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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
16 May 2024 (16.05.2024)

WIPO | PCT

(10) International Publication Number
WO 2024/103003 A2

(51) International Patent Classification:
C12Q 1/6886 (2018.01) *G16B 20/00* (2019.01)

(21) International Application Number:
PCT/US2023/079380

(22) International Filing Date:
10 November 2023 (10.11.2023)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/424,478 10 November 2022 (10.11.2022) US

(71) Applicant: **THE REGENTS OF THE UNIVERSITY OF CALIFORNIA** [US/US]; 111 Franklin Street, 12th Floor, Oakland, California 94607-5200 (US).

(72) Inventors: **AHITUV, Nadav**; 1473 Whitecliff Way, Walnut Creek, California 94596 (US). **YIZHAR-BARNEA, Ofer**; 835 Columba Lane, Foster City, California 94404 (US). **GEORGAKOPOULOS-SOARES, Ilias**; 568 Dickens Drive, Hummelstown, Pennsylvania 17036 (US). **HEMBERG, Martin**; 44 Bay State Road, Weston, Massachusetts 19103 (US). **MOURATIDIS, Ioannid**; Avenue des Grands Prox 152, 1150-Woluwe-Sain-Pierre, Brussels (BE).

(74) Agent: **ZURAWSKI, John A.** et al.; 1800 John F. Kennedy Boulevard, Suite 1010, Philadelphia, Pennsylvania 19103 (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,

(54) Title: SYSTEMS FOR MUTATION CALLER AND METHODS OF USING THE SAME

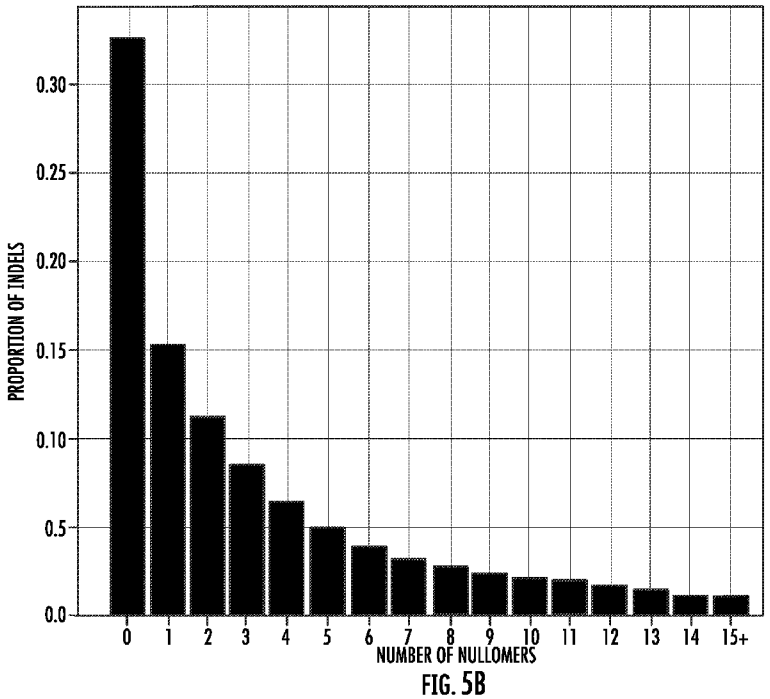


FIG. 5B

(57) Abstract: The present disclosure provides methods and compositions for the detection, identification, classification and characterization of cancer in general and cancer types in biological material. Sequences that are not found in the human reference genome or any set of genomic tiled regions, termed nullomers, which can resurface due to mutations, serve as biomarkers and are predictive of cancer. The invention also enables the identification of cancer subtype and the stratification of patients based on sample-specific vulnerabilities guiding treatment choice. For coding nullomers it also covers their use as neoantigens. The algorithms presented hereby can be applied to biological material including biopsy, cell-free DNA samples and RNA samples.

WO 2024/103003 A2

[Continued on next page]

TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(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:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

— *with sequence listing part of description (Rule 5.2(a))*

SYSTEMS FOR MUTATION CALLER AND METHODS OF USING THE SAME

CROSS-REFERENCE TO RELATED APPLICATIONS

This Application claims the benefit of U.S. Application No. 63/424,478, filed on November 10, 2022, the contents of which are hereby incorporated by reference in its entirety.

