



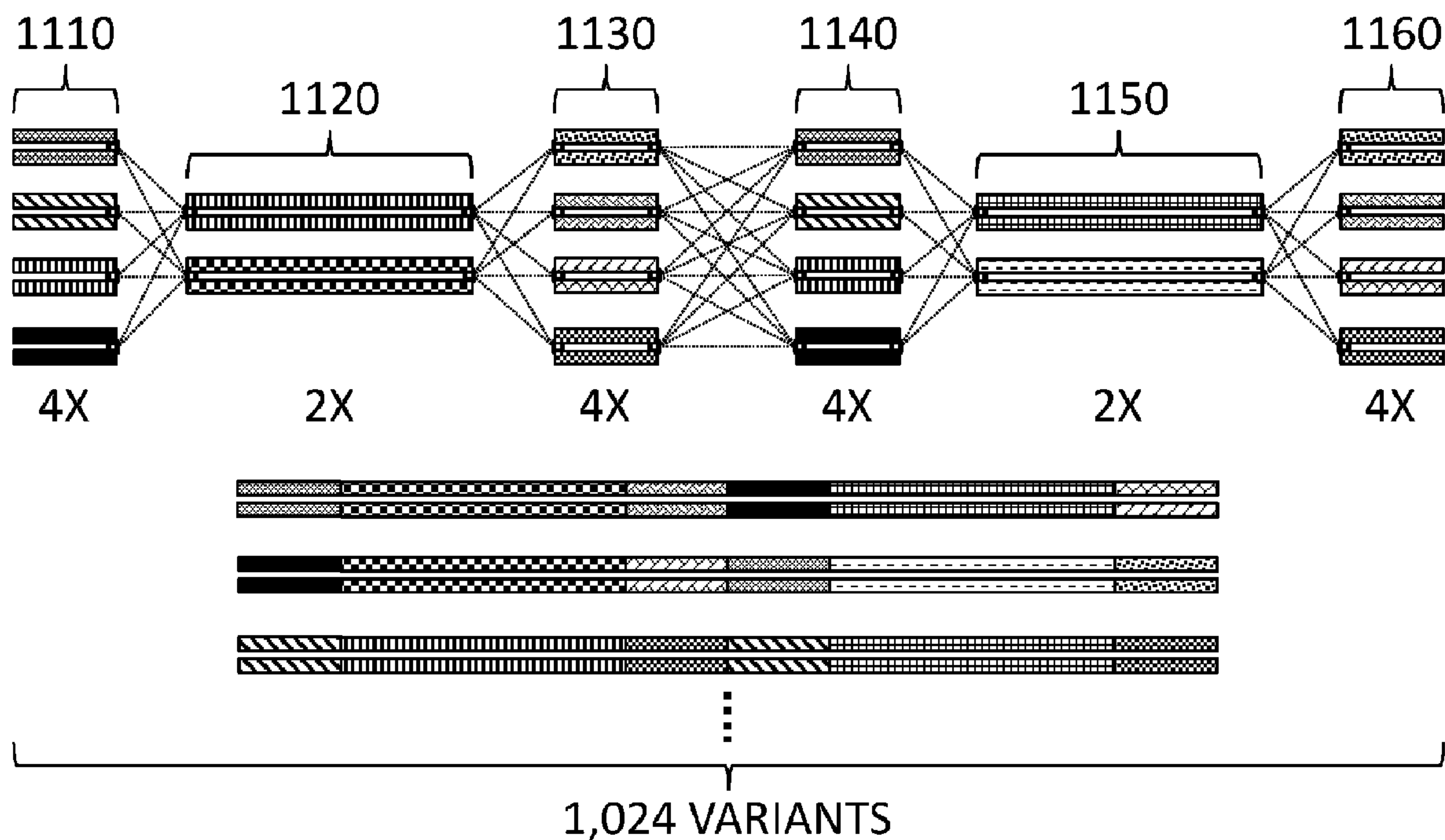
(12) **DEMANDE DE BREVET CANADIEN  
CANADIAN PATENT APPLICATION**

(13) **A1**

(86) Date de dépôt PCT/PCT Filing Date: 2018/03/14  
 (87) Date publication PCT/PCT Publication Date: 2018/09/20  
 (85) Entrée phase nationale/National Entry: 2019/09/12  
 (86) N° demande PCT/PCT Application No.: US 2018/022493  
 (87) N° publication PCT/PCT Publication No.: 2018/170169  
 (30) Priorité/Priority: 2017/03/15 (US62/471,810)

(51) Cl.Int./Int.Cl. *A61K 31/7088* (2006.01),  
*A61K 31/711* (2006.01), *C12N 15/10* (2006.01)  
 (71) Demandeur/Applicant:  
TWIST BIOSCIENCE CORPORATION, US  
 (72) Inventeurs/Inventors:  
COX, ANTHONY, US;  
CHEN, SIYUAN, US  
 (74) Agent: GOWLING WLG (CANADA) LLP

(54) Titre : BANQUES DE VARIANTS DE LA SYNAPSE IMMUNOLOGIQUE ET LEUR SYNTHÈSE  
 (54) Title: VARIANT LIBRARIES OF THE IMMUNOLOGICAL SYNAPSE AND SYNTHESIS THEREOF



**FIG. 11**

(57) **Abrégé/Abstract:**

Disclosed herein are methods for the generation of highly accurate nucleic acid libraries encoding for predetermined variants of a nucleic acid sequence. The nucleic acid sequence may encode for all or part of a reference domain of a CAR. The degree of

(57) **Abrégé(suite)/Abstract(continued):**

variation may be complete, resulting in a saturated variant library, or less than complete, resulting in a non-saturating library of variants. The variant nucleic acid libraries described herein may be designed for further processing by transcription or translation. The variant nucleic acid libraries described herein may be designed to generate variant RNA, DNA and/or protein populations. Further provided herein are method for identifying variant species with increased or decreased activities, with applications in regulating biological functions and the design of therapeutics for treatment or reduction of a disease, such as cancer.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property  
Organization  
International Bureau(10) International Publication Number  
**WO 2018/170169 A1**(43) International Publication Date  
20 September 2018 (20.09.2018)

## (51) International Patent Classification:

*A61K 31/7088* (2006.01)    *C12N 15/10* (2006.01)  
*A61K 31/711* (2006.01)

## (21) International Application Number:

PCT/US2018/022493

## (22) International Filing Date:

14 March 2018 (14.03.2018)

## (25) Filing Language:

English

## (26) Publication Language:

English

## (30) Priority Data:

62/471,810    15 March 2017 (15.03.2017)    US

(71) Applicant: **TWIST BIOSCIENCE CORPORATION**  
[US/US]; 455 Mission Bay Blvd. South, Suite 545, San Francisco, California 94158 (US).(72) Inventors: **COX, Anthony**; 455 Mission Bay Blvd. South, Suite 545, San Francisco, California 94158 (US). **CHEN, Siyuan**; 455 Mission Bay Blvd. South, Suite 545, San Francisco, California 94158 (US).(74) Agent: **HARBURGER, David**; Wilson Sonsini Goodrich & Rosati, 650 Page Mill Road, Palo Alto, California 94304 (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, JO, 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,

(54) Title: VARIANT LIBRARIES OF THE IMMUNOLOGICAL SYNAPSE AND SYNTHESIS THEREOF

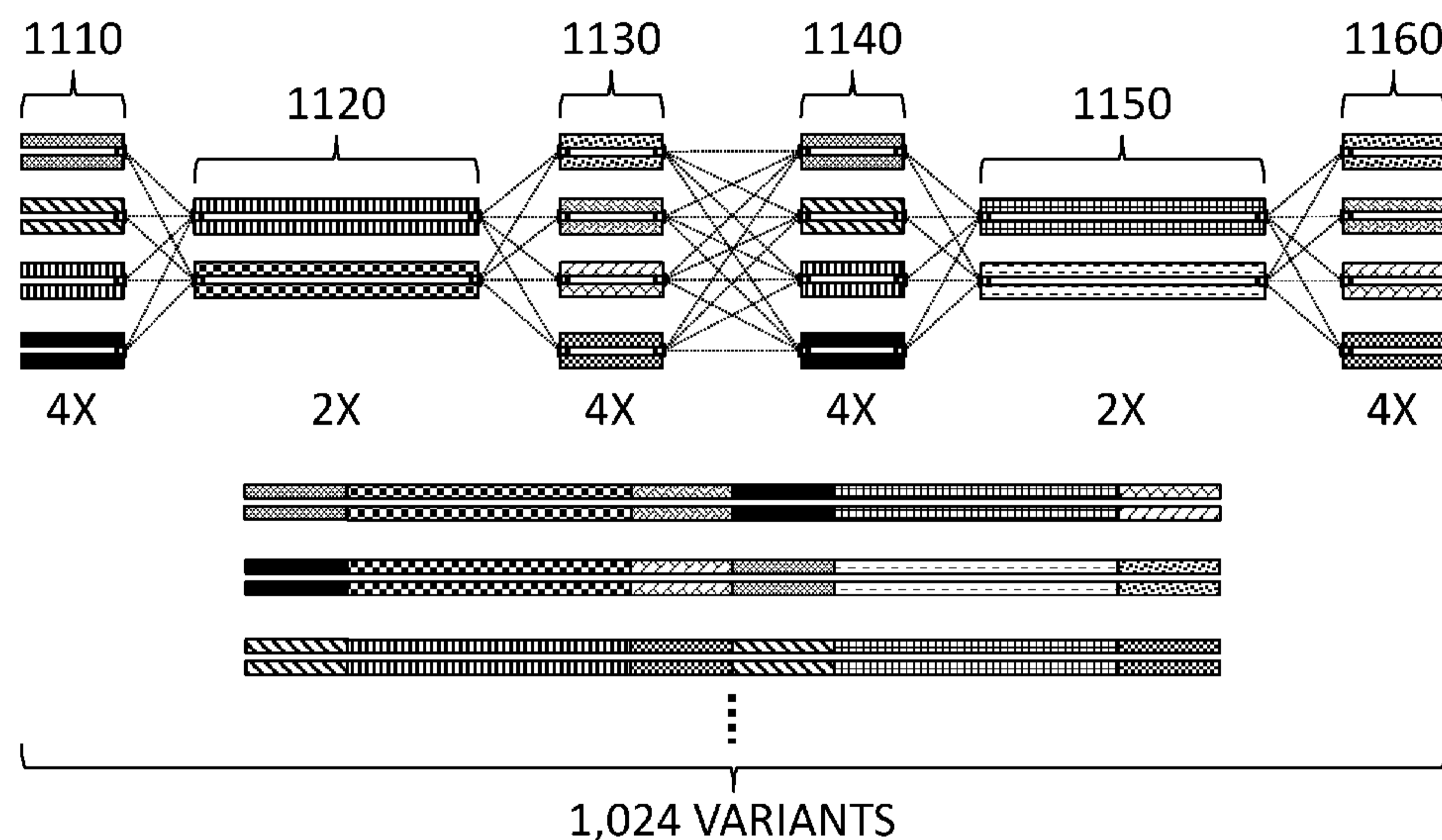


FIG. 11

(57) **Abstract:** Disclosed herein are methods for the generation of highly accurate nucleic acid libraries encoding for predetermined variants of a nucleic acid sequence. The nucleic acid sequence may encode for all or part of a reference domain of a CAR. The degree of variation may be complete, resulting in a saturated variant library, or less than complete, resulting in a non-saturating library of variants. The variant nucleic acid libraries described herein may be designed for further processing by transcription or translation. The variant nucleic acid libraries described herein may be designed to generate variant RNA, DNA and/or protein populations. Further provided herein are method for identifying variant species with increased or decreased activities, with applications in regulating biological functions and the design of therapeutics for treatment or reduction of a disease, such as cancer.

[Continued on next page]

**WO 2018/170169 A1** 

SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN,  
TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

**Declarations under Rule 4.17:**

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

**Published:**

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



## VARIANT LIBRARIES OF THE IMMUNOLOGICAL SYNAPSE AND SYNTHESIS THEREOF

### CROSS-REFERENCE

[0001] This application claims benefit of U.S. Provisional Patent Application No. 62/471,810 filed on March 15, 2017, which is incorporated herein by reference in its entirety.

### SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on March 9, 2018, is named 44854-733\_601\_SL.txt and is 19,476 bytes in size.

### BACKGROUND

[0003] The cornerstone of synthetic biology is the design, build, and test process—an iterative process that requires DNA, to be made accessible for rapid and affordable generation and optimization of these custom pathways and organisms. In the design phase, the A, C, T and G nucleotides that constitute DNA are formulated into the various gene sequences that would comprise the locus or the pathway of interest, with each sequence variant representing a specific hypothesis that will be tested. These variant gene sequences represent subsets of sequence space, a concept that originated in evolutionary biology and pertains to the totality of sequences that make up genes, genomes, transcriptome and proteome.

[0004] Many different variants are typically designed for each design-build-test cycle to enable adequate sampling of sequence space and maximize the probability of an optimized design. Though straightforward in concept, process bottlenecks around speed, throughput and quality of conventional synthesis methods dampen the pace at which this cycle advances, extending development time. The inability to sufficiently explore sequence space due to the high cost of acutely accurate DNA and the limited throughput of current synthesis technologies remains the rate-limiting step.

[0005] Beginning with the build phase, two processes are noteworthy: oligonucleotide synthesis and gene synthesis. Historically, synthesis of different gene variants was accomplished through molecular cloning. While robust, this approach is not scalable. Early chemical gene synthesis efforts focused on producing a large number of oligonucleotides with overlapping sequence homology. These were then pooled and subjected to multiple rounds of polymerase chain reaction (PCR), enabling concatenation of the overlapping oligonucleotides into a full length double stranded gene. A number of factors hinder this method, including time-consuming and labor-intensive construction, requirement of high volumes of phosphoramidites, an expensive raw

material, and production of nanomole amounts of the final product, significantly less than required for downstream steps, and a large number of separate oligonucleotides required one 96 well plate to set up the synthesis of one gene.

**[0006]** Synthesizing oligonucleotides on microarrays provided a significant increase in the throughput of gene synthesis. A large number of oligonucleotides could be synthesized on the microarray surface, then cleaved off and pooled together. Each oligonucleotide destined for a specific gene contains a unique barcode sequence that enabled that specific subpopulation of oligonucleotides to be depooled and assembled into the gene of interest. In this phase of the process, each subpool is transferred into one well in a 96 well plate, increasing throughput to 96 genes. While this is two orders of magnitude higher in throughput than the classical method, it still does not adequately support the design, build, test cycles that require thousands of sequences at one time due to a lack of cost efficiency and slow turnaround times. Thus, there is a need for more efficient generation of variant sequence libraries.

**[0007]** The immune system has the ability to seek and destroy harmful cells. T cells play an important role in such a process. As such, therapies comprising genetically modified T cells to teach the immune system to recognize and eliminate malignant cells are alternative therapies that may be used to treat diseases such as cancer. In particular, genetically modified T cells that express chimeric antigen receptors (CARs) may be designed to target deleterious cells, such as cancer cells. However, CAR therapies require further optimization due to their ability to cause undesirable effects such as on-target off-tumor toxicity, autoimmunity, host-graft toxicity, and sometimes death. Thus, there is a need for improved compositions and methods for cancer therapies utilizing CAR therapies.

### **INCORPORATION BY REFERENCE**

**[0008]** All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

### **BRIEF SUMMARY**

**[0009]** Provided herein are nucleic acid libraries, wherein the nucleic acid library comprises at least 10,000 nucleic acids, wherein each nucleic acid encodes for a preselected variant of a reference sequence that encodes for a chimeric antigen receptor or a functional domain thereof. Further provided herein are nucleic acid libraries, wherein the functional domain of the chimeric antigen receptor comprises an antigen recognition domain, a hinge domain, a transmembrane



domain, or an intracellular domain. Further provided herein are nucleic acid libraries, wherein the library comprises at least 1,000,000 nucleic acids. Further provided herein are nucleic acid libraries, wherein each nucleic acid is 100 bases to 2000 bases in length. Further provided herein are nucleic acid libraries, wherein each nucleic acid is attached to a vector sequence. Further provided herein are nucleic acid libraries, wherein the vector sequence is a viral vector sequence.

**[0010]** Provided herein are nucleic acid libraries, wherein the nucleic acid library comprises a plurality of nucleic acids, wherein each nucleic acid encodes for a preselected variant of a reference sequence that encodes for a chimeric antigen receptor or a functional domain thereof, wherein the plurality of nucleic acids comprises all variations for at least two positions in the reference sequence, and wherein the at least two positions encode sequences from 5 to 20 different codons. Further provided herein are nucleic acid libraries, wherein the functional domain of the chimeric antigen receptor comprises an antigen recognition domain, a hinge domain, a transmembrane domain, or an intracellular domain. Further provided herein are nucleic acid libraries, wherein the plurality of nucleic acids comprises all variations for 2, 3, 4, or 5 positions in the reference sequence. Further provided herein are nucleic acid libraries, wherein the at least two positions encode sequences from 10 to 20 different codons. Further provided herein are nucleic acid libraries, wherein the at least two positions encode sequences for about 10 different codons. Further provided herein are nucleic acid libraries, wherein each nucleic acid is 100 bases to 2000 bases in length.

**[0011]** Provided herein are oligonucleotide libraries, the oligonucleotide library comprising at least 10,000 oligonucleotides, wherein each oligonucleotide encodes for a preselected variant of a reference sequence that encodes for a region of a chimeric antigen receptor, wherein the region comprises a portion of an antigen recognition domain, a hinge domain, a transmembrane domain, or an intracellular domain. Further provided herein are oligonucleotide libraries, wherein the library comprises at least 1,000,000 oligonucleotides. Further provided herein are oligonucleotide libraries, wherein each oligonucleotide is 12 to 500 bases in length. Further provided herein are oligonucleotide libraries, wherein each oligonucleotide is attached to a vector sequence.

**[0012]** Provided herein are oligonucleotide libraries, wherein the oligonucleotide library comprises a plurality of oligonucleotides, wherein each oligonucleotide encodes for a preselected variant of a reference sequence that encodes for a region of a chimeric antigen receptor, wherein the region comprises a portion of an antigen recognition domain, a hinge domain, a transmembrane domain, or an intracellular domain, wherein each oligonucleotide is at least 12 bases in length, wherein the plurality of oligonucleotides comprises all variations for at least two positions in the reference sequence, and wherein the at least two positions encode sequences from 5 to 20 different



codons. Further provided herein are oligonucleotide libraries, wherein the plurality of oligonucleotides comprises all variations for 2, 3, 4, or 5 positions. Further provided herein are oligonucleotide libraries, wherein the at least two positions encode sequences from 10 to 20 different codons. Further provided herein are oligonucleotide libraries, wherein the at least two positions encode sequences for about 10 different codons. Further provided herein are oligonucleotide libraries, wherein each oligonucleotide is 12 to 500 bases in length.

**[0013]** Provided herein are nucleic acid libraries comprising at least 400 nucleic acids, wherein a first plurality of the at least 400 nucleic acids encodes for a variant of a reference sequence that encodes for an antigen recognition domain, and wherein a second plurality of the at least 400 nucleic acids encodes for a variant of a reference sequence that encodes for a hinge domain, a transmembrane domain, or an intracellular domain of a chimeric antigen receptor (CAR). Further provided herein are nucleic acid libraries, wherein each nucleic acid is at least 100 bases in length. Further provided herein are nucleic acid libraries, wherein each nucleic acid is 100 to 2000 bases in length. Further provided herein are nucleic acid libraries, wherein the first plurality of the at least 400 nucleic acids comprises all variations for at least two positions in the reference sequence. Further provided herein are nucleic acid libraries, wherein the first plurality of the at least 400 nucleic acids comprises all variations for 2, 3, 4, or 5 positions in the reference sequence. Further provided herein are nucleic acid libraries, wherein the second plurality of the at least 400 nucleic acids comprises all variations for at least two positions in the reference sequence. Further provided herein are nucleic acid libraries, wherein the second plurality of the at least 400 nucleic acids comprises all variations for 2, 3, 4, or 5 positions in the reference sequence. Further provided herein are nucleic acid libraries, wherein the at least two positions encode sequences from 5 to 20 different codons. Further provided herein are nucleic acid libraries, wherein the at least two positions encode sequences from 10 to 20 different codons. Further provided herein are nucleic acid libraries, wherein the at least two positions encode sequences for about 10 different codons.

**[0014]** Provided herein are methods of synthesizing a nucleic acid library, comprising: (a) providing a first set of preselected oligonucleotide sequences encoding for at least 400 sequences of a chimeric antigen receptor gene or gene fragment, wherein each sequence comprises at least one variation of at least two preselected codons for an amino acid residue in an antigen recognition domain; (b) synthesizing the first set of preselected oligonucleotide sequences; and (c) screening a first activity for proteins encoded by the first set of oligonucleotide sequences, wherein the first activity is specificity, avidity, affinity, stability, or expression. Further provided herein are methods of synthesizing a nucleic acid library, wherein the antigen is a cancer antigen. Further provided herein are methods of synthesizing a nucleic acid library, wherein the cancer antigen is MAGE A3,



MAGE A12, MAGE A2, MAGE A6, NY-ESO-1, or CEA. Further provided herein are methods of synthesizing a nucleic acid library, wherein each sequence comprises up to 100 variations at preselected codons for amino acid residues in the antigen recognition domain. Further provided herein are methods of synthesizing a nucleic acid library, wherein each sequence comprises up to 30 variations at preselected codons for amino acid residues in the antigen recognition domain. Further provided herein are methods of synthesizing a nucleic acid library further comprising (a) providing a second set of preselected oligonucleotide sequences encoding for at least one sequence of a chimeric antigen receptor gene or gene fragment, where each sequence comprises at least one variation at a preselected codon for an amino acid residue in the chimeric antigen receptor in a hinge domain, a transmembrane domain, or an intracellular domain; (b) synthesizing the second set of preselected oligonucleotide sequences; and (c) screening a second activity for proteins encoded by the second set of oligonucleotide sequences, wherein the second activity is specificity, avidity, affinity, stability, or expression, and wherein the second activity is different from the first activity.

**[0015]** Provided herein is an oligonucleotide library, the library comprising at least 10,000 variant oligonucleotides, wherein each variant oligonucleotide is at least 12 bases in length, wherein each variant oligonucleotide encodes for a variant of a reference sequence that encodes for an antigen recognition domain, a hinge domain, a transmembrane domain, or an intracellular domain of a chimeric antigen receptor, and wherein at least 80% of the variant oligonucleotides have no errors compared to the predetermined sequences without correcting errors. Further provided herein is an oligonucleotide library wherein the library comprises at least 1,000,000 variant oligonucleotides. Further provided herein is an oligonucleotide library, wherein each variant oligonucleotide is 20 to 500 bases in length. Further provided herein is an oligonucleotide library wherein at least 90% of the variant oligonucleotides have no errors compared to the predetermined sequences without correcting errors.

**[0016]** Provided herein is a nucleic acid library, the library comprising at least 10,000 variant nucleic acids, wherein each variant nucleic acid encodes for a variant of a reference sequence for a chimeric antigen receptor, wherein the variation occurs in the sequence for an antigen recognition domain, a hinge domain, a transmembrane domain, or an intracellular domain of a chimeric antigen receptor, and wherein at least 70% of the variant nucleic acids have no errors compared to the predetermined sequences without correcting errors. Further provided herein is a nucleic acid library wherein the library comprises about 10,000,000 variant nucleic acids. Further provided herein is a nucleic acid library wherein each variant nucleic acid is 600 to 3000 bases in length. Further provided herein is a nucleic acid library wherein each variant nucleic acid is attached to a



vector sequence. Further provided herein is a nucleic acid library wherein the vector sequence is a viral vector sequence.

**[0017]** Provided herein is a method for nucleic acid synthesis, the method comprising: providing predetermined sequences encoding for a plurality of non-identical oligonucleotides, wherein each of the non-identical oligonucleotides is at least 20 bases in length, and wherein the plurality of non-identical oligonucleotides encodes for up to 19 variants for each of at least 2 codons compared to at least one reference sequence, wherein the at least one reference sequence encodes for a region of an antigen recognition domain, a hinge domain, a transmembrane domain, or an intracellular domain of a chimeric antigen receptor; providing a structure having a surface, wherein the surface comprises clusters of loci for oligonucleotide extension, each cluster comprising 50 to 500 loci; synthesizing the plurality of non-identical oligonucleotides, wherein each of the non-identical oligonucleotides extends from a separate locus; and performing an amplification reaction to assemble a library of variant nucleic acids, wherein the amplification reaction comprises mixing the plurality of non-identical oligonucleotides from a single cluster with a DNA polymerase and the at least one reference sequence. Further provided herein is a method wherein each cluster further comprises additional oligonucleotides, and wherein the plurality of non-identical oligonucleotides and the additional oligonucleotides within said cluster collectively encode for a single gene and variants thereof. Further provided herein is a method wherein the surface of the structure comprises at least 6000 of the clusters. Further provided herein is a method wherein at least 80% of the variant nucleic acids have no errors compared to the predetermined sequences without correcting errors. Further provided herein is a method wherein at least 90% of the variant nucleic acids have no errors compared to the predetermined sequences without correcting errors. Further provided herein is a method wherein a replicate of each non-identical oligonucleotide extends from the surface at up to 5 loci. Further provided herein is a method wherein a replicate of each non-identical oligonucleotide extends from the surface at up to 3 loci. Further provided herein is a method wherein the plurality of non-identical oligonucleotides encodes for up to 19 variants for each of at least 3 codons compared to at least one reference sequence. Further provided herein is a method further comprising expressing the variant nucleic acid library, wherein expression of the variant nucleic acid library provides for a variant protein having improved affinity, improved gene expression, improved avidity, improved stability, or improved target specificity compared to the expression product of the reference sequence.

**[0018]** Provided herein is a method for nucleic acid synthesis, the method comprising: providing predetermined sequences encoding for a plurality of non-identical oligonucleotides, wherein each of the non-identical oligonucleotides is at least 20 bases in length, and wherein the



plurality of non-identical oligonucleotides encodes for up to 19 variants for each of at least 2 codons compared to at least one reference sequence, wherein the at least one reference sequence encodes for an epitope to a chimeric antigen receptor; providing a structure having a surface, wherein the surface comprises clusters of loci for oligonucleotide extension, each cluster comprising 50 to 500 loci; synthesizing the plurality of non-identical oligonucleotides, wherein each of the non-identical oligonucleotides extends from a separate locus; and performing an amplification reaction to assemble a library of variant nucleic acids, wherein the amplification reaction comprises mixing the plurality of non-identical oligonucleotides from a single cluster with a DNA polymerase and the at least one reference sequence. Further provided herein is a method wherein the surface of the structure comprises at least 6000 of the clusters. Further provided herein is a method wherein at least about 80% of the variant nucleic acids have no errors compared to the predetermined sequences without correcting errors. Further provided herein is a method wherein at least about 90% of the variant nucleic acids have no errors compared to the predetermined sequences without correcting errors. Further provided herein is a method wherein a replicate of each non-identical oligonucleotide extends from the surface at up to 5 loci. Further provided herein is a method wherein a replicate of each non-identical oligonucleotide extends from the surface at up to 3 loci. Further provided herein is a method wherein the plurality of non-identical oligonucleotides encodes for up to 19 variants for each of at least 3 codons compared to at least one reference sequence. Further provided herein is a method further comprising expressing the variant nucleic acid library, wherein expression of the variant nucleic acid library provides for a variant peptide having improved affinity, improved gene expression, improved avidity, improved stability, or improved target specificity compared to the expression product of the reference sequence. Further provided herein is a method wherein the at least one reference sequence encodes for the epitope to a chimeric antigen receptor and also for an epitope to a surface protein expressed by a non-lymphoid cell.

**[0019]** Provided herein is a method for nucleic acid synthesis of nucleic acid libraries for optimization assays, the method comprising: providing a first set of predetermined sequences encoding for a first plurality of non-identical oligonucleotides, wherein each of the non-identical oligonucleotides is at least 20 bases in length, and wherein the first plurality of non-identical oligonucleotides encodes for up to 19 variants for each of at least 3 codons compared to at least one reference sequence, wherein the at least one reference sequence encodes for a region of an antigen recognition domain, a hinge domain, a transmembrane domain, or an intracellular domain of a chimeric antigen receptor; providing a second set of predetermined sequences encoding for a second plurality of non-identical oligonucleotides, wherein each of the second non-identical



oligonucleotides is at least 20 bases in length, and wherein the second plurality of non-identical oligonucleotides encodes for up to 19 variants for each of at least 3 codons compared to at least one reference sequence, wherein the at least one reference sequence encodes for an epitope to the chimeric antigen receptor; providing a structure having a surface, wherein the surface comprises clusters of loci for oligonucleotide extension, each cluster comprising 50 to 500 loci; synthesizing both the first plurality of non-identical oligonucleotides and the second plurality of non-identical oligonucleotides, wherein each non-identical oligonucleotide extends from a separate locus; and performing an amplification reaction to assemble a first library of nucleic acids, wherein the amplification reaction comprises mixing the first plurality of non-identical oligonucleotides from a single cluster with a DNA polymerase and the at least one reference sequence; performing an amplification reaction to assemble a second library of variant nucleic acids, wherein the amplification reaction comprises mixing the second plurality of non-identical oligonucleotides from a single cluster with a DNA polymerase and the at least one reference sequence. Further provided herein is a method further comprising screening purified protein expression products of the second library of variant nucleic acids against a library of cells expressing the first library of variant nucleic acids, wherein the library of cells comprises populations of cells, each population of cells expressing a different protein expression product encoded by the first library of variant nucleic acids. Further provided herein is a method wherein the at least one reference sequence for an epitope to the chimeric antigen receptor also encodes for an epitope to a surface protein expressed by a non-lymphoid cell.

### BRIEF DESCRIPTION OF THE DRAWINGS

[0020] **Figures 1A-1D** depict a process workflow for the synthesis of variant biological molecules incorporating a PCR mutagenesis step.

[0021] **Figures 2A-2D** depict a process workflow for the generation of a nucleic acid comprising a nucleic acid sequence which differs from a reference nucleic acid sequence at a single predetermined codon site.

[0022] **Figures 3A-3F** depict an alternative workflow for the generation of a set of nucleic acid variants from a template nucleic acid, with each variant comprising a different nucleic acid sequence at a single codon position. Each variant nucleic acid encodes for a different amino acid at their single codon position, the different codons represented by X, Y, and Z.

[0023] **Figures 4A-4E** depict a reference amino acid sequence (**FIG. 4A**) having a number of amino acids, each residue indicated by a single circle, and variant amino acid sequences (**FIGS. 4B, 4C, 4D, & 4E**) generated using methods described herein. The reference amino acid sequence



and variant sequences are encoded by nucleic acids and variants thereof generated by processes described herein.

**[0024]** **Figures 5A-5B** depict a reference amino acid sequence (**FIG. 5A**) and a library of variant amino acid sequences (**FIG. 5B**), each variant comprising a single residue variant (indicated by an “X”). The reference amino acid sequence and variant sequences are encoded by nucleic acids and variants thereof generated by processes described herein. **Figure 5A** discloses SEQ ID NO: 42 and **Figure 5B** discloses SEQ ID NOS 43-49, respectively, in order of appearance.

**[0025]** **Figures 6A-6B** depict a reference amino acid sequence (**FIG. 6A**) and a library of variant amino acid sequences (**FIG. 6B**), each variant comprising two sites of single position variants. Each variant is indicated by differently patterned circles. The reference amino acid sequence and variant sequences are encoded by nucleic acids and variants thereof generated by processes described herein.

**[0026]** **Figures 7A-7B** depict a reference amino acid sequence (**FIG. 7A**) and a library of variant amino acid sequences (**FIG. 7B**), each variant comprising a stretch of amino acids (indicated by a box around the circles), each stretch having three sites of position variants (encoding for histidine) differing in sequence from the reference amino acid sequence. The reference amino acid sequence and variant sequences are encoded by nucleic acids and variants thereof generated by processes described herein.

**[0027]** **Figures 8A-8B** depict a reference amino acid sequence (**FIG. 8A**) and a library of variant amino acid sequences (**FIG. 8B**), each variant comprising two stretches of amino acid sequence (indicated by a box around the circles), each stretch having one site of single position variants (illustrated by the patterned circles) differing in sequence from reference amino acid sequence. The reference amino acid sequence and variant sequences are encoded by nucleic acids and variants thereof generated by processes described herein.

**[0028]** **Figures 9A-9B** depict a reference amino acid sequence (**FIG. 9A**) and a library of amino acid sequence variants (**FIG. 9B**), each variant comprising a stretch of amino acids (indicated by patterned circles), each stretch having a single site of multiple position variants differing in sequence from the reference amino acid sequence. In this illustration, 5 positions are varied where the first position has a 50/50 K/R ratio; the second position has a 50/25/25 V/L/S ratio, the third position has a 50/25/25 Y/R/D ratio, the fourth position has an equal ratio for all amino acids, and the fifth position has a 75/25 ratio for G/P. The reference amino acid sequence and variant sequences are encoded by nucleic acids and variants thereof generated by processes described herein.

[0029] **Figures 10A-10C** illustrate generations of chimeric antigen receptors (CARs). CARs comprise an extracellular single chain variable fragment (scFv), a hinge domain (H), a transmembrane domain (TM), and an intracellular domain. An intracellular domain of a first generation CAR comprises CD3 $\zeta$  alone (**FIG. 10A**). An intracellular domain of a second generation CAR comprises a costimulatory domain and CD3 $\zeta$  (**FIG. 10B**). An intracellular domain of a third generation CAR comprises two costimulatory domains and CD3 $\zeta$  (**FIG. 10C**). In **Figures 10D-10E**, various CAR binding arrangements are illustrated.

[0030] **Figure 11** depicts an exemplary number of variants produced by interchanging sections of two expression cassettes (e.g., promoters, open reading frames, and terminators) to generate a variant library of expression cassettes.

[0031] **Figure 12** presents a diagram of steps demonstrating an exemplary process workflow for gene synthesis as disclosed herein.

[0032] **Figure 13** illustrates an example of a computer system.

[0033] **Figure 14** is a block diagram illustrating an architecture of a computer system.

[0034] **Figure 15** is a diagram demonstrating a network configured to incorporate a plurality of computer systems, a plurality of cell phones and personal data assistants, and Network Attached Storage (NAS).

[0035] **Figure 16** is a block diagram of a multiprocessor computer system using a shared virtual address memory space.

[0036] **Figure 17** depicts a BioAnalyzer plot of PCR reaction products resolved by gel electrophoresis.

[0037] **Figure 18** depicts an electropherogram showing 96 sets of PCR products, each set of PCR products differing in sequence from a wild-type template nucleic acid at a single codon position, where the single codon position of each set is located at a different site in the wild-type template nucleic acid sequence. Each set of PCR products comprises 19 variant nucleic acids, each variant encoding for a different amino acid at their single codon position.

### DETAILED DESCRIPTION

[0038] The present disclosure employs, unless otherwise indicated, conventional molecular biology techniques, which are within the skill of the art. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of ordinary skill in the art.

[0039] **Definitions**

[0040] Throughout this disclosure, numerical features are presented in a range format. It should be understood that the description in range format is merely for convenience and brevity and



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

**[0041]** The terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting of any embodiment. As used herein, the singular forms “a,” “an” and “the” are intended to include the plural forms as well, unless the context clearly indicates otherwise. It will be further understood that the terms “comprises” and/or “comprising,” when used in this specification, specify the presence of stated features, integers, steps, operations, elements, and/or components, but do not preclude the presence or addition of one or more other features, integers, steps, operations, elements, components, and/or groups thereof. As used herein, the term “and/or” includes any and all combinations of one or more of the associated listed items.

**[0042]** Unless specifically stated or obvious from context, as used herein, the term “about” in reference to a number or range of numbers is understood to mean the stated number and numbers +/- 10% thereof, or 10% below the lower listed limit and 10% above the higher listed limit for the values listed for a range.

**[0043]** The term “nucleic acid” encompasses double- or triple-stranded nucleic acids, as well as single-stranded molecules. In double- or triple-stranded nucleic acids, the nucleic acid strands need not be coextensive (i.e., a double-stranded nucleic acid need not be double-stranded along the entire length of both strands). Nucleic acid sequences, when provided, are listed in the 5' to 3' direction, unless stated otherwise. Methods described herein provide for the generation of isolated nucleic acids. Methods described herein additionally provide for the generation of isolated and purified nucleic acids.

**[0044]** As used herein, the terms “preselected sequence,” “predefined sequence” or “predetermined sequence” are used interchangeably. The terms mean that the sequence of the polymer is known and chosen before synthesis or assembly of the polymer. In particular, various



aspects of the invention are described herein primarily with regard to the preparation of nucleic acids molecules, the sequence of the oligonucleotide or polynucleotide being known and chosen before the synthesis or assembly of the nucleic acid molecules.

