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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
03 April 2025 (03.04.2025)

WIPO | PCT

(10) International Publication Number
WO 2025/072084 A1

(51) International Patent Classification:
G16B 50/10 (2019.01) *G16B 50/20* (2019.01)
G16B 50/30 (2019.01) *G16B 40/00* (2019.01)
C12Q 1/6883 (2018.01) *G16B 20/40* (2019.01)

(21) International Application Number:
PCT/US2024/047949

(22) International Filing Date:
23 September 2024 (23.09.2024)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/540,325 25 September 2023 (25.09.2023) US

(71) Applicant: **FOUNDATION MEDICINE, INC.** [US/US]; 150 Second Street, Cambridge, Massachusetts 02141 (US).

(72) Inventors: **FRAMPTON, Garrett Michael**; 150 Second Street, Cambridge, Massachusetts 02141 (US). **JIN, Dexter**; 150 Second Street, Cambridge, Massachusetts 02141 (US).

(74) Agent: **MEAD, Katherine M.** et al.; Lee & Hayes, P.C., 601 W. Riverside Ave, Suite 1400, Spokane, Washington 99201 (US).

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

WO 2025/072084 A1

(54) Title: UPDATING RECORDS BASED ON CONSENSUS ANNOTATIONS OF GENETIC VARIANTS

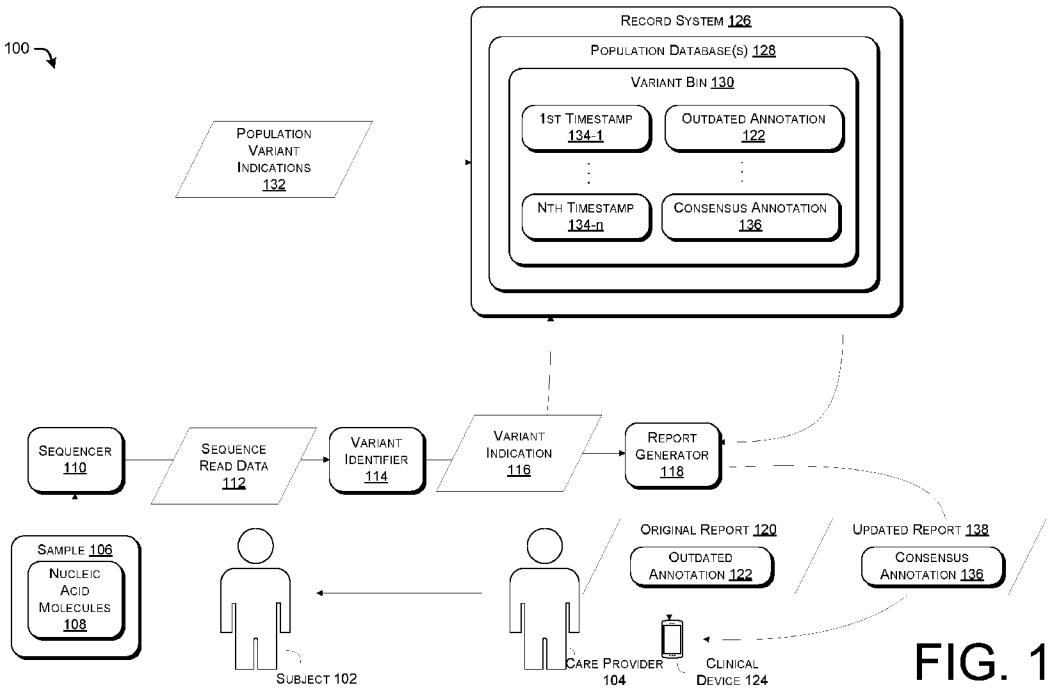

FIG. 1

(57) Abstract: Techniques for updating annotations of genetic variants are described. An example method includes identifying population records of a genetic variant shared by samples of a population of subjects, the population records respectively comprising indications of the genetic variant and annotations of the genetic variant. The example method further includes determining that a first set of the population records share a consensus annotation. The first set of the population records correspond to greater than a threshold number of subjects in the population. The example method includes updating a second set of the population records to include the consensus annotation. Further, a report based on the genetic variant and the consensus annotation is generated.

[Continued on next page]

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

Published:

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

UPDATING RECORDS BASED ON CONSENSUS ANNOTATIONS OF GENETIC VARIANTS

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This application claims the priority of U.S. Provisional App. No. 63/540,325, which is titled "Updating Records Based on Consensus Annotations of Genetic Variants," was filed on September 25, 2023, and is incorporated by reference herein in its entirety.

BACKGROUND

[0002] Genetic mutations and other variants are relevant to human health. In particular, specific types of genetic variants are associated with common characteristics. In some cases, a specific type of genetic variant may be associated with a particular type or subtype of cancer. In some examples, a specific type of genetic variant is associated with cancers that are resistant to some therapies and responsive to other therapies. These associations can be highly pertinent to cancer treatment and management.

[0003] In many cases, a sample obtained from a patient can be analyzed and sequenced in order to identify genetic variants of the patient. A report summarizing the genetic variants, and any known associations between the genetic variants and clinically relevant characteristics, can be generated and output. However, on-going cancer research leads to new relevant information being generated rapidly. Thus, the characteristics of a given genetic variant reported at one time may be quickly out-of-date due to additional research findings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0004] Various aspects of the disclosed methods, devices, and systems are set forth with particularity in the appended claims. A better understanding of the features and advantages of the disclosed methods, devices, and systems will be obtained by reference to the following detailed description of illustrative embodiments and the accompanying drawings, of which:

[0005] FIG. 1 illustrates an example environment for efficiently distributing updated annotations of genetic variants.

[0006] FIG. 2 illustrates example records stored in at least one patient database.

[0007] FIG. 3 illustrates an example report summarizing predicted categories of a cancer of a subject.

[0008] FIG. 4 illustrates an example process for updating annotations in patient records.

[0009] FIG. 5 illustrates an example environment for sequencing various nucleic acid molecules.

[0010] FIG. 6 illustrates one or more devices configured to perform various operations described herein.

DETAILED DESCRIPTION

[0011] Various implementations of the present disclosure relate to techniques for identifying and distributing consensus annotations associated with genetic variants. Records of multiple samples associated with the same genetic variant may be stored in one or more databases. Each record may include one or more annotations associated with the variant, which can describe associations between the variant and various pathologies (e.g., types or subtypes of cancers), whether the pathologies are responsive to treatments, whether the pathologies are resistant to treatments, and other clinically relevant characteristics. As additional instances of the genetic variant are identified in a population, additional records may be added. Earlier records may, however, indicate out-of-date annotations in view of later records. It may be beneficial to update the earlier records with updated annotations based on the later records.

[0012] While it may be possible to store the records in a live database that would automatically update earlier annotations based on later annotations continuously and in real-time as new data is received, the live database could

introduce problems to the records. In some cases, numerous records can be stored in the databases, such as millions of records of thousands of genetic variants from thousands of patient samples. In many cases, maintaining and updating a live database for this purpose could require prohibitive amounts of processing and memory resources. Further, it is possible that some recent annotations are inaccurate. In some examples, a live database may be designed that includes entries corresponding to variant identifiers, rather than entries corresponding to instances of observed variants. While this may minimize memory resources, it presents other problems. For instance, if there is a lack of a consensus on a given variant identifier, the entry may be frequently updated, leading to uncertainty. Automatically updating relevant records in a live database continuously and in real-time based on each recent annotation could therefore cause the storage and distribution of inaccurate annotations.

[0013] Various implementations of the present disclosure enable automatically updating patient records without the use of a live database and while minimizing inconsistency between recent annotations. Accordingly, implementations of the present disclosure can reduce the amount of processing and memory resources utilized to track records of genetic variants throughout a population. Further, implementations of the present disclosure relate to identifying whether recent annotations represent consensus annotations, prior to distributing the consensus annotations to earlier records that may not yet have included the consensus annotations or that may include older annotations that are now outdated in view of the consensus annotations. Thus, implementations may prevent the distribution of inaccurate annotations of genetic variants. For at least these reasons, implementations of the present disclosure provide significant improvements to the technical fields of bioinformatics, particularly related to genetic analysis.

[0014] Various analyses described herein cannot be performed in the human mind, or by pen and paper. For example, it would be impossible for a single human to track hundreds, thousands, or millions of records of various genetic variants identified from various patient samples. Further, it would be impossible for a single human to identify consensus annotations represented by recent records using various techniques described herein. In addition, it would be impossible for a single human to distribute consensus annotations to previous records of a particular genetic variant. Implementations of the present disclosure are therefore directed to practical applications that are deeply rooted in computer technology, by enabling actions that could not be performed without the use of a computer.

Example Definitions

[0015] As used herein, the terms "deoxyribonucleic acid," "DNA," "DNA molecule," and their equivalents, may refer to a polymer of nucleotides (also referred to as "nucleobases") containing deoxyribose. The nucleotides in DNA include cytosine (C), guanine (G), adenine (A), and thymine (T). Each DNA nucleotide includes a deoxyribose and a phosphate group. An example single-stranded DNA (ssDNA) molecule includes a chain of covalently bonded DNA nucleotides. In the example ssDNA molecule, the phosphate group of the mth nucleotide is covalently bonded to the deoxyribose of the (m-1)th nucleotide, wherein m is a positive integer greater than 2 and less than or equal to the number of DNA nucleotides in the chain. In various examples, DNA is double-stranded and includes two ssDNA molecules that are complementary to one another and coiled around each other in a double helix form. The nucleotides of one ssDNA molecule are hydrogen bonded to the nucleotides of the other ssDNA molecule. In particular, the pyrimidines (A and T) hydrogen bond to each other, and the purines (C and G) hydrogen bond to each other.

[0016] As used herein, the terms "ribonucleic acid," "RNA," "RNA molecule," and their equivalents, may refer to a polymer of nucleotides containing ribose. The nucleotides in RNA include cytosine (C), guanine (G), adenine (A),

and uracil (U). Each RNA nucleotide includes a ribose and a phosphate group. In an example RNA molecule, the phosphate group of the *n*th nucleotide is covalently bonded to the ribose of the (*n*-1)th nucleotide, wherein *n* is a positive integer greater than 2 and less than or equal to the number of RNA nucleotides in the chain. Messenger RNA (mRNA) is a type of RNA molecule that is synthesized (or "transcribed") by RNA polymerase (an enzyme) to be complementary to a gene encoded in a DNA sequence, and is also used by a ribosome to synthesize a polypeptide or protein. An mRNA is therefore an example of a "coding RNA." In various cases, intron sequences are removed from an mRNA via a process known as "RNA splicing." MicroRNA ("miRNA") are single-stranded RNA molecules that perform post-transcriptional gene expression regulation. For instance, a miRNA may bind to a complementary mRNA molecule, thereby cleaving, destabilizing, or otherwise preventing the mRNA molecule from being translated into a polypeptide or protein by a ribosome. In various examples, a miRNA has a length in a range of 21 to 23 RNA nucleotides. As used herein, the terms "non-coding RNA" may refer to a type of RNA that is not translated into a protein. Examples of non-coding RNA include miRNA, transfer RNA (tRNA), and ribosomal RNA (rRNA). The term "functional RNA," and its equivalents, may refer to any RNA molecule that impacts a biological process. For instance, functional RNA may include mRNA, miRNA, tRNA, rRNA, and the like.

[0017] As used herein, the term "base," and its equivalents, may refer to a monomer of a polymer. For example, a base of DNA or RNA is a nucleotide.

[0018] As used herein, the term "base pair," and its equivalents, may refer to a pair of complementary DNA nucleotides, which are hydrogen-bonded to one another in a double-stranded DNA molecule. For example, a base pair includes a first base in a first ssDNA and a second base in a second ssDNA, wherein the first and second bases are complementary and hydrogen-bonded to one another.

[0019] As used herein, the terms "nucleotide," "nucleobase," "nucleic acid," "nucleic acid molecule," and their equivalents, may refer to an organic molecule that includes a nitrogenous base, a sugar, and a phosphate group. In various cases, a nucleotide is a monomer of DNA or RNA. A nucleotide, for instance, is a chemical structure.

[0020] As used herein, the terms "3' end," "3-prime end," and their equivalents, may refer to a terminus of a single-stranded nucleotide polymer that includes a base whose third carbon in its deoxyribose or ribose is bound to a hydroxyl group while being unbound to another base.

[0021] As used herein, the terms "5' end," "5-prime end," and their equivalents, may refer to a terminus of a single-stranded nucleotide polymer that includes a base whose fifth carbon in its deoxyribose or ribose ring is unbound to another base. In some cases, the fifth carbon is bound to a phosphate group.

[0022] As used herein, the "length" of a polymer refers to a number of covalently bonded monomers that are included in the polymer. For instance, the length of a DNA molecule may be the number of covalently bonded nucleotides in at least one strand of the DNA molecule and/or the number of base pairs in the DNA molecule. In various examples, the length of an RNA molecule may be the number of covalently bonded nucleotides in the RNA molecule.

[0023] As used herein, the term "gene," and its equivalents, refers to a sequence of DNA nucleotides that is transcribed into a functional RNA. The functional RNA, for instance, is RNA that is translated into a polypeptide or protein (e.g., mRNA) or that has some other biological function (e.g., miRNA, tRNA, etc.). A gene is "expressed" when it is used as a template to generate a functional RNA. A subject, for instance, has numerous genes contained in the subject's genome. A gene may include both introns and exons. As used herein, the term "intron," and its

equivalents, may refer to a subset of DNA nucleotides in a gene that is not used to code for any functional RNA that is expressed by the organism. As used herein, the term "exon," and its equivalents, may refer to a subset of DNA nucleotides in a gene that is used to code for a functional RNA. For instance, an exon may encode a polypeptide or protein that is expressed by the organism. In various examples, a gene can be represented in data (e.g., as data representative of the sequence of DNA nucleotides in the gene) or as a chemical structure (e.g., as the sequence of DNA nucleotides itself).

[0024] As used herein, the term "genome," and its equivalents, refers to the aggregate of genes of a subject. In various cases, a genome represents the sequences of several linear DNA molecules that are present in a subject's chromosomes. A "reference genome" refers to an aggregation of genes of one or more reference subjects. In various cases, a genome is represented in data.

[0025] As used herein, the terms "pangenome," "pan-genome," "supragenome," and their equivalents, refers to an aggregate set of genes from multiple subgroups (e.g., strains) within a population (e.g., a clade) of subjects. A pangenome, for example, indicates genes that are present in all subjects within the population, as well as genes that are present in some of the subjects of the population. A pangenome is represented in data, for instance.

[0026] As used herein, the term "transcriptome," and its equivalents, refers to the aggregate of RNA sequences of a subject. In some cases, a transcriptome is limited to mRNA sequences. In various examples, a transcriptome is represented in data.

[0027] As used herein, the term "genomic DNA," "gDNA," "chromosomal DNA," and their equivalents, may refer to DNA molecules that are obtained from a chromosome and/or nucleus of a cell.

[0028] As used herein, the terms "DNA fragment," "fragment," and their equivalents, may refer to DNA molecules that are excised and/or broken off from a larger DNA molecule.

[0029] As used herein, the terms "cell-free DNA," "cfDNA," and their equivalents, may refer to DNA fragments that are non-encapsulated and obtained outside of cells within a sample (e.g., a liquid biopsy sample).

[0030] As used herein, the terms "circulating tumor DNA," "ctDNA," and their equivalents, may refer to a cfDNA molecule that originates from a cancer cell.

