthesizes the complementary strand “TGAC” and uses it to pull out (by hybridization) all instances of the specific key from the pool of keys in their shared one-time pad (OTP_{Alice&Bob}).
 2. Alice sequences the key and performs an xor between the key and the message, producing a ciphertext.
 3. Alice sends the ciphertext (digitally, or in encoded in DNA), to Bob.
- The used key is now depleted from the one-time pad, and is destroyed (by an active process such as DNase or chemical treatment or by throwing the DNA key into the environment)

Fig. 4



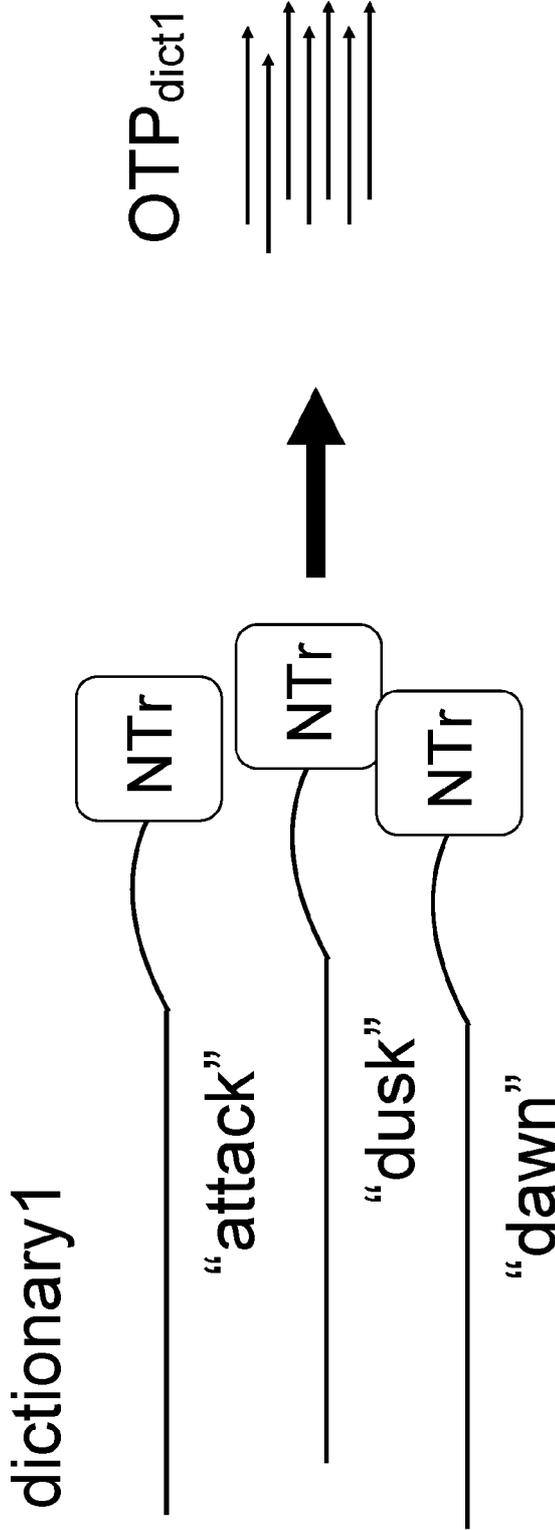
- Alice transmits the ciphertext to Bob (digitally or physically).

Fig. 5



- To decrypt the ciphertext from Alice, Bob synthesizes a complementary oligo “TGAC” and uses it to pull out (by hybridization) all instances of the specific key from the pool of keys in their shared one-time pad (OTP_{Alice&Bob}).
- Bob sequences the key and performs an xor operation between the key and the message, recovering the plaintext communication from Alice.
- The used key is now depleted from the one-time pad, and is destroyed (by an active process such as DNase or chemical treatment or by throwing the DNA key into the environment).

Fig. 6 One-time pad of dictionary keys



A one-time pad of dictionary keys is a special case of the one-time pad of nucleotide keys in which each initiator represents a word from the English dictionary.