**[0045]** Provided herein are methods and compositions for production of synthetic (i.e. de novo synthesized or chemically synthesized) polynucleotides. The term oligonucleotide, oligo, oligonucleic acid, and polynucleotide are defined to be synonymous throughout. Libraries of synthesized polynucleotides described herein may comprise a plurality of polynucleotides collectively encoding for one or more genes or gene fragments. In some instances, the polynucleotide library comprises coding or non-coding sequences. In some instances, the polynucleotide library encodes for a plurality of cDNA sequences. Reference gene sequences from which the cDNA sequences are based may contain introns, whereas cDNA sequences exclude introns. Polynucleotides described herein may encode for genes or gene fragments from an organism. Exemplary organisms include, without limitation, prokaryotes (e.g., bacteria) and eukaryotes (e.g., mice, rabbits, humans, and non-human primates). In some instances, the polynucleotide library comprises one or more polynucleotides, each of the one or more polynucleotides encoding sequences for multiple exons. Each polynucleotide within a library described herein may encode a different sequence, i.e., non-identical sequence. In some instances, each polynucleotide within a library described herein comprises at least one portion that is complementary to a sequence of another polynucleotide within the library. Polynucleotide sequences described herein may, unless stated otherwise, comprise DNA or RNA.

**[0046]** Provided herein are methods and compositions for production of synthetic (i.e. de novo synthesized) genes. Libraries comprising synthetic genes may be constructed by a variety of methods described in further detail elsewhere herein, such as PCA, non-PCA gene assembly methods or hierarchical gene assembly, combining (“stitching”) two or more double-stranded oligonucleotides to produce larger DNA units (i.e., a chassis). Libraries of large constructs may involve oligonucleotides that are at least 1, 1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 400, 500 kb long or longer. The large constructs may be bounded by an independently selected upper limit of about 5000, 10000, 20000 or 50000 base pairs. The synthesis of any number of polypeptide-segment encoding nucleotide sequences, including sequences encoding non-ribosomal peptides (NRPs), sequences encoding non-ribosomal peptide-synthetase (NRPS) modules and synthetic variants, polypeptide segments of other modular proteins, such as antibodies, polypeptide segments from other protein families, including non-coding DNA or RNA, such as regulatory sequences e.g. promoters, transcription factors, enhancers, siRNA, shRNA, RNAi, miRNA, small nucleolar RNA derived from microRNA, or any functional



or structural DNA or RNA unit of interest. The following are non-limiting examples of polynucleotides: coding or non-coding regions of a gene or gene fragment, intergenic DNA, loci (locus) defined from linkage analysis, exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, short interfering RNA (siRNA), short-hairpin RNA (shRNA), micro-RNA (miRNA), small nucleolar RNA, ribozymes, complementary DNA (cDNA), which is a DNA representation of mRNA, usually obtained by reverse transcription of messenger RNA (mRNA) or by amplification; DNA molecules produced synthetically or by amplification, genomic DNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers. cDNA encoding for a gene or gene fragment referred to herein, may comprise at least one region encoding for exon sequence(s) without an intervening intron sequence found in the corresponding genomic sequence. Alternatively, the corresponding genomic sequence to a cDNA may lack an intron sequence in the first place.

**[0047] Chimeric Antigen Receptor**

**[0048]** Chimeric antigen receptors (CARs) are generally engineered receptors that recognize a particular antigen such as an antigen whose epitope is unique to cancer cells. Referring to **FIGS. 10A-10C**, CARs comprise an antigen recognition domain derived from a single chain antibody (scFv), a hinge domain (H) or spacer, a transmembrane domain (TM) that anchors the CAR to the plasma membrane, and an intracellular domain that mediates T cell activation. An intracellular domain of a first generation CAR comprises CD3 $\zeta$  (**FIG. 10A**). An intracellular domain of a second generation CAR comprises a costimulatory domain and CD3 $\zeta$  (**FIG. 10B**). An intracellular domain of a third generation CAR comprises two costimulatory domains and CD3 $\zeta$  (**FIG. 10C**). Exemplary CAR binding schematics are depicted in **FIGS. 10D-10E**. The CAR may be expressed in a lymphoid cell, e.g., a T cell, and an epitope to the CAR may be expressed in a non-lymphoid cell, or in an in vitro matrix, such as a bead, gel or column. *See e.g., FIG. 10D*. In some arrangements, the CAR is capable of binding to an epitope of a “linker” protein or peptide, which is able to also serve as an epitope to a surface protein of another surface, e.g., a cell, bead, gel or column. *See e.g., FIG. 10E*.

**[0049] Engineering Variance in Chimeric Antigen Receptor**

**[0050]** Provided herein are methods for synthesis of variant nucleic acid libraries, wherein each variant nucleic acid encodes for a sequence that is varied in comparison to a reference domain within a CAR. For example, the varied sequence is a nucleic acid sequence that encodes for an antigen recognition domain, a hinge domain, a transmembrane domain, or an intracellular domain. In some instances, the intracellular domain comprises a signaling domain. In some instances, the



intracellular domain further comprises a costimulatory domain. In some instances, the variant nucleic acid libraries comprise variant sequences that encode for all domains of a CAR. The resulting nucleic acid library may be an oligonucleotide library, gene fragment library, or gene library. In some instances, the nucleic acid library is expressed in cells to generate a variant protein library.

**[0051]** Variant nucleic acid libraries generated by methods disclosed herein may encode for an antigen recognition domain. An antigen recognition domain may comprise a single chain variable fragment (scFv). In some instances, an antigen recognition domain comprises F(ab')<sub>2</sub>, Fab', Fab, or Fv. Variant nucleic acid libraries that encode for antigen recognition domains may be designed and synthesized for a particular tumor antigen. Exemplary tumor antigens include, but are not limited to,  $\alpha$ -Folate receptor, BCMA, CAIX, CD123, CD138, CD171, CD19, CD20, CD22, CD30, CD33, CD44, CD44v7/8, CEA, cMET, DNAM-1, EDB-F, EGFR, EGFRvIII, EGP-2, EGP-40, EpCAM, FAP, Folate-binding Protein, GD2, GD3, glypican-3, h5T4, HER2, HER3, IL-13, IL-13R-a2, kappa light chain, KDR, Lewis Y, LMP-1, MAGE-A1, mesothelin, MUC-1, MUC-16, PSCA, PSMA, ROR1, TAG-72, VEGFR, and VEGFR2. In some instances, about 25, 50, 100, 250, 500, 1000, or more than 1000 common coding gene sequences of an antigen recognition domain is selected for variation. Variation may include designing of at least 25, 50, 100, 250, 500, 1000, 2000, 5000, 10000, 20000, 50000, 100000, or more than 100000 variants for each of the common coding gene sequences selected.

**[0052]** The hinge domain and transmembrane domain may connect the antigen recognition domain to the intracellular domains and anchor the CAR in the T cell membrane. In some instances, variant nucleic acid libraries are generated that encode for the hinge domain. Exemplary hinge domains are immunoglobulin (e.g., IgG1, IgG2, IgG3, IgG4, IgA1, IgA2, IgM, IgD or IgE), CH2CH3 region of immunoglobulin and optionally portions of CD3, and CD8 $\alpha$ . In some cases, variant nucleic acid libraries are generated that encode for the transmembrane domain.

**[0053]** Variant nucleic acid libraries generated by methods disclosed herein may comprise sequences that are varied to an intracellular domain. In some instances, the intracellular domain comprises a costimulatory domain. In some instances, variant nucleic acid libraries comprise sequences that are varied to a reference costimulatory domain of a CAR. Exemplary costimulatory domains include, but not limited to, CD8, CD27, CD28, 4-1BB (CD137), ICOS, DAP10, OX40 (CD134) or fragments or combinations thereof. In some instances, about 100, 250, 500, 1000, or more than 1000 common coding gene sequences of a costimulatory domain is selected for variation. Variation includes designing of at least about 500, 1000, 2000, 5000, 10000, 20000,



50000, 100000, or more than 100000 variants for each of the common coding gene sequences selected. Exemplary costimulatory domain sequences are shown in **Table 1**.

**Table 1. Costimulatory Domains**

Name	SEQ ID NO	Accession Number	Nucleic Acid Sequence
<b>Human ICOS</b>	24	AAL40933.1	ATGAAGTCAGGCCTCTGGTATTTCTTTCTCTTCTGC TTGCGCATTAAAGTTTTAACAGGAGAAATCAATGG TTCTGCCAATTATGAGATGTTTATATTTTACAACG GAGGTGTACAAATTTTATGCAAATATCCTGACATT GTCCAGCAATTTAAAATGCAGTTGCTGAAAGGGG GGCAAATACTCTGCGATCTCACTAAGACAAAAGG AAGTGGAAACACAGTGTCCATTAAGAGTCTGAAA TTCTGCCATTCTCAGTTATCCAACAACAGTGTCTCT TTTTTTCTATACAACTTGGACCATTCTCATGCCAAC TATTACTTCTGCAACCTATCAATTTTTGATCCTCCT CCTTTTAAAGTAACTCTTACAGGAGGATATTTGCA TATTTATGAATCACAACCTTTGTTGCCAGCTGAAGT TCTGGTTACCCATAGGATGTGCAGCCTTTGTTGTA GTCTGCATTTTGGGATGCATACTTATTTGTTGGCTT ACAAAAAGAAGTATTCATCCAGTGTGCACGACC CTAACGGTGAATACATGTTTCATGAGAGCAGTGAA CACAGCCAAAAAATCTAGACTCACAGATGTGACC CTATAA
<b>Human OX40</b>	25	BAG60249.1.	ATGGTATCACATCGGTATCCTCGAATTCAAAGTAT CAAAGTACAATTTACCGAATATAAGAAGGAGAAA GGTTTCATCCTCACTTCCCAAAGGAGGATGAAAT CATGAAGGTGCAGAACAACTCAGTCATCATCAAC TGTGATGGGTTTTATCTCATCTCCCTGAAGGGCTA CTTCTCCAGGAAGTCAACATTAGCCTTCATTACC AGAAGGATGAGGAGCCCCTCTTCCAAGTGAAGAA GGTCAGGTCTGTCAACTCCTTGATGGTGGCCTCTC TGACTTACAAAGACAAAGTCTACTTGAATGTGACC ACTGACAATACCTCCCTGGATGACTTCCATGTGAA TGGCGGAGAACTGATTCTTATCCATCAAATCCTG GTGAATTCTGTGTCCTTTGA
<b>Human 4-1BB (CD137)</b>	26	AAR05440.1	ATGGGAAACAGCTGTTACAACATAGTAGCCACTCT GTTGCTGGTCCTCAACTTTGAGAGGACAAGATCAT TGCAGGATCCTTGTAAGTAACTGCCAGCTGGTACA TTCTGTGATAATAACAGGAATCAGATTTGCAGTCC CTGTCCTCCAAATAGTTTCTCCAGCGCAGGTGGAC AAAGGACCTGTGACATATGCAGGCAGTGTAAAGG TGTTTTCAGGACCAGGAAGGAGTGTTCCTCCACCA GCAATGCAGAGTGTGACTGCACTCCAGGGTTTCAC TGCCTGGGGGCAGGATGCAGCATGTGTGAACAGG ATTGTAAACAAGGTCAAGAACTGACAAAAAAGG TTGTAAAGACTGTTGCTTTGGGACATTTAACGATC AGAAACGTGGCATCTGTCGACCCTGGACAAACTG TTCTTTGGATGGAAAGTCTGTGCTTGTGAATGGGA CGAAGGAGAGGGACGTGGTCTGTGGACCATCTCC AGCCGACCTCTCTCCGGGAGCATCCTCTGTGACCC CGCCTGCCCTGCGAGAGAGCCAGGACACTCTCC GCAGATCATCTCCTTCTTTCTTGCCTGACGTCGA CTGCGTTGCTCTTCTGCTGTTCTTCTCACGCTCC GTTTCTCTGTTGTTAAACGGGGCAGAAAGAACTC



			CTGTATATATTCAAACAACCATTTATGAGACCAGT ACAAACTACTCAAGAGGAAGATGGCTGTAGCTGC CGATTTCCAGAAGAAGAAGAAGGAGGATGTGAAC TGTGA
<b>Human CD27</b>	27	AAR84239.1	ATGGCACGGCCACATCCCTGGTGGCTGTGCGTTCT GGGGACCCTGGTGGGGCTCTCAGCTACTCCAGCCC CCAAGAGCTGCCAGAGAGGCACTACTGGGCTCA GGGAAAGCTGTGCTGCCAGATGTGTGAGCCAGGA ACATTCCTCGTGAAGGACTGTGACCAGCATAGAA AGGCTGCTCAGTGTGATCCTTGCATAACCGGGGGTC TCCTTCTCTCCTGACCACCACACCCGGCCCCACTG TGAGAGCTGTGCGGCACTGTAACCTCTGGTCTTCTCG TTCGCAACTGCACCATCACTGCCAATGCTGAGTGT GCCTGTGCAATGGCTGGCAGTGCAGGGACAAGG AGTGCACCGAGTGTGATCCTCTTCAAACCCTTCG CTGACCGCTCGGTCGTCTCAGGCCCTGAGCCCACA CCCTCAGCCCACCCACTTACCTTATGTCAGTGAGA TGCTGGAGGCCAGGACAGCTGGGCACATGCAGAC TCTGGCTGACTTCAGGCAGCTGCCTGCCCGGACTC TCTCTACCCACTGGCCACCCCAAAGATCCCTGTGC AGCTCCGATTTTATTCGCATCCTTGTGATCTTCTCT GGAATGTTCTTGTTTTCACCCTGGCCGGGGCCCT GTTCTCCATCAACGAAGGAAATATAGATCAAAC AAAGGAGAAAGTCCTGTGGAGCCTGCAGAGCCTT GTCGTTACAGCTGCCCCAGGGAGGAGGGGCAG CACCATCCCCATCCAGGAGGATTACCGAAAACCG GAGCCTGCCTGCTCCCCCTGA
<b>Human CD8 alpha</b>	28	AAB04637.1	ATGGCCTTACCAGTGACCGCCTTGCTCCTGCCGCT GGCCTTGCTGCTCCACGCCGCCAGGCCGAGCCAGT TCCGGGTGTCGCCGCTGGATCGGACCTGGAACCTG GGCGAGACAGTGGAGCTGAAGTGCCAGGTGCTGC TGCCAACCCGACGTCGGGCTGCTCGTGGCTCTTC CAGCCGCGCGGGCGCCGCGCCAGTCCCACCTTCCT CCTATACCTCTCCCAAACAAGCCCAAGGCGGCC GAGGGGCTGGACACCCAGCGGTTCTCGGGCAAGA GGTTGGGGGACACCTTCGTCCTCACCTGAGCGAC TTCCGCCGAGAGAACGAGGGCTACTATTTCTGCTC GGCCCTGAGCAACTCCATCATGTACTTCAGCCACT TCGTGCCGGTCTTCCTGCCAGCGAAGCCCACCACG ACGCCAGCGCCGCGACCACCAACACCGGCGCCCA CCATCGCGTCGCAGCCCCTGTCCCTGCGCCCAGAG GCGTGCCGGCCAGCGGCGGGGGGCGCAGTGCACA CGAGGGGGCTGGACTTCGCCTGTGATATCTACATC TGGGCGCCCTTGGCCGGGACTTGTGGGGTCCCTTCT CCTGTCACTGGTTATCACCTTTACTGCAACCACA GGAACCGAAGACGTGTTTGC AAATGTCCCCGGCCT GTGGTCAAATCGGGAGACAAGCCCAGCCTTTCGG CGAGATACGTCTAA
<b>Human DAP10</b>	29	AAD47911.1	ATGATCCATCTGGGTACATCCTCTTCCTGCTTTTG CTCCCAGTGGCTGCAGCTCAGACGACTCCAGGAG AGAGATCATCACTCCCTGCCTTTTACCCTGGCACT TCAGGCTCTTGTTCGGATGTGGGTCCCTCTCTCTG CCGCTCCTGGCAGGCCTCGTGGCTGCTGATGCGGT GGCATCGCTGCTCATCGTGGGGGGCGGTGTTCTGT GCGCACGCCACGCCGACGCCCGCCCAAGAAGA TGGCAAAGTCTACATCAACATGCCAGGCAGGGGC



			TGA
--	--	--	-----

**[0054]** Variant nucleic acid libraries as described herein may comprise variants of the intracellular domain. In some instances, the intracellular domain comprises a signaling domain. In some instances, variant nucleic acid libraries generated by methods disclosed herein comprise sequences that are varied when compared to the signaling domain. In some instances, the signaling domain comprises an immunoreceptor tyrosine-based activation motif (ITAM). In some instances, the signaling domain comprises a domain derived from TCR zeta, FcR gamma, FcR beta, CD3 gamma, CD3 delta, CD3 epsilon, CD5, CD22, CD79a, CD79b or CD66d. In some cases, the signaling domain for T-cell activation comprises a domain derived from CD3 $\zeta$ . In some instances, about 100, 250, 500, 1000, or more than 1000 common coding gene sequences of a signaling domain for T cell activation is selected for variation. Variation includes designing of at least about 500, 1000, 2000, 5000, 10000, 20000, 50000, 100000, or more than 100000 variants for each of the common coding gene sequences selected. Exemplary signaling domain sequences are shown in **Table 2**.

**Table 2. Signaling Domain Sequences**

Name	SEQ ID NO	Accession Number	Nucleic Acid Sequence
Human CD3 gamma	30	AAI13831.1	ATGGAACAGGGGAAGGGCCTGGCTGTCCTCATCCTGGCT ATCATTCTTCTTCAAGGTACTTTGGCCCAGTCAATCAAAG GAAACCACTTGGTTAAGGTGTATGACTATCAAGAAGATG GTTTCGGTACTTCTGACTTGTGATGCAGAAGCCAAAAATAT CACATGGTTTAAAGATGGGAAGATGATCGGCTTCCTAAC TGAAGATAAAAAAAAAATGGAATCTGGGAAGTAATGCCAA GGACCCTCGAGGGATGTATCAGTGTAAGGATCACAGAA CAAGTCAAACCACTCCAAGTGTATTACAGAATGTGTCA GAACTGCATTGAACTAAATGCAGCCACCATATCTGGCTTT CTCTTTGCTGAAATCGTCAGCATTTCGTCCTTGCTGTTGG GGTCTACTTCATTGCTGGACAGGATGGAGTTCGCCAGTCG AGAGCTTCAGACAAGCAGACTCTGTTGCCAATGACCAG CTCTACCAGCCCCTCAAGGATCGAGAAGATGACCAGTAC AGCCACCTTCAAGGAAACCAGTTGAGGAGGAATTGA
Human CD3 epsilon	31	EAW67364.1	ATGCAGTCGGGCACTCACTGGAGAGTTCTGGGCCTCTGCC TCTTATCAGTTGGCGTTTGGGGGCAAGATGGTAATGAAG AAATGGGTGGTATTACACAGACACCATATAAAGTCTCCA TCTCTGGAACCACAGTAATATTGACATGCCCTCAGTATCC TGGATCTGAAATACTATGGCAACACAATGATAAAAACAT AGGCGGTGATGAGGATGATAAAAACATAGGCAGTGATGA GGATCACCTGTCACTGAAGGAATTTTCAGAATTGGAGCA AAGTGGTTATTATGTCTGCTACCCAGAGGAAGCAAACC AGAAGATGCGAACTTTTATCTCTACCTGAGGGCAAGAGT GTGTGAGAACTGCATGGAGATGGATGTGATGTCGGTGCC CACAATTGTCATAGTGGACATCTGCATCACTGGGGGCTTG CTGCTGCTGGTTTACTACTGGAGCAAGAATAGAAAGGCC AAGGCCAAGCCTGTGACACGAGGAGCGGGTGCTGGCGGC AGGCAAAGGGGACAAAACAAGGAGAGGCCACCACCTGT



			TCCCAACCCAGACTATGAGCCCATCCGGAAAGGCCAGCG GGACCTGTATTCTGGCCTGAATCAGAGACGCATCTGA
<b>Human CD79 alpha</b>	32	AAB20812.1	ATGCCTGGGGGTCCAGGAGTCCTCCAAGCTCTGCCTGCCA CCATCTTCCTCCTCTTCCTGCTGTCTGCTGTCTACCTGGGC CCTGGGTGCCAGGCCCTGTGGATGCACAAGGTCCCAGCA TCATTGATGGTGAGCCTGGGGGAAGACGCCCACTTCCAA TGCCCGCACAATAGCAGCAACAACGCCAACGTCACCTGG TGGCGCGTCCTCCATGGCAACTACACGTGGCCCCCTGAGT TCTTGGGCCCCGGGCGAGGACCCCAATGGTACGCTGATCA TCCAGAATGTGAACAAGAGCCATGGGGGCATATACGTGT GCCGGGTCCAGGAGGGCAACGAGTCATACCAGCAGTCCT GCGGCACCTACCTCCGCGTGCGCCAGCCGCCCCCAGGC CCTTCCTGGACATGGGGGAGGGCACCAAGAACCGAATCA TCACAGCCGAGGGGATCATCCTCCTGTTCTGCGCGGTGGT GCCTGGGACGCTGCTGCTGTTTCAGGAAACGATGGCAGAA CGAGAAGCTCGGGTTGGATGCCGGGGATGAATATGAAGA TGAAAACCTTTATGAAGGCCTGAACCTGGACGACTGCTC CATGTATGAGGACATCTCCCGGGGCCTCCAGGGCACCTA CCAGGATGTGGGCAGCCTCAACATAGGAGATGTCCAGCT GGAGAAGCCGTGA
<b>Human CD79 beta</b>	33	NM_000626.3	AGGGGACAGGCTGCAGCCGGTGCAGTTACACGTTTTCT CCAAGGAGCCTCGGACGTTGTCACGGGTTTGGGGTCCGG GACAGAGCGGTGACCATGGCCAGGCTGGCGTTGTCTCCT GTGCCAGCCACTGGATGGTGGCGTTGCTGCTGCTGCTCT CAGCTGAGCCAGTACCAGCAGCCAGATCGGAGGACCGGT ACCGGAATCCCAAAGGTAGTGCTTGTTCGCGGATCTGGC AGAGCCCACGTTTCATAGCCAGGAAACGGGGCTTCACGG TGAAAATGCACTGCTACATGAACAGCGCCTCCGGCAATG TGAGCTGGCTCTGGAAGCAGGAGATGGACGAGAATCCCC AGCAGCTGAAGCTGGAAAAGGGCCGCATGGAAGAGTCCC AGAACGAATCTCTCGCCACCCTCACCATCCAAGGCATCC GGTTTGAGGACAATGGCATCTACTTCTGTCAGCAGAAGT GCAACAACACCTCGGAGGTCTACCAGGGCTGCGGCACAG AGCTGCGAGTCATGGGATTCAGCACCTTGGCACAGCTGA AGCAGAGGAACACGCTGAAGGATGGTATCATCATGATCC AGACGCTGCTGATCATCCTCTTCATCATCGTGCCTATCTT CCTGCTGCTGGACAAGGATGACAGCAAGGCTGGCATGGA GGAAGATCACACCTACGAGGGCCTGGACATTGACCAGAC AGCCACCTATGAGGACATAGTGACGCTGCGGACAGGGGA AGTGAAGTGGTCTGTAGGTGAGCACCCAGGCCAGGAGTG AGAGCCAGGTCGCCCCATGACCTGGGTGCAGGCTCCCTG GCCTCAGTGACTGCTTCGGAGCTGCCTGGCTCATGGCCCA ACCCCTTTCCTGGACCCCCAGCTGGCCTCTGAAGCTGGC CCACCAGAGCTGCCATTTGTCTCCAGCCCCTGGTCCCAG CTCTTGCCAAAGGGCCTGGAGTAGAAGGACAACAGGGCA GCAACTTGAGGGGAGTTCTCTGGGGATGGACGGGACCCA GCCTTCTGGGGGTGCTATGAGGTGATCCGTCCCCACACAT GGGATGGGGGAGGCAGAGACTGGTCCAGAGCCCGCAA TGGACTCGGAGCCGAGGGCCTCCCAGCAGAGCTTGGGAA GGGCCATGGACCCAACCTGGGCCCCAGAAGAGCCACAGGA ACATCATTCCTCTCCCGCAACCACTCCCACCCAGGGAGG CCCTGGCCTCCAGTGCTTCCCCCGTGGAATAAACGGTGT GTCCTGAGAAACCACACAAAAAAA
<b>Human CD3 zeta</b>	34	AAY57330.1	ATGAAGTGGAAGGCGCTTTTCACCGCGGCCATCCTGCAG GCACAGTTGCCGATTACAGAGGCACAGAGCTTTGGCCTG CTGGATCCCAAACCTCTGCTACCTGCTGGATGGAATCCTCT



			TCATCTATGGTGTTCATTCTCACTGCCTTGTTTCCTGAGAGT GAAGTTCAGCAGGAGCGCAGACGCCCCGCGTACCAGCA GGGCCAGAACCAGCTCTATAACGAGCTCAATCTAGGACG AAGAGAGGAGTACGATGTTTTGGACAAGAGACGTGGCCG GGACCCTGAGATGGGGGGAAAGCCGCAGAGAAGGAAGA ACCCTCAGGAAGGCCTGTACAATGAACTGCAGAAAGATA AGATGGCGGAGGCCTACAGTGAGATTGGGATGAAAGGCG AGCGCCGGAGGGGCAAGGGGACGATGGCCTTTACCAGG GTCTCAGTACAGCCACCAAGGACACCTACGACGCCCTTC ACATGCAGGCCCTGCCCCCTCGCTAA
--	--	--	---

**[0055]** Methods provided herein for the synthesis of a variant nucleic acid library that is varied in comparison to a reference domain within a CAR may include: de novo synthesis of a variant library of primer oligonucleotides followed by PCR mutagenesis, or de novo synthesis of multiple fragments of the variant version for annealing and assembly including polymerase chain assembly.

**[0056]** A nucleic acid library as described herein may comprise a plurality of nucleic acids that are varied in comparison to a reference domain within a reference CAR, and variation is generated by de novo synthesis of a variant library of primer oligonucleotides followed by PCR mutagenesis. In some instances, oligonucleotides are synthesized on a surface, wherein each oligonucleotide encodes for a predetermined sequence. In some instances, the predetermined sequence is a predetermined variant of a reference nucleic acid sequence that encodes for an antigen recognition domain, a hinge domain, a transmembrane domain, or an intracellular domain of a CAR. In some instances, oligonucleotide primers are de novo synthesized for use in a series of PCR reactions to generate a library of oligonucleotide variants of a reference nucleic acid sequence that encodes for an antigen recognition domain, a hinge domain, a transmembrane domain, or an intracellular domain of a CAR. In some instances, the oligonucleotide primers are used for amplification from a reference nucleic acid sequence to produce a library of variant nucleic acids encoding for an antigen recognition domain, a hinge domain, a transmembrane domain, or an intracellular domain of a CAR.

**[0057]** In some instances, a variant nucleic acid library varied in comparison to a reference domain within a CAR is generated by de novo synthesis of multiple fragments of the variant version of the common sequence for annealing and assembly including polymerase chain assembly. In some instances, a surface is used for de novo synthesis of multiple fragments of a nucleic acid sequence, wherein at least one of the fragments is synthesized in multiple versions. The nucleic acid sequence may be a variant of a reference nucleic acid sequence that encodes for an antigen recognition domain, a hinge domain, a transmembrane domain, or an intracellular domain of a CAR. In some instances, the fragments are subject to hybridization to generate a CAR variant



library. In some instances, the synthesized fragments are amplified and subject to ligation or hybridization to generate a CAR variant library.

**[0058]** In some instances, a CAR variant library comprises nucleic acid sequences that encode for an antigen recognition domain, a hinge domain, a transmembrane domain, or an intracellular domain including a stimulatory domain or a signaling domain of a CAR. In some instances, a CAR variant library comprises nucleic acid sequences that encodes for a single domain of a CAR. In some instances, a CAR variant library comprises nucleic acid sequences that encodes for multiple domains of a CAR. For example, a CAR variant library comprises nucleic acid sequences that encode for all domains of a CAR.

**[0059]** Libraries comprising nucleic acids encoding for variant CARs as described herein comprise various lengths of amino acids when translated. In some instances, the length of each of the amino acid fragments or average length of the amino acid synthesized may be at least or about 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, or more than 150 amino acids. In some instances, the length of the amino acid is in a range of about 15 to 150, 20 to 145, 25 to 140, 30 to 135, 35 to 130, 40 to 125, 45 to 120, 50 to 115, 55 to 110, 60 to 110, 65 to 105, 70 to 100, or 75 to 95 amino acids. In some instances, the length of the amino acid is in a range of about 22 to about 75 amino acids.

**[0060]** A number of variant sequences for the at least one region of the CAR chain for variation are de novo synthesized using methods as described herein. In some instances, a number of variant sequences is de novo synthesized for an antigen recognition domain, a hinge domain, a transmembrane domain, or an intracellular domain of a CAR. In some instances, a number of variant sequences is de novo synthesized for a stimulatory domain or a signaling domain of a CAR. The number of variant sequences may be at least or about 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, or more than 500 sequences. In some instances, the number of variant sequences is at least or about 500, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10,000, 20,000, 30,000, 40,000, 50,000, 60,000, 70,000, 80,000, 90,000, 100,000, 200,000, 300,000, 400,000, 500,000, 600,000, 700,000, 800,000, 900,000, 1 million, or more than 1 million sequences. In some instances, the number of variant sequences is in a range of about 10 to 500, 25 to 475, 50 to 450, 75 to 425, 100 to 400, 125 to 375, 150 to 350, 175 to 325, 200 to 300, 225 to 375, 250 to 350, or 275 to 325 sequences.

**[0061]** Provided herein are variant nucleic acids encoding for variant CARs, wherein the variant CARs are antigen specific. In some instances, the antigen is involved in or associated with a disease, disorder, or condition. For example, the antigen is associated with a proliferative disease, a



tumorous disease, an inflammatory disease, an immunological disorder, an autoimmune disease, an infectious disease, a viral disease, an allergic reaction, a parasitic reaction, a graft-versus-host disease or a host-versus-graft disease. In some instances, the antigen is an antigen expressed on a tumor cell. In some instances, an antigen is associated with a pathogen such as a virus or bacterium.

**[0062]** In some instances, the variant CARs recognize antigens that are tissue-restricted. For example, the variant CARs are restricted non-vital cell lineages or tissues. In some instances, the variant CARs recognize antigens from mutated gene products.

**[0063]** Provided herein are variant CAR libraries, wherein the variant CARs encode for variants in an antigen binding interface. In some instances, residues for variation are preselected or predicted residues that contact an antigen. In some instances, the antigen binding interface is the antigen recognition domain. In some instances, residues for variation are preselected or predicted residues located in the binding pocket.

**[0064]** Variant CAR libraries as described herein comprise one or more mutations in a library. In some instances, the CAR variant libraries are single variant libraries comprising variants at a single site across the library. In some instances, the CAR variant libraries are multiple variant libraries comprising variants at a number of sites. In some instances, the number of sites is 2, 3, 4, 5, 6, 7, 8, 9, 10, or more than 10 sites.

**[0065]** Provided here are libraries where one or more preselected codons in a CAR gene or gene fragment encode for variant amino acid residues to generate variation in resulting variant CAR protein libraries. In some instances, up to 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 amino acid residues are varied. In some instances, up to 30 amino acid residues are varied. In some instances, up to 5 amino acid residues are varied. In some instances, up to 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, or more than 1000 amino acid residues are varied. In some instances, at least or about 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 amino acid residues are varied. In some instances, at least or about 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, or more than 1000 amino acid residues are varied. In some instances, all amino acid residues in a preselected region are varied. In some instances, variant CAR libraries are highly diverse. In some instances, the libraries comprise at least or about  $10^6$ ,  $10^8$ ,  $10^9$ ,  $10^{10}$ ,  $10^{11}$ ,  $10^{12}$ , or  $10^{13}$  variants. In some instances, the libraries comprise at least or about  $10^9$  variants.

**[0066]** Subsequent to synthesis of variant nucleic acid libraries encoding for a reference domain within a CAR, variant nucleic acid libraries may be inserted into a vector sequence. In some instances, the vector sequence is expressed (e.g., by electroporation, transfection, or transduction) in cells and functional consequences are determined. CAR domains that may be varied are an antigen recognition domain, a hinge domain, a transmembrane domain, or an intracellular domain.

In some instances, the intracellular domain comprises a signaling domain or a costimulatory domain. In some instances, the variant nucleic acid libraries comprise variant sequences that encode for all domains of a CAR. Functional consequences that may be measured include binding affinity and binding strength. In some instances, variant nucleic acid libraries generated by methods disclosed herein result in increased binding strength of an antigen recognition domain to a tumor antigen associated with a particular cancer of interest. The cancer may be a solid cancer or a hematologic cancer. In some instances, the cancer is bladder cancer, lung cancer, brain cancer, melanoma, breast cancer, Non-Hodgkin lymphoma, cervical cancer, ovarian cancer, colorectal cancer, pancreatic cancer, esophageal cancer, prostate cancer, kidney cancer, skin cancer, leukemia, thyroid cancer, liver cancer, or uterine cancer. Exemplary antigens for use in binding assays include, without limitation, those provided in **Table 3**.

**Table 3. Antigens**

Antigen peptide name	Protein Sequence	SEQ ID NO	Disease Target
CD19	KVSAVTLAY	35	B-cell malignancy
CD20	RPKSNIVLL	36	B-cell malignancy
CEA peptide	IMIGVLVGV	37	Metastatic colorectal cancer
HER2	IISAVVGIL	38	Breast cancer, lung cancer, prostate cancer, glioma
NY-ESO-1	SLLMWITQV	39	Many cancers including melanoma and sarcoma
PSMA	KYADKIYSI	40	Prostate cancer

**[0067]** Variant nucleic acid libraries generated using methods described herein may be screened to select a modified CAR domain sequence providing for a protein complex with improved affinity (measure of the strength of interaction between an epitope and an antibody's antigen binding site) for a tumor antigen. Additional functional considerations, such as variant gene expression, avidity, stability, and target cell (i.e. cancer cell) specificity may also be assessed. In some instances, the increased specificity of a CAR provides for reduced cross-reactivity to non-cancer associated antigens compared to a reference non-variant CAR. In some instances, the



variant CAR libraries are screened for localization within a cell. In some instances, the variant CAR libraries are screened to identify properly localized variant CARs.

**[0068]** Increased specificity of a variant CAR to a tumor antigen may result in decreased toxicity. Toxicity in some instances that results include cytokine release syndrome (CRS), macrophage activation syndrome (MAS), on-target off-tumor toxicity, autoimmunity, host-graft toxicity, neurotoxicity, and tumor lysis syndrome. Variant nucleic acid libraries encoding for an antigen recognition domain may be inserted into a vector sequence, expressed in cells, and screened for functional consequences such as decreased toxicity. For example, variant nucleic acid libraries are screened for increased cytokine (e.g., interleukin-6 (IL-6), interferon- $\gamma$ , tumor necrosis factor, IL-2, IL-2-receptor- $\alpha$ , IL-8, and IL-10) release from cells or increased cytokine expression as a measure of CRS.

**[0069]** In some instances, variant nucleic acid libraries generated using methods described herein are screened to select a modified CAR domain sequence providing for improved T cell function or improved T cell activation. In some instances, variant CARs are used to screen T cell pathways affecting cell growth, survival, cellular differentiation, cell death, or intracellular signaling modulation. For example, the serine/threonine kinase Akt has been shown to improve T-cell function and survival resulting in resistance to tumor inhibitory mechanisms. In some cases, variant CARs are used to screen T cell pathways that improve resistance to tumor inhibitory mechanisms. In some instances, variant CARs generated using methods described herein result in enhanced immune response towards a target cell such as a cancer cell. For example, a variant CAR may increase cytotoxic activity of a cytotoxic cell (e.g., a Natural Killer cell or cytotoxic T lymphocyte) towards a target cell such as a cancer cell. In some instances, variant CARs generated using methods described herein result in enhanced immune response towards a target cell such as a cancer cell, thus resulting in increased death of a cancer cell.

**[0070]** In some instances, variant CAR libraries that are expressed in cells are used to identify variant CARs with improved variant gene expression, avidity, stability, affinity, or specificity. For example, a first variant CAR library is generated that comprises improved specificity to a tumor antigen. In some instances, following identification of variant CARs with improved gene expression, avidity, stability, affinity, or specificity, those variant CARs are further varied to produce a second library of variant CARs with a second improvement. For example, variant CARs with improved specificity are further varied to identify variants further comprising improved stability. In some instances, the second library comprises variants in a same region as the first library. In some instances, the second library comprises variants in a different region of the first library. For example, the first library may comprise variants in an antigen recognition domain of



the CAR and the second library may comprise variants in a hinge domain, a transmembrane domain, or an intracellular domain of the CAR. In some instances, a number of variant CAR libraries are generated. In some instances, the number of variant CAR libraries is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or more than 12 variant libraries. In some instances, the variant libraries comprise at least or about  $10^1$ ,  $10^2$ ,  $10^3$ ,  $10^4$ ,  $10^5$ , or  $10^6$  variants. In some instances, each of the variant libraries independently has improvements in gene expression, avidity, stability, affinity, or specificity.

**[0071]** Variant CAR libraries may be engineered into cells and introduced into a subject. In some instances, the cells are autologous, meaning derived from a subject's own cells. Alternately, cells expressing variant CARs are allogeneic, meaning derived from another subject with a similar tissue type. In some instances, cells are tailored to the subject. In some instances, cells are compatible with local tissue.