[0031] As used herein, the terms "end motif," "terminal sequences," and their equivalents, may refer to a sequence of nucleotides extending from a 3' or 5' end of a DNA or RNA molecule. In various cases, the end motif is shorter than a length of the DNA or RNA molecule. For example, the end motif may have a length in a range of 5 to 30 bases or base pairs, a range of 3 to 30 bases or base pairs, or a range of 1 to 30 base pairs.

[0032] As used herein, the term "promoter," and its equivalents, may refer to a portion of a DNA molecule that binds one or more proteins in order to initiate transcription of a gene. For example, the promotor is located "upstream" of the gene. For example, the promotor is located between the 5' end of the DNA molecule and the gene. A promotor may include one or more binding sites for RNA polymerase, and/or one or more transcription factor binding sites. In some examples, a promotor includes one or more CpG islands. A promoter, for instance, includes a transcription start site.

[0033] As used herein, the terms "CpG island," "CGI," "CpG site," and their equivalents, may refer to a continuous portion of a DNA molecule whose sequence includes greater than a threshold amount (e.g., greater than 50%) of G-C base pairs.

[0034] As used herein, the term "enhancer," and its equivalents, may refer to a portion of a DNA molecule that binds one or more proteins in order to increase the chance that a gene will be transcribed. For instance, an enhancer includes one or more transcription factor binding sites. In various cases, an enhancer includes one or more CpG islands.

[0035] As used herein, the term "cancer," and its equivalents, may refer to a condition of a subject in which particular cells (referred to as "cancer cells") divide uncontrollably in the subject's body. In some cases, a cancer is characterized by a location or tissue type from which the cancer cells originated. In some examples, a cancer is characterized by a location or tissue type in which the cancer cells are located.

[0036] As used herein, the terms "tumor," "neoplasm," and their equivalents, may refer to a mass of tissue including cancer cells.

[0037] As used herein, the terms "tissue of origin," "tissue origin," and their equivalents, refers to a differentiated type of tissue from which cancer cells in the body of a subject began dividing uncontrollably in the subject's body.

[0038] As used herein, the terms "liquid biopsy," "fluid biopsy," and their equivalents, may refer to a process of obtaining a fluid sample from a subject's body. The sample, for instance, can be referred to as a "liquid biopsy sample." Examples of fluids that are sampled from the body include blood, plasma, cerebrospinal fluid, sputum, stool, urine, lymphatic fluid, and saliva.

[0039] As used herein, the term "tissue biopsy," and its equivalents, may refer to a process of obtaining a sample of cells from a subject's body. A tissue biopsy, in various cases, is performed by cutting a mass of cells from the subject's body. For instance, a tissue biopsy is a procedure performed by a surgeon, interventional radiologist, interventional cardiologist, or other specialized clinician. The term "tissue" or "tissue biopsy sample" can be used to refer to the sample of cells obtained using a tissue biopsy.

[0040] As used herein, the term "subject," and its equivalents, may refer to a human or non-human animal. A subject that is receiving care from at least one care provider may be referred to as a "patient."

[0041] As used herein, the term "variant," and its equivalents, may refer to a difference between a subject genetic sequence and a reference sequence. For instance, a variant may correspond to a difference between one or more nucleotides in a genome of a subject and one or more corresponding nucleotides in at least one reference genome or pangenome. A variant may be characterized by its identity (e.g., what nucleotides are different), its position (e.g., where are the nucleotides located in the genome, what chromosome contains the nucleotides, what gene contains the nucleotides, etc.), its length (e.g., how many nucleotides are different from the reference sequence), its type (e.g., substitution, insertion, deletion, copy number alternation, rearrangement of fusion, etc.), and other features that indicates its significance and/or relevance. In some cases, a variant represents any apparent alteration in a sequence that has been read from a nucleic acid molecule with respect to the reference sequence, such as reads cleaved by restriction enzymes (RE). In various examples, a variant can be represented in data (e.g., by data characterizing the variant) or as a chemical structure (e.g., the nucleotides themselves). As used herein, the term "mutation," and its equivalents, may refer to a change in a gene.

[0042] As used herein, the term "substitution," and its equivalents, can refer to a nucleotide in a subject sequence that is different than an equivalent nucleotide (e.g., a nucleotide at the same position) in a reference sequence.

[0043] As used herein, the term "insertion," and its equivalents, can refer to a nucleotide in a subject sequence that is added with respect to a reference sequence.

[0044] As used herein, the term "deletion," and its equivalents, can refer to the removal of a nucleotide from a nucleotide sequence.

[0045] As used herein, the terms "copy number alternation," "CNA," "copy number variation," "CNV," and their equivalents, can refer to a portion of a reference sequence that is repeated.

[0046] As used herein, the terms "rearrangement of fusion," "fusion rearrangement," "translocation," and their equivalents, can refer to a change in the relative position of one or more portions of a reference sequence, thereby generating a gene that was not present in the reference sequence.

[0047] As used herein, the term "sequencing," and its equivalents, may refer to a process of identifying the order and identity of monomers in a polymer chain, such as the order and identity of nucleotides in a DNA or RNA molecule. The terms "whole genome sequencing," "WGS," and their equivalents, may refer to the process of sequencing an entire genome of a subject, including the introns and exons of the genes of the subject. The term "whole exome sequencing," and its equivalents, may refer to the process of sequencing all exomes of a subject. The term "targeted sequencing," and its equivalents, may refer to the process of sequencing a portion of the genome of a subject, such as sequencing a single gene of the subject. For instance, targeted sequencing may entail detecting the presence and/or sequences of a panel of predetermined genes. Various techniques can be utilized to sequence a DNA or RNA molecule, such as massively parallel sequencing (MPS), nanopore sequencing, direct sequencing, Sanger sequencing, or next-generation sequencing. In various cases, sequencing is performed on physical molecules (e.g., RNA or DNA) and is used to generate data.

[0048] As used herein, the terms "next-generation sequencing," "NGS," and their equivalents, may refer to second-, third-, or greater generation nucleic acid sequencing techniques. Examples of NGS include, for instance, massive parallel sequencing and nanopore sequencing.

[0049] As used herein, the terms "massive parallel sequencing," "massively parallel sequencing," "MPS," and their equivalents, may refer to a technique for simultaneously performing multiple reactions that can be used to identify the order and identity of monomers in multiple polymer chains. In particular cases, massive parallel sequencing can be performed using sequencing-by-synthesis on clonally amplified DNA molecules that are located in spatially separated regions, which are individually monitored by sensors.

[0050] As used herein, the term "nanopore sequencing," and its equivalents, may refer to a technique for identifying the order and identity of monomers in a polymer chain by transporting the polymer chain from a first space to a second space, wherein the first space and the second space are separated by a substrate, by directing the polymer chain through a small hole (known as a "nanopore") embedded in the substrate, and monitoring a relative electrical signal (e.g., a voltage or current) between the first space and the second space.

[0051] As used herein, the term "sensor," and its equivalents, may refer to a physical device or other apparatus that is configured to detect one or more detection signals.

[0052] As used herein, the term "detection signal," and its equivalents, may refer to a physical signal that can be identified, characterized, or otherwise perceived by a sensor.

[0053] As used herein, the term "sequence read data," and its equivalents, may refer to data that is indicative of an order and identity of monomers in a polymer, such as the order and identity of nucleotides in a DNA or RNA sequence. In various implementations, sequence read data is generated via a sequencing operation.

[0054] As used herein, the term "image," and its equivalents, may refer to 2D or 3D array of data indicative of an array of pixels or voxels.

[0055] As used herein, the term "ligating," and its equivalents, may refer to a process of joining two molecules together, for example, with a chemical bond.

[0056] As used herein, the term "adapter," and its equivalents, may refer to an oligonucleotide that can be ligated to a target nucleic acid molecule. In various cases, an adapter prepares the target nucleic acid molecule for sequencing.

[0057] As used herein, the term "bait molecule," and its equivalents, may refer to a nucleic acid molecule having a region that is complementary to a region of a target molecule (e.g., cfDNA). A bait molecule includes, for instance, a nucleic acid molecule that can hybridize to (*i.e.*, is complementary to) a target molecule can be used to capture the target molecule. In some instances, the bait molecule is a capture oligonucleotide (or capture probe). In some instances, the bait molecule is suitable for solution phase hybridization to the target molecule. In some instances, the bait molecule is suitable for solid phase hybridization to the target molecule. In some instances, the bait molecule is suitable for both solution-phase and solid-phase hybridization to the target molecule. The design and construction of bait molecules is described in more detail in, *e.g.*, International Patent Application Publication No. WO 2020/236941.

[0058] As used herein, the term "amplifying," and its equivalents, may refer to a process of generating copies of a target molecule, such as a nucleic acid molecule.

[0059] As used herein, the term "hybridization," and its equivalents, may refer to a process by which to complementary single-stranded nucleic acid molecules bind to one another, thereby forming a double-stranded nucleic acid molecule. In certain examples, the double-stranded nature of the nucleic acid molecule is maintained under stringent hybridization conditions. Exemplary stringent hybridization conditions include an overnight incubation at 42 °C in a solution including 50% formamide, 5XSSC (750 mM NaCl, 75 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5XDenhardt's solution, 10% dextran sulfate, and 20 µg/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1XSSC at 50 °C.

[0060] As used herein, the term "complementary," and its equivalents, may refer to a state of two single-stranded nucleic acid molecules with respective sequences that cause the nucleic acid molecules to spontaneously hybridize to one another. One nucleic acid molecule, for instance, may have a sequence that causes each nucleic acid to hydrogen bond to a respective nucleic acid in the other nucleic acid molecule.

[0061] As used herein, the terms "therapy," "treatment," and their equivalents, may refer to a composition or process that can be used to remediate a health problem. Cancer therapies, for instance, include surgery, radiotherapy, chemotherapy, immunotherapy, cell-based therapies, and the like. Examples of cancer therapies include abemaciclib (Verzenio), abiraterone acetate (Zytiga), acalabrutinib (Calquence), ado-trastuzumab emtansine (Kadcyla), afatinib dimaleate (Gilotrif), aldesleukin (Proleukin), alectinib (Alecensa), alemtuzumab (Campath), alitretinoin (Panretin), alpelisib (Piqray), amivantamab-vmjw (Rybrevant), anastrozole (Arimidex), apalutamide (Erleada), asciminib hydrochloride (Scemblix), atezolizumab (Tecentriq), avapritinib (Ayvakit), avelumab (Bavencio), axicabtagene ciloleucel (Yescarta), axitinib (Inlyta), belantamab mafodotin-blmf (Blenrep), belimumab (Benlysta), belinostat (Beleodaq), belzutifan (Welireg), bevacizumab (Avastin), bexarotene (Targretin), binimetinib (Mektovi), blinatumomab (Blincyto), bortezomib (Velcade), bosutinib (Bosulif), brentuximab vedotin (Adcetris), brexucabtagene autoleucel (Tecartus), brigatinib (Alunbrig), cabazitaxel (Jevtana), cabozantinib (Cabometyx), cabozantinib

(Cabometyx, Cometriq), canakinumab (Ilaris), capmatinib hydrochloride (Tabrecta), carfilzomib (Kyprolis), cemiplimab-rwlc (Libtayo), ceritinib (LDK378/Zykadia), cetuximab (Erbitux), cobimetinib (Cotellic), copanlisib hydrochloride (Aliqopa), crizotinib (Xalkori), dabrafenib (Tafinlar), dacomitinib (Vizimpro), daratumumab (Darzalex), daratumumab and hyaluronidase-fihj (Darzalex Faspro), darolutamide (Nubeqa), dasatinib (Sprycel), denileukin diftitox (Ontak), denosumab (Xgeva), dinutuximab (Unituxin), dostarlimab-gxly (Jemperli), durvalumab (Imfinzi), duvelisib (Copiktra), elotuzumab (Empliciti), enasidenib mesylate (Idhifa), encorafenib (Braftovi), enfortumab vedotin-ejfv (Padcev), entrectinib (Rozlytrek), enzalutamide (Xtandi), erdafitinib (Balversa), erlotinib (Tarceva), everolimus (Afinitor), exemestane (Aromasin), fam-trastuzumab deruxtecan-nxki (Enhertu), fedratinib hydrochloride (Inrebic), fulvestrant (Faslodex), gefitinib (Iressa), gemtuzumab ozogamicin (Mylotarg), gilteritinib (Xospata), glasdegib maleate (Daurismo), hyaluronidase-zzxf (Phesgo), ibrutinib (Imbruvica), ibritumomab tiuxetan (Zevalin), idecabtagene vicleucel (Abecma), idelalisib (Zydelig), imatinib mesylate (Gleevec), infigratinib phosphate (Truseltiq), inotuzumab ozogamicin (Besponsa), iobenguane I131 (Azedra), ipilimumab (Yervoy), isatuximab-irfc (Sarclisa), ivosidenib (Tibsovo), ixazomib citrate (Ninlaro), lanreotide acetate (Somatuline Depot), lapatinib (Tykerb), larotrectinib sulfate (Vitrakvi), Lenvatinib mesylate (Lenvima), letrozole (Femara), lisocabtagene maraleucel (Breyanzi), loncastuximab tesirine-lpyl (Zynlonta), lorlatinib (Lorbrena), lutetium Lu 177-dotatate (Lutathera), margetuximabcmkb (Margenza), midostaurin (Rydapt), mobocertinib succinate (Exkivity), mogamulizumab-kpkc (Poteligeo), moxetumomab pasudotox-tdfk (Lumoxiti), naxitamab-gqgk (Danyelza), necitumumab (Portrazza), neratinib maleate (Nerlynx), nilotinib (Tasigna), niraparib tosylate monohydrate (Zejula), nivolumab (Opdivo), obinutuzumab (Gazyva), ofatumumab (Arzerra), olaparib (Lynparza), olaratumab (Lartruvo), osimertinib (Tagrisso), palbociclib (Ibrance), panitumumab (Vectibix), panobinostat (Farydak), pazopanib (Votrient), pembrolizumab (Keytruda), pemigatinib (Pemazyre), pertuzumab (Perjeta), pexidartinib hydrochloride (Turalio), polatuzumab vedotin-piiq (Polivy), ponatinib hydrochloride (Iclusig), pralatrexate (Folotyn), pralsetinib (Gavreto), radium 223 dichloride (Xofigo), ramucirumab (Cyramza), regorafenib (Stivarga), ribociclib (Kisqali), ripretinib (Qinlock), rituximab (Rituxan), rituximab and hyaluronidase human (Rituxan Hycela), romidepsin (Istodax), rucaparib camsylate (Rubraca), ruxolitinib phosphate (Jakafi), sacituzumab govitecanhziy (Trodelvy), seliciclib, selinexor (Xpovio), selpercatinib (Retevmo), selumetinib sulfate (Koselugo), siltuximab (Sylvant), sipuleucel-T (Provenge), sirolimus protein-bound particles (Fyarro), sonidegib (Odomzo), sorafenib (Nexavar), sotorasib (Lumakras), sunitinib (Sutent), tafasitamab-cxix (Monjuvi), tagraxofusp-erzs (Elzonris), talazoparib tosylate (Talzenna), tamoxifen (Nolvadex), tazemetostat hydrobromide (Tazverik), tebentafusp-tebn (Kimmtrak), temsirolimus (Torisel), tepotinib hydrochloride (Tepmetko), tisagenlecleucel (Kymriah), tisotumab vedotin-tftv (Tivdak), tocilizumab (Actemra), tofacitinib (Xeljanz), tositumomab (Bexxar), trametinib (Mekinist), trastuzumab (Herceptin), tretinoin (Vesanoid), tivozanib hydrochloride (Fotivda), toremifene (Fareston), tucatinib (Tukysa), umbralisib tosylate (Ukoniq), vandetanib (Caprelsa), vemurafenib (Zelboraf), venetoclax (Venclexta), vismodegib (Erivedge), vorinostat (Zolinza), zanubrutinib (Brukinsa), ziv-aflibercept (Zaltrap), and combinations thereof. Examples of cancer therapies also include targeted antibody-based therapies (antibody-drug conjugates, antibody-radioisotope conjugates, and targeted immune cell therapies (e.g., immune effector cells genetically modified to express a chimeric antigen receptor (CAR).