REFERENCE TO SEQUENCE LISTING

The Sequence Listing submitted as an XML file named "UCAL-026-PCT" created on November 10, 2023, and having a size of 151 KB is hereby incorporated by reference pursuant to 37 C.F.R. § 1.52(e)(5).

TECHNOLOGY FIELD

The present disclosure relates to the identification of prognostic and diagnostic cancer biomarkers in biological material and the characterization of tumor subtype, vulnerabilities and therapeutic strategies, from the resurfacing of nullomers.

BACKGROUND

Cancer is the second leading cause of death worldwide ("Cancer" n.d.), and for most cancer types, survivability is significantly higher if the tumor is detected at an early stage (Hawkes 2019; Etzioni et al. 2003). Currently mass population screening is applicable only for breast and cervical cancers and utilizes physical tests like mammography and cytology screens. Detection for other cancer types, done both *en masse* and in a low and affordable resource setting, still poses a major challenge for the scientific and clinical communities ("Cancer" n.d.). In particular, a major hurdle is to single-out cancer biomarkers for the detection of cancer development at its earliest stage for patient stratification and improvement of patients' outcome by providing personalized treatments.

Circulating cell-free DNA (cfDNA) is an emerging and promising resource for cancer diagnostics and prognostics (Bronkhorst, Ungerer, and Holdenrieder 2019; Heitzer, Auinger, and Speicher 2020). It has a short life span (16 minutes to 2.5 hours), which makes it a highly temporal indicator of various processes occurring in the subject's body and with advances in sequencing technologies, can be rapidly analyzed. Analysis of cell-free tumor DNA (ctDNA, liquid biopsy) has become a prospective minimally invasive tool to screen the population and to monitor patients

already diagnosed with cancer. To distinguish cancerous cells, their tissue of origin and cancer type, current technologies rely on sequencing to resolve somatic mutations (Zill et al. 2018) and epigenetic marks, such as DNA methylation or histone modifications that can determine the cancerous tissue (Saghafinia et al. 2018; Sadeh et al. 2021). However, ctDNA still has many hurdles and caveats that need to be overcome (Barbany et al. 2019). Some of the major hurdles include: 1) cfDNA is fragmented (180-360 base pairs) making its collection and extraction more challenging and the tumor-derived DNA makes up only a small portion (estimated to be around 0.4%) warranting the need for extremely sensitive biomarkers that can easily detect the presence of cancerous cells; 2) prior knowledge of specific mutations or methylation marks is required for targeted screening, and consequently the main focus has been on coding mutations which only constitute a small fraction of mutations; 3) cfDNA mutation and epigenetic diagnosis could be confounded by somatic alterations in white blood cells (Razavi et al. 2019); 4) the diagnostic techniques used to detect methylation or histone marks are technologically complex and can have low sensitivity and specificity (Ji et al. 2014; Worm Ørntoft 2018; Warton and Samimi 2015; Bronkhorst, Ungerer, and Holdenrieder 2019) and 5) to provide the most optimal cancer treatment, it needs to be diagnosed at preliminary stages when the tumor is small (~5mm in diameter). At these stages, the tumor produces minute levels of ctDNA that are difficult to detect using current methods (Bronkhorst, Ungerer, and Holdenrieder 2019).

Nullomers are short DNA sequences (11-18 base pairs) that are absent from the human genome (Hampikian and Andersen 2006; Vergni and Santoni 2016). While the absence of nullomeric sequences could be due to chance, we and others have shown that a significant proportion of them is under negative selection pressures (Georgakopoulos-Soares et al. 2020; Vergni and Santoni 2016), suggesting that they could have a deleterious effect on the genome. Experimental evidence was also provided through the observation that two out of three nullomers led to lethality in several cancerous cell types when delivered as synthetic peptides (Alileche et al. 2012; Alileche and Hampikian 2017).

As nullomers do not exist in a human genome, their appearance due to mutagenesis followed by clonal expansion could be exploited as a diagnostic method for diseases associated with a mutational burden, such as cancer.