**[0072]** In some instances, variant nucleic acid libraries as described herein comprise about 50-100000, 100-75000, 250-50000, 500-25000, 1000-15000, 2000-10000, or 4000-8000 sequences. In some instances, variant nucleic acid libraries comprise 500 sequences. In some cases, variant nucleic acid libraries comprise 5000 sequences. In some instances, variant nucleic acid libraries comprise 15000 sequences. In some cases, variant nucleic acid libraries comprise at least 50, 100, 150, 500, 1000, 2000, 5000, 10000, 20000, 50000, 100000, 200000, 400000, 800000, 1000000, or more than 1000000 sequences. Variant nucleic acid libraries may comprise at most 50, 100, 500, 1000, 2000, 5000, 10000, 20000, 50000, 100000, 200000, 400000, 800000, or 1000000 sequences. The total variant library may result in  $10^6$ ,  $10^8$ ,  $10^{10}$ ,  $10^{11}$ ,  $10^{12}$ ,  $10^{13}$  or more different nucleic acids.

**[0073]** In some cases, the length of each of the oligonucleotide fragments or average length of the oligonucleotides synthesized may be at least or about at least 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 150, 200, 300, 400, 500, 2000 nucleotides, or more. The length of each of the oligonucleotide fragments or average length of the oligonucleotides synthesized may be at most or about at most 2000, 500, 400, 300, 200, 150, 100, 50, 45, 35, 30, 25, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10 nucleotides, or less. The length of each of the oligonucleotide fragments or average length of the oligonucleotides synthesized may fall from 10-2000, 10-500, 9-400, 11-300, 12-200, 13-150, 14-100, 15-50, 16-45, 17-40, 18-35, 19-25.

**[0074] Variant Library Synthesis**

**[0075]** Methods described herein provide for synthesis of a library of oligonucleotides each encoding for a predetermined variant of at least one predetermined reference nucleic acid sequence. In some instances, the oligonucleotides synthesized herein are isolated, purified, or isolated and



purified. In some cases, the predetermined reference sequence is a nucleic acid sequence encoding for a protein, and the variant library comprises sequences encoding for variation of at least a single codon such that a plurality of different variants of a single residue in the subsequent protein encoded by the synthesized nucleic acid are generated by standard translation processes. The synthesized specific alterations in the nucleic acid sequence may be introduced by incorporating nucleotide changes into overlapping or blunt ended oligonucleotide primers. Alternatively, a population of oligonucleotides may collectively encode for a long nucleic acid (e.g., a gene) and variants thereof. In this arrangement, the population of oligonucleotides may be hybridized and subject to standard molecular biology techniques to form the long nucleic acid (e.g., a gene) and variants thereof. When the long nucleic acid (e.g., a gene) and variants thereof are expressed in cells, a variant protein library is generated. Similarly, provided here are methods for synthesis of variant libraries encoding for RNA sequences (e.g., miRNA, shRNA, and mRNA) or DNA sequences (e.g., enhancer, promoter, UTR, and terminator regions). In some instances, the sequences are exon sequences or coding sequences. In some instances, the sequences do not comprise intron sequences. Also provided here are downstream applications for variants selected out of the libraries synthesized using methods described here. Downstream applications include identification of variant nucleic acid or protein sequences with enhanced biologically relevant functions, e.g., biochemical affinity, enzymatic activity, changes in cellular activity, and for the treatment or prevention of a disease state.

**[0076]**     *Synthesis followed by PCR Mutagenesis*

**[0077]**     A first process for synthesis of a variant library of oligonucleotides is for PCR mutagenesis (saturating or non-saturating) methods. In this workflow, a plurality of oligonucleotides is synthesized, wherein each oligonucleotide encodes for a predetermined sequence which is a predetermined variant of a reference nucleic acid sequence. Referring to the figures, an exemplary workflow is depicted in **FIGS. 1A-1D**, wherein oligonucleotides are generated on a surface. **FIG. 1A** depicts an expansion view of a single cluster of a surface with 121 loci. Each oligonucleotide depicted in **FIG. 1B** is a primer that may be used for amplification from a reference nucleic acid sequence to produce a library of variant long nucleic acids, **FIG. 1C**. The library of variant long nucleic acids is then, optionally, subject to transcription and or translation to generate a variant RNA or protein library, **FIG. 1D**. In this exemplary illustration, a device having a substantially planar surface used for de novo synthesis of oligonucleotides is depicted, **FIG. 1A**. In some instances, the device comprises a cluster of loci, wherein each locus is a site for oligonucleotide extension. In some instances, a single cluster comprises all the



oligonucleotide variants needed to generate a desired variant sequence library. In an alternative arrangement, a plate comprises a field of loci which are not segregated into clusters.

**[0078]** In some instances, oligonucleotides synthesized within a cluster (e.g., as seen in **FIG. 1A**) are amplified by PCR. Such an arrangement may provide for improved oligonucleotide representation compared to amplification of non-identical oligonucleotides across an entire plate without a clustered arrangement. In some instances, amplification of oligonucleotides synthesized on surfaces of loci within a cluster overcomes negative effects on representation due to repeated synthesis of large oligonucleotide populations having oligonucleotides with heavy GC content. In some instances, a cluster described herein, comprises about 50-1000, 75-900, 100-800, 125-700, 150-600, 200-500, 50-500 or 300-400 discrete loci. In some instances, a loci is a spot, well, microwell, channel, or post. In some instances, each cluster has at least 1X, 2X, 3X, 4X, 5X, 6X, 7X, 8X, 9X, 10X, or more redundancy of separate features supporting extension of oligonucleotides having an identical sequence. In some instances, 1X redundancy means not having oligonucleotides with the same sequence.

**[0079]** A de novo synthesized oligonucleotide library described herein may comprise a plurality of oligonucleotides, each with at least one variant sequence at a first position, position “x”, and each variant oligonucleotide is used as a primer in a first round of PCR to generate a first extension product. In this example, position “x” in a first oligonucleotide **220** encodes for a variant codon sequence, i.e., one of 19 possible variants from a reference sequence. *See FIG. 2A.* A second oligonucleotide **225** comprising a sequence overlapping that of the first oligonucleotide is also used as a primer in a separate round of PCR to generate a second extension product. In addition, outer primers **215, 230** may be used for amplification of a fragment from a long nucleic acid sequence. The resultant amplification products are fragments of the long nucleic acid sequence **235, 240**. *See FIG. 2B.* The fragments of the long nucleic acid sequence **235, 240** are then hybridized, and subject to an extension reaction to form a variant of the long nucleic acid **245**. *See FIG. 2C.* The overlapping ends of the first and second extension products may serve as a primer of a second round of PCR, thereby generating a third extension product (**FIG. 2D**) that contains the variant. To increase the yield, the variant of the long nucleic acid is amplified in a reaction including a DNA polymerase, amplification reagents, and the outer primers **215, 230**. In some instances, the second oligonucleotide comprises a sequence adjacent to, but not including, the variant site. In an alternative arrangement, a first oligonucleotide is generated that has a region that overlaps with a second oligonucleotide. In this scenario, the first oligonucleotide is synthesized with variation at a single codon for up to 19 variants. The second oligonucleotide does not comprise a variant sequence. Optionally, a first population comprises the first oligonucleotide



variants and additional oligonucleotides encoding for variants at a different codon site. Alternatively, the first oligonucleotide and the second oligonucleotide may be designed for blunt end ligation.

**[0080]** An alternative mutagenesis PCR method is depicted in **FIGS. 3A-3F**. In such a process, a template nucleic acid molecule **300** comprising a first and second strand **305, 310** is amplified in a PCR reaction containing a first primer **315** and a second primer **320** (**FIG. 3A**). The amplification reaction includes uracil as a nucleotide reagent. A uracil-labeled extension product **325** (**FIG. 3B**) is generated, optionally purified, and serves as a template for a subsequent PCR reaction using a first oligonucleotide **335** and a plurality of second oligonucleotides **330** to generate first extension products **340** and **345** (**FIGS. 3C-3D**). In this process, a plurality of second oligonucleotide **330** comprises oligonucleotides encoding for variant sequences (denoted as *X*, *Y*, and *Z*, in **FIG. 3C**). The uracil-labeled template nucleic acid is digested by a uracil-specific excision reagent, e.g., USER digest available commercially from New England Biolabs. Variant **335** and different codons **330** with variants *X*, *Y*, and *Z* are added and a limited PCR step is performed to generate **FIG. 3D**. After the uracil-containing template is digested, the overlapping ends of the extension products serve to prime a PCR reaction with the first extension products **340** and **345** acting as primers in combination with a first outer primer **350** and a second outer primer **355**, thereby generating a library of nucleic acid molecules **360** containing a plurality of variants *X*, *Y*, and *Z* at the variant site **FIG. 3F**.

**[0081]** *De novo synthesis of a population with variant and non-variant portions of a long nucleic acid*

**[0082]** In a second process for synthesis of a variant library, a surface is used for de novo synthesis of multiple fragments of a long nucleic acid, wherein at least one of the fragments is synthesized in multiple versions, each version being of a different variant sequence. In this arrangement, all of the fragments needed to assemble a library of variant long range nucleic acids are de novo synthesized. The synthesized fragments may have overlapping sequence such that, following synthesis, the fragment library is subject to hybridization. Following hybridization, an extension reaction may be performed to fill in any complementary gaps.

**[0083]** Alternatively, the synthesized fragments may be amplified with primers and then subject to either blunt end ligation or overlapping hybridization. In some instances, the device comprises a cluster of loci, wherein each locus is a site for oligonucleotide extension. In some instances, a single cluster comprises all the oligonucleotide variants and other fragment sequences of a predetermined long nucleic acid to generate a desired variant nucleic acid sequence library. The

cluster may comprise about 50 to 500 loci. In some arrangements, a cluster comprises greater than 500 loci.

**[0084]** Each individual oligonucleotide in the first oligonucleotide population may be generated on a separate, individually addressable locus of a cluster. One oligonucleotide variant may be represented by a plurality of individually addressable loci. Each variant in the first oligonucleotide population may be represented 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more times. In some instances, each variant in the first oligonucleotide population is represented at 3 or less loci. In some instances, each variant in the first oligonucleotide population is represented at two loci. In some instances, each variant in the first oligonucleotide population is represented at only a single locus.

**[0085]** Methods are provided herein to generate nucleic acid libraries with reduced redundancy. In some instances, variant oligonucleotides may be generated without the need to synthesize the variant oligonucleotide more than 1 time to obtain the desired variant oligonucleotide. In some instances, the present disclosure provides methods to generate variant oligonucleotides without the need to synthesize the variant oligonucleotide more than 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more times to generate the desired variant oligonucleotide.

**[0086]** Variant oligonucleotides may be generated without the need to synthesize the variant oligonucleotide at more than 1 discrete site to obtain the desired variant oligonucleotide. The present disclosure provides methods to generate variant oligonucleotides without the need to synthesize the variant oligonucleotide at more than 1 site, 2 sites, 3 sites, 4 sites, 5 sites, 6 sites, 7 sites, 8 sites, 9 sites, or 10 sites. In some instances, an oligonucleotide is synthesized in at most 6, 5, 4, 3, 2, or 1 discrete sites. The same oligonucleotide may be synthesized in 1, 2, or 3 discrete loci on a surface.

**[0087]** In some instances, the amount of loci representing a single variant oligonucleotide is a function of the amount of nucleic acid material required for downstream processing, e.g., an amplification reaction or cellular assay. In some instances, the amount of loci representing a single variant oligonucleotide is a function of the available loci in a single cluster.

**[0088]** Provided herein are methods for generation of a library of oligonucleotides comprising variant oligonucleotides differing at a plurality of sites in a reference nucleic acid. In such cases, each variant library is generated on an individually addressable locus within a cluster of loci. It will be understood that the number of variant sites represented by the oligonucleotide library will be determined by the number of individually addressable loci in the cluster and the number of desired variants at each site. In some instances, each cluster comprises about 50 to 500 loci. In some instances, each cluster comprises 100 to 150 loci.



**[0089]** In an exemplary arrangement, 19 variants are represented at a variant site corresponding to codons encoding for each of the 19 possible variant amino acids. In another exemplary case, 61 variants are represented at a variant site corresponding to triplets encoding for each of the 19 possible variant amino acids. In a non-limiting example, a cluster comprises 121 individually addressable loci. In this example, an oligonucleotide population comprises 6 replicates each of a single-site variant (6 replicates x 1 variant site x 19 variants = 114 loci), 3 replicates each of a double-site variant (3 replicates x 2 variant sites x 19 variants = 114 loci), or 2 replicates each of a triple-site variant (2 replicates x 3 variant sites x 19 variants = 114 loci). In some instances, an oligonucleotide population comprises variants at four, five, six or more than six variant sites.

**[0090]** Provided herein are methods and compositions for production of synthetic (i.e. de novo synthesized or chemically synthesized) oligonucleotides. Libraries of synthesized oligonucleotides described herein may comprise a plurality of oligonucleotides collectively encoding for one or more genes or gene fragments. In some instances, the oligonucleotide library comprises coding or non-coding sequences. In some instances, the oligonucleotide library encodes for a plurality of cDNA sequences. In some instances, the oligonucleotide library comprises one or more oligonucleotides, each of the one or more oligonucleotides encoding sequences for multiple exons. Each oligonucleotide within a library described herein may encode a different sequence, i.e., non-identical sequence. In some instances, each oligonucleotide within a library described herein comprises at least one portion that is complementary to a sequence of another oligonucleotide within the library. Oligonucleotide sequences described herein may, unless stated otherwise, comprise DNA or RNA.

**[0091]** Provided herein are methods and compositions for production of synthetic (i.e. de novo synthesized) genes. Libraries comprising synthetic genes may be constructed by a variety of methods described in further detail elsewhere herein, such as PCA, non-PCA gene assembly methods or hierarchical gene assembly, combining (“stitching”) two or more double-stranded oligonucleotides to produce larger DNA units (i.e., a chassis). Libraries of large constructs may involve oligonucleotides that are at least 1, 1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 400, 500 kb long or longer. The large constructs may be bound by an independently selected upper limit of about 5000, 10000, 20000 or 50000 base pairs. The synthesis of any number of polypeptide-segment encoding nucleotide sequences may include sequences encoding non-ribosomal peptides (NRPs), sequences encoding non-ribosomal peptide-synthetase (NRPS) modules and synthetic variants, polypeptide segments of other modular proteins, such as antibodies, polypeptide segments from other protein families, including non-coding DNA or RNA, such as regulatory sequences e.g. promoters, transcription factors, enhancers,



siRNA, shRNA, RNAi, miRNA, small nucleolar RNA derived from microRNA, or any functional or structural DNA or RNA unit of interest. The following are non-limiting examples of oligonucleotides: coding or non-coding regions of a gene or gene fragment, intergenic DNA, loci (locus) defined from linkage analysis, exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, short interfering RNA (siRNA), short-hairpin RNA (shRNA), micro-RNA (miRNA), small nucleolar RNA, ribozymes, cDNA, which is a DNA representation of mRNA, usually obtained by reverse transcription of messenger RNA (mRNA) or by amplification; DNA molecules produced synthetically or by amplification, genomic DNA, recombinant oligonucleotides, branched oligonucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers. In the context of cDNA, the term gene or gene fragment refers to a DNA nucleic acid sequence comprising at least one region encoding for exon sequences without an intervening intron sequence.

**[0092]** In various embodiments, methods and compositions described herein relate to a library of genes. The gene library may comprise a plurality of subsegments. In one or more subsegments, the genes of the library may be covalently linked together. In one or more subsegments, the genes of the library may encode for components of a first metabolic pathway with one or more metabolic end products. In one or more subsegments, genes of the library may be selected based on the manufacturing process of one or more targeted metabolic end products. The one or more metabolic end products may comprise a biofuel. In one or more subsegments, the genes of the library may encode for components of a second metabolic pathway with one or more metabolic end products. The one or more end products of the first and second metabolic pathways may comprise one or more shared end products. In some cases, the first metabolic pathway comprises an end product that is manipulated in the second metabolic pathway.

**[0093]** *Codon variation*

**[0094]** Variant oligonucleotide libraries described herein may comprise a plurality of oligonucleotides, wherein each oligonucleotide encodes for a variant codon sequence compared to a reference nucleic acid sequence. In some instances, each oligonucleotide of a first oligonucleotide population contains a variant at a single variant site. In some instances, the first oligonucleotide population contains a plurality of variants at a single variant site such that the first oligonucleotide population contains more than one variant at the same variant site. The first oligonucleotide population may comprise oligonucleotides collectively encoding multiple codon variants at the same variant site. The first oligonucleotide population may comprise oligonucleotides collectively encoding up to 19 or more codons at the same position. The first oligonucleotide population may comprise oligonucleotides collectively encoding up to 60 variant



triplets at the same position, or the first oligonucleotide population may comprise oligonucleotides collectively encoding up to 61 different triplets of codons at the same position. Each variant may encode for a codon that results in a different amino acid during translation. **Table 4** provides a listing of each codon possible (and the representative amino acid) for a variant site.

**Table 4. List of codons and amino acids**

Amino Acids	One letter code	Three letter code	Codons					
Alanine	A	Ala	GCA	GCC	GCG	GCT		
Cysteine	C	Cys	TGC	TGT				
Aspartic acid	D	Asp	GAC	GAT				
Glutamic acid	E	Glu	GAA	GAG				
Phenylalanine	F	Phe	TTC	TTT				
Glycine	G	Gly	GGA	GGC	GGG	GGT		
Histidine	H	His	CAC	CAT				
Isoleucine	I	Iso	ATA	ATC	ATT			
Lysine	K	Lys	AAA	AAG				
Leucine	L	Leu	TTA	TTG	CTA	CTC	CTG	CTT
Methionine	M	Met	ATG					
Asparagine	N	Asn	AAC	AAT				
Proline	P	Pro	CCA	CCC	CCG	CCT		
Glutamine	Q	Gln	CAA	CAG				
Arginine	R	Arg	AGA	AGG	CGA	CGC	CGG	CGT
Serine	S	Ser	AGC	AGT	TCA	TCC	TCG	TCT
Threonine	T	Thr	ACA	ACC	ACG	ACT		

Valine	V	Val	GTA	GTC	GTG	GTT
Tryptophan	W	Trp	TGG			
Tyrosine	Y	Tyr	TAC	TAT		

**[0095]** Provided herein are nucleic acid libraries that may comprise a plurality of nucleic acids, wherein each nucleic acid encodes for a variant codon sequence compared to a reference nucleic acid sequence. In some instances, each nucleic acid of a first nucleic acid population contains a variant at a single variant site. In some instances, the first nucleic acid population contains a plurality of variants at a single variant site such that the first nucleic acid population contains more than one variant at the same variant site. The first nucleic acid population may comprise nucleic acids collectively encoding multiple codon variants at the same variant site. The first oligonucleotide population may comprise nucleic acids collectively encoding up to 19 or more codons at the same position. The first nucleic acid population may comprise nucleic acids collectively encoding up to 60 variant triplets at the same position, or the first nucleic acid population may comprise oligonucleotides collectively encoding up to 61 different triplets of codons at the same position. Each variant may encode for a codon that results in a different amino acid during translation.

**[0096]** An oligonucleotide population may comprise varied oligonucleotides collectively encoding up to 20 codon variations at multiple positions. In such cases, each oligonucleotide in the population comprises variation for codons at more than one position in the same oligonucleotide. In some instances, each oligonucleotide in the population comprises variation for codons at 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more codons in a single oligonucleotide. In some instances, each variant long nucleic acid comprises variation for codons at 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 or more codons in a single long nucleic acid. In some instances, the variant oligonucleotide population comprises variation for codons at 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 or more codons in a single oligonucleotide. In some instances, the variant oligonucleotide population comprises variation for codons in at least about 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 or more codons in a single long nucleic acid.

**[0097]** Provided herein are processes where a second oligonucleotide population is generated on a second cluster containing a plurality of individually addressable loci. The second oligonucleotide population may comprise a plurality of second oligonucleotides that are constant for each codon position (i.e., encode the same amino acid at each position). The second



oligonucleotide may overlap with at least a portion of the first oligonucleotides. In some instances, the second oligonucleotides do not contain the variant site represented on the first oligonucleotides. Alternatively, the second oligonucleotide population may comprise a plurality of second oligonucleotides that contain at least one variant for one or more codon positions.

**[0098]** Provided herein are methods for synthesizing a library of oligonucleotides where a single population of oligonucleotides is generated comprising variants at multiple codon positions. A first oligonucleotide population may be generated on a first cluster containing a plurality of individually addressable loci. In such cases, the first oligonucleotide population comprises variants at different codon positions. In some instances, the different sites are consecutive (i.e., encoding consecutive amino acids). For example, the first oligonucleotide population comprises variants in two consecutive codon positions, encoding up to 19 variants at a position. In some instances, the first oligonucleotide population comprises variants in two consecutive codon positions, encoding from about 1 to about 19 variants at a position. In some instances, about 38 oligonucleotides are synthesized. In some instances, the oligonucleotides comprising variants in two consecutive codon positions comprise a range of about 36 to about 66 bases in length. A first oligonucleotide population may comprise varied oligonucleotides collectively encoding up to 19 codon variants at the same, or additional variant site. A first oligonucleotide population may include a plurality of first oligonucleotides that contains up to 19 variants at position *x*, up to 19 variants at position *y*, and up to 19 variants at position *z*. In such an arrangement, each variant encodes a different amino acid such that up to 19 amino acid variants are encoded at each of the different variant sites. In an additional instance, a second oligonucleotide population is generated on a second cluster containing a plurality of individually addressable loci. The second oligonucleotide population may comprise a plurality of second oligonucleotides that are constant for each codon position (i.e., encode the same amino acid at each position). The second oligonucleotides may overlap with at least a portion of the first oligonucleotides. The second oligonucleotides may not contain the variant site represented on the first oligonucleotides.

**[0099]** Variant nucleic acid libraries generated by processes described herein provide for the generation of variant protein libraries. In a first exemplary arrangement, a template oligonucleotide encodes for a sequence that, when transcribed and translated, results in a reference amino acid sequence (**FIG. 4A**) having a number of codon positions, indicated by a single circle. Oligonucleotide variants of the template may be generated using methods described herein. In some instances, a single variant is present in the oligonucleotide, resulting in a single amino acid sequence (**FIG. 4B**). In some instances, more than one variant is present in the oligonucleotide, wherein the variants are separated by one or more codons, resulting in a protein with spacing



between variant residues (**FIG. 4C**). In some instances, more than one variant is present in the oligonucleotide, wherein the variants are sequential and adjacent or consecutive to one another, resulting in spaced variant stretches of residues (**FIG. 4D**). In some instances, two stretches of variants are present in the oligonucleotide, wherein each stretch of variants comprises sequential and adjacent or consecutive variants (**FIG. 4E**).

**[00100]** Provided herein are methods to generate a library of oligonucleotide variants, wherein each variant comprises a single position codon variant. In one instance, a template oligonucleotide has a number of codon positions wherein exemplary amino acid residues are indicated by circles with their respective one letter code protein codon, **FIG. 5A**. **FIG. 5B** depicts a library of amino acid variants encoded by a library of variant nucleic acids, wherein each variant comprises a single position variant, indicated by an "X", located at a different single site. A first position variant has any codon to replace alanine, a second variant with any codon encoded by the library of variant nucleic acids to replace tryptophan, a third variant with any codon to replace isoleucine, a fourth variant with any codon to replace lysine, a fifth variant with any codon to replace arginine, a sixth variant with any codon to replace glutamic acid, and a seventh variant with any codon to replace glutamine. When all or less than all codon variants are encoded by the variant nucleic acid library, a resulting corresponding population of amino acid sequence variants is generated following protein expression (i.e., standard cellular events of DNA transcription followed by translation and processing events).

**[00101]** In some arrangements, a library is generated with multiple sites of single position variants. As depicted in **FIG. 6A**, a wild-type template is provided. **FIG. 6B** depicts the resultant amino acid sequence with two sites of single position codon variants, wherein each codon variant encoding for a different amino acid is indicated by differently patterned circles.

**[00102]** Provided herein are methods to generate a library having a stretch of multiple site, single position variants. Each stretch of oligonucleotide or nucleic acid may have 1, 2, 3, 4, 5, or more variants. Each stretch of oligonucleotide or nucleic acid may have at least 1 variant. Each stretch of oligonucleotide or nucleic acid may have at least 2 variants. Each stretch of oligonucleotide or nucleic acid may have at least 3 variants. For example, a stretch of 5 oligonucleotides or nucleic acids may have 1 variant. A stretch of 5 oligonucleotides or nucleic acids may have 2 variants. A stretch of 5 oligonucleotides or nucleic acids may have 3 variants. A stretch of 5 oligonucleotides or nucleic acids may have 4 variants. For example, a stretch of 4 oligonucleotides or nucleic acids may have 1 variant. A stretch of 4 oligonucleotides or nucleic acids may have 2 variants. A stretch of 4 oligonucleotides or nucleic acids may have 3 variants. A stretch of 4 oligonucleotides or nucleic acids may have 4 variants.



**[00103]** In some instances, single position variants may all encode for the same amino acid, e.g. a histidine. As depicted in **FIG. 7A**, a reference amino acid sequence is provided. In this arrangement, a stretch of an oligonucleotide encodes for multiple sites of single position variants and, when expressed, results in an amino acid sequence having all single position variants encoding for a histidine, **FIG. 7B**. In some embodiments, a variant library synthesized by methods described herein does not encode for more than 4 histidine residues in a resultant amino acid sequence.

**[00104]** In some instances, a variant library of nucleic acids generated by methods described herein provides for expression of amino acid sequences having separate stretches of variation. A template amino acid sequence is depicted in **FIG. 8A**. A stretch of oligonucleotides may have only 1 variant codon in two stretches and, when expressed, result in an amino acid sequence depicted in **FIG. 8B**. Variants are depicted in **FIG. 8B** by the differently patterned circles to indicate variation in amino acids at different positions in a single stretch.

**[00105]** Provided herein are methods and devices to synthesize oligonucleotide or nucleic acid libraries with 1, 2, 3, or more codon variants, wherein the variant for each site is selectively controlled. The ratio of two amino acids for a single site variant may be about 1:100, 1:50, 1:10, 1:5, 1:3, 1:2, 1:1. The ratio of three amino acids for a single site variant may be about 1:1:100, 1:1:50, 1:1:20, 1:1:10, 1:1:5, 1:1:3, 1:1:2, 1:1:1, 1:10:10, 1:5:5, 1:3:3, or 1:2:2. **FIG. 9A** depicts a wild-type reference amino acid sequence encoded by a wild-type nucleic acid sequence. **FIG. 9B** depicts a library of amino acid variants, wherein each variant comprising a stretch of sequence (indicated by the patterned circles), wherein each position may have a certain ratio of amino acids in the resultant variant protein library. The resultant variant protein library is encoded by a variant nucleic acid library generated by methods described herein. In this illustration, 5 positions are varied: the first position **900** has a 50/50 K/R ratio; the second position **910** has a 50/25/25 V/L/S ratio, the third position **920** has a 50/25/25 Y/R/D ratio, the fourth position **930** has an equal ratio for all 20 amino acids, and the fifth position **940** has a 75/25 ratio for G/P. The ratios described herein are exemplary only.

**[00106]** *Variation in expression cassettes*

**[00107]** In some instances, a synthesized variant library is generated which encodes for a portion of an expression construct. Exemplary portions of an expression construct include the promoter, open reading frame, and termination region. In some instances, the expression construct encodes for one, two, three or more expression cassettes. An oligonucleotide library may be generated, encoding for codon variation at a single site or multiple sites in separate regions that make up portions of an expression construct cassette, as depicted in **FIG. 11**. To generate a two construct expressing cassette, variant oligonucleotides were synthesized encoding at least a portion of a



variant sequence of a first promoter **1110**, first open reading frame **1120**, first terminator **1130**, second promoter **1140**, second open reading frame **1150**, or second terminator sequence **1160**. After rounds of amplification, as described in previous examples, a library of 1,024 expression constructs was generated. **FIG. 11** provides but one example arrangement. In some instances, additional regulator sequences, such as untranslated regulatory region (UTR) or an enhancer region, are also included in an expression cassette referred to herein. An expression cassette may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more components for which variant sequences are generated by methods described herein. In some instances, the expression construct comprises more than one gene in a multicistronic vector. In one example, the synthesized DNA nucleic acids are inserted into viral vectors (e.g., a lentivirus) and then packaged for transduction into cells, or non-viral vectors for transfer into cells, followed by screening and analysis.

**[00108]** Expression vectors for inserting nucleic acids disclosed herein comprise mammalian cells, e.g., human, non-human primate, pig, rabbit and mouse. Exemplary expression vectors include, without limitation, mammalian expression vectors: pSF-CMV-NEO-NH2-PPT-3XFLAG, pSF-CMV-NEO-COOH-3XFLAG, pSF-CMV-PURO-NH2-GST-TEV, pSF-OXB20-COOH-TEV-FLAG(R)-6His (“6His” disclosed as SEQ ID NO: 41), pCEP4 pDEST27, pSF-CMV-Ub-KrYFP, pSF-CMV-FMDV-daGFP, pEF1a-mCherry-N1 Vector, pEF1a-tdTomato Vector, pSF-CMV-FMDV-Hygro, pSF-CMV-PGK-Puro, pMCP-tag(m), and pSF-CMV-PURO-NH2-CMYC. Nucleic acids synthesized by methods described herein may be transferred into cells by various methods known in the art, including, without limitation, transfection, transduction, and electroporation. Exemplary cellular functions tested include, without limitation, changes in cellular proliferation, death, migration/adhesion, metabolic, and cell-signaling activity.

**[00109] Highly Parallel Nucleic Acid Synthesis**

**[00110]** Provided herein is a platform approach utilizing miniaturization, parallelization, and vertical integration of the end-to-end process from oligonucleotide synthesis to gene assembly within nanowells on silicon to create a revolutionary synthesis platform. Devices described herein provide, with the same footprint as a 96-well plate, a silicon synthesis platform capable of increasing throughput by a factor of up to 1,000 or more compared to traditional synthesis methods, with production of up to approximately 1,000,000 or more oligonucleotides, or 10,000 or more genes in a single highly-parallelized run.

**[00111]** With the advent of next-generation sequencing, high resolution genomic data has become an important factor for studies that delve into the biological roles of various genes in both normal biology and disease pathogenesis. At the core of this research is the central dogma of molecular biology and the concept of “residue-by-residue transfer of sequential information.”



Genomic information encoded in the DNA is transcribed into a message that is then translated into the protein that is the active product within a given biological pathway.

**[00112]** Another exciting area of study is on the discovery, development and manufacturing of therapeutic molecules focused on a highly-specific cellular target. High diversity DNA sequence libraries are at the core of development pipelines for targeted therapeutics. Gene mutants are used to express proteins in a design, build, and test protein engineering cycle that ideally culminates in an optimized gene for high expression of a protein with high affinity for its therapeutic target. As an example, consider the binding pocket of a receptor. The ability to test all sequence permutations of all residues within the binding pocket simultaneously will allow for a thorough exploration, increasing chances of success. Saturation mutagenesis, in which a researcher attempts to generate all possible mutations at a specific site within the receptor, represents one approach to this development challenge. Though costly and time and labor-intensive, it enables each variant to be introduced into each position. In contrast, combinatorial mutagenesis, where a few selected positions or short stretch of DNA may be modified extensively, generates an incomplete repertoire of variants with biased representation.

**[00113]** To accelerate the drug development pipeline, a library with the desired variants available at the intended frequency in the right position available for testing—in other words, a precision library, enables reduced costs as well as turnaround time for screening. Provided herein are methods for synthesizing nucleic acid synthetic variant libraries which provide for precise introduction of each intended variant at the desired frequency. To the end user, this translates to the ability to not only thoroughly sample sequence space but also be able to query these hypotheses in an efficient manner, reducing cost and screening time. Genome-wide editing may elucidate important pathways, libraries where each variant and sequence permutation may be tested for optimal functionality, and thousands of genes may be used to reconstruct entire pathways and genomes to re-engineer biological systems for drug discovery.

**[00114]** In a first example, a drug itself may be optimized using methods described herein. For example, to improve a specified function of an antibody, a variant oligonucleotide library encoding for a portion of the antibody is designed and synthesized. A variant oligonucleotide library for the antibody may then be generated by processes described herein (e.g., PCR mutagenesis followed by insertion into a vector). The antibody is then expressed in a production cell line and screened for enhanced activity. Example screens include examining modulation in binding affinity to an antigen, stability, or effector function (e.g., ADCC, complement, or apoptosis). Exemplary regions to optimize the antibody include, without limitation, the Fc region, Fab region, variable region of



the Fab region, constant region of the Fab region, variable domain of the heavy chain or light chain ( $V_H$  or  $V_L$ ), and specific complementarity-determining regions (CDRs) of  $V_H$  or  $V_L$ .

**[00115]** Nucleic acid libraries synthesized by methods described herein may be expressed in various cells associated with a disease state. Cells associated with a disease state include cell lines, tissue samples, primary cells from a subject, cultured cells expanded from a subject, or cells in a model system. Exemplary cells include without limitation, prokaryotic and eukaryotic cells. Exemplary eukaryotic cells include, without limitation, animal, plant, and fungal cells. Exemplary animal cells include, without limitation, insect, fish and mammalian cells. Exemplary mammalian cells include mouse, human, and primate cells. Exemplary model systems include, without limitation, plant and animal models of a disease state. Exemplary animals include, without limitation, mice, rabbits, primates, fish, and insects. Exemplary plants include, without limitation, a monocot and dicot. Exemplary plants also include, without limitation, microalgae, kelp, cyanobacteria, and green, brown and red algae, wheat, tobacco, and corn, rice, cotton, vegetables, and fruit.

**[00116]** To identify a variant molecule associated with prevention, reduction or treatment of a disease state, a variant nucleic acid library described herein is expressed in a cell associated with a disease state, or one in which a cell a disease state may be induced. In some instances, an agent is used to induce a disease state in cells. Exemplary tools for disease state induction include, without limitation, a Cre/Lox recombination system, LPS inflammation induction, and streptozotocin to induce hypoglycemia. The cells associated with a disease state may be cells from a model system or cultured cells, as well as cells from a subject having a particular disease condition. Exemplary disease conditions include a bacterial, fungal, viral, autoimmune, or proliferative disorder (e.g., cancer). In some instances, the variant nucleic acid library is expressed in the model system, cell line, or primary cells derived from a subject, and screened for changes in at least one cellular activity. Exemplary cellular activities include, without limitation, proliferation, cycle progression, cell death, adhesion, migration, reproduction, cell signaling, energy production, oxygen utilization, metabolic activity, and aging, response to free radical damage, or any combination thereof.

**[00117] Substrates**

**[00118]** Provided herein are substrates comprising a plurality of clusters, wherein each cluster comprises a plurality of loci that support the attachment and synthesis of oligonucleotides. The term "locus" as used herein refers to a discrete region on a structure which provides support for oligonucleotides encoding for a single predetermined sequence to extend from the surface. In some instances, a locus is on a two dimensional surface, e.g., a substantially planar surface. In some instances, a locus refers to a discrete raised or lowered site on a surface e.g., a well, microwell,



channel, or post. In some instances, a surface of a locus comprises a material that is actively functionalized to attach to at least one nucleotide for oligonucleotide synthesis, or preferably, a population of identical nucleotides for synthesis of a population of oligonucleotides. In some instances, oligonucleotide refers to a population of oligonucleotides encoding for the same nucleic acid sequence. In some instances, a surface of a device is inclusive of one or a plurality of surfaces of a substrate.

**[00119]** Average error rates for oligonucleotides synthesized within a library using the systems and methods provided may be less than 1 in 1000, less than 1 in 1250, less than 1 in 1500, less than 1 in 2000, less than 1 in 3000 or less often. In some instances, average error rates for oligonucleotides synthesized within a library using the systems and methods provided are less than 1/500, 1/600, 1/700, 1/800, 1/900, 1/1000, 1/1100, 1/1200, 1/1250, 1/1300, 1/1400, 1/1500, 1/1600, 1/1700, 1/1800, 1/1900, 1/2000, 1/3000, or less. In some instances, average error rates for oligonucleotides synthesized within a library using the systems and methods provided are less than 1/1000.

**[00120]** In some instances, aggregate error rates for oligonucleotides synthesized within a library using the systems and methods provided are less than 1/500, 1/600, 1/700, 1/800, 1/900, 1/1000, 1/1100, 1/1200, 1/1250, 1/1300, 1/1400, 1/1500, 1/1600, 1/1700, 1/1800, 1/1900, 1/2000, 1/3000, or less compared to the predetermined sequences. In some instances, aggregate error rates for oligonucleotides synthesized within a library using the systems and methods provided herein are less than 1/500 or less compared to the predetermined sequences. In some instances, aggregate error rates for oligonucleotides synthesized within a library using the systems and methods provided are less than 1/500, 1/600, 1/700, 1/800, 1/900, or 1/1000. In some instances, aggregate error rates for oligonucleotides synthesized within a library using the systems and methods provided are less than 1/1000.

**[00121]** In some instances, an error correction enzyme may be used for oligonucleotides synthesized within a library using the systems and methods provided herein. In some instances, aggregate error rates for oligonucleotides with error correction may be less than 1/500, 1/600, 1/700, 1/800, 1/900, 1/1000, 1/1100, 1/1200, 1/1300, 1/1400, 1/1500, 1/1600, 1/1700, 1/1800, 1/1900, 1/2000, 1/3000, or less compared to the predetermined sequences. In some instances, aggregate error rates with error correction for oligonucleotides synthesized within a library using the systems and methods provided may be less than 1/500, 1/600, 1/700, 1/800, 1/900, or 1/1000. In some instances, aggregate error rates with error correction for oligonucleotides synthesized within a library using the systems and methods provided may be less than 1/1000.