- A pool of initiators is synthesized such that each oligo encodes a word (i.e. from the English dictionary) in DNA.
- NTrs are used to extend each initiator to produce a nucleotide polymer of random sequence. Then, a universal 3' primer is ligated to every strand.
- PCR is performed, using universal primers, to amplify a subset (or all) of the strands, such that any aliquot of this amplified mixture will have full representation of each key. Aliquots of this amplified mixture are distributed between communicating parties.
- This becomes a one-time pad of nucleotide keys as a dictionary, which is aliquotted between communicating parties.
- Each set of dictionary keys can be prefixed with a barcode (i.e., ATCG = dictionary 1, CTAG = dictionary 2, etc.) such that multiple pads can be generated and pooled.

Fig. 7

Example of using biological one-time pad of dictionary keys



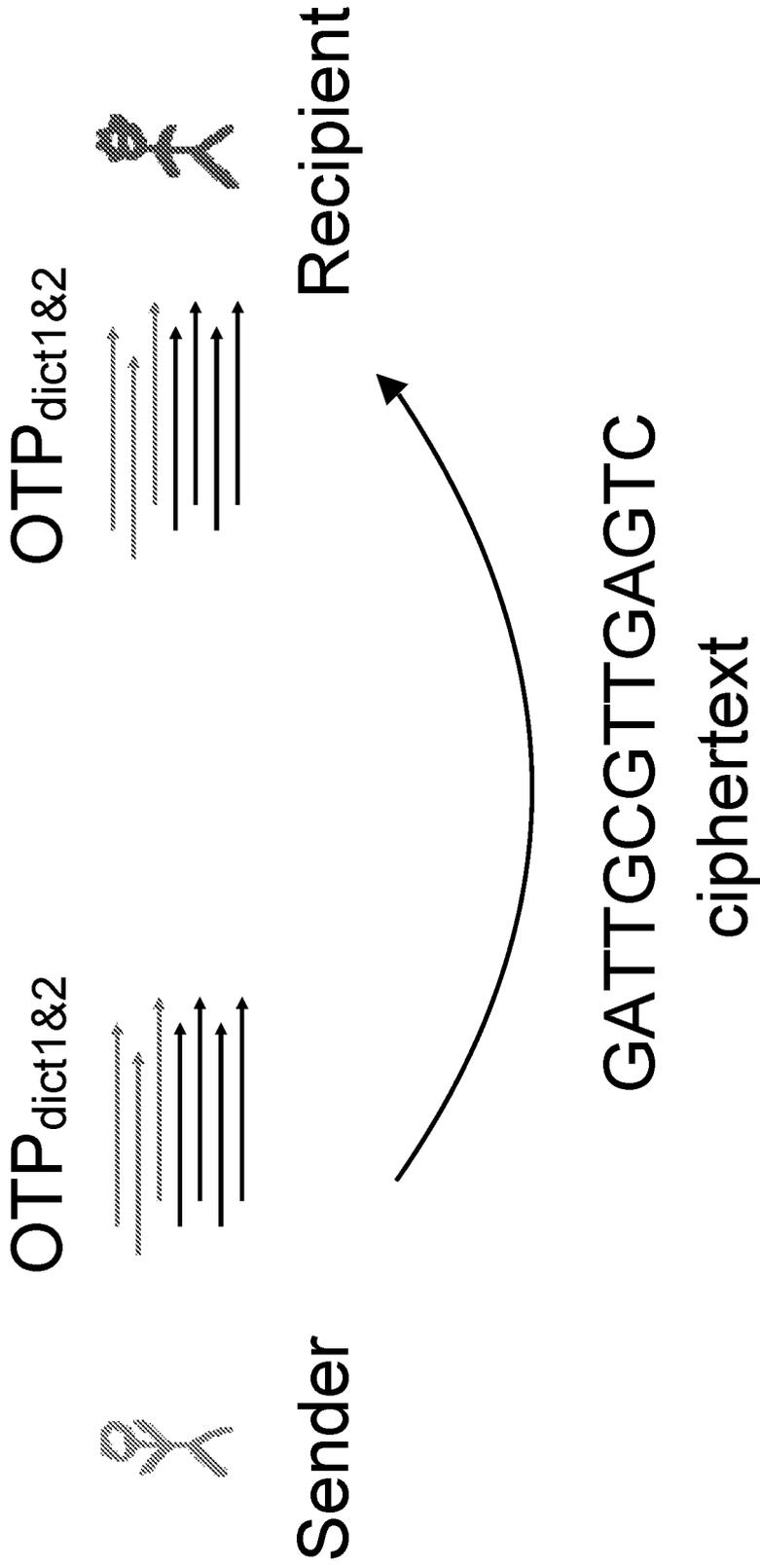
- This one-time pad consists of a library of one-time use dictionary keys.
- Alice and Bob have the same one-time pad and both agree on an order with which to use the keys.
- Alice and Bob both have DNA sequencing and synthesis capabilities.
- Alice uses a set of one-time dictionary keys from the one-time pad to encrypt a message she wishes to send to Bob.