[0062] As used herein, the term “treatment-responsive,” and its equivalents, may refer to a type of cancer cells that can be substantially killed using a predetermined type of therapy. For example, cancer cells of a subject may be

responsive to a particular treatment if, after the subject is administered the treatment, the cancer cells are diminished by a particular progression level (e.g., radiographic progression level, marker-based progression level, such as prostate-specific antigen (PSA) progression, etc.). Accordingly, the responsiveness of the cells to the type of therapy may indicate the effectiveness of that therapy.

[0063] As used herein, the term "treatment-resistant," and its equivalents, may refer to a type of cancer that cannot be substantially killed using a predetermined type of therapy.

[0064] As used herein, the term "metastasis profile," and its equivalents, may refer to a propensity of a type of cancer to metastasize into one or more differentiated tumor types besides the cancer's tissue origin. In some implementations, the metastasis profile can further indicate the type of tissue in which the cancer can or is likely to metastasize.

[0065] As used herein, the term "clinical trial," and its equivalents, may refer to a research study used to evaluate a hypothesis based on participation by one or more subjects. In various examples, a clinical trial can be used to assess the efficacy and/or safety of a proposed therapy. A clinical trial may be performed in furtherance of approval of a treatment by a regulatory authority (e.g., the United States Food & Drug Administration (FDA)).

Description of Example Implementations

[0066] Various implementations of the present disclosure will now be described with reference to the accompanying Figures.

[0067] FIG. 1 illustrates an example environment 100 for efficiently distributing updated annotations of genetic variants. A subject 102, for instance, presents to a clinical environment with at least one symptom or characteristic associated with cancer. The subject 102 may be a patient of the clinical environment. In various cases, the subject 102 is a human, but implementations are not so limited. For instance, the subject 102 could be a non-human animal. In various implementations, the clinical environment includes a hospital, a health clinic, or a veterinary clinic.

[0068] A care provider 104, in various cases, obtains a sample 106 from the subject 102. In various implementations, the care provider 104 is a clinician responsible for caring for the subject 102. For instance, the care provider 104 is a physician, a physician's assistant, a nurse, a resident, a medical student, a medical technician, or the like.

[0069] The sample 106, in various cases, can be a liquid biopsy sample, a tissue biopsy sample, or a combination thereof. In some examples, the sample 106 is obtained by obtaining a fluid from the subject 102 in a liquid biopsy procedure. Examples of fluids in the sample 106 include blood, plasma, cerebrospinal fluid, sputum, stool, urine, lymphatic fluid, or saliva. For instance, the sample 106 is obtained by intravenously extracting blood and/or plasma from the subject 102. A liquid biopsy sample of the subject 102, for instance, may include circulating tumor cells (CTCs). In some cases, a tissue biopsy sample is obtained by physically excising a portion of a tissue of the subject 102. For instance, a portion of a lesion in the body of the subject 102 may be surgically removed from the subject 102. In some cases, the lesion is a tumor.

[0070] The sample 106 includes nucleic acid molecules 108. In various cases, the nucleic acid molecules 108 include DNA and/or RNA. In some cases, the nucleic acid molecules 108 are disposed inside of cells in the sample 106. For instance, the nucleic acid molecules 108 may include cellular DNA and/or RNA. In some examples, the nucleic acid molecules 108 are disposed outside of cells of the subject 102. For example, the nucleic acid molecules 108 may include cell-free DNA (cfDNA) and/or circulating tumor DNA (ctDNA). In some instances, the nucleic acid molecules 108 include mRNA, microRNA, non-coding RNA, functional RNA, or any combination thereof.

[0071] In various implementations, sequences of the nucleic acid molecules 108 in the sample 106 are indicative of a cancer of the subject 102. For example, the nucleic acid molecules 108 may include ctDNA released from cells in a tumor of the subject 102. In some examples, the nucleic acid molecules 108 include genomic DNA (gDNA) of the subject 102, which may include sequences indicative of a type of cancer that the subject 102 has developed or is predisposed toward.

[0072] In various implementations, the nucleic acid molecules 108 are sequenced by a sequencer 110. According to some examples, the nucleic acid molecules 108 are extracted from the sample 106. Various types of techniques can be used by the sequencer 110 to sequence the nucleic acid molecules 108. For example, the sequencer 110 may perform Sanger sequencing, second-generation sequencing (e.g., sequencing-by-synthesis), third-generation sequencing (e.g., nanopore sequencing), or any combination thereof. In some cases, the sequencer 110 identifies methylated cytosines in the nucleic acid molecules 108 using methylation sequencing (methyl-seq). In some examples, the sequencer 110 generates complementary DNA (cDNA) to RNA molecules in the nucleic acid molecules 108 and sequences the cDNA in order to identify the sequences of the RNA.

[0073] In some cases, the sequencer 110 divides the nucleic acid molecules 108 into fragments. The sequencer 110, in some instances, ligates adapters onto the fragments and/or nucleic acid molecules 108. For instance, the adapters include amplification primers, flow cell adaptor sequences, substrate adapter sequences, sample index sequences, or any combination thereof. In various examples, the sequencer 110 amplifies the fragments and/or nucleic acid molecules 108. For instance, the sequencer 110 performs a polymerase chain reaction (PCR) amplification technique, a non-PCR amplification technique, an isothermal amplification technique, or any combination thereof. In various cases, the sequencer 110 captures the amplified molecules onto a substrate and performs sequencing-by-synthesis in order to identify the sequences of the amplified molecules. For example, the sequencer 110 includes one or more photosensors configured to detect light signals emitted by fluorescently tagged nucleotide triphosphates (NTPs) that are added to nucleic acid molecules synthesized by DNA polymerase using the amplified molecules as templates. In various cases, the sequencer 110 includes one or more sensors that detect signals indicative of the sequences of the nucleic acid molecules 108.

[0074] The sequencer 110 outputs sequence read data 112 indicative of sequences of the nucleic acid molecules 108. In various cases, the sequence read data 112 indicates an identity and/or order of bases in the nucleic acid molecules 108. In some examples, the sequence read data 112 indicates the presence and/or location of methylated cytosines in the nucleic acid molecules 108.

[0075] A variant identifier 114 identifies the presence of a variant in the nucleic acid molecules 108 by analyzing the sequence read data 112. The variant identifier 114 may compare the sequences indicated in the sequence read data 112 to one or more reference sequences, such as one or more reference genomes and/or a pangenome. In various cases, the variant identifier 114 is configured to map the sequences indicated in the sequence read data 112 to regions in the reference sequence(s). For instance, the variant identifier 114 may determine that a sequence indicated in the sequence read data 112 is indicative of the subject's 102 copy of a predetermined gene. The variant identifier 114, in some cases, may identify one or more differences between a sequence in the sequence read data 112 and the reference sequence(s). In various cases, the difference(s) is representative of one or more variants indicated by the sequence read data 112.

[0076] The variant identifier 114 outputs an variant indication 116 based on the identified variant. The variant indication 116, for instance, may indicate a position and/or type of variant identified from the sequence read data 112. Examples of variants identified in the variant indication 116 include nucleotide substitutions, nucleotide additions, nucleotide deletions, structural variants, and copy number variants. The variant indication 116 may indicate a chromosome on which the variant is observed, a position on the chromosome where the variant was detected, a region (e.g., a gene, promoter, enhancer, transcription factor binding site, hotspot, etc.) where the variant was detected, or any combination thereof.

[0077] According to some examples, the variant indication 116 further includes one or more annotations. In some cases, the annotation(s) are prestored in the variant identifier 114. For example, the variant identifier 114 may store or otherwise access a table of annotations associated with various variants. The table, for instance, is indexed according to variant. When the variant identifier 114 identifies a specific variant in the sequence read data 112, the variant identifier 114 may identify an entry of the table corresponding to the identified variant, which may also include the annotation(s). In some cases, the annotations stored in the variant identifier 114 are received from one or more external devices. For instance, one or more clinical providers may input the annotations into the variant identifier 114. Accordingly, in some examples, the annotations may be provided as user input via an external device, for instance based on an analysis of the sequence read data 112 associated with the variant.

[0078] The annotation(s), for instance, indicate characteristics of the variant. In some cases, the annotation(s) indicates an association between the variant and at least one disease, such as at least one type of cancer. In some examples, the annotation(s) include an indication of an effective therapy for the disease(s), such as a treatment that the disease(s) is responsive to. In some examples, the annotation(s) indicate an expected progression of the disease(s) associated with the variant, such as a metastasis profile associated with the variant. In some examples, the annotation(s) include an identifier of the variant, such as a name that is used to refer to the variant in the medical or research community. In some cases, the annotation(s) include an association between the variant and a particular ancestry (e.g., whether the variant is typically associated with individuals having Ashkenazi heritage). The annotation(s), in some examples, indicate whether the variant is a germline variant. The annotation(s) may include a class identifier, indicating a class or category of the variant. In various cases, the annotation(s) include other information, such as custom notes associated with the variant.

[0079] In various cases, the variant indication 116 is provided to a report generator 118. The report generator 118 generates an original report 120 based on the variant indication 116. The original report 120 can be printed, output, transmitted over a network, displayed via a user interface, or otherwise provided to one or more end users or devices. As an example, the original report 120 can be provided to a clinical device 124. Based on reviewing the original report 120, the care provider 104 may diagnose and/or initiate a treatment of the subject 102. In some cases, the care provider 104 communicates a prognosis of the subject 102 based on the original report 120.

[0080] The original report 120, in various cases, includes an outdated annotation 122 corresponding to the identified variant. For instance, the outdated annotation 122 may indicate clinically acceptable information associated with the identified variant at the time of the original report 120 . However, in various cases, additional data related to the identified variant becomes available over time. For example, additional research studies, clinical trials, longitudinal studies, and other research associated with the subject 102 and/or other subjects in a population,

may indicate that the outdated annotation 122 is erroneous or incomplete. In some cases, it may be harmful to the subject 102 if the care provider 104 relies on the outdated annotation 122 to manage care of the subject 102 after additional information relevant to the variant of the subject 102 becomes available.

[0081] In various implementations of the present disclosure, a record system 126 is configured to track variants across a population. The record system 126 includes one or more population databases 128 that store various indications of variants of a population. In various cases, the population includes one or more individuals other than the subject 102. The population databases 128, for instance, store various information about variants of the individuals in the population. The population databases 128 may indicate the variants, as well as annotations associated with the variants. For instance, the variant identifier 114 may access the population databases 128 in order to identify annotations, such as the outdated annotation 122, to provide to the report generator 118.

[0082] In particular cases, the population database(s) 128 include a variant bin 130. For instance, the record system 126 may sort various records into collections of records (or "bins") according to variant. Thus, the variant bin 130 may correspond to the specific type of variant of the subject 102 that was identified by the variant identifier 114 and indicated in the original variant indication 116. The variant bin 130 includes *n* records (also referred to as "patient records") corresponding to the variant, where *n* is an integer greater than one. In some examples, the *n* records correspond to n individuals in the population that share the same type of variant.

[0083] In various cases, the record system 126 generates the first record among the *n* records based on the original variant indication 116. The record system 126, in various cases, generates additional records among the *n* records based on population variant indications 132. The population variant indications 132 may be received by the record system 126 from one or more external devices. In various cases, the population variant indications 132 identify variants observed from the population. The population variant indications 132 may further indicate annotations of the variants observed from the population.

[0084] The records stored in the variant bin 130 include first to nth timestamps 134-1 to 134-n. In various cases, the first to nth timestamps 134-1 to 134-n indicate times associated with the observed variant of the subject 102 and the population. For example, the first to nth timestamps 134-1 to 134-n indicate times at which the variant was observed in samples of the population (e.g., the time at which the variant identifier 114 identified the variant), times at which the variant was reported to the record system 126 (e.g., times at which the variant indication 116 and the population variant indications 132 were received by the record system 126), times at which the annotations were generated (e.g., times at which a study indicating that a variant is associated with a treatment-responsive therapy was published), times at which reports indicating the variants were output (e.g., a time at which the original report 120 was output to the clinical device 124), or any combination thereof.

[0085] Each of the n records also includes one or more annotations associated with the variant. For example, the first record, corresponding to the subject 102, includes the outdated annotation 122. In various cases, annotations stored in the variant bin 130 for additional records may be different than the outdated annotation 122.

[0086] In various implementations, the record system 126 may monitor the population database(s) 128 and/or the variant bin 130 in order to identify a consensus annotation 136. For instance, the record system 126 may review the annotations stored in the variant bin 130 at a predetermined frequency (e.g., every day, week, month, etc.) and/or in response to an event (e.g., in response to a new record being added to the variant bin 130). The nth record, for

example, includes the consensus annotation 136. The consensus annotation 136, in various implementations, includes an annotation that is shared by multiple records stored in the variant bin 130. In particular cases, the consensus annotation 136 is stored in greater than a threshold number (e.g., m, wherein m is a positive integer) of records stored in the variant bin 130. The threshold number of records, for example, may be the mth most recent records stored in the variant bin 130. For instance, the record system 126 may identify the consensus annotation 136 among records whose associated timestamps 134-1 to 134-n indicate times later than a threshold time point. Thus, in some examples, the consensus annotation 136 may represent a commonly accepted annotation in a set of recent records due to recently accepted research. The consensus annotation 136 may accordingly differ from annotations, such as the outdated annotation 122, that are included in older records.

[0087] Once the consensus annotation 136 is identified, the record system 126 is configured to update annotations in the variant bin 130 based on the consensus annotation 136. In some examples, the record system 126 rewrites annotations in the variant bin 130 that are not the consensus annotation 136. For example, the record system 126 may replace the outdated annotation 122 in the first record with the consensus annotation 136. In some cases, the record system 126 appends annotations in the variant bin 130 that do not include the consensus annotation 136. In some examples, the record system 126 may also, or alternately, append the consensus annotation 136 to records that already include the outdated annotation 122, such that the records include both the outdated annotation 122 and the consensus annotation 136. Accordingly, a user or system can have access to both the outdated annotation 122 and the newer appended consensus annotation 136 in a particular record when the user or system accesses that particular record, and can determine how the annotations were updated over time. In these examples, the record system 126 may flag the outdated annotation 122 as being an older annotation, and flag the appended consensus annotation 136 as being a current annotation.

[0088] In various cases, the record system 126 may further output an indication of the consensus annotation 136. For instance, in response to modifying the first record to include the consensus annotation 136, the record system 126 may transmit an indication of the consensus annotation 136 to the report generator 118.

[0089] The report generator 118, in various cases, may generate an updated report 138 based on the consensus annotation 136. The report generator 118 can be printed, output, transmitted over a network, displayed via a user interface, or otherwise provided to one or more end users or devices. For example, the report generator 118 may output the updated report 138 to the clinical device 124, or some other clinical deice associated with the subject 102 and/or the care provider 104. Accordingly, the subject 102 and/or care provider 104 may update a treatment, diagnosis, prognosis, or any combination thereof, based on the consensus annotation 136.