**[00122]** Error rate may limit the value of gene synthesis for the production of libraries of gene variants. With an error rate of 1/300, about 0.7% of the clones in a 1500 base pair gene will be correct. As most of the errors from oligonucleotide synthesis result in frame-shift mutations, over 99% of the clones in such a library will not produce a full-length protein. Reducing the error rate by 75% would increase the fraction of clones that are correct by a factor of 40. The methods and compositions of the disclosure allow for fast de novo synthesis of large oligonucleotide and gene libraries with error rates that are lower than commonly observed gene synthesis methods both due to the improved quality of synthesis and the applicability of error correction methods that are enabled in a massively parallel and time-efficient manner. Accordingly, libraries may be synthesized with base insertion, deletion, substitution, or total error rates that are under 1/300, 1/400, 1/500, 1/600, 1/700, 1/800, 1/900, 1/1000, 1/1250, 1/1500, 1/2000, 1/2500, 1/3000, 1/4000, 1/5000, 1/6000, 1/7000, 1/8000, 1/9000, 1/10000, 1/12000, 1/15000, 1/20000, 1/25000, 1/30000, 1/40000, 1/50000, 1/60000, 1/70000, 1/80000, 1/90000, 1/100000, 1/125000, 1/150000, 1/200000, 1/300000, 1/400000, 1/500000, 1/600000, 1/700000, 1/800000, 1/900000, 1/1000000, or less, across the library, or across more than 80%, 85%, 90%, 93%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.8%, 99.9%, 99.95%, 99.98%, 99.99%, or more of the library. The methods and compositions of the disclosure further relate to large synthetic oligonucleotide and gene libraries with low error rates associated with at least 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 93%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.8%, 99.9%, 99.95%, 99.98%, 99.99%, or more of the oligonucleotides or genes in at least a subset of the library to relate to error free sequences in comparison to a predetermined/preselected sequence. In some instances, at least 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 93%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.8%, 99.9%, 99.95%, 99.98%, 99.99%, or more of the oligonucleotides or genes in an isolated volume within the library have the same sequence. In some instances, at least 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 93%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.8%, 99.9%, 99.95%, 99.98%, 99.99%, or more of any oligonucleotides or genes related with more than 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9% or more similarity or identity have the same sequence. In some instances, the error rate related to a specified locus on an oligonucleotide or gene is optimized. Thus, a given locus or a plurality of selected loci of one or more oligonucleotides or genes as part of a large library may each have an error rate that is less than 1/300, 1/400, 1/500, 1/600, 1/700, 1/800, 1/900, 1/1000, 1/1250, 1/1500, 1/2000, 1/2500, 1/3000, 1/4000, 1/5000, 1/6000, 1/7000, 1/8000, 1/9000, 1/10000, 1/12000, 1/15000, 1/20000, 1/25000, 1/30000, 1/40000, 1/50000, 1/60000, 1/70000, 1/80000, 1/90000, 1/100000, 1/125000, 1/150000, 1/200000, 1/300000, 1/400000, 1/500000, 1/600000, 1/700000, 1/800000, 1/900000, 1/1000000, or less. In various instances, such error



optimized loci may comprise at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 30000, 50000, 75000, 100000, 500000, 1000000, 2000000, 3000000 or more loci. The error optimized loci may be distributed to at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 30000, 75000, 100000, 500000, 1000000, 2000000, 3000000 or more oligonucleotides or genes.

**[00123]** The error rates may be achieved with or without error correction. The error rates may be achieved across the library, or across more than 80%, 85%, 90%, 93%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.8%, 99.9%, 99.95%, 99.98%, 99.99%, or more of the library.

**[00124]** Provided herein are structures that may comprise a surface that supports the synthesis of a plurality of oligonucleotides having different predetermined sequences at addressable locations on a common support. In some instances, a device provides support for the synthesis of more than 2,000; 5,000; 10,000; 20,000; 30,000; 50,000; 75,000; 100,000; 200,000; 300,000; 400,000; 500,000; 600,000; 700,000; 800,000; 900,000; 1,000,000; 1,200,000; 1,400,000; 1,600,000; 1,800,000; 2,000,000; 2,500,000; 3,000,000; 3,500,000; 4,000,000; 4,500,000; 5,000,000; 10,000,000 or more non-identical oligonucleotides. In some instances, the device provides support for the synthesis of more than 2,000; 5,000; 10,000; 20,000; 30,000; 50,000; 75,000; 100,000; 200,000; 300,000; 400,000; 500,000; 600,000; 700,000; 800,000; 900,000; 1,000,000; 1,200,000; 1,400,000; 1,600,000; 1,800,000; 2,000,000; 2,500,000; 3,000,000; 3,500,000; 4,000,000; 4,500,000; 5,000,000; 10,000,000 or more oligonucleotides encoding for distinct sequences. In some instances, at least a portion of the oligonucleotides have an identical sequence or are configured to be synthesized with an identical sequence.

**[00125]** Provided herein are methods and devices for manufacture and growth of oligonucleotides about 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, or 2000 bases in length. In some instances, the length of the oligonucleotide formed is about 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, or 225 bases in length. An oligonucleotide may be at least 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 bases in length. An oligonucleotide may be from 10 to 225 bases in length, from 12 to 100 bases in length, from 20 to 150 bases in length, from 20 to 130 bases in length, from 30 to 100 bases in length, or from 30 to 70 bases in length.



**[00126]** In some instances, oligonucleotides are synthesized on distinct loci of a substrate, wherein each locus supports the synthesis of a population of oligonucleotides. In some instances, each locus supports the synthesis of a population of oligonucleotides having a different sequence than a population of oligonucleotides grown on another locus. In some instances, the loci of a device are located within a plurality of clusters. In some instances, a device comprises at least 10, 500, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 11000, 12000, 13000, 14000, 15000, 20000, 30000, 40000, 50000 or more clusters. In some instances, a device comprises more than 2,000; 5,000; 10,000; 100,000; 200,000; 300,000; 400,000; 500,000; 600,000; 700,000; 800,000; 900,000; 1,000,000; 1,100,000; 1,200,000; 1,300,000; 1,400,000; 1,500,000; 1,600,000; 1,700,000; 1,800,000; 1,900,000; 2,000,000; 300,000; 400,000; 500,000; 600,000; 700,000; 800,000; 900,000; 1,000,000; 1,200,000; 1,400,000; 1,600,000; 1,800,000; 2,000,000; 2,500,000; 3,000,000; 3,500,000; 4,000,000; 4,500,000; 5,000,000; or 10,000,000 or more distinct loci. In some instances, a device comprises about 10,000 distinct loci. The amount of loci within a single cluster is varied in different instances. In some instances, each cluster includes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 130, 150, 200, 300, 400, 500 or more loci. In some instances, each cluster includes about 50-500 loci. In some instances, each cluster includes about 100-200 loci. In some instances, each cluster includes about 100-150 loci. In some instances, each cluster includes about 109, 121, 130 or 137 loci. In some instances, each cluster includes about 19, 20, 61, 64 or more loci.

**[00127]** The number of distinct oligonucleotides synthesized on a device may be dependent on the number of distinct loci available in the substrate. In some instances, the density of loci within a cluster of a device is at least or about 1 locus per  $\text{mm}^2$ , 10 loci per  $\text{mm}^2$ , 25 loci per  $\text{mm}^2$ , 50 loci per  $\text{mm}^2$ , 65 loci per  $\text{mm}^2$ , 75 loci per  $\text{mm}^2$ , 100 loci per  $\text{mm}^2$ , 130 loci per  $\text{mm}^2$ , 150 loci per  $\text{mm}^2$ , 175 loci per  $\text{mm}^2$ , 200 loci per  $\text{mm}^2$ , 300 loci per  $\text{mm}^2$ , 400 loci per  $\text{mm}^2$ , 500 loci per  $\text{mm}^2$ , 1,000 loci per  $\text{mm}^2$  or more. In some instances, a device comprises from about 10 loci per  $\text{mm}^2$  to about 500  $\text{mm}^2$ , from about 25 loci per  $\text{mm}^2$  to about 400  $\text{mm}^2$ , from about 50 loci per  $\text{mm}^2$  to about 500  $\text{mm}^2$ , from about 100 loci per  $\text{mm}^2$  to about 500  $\text{mm}^2$ , from about 150 loci per  $\text{mm}^2$  to about 500  $\text{mm}^2$ , from about 10 loci per  $\text{mm}^2$  to about 250  $\text{mm}^2$ , from about 50 loci per  $\text{mm}^2$  to about 250  $\text{mm}^2$ , from about 10 loci per  $\text{mm}^2$  to about 200  $\text{mm}^2$ , or from about 50 loci per  $\text{mm}^2$  to about 200  $\text{mm}^2$ . In some instances, the distance from the centers of two adjacent loci within a cluster is from about 10  $\mu\text{m}$  to about 500  $\mu\text{m}$ , from about 10  $\mu\text{m}$  to about 200  $\mu\text{m}$ , or from about 10  $\mu\text{m}$  to about 100  $\mu\text{m}$ . In some instances, the distance from two centers of adjacent loci is greater than about 10  $\mu\text{m}$ , 20  $\mu\text{m}$ , 30  $\mu\text{m}$ , 40  $\mu\text{m}$ , 50  $\mu\text{m}$ , 60  $\mu\text{m}$ , 70  $\mu\text{m}$ , 80  $\mu\text{m}$ , 90  $\mu\text{m}$  or 100  $\mu\text{m}$ . In some instances, the distance from the centers of two adjacent loci is less than about 200  $\mu\text{m}$ , 150  $\mu\text{m}$ , 100  $\mu\text{m}$ , 80  $\mu\text{m}$ ,



70  $\mu\text{m}$ , 60  $\mu\text{m}$ , 50  $\mu\text{m}$ , 40  $\mu\text{m}$ , 30  $\mu\text{m}$ , 20  $\mu\text{m}$  or 10  $\mu\text{m}$ . In some instances, each locus has a width of about 0.5  $\mu\text{m}$ , 1  $\mu\text{m}$ , 2  $\mu\text{m}$ , 3  $\mu\text{m}$ , 4  $\mu\text{m}$ , 5  $\mu\text{m}$ , 6  $\mu\text{m}$ , 7  $\mu\text{m}$ , 8  $\mu\text{m}$ , 9  $\mu\text{m}$ , 10  $\mu\text{m}$ , 20  $\mu\text{m}$ , 30  $\mu\text{m}$ , 40  $\mu\text{m}$ , 50  $\mu\text{m}$ , 60  $\mu\text{m}$ , 70  $\mu\text{m}$ , 80  $\mu\text{m}$ , 90  $\mu\text{m}$  or 100  $\mu\text{m}$ . In some instances, each locus has a width of about 0.5  $\mu\text{m}$  to 100 $\mu\text{m}$ , about 0.5  $\mu\text{m}$  to 50  $\mu\text{m}$ , about 10  $\mu\text{m}$  to 75  $\mu\text{m}$ , or about 0.5  $\mu\text{m}$  to 50  $\mu\text{m}$ .

**[00128]** In some instances, the density of clusters within a device is at least or about 1 cluster per 100  $\text{mm}^2$ , 1 cluster per 10  $\text{mm}^2$ , 1 cluster per 5  $\text{mm}^2$ , 1 cluster per 4  $\text{mm}^2$ , 1 cluster per 3  $\text{mm}^2$ , 1 cluster per 2  $\text{mm}^2$ , 1 cluster per 1  $\text{mm}^2$ , 2 clusters per 1  $\text{mm}^2$ , 3 clusters per 1  $\text{mm}^2$ , 4 clusters per 1  $\text{mm}^2$ , 5 clusters per 1  $\text{mm}^2$ , 10 clusters per 1  $\text{mm}^2$ , 50 clusters per 1  $\text{mm}^2$  or more. In some instances, a device comprises from about 1 cluster per 10  $\text{mm}^2$  to about 10 clusters per 1  $\text{mm}^2$ . In some instances, the distance from the centers of two adjacent clusters is less than about 50  $\mu\text{m}$ , 100  $\mu\text{m}$ , 200  $\mu\text{m}$ , 500  $\mu\text{m}$ , 1000  $\mu\text{m}$ , or 2000  $\mu\text{m}$  or 5000  $\mu\text{m}$ . In some instances, the distance from the centers of two adjacent clusters is from about 50  $\mu\text{m}$  to about 100  $\mu\text{m}$ , from about 50  $\mu\text{m}$  to about 200  $\mu\text{m}$ , from about 50  $\mu\text{m}$  to about 300  $\mu\text{m}$ , from about 50  $\mu\text{m}$  to about 500  $\mu\text{m}$ , and from about 100  $\mu\text{m}$  to about 2000  $\mu\text{m}$ . In some instances, the distance from the centers of two adjacent clusters is from about 0.05  $\text{mm}$  to about 50  $\text{mm}$ , from about 0.05  $\text{mm}$  to about 10  $\text{mm}$ , from about 0.05  $\text{mm}$  to about 5  $\text{mm}$ , from about 0.05  $\text{mm}$  to about 4  $\text{mm}$ , from about 0.05  $\text{mm}$  to about 3  $\text{mm}$ , from about 0.05  $\text{mm}$  to about 2  $\text{mm}$ , from about 0.1  $\text{mm}$  to 10  $\text{mm}$ , from about 0.2  $\text{mm}$  to 10  $\text{mm}$ , from about 0.3  $\text{mm}$  to about 10  $\text{mm}$ , from about 0.4  $\text{mm}$  to about 10  $\text{mm}$ , from about 0.5  $\text{mm}$  to 10  $\text{mm}$ , from about 0.5  $\text{mm}$  to about 5  $\text{mm}$ , or from about 0.5  $\text{mm}$  to about 2  $\text{mm}$ . In some instances, each cluster has a diameter or width along one dimension of about 0.5 to 2  $\text{mm}$ , about 0.5 to 1  $\text{mm}$ , or about 1 to 2  $\text{mm}$ . In some instances, each cluster has a diameter or width along one dimension of about 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9 or 2  $\text{mm}$ . In some instances, each cluster has an interior diameter or width along one dimension of about 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.1, 1.15, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9 or 2  $\text{mm}$ .

**[00129]** Structures for oligonucleotide synthesis for use with devices, compositions, systems, and methods as described herein comprise a variety of sizes. A device may be about the size of a standard 96 well plate, for example from about 100 and 200  $\text{mm}$  by about 50 and 150  $\text{mm}$ . In some instances, a device has a diameter less than or equal to about 1000  $\text{mm}$ , 500  $\text{mm}$ , 450  $\text{mm}$ , 400  $\text{mm}$ , 300  $\text{mm}$ , 250  $\text{mm}$ , 200  $\text{mm}$ , 150  $\text{mm}$ , 100  $\text{mm}$  or 50  $\text{mm}$ . In some instances, the diameter of a device is from about 25  $\text{mm}$  to 1000  $\text{mm}$ , from about 25  $\text{mm}$  to about 800  $\text{mm}$ , from about 25  $\text{mm}$  to about 600  $\text{mm}$ , from about 25  $\text{mm}$  to about 500  $\text{mm}$ , from about 25  $\text{mm}$  to about 400  $\text{mm}$ , from about 25  $\text{mm}$  to about 300  $\text{mm}$ , or from about 25  $\text{mm}$  to about 200. Non-limiting examples of device size include about 300  $\text{mm}$ , 200  $\text{mm}$ , 150  $\text{mm}$ , 130  $\text{mm}$ , 100  $\text{mm}$ , 76  $\text{mm}$ , 51  $\text{mm}$  and 25  $\text{mm}$ . In some instances, a device has a planar surface area of at least about 100  $\text{mm}^2$ ; 200  $\text{mm}^2$ ; 500



mm<sup>2</sup>; 1,000 mm<sup>2</sup>; 2,000 mm<sup>2</sup>; 5,000 mm<sup>2</sup>; 10,000 mm<sup>2</sup>; 12,000 mm<sup>2</sup>; 15,000 mm<sup>2</sup>; 20,000 mm<sup>2</sup>; 30,000 mm<sup>2</sup>; 40,000 mm<sup>2</sup>; 50,000 mm<sup>2</sup> or more. In some instances, the thickness of a device is from about 50 mm to about 2000 mm, from about 50 mm to about 1000 mm, from about 100 mm to about 1000 mm, from about 200 mm to about 1000 mm, or from about 250 mm to about 1000 mm. Non-limiting examples of device thickness include 275 mm, 375 mm, 525 mm, 625 mm, 675 mm, 725 mm, 775 mm and 925 mm. In some instances, the thickness of a device varies with diameter and depends on the composition of the substrate. For example, a device comprising materials other than silicon has a different thickness than a silicon device of the same diameter. Device thickness may be determined by the mechanical strength of the material used and the device must be thick enough to support its own weight without cracking during handling. In some instances, a structure comprises a plurality of devices described herein.

**[00130] Surface Materials**

**[00131]** Provided herein is a device comprising a surface, wherein the surface is modified to support oligonucleotide synthesis at predetermined locations and with a resulting low error rate, a low dropout rate, a high yield, and a high oligo representation. In some embodiments, surfaces of a device for oligonucleotide synthesis provided herein are fabricated from a variety of materials capable of modification to support a de novo oligonucleotide synthesis reaction. In some cases, the devices are sufficiently conductive, *e.g.*, are able to form uniform electric fields across all or a portion of the device. A device described herein may comprise a flexible material. Exemplary flexible materials include, without limitation, modified nylon, unmodified nylon, nitrocellulose, and polypropylene. A device described herein may comprise a rigid material. Exemplary rigid materials include, without limitation, glass, fuses silica, silicon, silicon dioxide, silicon nitride, plastics (for example, polytetrafluoroethylene, polypropylene, polystyrene, polycarbonate, and blends thereof, and metals (for example, gold, platinum). Devices disclosed herein may be fabricated from a material comprising silicon, polystyrene, agarose, dextran, cellulosic polymers, polyacrylamides, polydimethylsiloxane (PDMS), glass, or any combination thereof. In some cases, a device disclosed herein is manufactured with a combination of materials listed herein or any other suitable material known in the art.

**[00132]** In some cases, a device disclosed herein comprises a silicon dioxide base and a surface layer of silicon oxide. Alternatively, the device may have a base of silicon oxide. A surface of the device provided here may be textured, resulting in an increase in overall surface area for oligonucleotide synthesis. Devices disclosed herein may comprise at least 5 %, 10%, 25%, 50%, 80%, 90%, 95%, or 99% silicon. A device disclosed herein may be fabricated from a silicon on insulator (SOI) wafer.



**[00133]** Substrates, devices and reactors provided herein are fabricated from any variety of materials suitable for the methods and compositions described herein. In certain instances, device materials are fabricated to exhibit a low level of nucleotide binding. In some instances, device materials are modified to generate distinct surfaces that exhibit a high level of nucleotide binding. In some instances, device materials are transparent to visible and/or UV light. In some instances, device materials are sufficiently conductive, *e.g.*, are able to form uniform electric fields across all or a portion of a substrate. In some instances, conductive materials are connected to an electric ground. In some instances, the device is heat conductive or insulated. In some instances, the materials are chemical resistant and heat resistant to support chemical or biochemical reactions, for example oligonucleotide synthesis reaction processes. In some instances, a device comprises flexible materials. Flexible materials include, without limitation, modified nylon, unmodified nylon, nitrocellulose, polypropylene, and the like. In some instances, a device comprises rigid materials. Rigid materials include, without limitation, glass, fused silica, silicon, silicon dioxide, silicon nitride, plastics (for example, polytetrafluoroethylene, polypropylene, polystyrene, polycarbonate, and blends thereof, and the like), and metals (for example, gold, platinum, and the like). In some instances, a device is fabricated from a material comprising silicon, polystyrene, agarose, dextran, cellulosic polymers, polyacrylamides, polydimethylsiloxane (PDMS), glass, or any combination thereof. In some instances, a device is manufactured with a combination of materials listed herein or any other suitable material known in the art.

**[00134] Surface Architecture**

**[00135]** Provided herein are devices comprising raised and/or lowered features. One benefit of having such features is an increase in surface area to support oligonucleotide synthesis. In some instances, a device having raised and/or lowered features is referred to as a three-dimensional substrate. In some instances, a three-dimensional device comprises one or more channels. In some instances, one or more loci comprise a channel. In some instances, the channels are accessible to reagent deposition via a deposition device such as an oligonucleotide synthesizer. In some instances, reagents and/or fluids collect in a larger well in fluid communication with one or more channels. For example, a device comprises a plurality of channels corresponding to a plurality of loci with a cluster, and the plurality of channels are in fluid communication with one well of the cluster. In some methods, a library of oligonucleotides is synthesized in a plurality of loci of a cluster.

**[00136]** In some instances, the structure is configured to allow for controlled flow and mass transfer paths for oligonucleotide synthesis on a surface. In some instances, the configuration of a device allows for the controlled and even distribution of mass transfer paths, chemical exposure



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

**[00137]** Provided herein are methods to synthesize an amount of DNA of 1 fM, 5 fM, 10 fM, 25 fM, 50 fM, 75 fM, 100 fM, 200 fM, 300 fM, 400 fM, 500 fM, 600 fM, 700 fM, 800 fM, 900 fM, 1 pM, 5 pM, 10 pM, 25 pM, 50 pM, 75 pM, 100 pM, 200 pM, 300 pM, 400 pM, 500 pM, 600 pM, 700 pM, 800 pM, 900 pM, or more. In some instances, an oligonucleotide library may span the length of about 1 %, 2 %, 3 %, 4 %, 5 %, 10 %, 15 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 95 %, or 100 % of a gene. A gene may be varied up to about 1 %, 2 %, 3 %, 4 %, 5 %, 10 %, 15 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 85%, 90 %, 95 %, or 100 %.

**[00138]** Non-identical oligonucleotides may collectively encode a sequence for at least 1 %, 2 %, 3 %, 4 %, 5 %, 10 %, 15 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 85%, 90 %, 95 %, or 100 % of a gene. In some instances, an oligonucleotide may encode a sequence of 50 %, 60 %, 70 %, 80 %, 85%, 90 %, 95 %, or more of a gene. In some instances, an oligonucleotide may encode a sequence of 80 %, 85%, 90 %, 95 %, or more of a gene.

**[00139]** In some instances, segregation is achieved by physical structure. In some instances, segregation is achieved by differential functionalization of the surface generating active and passive regions for oligonucleotide synthesis. Differential functionalization is also achieved by alternating the hydrophobicity across the device surface, thereby creating water contact angle effects that cause beading or wetting of the deposited reagents. Employing larger structures may decrease splashing and cross-contamination of distinct oligonucleotide synthesis locations with reagents of the neighboring spots. In some instances, a device, such as an oligonucleotide synthesizer, is used to deposit reagents to distinct oligonucleotide synthesis locations. Substrates having three-dimensional features are configured in a manner that allows for the synthesis of a large number of oligonucleotides (*e.g.*, more than about 10,000) with a low error rate (*e.g.*, less than about 1:500, 1:1000, 1:1500, 1:2,000; 1:3,000; 1:5,000; or 1:10,000). In some instances, a device comprises features with a density of about or greater than about 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 300, 400 or 500 features per mm<sup>2</sup>.

**[00140]** A well of a device may have the same or different width, height, and/or volume as another well of the substrate. A channel of a device may have the same or different width, height,



and/or volume as another channel of the substrate. In some instances, the width of a cluster is from about 0.05 mm to about 50 mm, from about 0.05 mm to about 10 mm, from about 0.05 mm to about 5 mm, from about 0.05 mm to about 4 mm, from about 0.05 mm to about 3 mm, from about 0.05 mm to about 2 mm, from about 0.05 mm to about 1 mm, from about 0.05 mm to about 0.5 mm, from about 0.05 mm to about 0.1 mm, from about 0.1 mm to 10 mm, from about 0.2 mm to about 10 mm, from about 0.3 mm to about 10 mm, from about 0.4 mm to about 10 mm, from about 0.5 mm to about 10 mm, from about 0.5 mm to about 5 mm, or from about 0.5 mm to about 2 mm. In some instances, the width of a well comprising a cluster is from about 0.05 mm to about 50 mm, from about 0.05 mm to about 10 mm, from about 0.05 mm to about 5 mm, from about 0.05 mm to about 4 mm, from about 0.05 mm to about 3 mm, from about 0.05 mm to about 2 mm, from about 0.05 mm to about 1 mm, from about 0.05 mm to about 0.5 mm, from about 0.05 mm to about 0.1 mm, from about 0.1 mm to about 10 mm, from about 0.2 mm to about 10 mm, from about 0.3 mm to about 10 mm, from about 0.4 mm to about 10 mm, from about 0.5 mm to about 10 mm, from about 0.5 mm to about 5 mm, or from about 0.5 mm to about 2 mm. In some instances, the width of a cluster is less than or about 5 mm, 4 mm, 3 mm, 2 mm, 1 mm, 0.5 mm, 0.1 mm, 0.09 mm, 0.08 mm, 0.07 mm, 0.06 mm or 0.05 mm. In some instances, the width of a cluster is from about 1.0 to about 1.3 mm. In some instances, the width of a cluster is about 1.150 mm. In some instances, the width of a well is less than or about 5 mm, 4 mm, 3 mm, 2 mm, 1 mm, 0.5 mm, 0.1 mm, 0.09 mm, 0.08 mm, 0.07 mm, 0.06 mm or 0.05 mm. In some instances, the width of a well is from about 1.0 to about 1.3 mm. In some instances, the width of a well is about 1.150 mm. In some instances, the width of a cluster is about 0.08 mm. In some instances, the width of a well is about 0.08 mm. The width of a cluster may refer to clusters within a two-dimensional or three-dimensional substrate.

**[00141]** In some instances, the height of a well is from about 20 um to about 1000 um, from about 50 um to about 1000 um, from about 100 um to about 1000 um, from about 200 um to about 1000 um, from about 300 um to about 1000 um, from about 400 um to about 1000 um, or from about 500 um to about 1000 um. In some instances, the height of a well is less than about 1000 um, less than about 900 um, less than about 800 um, less than about 700 um, or less than about 600 um.

**[00142]** In some instances, a device comprises a plurality of channels corresponding to a plurality of loci within a cluster, wherein the height or depth of a channel is from about 5 um to about 500 um, from about 5 um to about 400 um, from about 5 um to about 300 um, from about 5 um to about 200 um, from about 5 um to about 100 um, from about 5 um to about 50 um, or from about 10 um to about 50 um. In some instances, the height of a channel is less than 100 um, less than 80 um, less than 60 um, less than 40 um or less than 20 um.



**[00143]** In some instances, the diameter of a channel, locus (*e.g.*, in a substantially planar substrate) or both channel and locus (*e.g.*, in a three-dimensional device wherein a locus corresponds to a channel) is from about 1  $\mu\text{m}$  to about 1000  $\mu\text{m}$ , from about 1  $\mu\text{m}$  to about 500  $\mu\text{m}$ , from about 1  $\mu\text{m}$  to about 200  $\mu\text{m}$ , from about 1  $\mu\text{m}$  to about 100  $\mu\text{m}$ , from about 5  $\mu\text{m}$  to about 100  $\mu\text{m}$ , or from about 10  $\mu\text{m}$  to about 100  $\mu\text{m}$ , for example, about 90  $\mu\text{m}$ , 80  $\mu\text{m}$ , 70  $\mu\text{m}$ , 60  $\mu\text{m}$ , 50  $\mu\text{m}$ , 40  $\mu\text{m}$ , 30  $\mu\text{m}$ , 20  $\mu\text{m}$  or 10  $\mu\text{m}$ . In some instances, the diameter of a channel, locus, or both channel and locus is less than about 100  $\mu\text{m}$ , 90  $\mu\text{m}$ , 80  $\mu\text{m}$ , 70  $\mu\text{m}$ , 60  $\mu\text{m}$ , 50  $\mu\text{m}$ , 40  $\mu\text{m}$ , 30  $\mu\text{m}$ , 20  $\mu\text{m}$  or 10  $\mu\text{m}$ . In some instances, the distance from the center of two adjacent channels, loci, or channels and loci is from about 1  $\mu\text{m}$  to about 500  $\mu\text{m}$ , from about 1  $\mu\text{m}$  to about 200  $\mu\text{m}$ , from about 1  $\mu\text{m}$  to about 100  $\mu\text{m}$ , from about 5  $\mu\text{m}$  to about 200  $\mu\text{m}$ , from about 5  $\mu\text{m}$  to about 100  $\mu\text{m}$ , from about 5  $\mu\text{m}$  to about 50  $\mu\text{m}$ , or from about 5  $\mu\text{m}$  to about 30  $\mu\text{m}$ , for example, about 20  $\mu\text{m}$ .

**[00144] Surface Modifications**

**[00145]** In various instances, surface modifications are employed for the chemical and/or physical alteration of a surface by an additive or subtractive process to change one or more chemical and/or physical properties of a device surface or a selected site or region of a device surface. For example, surface modifications include, without limitation, (1) changing the wetting properties of a surface, (2) functionalizing a surface, *i.e.*, providing, modifying or substituting surface functional groups, (3) defunctionalizing a surface, *i.e.*, removing surface functional groups, (4) otherwise altering the chemical composition of a surface, *e.g.*, through etching, (5) increasing or decreasing surface roughness, (6) providing a coating on a surface, *e.g.*, a coating that exhibits wetting properties that are different from the wetting properties of the surface, and/or (7) depositing particulates on a surface.

**[00146]** In some instances, the addition of a chemical layer on top of a surface (referred to as an adhesion promoter) facilitates structured patterning of loci on a surface of a substrate. Exemplary surfaces for application of adhesion promotion include, without limitation, glass, silicon, silicon dioxide and silicon nitride. In some instances, the adhesion promoter is a chemical with a high surface energy. In some instances, a second chemical layer is deposited on a surface of a substrate. In some instances, the second chemical layer has a low surface energy. In some instances, surface energy of a chemical layer coated on a surface supports localization of droplets on the surface. Depending on the patterning arrangement selected, the proximity of loci and/or area of fluid contact at the loci are alterable.

**[00147]** In some instances, a device surface, or resolved loci, onto which nucleic acids or other moieties are deposited, *e.g.*, for oligonucleotide synthesis, are smooth or substantially planar (*e.g.*, two-dimensional) or have irregularities, such as raised or lowered features (*e.g.*, three-dimensional



features). In some instances, a device surface is modified with one or more different layers of compounds. Such modification layers of interest include, without limitation, inorganic and organic layers such as metals, metal oxides, polymers, small organic molecules and the like. Non-limiting polymeric layers include peptides, proteins, nucleic acids or mimetics thereof (*e.g.*, peptide nucleic acids and the like), polysaccharides, phospholipids, polyurethanes, polyesters, polycarbonates, polyureas, polyamides, polyethyleneamines, polyarylene sulfides, polysiloxanes, polyimides, polyacetates, and any other suitable compounds described herein or otherwise known in the art. In some instances, polymers are heteropolymeric. In some instances, polymers are homopolymeric. In some instances, polymers comprise functional moieties or are conjugated.

**[00148]** In some instances, resolved loci of a device are functionalized with one or more moieties that increase and/or decrease surface energy. In some instances, a moiety is chemically inert. In some instances, a moiety is configured to support a desired chemical reaction, for example, one or more processes in an oligonucleotide synthesis reaction. The surface energy, or hydrophobicity, of a surface is a factor for determining the affinity of a nucleotide to attach onto the surface. In some instances, a method for device functionalization may comprise: (a) providing a device having a surface that comprises silicon dioxide; and (b) silanizing the surface using, a suitable silanizing agent described herein or otherwise known in the art, for example, an organofunctional alkoxysilane molecule.

**[00149]** In some instances, the organofunctional alkoxysilane molecule comprises dimethylchloro-octodecyl-silane, methylchloro-octodecyl-silane, trichloro-octodecyl-silane, trimethyl-octodecyl-silane, triethyl-octodecyl-silane, or any combination thereof. In some instances, a device surface comprises functionalized with polyethylene/polypropylene (functionalized by gamma irradiation or chromic acid oxidation, and reduction to hydroxyalkyl surface), highly crosslinked polystyrene-divinylbenzene (derivatized by chloromethylation, and aminated to benzylamine functional surface), nylon (the terminal aminoethyl groups are directly reactive), or etched with reduced polytetrafluoroethylene. Other methods and functionalizing agents are described in U.S. Patent No. 5474796, which is herein incorporated by reference in its entirety.

**[00150]** In some instances, a device surface is functionalized by contact with a derivatizing composition that contains a mixture of silanes, under reaction conditions effective to couple the silanes to the device surface, typically via reactive hydrophilic moieties present on the device surface. Silanization generally covers a surface through self-assembly with organofunctional alkoxysilane molecules.



**[00151]** A variety of siloxane functionalizing reagents may further be used as currently known in the art, *e.g.*, for lowering or increasing surface energy. The organofunctional alkoxy silanes may be classified according to their organic functions.

**[00152]** Provided herein are devices that may contain patterning of agents capable of coupling to a nucleoside. In some instances, a device may be coated with an active agent. In some instances, a device may be coated with a passive agent. Exemplary active agents for inclusion in coating materials described herein include, without limitation, N-(3-triethoxysilylpropyl)-4-hydroxybutyramide (HAPS), 11-acetoxyundecyltriethoxysilane, n-decyltriethoxysilane, (3-aminopropyl)trimethoxysilane, (3-aminopropyl)triethoxysilane, 3-glycidoxypropyltrimethoxysilane (GOPS), 3-iodo-propyltrimethoxysilane, butyl-aldehyd-trimethoxysilane, dimeric secondary aminoalkyl siloxanes, (3-aminopropyl)-diethoxy-methylsilane, (3-aminopropyl)-dimethyl-ethoxysilane, and (3-aminopropyl)-trimethoxysilane, (3-glycidoxypropyl)-dimethyl-ethoxysilane, glycidoxy-trimethoxysilane, (3-mercaptopropyl)-trimethoxysilane, 3-4 epoxycyclohexyl-ethyltrimethoxysilane, and (3-mercaptopropyl)-methyl-dimethoxysilane, allyl trichlorochlorosilane, 7-oct-1-enyl trichlorochlorosilane, or bis (3-trimethoxysilylpropyl) amine.

**[00153]** Exemplary passive agents for inclusion in a coating material described herein include, without limitation, perfluorooctyltrichlorosilane; tridecafluoro-1,1,2,2-tetrahydrooctyl)trichlorosilane; 1H, 1H, 2H, 2H-fluorooctyltriethoxysilane (FOS); trichloro(1H, 1H, 2H, 2H - perfluorooctyl)silane; tert-butyl-[5-fluoro-4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)indol-1-yl]-dimethyl-silane; CYTOP™; Fluorinert™; perfluorooctyltrichlorosilane (PFOTCS); perfluorooctyldimethylchlorosilane (PFODCS); perfluorodecyltriethoxysilane (PFDTES); pentafluorophenyl-dimethylpropylchloro-silane (PFPTES); perfluorooctyltriethoxysilane; perfluorooctyltrimethoxysilane; octylchlorosilane; dimethylchloro-octodecyl-silane; methyldichloro-octodecyl-silane; trichloro-octodecyl-silane; trimethyl-octodecyl-silane; triethyl-octodecyl-silane; or octadecyltrichlorosilane.

**[00154]** In some instances, a functionalization agent comprises a hydrocarbon silane such as octadecyltrichlorosilane. In some instances, the functionalizing agent comprises 11-acetoxyundecyltriethoxysilane, n-decyltriethoxysilane, (3-aminopropyl)trimethoxysilane, (3-aminopropyl)triethoxysilane, glycidyloxypropyl/trimethoxysilane and N-(3-triethoxysilylpropyl)-4-hydroxybutyramide.

**[00155] Oligonucleotide Synthesis**

**[00156]** Methods of the current disclosure for oligonucleotide synthesis may include processes involving phosphoramidite chemistry. In some instances, oligonucleotide synthesis comprises coupling a base with phosphoramidite. Oligonucleotide synthesis may comprise coupling a base by



deposition of phosphoramidite under coupling conditions, wherein the same base is optionally deposited with phosphoramidite more than once, i.e., double coupling. Oligonucleotide synthesis may comprise capping of unreacted sites. In some instances, capping is optional. Oligonucleotide synthesis may also comprise oxidation or an oxidation step or oxidation steps. Oligonucleotide synthesis may comprise deblocking, detritylation, and sulfurization. In some instances, oligonucleotide synthesis comprises either oxidation or sulfurization. In some instances, between one or each step during an oligonucleotide synthesis reaction, the device is washed, for example, using tetrazole or acetonitrile. Time frames for any one step in a phosphoramidite synthesis method may be less than about 2 min, 1 min, 50 sec, 40 sec, 30 sec, 20 sec and 10 sec.

**[00157]** Oligonucleotide synthesis using a phosphoramidite method may comprise a subsequent addition of a phosphoramidite building block (*e.g.*, nucleoside phosphoramidite) to a growing oligonucleotide chain for the formation of a phosphite triester linkage. Phosphoramidite oligonucleotide synthesis proceeds in the 3' to 5' direction. Phosphoramidite oligonucleotide synthesis allows for the controlled addition of one nucleotide to a growing nucleic acid chain per synthesis cycle. In some instances, each synthesis cycle comprises a coupling step.