Fig. 8



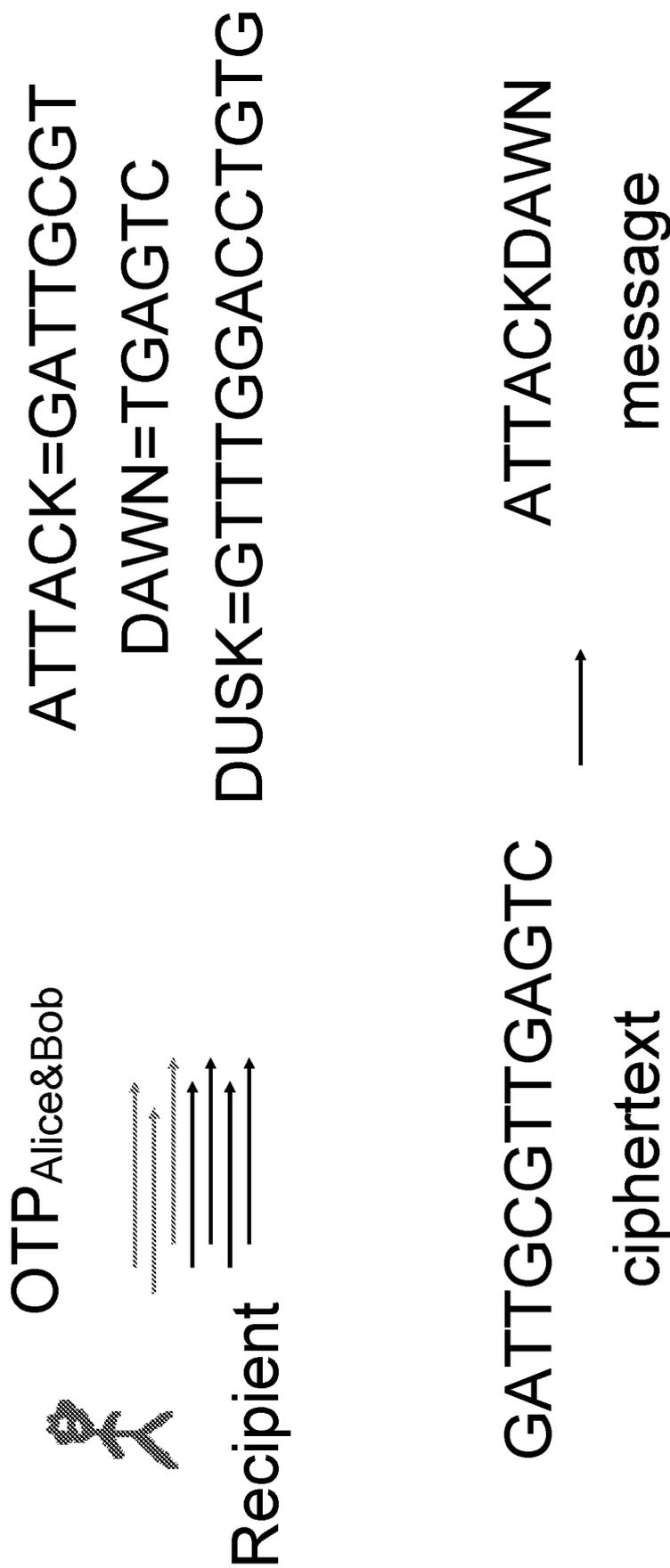
- Alice and Bob agree to use the key prefixed (barcoded) with "ACTG", marked in red.
- Alice synthesizes the complementary strand "TGAC" and uses it to pull out (by hybridization) all instances of the appropriate one-time key their shared one-time pad (OTP_{Alice&Bob}).
- Alice sequences the dictionary of keys and performs a substitution of each word with the corresponding key. i.e.,
- Alice sends the cipher text (digitally, or in encoded in DNA), to Bob.
- The used dictionary keys are now depleted from the one-time pad, and is destroyed (by an active process such as DNase or chemical treatment or by throwing the DNA key into the environment).