[0090] FIG. 1 illustrates various elements that can be embodied in one or more computing devices. For example, at least a portion of the functions of the sequencer 110, the variant identifier 114, the report generator 118, the clinical device 124, the record system 126, the population database(s) 128, the variant bin 130, or any combination thereof, are performed by one or more processors in at least one computing device. Examples of computing devices include server computers, desktop computers, laptop computers, tablet computers, mobile phones, wearable devices, Internet of Things (IoT) devices, and the like. In various cases, instructions for performing at least a portion of the functions of these elements are stored in memory and/or in a non-transitory computer readable medium. The instructions, for instance, are executed by the processor(s).

[0091] FIG. 1 also illustrates various types of data. For example, the sequence read data 112, the original variant indication 116, the original report 120, the outdated annotation 122, the variant bin 130, the population variant indications 132, the first to nth timestamps 134-1 to 134-n, the consensus annotation 136, the updated report 138, or any combination thereof, includes data. The various types of data illustrated in FIG. 1 may be stored, such as in memory or in non-transitory computer readable media. In various implementations, at least a portion of the data is transmitted or otherwise output by one or more computing devices. For example, a computing device may transmit one or more communication signals to another computing device, wherein the communication signal(s) encode at least a portion of the data. Examples of communication signals include electromagnetic signals, optical signals, ultrasonic signals, optical signals, and electrical signals. For example, communication signals can be transmitted wirelessly and/or in a wired fashion. The communication signals, for instance, are transmitted over one or more wireless channels and/or one or more wired channels (e.g., optical cabling, electrical cabling, etc.). In various cases, the communication signal(s) are transmitted over one or more communication networks. A communication network, for instance, may be defined according to one or more physical channels, such as one or more frequency spectra. In some cases, a communication network is defined according to one or more communication protocols and/or standards. Examples of communication networks include fiber optic networks, Institute of Electrical and Electronics Engineers (IEEE) networks (e.g., WI-FI™ networks, WiMAX networks, BLUETOOTH™ networks, etc.), cellular networks (e.g., a 3rd Generation Partnership Project (3GPP) radio network, such as a Long Term Evolution (LTE) network, a New Radio (NR) network; or a cellular core network such as a 3rd Generation (3G) core, a 4th Generation (4G) core, a 5th Generation (5G) core, etc.), ultrasonic networks, and the like. In some cases, the data is broadcasted from one device to multiple other devices. In some cases, the data is unicasted from one device to another device. For instance, various forms of data described herein may be transmitted via a peer-to-peer (P2P) connection.

[0092] A particular example will now be described with reference to FIG. 1. In this example, the subject 102 may present to the care provider 104 with breast cancer. The care provider 104 may order, or perform, a tissue biopsy procedure on a breast tumor of the subject 102 in order to obtain the sample 106. The nucleic acid molecules 108 may be obtained from cancer cells in the sample 106 of the tumor. In various cases, the sequencer 110 generates sequence read data 112 indicating sequences of the nucleic acid molecules 108. The variant identifier 114 may identify, based on the sequence read data 112, that the cancer cells in the tumor of the subject 102 have a genetic variant associated with HER2 negative status. In some cases, the variant identifier 114 further determines, with reference to one or more databases, that the genetic variant identified is associated with cancers that are resistant to a first immunotherapy. The report generator 118 may indicate, in the original report 120, the genetic variant. The outdated annotation 122 my indicate that the genetic variant is associated with cancers that are resistant to the first immunotherapy. The care provider 104 may review the original report 120 and recommend a chemotherapy regimen for the subject 102 based on the original report 120.

[0093] Meanwhile, the record system 126 further stores the outdated annotation 122 in a first record of the variant bin 130 associated with the subject 102, wherein the variant bin 130 stores records associated with the genetic variant associated with HER2 negative status. The first record may further include the first timestamp 134-1, which indicates the time at which the original report 120 was output to the clinical device 124. In the weeks following the time indicated in the first timestamp 134-1, a clinical trial may demonstrate the effectiveness of a second

immunotherapy on cancers associated with the genetic variant associated with the variant bin 130. Additional records are added to the variant bin 130, most of which may indicate that the variant is associated with cancers that are responsive to the second immunotherapy. At some point, the record system 126 may identify the consensus annotation 136 indicating the responsiveness of the second immunotherapy. Upon identifying the consensus annotation 136, the record system 126 may edit the first record (associated with the subject 102) to include the consensus annotation 136.

[0094] In addition, the record system 126 may output an indication of the consensus annotation 136 to the report generator 118. In response, the report generator 118 may generate and output the updated report 138 that includes the consensus annotation 136 indicating the responsiveness of the second immunotherapy. In some examples, the clinical device 124 outputs a push notification indicating the updated report 138 to the care provider 104. Upon selecting the push notification, the care provider 104 may view the updated report 138. Accordingly, the care provider 104 may update the treatment regimen of the subject 102 to include the second immunotherapy, which may result in superior outcomes for the subject 102.

[0095] FIG. 2 illustrates example records 200 stored in at least one patient database. For example, the example records 200 include various records stored in the population database(s) 128 described above with reference to FIG. 1. In this example, thirty-four records are included in the example records 200. In various cases, a record system (e.g., the record system 126) is configured to identify a consensus annotation based on the example records 200.

[0096] The example records 200 include timestamps 202 associated with each observed variant. In various cases, the timestamps 202 indicate the date and time at which reports on the observed variants were output.

[0097] The example records 200 further identify the observed variants by indicating positions 204 and types 206 of the observed variants. In this example, the example records 200 indicate variants at the same position, but implementations are not so limited. In various cases, multiple types 206 of variants are indicated in the example records 200. For instance, record six and record twenty-two are C>G variants, whereas the rest of the records are C>T variants. In various implementations, the variants are binned. For example, records six and record twenty-two are omitted from a bin corresponding to the C>T variant.

[0098] The record system may be configured to identify the consensus annotation based on annotations 208 of the portion of the example records 200 associated with the bin. For example, multiple early records among the example records 200 include annotations 208 indicating that the variant is associated with cancers that are “resistant to slugimib.” However, the majority of the later records occurring after September 1, 2021 include annotations 208 indicating that the variant is associated with cancers that are “resistant to slugimib” and “responsive to titanamib.” In various cases, the record system determines that greater than a threshold number (e.g., ten) records with timestamps after a threshold point in time (e.g., September 1, 2021) include the same annotation. The record system may therefore determine that the same annotation is a consensus annotation.

[0099] Upon identify the consensus annotation, the record system may update other records among the example records 200 to indicate the consensus annotation. For instance, the record system may rewrite at least one of the annotations 208 in records 1-5, 7-11, 21, or 23-25 to include the consensus annotation. In some cases, the record system further updates the timestamps 202 to indicate the time at which the example records 200 were rewritten to include the consensus annotation.

[0100] FIG. 3 illustrates an example report 300 summarizing predicted categories of a cancer of a subject. In various cases, the report 300 is the original report 120 or the updated report 138 described above with reference to FIG. 1. The report 300, for instance, may be displayed to a patient and/or care provider. In some cases, the report 300 is generated based on features of a sample (e.g., a liquid biopsy sample) obtained from the subject.

[0101] The report 300 includes a tissue origin 302 of the cancer. The tissue origin 302, for instance, indicates a histological tissue type 304, a primary site 306, cell subtype 307, or any combination, of the cancer.

[0102] In various cases, the report 300 includes one or more therapy indicators 308. For instance, the therapy indicator(s) 308 convey whether the cancer is predicted to be resistant to one or more predetermined therapies and/or whether the cancer is predicted to be responsive to one or more predetermined therapies.

[0103] In some examples, the report 300 includes one or more prognostic indicators 310. The prognostic indicator(s) 310, for instance, indicate a prognosis of the subject in view of the categorized cancer. For example, the prognostic indicator(s) 310 may indicate a survivability, a recoverability, a quality of life indicator, or other information indicative of the prognosis of the subject.

[0104] The report 300 may include a trial qualification 312 of the subject. The trial qualification 312, for instance, indicates whether the subject is predicted to qualify for a predetermined clinical trial.

[0105] The report 300, in various implementations, includes a metastasis profile 314 of the subject. The metastasis profile 314, for instance, indicates a likelihood that the cancer will metastasize (e.g., at a particular point in time), one or more tissues in which the cancer is predicted to metastasize, or the like.

[0106] In various cases, the report 300 includes recommended follow-up tests 316. For example, the report 300 may include a recommendation to perform whole genome sequencing on the subject, particularly in cases if the cancer cannot be categorized above a threshold certainty.

[0107] The report 300 may include a genomic profile 318 of the subject. In various cases, the genomic profile 318 includes or is generated based on the results of genetic analyses of the subject.

[0108] FIG. 4 illustrates an example process 400 for updating annotations in patient records. The process 400, for instance, is performed by an entity that includes at least one processor, at least one database, at least one computing device, the record system 126, or any combination thereof.

[0109] At 402, the entity identifies a consensus annotation of a genetic variant of a first set of population records. In various cases, the consensus annotation is shared by the first set of population records. According to some examples, the consensus annotation is shared by greater than a threshold number of the first set of population records.

[0110] The genetic variant, in various cases, may include nucleotide substitution, a nucleotide addition, a nucleotide deletion, a structural variant, or a number of copies of a predetermined nucleotide sequence. In some cases, the entity groups, among population records indicating multiple types of genetic variants, population records indicating the genetic variant associated with the first set of population records. For instance, the population records indicate their respective genetic variants, and the entity bins the population records by genetic variant. In some cases, the population records indicate at least one of a chromosome where the genetic variant is detected, a position on the chromosome where the genetic variant is detected, a type of the genetic variant, or any combination thereof.

[0111] At 404, the entity updates a second set of population records based on the consensus annotation. In various cases, the second set of population records correspond to the same genetic variant. For instance, the

second set of population records are grouped with the first set of population records. However, the second set of population records may have excluded the consensus annotation prior to performance of 404.

[0112] In some examples, the second set of population records was generated before the first set of population records. For instance, the first set of population records and the second set of population records include respective timestamps, wherein the timestamps of the first set of population records indicate later times than the timestamps of the second set of population records. In various implementations, the first set of population records are associated with times after a threshold time. For instance, the second set of population records are associated with at least one time before the threshold time.

[0113] At 406, the entity generates a report of a subject corresponding to the second set of population records based on the consensus annotation. For instance, the subject corresponds to one of the second set of population records. The report, in various cases, indicates the consensus annotation. In some examples, the report is output to at least one external device. In various implementations, the second set of population records can be updated without the use of a linked database.

[0114] FIG. 5 illustrates an example environment 500 for sequencing various nucleic acid molecules 502. In various implementations, the nucleic acid molecules 502 include cfDNA and/or gDNA. For instance, the nucleic acid molecules 502 may include ctDNA. The nucleic acid molecules 502, in various cases, are extracted from a sample, such as a biological sample obtained from a subject. In some implementations, the nucleic acid molecules 502 include DNA that is complementary to RNA present in the sample.

[0115] The nucleic acid molecules 502, in various cases, are ligated with adapters 504. For examples, the adapters 504 are hybridized to the nucleic acid molecules 502. The adapters 504, for example, include additional nucleic acid molecules. In various implementations, the adapters 504 have a shorter length than the nucleic acid molecules 502 being sequenced. For instance, the adapters 504 include amplification primers, flow cell adapter sequences, substrate adapter sequences, or sample index sequences. Although FIG. 5 illustrates adapters 504 being ligated to one end of each of the nucleic acid molecules 502, implementations are not so limited. For example, the adapters 504 may be ligated to both ends of each of the nucleic acid molecules 502.

[0116] In various examples, the nucleic acid molecules 502 ligated with the adapters 504 are amplified in order to generate amplified molecules 506. Various amplification techniques can be performed. For instance, the amplified molecules 506 are generated using PCR, a non-PCR amplification technique, an isothermal amplification technique, or any combination thereof.

[0117] Amplified molecules 506 may be captured by bait molecules 510 and sequenced. In some implementations, the amplified molecules 506 are sequenced via sequencing-by-synthesis. In various cases, fluorescently tagged deoxyribonucleotide triphosphates (dNTP) 512 are utilized to synthesize a strand that is complementary to DNA strands bound to the substrate 508. When a dNTP 512 is added to the strand (e.g., by an enzyme), the dNTP 512 emits an optical signal 514. In various implementations, the frequency of the optical signal 514 is dependent on the type of dNTP 512 from which the optical signal 514 is emitted. By detecting the optical signals 514 as the strand is being synthesized, the sequence of the original nucleic acid molecules 502 can be derived.

[0118] In some implementations, the amplified molecules 506 are sequenced via nanopore sequencing. For instance, the amplified molecules 506 are directed through a nanopore 516 extending through a substrate 518. In

various cases, the amplified molecules 506 are negatively charged, such that they can be directed through the nanopore 516 by imposing an electrical field across the substrate 518. In various cases, the amplified molecules 506 and the nanopore 516 are in the presence of a charged solution. Thus, charged solutes traveling through the nanopore 516 can be monitored by reviewing an electrical signal (e.g., a current) sensed between electrodes 520 on either side of the substrate 518. As an amplified molecule 506 is directed through the nanopore 516, the individual bases within the amplified molecule 506 will block the nanopore 516, which may decrease the amount of charged solutes traveling through the nanopore 516 and consequently, the magnitude of the electrical signal detected by the electrodes 520. Each of the four types of bases within the amplified molecules 506, may block the nanopore 516 to a different extent. Therefore, the sequence of the nucleic acid molecules 502 can be derived by analyzing the measured electrical signal with respect to time as the amplified molecules 506 are directed through the nanopore 516.

[0119] FIG. 6 illustrates one or more devices 600 configured to perform various operations described herein. The device(s) 600 include one or more processor(s) 602. In some implementations, the processor(s) 602 includes a central processing unit (CPU), a graphics processing unit (GPU), both CPU and GPU, or other processing unit or component known in the art.

[0120] The processor(s) 602 is operably connected to memory 604. In various implementations, the memory 604 is volatile (such as random access memory (RAM)), non-volatile (such as read only memory (ROM), flash memory, etc.) or some combination of the two. The memory 604 stores instructions that, when executed by the processor(s) 602, causes the processor(s) 602 to perform various operations. In various examples, the memory 604 stores methods, threads, processes, applications, objects, modules, any other sort of executable instruction, or a combination thereof. In some cases, the memory 604 stores files, databases, or a combination thereof. In some examples, the memory 604 includes, but is not limited to, RAM, ROM, electrically erasable programmable read-only memory (EEPROM), flash memory, or any other memory technology. In some examples, the memory 604 includes one or more of CD-ROMs, digital versatile discs (DVDs), content-addressable memory (CAM), or other optical storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other (e.g., non-transitory) medium which can be used to store the desired information and which can be accessed by the processor(s) 602. For instance, the memory 604 stores instructions that, when executed by the processor(s) 602, causes the processor(s) 602 to perform operations of the variant identifier 114, the report generator 118, the record system 126, or any combination thereof.