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

**[00158]** Following coupling, phosphoramidite oligonucleotide synthesis methods optionally comprise a capping step. In a capping step, the growing oligonucleotide is treated with a capping agent. A capping step is useful to block unreacted substrate-bound 5'-OH groups after coupling from further chain elongation, preventing the formation of oligonucleotides with internal base



deletions. Further, phosphoramidites activated with 1H-tetrazole may react, to a small extent, with the O6 position of guanosine. Without being bound by theory, upon oxidation with I<sub>2</sub>/water, this side product, possibly via O6-N7 migration, may undergo depurination. The apurinic sites may end up being cleaved in the course of the final deprotection of the oligonucleotide thus reducing the yield of the full-length product. The O6 modifications may be removed by treatment with the capping reagent prior to oxidation with I<sub>2</sub>/water. In some instances, inclusion of a capping step during oligonucleotide synthesis decreases the error rate as compared to synthesis without capping. As an example, the capping step comprises treating the substrate-bound oligonucleotide with a mixture of acetic anhydride and 1-methylimidazole. Following a capping step, the device is optionally washed.

**[00159]** In some instances, following addition of a nucleoside phosphoramidite, and optionally after capping and one or more wash steps, the device bound growing oligonucleotide is oxidized. The oxidation step comprises the phosphite triester is oxidized into a tetracoordinated phosphate triester, a protected precursor of the naturally occurring phosphate diester internucleoside linkage. In some instances, oxidation of the growing oligonucleotide is achieved by treatment with iodine and water, optionally in the presence of a weak base (*e.g.*, pyridine, lutidine, collidine). Oxidation may be carried out under anhydrous conditions using, *e.g.* tert-Butyl hydroperoxide or (1S)-(+)-(10-camphorsulfonyl)-oxaziridine (CSO). In some methods, a capping step is performed following oxidation. A second capping step allows for device drying, as residual water from oxidation that may persist may inhibit subsequent coupling. Following oxidation, the device and growing oligonucleotide is optionally washed. In some instances, the step of oxidation is substituted with a sulfurization step to obtain oligonucleotide phosphorothioates, wherein any capping steps may be performed after the sulfurization. Many reagents are capable of the efficient sulfur transfer, including but not limited to 3-(Dimethylaminomethylidene)amino)-3H-1,2,4-dithiazole-3-thione, DDTT, 3H-1,2-benzodithiol-3-one 1,1-dioxide, also known as Beaucage reagent, and N,N,N'N'-Tetraethylthiuram disulfide (TETD).

**[00160]** In order for a subsequent cycle of nucleoside incorporation to occur through coupling, the protected 5' end of the device bound growing oligonucleotide is removed so that the primary hydroxyl group is reactive with a next nucleoside phosphoramidite. In some instances, the protecting group is DMT and deblocking occurs with trichloroacetic acid in dichloromethane. Conducting detritylation for an extended time or with stronger than recommended solutions of acids may lead to increased depurination of solid support-bound oligonucleotide and thus reduces the yield of the desired full-length product. Methods and compositions of the disclosure described herein provide for controlled deblocking conditions limiting undesired depurination reactions. In



some instances, the device bound oligonucleotide is washed after deblocking. In some instances, efficient washing after deblocking contributes to synthesized oligonucleotides having a low error rate.

**[00161]** Methods for the synthesis of oligonucleotides typically involve an iterating sequence of the following steps: application of a protected monomer to an actively functionalized surface (*e.g.*, locus) to link with either the activated surface, a linker or with a previously deprotected monomer; deprotection of the applied monomer so that it is reactive with a subsequently applied protected monomer; and application of another protected monomer for linking. One or more intermediate steps include oxidation or sulfurization. In some instances, one or more wash steps precede or follow one or all of the steps.

**[00162]** Methods for phosphoramidite-based oligonucleotide synthesis comprise a series of chemical steps. In some instances, one or more steps of a synthesis method involve reagent cycling, where one or more steps of the method comprise application to the device of a reagent useful for the step. For example, reagents are cycled by a series of liquid deposition and vacuum drying steps. For substrates comprising three-dimensional features such as wells, microwells, channels and the like, reagents are optionally passed through one or more regions of the device via the wells and/or channels.

**[00163]** Methods and systems described herein relate to oligonucleotide synthesis devices for the synthesis of oligonucleotides. The synthesis may be in parallel. For example, at least or about at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 1000, 10000, 50000, 75000, 100000 or more oligonucleotides may be synthesized in parallel. The total number oligonucleotides that may be synthesized in parallel may be from 2-100000, 3-50000, 4-10000, 5-1000, 6-900, 7-850, 8-800, 9-750, 10-700, 11-650, 12-600, 13-550, 14-500, 15-450, 16-400, 17-350, 18-300, 19-250, 20-200, 21-150, 22-100, 23-50, 24-45, 25-40, 30-35. Those of skill in the art appreciate that the total number of oligonucleotides synthesized in parallel may fall within any range bound by any of these values, for example 25-100. The total number of oligonucleotides synthesized in parallel may fall within any range defined by any of the values serving as endpoints of the range. Total molar mass of oligonucleotides synthesized within the device or the molar mass of each of the oligonucleotides may be at least or at least about 10, 20, 30, 40, 50, 100, 250, 500, 750, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 25000, 50000, 75000, 100000 picomoles, or more. The length of each of the oligonucleotides or average length of the oligonucleotides within the device may be at least or about at least 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 150, 200, 300, 400, 500 nucleotides, or more. The length of each of the oligonucleotides or



average length of the oligonucleotides within the device may be at most or about at most 500, 400, 300, 200, 150, 100, 50, 45, 35, 30, 25, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10 nucleotides, or less. The length of each of the oligonucleotides or average length of the oligonucleotides within the device may fall from 10-500, 9-400, 11-300, 12-200, 13-150, 14-100, 15-50, 16-45, 17-40, 18-35, 19-25. Those of skill in the art appreciate that the length of each of the oligonucleotides or average length of the oligonucleotides within the device may fall within any range bound by any of these values, for example 100-300. The length of each of the oligonucleotides or average length of the oligonucleotides within the device may fall within any range defined by any of the values serving as endpoints of the range.

**[00164]** Methods for oligonucleotide synthesis on a surface provided herein allow for synthesis at a fast rate. As an example, at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 55, 60, 70, 80, 90, 100, 125, 150, 175, 200 nucleotides per hour, or more are synthesized. Nucleotides include adenine, guanine, thymine, cytosine, uridine building blocks, or analogs/modified versions thereof. In some instances, libraries of oligonucleotides are synthesized in parallel on a substrate. For example, a device comprising about or at least about 100; 1,000; 10,000; 30,000; 75,000; 100,000; 1,000,000; 2,000,000; 3,000,000; 4,000,000; or 5,000,000 resolved loci is able to support the synthesis of at least the same number of distinct oligonucleotides, wherein each oligonucleotide encoding a distinct sequence is synthesized on a resolved locus. In some instances, a library of oligonucleotides is synthesized on a device with low error rates described herein in less than about three months, two months, one month, three weeks, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 days, 24 hours or less. In some instances, larger nucleic acids assembled from an oligonucleotide library synthesized with a low error rate using the substrates and methods described herein are prepared in less than about three months, two months, one month, three weeks, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 days, 24 hours or less.

**[00165]** In some instances, methods described herein provide for generation of a library of oligonucleotides comprising variant oligonucleotides differing at a plurality of codon sites. In some instances, an oligonucleotide may have 1 site, 2 sites, 3 sites, 4 sites, 5 sites, 6 sites, 7 sites, 8 sites, 9 sites, 10 sites, 11 sites, 12 sites, 13 sites, 14 sites, 15 sites, 16 sites, 17 sites, 18 sites, 19 sites, 20 sites, 30 sites, 40 sites, 50 sites, or more of variant codon sites.

**[00166]** In some instances, the one or more sites of variant codon sites may be adjacent. In some instances, the one or more sites of variant codon sites may not be adjacent and separated by 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more codons.



**[00167]** In some instances, an oligonucleotide may comprise multiple sites of variant codon sites, wherein all the variant codon sites are adjacent to one another, forming a stretch of variant codon sites. In some instances, an oligonucleotide may comprise multiple sites of variant codon sites, wherein none the variant codon sites are adjacent to one another. In some instances, an oligonucleotide may comprise multiple sites of variant codon sites, wherein some the variant codon sites are adjacent to one another, forming a stretch of variant codon sites, and some of the variant codon sites are not adjacent to one another.

**[00168]** Referring to the Figures, **FIG. 12** illustrates an exemplary process workflow for synthesis of nucleic acids (e.g., genes) from shorter oligonucleotides. The workflow is divided generally into phases: (1) de novo synthesis of a single stranded oligonucleotide library, (2) joining oligonucleotides to form larger fragments, (3) error correction, (4) quality control, and (5) shipment. Prior to de novo synthesis, an intended nucleic acid sequence or group of nucleic acid sequences is preselected. For example, a group of genes is preselected for generation.

**[00169]** Once large oligonucleotides for generation are selected, a predetermined library of oligonucleotides is designed for de novo synthesis. Various suitable methods are known for generating high density oligonucleotide arrays. In the workflow example, a device surface layer **1201** is provided. In the example, chemistry of the surface is altered in order to improve the oligonucleotide synthesis process. Areas of low surface energy are generated to repel liquid while areas of high surface energy are generated to attract liquids. The surface itself may be in the form of a planar surface or contain variations in shape, such as protrusions or microwells which increase surface area. In the workflow example, high surface energy molecules selected serve a dual function of supporting DNA chemistry, as disclosed in International Patent Application Publication WO/2015/021080, which is herein incorporated by reference in its entirety.

**[00170]** In situ preparation of oligonucleotide arrays is generated on a solid support and utilizes single nucleotide extension process to extend multiple oligomers in parallel. A material deposition device, such as an oligonucleotide synthesizer, is designed to release reagents in a step wise fashion such that multiple oligonucleotides extend, in parallel, one residue at a time to generate oligomers with a predetermined nucleic acid sequence **1202**. In some instances, oligonucleotides are cleaved from the surface at this stage. Cleavage includes gas cleavage, e.g., with ammonia or methylamine.

**[00171]** The generated oligonucleotide libraries are placed in a reaction chamber. In this exemplary workflow, the reaction chamber (also referred to as “nanoreactor”) is a silicon coated well, containing PCR reagents and lowered onto the oligonucleotide library **1203**. Prior to or after the sealing **1204** of the oligonucleotides, a reagent is added to release the oligonucleotides from the substrate. In the exemplary workflow, the oligonucleotides are released subsequent to sealing of



the nanoreactor **1205**. Once released, fragments of single stranded oligonucleotides hybridize in order to span an entire long range sequence of DNA. Partial hybridization **1205** is possible because each synthesized oligonucleotide is designed to have a small portion overlapping with at least one other oligonucleotide in the pool.

**[00172]** After hybridization, a PCA reaction is commenced. During the polymerase cycles, the oligonucleotides anneal to complementary fragments and gaps are filled in by a polymerase. Each cycle increases the length of various fragments randomly depending on which oligonucleotides find each other. Complementarity amongst the fragments allows for forming a complete large span of double stranded DNA **1206**.

**[00173]** After PCA is complete, the nanoreactor is separated from the device **1207** and positioned for interaction with a device having primers for PCR **1208**. After sealing, the nanoreactor is subject to PCR **1209** and the larger nucleic acids are amplified. After PCR **1210**, the nanochamber is opened **1211**, error correction reagents are added **1212**, the chamber is sealed **1213** and an error correction reaction occurs to remove mismatched base pairs and/or strands with poor complementarity from the double stranded PCR amplification products **1214**. The nanoreactor is opened and separated **1215**. Error corrected product is next subject to additional processing steps, such as PCR and molecular bar coding, and then packaged **1222** for shipment **1223**.

**[00174]** In some instances, quality control measures are taken. After error correction, quality control steps include for example interaction with a wafer having sequencing primers for amplification of the error corrected product **1216**, sealing the wafer to a chamber containing error corrected amplification product **1217**, and performing an additional round of amplification **1218**. The nanoreactor is opened **1219** and the products are pooled **1220** and sequenced **1221**. After an acceptable quality control determination is made, the packaged product **1222** is approved for shipment **1223**.

**[00175]** In some instances, a nucleic acid generated by a workflow such as that in **FIG. 12** is subject to mutagenesis using overlapping primers disclosed herein. In some instances, a library of primers is generated by in situ preparation on a solid support and utilize single nucleotide extension process to extend multiple oligomers in parallel. A deposition device, such as an oligonucleotide synthesizer, is designed to release reagents in a step wise fashion such that multiple oligonucleotides extend, in parallel, one residue at a time to generate oligomers with a predetermined nucleic acid sequence **1202**.

**[00176]** **Computer systems**

**[00177]** Any of the systems described herein, may be operably linked to a computer and may be automated through a computer either locally or remotely. In various instances, the methods and



systems of the disclosure may further comprise software programs on computer systems and use thereof. Accordingly, computerized control for the synchronization of the dispense/vacuum/refill functions such as orchestrating and synchronizing the material deposition device movement, dispense action and vacuum actuation are within the bounds of the disclosure. The computer systems may be programmed to interface between the user specified base sequence and the position of a material deposition device to deliver the correct reagents to specified regions of the substrate.

[00178] The computer system **1300** illustrated in **FIG. 13** may be understood as a logical apparatus that can read instructions from media **1311** and/or a network port **1305**, which can optionally be connected to server **1309** having fixed media **1312**. The system, such as shown in **FIG. 13** can include a CPU **1301**, disk drives **1303**, optional input devices such as keyboard **1315** and/or mouse **1316** and optional monitor **1307**. Data communication may be achieved through the indicated communication medium to a server at a local or a remote location. The communication medium may include any means of transmitting and/or receiving data. For example, the communication medium may be a network connection, a wireless connection or an internet connection. Such a connection may provide for communication over the World Wide Web. It is envisioned that data relating to the present disclosure may be transmitted over such networks or connections for reception and/or review by a party **1322** as illustrated in **FIG. 13**.

[00179] **FIG. 14** is a block diagram illustrating a first example architecture of a computer system **1400** that may be used in connection with example instances of the present disclosure. As depicted in **FIG. 14**, the example computer system may include a processor **1402** for processing instructions. Non-limiting examples of processors include: Intel Xeon™ processor, AMD Opteron™ processor, Samsung 32-bit RISC ARM 1176JZ(F)-S v1.0™ processor, ARM Cortex-A8 Samsung S5PC100™ processor, ARM Cortex-A8 Apple A4™ processor, Marvell PXA 930™ processor, or a functionally-equivalent processor. Multiple threads of execution may be used for parallel processing. In some instances, multiple processors or processors with multiple cores may also be used, whether in a single computer system, in a cluster, or distributed across systems over a network comprising a plurality of computers, cell phones, and/or personal data assistant devices.

[00180] As illustrated in **FIG. 14**, a high speed cache **1404** may be connected to, or incorporated in, the processor **1402** to provide a high speed memory for instructions or data that have been recently, or are frequently, used by processor **1402**. The processor **1402** is connected to a north bridge **1406** by a processor bus **1408**. The north bridge **1406** is connected to random access memory (RAM) **1410** by a memory bus **1412** and manages access to the RAM **1410** by the processor **1402**. The north bridge **1406** is also connected to a south bridge **1414** by a chipset bus



**1416.** The south bridge **1414** is, in turn, connected to a peripheral bus **1418**. The peripheral bus may be, for example, PCI, PCI-X, PCI Express, or other peripheral bus. The north bridge and south bridge are often referred to as a processor chipset and manage data transfer between the processor, RAM, and peripheral components on the peripheral bus **1418**. In some alternative architectures, the functionality of the north bridge may be incorporated into the processor instead of using a separate north bridge chip. In some instances, system **1400** may include an accelerator card **1422** attached to the peripheral bus **1418**. The accelerator may include field programmable gate arrays (FPGAs) or other hardware for accelerating certain processing. For example, an accelerator may be used for adaptive data restructuring or to evaluate algebraic expressions used in extended set processing.

**[00181]** Software and data are stored in external storage **1424** and may be loaded into RAM **1410** and/or cache **1404** for use by the processor. The system **1400** includes an operating system for managing system resources; non-limiting examples of operating systems include: Linux, Windows<sup>TM</sup>, MACOS<sup>TM</sup>, BlackBerry OS<sup>TM</sup>, iOS<sup>TM</sup>, and other functionally-equivalent operating systems, as well as application software running on top of the operating system for managing data storage and optimization in accordance with example instances of the present disclosure. In this example, system **1400** also includes network interface cards (NICs) **1420** and **1421** connected to the peripheral bus for providing network interfaces to external storage, such as Network Attached Storage (NAS) and other computer systems that may be used for distributed parallel processing.

**[00182]** **FIG. 15** is a diagram showing a network **1500** with a plurality of computer systems **1502a**, and **1502b**, a plurality of cell phones and personal data assistants **1502c**, and Network Attached Storage (NAS) **1504a**, and **1504b**. In example instances, systems **1502a**, **1502b**, and **1502c** may manage data storage and optimize data access for data stored in Network Attached Storage (NAS) **1504a** and **1504b**. A mathematical model may be used for the data and be evaluated using distributed parallel processing across computer systems **1502a**, and **1502b**, and cell phone and personal data assistant systems **1502c**. Computer systems **1502a**, and **1502b**, and cell phone and personal data assistant systems **1502c** may also provide parallel processing for adaptive data restructuring of the data stored in Network Attached Storage (NAS) **1504a** and **1504b**. **FIG. 15** illustrates an example only, and a wide variety of other computer architectures and systems may be used in conjunction with the various instances of the present disclosure. For example, a blade server may be used to provide parallel processing. Processor blades may be connected through a back plane to provide parallel processing. Storage may also be connected to the back plane or as Network Attached Storage (NAS) through a separate network interface. In some example instances,



processors may maintain separate memory spaces and transmit data through network interfaces, back plane or other connectors for parallel processing by other processors. In other instances, some or all of the processors may use a shared virtual address memory space.

**[00183]** FIG. 16 is a block diagram of a multiprocessor computer system 1600 using a shared virtual address memory space in accordance with an example instance. The system includes a plurality of processors 1602a-f that may access a shared memory subsystem 1604. The system incorporates a plurality of programmable hardware memory algorithm processors (MAPs) 1606a-f in the memory subsystem 1604. Each MAP 1606a-f may comprise a memory 1608a-f and one or more field programmable gate arrays (FPGAs) 1610a-f. The MAP provides a configurable functional unit and particular algorithms or portions of algorithms may be provided to the FPGAs 1610a-f for processing in close coordination with a respective processor. For example, the MAPs may be used to evaluate algebraic expressions regarding the data model and to perform adaptive data restructuring in example instances. In this example, each MAP is globally accessible by all of the processors for these purposes. In one configuration, each MAP may use Direct Memory Access (DMA) to access an associated memory 1608a-f, allowing it to execute tasks independently of, and asynchronously from the respective microprocessor 1602a-f. In this configuration, a MAP may feed results directly to another MAP for pipelining and parallel execution of algorithms.

**[00184]** The above computer architectures and systems are examples only, and a wide variety of other computer, cell phone, and personal data assistant architectures and systems may be used in connection with example instances, including systems using any combination of general processors, co-processors, FPGAs and other programmable logic devices, system on chips (SOCs), application specific integrated circuits (ASICs), and other processing and logic elements. In some instances, all or part of the computer system may be implemented in software or hardware. Any variety of data storage media may be used in connection with example instances, including random access memory, hard drives, flash memory, tape drives, disk arrays, Network Attached Storage (NAS) and other local or distributed data storage devices and systems.

**[00185]** In example instances, the computer system may be implemented using software modules executing on any of the above or other computer architectures and systems. In other instances, the functions of the system may be implemented partially or completely in firmware, programmable logic devices such as field programmable gate arrays (FPGAs) as referenced in FIG. 16, system on chips (SOCs), application specific integrated circuits (ASICs), or other processing and logic elements. For example, the Set Processor and Optimizer may be implemented with hardware acceleration through the use of a hardware accelerator card, such as accelerator card 1422 illustrated in FIG. 14.



[00186] The following examples are set forth to illustrate more clearly the principle and practice of embodiments disclosed herein to those skilled in the art and are not to be construed as limiting the scope of any claimed embodiments. Unless otherwise stated, all parts and percentages are on a weight basis.

### EXAMPLES

[00187] The following examples are given for the purpose of illustrating various embodiments of the disclosure and are not meant to limit the present disclosure in any fashion. The present examples, along with the methods described herein are presently representative of preferred embodiments, are exemplary, and are not intended as limitations on the scope of the disclosure. Changes therein and other uses which are encompassed within the spirit of the disclosure as defined by the scope of the claims will occur to those skilled in the art.

[00188] **Example 1: Functionalization of a device surface**

[00189] A device was functionalized to support the attachment and synthesis of a library of oligonucleotides. The device surface was first wet cleaned using a piranha solution comprising 90% H<sub>2</sub>SO<sub>4</sub> and 10% H<sub>2</sub>O<sub>2</sub> for 20 minutes. The device was rinsed in several beakers with DI water, held under a DI water gooseneck faucet for 5 min, and dried with N<sub>2</sub>. The device was subsequently soaked in NH<sub>4</sub>OH (1:100; 3 mL:300 mL) for 5 min, rinsed with DI water using a handgun, soaked in three successive beakers with DI water for 1 min each, and then rinsed again with DI water using the handgun. The device was then plasma cleaned by exposing the device surface to O<sub>2</sub>. A SAMCO PC-300 instrument was used to plasma etch O<sub>2</sub> at 250 watts for 1 min in downstream mode.

[00190] The cleaned device surface was actively functionalized with a solution comprising N-(3-triethoxysilylpropyl)-4-hydroxybutyramide using a YES-1224P vapor deposition oven system with the following parameters: 0.5 to 1 torr, 60 min, 70 °C, 135 °C vaporizer. The device surface was resist coated using a Brewer Science 200X spin coater. SPR<sup>TM</sup> 3612 photoresist was spin coated on the device at 2500 rpm for 40 sec. The device was pre-baked for 30 min at 90 °C on a Brewer hot plate. The device was subjected to photolithography using a Karl Suss MA6 mask aligner instrument. The device was exposed for 2.2 sec and developed for 1 min in MSF 26A. Remaining developer was rinsed with the handgun and the device soaked in water for 5 min. The device was baked for 30 min at 100 °C in the oven, followed by visual inspection for lithography defects using a Nikon L200. A descum process was used to remove residual resist using the SAMCO PC-300 instrument to O<sub>2</sub> plasma etch at 250 watts for 1 min.



**[00191]** The device surface was passively functionalized with a 100  $\mu$ L solution of perfluorooctyltrichlorosilane mixed with 10  $\mu$ L light mineral oil. The device was placed in a chamber, pumped for 10 min, and then the valve was closed to the pump and left to stand for 10 min. The chamber was vented to air. The device was resist stripped by performing two soaks for 5 min in 500 mL NMP at 70 °C with ultrasonication at maximum power (9 on Crest system). The device was then soaked for 5 min in 500 mL isopropanol at room temperature with ultrasonication at maximum power. The device was dipped in 300 mL of 200 proof ethanol and blown dry with N<sub>2</sub>. The functionalized surface was activated to serve as a support for oligonucleotide synthesis.

**[00192] Example 2: Synthesis of a 50-mer sequence on an oligonucleotide synthesis device**

**[00193]** A two dimensional oligonucleotide synthesis device was assembled into a flowcell, which was connected to a flowcell (Applied Biosystems (ABI394 DNA Synthesizer"). The two-dimensional oligonucleotide synthesis device was uniformly functionalized with N-(3-TRIETHOXYSILYL)PROPYL)-4-HYDROXYBUTYRAMIDE (Gelest) was used to synthesize an exemplary oligonucleotide of 50 bp ("50-mer oligonucleotide") using oligonucleotide synthesis methods described herein.

**[00194]** The sequence of the 50-mer was as described in SEQ ID NO.: 20.

5'AGACAATCAACCATTTGGGGTGGACAGCCTTGACCTCTAGACTTCGGCAT##TTTTTTT TTT3' (SEQ ID NO.: 20), where # denotes Thymidine-succinyl hexamide CED phosphoramidite (CLP-2244 from ChemGenes), which is a cleavable linker enabling the release of oligos from the surface during deprotection.

**[00195]** The synthesis was done using standard DNA synthesis chemistry (coupling, capping, oxidation, and deblocking) according to the protocol in **Table 5** and an ABI synthesizer.

**Table 5: Synthesis Protocol**

<b>Table 5</b>		
<b>General DNA Synthesis Process Name</b>	<b>Process Step</b>	<b>Time (sec)</b>
<b>WASH</b> (Acetonitrile Wash Flow)	Acetonitrile System Flush	4
	Acetonitrile to Flowcell	23
	N <sub>2</sub> System Flush	4
	Acetonitrile System Flush	4
<b>DNA BASE ADDITION</b> (Phosphoramidite + Activator Flow)	Activator Manifold Flush	2
	Activator to Flowcell	6
	Activator + Phosphoramidite to Flowcell	6
	Activator to Flowcell	0.5
	Activator + Phosphoramidite to Flowcell	5

<b>Table 5</b>		
<b>General DNA Synthesis Process Name</b>	<b>Process Step</b>	<b>Time (sec)</b>
	Activator to Flowcell	0.5
	Activator + Phosphoramidite to Flowcell	5
	Activator to Flowcell	0.5
	Activator + Phosphoramidite to Flowcell	5
	Incubate for 25sec	25
<b>WASH</b> (Acetonitrile Wash Flow)	Acetonitrile System Flush	4
	Acetonitrile to Flowcell	15
	N2 System Flush	4
	Acetonitrile System Flush	4
<b>DNA BASE ADDITION</b> (Phosphoramidite + Activator Flow)	Activator Manifold Flush	2
	Activator to Flowcell	5
	Activator + Phosphoramidite to Flowcell	18
	Incubate for 25sec	25
<b>WASH</b> (Acetonitrile Wash Flow)	Acetonitrile System Flush	4
	Acetonitrile to Flowcell	15
	N2 System Flush	4
	Acetonitrile System Flush	4
<b>CAPPING</b> (CapA+B, 1:1, Flow)	CapA+B to Flowcell	15
<b>WASH</b> (Acetonitrile Wash Flow)	Acetonitrile System Flush	4
	Acetonitrile to Flowcell	15
	Acetonitrile System Flush	4
<b>OXIDATION</b> (Oxidizer Flow)	Oxidizer to Flowcell	18
<b>WASH</b> (Acetonitrile Wash Flow)	Acetonitrile System Flush	4
	N2 System Flush	4
	Acetonitrile System Flush	4
	Acetonitrile to Flowcell	15
	Acetonitrile System Flush	4
	Acetonitrile to Flowcell	15
	N2 System Flush	4
	Acetonitrile System Flush	4
	Acetonitrile to Flowcell	23
	N2 System Flush	4
	Acetonitrile System Flush	4
<b>DEBLOCKING</b> (Deblock Flow)	Deblock to Flowcell	36
<b>WASH</b> (Acetonitrile Wash Flow)	Acetonitrile System Flush	4
	N2 System Flush	4
	Acetonitrile System Flush	4
	Acetonitrile to Flowcell	18



<b>Table 5</b>		
<b>General DNA Synthesis Process Name</b>	<b>Process Step</b>	<b>Time (sec)</b>
	N2 System Flush	4.13
	Acetonitrile System Flush	4.13
	Acetonitrile to Flowcell	15

**[00196]** The phosphoramidite/activator combination was delivered similar to the delivery of bulk reagents through the flowcell. No drying steps were performed as the environment stays "wet" with reagent the entire time.

**[00197]** The flow restrictor was removed from the ABI 394 synthesizer to enable faster flow. Without flow restrictor, flow rates for amidites (0.1M in ACN), Activator, (0.25M Benzoylthiotetrazole ("BTT"; 30-3070-xx from GlenResearch) in ACN), and Ox (0.02M I2 in 20% pyridine, 10% water, and 70% THF) were roughly ~100uL/sec, for acetonitrile ("ACN") and capping reagents (1:1 mix of CapA and CapB, wherein CapA is acetic anhydride in THF/Pyridine and CapB is 16% 1-methylimidazole in THF), roughly ~200uL/sec, and for Deblock (3% dichloroacetic acid in toluene), roughly ~300uL/sec (compared to ~50uL/sec for all reagents with flow restrictor). The time to completely push out Oxidizer was observed, the timing for chemical flow times was adjusted accordingly and an extra ACN wash was introduced between different chemicals. After oligonucleotide synthesis, the chip was deprotected in gaseous ammonia overnight at 75 psi. Five drops of water were applied to the surface to recover oligonucleotides. The recovered oligonucleotides were then analyzed on a BioAnalyzer small RNA chip (data not shown).

**[00198] Example 3: Synthesis of a 100-mer sequence on an oligonucleotide synthesis device**

**[00199]** The same process as described in Example 2 for the synthesis of the 50-mer sequence was used for the synthesis of a 100-mer oligonucleotide ("100-mer oligonucleotide"; 5' CGGGATCCTTATCGTCATCGTCGTACAGATCCCGACCCATTTGCTGTCCACCAGTCATGCTAGCCATACCATGATGATGATGATGATGAGAACCCCGCAT##TTTTTTTTTTT3', where # denotes Thymidine-succinyl hexamide CED phosphoramidite (CLP-2244 from ChemGenes); SEQ ID NO.: 21) on two different silicon chips, the first one uniformly functionalized with N-(3-TRIETHOXY-SILYL-PROPYL)-4-HYDROXYBUTYRAMIDE and the second one functionalized with 5/95 mix of 11-acetoxyundecyltriethoxysilane and n-decyltriethoxysilane, and the oligonucleotides extracted from the surface were analyzed on a BioAnalyzer instrument (data not shown).

**[00200]** All ten samples from the two chips were further PCR amplified using a forward (5'ATGCGGGGTTCTCATCATC3'; SEQ ID NO.: 22) and a reverse

(5'CGGGATCCTTATCGTCATCG3'; SEQ ID NO.: 23) primer in a 50uL PCR mix (25uL NEB Q5 mastermix, 2.5uL 10uM Forward primer, 2.5uL 10uM Reverse primer, 1uL oligonucleotide extracted from the surface, and water up to 50uL) using the following thermalcycling program:

98 °C, 30 sec

98 °C, 10 sec; 63 °C, 10 sec; 72 °C, 10 sec; repeat 12 cycles

72 °C, 2min

**[00201]** The PCR products were also run on a BioAnalyzer (data not shown), demonstrating sharp peaks at the 100-mer position. Next, the PCR amplified samples were cloned, and Sanger sequenced. **Table 6** summarizes the results from the Sanger sequencing for samples taken from spots 1-5 from chip 1 and for samples taken from spots 6-10 from chip 2.

**Table 6: Sequencing Results**

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

**[00202]** Thus, the high quality and uniformity of the synthesized oligonucleotides were repeated on two chips with different surface chemistries. Overall, 89%, corresponding to 233 out of 262 of the 100-mers that were sequenced were perfect sequences with no errors.

**[00203]** Finally, **Table 7** summarizes error characteristics for the sequences obtained from the oligonucleotides samples from spots 1-10.

**Table 7: Error Characteristics**

Sample ID/Spot no.	OSA_0 046/1	OSA_0 047/2	OSA_0 048/3	OSA_0 049/4	OSA_0 050/5	OSA_0 051/6	OSA_0 052/7	OSA_0 053/8	OSA_0 054/9	OSA_00 55/10
Total Sequences	32	32	32	32	32	32	32	32	32	32
Sequencing Quality	25 of 28	27 of 27	26 of 30	21 of 23	25 of 26	29 of 30	27 of 31	29 of 31	28 of 29	25 of 28



Oligo Quality	23 of 25	25 of 27	22 of 26	18 of 21	24 of 25	25 of 29	22 of 27	28 of 29	26 of 28	20 of 25
ROI Match Count	2500	2698	2561	2122	2499	2666	2625	2899	2798	2348
ROI Mutation	2	2	1	3	1	0	2	1	2	1
ROI Multi Base Deletion	0	0	0	0	0	0	0	0	0	0
ROI Small Insertion	1	0	0	0	0	0	0	0	0	0
ROI Single Base Deletion	0	0	0	0	0	0	0	0	0	0
Large Deletion Count	0	0	1	0	0	1	1	0	0	0
Mutation: G>A	2	2	1	2	1	0	2	1	2	1
Mutation: T>C	0	0	0	1	0	0	0	0	0	0
ROI Error Count	3	2	2	3	1	1	3	1	2	1
ROI Error Rate	Err: ~1 in 834	Err: ~1 in 1350	Err: ~1 in 1282	Err: ~1 in 708	Err: ~1 in 2500	Err: ~1 in 2667	Err: ~1 in 876	Err: ~1 in 2900	Err: ~1 in 1400	Err: ~1 in 2349
ROI Minus Primer Error Rate	MP Err: ~1 in 763	MP Err: ~1 in 824	MP Err: ~1 in 780	MP Err: ~1 in 429	MP Err: ~1 in 1525	MP Err: ~1 in 1615	MP Err: ~1 in 531	MP Err: ~1 in 1769	MP Err: ~1 in 854	MP Err: ~1 in 1451

**[00204] Example 4: Generation of a nucleic acid library by single-site, single position mutagenesis**

**[00205]** Oligonucleotide primers were de novo synthesized for use in a series of PCR reactions to generate a library of oligonucleotide variants of a template nucleic acid, see **FIGS. 2A-2D**. Four types of primers were generated in **FIG. 2A**: an outer 5' primer **215**, an outer 3' primer **230**, an inner 5' primer **225**, and an inner 3' primer **220**. The inner 5' primer / first oligonucleotide **220** and an inner 3' primer / second oligonucleotide **225** were generated using an oligonucleotide synthesis method as generally outlined in **Table 6**. The inner 5' primer / first oligonucleotide **220** represents a set of up to 19 primers of predetermined sequence, where each primer in the set differs from another at a single codon, in a single site of the sequence.

**[00206]** Oligonucleotide synthesis was performed on a device having at least two clusters, each cluster having 121 individually addressable loci.

[00207] The inner 5' primer **225** and the inner 3' primer **220** were synthesized in separate clusters. The inner 5' primer **225** was replicated 121 times, extending on 121 loci within a single cluster. For inner 3' primer **220**, each of the 19 primers of variant sequences were each extended on 6 different loci, resulting in the extension of 114 oligonucleotides on 114 different loci.

[00208] Synthesized oligonucleotides were cleaved from the surface of the device and transferred to a plastic vial. A first PCR reaction was performed, using fragments of the long nucleic acid sequence **235, 240** to amplify the template nucleic acid, as illustrated in **FIG. 2B**. A second PCR reaction was performed using primer combination and the products of the first PCR reaction as a template, as illustrated in **FIGS. 2C-2D**. Analysis of the second PCR products was conducted on a BioAnalyzer, as shown in the trace of **FIG. 17**.

[00209] **Example 5: Generation of a nucleic acid library comprising 96 different sets of single position variants**

[00210] Four sets of primers, as generally shown in **FIG. 2A** and addressed in Example 2, were generated using de novo oligonucleotide synthesis. For the inner 5' primer **220**, 96 different sets of primers were generated, each set of primers targeting a different single codon positioned within a single site of the template nucleic acid. For each set of primers, 19 different variants were generated, each variant comprising a codon encoding for a different amino acid at the single site. Two rounds of PCR were performed using the generated primers, as generally shown in **FIGS. 2A-2D** and described in Example 2. The 96 sets of amplification products were visualized in an electropherogram (**FIG. 18**), which was used to calculate a 100% amplification success rate.

[00211] **Example 6: Generation of a nucleic acid library comprising 500 different sets of single position variants**

[00212] Four sets of primers, as generally shown in **FIG. 2A** and addressed in Example 2, were generated using de novo oligonucleotide synthesis. For the inner 5' primer **220**, 500 different sets of primers were generated, each set of primers targeting a different single codon positioned within a single site of the template nucleic acid. For each set of primers, 19 different variants were generated, each variant comprising a codon encoding for a different amino acid at the single site. Two rounds of PCR were performed using the generated primers, as generally shown in **FIG. 2A** and described in Example 2. Electropherograms display each of the 500 sets of PCR products having a population of nucleic acids with 19 variants at a different single site (data not shown). A comprehensive sequencing analysis of the library showed a greater than 99% success rate across preselected codon mutations (sequence trace and analysis data not shown).



**[00213] Example 7: Single-site mutagenesis primers for 1 position**

**[00214]** An example of codon variation design is provided in **Table 8** for Yellow Fluorescent Protein. In this case, a single codon from a 50-mer of the sequence is varied 19 times. Variant nucleic acid sequence is indicated by bold letters. The wild type primer sequence is:

ATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT (SEQ ID NO.: 1).

In this case, the wild type codon encodes for valine, indicated by underline in SEQ ID NO.: 1.

Therefore the 19 variants below exclude a codon encoding for valine. In an alternative example, if all triplets are to be considered, then all 60 variants would be generated, including an alternative sequence for the wild type codon.