Fig. 9



- Alice transmits the ciphertext to Bob (digitally or physically).

Fig. 10



- Bob sequences the key dictionary and performs local alignments between the key and the message, recovering the plaintext communication from Alice.
- The used key is now depleted from the one-time pad, and is destroyed (by an active process such as DNase or chemical treatment or by throwing the DNA key into the environment).

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2017/029751

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C 12Q 1/68; G06N 3/00; G06N 3/1 2 (201 7.01)
CPC - C 12Q 1/68; G06F 19/22; G06N 3/1 23 (201 7.02)

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History document

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

USPC - 435/6; 506/2; 536/23.1 (keyword delimited)

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

See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X -- Y	WO 2000/068431 A2 (MOUNT SINAI SCHOOL OF MEDICINE OF NEW YORK UNIVERSITY) 16 November 2000 (16.1 1.2000) entire document	1-5, 7, 16-19 ----- 6, 8-15, 20-28, 42, 43
X -- Y	US 2015/026931 3 A1 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE) 24 September 2015 (24.09.2015) entire document	38 ----- 37, 39-43
Y	WO 2016/053891 A1 (THE REGENTS OF THE UNIVERSITY OF CALIFORNIA) 07 April 2016 (07.04.2016) entire document	6
Y	WO 2012/103154 A1 (NUGEN TECHNOLOGIES, INC. et al) 02 August 2012 (02.08.2012) entire document	8-12
Y	US 2016/0010152 A1 (SHAFER et al) 14 January 2016 (14.01.2016) entire document	13
Y	US 2008/0160529 A1 (BROW et al) 03 July 2008 (03.07.2008) entire document	14, 21
Y	US 2015/0125949 A1 (GENEART AG) 07 May 2015 (07.05.2015) entire document	15, 23-28
Y	MEYER et al. "Parallel tagged sequencing on the 454 platform," Nature Protocols, 31 January 2008 (31.01.2008), Vol. 3, Pgs. 267-278. entire document	20-22
Y	US 2003/0087240 A1 (WHITCOMBE et al) 08 May 2003 (08.05.2003) entire document	22

Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 31 July 2017	Date of mailing of the international search report 16 AUG 2017
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, VA 22313-1450 Facsimile No. 571-273-8300	Authorized officer Blaine R. Copenheaver PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US20 17/029751

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CHEN et al. "Reverse Transcriptase Adds Nontemplated Nucleotides to cDNAs During 5'-RACE and Primer Extension", BioTechniques, 01 March 2001 (01.03.2001), Vol. 30, Pgs. 574-582. entire document	24
Y	PARKHOMCHUK et al. "Transcriptome analysis by strand-specific sequencing of complementary DNA," Nucleic Acids Research, 20 July 2009 (20.07.2009), Vol. 37, Pgs. 1-7. entire document	27, 41
Y	US 2014/0255939 A1 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE et al) 11 September 2014 (11.09.2014) entire document	37, 39-42
Y	US 2014/0256568 A1 (RAINDANCE TECHNOLOGIES, INC.) 11 September 2014 (11.09.2014) entire document	39, 40