[0121] The processor(s) 602 is operably connected to one or more input devices 606 and one or more output devices 608. Collectively, the input device(s) 606 and the output device(s) 608 function as an interface between at least one user and the device(s) 600. The input device(s) 606 is configured to receive an input from a user and includes at least one of a keypad, a cursor control, a touch-sensitive display, a voice input device (e.g., a microphone), a haptic feedback device (e.g., a gyroscope), or any combination thereof. The output device(s) 608 includes at least one of a display, a speaker, a haptic output device, a printer, or any combination thereof. In various examples, the processor(s) 602 causes a display among the input device(s) 606 to visually output various data described herein. In some implementations, the input device(s) 606 includes one or more touch sensors, the output device(s) 608 includes a display screen, and the touch sensor(s) are integrated with the display screen.

[0122] In various implementations, the processor(s) 602 is operably connected to one or more transceivers 610 that transmit and/or receive data over one or more communication networks 612. For example, the transceiver(s)

610 includes a network interface card (NIC), a network adapter, a local area network (LAN) adapter, or a physical, virtual, or logical address to connect to the various external devices and/or systems. In various examples, the transceiver(s) 610 includes any sort of wireless transceivers capable of engaging in wireless communication (e.g., radio frequency (RF) communication). For example, the communication network(s) 612 includes one or more wireless networks that include a 3rd Generation Partnership Project (3GPP) network, such as a Long Term Evolution (LTE) radio access network (RAN) (e.g., over one or more LTE bands), a New Radio (NR) RAN (e.g., over one or more NR bands), or a combination thereof. In some cases, the transceiver(s) 610 includes other wireless modems, such as a modem for engaging in WI-FI®, WIGIG®, WIMAX®, BLUETOOTH®, or infrared communication over the communication network(s) 612.

[0123] The device(s) 600 may further include the sequencer 110. In various implementations, the sequencer 110 includes one or more fluidic circuits 614 configured to receive a sample 616 derived from a subject 618. The sequencer 110, in various cases, may be configured to generate data indicative of one or more sequences of nucleic acid molecules (e.g., DNA and/or RNA) present in the sample 616. In various cases, the sequencer 110 introduces one or more reagents 619 to the fluidic circuit(s) 614 in order to prepare for and perform sequencing of the nucleic acid molecules. Further, the sequencer 110 may include one or more sensors 620 configured to measure or otherwise detect detection signals from the fluidic circuit(s) 614, which may be indicative of the sequences of the nucleic acid molecules. According to various implementations, the sensor(s) 620 may further include one or more ADCs. The sequencer 110, in various cases, outputs sequence read data to the processor(s) 602 for additional processing.

Example Clauses

1. A method, including: providing a plurality of nucleic acid molecules obtained from a sample from a subject in a population; ligating one or more adapters onto one or more nucleic acid molecules from the plurality of nucleic acid molecules; amplifying the one or more ligated nucleic acid molecules from the plurality of nucleic acid molecules; capturing all or a subset of the amplified nucleic acid molecules; sequencing, by a sequencer, the captured nucleic acid molecules to obtain a plurality of sequence reads that represent the captured nucleic acid molecules; receiving, at one or more processors, sequence read data for the plurality of sequence reads; identifying, using the one or more processors, a genetic variant by comparing the sequence read data to a reference sequence; generating, using the one or more processors, an example record including an indication of the genetic variant and an example annotation indicating information associated with the genetic variant; storing the example record in one or more databases; identifying, using the one or more processors, population records, corresponding to the genetic variant, from the one or more databases, wherein the population records include the example record, and wherein individual population records, in the population records, are associated with annotations indicating information associated with the genetic variant; determining, using the one or more processors, that a first set of the population records include at least a threshold number of population records; identifying, using the one or more processors and in response to determining that the first set of the population records includes at least the threshold number of population records, a consensus annotation associated with the genetic variant that is indicated by the first set of the population records; and adding the consensus annotation to a second set of the population records, the second set of the population records being different from the first set of the population records.

2. The method of clause 1, wherein: timestamp data indicates that the first set of the population records was generated after the second set of the population records, the first set of the population records includes a different annotation than the second set of the population records, and the one or more processors identify the different annotation from the first set of the population records as the consensus annotation.

3. The method of clause 1 or 2, wherein the second set of the population records includes at least one population record that is also within the first set of the population records.

4. The method of any of clauses 1 to 3, wherein identifying the population records corresponding to the genetic variant includes: identifying, using the one or more processors, a set of database records that: correspond to different genetic variants, and indicate variant attributes of the different genetic variants; grouping, using the one or more processors, the set of database records into a plurality of bins based on one or more of the variant attributes; and selecting, using the one or more processors, a group of database records associated with a bin, of the plurality of bins, as the population records corresponding to the genetic variant.

5. The method of any of clauses 1 to 4, wherein the second set of the population records includes the example record, the method further including: in response to identifying the genetic variant, causing, using the one or more processors, transmission of an original report to an external device associated with the subject or a care provider, the original report indicating the genetic variant and an original annotation associated with the genetic variant; and in response to adding the consensus annotation to the second set of the population records, causing, using the one or more processors, transmission of an updated report to the external device, the updated report indicating the genetic variant and the consensus annotation.

6. A method, including: identifying, using one or more processors, population records of a genetic variant shared by samples of a population of subjects, the population records respectively including indications of the genetic variant and annotations of the genetic variant; determining, using the one or more processors, that a first set of the population records share a consensus annotation among the annotations, the first set of the population records corresponding to greater than a threshold number of subjects in the population; in response to determining that the first set of the population records includes the consensus annotation, updating, using the one or more processors, a second set of the population records to include the consensus annotation; and in response to updating the second set of the population records: identifying, using the one or more processors, an example subject among the subjects corresponding to an example population record among the second set of population records; and generating, using the one or more processors, a report based on the genetic variant and the consensus annotation.

7. The method of clause 6, wherein identifying the population records of the genetic variant shared by the samples of the population of subjects includes: identifying, using the one or more processors, sequence read data of an example sample obtained from the example subject; identifying, using the one or more processors, the genetic variant by analyzing the sequence read data; and generating, using the one or more processors, the example population record indicating the genetic variant and an annotation of the genetic variant.

8. The method of clause 7, wherein identifying the sequence read data includes: providing a plurality of nucleic acid molecules obtained from the example sample; ligating one or more adapters onto one or more nucleic acid molecules from the plurality of nucleic acid molecules; amplifying the one or more ligated nucleic acid molecules from the plurality of nucleic acid molecules; capturing all or a subset of the amplified nucleic acid molecules; and

sequencing, by a sequencer, the captured nucleic acid molecules to obtain a plurality of sequence reads that represent the captured nucleic acid molecules, thereby generating the sequence read data.

9. The method of clause 8, wherein the one or more adapters include amplification primers, flow cell adaptor sequences, substrate adapter sequences, or sample index sequences.

10. The method of clause 8 or 9, wherein the captured nucleic acid molecules are captured from the amplified nucleic acid molecules by hybridization to one or more bait molecules.

11. The method of clause 10, wherein the one or more bait molecules include one or more nucleic acid molecules, each including a region that is complementary to a region of a captured nucleic acid molecule.

12. The method of any of clauses 8 to 11, wherein amplifying the one or more ligated nucleic acid molecules includes performing a polymerase chain reaction (PCR) amplification technique, a non-PCR amplification technique, or an isothermal amplification technique.

13. The method of any of clauses 8 to 12, wherein sequencing the captured nucleic acid molecules includes use of a massively parallel sequencing (MPS) technique, whole genome sequencing (WGS), whole exome sequencing, targeted sequencing, direct sequencing, or Sanger sequencing.

14. The method of any of clauses 8 to 13, wherein sequencing the captured nucleic acid molecules includes next generation sequencing (NGS).

15. The method of any of clauses 8 to 14, wherein the sequencer includes a next generation sequencer.

16. The method of any of clauses 8 to 15, wherein sequencing the captured nucleic acid molecules includes sequencing-by-synthesis or nanopore sequencing.

17. The method of any of clauses 7 to 16, further including: generating ligated molecules by ligating adaptors onto nucleic acid molecules of the example sample; generating amplified ligated molecules by amplifying the ligated molecules; generating, using the amplified ligated molecules, detection signals; detecting, by at least one sensor, the detection signals; and generating, using the one or more processors, the sequence read data based on the detection signals.

18. The method of clause 17, wherein the detection signals include electrical signals and/or optical signals.

19. The method of clause 17 or 18, wherein generating, using the amplified ligated molecules, the detection signals includes simultaneously: synthesizing, by a polymerase using fluorescently tagged nucleotide triphosphates (NTPs), a synthesized nucleic acid molecule that is complementary to one of the amplified ligated molecules, and wherein detecting, by the at least one sensor, the detection signals includes: detecting, by at least one optical sensor, optical signals emitted by the fluorescently tagged NTPs upon binding to the synthesized nucleic acid molecule.

20. The method of any of clauses 17 to 19, wherein generating, using the amplified ligated molecules, the detection signals includes simultaneously: directing the amplified ligated molecules through a nanopore extending from a first space to a second space through a substrate, and wherein detecting, by the at least one sensor, the detection signals includes: detecting, by sensors disposed in the first space and the second space, an electrical signal over time.

21. The method of any of clauses 17 to 20, wherein the sequence read data indicates a full genome or RNA transcriptome of the example sample.

22. The method of any of clauses 17 to 21, wherein the sequence read data indicates a whole exome of the example sample.

23. The method of any of clauses 17 to 22, wherein the sequence read data indicates a predetermined panel of genes of the of the example sample.

24. The method of any of clauses 17 to 23, further including: receiving, by the one or more processors, the example sample.

25. The method of clause 24, wherein the example sample includes a tissue biopsy sample, a liquid biopsy sample, or a normal control.

26. The method of clause 24 or 25, wherein the example sample is a liquid biopsy sample and includes blood, plasma, cerebrospinal fluid, sputum, stool, urine, lymphatic fluid, or saliva.

27. The method of clause 24, wherein the example sample is a liquid biopsy sample and includes circulating tumor cells (CTCs).

28. The method of any of clauses 24 to 27, wherein the example sample is a liquid biopsy sample and includes cell-free DNA (cfDNA), circulating tumor DNA (ctDNA), or any combination thereof.

29. The method of any of clauses 24 to 28, further including: extracting the nucleic acid molecules from the example sample, the nucleic acid molecules including DNA or RNA.

30. The method of clause 29, wherein the DNA includes genomic DNA or cDNA.

31. The method of clause 29 or 30, wherein the RNA includes messenger RNA, microRNA, or non-coding RNA.

32. The method of any of clauses 17 to 31, further including: generating, using the one or more processors, and based on the sequence read data, a collection of subject-specific population records indicating genetic variants of the example sample, the collection of subject-specific population records including the example population record, the genetic variants including the genetic variant; and storing the collection of subject-specific population records in one or more databases.

33. The method of any of clauses 6 to 32, further including: outputting, by the one or more processors, the report to an external device associated with at least one of the example subject or a care provider of the example subject.

34. The method of any of clauses 6 to 33, further including: displaying, by the one or more processors, the report via a user interface of a digital system.

35. The method of any of clauses 6 to 34, further including: causing, by the one or more processors, the report to be printed into a printed copy of the report; and causing, by the one or more processors, the printed copy of the report to be disseminated to at least one of the example subject or a care provider of the example subject.

36. The method of any of clauses 6 to 35, wherein: the first set of the population records was generated after the second set of the population records, the first set of the population records includes a different annotation than the second set of the population records, and the method further includes identifying, using one or more processors, the different annotation from the first set of the population records as the consensus annotation.

37. The method of clause 36, further including: determining, using the one or more processors, that the first set of the population records was generated after the second set of the population records based at least in part on timestamp data indicating times at which the population records, in the first set of the population records and the second set of the population records, were generated.

38. The method of clause 36 or 37, wherein determining that the first set of the population records share the

consensus annotation among the annotations includes: ordering, using the one or more processors, the population records into a temporal order based on timestamps indicating times at which the population records were generated, wherein instances of the timestamps associated with the first set of the population records are recent timestamps, after a threshold time, in the temporal order.

39. The method of clause 38, wherein the second set of the population records have second instances of the timestamps before the threshold time.

40. The method of any of clauses 6 to 39, the genetic variant being a particular genetic variant among a plurality of genetic variants, wherein identifying the population records of the genetic variant shared by the samples of the population of subjects includes: grouping, using the one or more processors, a plurality of records of the plurality of genetic variants of the samples into one or more bins respectively corresponding to the plurality of the genetic variants, an example bin among the one or more bins including the population records corresponding to the particular genetic variant.

41. The method of any of clauses 6 to 40, wherein the example population record includes: an identity field indicating at least one of a chromosome where the genetic variant is detected, a position on the chromosome where the genetic variant is detected, or a type of the genetic variant.

42. The method of clause 41, wherein the type of the genetic variant includes a nucleotide substitution, a nucleotide addition, a nucleotide deletion, a structural variant, or a number of copies of a predetermined nucleotide sequence.

43. The method of any of clauses 6 to 42, wherein the annotations include at least one of: an association between the genetic variant and at least one disease, an indication of an effective therapy for the at least one disease, disease progression information associated with the at least one disease, an identifier of the genetic variant, an ancestry association, an indication that the genetic variant is a germline variant, a class identifier of a class or category of the variant, or a custom annotation type.

44. The method of any of clauses 6 to 43, wherein the second set of the population records includes at least one population record that is also within the first set of the population records.

45. The method of any of clauses 6 to 44, wherein updating the second set of the population records to include the consensus annotation includes: adding, using the one or more processors, the consensus annotation to the example population record.

46. The method of any of clauses 6 to 45, wherein updating the second set of the population records to include the consensus annotation includes: replacing, using the one or more processors, an outdated annotation in the example population record with the consensus annotation.

47. The method of any of clauses 6 to 46, further including: adding or updating, using the one or more processors, timestamp data in the second set of the population records to indicate a time at which the second set of the population records was updated to include the consensus annotation.

48. The method of any of clauses 6 to 47, the report being an updated report, the method further including: identifying, using the one or more processors, an outdated annotation in the example population record; and generating, using the one or more processors, an original report indicating the genetic variant and the outdated annotation, wherein updating the second set of the population records is performed after generating the original report.

49. The method of any of clauses 6 to 48, wherein the population records are stored in one or more non-linked databases.

50. The method of any of clauses 6 to 49, wherein the population records are stored in respective databases.

51. The method of any of clauses 6 to 50, wherein generating the report based on the genetic variant and the consensus annotation includes: identifying, using the one or more processors, and based on the consensus annotation, a recommended therapy to treat a pathology of the example subject, the report indicating the recommended therapy.

52. The method of any of clauses 6 to 51, wherein generating the report based on the genetic variant and the consensus annotation includes: determining, using the one or more processors, and based on the consensus annotation, whether the example subject satisfies one or more criteria for a clinical trial, the report indicating whether the example subject satisfies the one or more criteria.

53. The method of any of clauses 6 to 52, wherein: the annotations include a plurality of annotation fields, the consensus annotation is associated with a first annotation field of the plurality of annotation fields, and updating the second set of the population records to include the consensus annotation includes updating, using the one or more processors, instances of the first annotation field, associated with the second set of the population records, to indicate the consensus annotation.

54. The method of any of clauses 6 to 53, wherein individual population records, of the population records, respectively correspond to individual subjects within the population of subjects.

55. The method of any of clauses 6 to 54, wherein a timestamp associated with the example population record includes: a time field indicating a time at which information was added to an annotation of the example population record.

56. The method of any of clauses 6 to 55, wherein updating the second set of population records to include the consensus annotation includes: appending, using the one or more processors, the consensus annotation to an annotation field of the example population record.