**Table 8. Variant Sequences**

SEQ ID NO.	Variant sequence	Variant codon
2	atg <b>TT</b> TAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	F
3	atg <b>TTA</b> AGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	L
4	atg <b>ATT</b> AGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	I
5	atg <b>TCT</b> AGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	S
6	atg <b>CCT</b> AGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	P
7	atg <b>ACT</b> AGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	T
8	atg <b>GCT</b> AGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	A
9	atg <b>TAT</b> AGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	Y
10	atg <b>CAT</b> AGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	H
11	atg <b>CAA</b> AGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	Q
12	atg <b>AAT</b> AGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	N
13	atg <b>AAA</b> AGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	K
14	atg <b>GAT</b> AGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	D
15	atg <b>GAA</b> AGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	E
16	atg <b>TGT</b> AGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	C
17	atg <b>TGG</b> AGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	W
18	atg <b>CGT</b> AGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	R
19	atg <b>GGT</b> AGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	G

**[00215] Example 8: Single site, dual position nucleic acid variants**

[00216] De novo oligonucleotide synthesis was performed under conditions similar to those described in Example 2. A single cluster on a device was generated which contained synthesized predetermined variants of a nucleic acid for 2 consecutive codon positions at a single site, each position being a codon encoding for an amino acid. In this arrangement, 19 variants/per position were generated for 2 positions with 3 replicates of each nucleic acid, resulting in 114 nucleic acids synthesized.

[00217] **Example 9: Multiple site, dual position nucleic acid variants**

[00218] De novo oligonucleotide synthesis was performed under conditions similar to those described in Example 2. A single cluster on a device was generated which contained synthesized predetermined variants of a nucleic acid for 2 non-consecutive codon positions, each position being a codon encoding for an amino acid. In this arrangement, 19 variants/per position were generated for 2 positions.

[00219] **Example 10: Single stretch, triple position nucleic acid variants**

[00220] De novo oligonucleotide synthesis was performed under conditions similar to those described in Example 2. A single cluster on a device was generated which contained synthesized predetermined variants of a reference nucleic acid for 3 consecutive codon positions. In the 3 consecutive codon position arrangement, 19 variants/ per position were generated for 3 positions with 2 replicates of each nucleic acid, and resulted in 114 nucleic acids synthesized.

[00221] **Example 11: Multiple site, triple position nucleic acid variants**

[00222] De novo oligonucleotide synthesis was performed under conditions similar to those described in Example 2. A single cluster on a device was generated which contains synthesized predetermined variants of a reference nucleic acid for at least 3 non-consecutive codon positions. Within a predetermined region, the location of codons encoding for 3 histidine residues were varied.

[00223] **Example 12: Multiple site, multiple position nucleic acid variants**

[00224] De novo oligonucleotide synthesis was performed under conditions similar to those described in Example 2. A single cluster on a device was generated which contained synthesized predetermined variants of a reference nucleic acid for 1 or more codon positions in 1 or more stretches. Five positions were varied in the library. The first position encoded codons for a resultant 50/50 K/R ratio in the expressed protein; the second position encoded codons for a resultant 50/25/25 V/L/S ratio in the expressed protein, the third position encoded codons for a resultant a 50/25/25 Y/R/D ratio in the expressed protein, the fourth position encoded codons for a resultant an equal ratio for all amino acids in the expressed protein, and the fifth position encoded codons for a resultant a 75/25 G/P ratio in the expressed protein.



**[00225] Example 13: Modular plasmid components for expressing diverse peptides**

**[00226]** A nucleic acid library is generated as in Examples 4-6 and 8-12, encoding for codon variation at a single site or multiple sites for each of separate regions that make up portions of an expression construct cassette, as depicted in **FIG. 11**. To generate a two construct expressing cassette, variant oligonucleotides were synthesized encoding at least a portion of a variant sequence of a first promoter **1110**, first open reading frame **1120**, first terminator **1130**, second promoter **1140**, second open reading frame **1150**, or second terminator sequence **1160**. After rounds of amplification, as described in previous examples, a library of 1,024 expression constructs is generated.

**[00227] Example 14: Multiple site, single position variants**

**[00228]** A nucleic acid library is generated as in Examples 4-6 and 8-12, encoding for codon variation at a single site or multiple sites in a region encoding for at least a portion of nucleic acid. A library of oligonucleotide variants is generated, wherein the library consists of multiple site, single position variants. See, for example, **FIG. 6B**.

**[00229] Example 15: Variant library synthesis**

**[00230]** De novo oligonucleotide synthesis is performed under conditions similar to those described in Example 2. At least 30,000 non-identical oligonucleotides are de novo synthesized, wherein each of the non-identical oligonucleotides encodes for a different codon variant of an amino acid sequence. The synthesized at least 30,000 non-identical oligonucleotides have an aggregate error rate of less than 1 in 1:000 bases compared to predetermined sequences for each of the at least 30,000 non-identical oligonucleotides. The library is used for PCR mutagenesis of a long nucleic acid and at least 30,000 non-identical variant nucleic acids are formed.

**[00231] Example 16: Cluster-based Variant library synthesis**

**[00232]** De novo oligonucleotide synthesis is performed under conditions similar to those described in Example 2. A single cluster on a device is generated which contained synthesized predetermined variants of a reference oligonucleotide for 2 codon positions. In the 2 consecutive codon position arrangement, 19 variants/ per position were generated for the 2 positions with 2 replicates of each oligonucleotide, and resulted in 38 oligonucleotides synthesized. Each variant sequence is 40 bases in length. In the same cluster, additional non-variant oligonucleotide sequences are generated, where the additional non-variant oligonucleotides and the variant oligonucleotides collective encode for 38 variants of the coding sequence of a gene. Each of the oligonucleotides has at least one region reverse complementary to another of the oligonucleotides. The oligonucleotides in the cluster are released by gaseous ammonia cleavage. A pin comprising water contacts the cluster, picks up the oligonucleotides, and moves the oligonucleotides to a small

vial. The vial also contains DNA polymerase reagents for a polymerase cycling assembly (PCA) reaction. The oligonucleotides anneal, gaps are filled in by an extension reaction, and resultant double-stranded DNA molecules are formed, forming a variant nucleic acid library. The variant nucleic acid library is, optionally, subjected to restriction enzyme is then ligated into expression vectors.

**[00233] Example 17: Generation of a variant nucleic acid library**

**[00234]** CAR variant libraries are generated, as described in Examples 8-16, to have variance in an antigen recognition domain, a hinge domain, a transmembrane domain, or an intracellular domain compared to a reference sequence. Alternatively, CAR variant libraries are also generated to have variance in all domains of a CAR. The CAR variant libraries are screened against a known tumor antigen. Variant sequences that result in increased T cell activation and/or T cell binding are identified. The tumor antigen may be expressed on a primary cultured cell, non-primary cultured cell, or in a non-cellular setting, such as on a plate, beads, slurry, or column.

**[00235] Example 18: Expression and screening of a variant CAR library**

**[00236]** The library of variant CAR genes described in Example 17 are transferred into mammalian cells to generate a library of cell populations, each cell population expressing a different CAR variant protein. The protein library is screened for CAR complexes with improved affinity (measure of the strength of interaction between an epitope and an antibody's antigen binding site) for a tumor antigen, such as HER2 or NY-ESO-1. Additional functional considerations, such as variant gene expression, avidity (measure of the overall strength of an antibody-antigen complex), stability, and target specificity are also assessed.

**[00237] Example 19: Manufacturing and delivery of engineered T cells**

**[00238]** T cells are harvested from a subject diagnosed with cancer and are genetically engineered with a CAR selected after performing analysis in Example 18. After a brief period of *in vitro* expansion and passing of product-specific release criteria, the engineered T-cells are administered to the same subject.

**[00239] Example 20: Variant peptide library**

**[00240]** A nucleic acid library is generated as in Examples 4-6 and 8-12, encoding for codon variation at a single site or multiple sites for common tumor antigens. After rounds of amplification, as described in previous examples, libraries of 1,024 expression constructs for each of the antigen variants are generated. The libraries are transferred into mammalian cells to generate a library of cell populations and are screened for improved binding affinity and target specificity.



**[00241] Example 21: Variant peptide library and variant library of CARs**

**[00242]** A nucleic acid is generated as in Examples 4-6 and 8-12, encoding for codon variation at a single site or multiple sites for an antigen recognition domain of a CAR. A nucleic acid library is generated as in Example 20, encoding for codon variation at a single site or multiple sites for common tumor antigens. After rounds of amplification, as described in previous examples, libraries of 1,024 expression constructs for each of the CAR variants and the antigen variants are generated. The libraries are transferred into mammalian cells to generate a library of cell populations and are screened for improved binding affinity and target specificity.

**[00243] Example 22: Variant CARs**

**[00244]** A nucleic acid library is generated as in Examples 4-6 and 8-12 to generate nucleic acids encoding for variant CARs. Variants are generated at three sites to generate a library comprising about  $10^3$  variant CARs.

**[00245]** The variant CARs are screened *in vitro* against tumor antigens for specificity of the variant CARs to the tumor antigen. Variant CARs that are highly specific for the tumor antigen are then further variegated to generate a second library. The second library comprises variation in residues located in the hinge domain, the transmembrane domain, or the intracellular domain of the CAR. The second library is expressed in cells and screened *in vitro* for a second improvement, for example, avidity, stability, affinity, or expression.

**[00246]** Select CAR genes or gene fragment variants having desired features after screening products from the first and/or second variant libraries are selected for development of a potential therapeutic.

**[00247]** While preferred embodiments of the present disclosure have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the disclosure. It should be understood that various alternatives to the embodiments of the disclosure described herein may be employed in practicing the disclosure. It is intended that the following claims define the scope of the disclosure and that methods and structures within the scope of these claims and their equivalents be covered thereby.

## CLAIMS

### WHAT IS CLAIMED IS:

1. A nucleic acid library, the nucleic acid library comprising at least 10,000 nucleic acids, wherein each nucleic acid encodes for a preselected variant of a reference sequence that encodes for a chimeric antigen receptor or a functional domain thereof.
2. The nucleic acid library of claim 1, wherein the functional domain of the chimeric antigen receptor comprises an antigen recognition domain, a hinge domain, a transmembrane domain, or an intracellular domain.
3. The nucleic acid library of claim 1, wherein the library comprises at least 1,000,000 nucleic acids.
4. The nucleic acid library of claim 1, wherein each nucleic acid is 100 bases to 2000 bases in length.
5. The nucleic acid library of claim 1, wherein each nucleic acid is attached to a vector sequence.
6. The nucleic acid library of claim 5, wherein the vector sequence is a viral vector sequence.
7. A nucleic acid library, the nucleic acid library comprising a plurality of nucleic acids, wherein each nucleic acid encodes for a preselected variant of a reference sequence that encodes for a chimeric antigen receptor or a functional domain thereof, wherein the plurality of nucleic acids comprises all variations for at least two positions in the reference sequence, and wherein the at least two positions encode sequences from 5 to 20 different codons.
8. The nucleic acid library of claim 7, wherein the functional domain of the chimeric antigen receptor comprises an antigen recognition domain, a hinge domain, a transmembrane domain, or an intracellular domain.
9. The nucleic acid library of claim 7, wherein the plurality of nucleic acids comprises all variations for 2, 3, 4, or 5 positions in the reference sequence.
10. The nucleic acid library of claim 7, wherein the at least two positions encode sequences from 10 to 20 different codons.
11. The nucleic acid library of claim 7, wherein the at least two positions encode sequences for about 10 different codons.
12. The nucleic acid library of claim 7, wherein each nucleic acid is 100 bases to 2000 bases in length.



13. An oligonucleotide library, the oligonucleotide library comprising at least 10,000 oligonucleotides, wherein each oligonucleotide encodes for a preselected variant of a reference sequence that encodes for a region of a chimeric antigen receptor, wherein the region comprises a portion of an antigen recognition domain, a hinge domain, a transmembrane domain, or an intracellular domain.

14. The oligonucleotide library of claim 13, wherein the library comprises at least 1,000,000 oligonucleotides.

15. The oligonucleotide library of claim 13, wherein each oligonucleotide is 12 to 500 bases in length.

16. The oligonucleotide library of claim 13, wherein each oligonucleotide is attached to a vector sequence.

17. An oligonucleotide library, the oligonucleotide library comprising a plurality of oligonucleotides, wherein each oligonucleotide encodes for a preselected variant of a reference sequence that encodes for a region of a chimeric antigen receptor, wherein the region comprises a portion of an antigen recognition domain, a hinge domain, a transmembrane domain, or an intracellular domain, wherein each oligonucleotide is at least 12 bases in length, wherein the plurality of oligonucleotides comprises all variations for at least two positions in the reference sequence, and wherein the at least two positions encode sequences from 5 to 20 different codons.

18. The oligonucleotide library of claim 17, wherein the plurality of oligonucleotides comprises all variations for 2, 3, 4, or 5 positions.

19. The oligonucleotide library of claim 17, wherein the at least two positions encode sequences from 10 to 20 different codons.

20. The oligonucleotide library of claim 17, wherein the at least two positions encode sequences for about 10 different codons.

21. The oligonucleotide library of claim 17, wherein each oligonucleotide is 12 to 500 bases in length.

22. A nucleic acid library comprising at least 400 nucleic acids, wherein a first plurality of the at least 400 nucleic acids encodes for a variant of a reference sequence that encodes for an antigen recognition domain, and wherein a second plurality of the at least 400 nucleic acids encodes for a variant of a reference sequence that encodes for a hinge domain, a transmembrane domain, or an intracellular domain of a chimeric antigen receptor (CAR).

23. The nucleic acid library of claim 22, wherein each nucleic acid is at least 100 bases in length.

24. The nucleic acid library of claim 22, wherein each nucleic acid is 100 to 2000 bases in length.

25. The nucleic acid library of claim 22, wherein the first plurality of the at least 400 nucleic acids comprises all variations for at least two positions in the reference sequence.

26. The nucleic acid library of claim 22, wherein the first plurality of the at least 400 nucleic acids comprises all variations for 2, 3, 4, or 5 positions in the reference sequence.

27. The nucleic acid library of claim 22, wherein the second plurality of the at least 400 nucleic acids comprises all variations for at least two positions in the reference sequence.

28. The nucleic acid library of claim 22, wherein the second plurality of the at least 400 nucleic acids comprises all variations for 2, 3, 4, or 5 positions in the reference sequence.

29. The nucleic acid library of claim 25, wherein the at least two positions encode sequences from 5 to 20 different codons.

30. The nucleic acid library of claim 25, wherein the at least two positions encode sequences from 10 to 20 different codons.

31. The nucleic acid library of claim 25, wherein the at least two positions encode sequences for about 10 different codons.

32. A method of synthesizing a nucleic acid library, comprising:

a. providing a first set of preselected oligonucleotide sequences encoding for at least 400 sequences of a chimeric antigen receptor gene or gene fragment, wherein each sequence comprises at least one variation of at least two preselected codons for an amino acid residue in an antigen recognition domain;

b. synthesizing the first set of preselected oligonucleotide sequences; and

c. screening a first activity for proteins encoded by the first set of oligonucleotide sequences, wherein the first activity is specificity, avidity, affinity, stability, or expression.

33. The method of claim 32, wherein the antigen is a cancer antigen.

34. The method of claim 33, wherein the cancer antigen is MAGE A3, MAGE A12, MAGE A2, MAGE A6, NY-ESO-1, or CEA.

35. The method of claim 32, wherein each sequence comprises up to 100 variations at preselected codons for amino acid residues in the antigen recognition domain.

36. The method of claim 32, wherein each sequence comprises up to 30 variations at preselected codons for amino acid residues in the antigen recognition domain.

37. The method of claim 32, further comprising

a. providing a second set of preselected oligonucleotide sequences encoding for at least one sequence of a chimeric antigen receptor gene or gene fragment, where each sequence



comprises at least one variation at a preselected codon for an amino acid residue in the chimeric antigen receptor in a hinge domain, a transmembrane domain, or an intracellular domain;

- b. synthesizing the second set of preselected oligonucleotide sequences; and
- c. screening a second activity for proteins encoded by the second set of oligonucleotide sequences, wherein the second activity is specificity, avidity, affinity, stability, or expression, and wherein the second activity is different from the first activity.



FIG. 1D

FIG. 1C

FIG. 1B

FIG. 1A



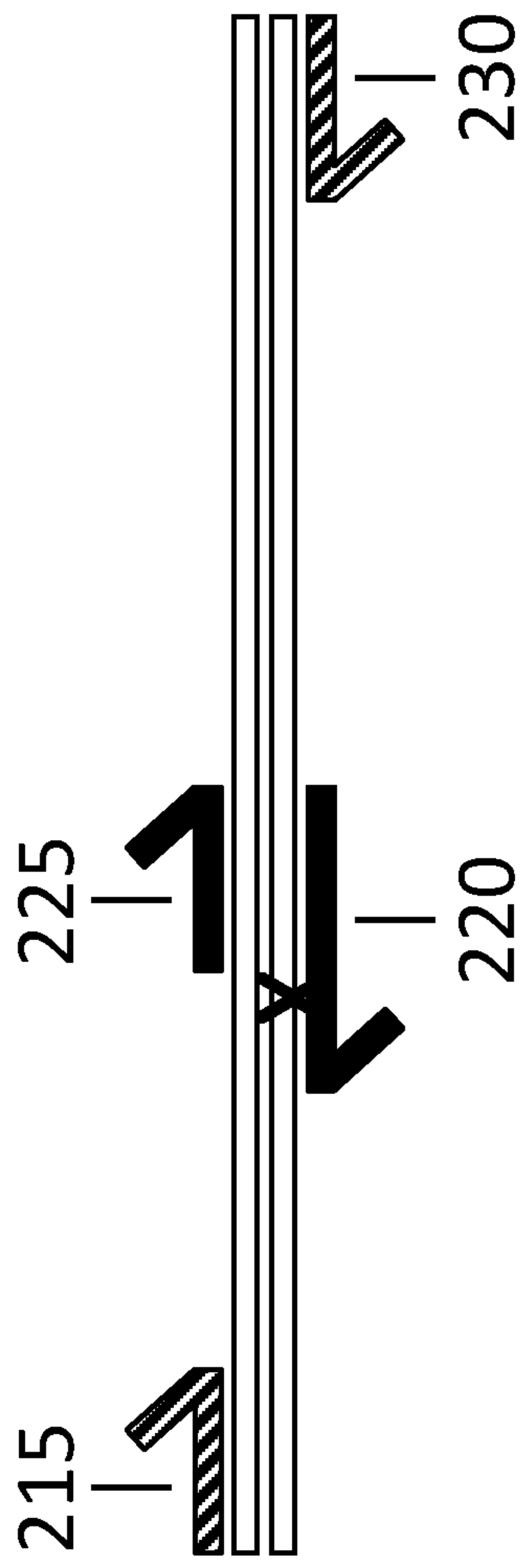


FIG. 2A

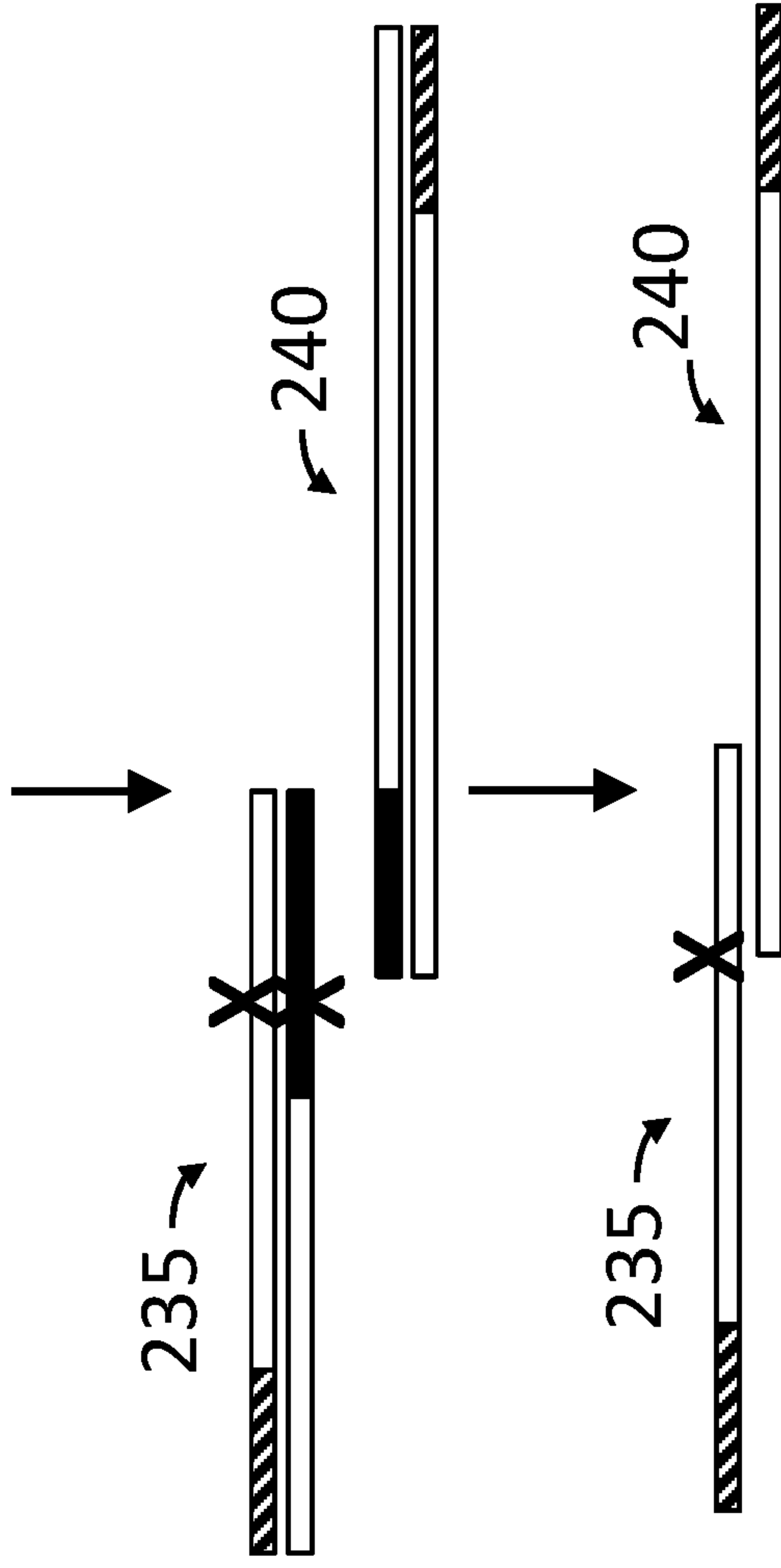


FIG. 2B

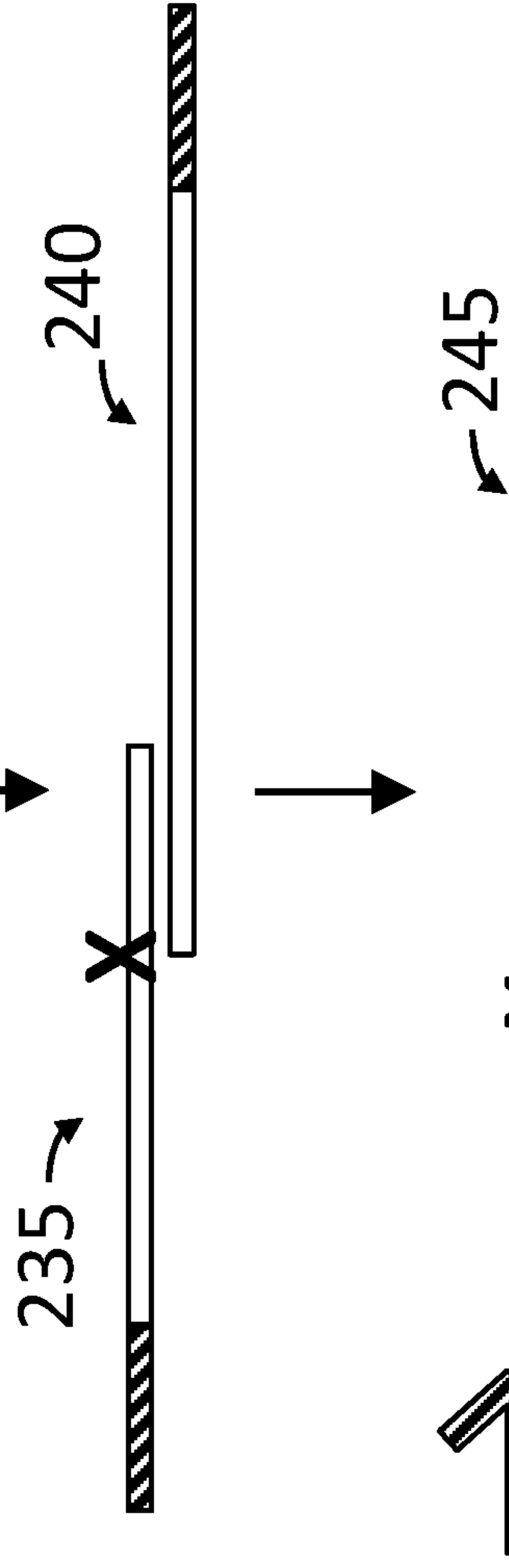


FIG. 2C

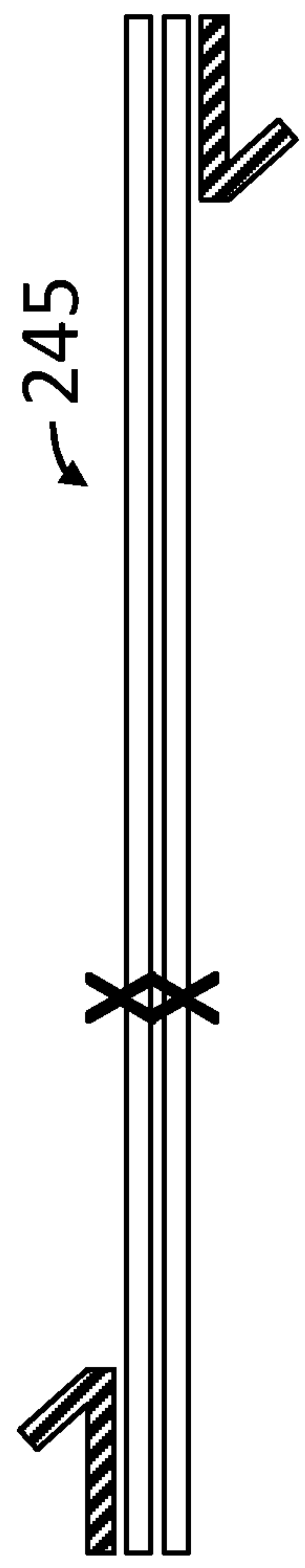


FIG. 2D

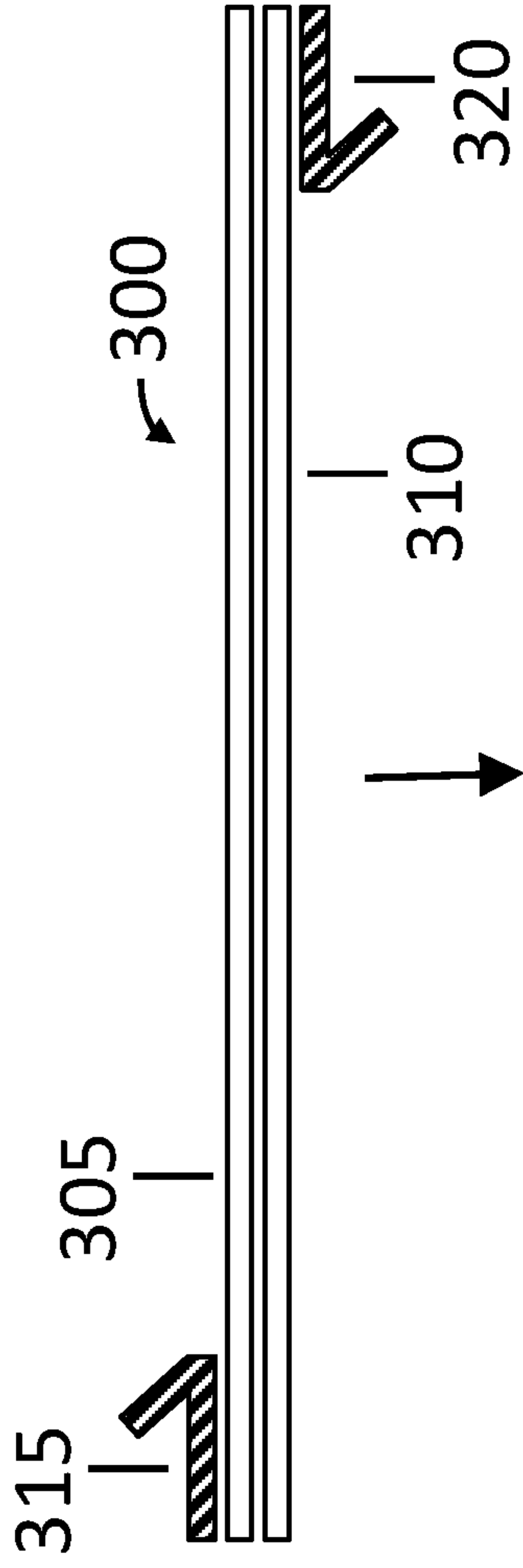


FIG. 3A

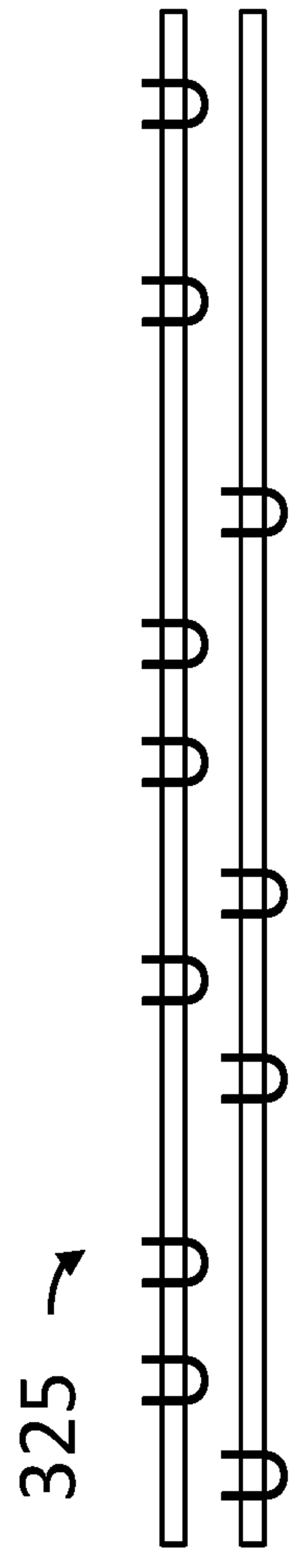


FIG. 3B

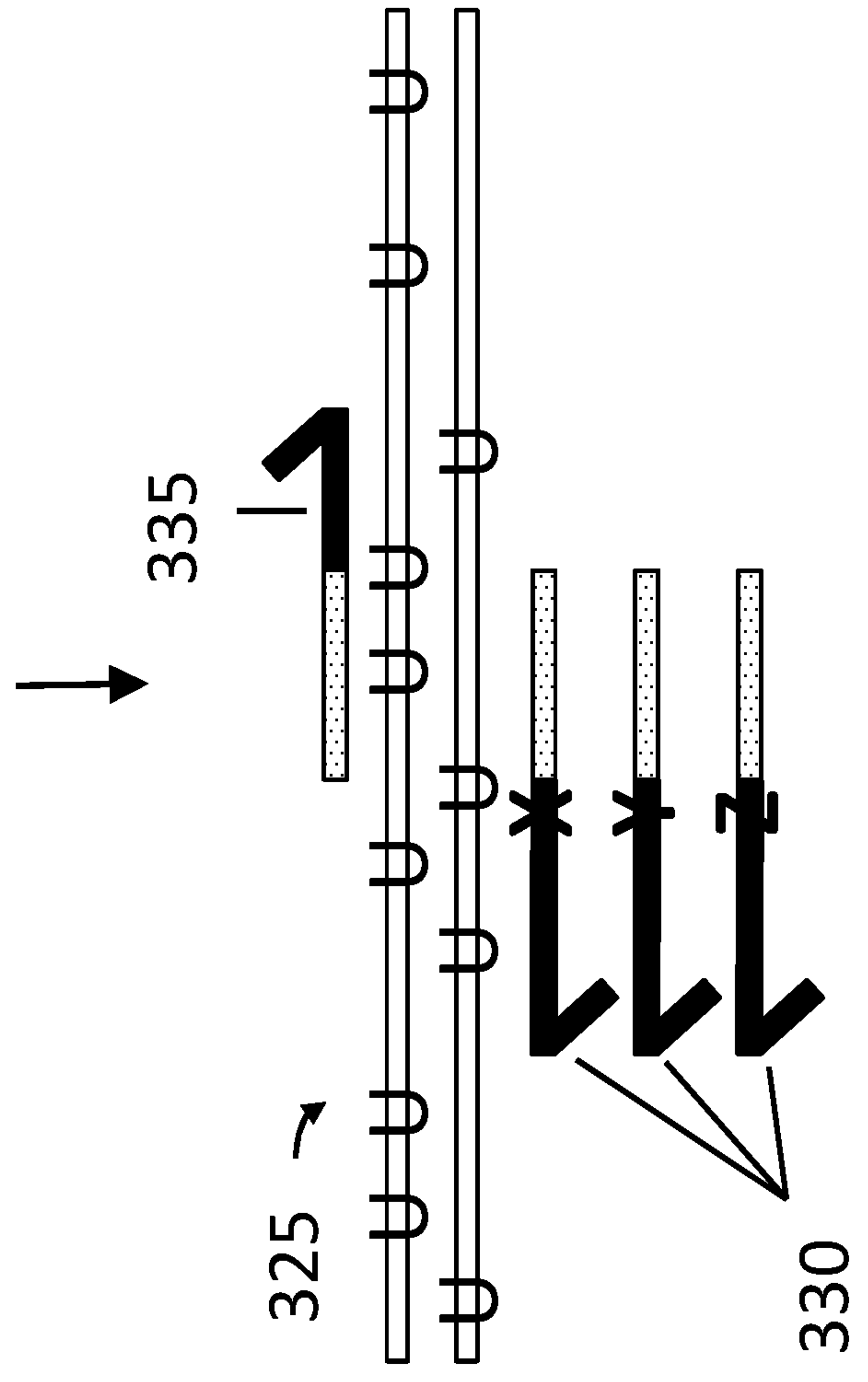


FIG. 3C



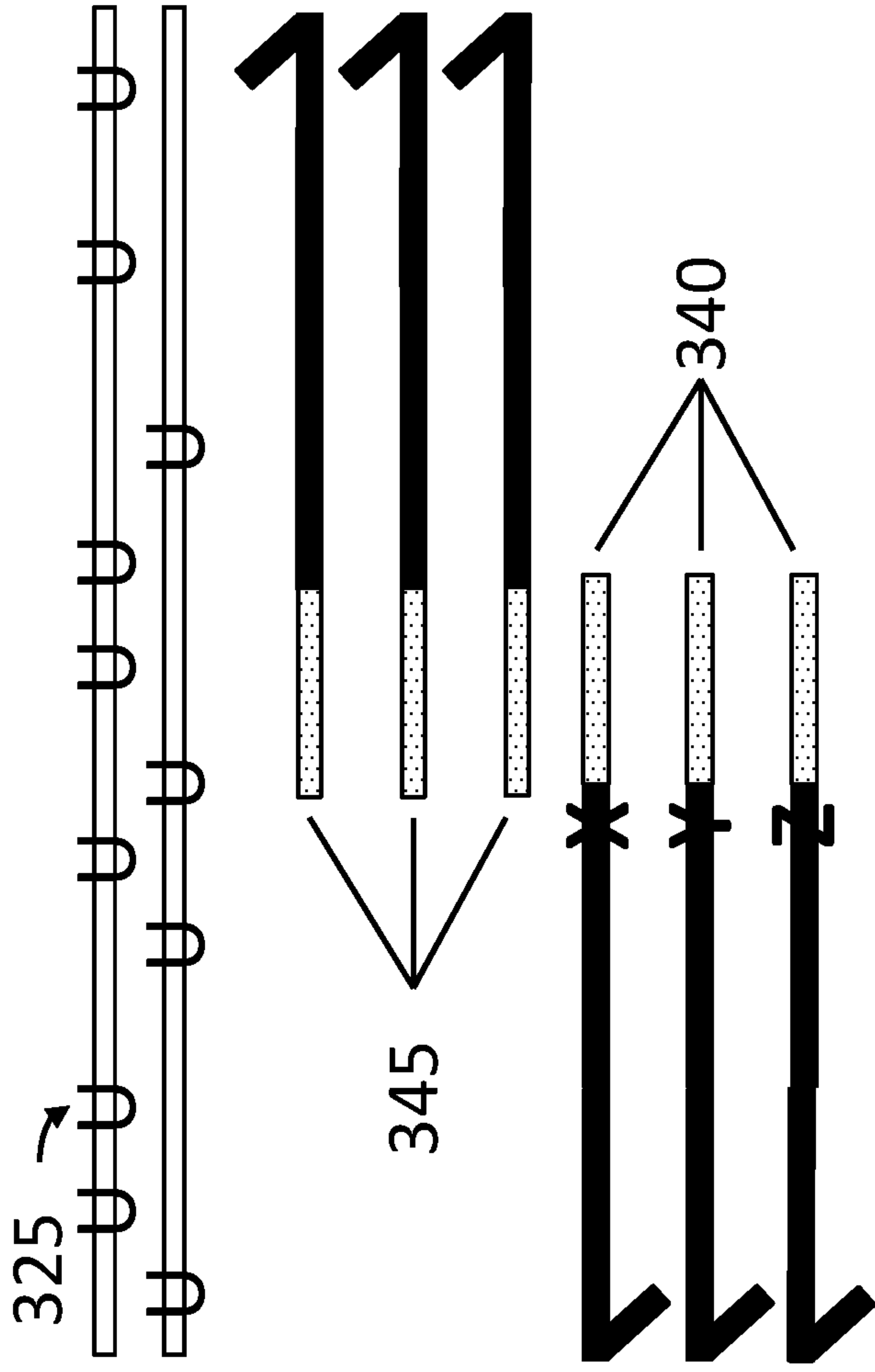


FIG. 3D

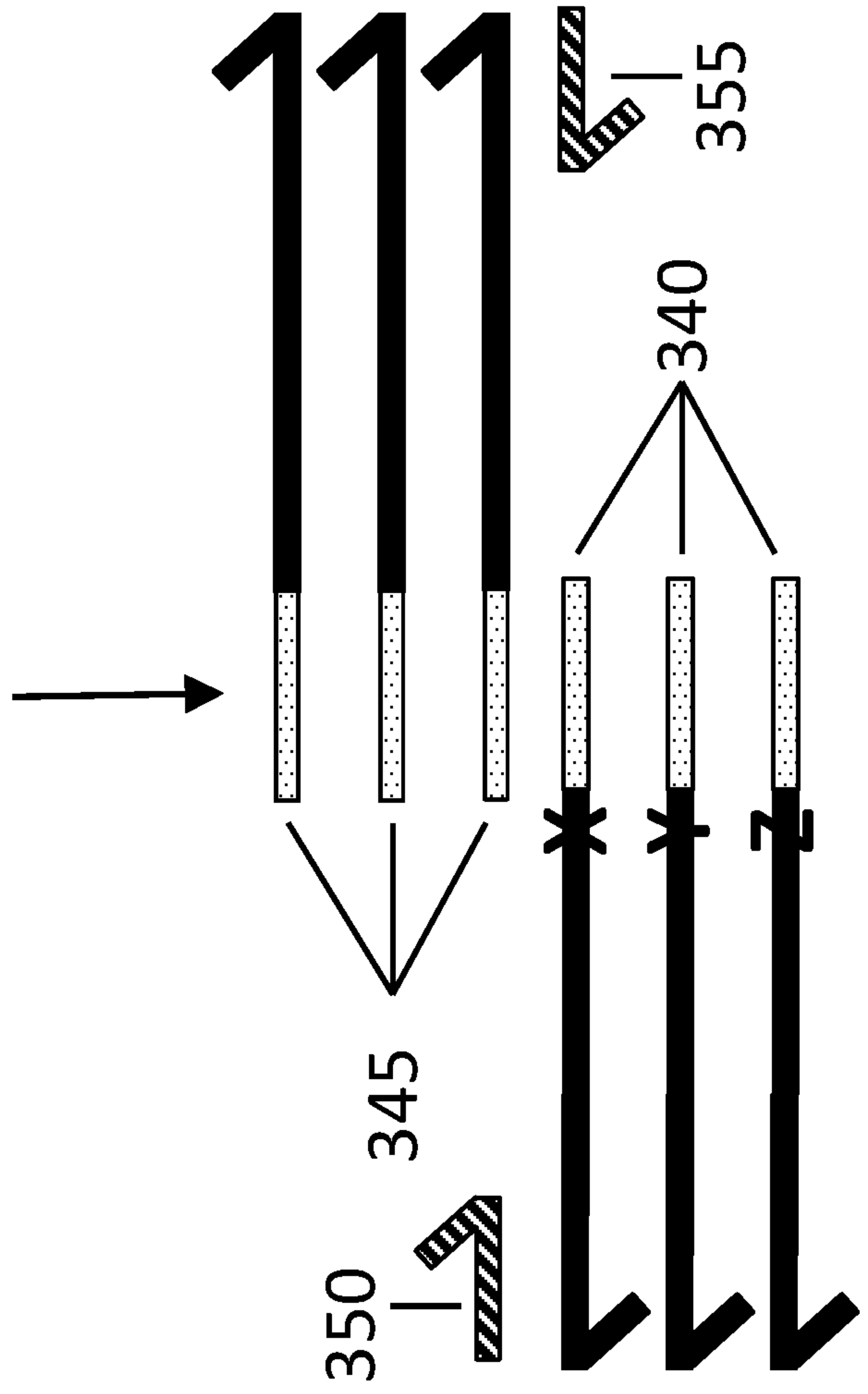


FIG. 3E

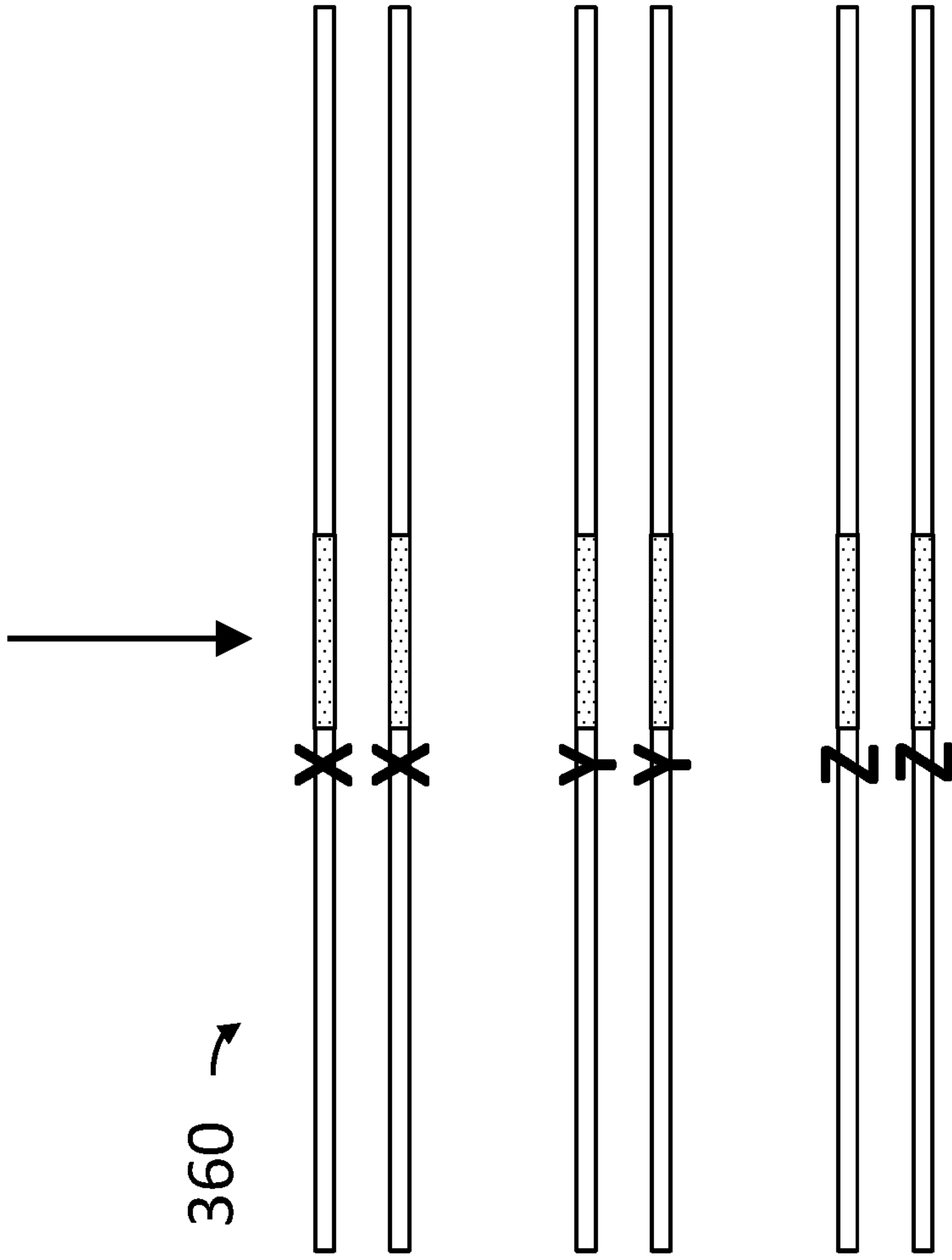


FIG. 3F



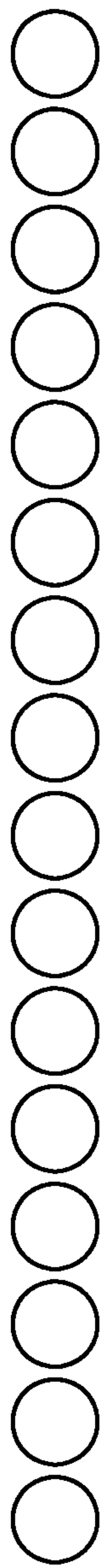


FIG. 4A

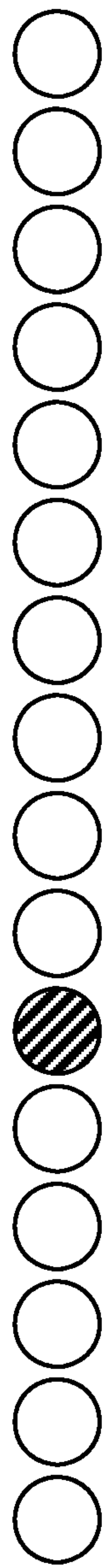


FIG. 4B

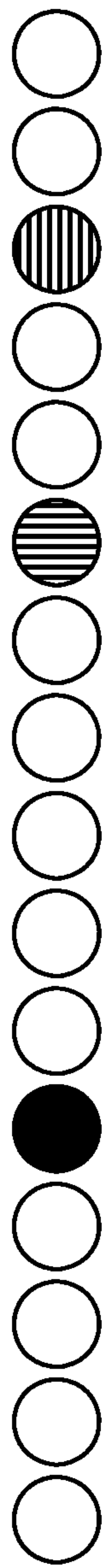


FIG. 4C

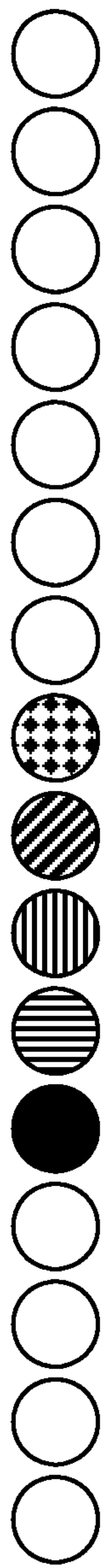


FIG. 4D

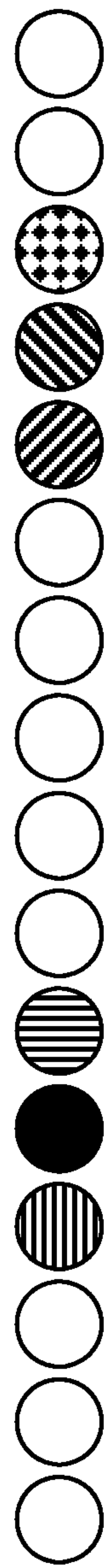
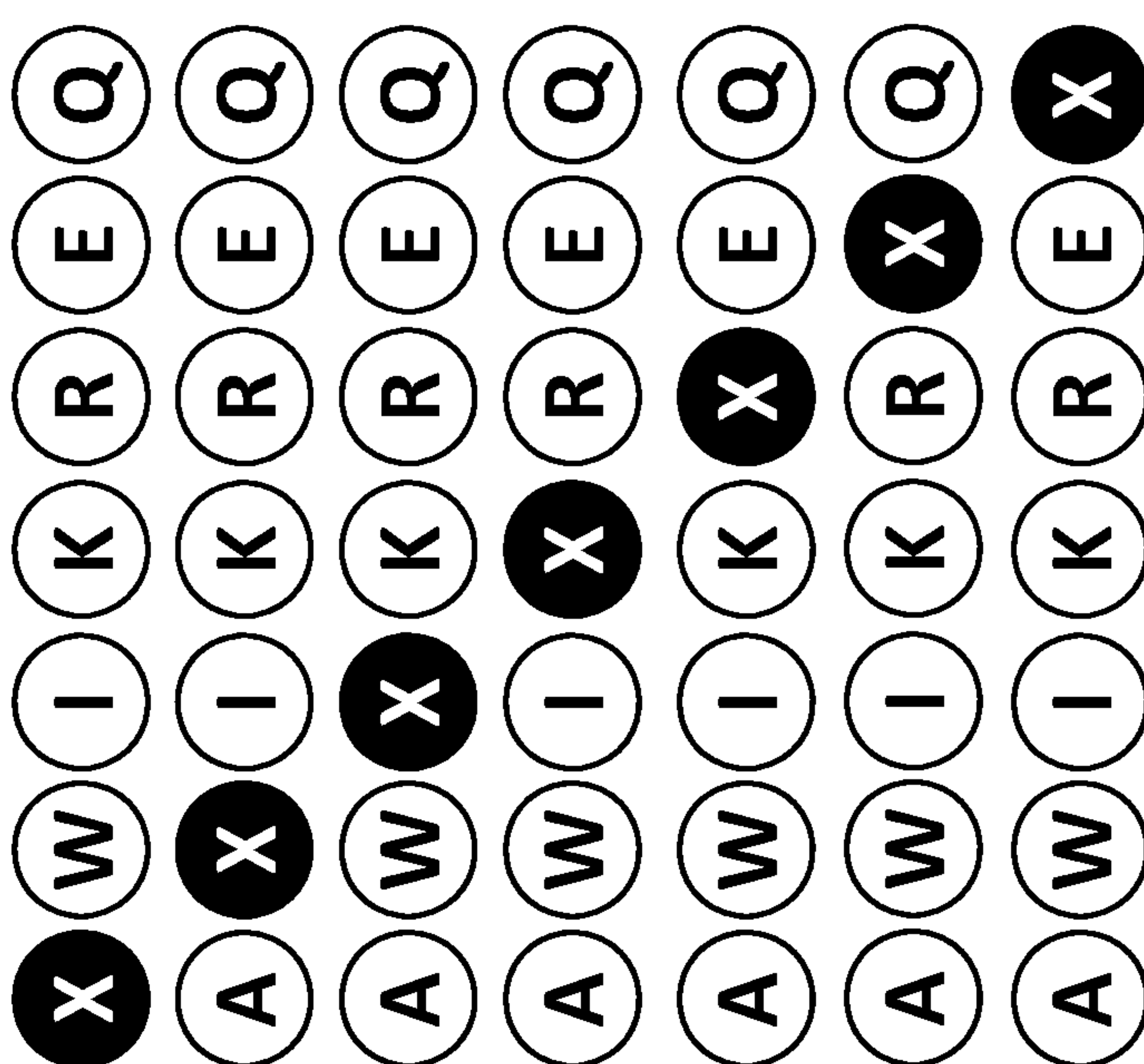
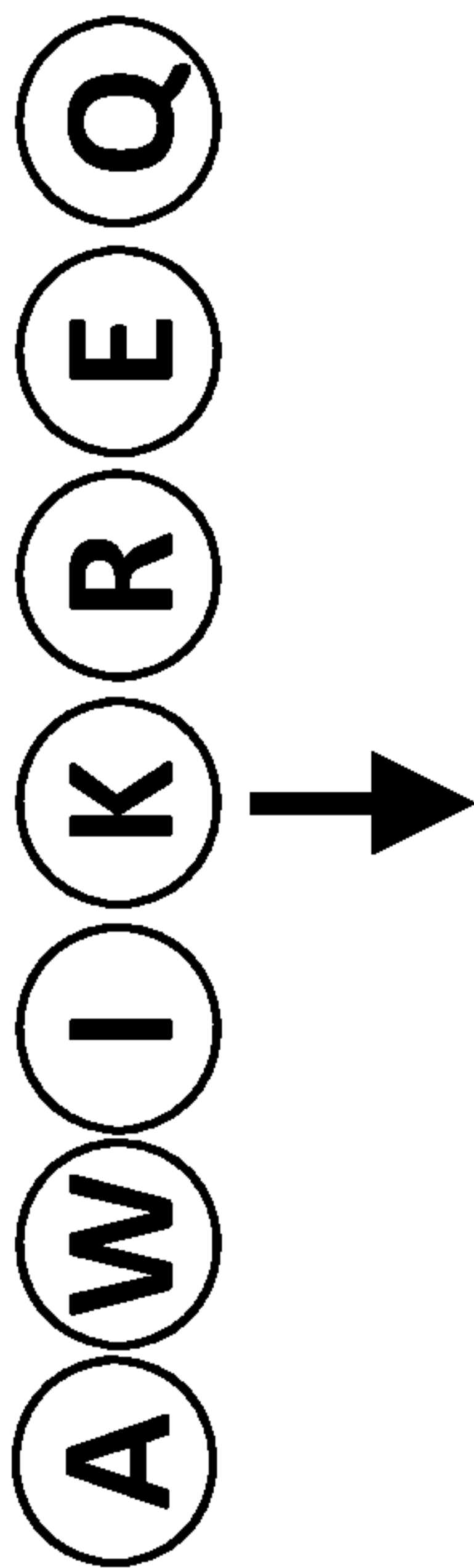


FIG. 4E



**X** = ANY CODON



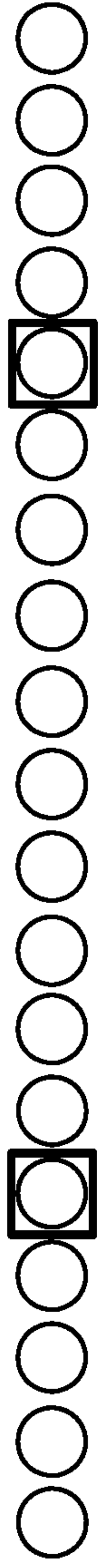


FIG. 6A

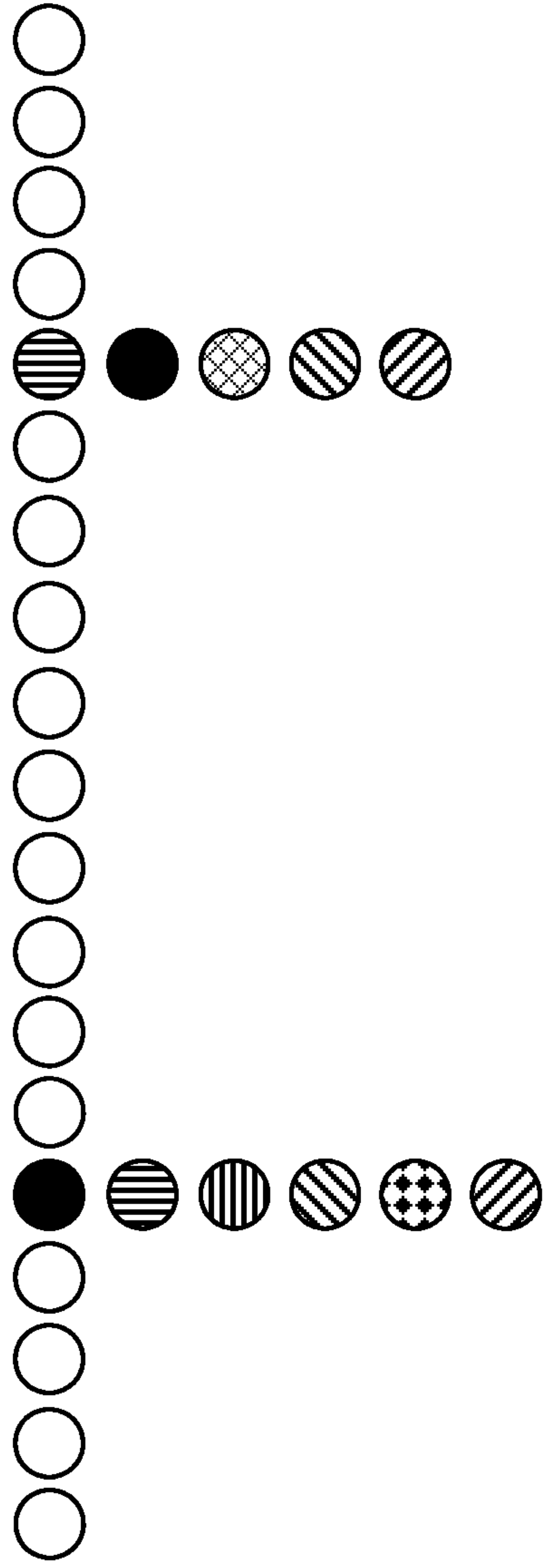


FIG. 6B



FIG. 7A

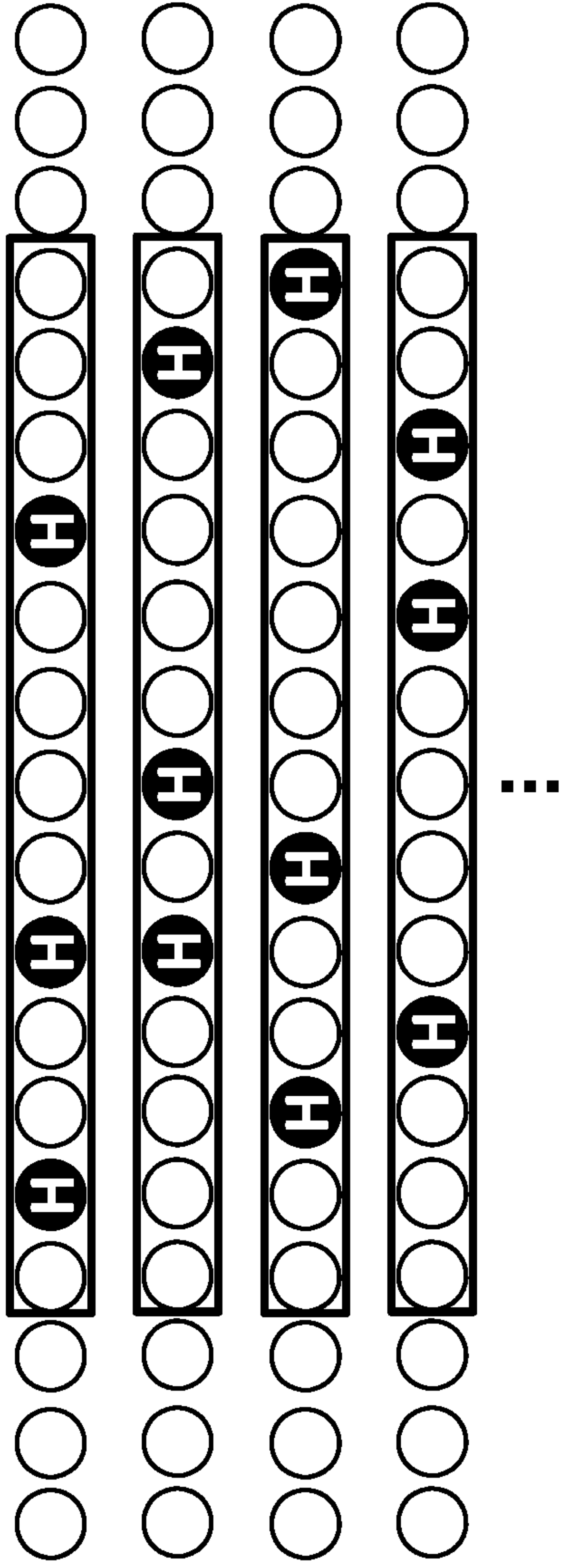


FIG. 7B



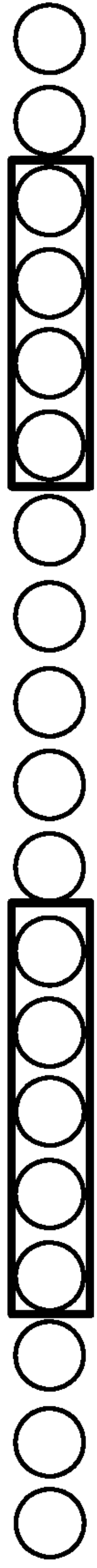


FIG. 8A

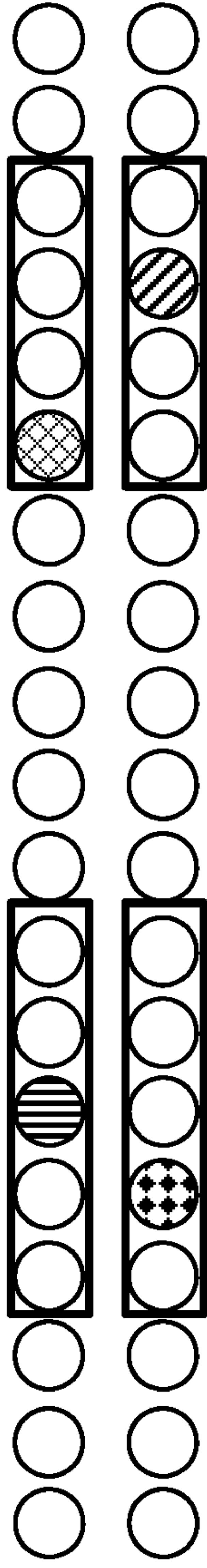
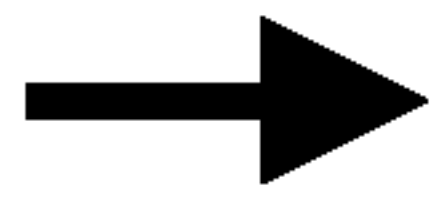


FIG. 8B

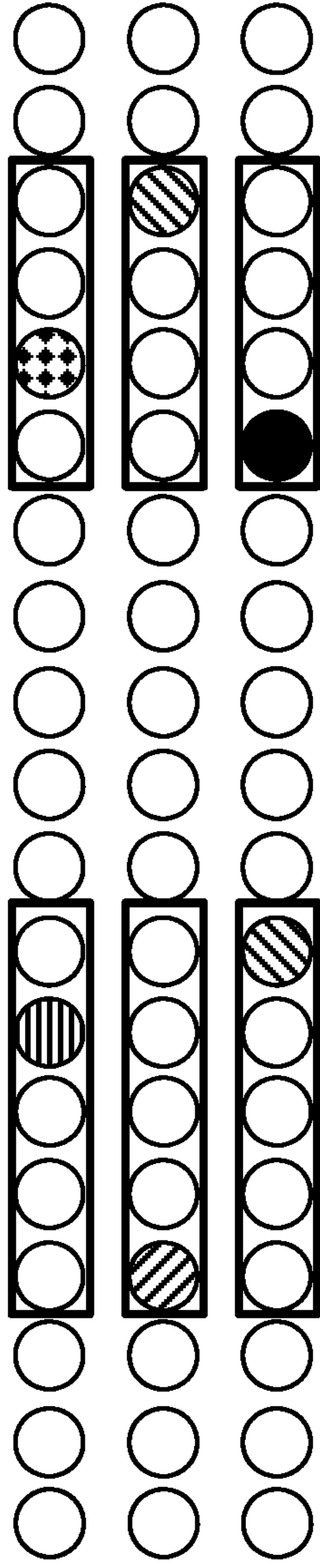




FIG. 9A

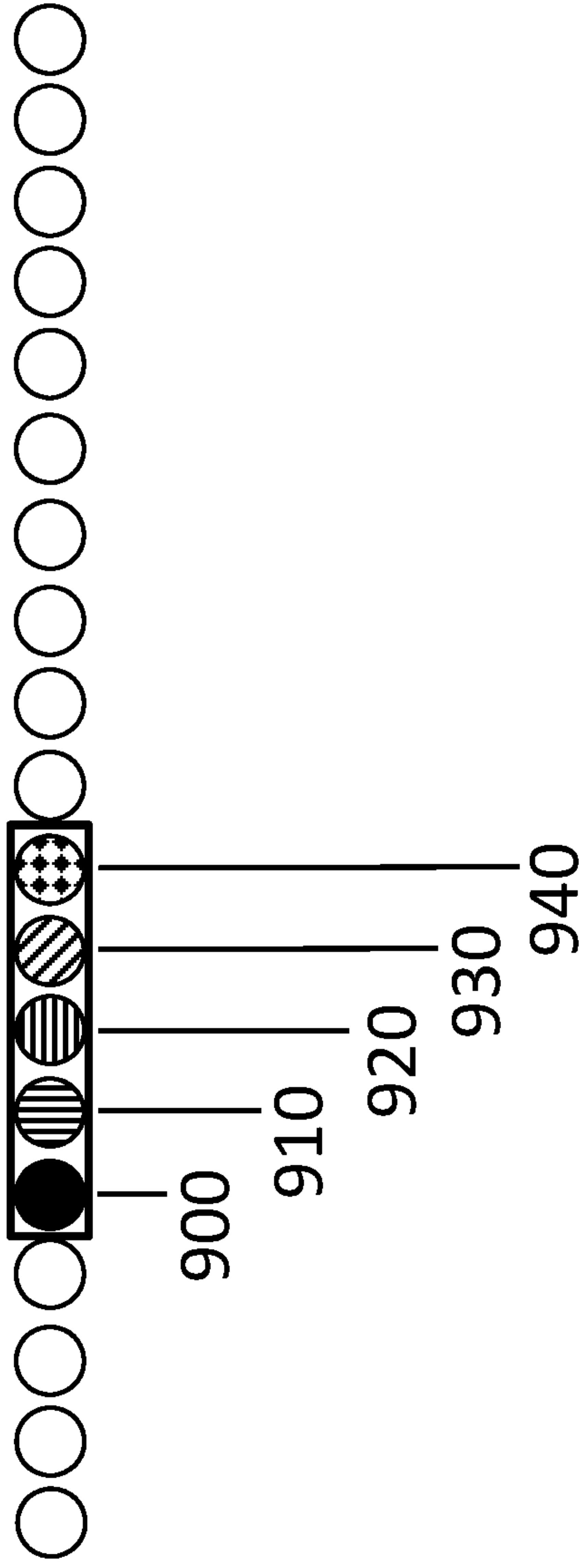
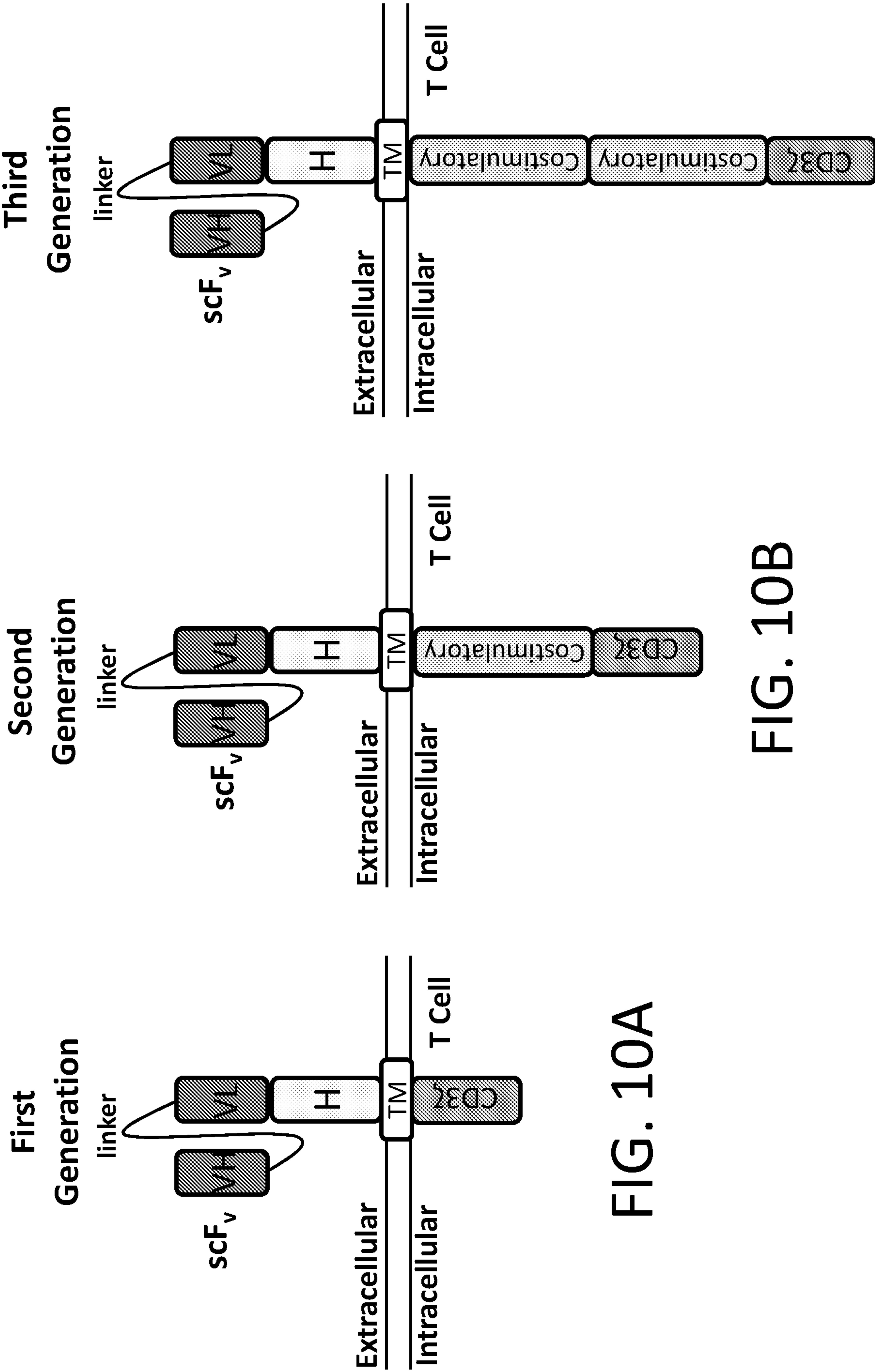