57. The method of any of clauses 6 to 56, wherein updating the second set of population records to include the consensus annotation includes: overwriting, using the one or more processors, existing data, within an annotation field of the example population record, with the consensus annotation.

58. The method of any of clauses 6 to 57, wherein updating the second set of population records to include the consensus annotation includes adding, using the one or more processors, the consensus annotation to the example population record, and the method further includes: generating, using the one or more processors, a second report indicating prognostic information of an individual based at least in part on the consensus annotation added to the example population record.

59. The method of any of clauses 6 to 58, further including: generating, using the one or more processors, the population records of the genetic variant.

60. A system, including: at least one processor; and memory storing instructions that, when executed by the at least one processor, cause the at least one processor to perform operations including: the method of any of clauses 6 to 59.

61. A non-transitory computer-readable medium storing instructions for performing operations that include: the method of any of clauses 6 to 59.

Conclusion

[0124] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference in their entirety to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference in its entirety. In the event of a conflict between a term herein and a term in an incorporated reference, the term herein controls.

[0125] The features disclosed in the foregoing description, or the following claims, or the accompanying drawings, expressed in their specific forms or in terms of a means for performing the disclosed function, or a method or process for attaining the disclosed result, as appropriate, may, separately, or in any combination of such features, be used for realizing implementations of the disclosure in diverse forms thereof.

[0126] As will be understood by one of ordinary skill in the art, each implementation disclosed herein can comprise, consist essentially of or consist of its particular stated element, step, or component. Thus, the terms "include" or "including" should be interpreted to recite: "comprise, consist of, or consist essentially of." The transition term "comprise" or "comprises" means has, but is not limited to, and allows for the inclusion of unspecified elements, steps, ingredients, or components, even in major amounts. The transitional phrase "consisting of" excludes any element, step, ingredient or component not specified. The transition phrase "consisting essentially of" limits the scope of the implementation to the specified elements, steps, ingredients or components and to those that do not materially affect the implementation. As used herein, the term "based on" is equivalent to "based at least partly on," unless otherwise specified.

[0127] Unless otherwise indicated, all numbers expressing quantities, properties, conditions, and so forth used in the specification and claims are to be understood as being modified in all instances by the term "about." Accordingly, unless indicated to the contrary, the numerical parameters set forth in the specification and attached claims are approximations that may vary depending upon the desired properties sought to be obtained by the present disclosure. At the very least, and not as an attempt to limit the application of the doctrine of equivalents to the scope of the claims, each numerical parameter should at least be construed in light of the number of reported significant digits and by applying ordinary rounding techniques. When further clarity is required, the term "about" has the meaning reasonably ascribed to it by a person skilled in the art when used in conjunction with a stated numerical value or range, i.e., denoting somewhat more or somewhat less than the stated value or range, to within a range of ±20% of the stated value; ±19% of the stated value; ±18% of the stated value; ±17% of the stated value; ±16% of the stated value; ±15% of the stated value; ±14% of the stated value; ±13% of the stated value; ±12% of the stated value; ±11% of the stated value; ±10% of the stated value; ±9% of the stated value; ±8% of the stated value; ±7% of the stated value; ±6% of the stated value; ±5% of the stated value; ±4% of the stated value; ±3% of the stated value; ±2% of the stated value; or ±1% of the stated value.

[0128] Notwithstanding that the numerical ranges and parameters setting forth the broad scope of the disclosure are approximations, the numerical values set forth in the specific examples are reported as precisely as possible. Any numerical value, however, inherently contains certain errors necessarily resulting from the standard deviation found in their respective testing measurements.

[0129] The terms "a," "an," "the," and similar referents used in the context of describing implementations (especially in the context of the following claims) are to be construed to cover both the singular and the plural, unless otherwise

indicated herein or clearly contradicted by context. Recitation of ranges of values herein is merely intended to serve as a shorthand method of referring individually to each separate value falling within the range. Unless otherwise indicated herein, each individual value is incorporated into the specification as if it were individually recited herein. All methods described herein can be performed in any suitable order unless otherwise indicated herein or otherwise clearly contradicted by context. The use of any and all examples, or exemplary language (e.g., "such as") provided herein is intended merely to better illuminate implementations of the disclosure and does not pose a limitation on the scope of the disclosure. No language in the specification should be construed as indicating any non-claimed element essential to the practice of implementations of the disclosure.

[0130] Groupings of alternative elements or implementations disclosed herein are not to be construed as limitations. Each group member may be referred to and claimed individually or in any combination with other members of the group or other elements found herein. It is anticipated that one or more members of a group may be included in, or deleted from, a group for reasons of convenience and/or patentability. When any such inclusion or deletion occurs, the specification is deemed to contain the group as modified thus fulfilling the written description of all Markush groups used in the appended claims.

[0131] Unless otherwise indicated, the practice of the present disclosure can employ conventional techniques of immunology, molecular biology, microbiology, cell biology and recombinant DNA. These methods are described in the following publications. See, e.g., Sambrook, et al. Molecular Cloning: A Laboratory Manual, 2nd Edition (1989); F. M. Ausubel, et al. eds., Current Protocols in Molecular Biology, (1987); the series Methods IN Enzymology (Academic Press, Inc.); M. MacPherson, et al., PCR: A Practical Approach, IRL Press at Oxford University Press (1991); MacPherson et al., eds. PCR 2: Practical Approach, (1995); Harlow and Lane, eds. Antibodies, A Laboratory Manual, (1988); and R. I. Freshney, ed. Animal Cell Culture (1987).

[0132] Tumor mutational burden (TMB) is a measure of the number of mutations carried by tumor cells. By comparing DNA sequences from a patient's healthy tissues and tumor cells, the number of acquired somatic mutations present in tumors, but not in normal tissues, may be determined. In some instances, driver mutations may be excluded from a TMB calculation.

[0133] In certain examples, "tumor mutational burden" or "TMB" refers to the number of somatic mutations in a tumor's genome and/or the number of somatic mutations per area of the tumor's genome. In some embodiments, TMB, as used herein, refers to the number of somatic mutations per megabase (Mb) of DNA sequenced. In some embodiments, germline (inherited) variants are excluded when determining TMB, given that the immune system has a higher likelihood of recognizing these as self. In various cases, driver mutations are excluded from a TMB calculation.

[0134] Microsatellites are highly polymorphic DNA-repeat regions. In certain examples, "microsatellite" refers to a repetitive nucleic acid having repeat units of less than about 10 base pairs or nucleotides in length. In certain examples, a microsatellite refers to a tract of tandemly repeated (i.e. adjacent) DNA motifs ranging from one to six or up to ten nucleotides, with each motif repeated 5 to 50 repeated times. "Microsatellite instability" refers to genetic instability in the microsatellite regions. Cancer patients with microsatellite instability classified as being high (MSI-H or MSI-High) frequently exhibit an accumulation of somatic mutations in tumor cells that leads to a range of molecular and biological changes including high tumor mutational burden, increased expression of neoantigens and abundant tumor-infiltrating lymphocytes. Chang et al. "Microsatellite Instability: A Predictive Biomarker for Cancer Immunotherapy," Appl

Immunohistochem Mol Morphol, 26(2):e15-e21 (2018). These changes have been linked to increased sensitivity to checkpoint inhibitor drugs, such as pembrolizumab, which is used to treat advanced melanoma, head and neck squamous cell carcinoma, non-small cell lung cancer (NSCLC), and classical Hodgkin lymphoma.

[0135] A viral status test refers to a test that identifies the presence of viral RNA or DNA in a subject. The test can identify viral load and/or viral identity. For example, the viral status test can identify the presence of viral RNA or DNA associated with the occurrence of certain cancers. Examples of such viruses include Hepatitis B Virus (HBV) and Hepatitis C Virus (HCV), Kaposi Sarcoma-Associated Herpesvirus (KSHV), Merkel Cell Polyomavirus (MCV), Human Papillomavirus (HPV), Human Immunodeficiency Virus Type 1 (HIV-1, or HIV), Human T-Cell Lymphotropic Virus Type 1 (HTLV-1), and Epstein-Barr Virus (EBV).

[0136] Cancer "hotspot" mutations give rise to oncological outcomes. PhyloP, SIFT, Grantham, COSMIC and PolyPhen-2 are in silico tools that can be used to assess pathogenicity of identified variants. Exemplary hotspot genes and mutations include EGFR exon 19 activating mutation, EGFR exon 19 deletion, EGFR exon 19 insertion, EGFR exon 19 sensitizing mutation, EGFR exon 20 activation mutation, EGFR exon 20 insertion, EGFR G719 mutation, EGFR L858R mutation, EGFR L861 mutation, EGFR S768 mutation, EGFR T790M mutation, C797 mutation, KIT activating mutation, KRAS activating mutation, MET activating mutation, NRAS activating mutation, PMS2 promoter mutations, among many others. Hotspot mutations also occur in the following genes: AKT2, BRCA1, BRCA2, ERC1, NSD1, POLH, PPM1G, PTEN, RAD18, RAD51, RAD51B, RB1, TERT, TP53, TP53Bp1, ALK, ARMT1, ATAD5, ATG7, ATIC, AXL, BIRC6, BRD3, BRD4, CAPRIN1, CCAR2, CCDC6, CDK5RAP2, CHD9, CIT, CTNNB1, CUL1, EBF1, EIF3E, HIP1, HMGA2, IRF2BP2, NOTCH1, NOTCH4, NPM1, OFD1, TACC1, TACC3, TERF2, TMEM106B, UBE2L3, USP10, WRDR48, YAP1, ZEB2, and ZMYND8.

[0137] A "DNA methylation test" refers to an assay, which can be commercially available, for distinguishing methylated versus unmethylated cytosine loci in DNA. Techniques for measuring cytosine methylation include bisulfite-based methylation assays. The addition of bisulfite to DNA results in the methylation of unmethylated cytosine and its ultimate conversion to the nucleotide uracil. Uracil has similar binding properties to thiamine in the DNA sequence. Previously methylated cytosine does not undergo similar chemical conversion on exposure to bisulfite. Bisulfite assays can thus be used to discriminate previously methylated versus unmethylated cytosine.

[0138] An exemplary quantitative methylation detection assay combines bisulfite treatment and restriction analysis COBRA, which uses methylation sensitive restriction endonucleases, gel electrophoresis, and detection based on labeled hybridization probes. (Ziong and Laird, Nucleic Acid Res. 1997 25; 2532-4). Another exemplary detection assay is the methylation specific polymerase chain reaction PCR (MSPCR) for amplification of DNA segments of interest. This assay can be performed after sodium bisulfite conversion of cytosine and uses methylation sensitive probes. Other detection assays include the Quantitative Methylation (QM) assay, which combines PCR amplification with fluorescent probes designed to bind to putative methylation sites; MethyLight™ (Qiagen, Redwood City, CA) a quantitative methylation detection assay that uses fluorescence-based PCR (Eads, et al., Cancer Res. 1999; 59:2302-2306); and Ms-SNuPE, a quantitative technique for determining differences in methylation levels in CpG sites. As with other techniques, Ms-SNuPE also requires bisulfite treatment to be performed first, leading to the conversion of unmethylated cytosine to uracil while methyl cytosine is unaffected. PCR primers specific for bisulfite converted DNA are then used to amplify the target sequence of interest. The amplified PCR product is isolated and

used to quantitate the methylation status of the CpG site of interest. (Gonzalgo and Jones Nuclei Acids Res1997; 25:252-31).

[0139] In particular embodiments, pyrosequencing can be used to detect marker methylation. Pyrosequencing is a method of DNA sequencing that relies on detection of the release of pyrophosphates as DNA is synthesized (and is therefore a "sequencing by synthesis" technique). To assess methylation by pyrosequencing, a DNA sample can be incubated with sodium bisulfite, converting unmethylated cytosine to uracil. The presence of uracil will result in thymine incorporation during PCR amplification. Therefore, sequencing results that include thymine at a nucleotide position that is known to encode cytosine can be interpreted as unmethylated sites. In contrast cytosines present in the sequencing results indicate that the site was methylated in the original DNA sample, because methylation protects cytosine from conversion to uracil upon treatment. Bisulfite treatment can also be performed on control samples with known methylation patterns, to reduce or eliminate false positive results. Commercially available pyrosequencing machines include Pyro Mark Q96 (Qiagen, Hilden, Germany). For more details on methods to use pyrosequencing for measurement of methylation, see Delaney et al. Methods Mol Biol. 2015 1343: 249-264. Pyrosequencing is especially useful for detecting methylation in the CpG sites within genes.

[0140] In particular embodiments, a protein marker is detected by contacting a sample with reagents (e.g., antibodies), generating complexes of reagent and marker(s), and detecting the complexes. Particular embodiments for detecting and measuring protein levels can use methods including agglutination, chemiluminescence, electro-chemiluminescence (ECL), enzyme-linked immunoassays (ELISA), immunoassay, immunoblotting, immunodiffusion, immunoelectrophoresis, immunofluorescence, immunohistochemistry, immunoprecipitation, mass-spectrometry, and western blot. See also, e.g., E. Maggio, Enzyme-Immunoassay (1980), CRC Press, Inc., Boca Raton, Fla; and U.S. Pat. Nos. 4,727,022; 4,659,678; 4,376,110; 4,275,149; 4,233,402; and 4,230,797.

[0141] Read depth refers to the number of times that a specific genomic site is sequenced during a sequencing run.

[0142] Certain implementations are described herein, including the best mode known to the inventors for carrying out implementations of the disclosure. Of course, variations on these described implementations will become apparent to those of ordinary skill in the art upon reading the foregoing description. The inventor expects skilled artisans to employ such variations as appropriate, and the inventors intend for implementations to be practiced otherwise than specifically described herein. Accordingly, the scope of this disclosure includes all modifications and equivalents of the subject matter recited in the claims appended hereto as permitted by applicable law. Moreover, any combination of the above-described elements in all possible variations thereof is encompassed by implementations of the disclosure unless otherwise indicated herein or otherwise clearly contradicted by context.

CLAIMS

What is claimed is:

1. A method, comprising:

identifying, using one or more processors, population records of a genetic variant shared by samples of a population of subjects, the population records respectively comprising indications of the genetic variant and annotations of the genetic variant;

determining, using the one or more processors, that a first set of the population records share a consensus annotation among the annotations, the first set of the population records corresponding to greater than a threshold number of subjects in the population;

in response to determining that the first set of the population records comprises the consensus annotation, updating, using the one or more processors, a second set of the population records to comprise the consensus annotation; and

in response to updating the second set of the population records:

identifying, using the one or more processors, an example subject among the subjects corresponding to an example population record among the second set of population records; and

generating, using the one or more processors, a report based on the genetic variant and the consensus annotation.

2. The method of claim 1, wherein identifying the population records of the genetic variant shared by the samples of the population of subjects comprises:

identifying, using the one or more processors, sequence read data of an example sample obtained from the example subject;

identifying, using the one or more processors, the genetic variant by analyzing the sequence read data; and

generating, using the one or more processors, the example population record indicating the genetic variant and an annotation of the genetic variant.

3. The method of claim 1, wherein:

the first set of the population records was generated after the second set of the population records,

the first set of the population records includes a different annotation than the second set of the population records, and

the method further comprises identifying, using one or more processors, the different annotation from the first set of the population records as the consensus annotation.

4. The method of claim 3, further comprising:

determining, using the one or more processors, that the first set of the population records was generated after the second set of the population records based at least in part on timestamp data indicating times at which the population records, in the first set of the population records and the second set of the population records, were generated.