FIG. 9B





Third Generation

Second Generation

First Generation

FIG. 10B

FIG. 10A

FIG. 10C

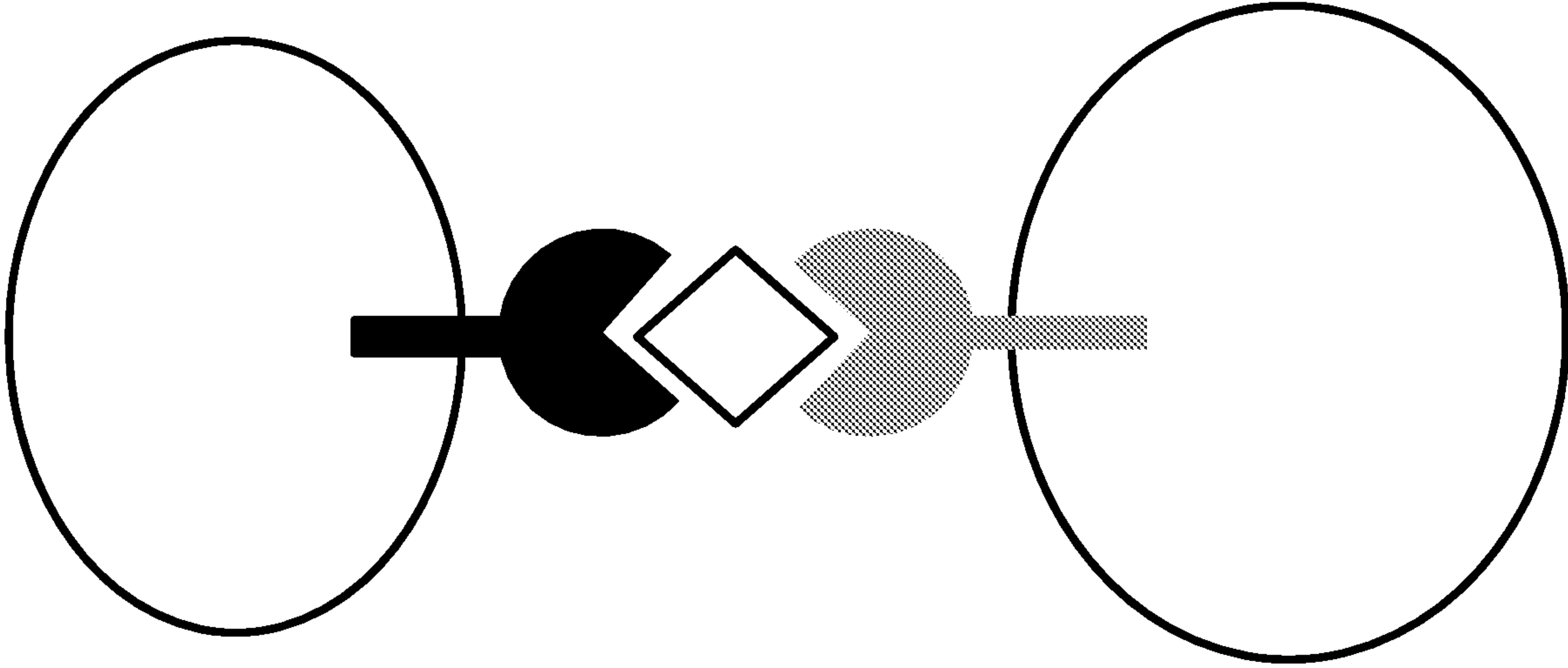


FIG. 10E

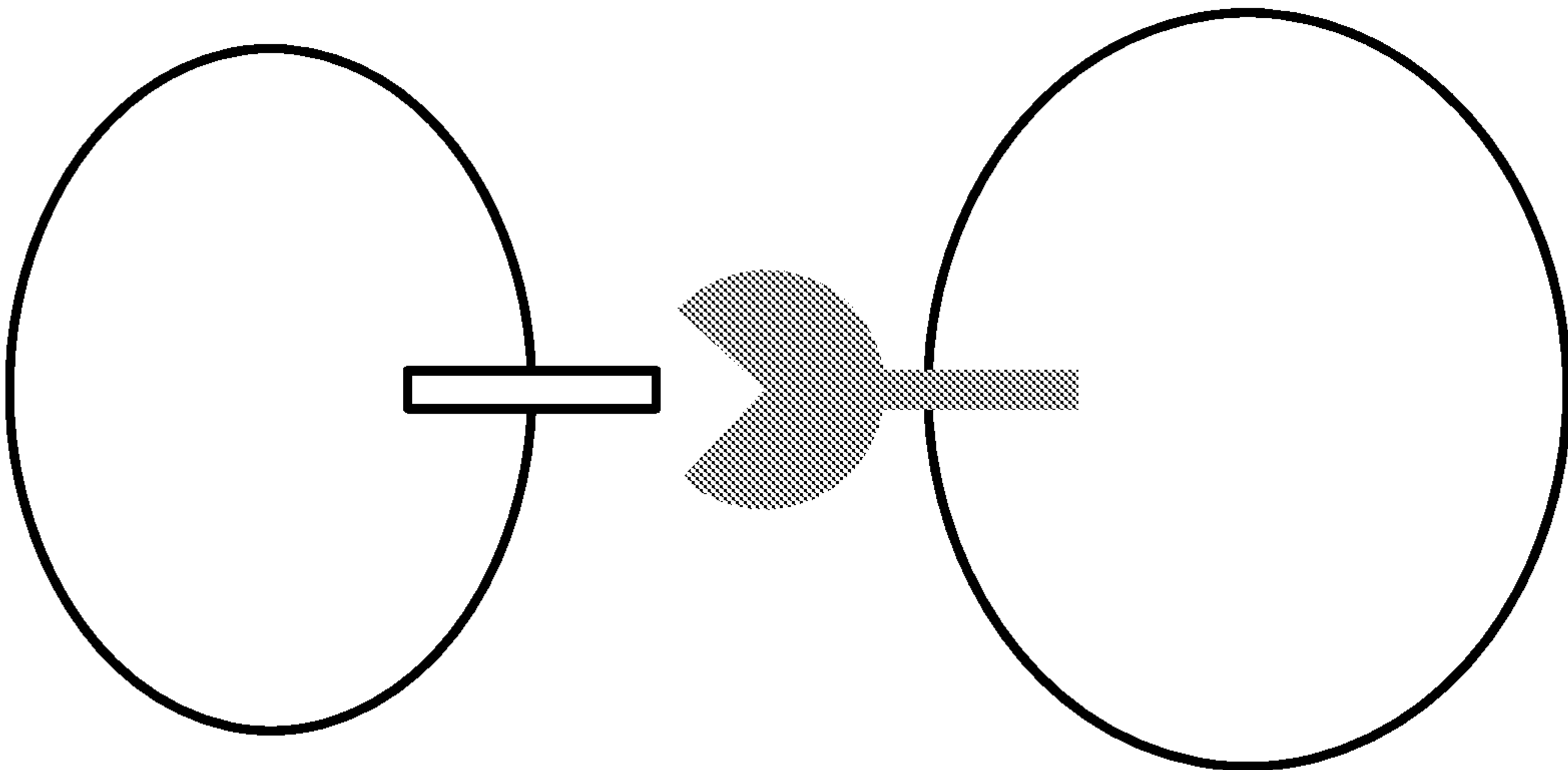


FIG. 10D



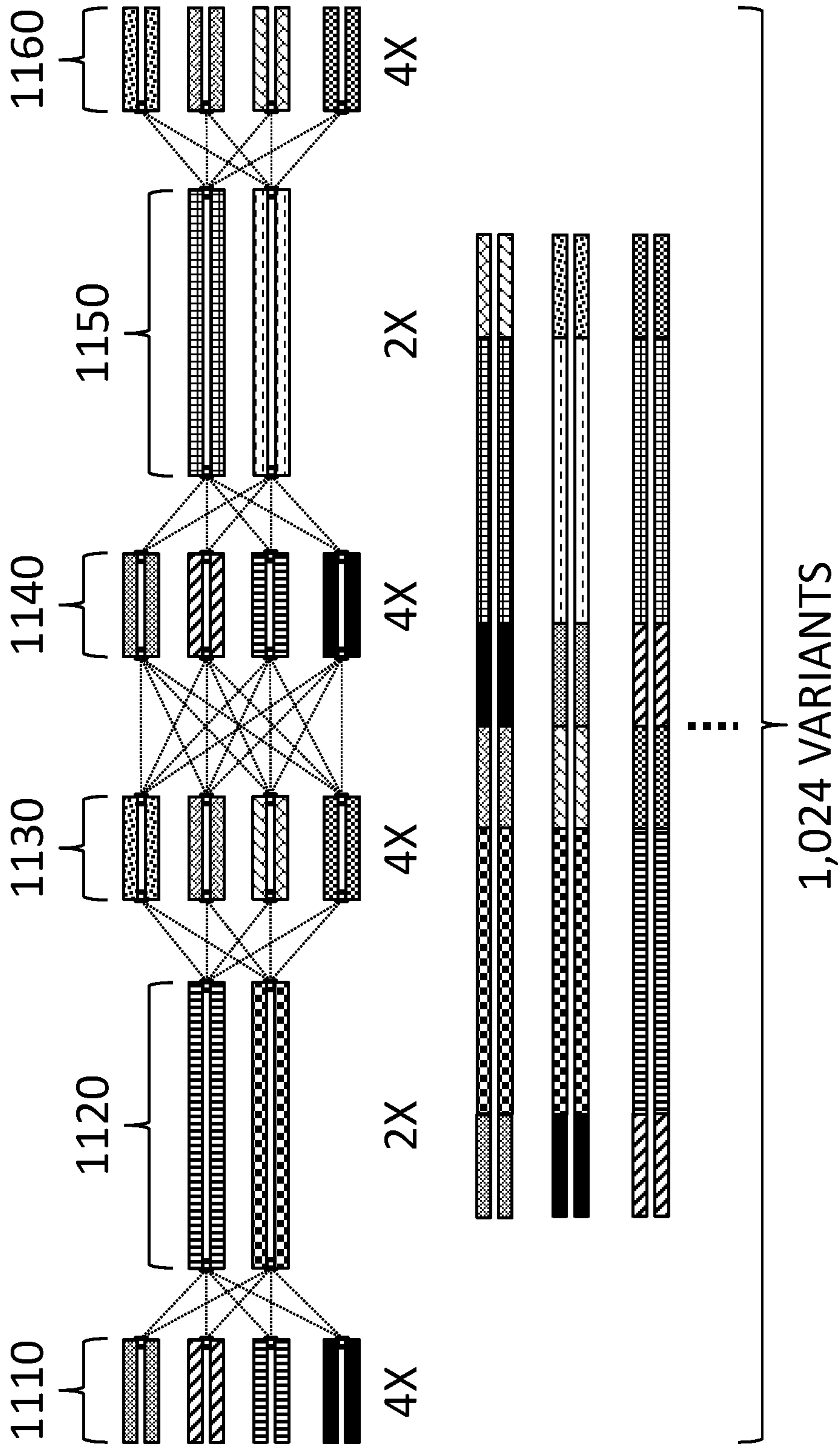


FIG. 11

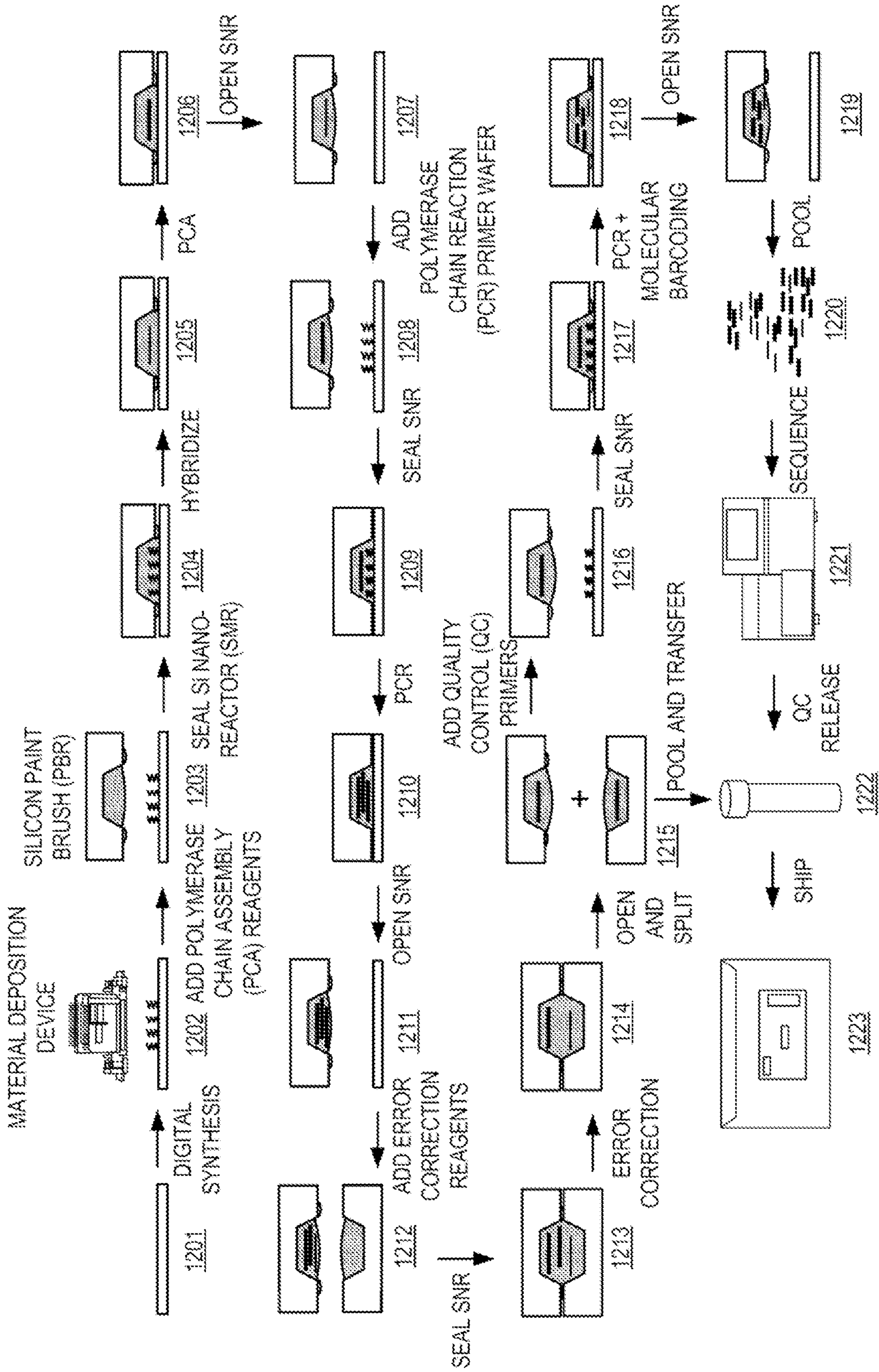


FIG. 12



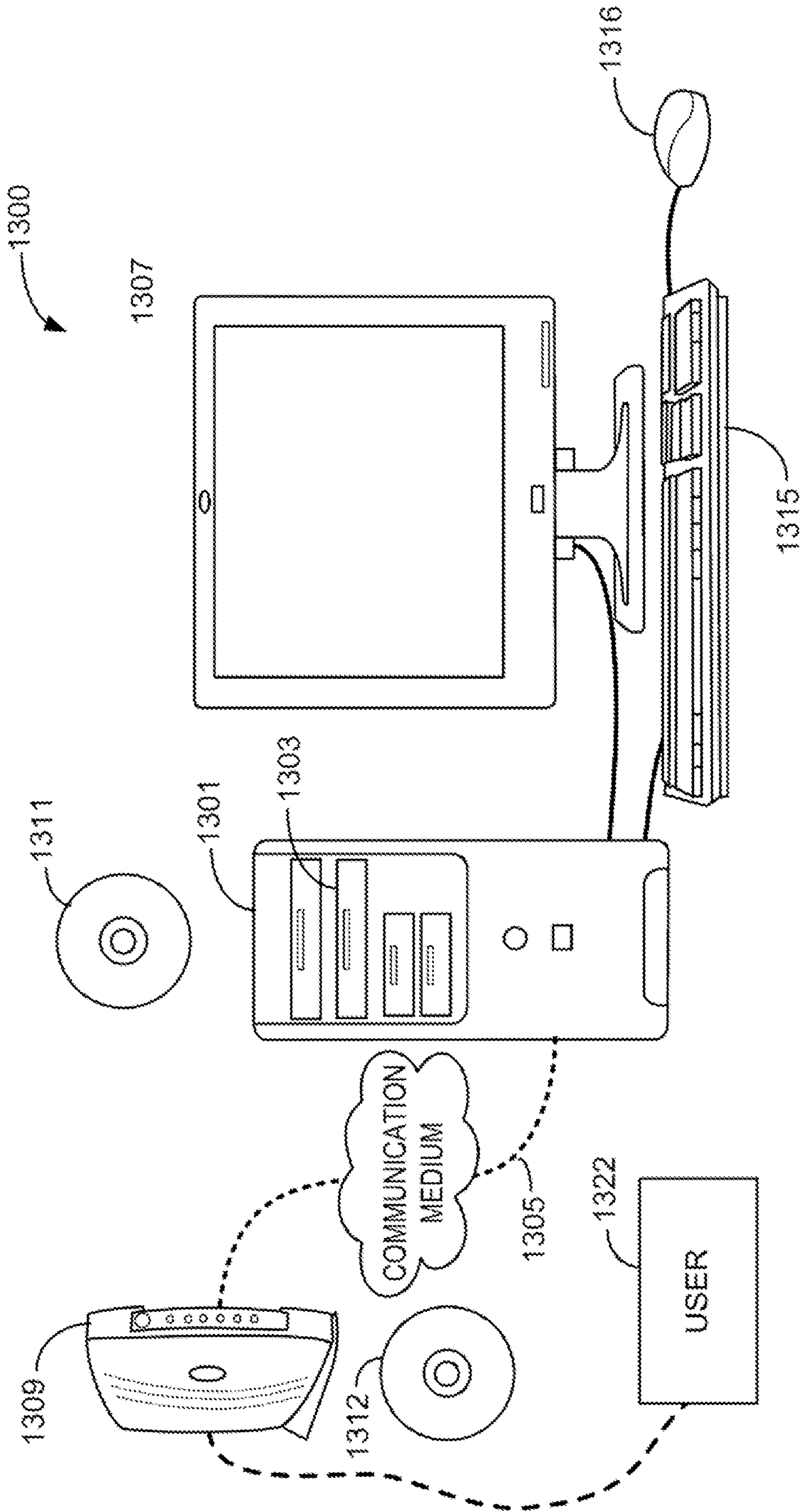


FIG. 13

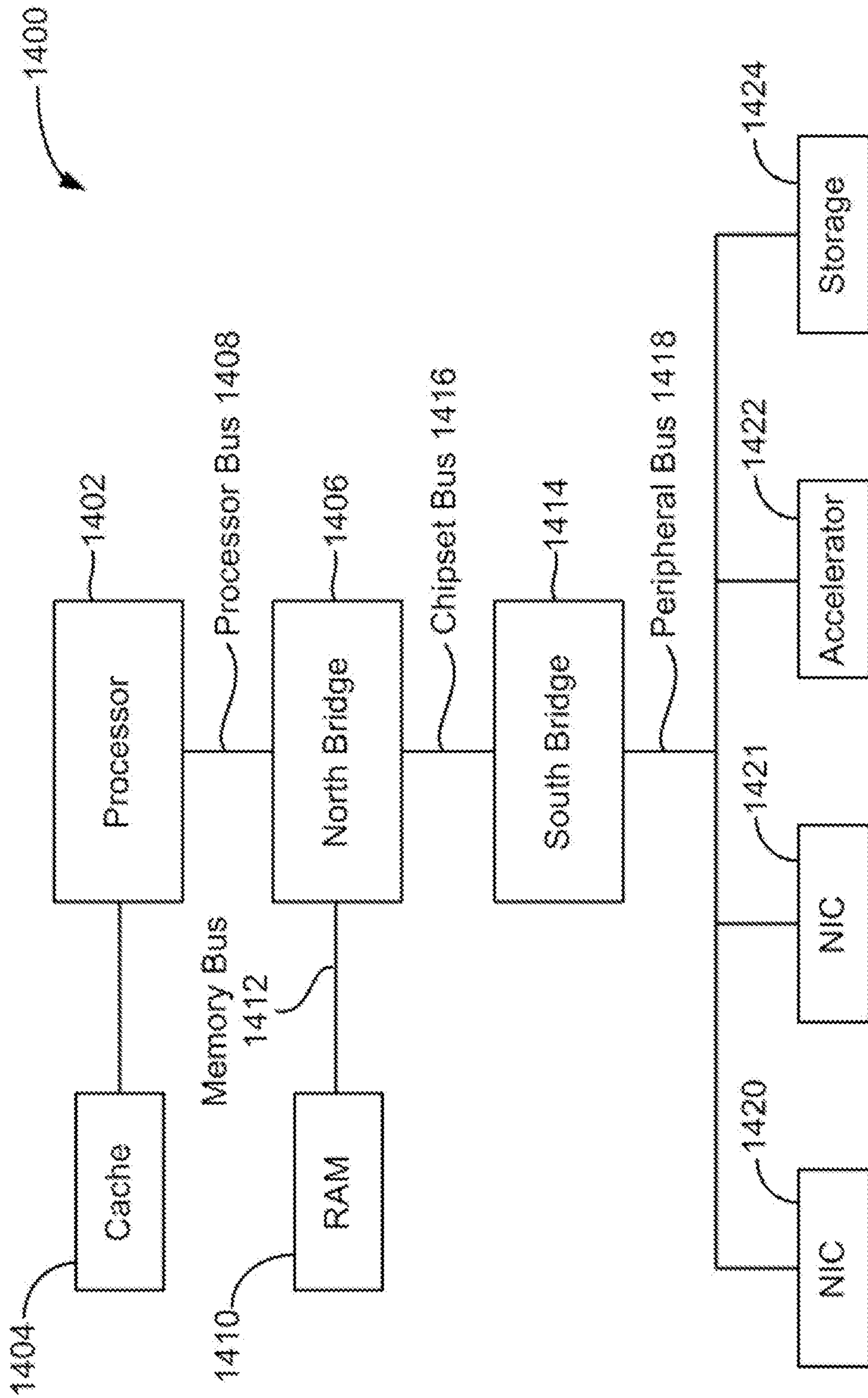


FIG. 14



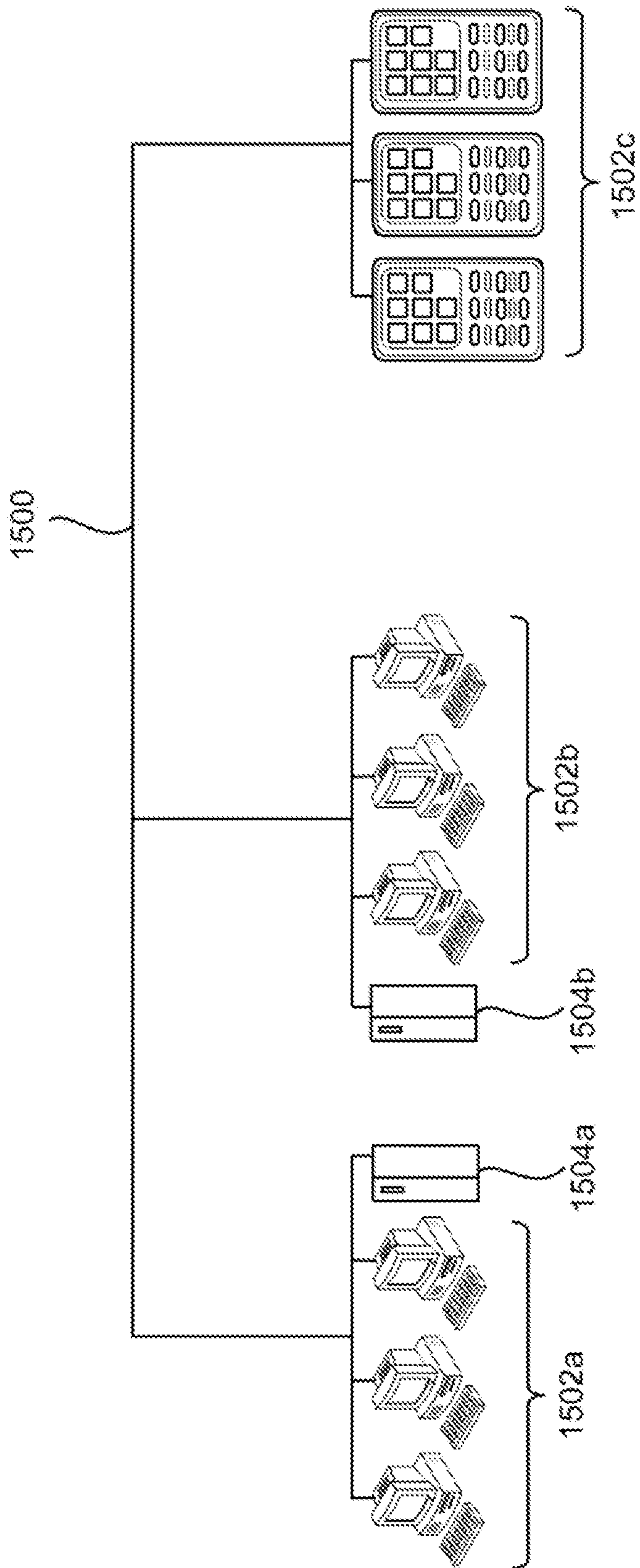


FIG. 15

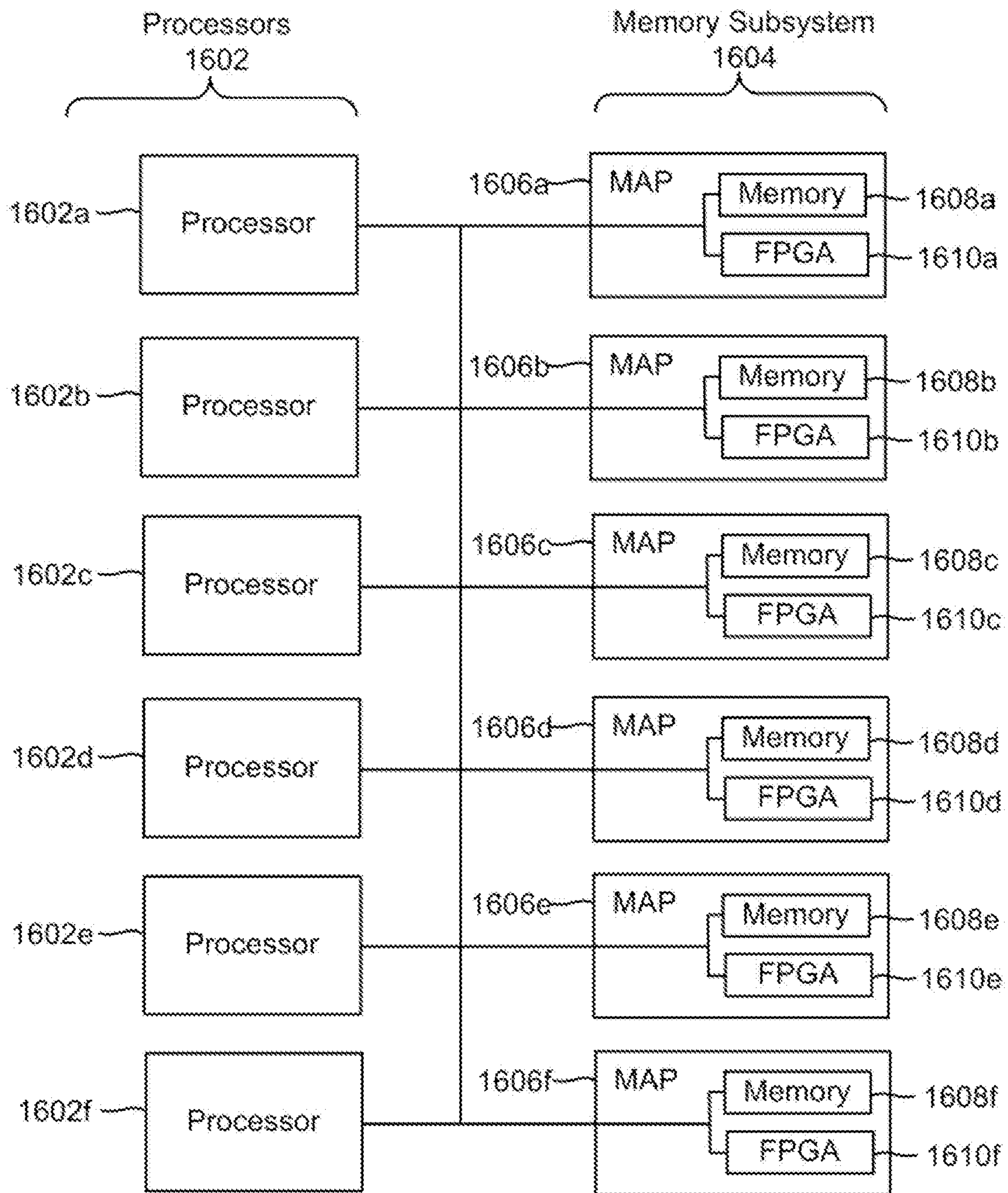


FIG. 16



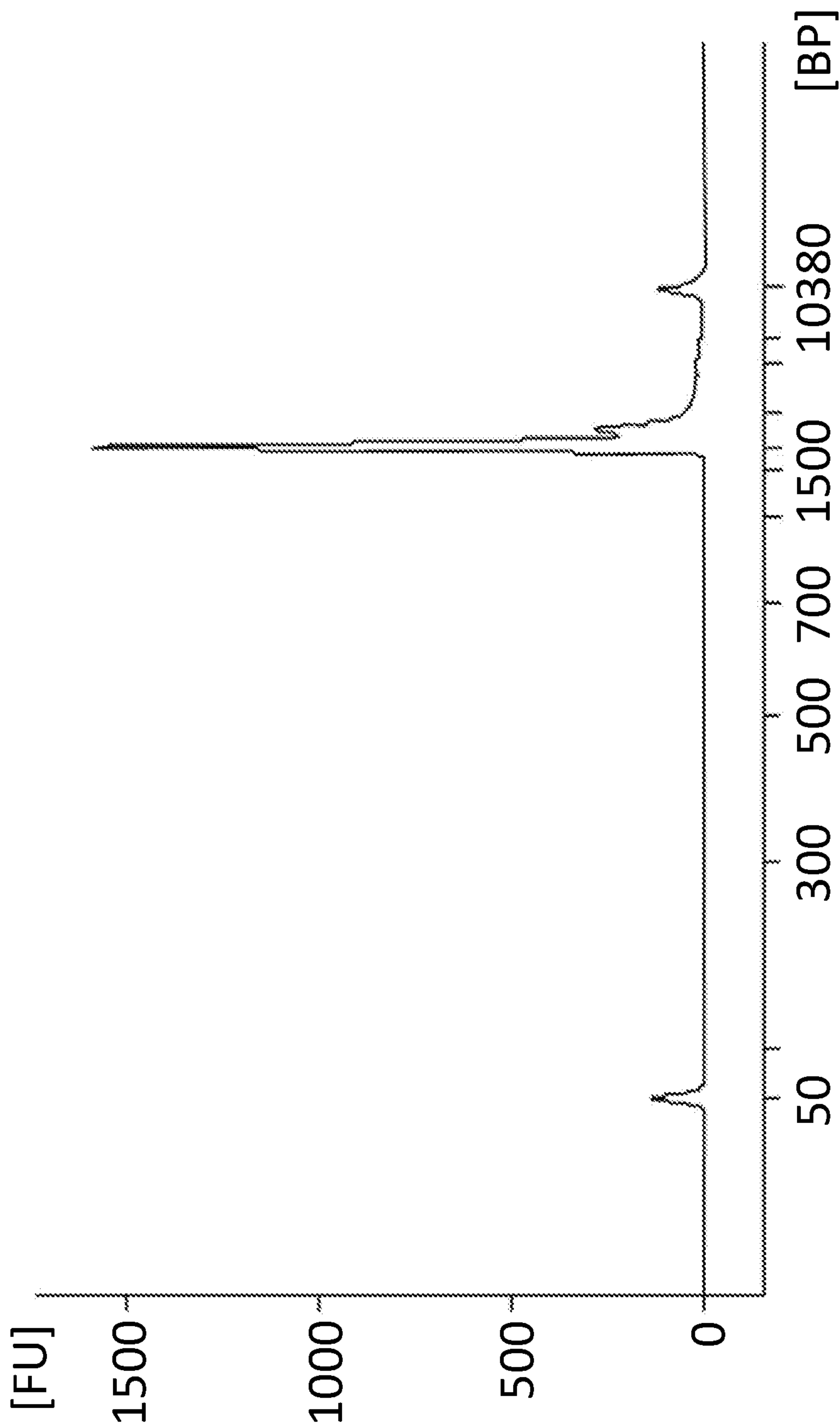
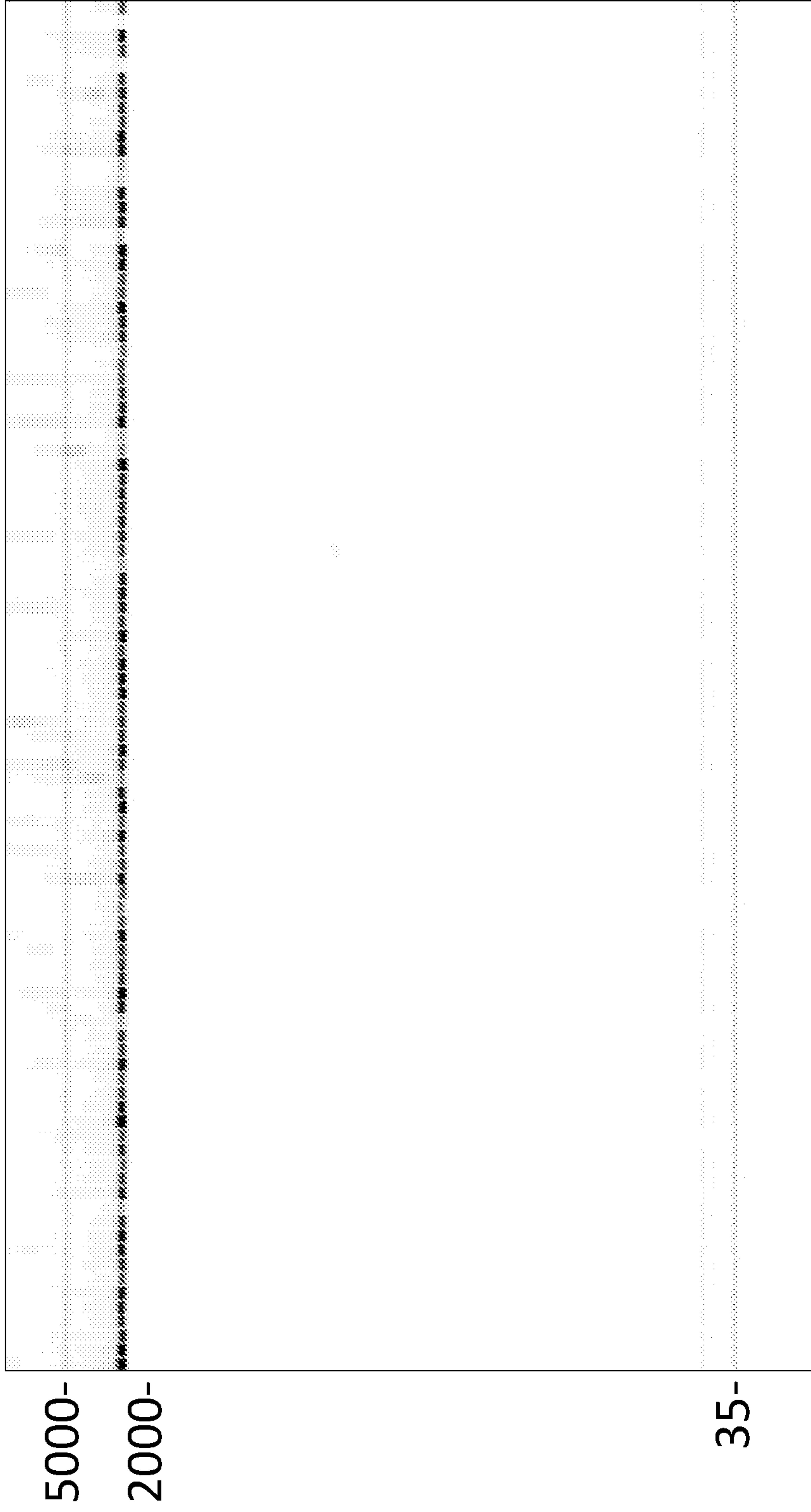


FIG. 17



SAMPLES  
FIG. 18



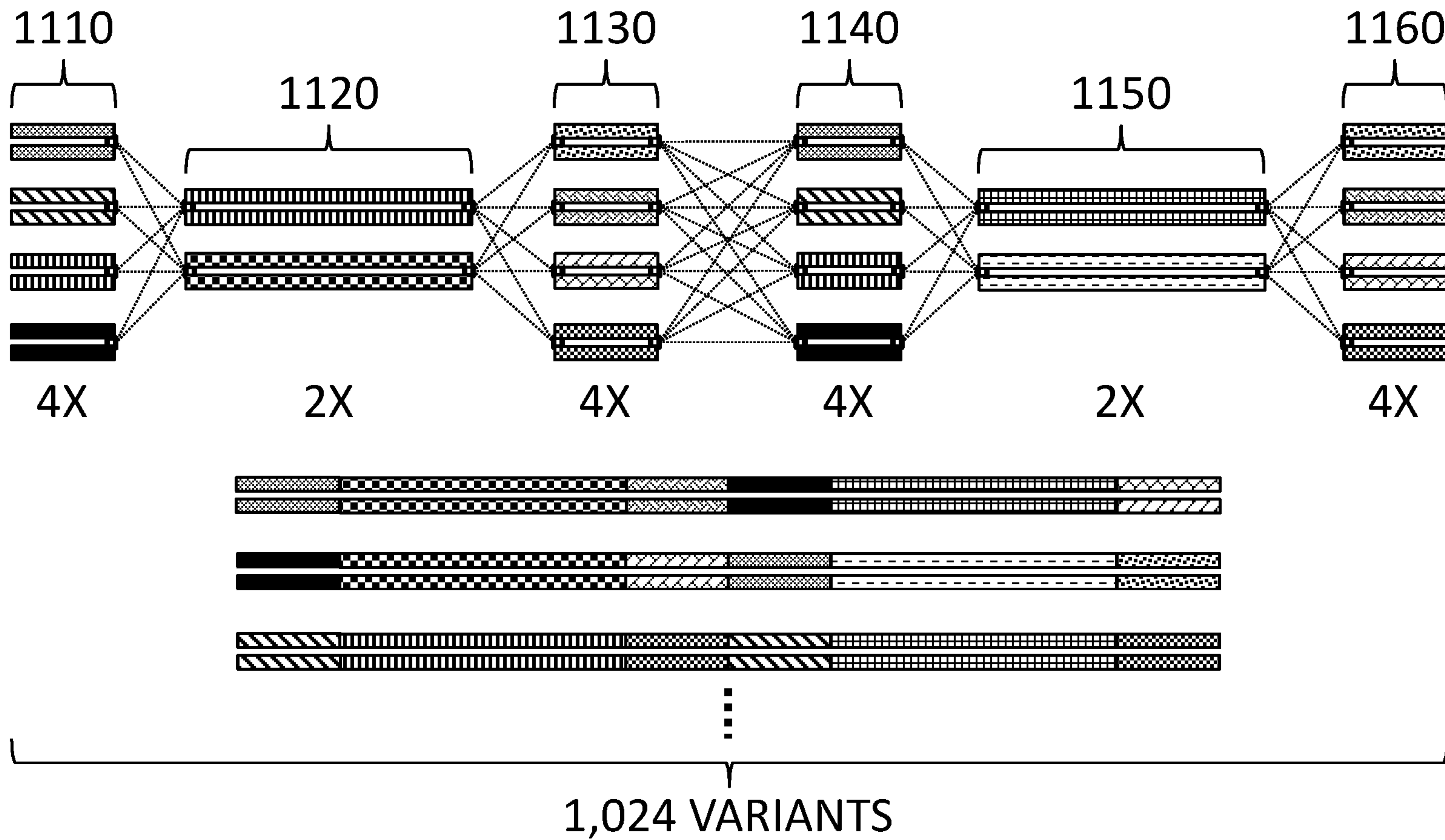


FIG. 11