5. The method of claim 3, wherein determining that the first set of the population records share the consensus annotation among the annotations comprises:

ordering, using the one or more processors, the population records into a temporal order based on timestamps indicating times at which the population records were generated,

wherein instances of the timestamps associated with the first set of the population records are recent timestamps, after a threshold time, in the temporal order, and

wherein the second set of the population records have second instances of the timestamps before the threshold time.

6. The method of claim 1, the genetic variant being a particular genetic variant among a plurality of genetic variants, wherein identifying the population records of the genetic variant shared by the samples of the population of subjects comprises:

grouping, using the one or more processors, a plurality of records of the plurality of genetic variants of the samples into one or more bins respectively corresponding to the plurality of the genetic variants, an example bin among the one or more bins comprising the population records corresponding to the particular genetic variant.

7. The method of claim 1, wherein the example population record comprises:

an identity field indicating at least one of a chromosome where the genetic variant is detected, a position on the chromosome where the genetic variant is detected, or a type of the genetic variant, and

wherein the type of the genetic variant comprises a nucleotide substitution, a nucleotide addition, a nucleotide deletion, a structural variant, or a number of copies of a predetermined nucleotide sequence.

8. The method of claim 1, wherein the annotations comprise at least one of:

an association between the genetic variant and at least one disease,

an indication of an effective therapy for the at least one disease,

disease progression information associated with the at least one disease,

an identifier of the genetic variant,

an ancestry association,

an indication that the genetic variant is a germline variant,

a class identifier of a class or category of the variant, or

a custom annotation type.

9. The method of claim 1, wherein updating the second set of the population records to comprise the consensus annotation comprises:

adding, using the one or more processors, the consensus annotation to the example population record; or

replacing, using the one or more processors, an outdated annotation in the example population record with the consensus annotation.

10. The method of claim 1, further comprising:

adding or updating, using the one or more processors, timestamp data in the second set of the population records to indicate a time at which the second set of the population records was updated to comprise the consensus annotation.

11. The method of claim 1, the report being an updated report, the method further comprising:

identifying, using the one or more processors, an outdated annotation in the example population record; and

generating, using the one or more processors, an original report indicating the genetic variant and the outdated annotation,

wherein updating the second set of the population records is performed after generating the original report.

12. The method of claim 1, wherein the population records are stored in one or more non-linked databases.

13. The method of claim 1, wherein the population records are stored in respective databases.

14. The method of claim 1, wherein:

the annotations comprise a plurality of annotation fields,

the consensus annotation is associated with a first annotation field of the plurality of annotation fields, and

updating the second set of the population records to comprise the consensus annotation comprises updating, using the one or more processors, instances of the first annotation field, associated with the second set of the population records, to indicate the consensus annotation.

15. A system, comprising:

at least one processor; and

memory storing instructions that, when executed by the at least one processor, cause the at least one processor to perform operations comprising:

identifying population records of a genetic variant shared by samples of a population of subjects, the population records respectively comprising indications of the genetic variant and annotations of the genetic variant;

determining that a first set of the population records share a consensus annotation among the annotations, the first set of the population records corresponding to greater than a threshold number of subjects in the population;

in response to determining that the first set of the population records comprises the consensus annotation, updating a second set of the population records to comprise the consensus annotation; and

in response to updating the second set of the population records:

identifying an example subject among the subjects corresponding to an example population record among the second set of population records; and

generating a report based on the genetic variant and the consensus annotation.

16. A method, comprising:

providing a plurality of nucleic acid molecules obtained from a sample from a subject in a population;

ligating one or more adapters onto one or more nucleic acid molecules from the plurality of nucleic acid molecules;

amplifying the one or more ligated nucleic acid molecules from the plurality of nucleic acid molecules;

capturing all or a subset of the amplified nucleic acid molecules;

sequencing, by a sequencer, the captured nucleic acid molecules to obtain a plurality of sequence reads that represent the captured nucleic acid molecules;

receiving, at one or more processors, sequence read data for the plurality of sequence reads;

identifying, using the one or more processors, a genetic variant by comparing the sequence read data to a reference sequence;

generating, using the one or more processors, an example record comprising an indication of the genetic variant and an example annotation indicating information associated with the genetic variant;

storing the example record in one or more databases;

identifying, using the one or more processors, population records, corresponding to the genetic variant, from the one or more databases, wherein the population records comprise the example record, and wherein individual population records, in the population records, are associated with annotations indicating information associated with the genetic variant;

determining, using the one or more processors, that a first set of the population records comprise at least a threshold number of population records;

identifying, using the one or more processors and in response to determining that the first set of the population records comprises at least the threshold number of population records, a consensus annotation associated with the genetic variant that is indicated by the first set of the population records; and

adding the consensus annotation to a second set of the population records, the second set of the population records being different from the first set of the population records.

17. The method of claim 16, wherein:

timestamp data indicates that the first set of the population records was generated after the second set of the population records,

the first set of the population records includes a different annotation than the second set of the population records, and

the one or more processors identify the different annotation from the first set of the population records as the consensus annotation.

18. The method of claim 16, wherein the second set of the population records includes at least one population record that is also within the first set of the population records.

19. The method of claim 16, wherein identifying the population records corresponding to the genetic variant comprises:

identifying, using the one or more processors, a set of database records that:

correspond to different genetic variants, and

indicate variant attributes of the different genetic variants;

grouping, using the one or more processors, the set of database records into a plurality of bins based on one or more of the variant attributes; and

selecting, using the one or more processors, a group of database records associated with a bin, of the plurality of bins, as the population records corresponding to the genetic variant.

20. The method of claim 16, wherein the second set of the population records comprises the example record, the method further comprising:

in response to identifying the genetic variant, causing, using the one or more processors, transmission of an original report to an external device associated with the subject or a care provider, the original report indicating the genetic variant and an original annotation associated with the genetic variant; and

in response to adding the consensus annotation to the second set of the population records, causing, using the one or more processors, transmission of an updated report to the external device, the updated report indicating the genetic variant and the consensus annotation.

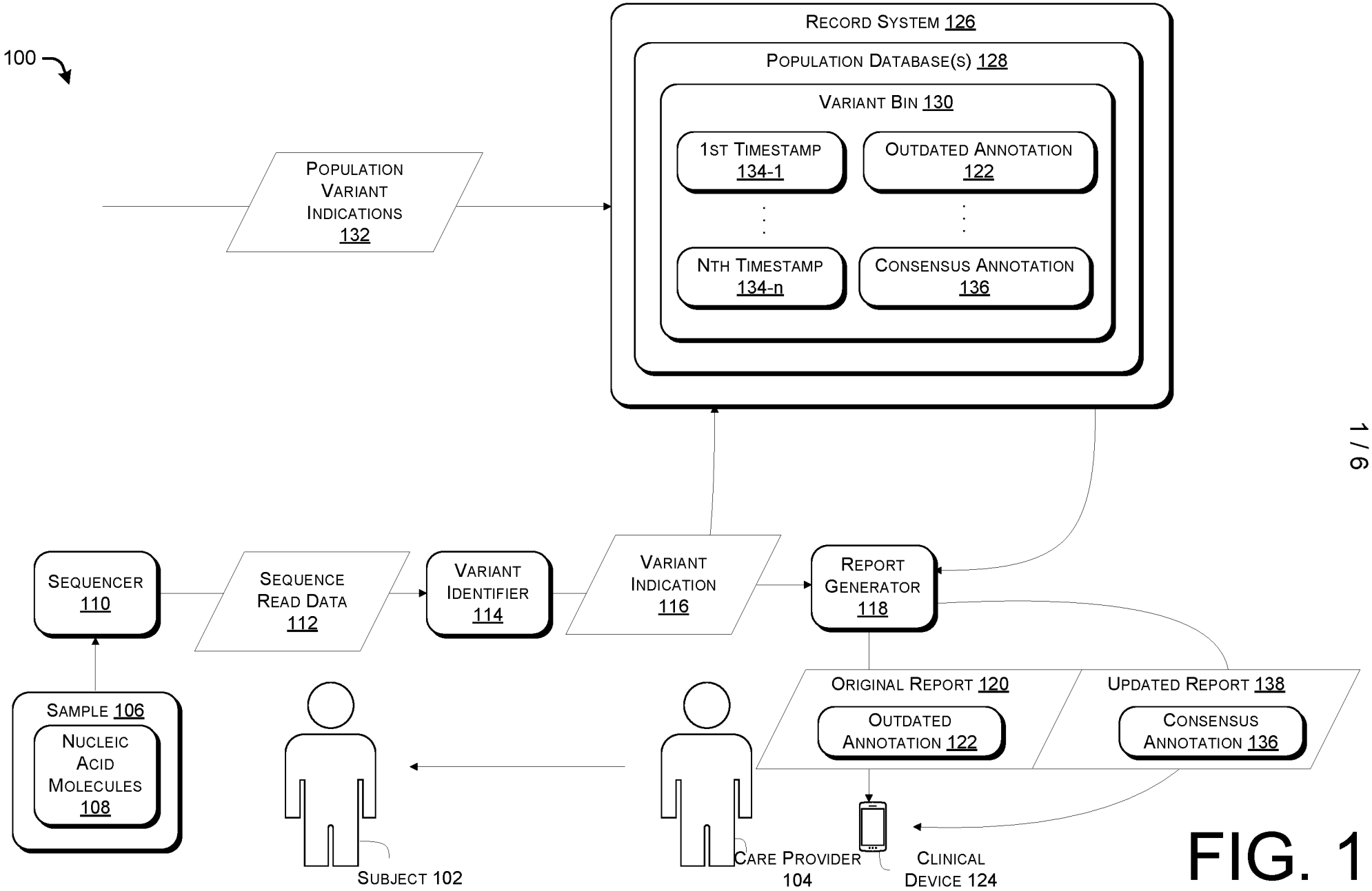
100
Record System 126
Population Database(s) 128
Variant Bin 130
1st Timestamp 134-1
Outdated Annotation 122
Nth Timestamp 134-n
Consensus Annotation 136
Population Variant Indications 132
Sequencer 110
Sequence Read Data 112
Variant Identifier 114
Variant Indication 116
Report Generator 118
Sample 106
Nucleic Acid Molecules 108
Original Report 120
Outdated Annotation 122
Updated Report 138
Consensus Annotation 136
Subject 102
Care Provider 104
Clinical Device 124

FIG. 1

200

202 204 206 208

No.	Timestamp	Position	Type	Annotation
1	1/23/2021 10:49	chr17:g.7674220	C>T	Resistant to slugimib
2	2/12/2021 14:39	chr17:g.7674220	C>T	Resistant to slugimib
3	3/4/2021 18:28	chr17:g.7674220	C>T	Resistant to slugimib
4	3/24/2021 22:18	chr17:g.7674220	C>T	Resistant to slugimib
5	4/14/2021 2:07	chr17:g.7674220	C>T	Resistant to slugimib
~~6~~	~~5/4/2021 5:57~~	~~chr17:g.7674220~~	~~C>G~~	~~Germ line mutation~~
7	5/24/2021 9:46	chr17:g.7674220	C>T	Resistant to slugimib
8	6/13/2021 13:36	chr17:g.7674220	C>T	Resistant to slugimib
9	7/3/2021 17:25	chr17:g.7674220	C>T	Resistant to slugimib
10	7/23/2021 21:15	chr17:g.7674220	C>T	Resistant to slugimib
11	8/13/2021 1:04	chr17:g.7674220	C>T	Resistant to slugimib
12	9/2/2021 4:54	chr17:g.7674220	C>T	Resistant to slugimib; responsive to titanamib
13	1/23/2021 10:49	chr17:g.7674220	C>T	Resistant to slugimib; responsive to titanamib
14	2/12/2021 14:39	chr17:g.7674220	C>T	Resistant to slugimib; responsive to titanamib
15	3/4/2021 18:28	chr17:g.7674220	C>T	Resistant to slugimib; responsive to titanamib
16	1/23/2021 10:49	chr17:g.7674220	C>T	Resistant to slugimib; responsive to titanamib
17	2/12/2021 14:39	chr17:g.7674220	C>T	Resistant to slugimib; responsive to titanamib
18	3/4/2021 18:28	chr17:g.7674220	C>T	Resistant to slugimib; responsive to titanamib
19	3/24/2021 22:18	chr17:g.7674220	C>T	Resistant to slugimib; responsive to titanamib
20	4/14/2021 2:07	chr17:g.7674220	C>T	Resistant to slugimib; responsive to titanamib
21	5/4/2021 5:57	chr17:g.7674220	C>T	Responsive to titanamib
~~22~~	~~5/24/2021 9:46~~	~~chr17:g.7674220~~	~~C>G~~	~~Germ line mutation~~
23	6/13/2021 13:36	chr17:g.7674220	C>T	Responsive to titanamib; associated with ovarian cancer
24	7/3/2021 17:25	chr17:g.7674220	C>T	Responsive to titanamib
25	7/23/2021 21:15	chr17:g.7674220	C>T	Responsive to titanamib; associated with ovarian cancer
26	8/13/2021 1:04	chr17:g.7674220	C>T	Resistant to slugimib; responsive to titanamib
27	9/2/2021 4:54	chr17:g.7674220	C>T	Resistant to slugimib; responsive to titanamib
28	1/23/2021 10:49	chr17:g.7674220	C>T	Resistant to slugimib; responsive to titanamib
29	2/12/2021 14:39	chr17:g.7674220	C>T	Resistant to slugimib; responsive to titanamib
30	3/4/2021 18:28	chr17:g.7674220	C>T	Resistant to slugimib; responsive to titanamib
31	1/23/2021 10:49	chr17:g.7674220	C>T	Resistant to slugimib; responsive to titanamib
32	2/12/2021 14:39	chr17:g.7674220	C>T	Resistant to slugimib; responsive to titanamib
33	3/4/2021 18:28	chr17:g.7674220	C>T	Resistant to slugimib; responsive to titanamib
34	3/24/2021 22:18	chr17:g.7674220	C>T	Resistant to slugimib; responsive to titanamib

FIG. 2

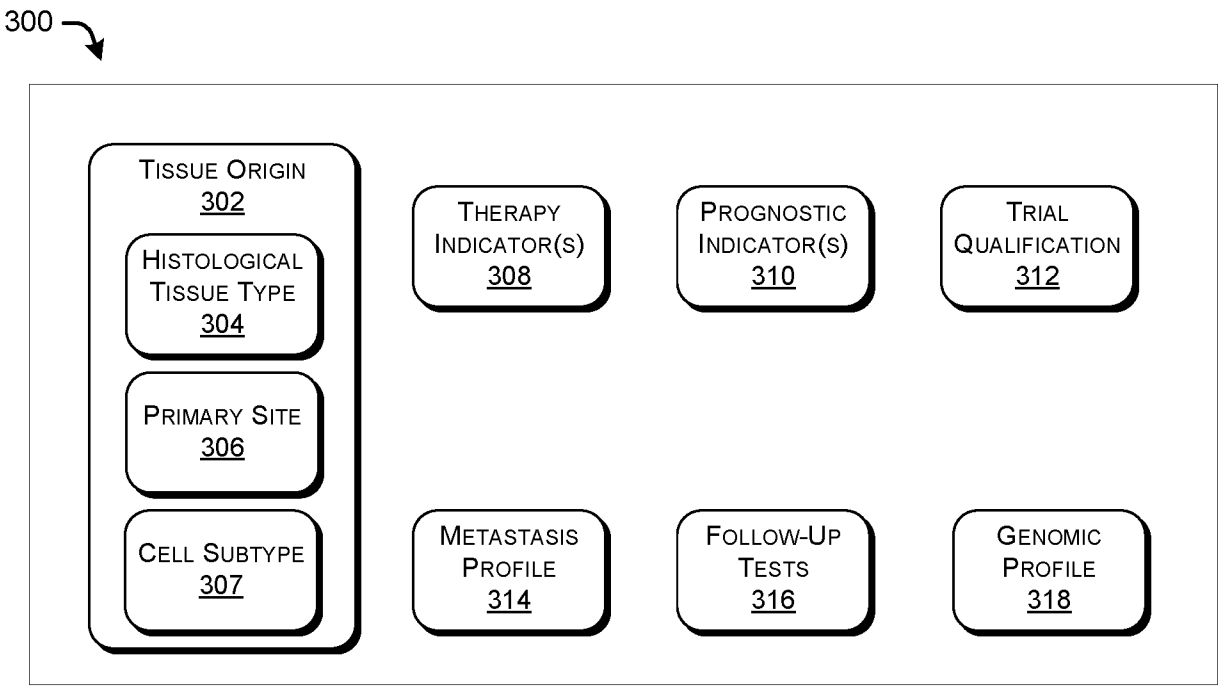
300
TISSUE ORIGIN
302
HISTOLOGICAL TISSUE TYPE
304
PRIMARY SITE
306
CELL SUBTYPE
307
THERAPY INDICATOR(S)
308
PROGNOSTIC INDICATOR(S)
310
TRIAL QUALIFICATION
312
METASTASIS PROFILE
314
FOLLOW-UP TESTS
316
GENOMIC PROFILE
318

FIG. 3

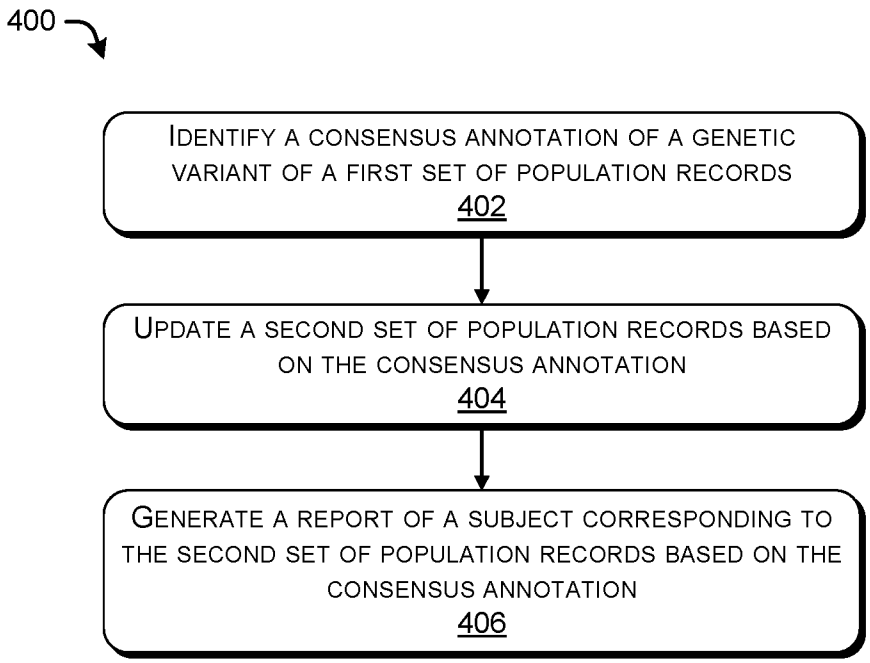
400
IDENTIFY A CONSENSUS ANNOTATION OF A GENETIC VARIANT OF A FIRST SET OF POPULATION RECORDS
402
UPDATE A SECOND SET OF POPULATION RECORDS BASED ON THE CONSENSUS ANNOTATION
404
GENERATE A REPORT OF A SUBJECT CORRESPONDING TO THE SECOND SET OF POPULATION RECORDS BASED ON THE CONSENSUS ANNOTATION
406

FIG. 4

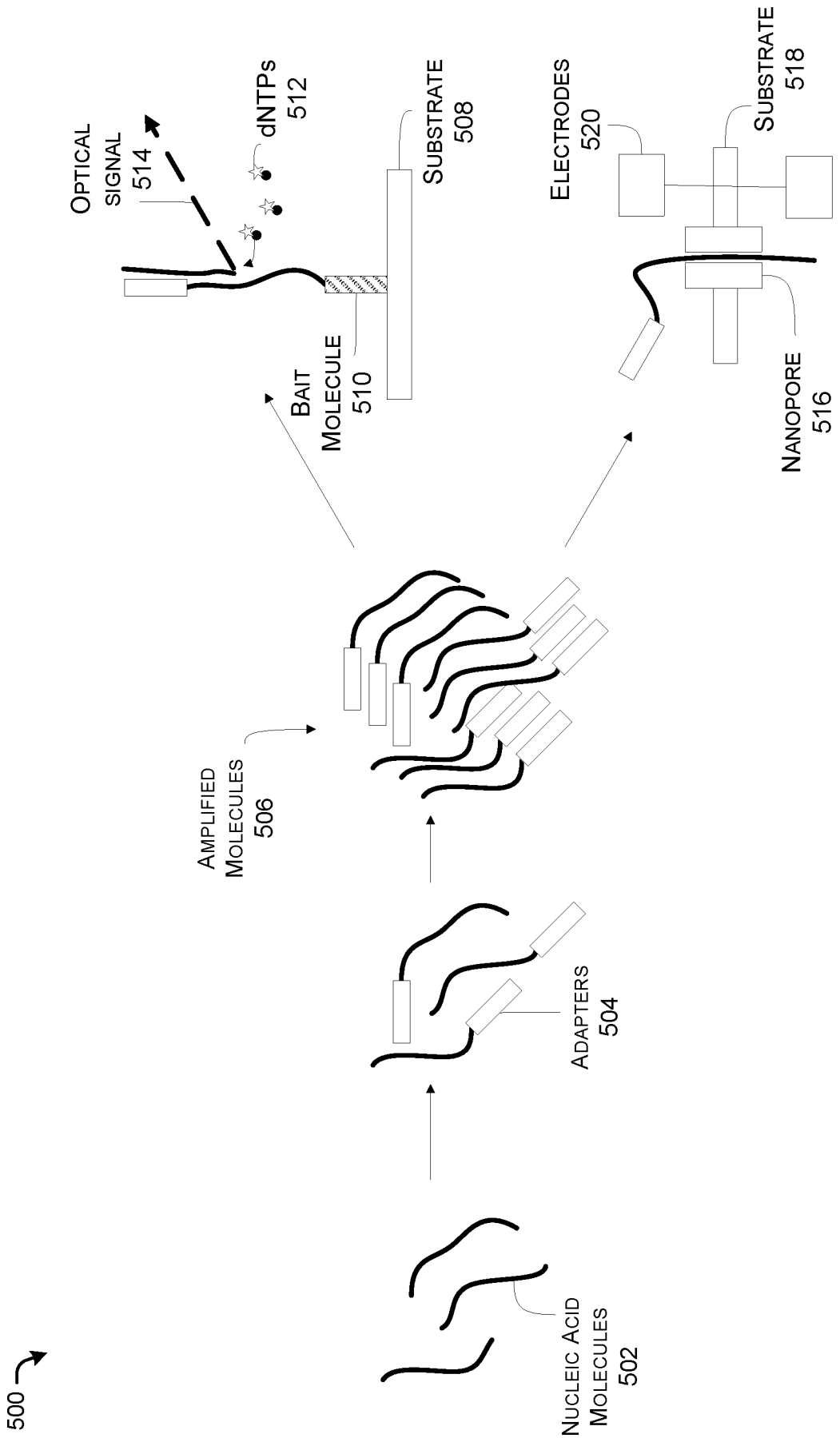
500
NUCLEIC ACID MOLECULES 502
ADAPTERS 504
AMPLIFIED MOLECULES 506
BAIT MOLECULE 510
SUBSTRATE 508
dNTPs 512
OPTICAL SIGNAL 514
NANOPORE 516
ELECTRODES 520
SUBSTRATE 518

FIG. 5

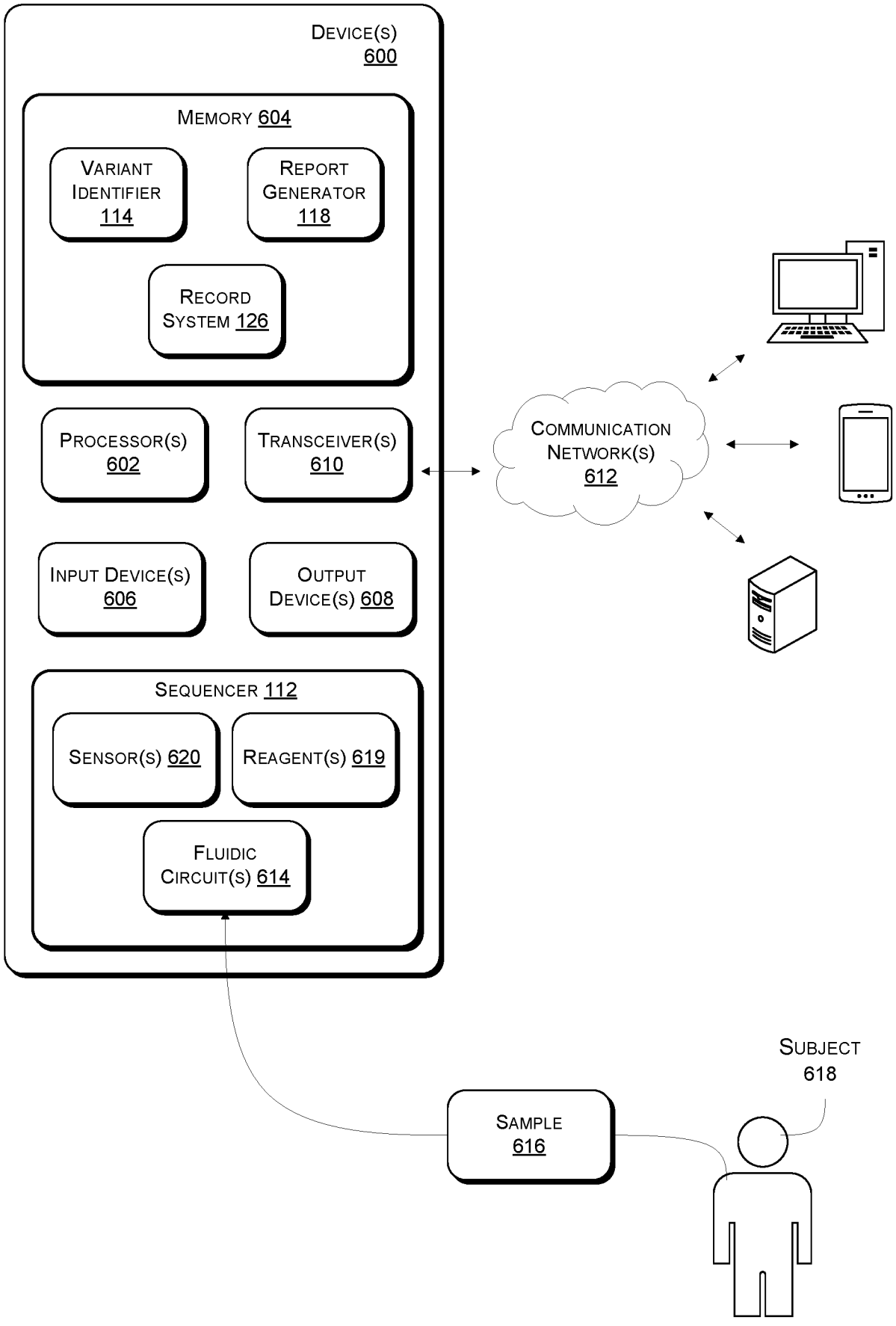
DEVICE(S)
600
MEMORY 604
VARIANT IDENTIFIER 114
REPORT GENERATOR 118
RECORD SYSTEM 126
PROCESSOR(S) 602
TRANSCEIVER(S) 610
INPUT DEVICE(S) 606
OUTPUT DEVICE(S) 608
SEQUENCER 112
SENSOR(S) 620
REAGENT(S) 619
FLUIDIC CIRCUIT(S) 614
COMMUNICATION NETWORK(S) 612
SAMPLE 616
SUBJECT 618

FIG. 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 24/47949

A. CLASSIFICATION OF SUBJECT MATTER

IPC - INV. G16B 50/10, G16B 50/30, C12Q 1/6883, G16B 50/20, G16B 40/00, G16B 20/40 (2024.01)
ADD. G06N 20/00, G16H 20/00, C12Q 1/6806, C12N 15/10 (2024.01)

CPC - INV. G16B 50/10, G16B 50/30, C12Q 1/6883, G16B 50/20, G16B 40/00, G16B 20/40

ADD. G06N 20/00, G16H 20/00, C12Q 1/6806, C12N 15/10

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Search History document

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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2020/0395100 A1 (Guardant Health Inc), 17 December 2020 (17.12.2020), entire document	1-20
A	US 2015/0154354 A1 (THE SCRIPPS RESEARCH INSTITUTE), 04 June 2015 (04.06.2015), entire document	1-20
A	US 2019/0390253 A1 (Guardant Health, Inc), 26 December 2019 (26.12.2019), entire document	1-20
A	US 2021/0233664 A1 (Tempus Labs), 29 July 2021 (29.07.2021), entire document	1-20
A	US 2022/0399087 A1 (Rady Children's Hospital Research Center et al.), 15 December 2022 (15.12.2022), entire document	1-20
A	US 2019/0205502 A1 (Regeneron Pharmaceuticals, Inc), 04 July 2019 (04.07.2019), entire document	1-20
A	US 2014/0088942 A1 (AMBRY GENETICS), 27 March 2014 (27.03.2014), entire document	1-20
A	US 2020/0135296 A1 (Ancestry.com DNA, LLC), 30 April 2020 (30.04.2020), entire document	1-20
A	Pan et al. " VARAdb: a comprehensive variation annotation database for human." In: Nucleic Acids Research, 2021, Vol. 49, Database issue D1431–D1444, 23 October 2020, [online] [retrieved on 08 November 2024] Retrieved from the Internet < URL: https://academic.oup.com/nar/article/49/D1/D1431/5937088 >, entire document	1-20

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

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

Date of the actual completion of the international search	Date of mailing of the international search report
09 November 2024	NOV 27 2024
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Kari Rodriquez Telephone No. PCT Helpdesk: 571-272-4300

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 24/47949

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Atzeni et al. "VariantAlert: A web-based tool to notify updates in genetic variant annotations." In: Human Mutation. 2022;43:1808–1815, [online] [retrieved on 08 November 2024] Retrieved from the Internet < URL: https://pubmed.ncbi.nlm.nih.gov/36300680/>, entire document	1-20
A	Chang et al. "wANNOVAR: annotating genetic variants for personal genomes via the web." In: Med Genet. 2012 July ; 49(7): 433–436, [online] [retrieved on 08 November 2024] Retrieved from the Internet < URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC3556337/ > entire document	1-20
A	Pedersen et al. "Vcfanno: fast, flexible annotation of genetic variants." In: Genome Biology (2016) 17:118, [online] [retrieved on 08 November 2024] Retrieved from the Internet < URL: https://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-0973-5 > entire document	1-20