SUMMARY OF EMBODIMENTS

The disclosure relates to determining whether nullomers could be used as a diagnostic tool to detect cancer and various additional tumor features. Throughout this manuscript, we refer to nullomers from tumor genomes as neomers to distinguish them from the more general category. We first analyzed The Cancer Genome Atlas (TCGA;20) database finding neomers created by somatic mutations that detected cancer subtypes with higher accuracy than leading methods as well as additional cancer features. Further analyses of cfDNA whole-genome sequencing (WGS) datasets found that these neomers can also be used to detect cancer subtypes. Using WGS on cfDNA from individual sequence data from subjects with colorectal, pancreatic, prostate, ovarian or lung cancer and normal controls found an enrichment for neomers. Finally, utilizing both promoter assays and a massively parallel reporter assay (MPRA), we show that neomers alter regulatory activity of tumors and can be used to detect cancer-associated mutations in gene regulatory elements. Combined, our results show that neomers can be used as a rapid, sensitive, specific and simple cancer diagnostic tool and also aid in the identification of gene regulatory driver mutations associated with cancer.

The disclosure relates to methods and compositions for the detection, identification, classification and characterization of cancer in general and cancer types in biological material such as solid tumors. The disclosure also relates to a method of identifying a neomer from a sample from subject comprising:

(a) identifying the presence of one or a plurality of nullomers in a dataset;

(b) identifying nullomers from a sample of a subject;

(c) annotating the nullomers from the subject as being specific to the subject by comparing the nullomers from the sample with nucleic acid sequences from a control subject;

(d) assigning a cancer type to a nullomer with a corresponding cancer type of the patient;

(e) creating a library of neomer that correspond to a cancer type by repeating steps (c) and (d).

The disclosure relates to a method of creating a library of neomers that correspond to a cancer type from a sample of a subject comprising:

(a) identifying the presence of one or a plurality of nullomers in a dataset;

(b) identifying nullomers from a sample of a subject;

(c) annotating the nullomers from the subject as being specific to the subject by comparing the nullomers from the sample with nucleic acid sequences from a control subject;

(d) assigning a cancer type to a nullomer with a corresponding cancer type of the patient;

(e) creating a library of neomers that correspond to a cancer type by repeating steps (c) and (d) if the frequency of the neomer corresponds to the presence of a tumor from one or a plurality of subject from the dataset.

The disclosure relates to a method of preparing a sample comprising:

(a) isolating a plurality of nucleic acids from the sample; (b) contacting the nucleic acids to one or a plurality of probes specific for one or a plurality of neomers; (c) detecting the presence of the probes associated with the one or plurality of neomers; and (d) correlating the presence or quantity of probes with the likelihood of the presence or quantity of neomer in the sample. In some embodiments, the one or plurality of probes comprise a complementary nucleic acid sequence bound to or associated with a fluorescent molecule, radioactive isotope or chemiluminescent molecule. In some embodiments, the step of detecting is performed by mass spectrometry. In some embodiments, the step of detecting and/or correlating the presence or quantity of probes with the likelihood of the presence or quantity of neomer in the sample is performed by analyzing the sample for the presence of a neomer.

The disclosure provides a method of identifying one or a plurality of nullomers in a sample comprising: (a) isolating a plurality of nucleic acids from the sample; (b) contacting the nucleic acids to one or a plurality of probes specific for one or a plurality of nullomers; (c) detecting the presence of the probes associated with the one or plurality of nullomers; and (d) correlating the presence or quantity of probes with the likelihood of the presence or quantity of nullomers in the sample. In some embodiments, the one or plurality of probes comprise a complementary nucleic acid sequence bound to or associated with a fluorescent molecule, radioactive isotope or chemiluminescent molecule. In some embodiments, the step of detecting is performed by mass spectrometry.

In some embodiments, the method further comprises, prior to step (b), disassociating a plurality of double stranded nucleic acid sequences comprising at least one nullomer by exposing the double-stranded nucleic acid sequences to a predetermined melting temperature for a period of time sufficient to create single stranded nullomer, annealing at least one primer to the nullomer, and allowing a sufficient period of time to extend the primer in the presence of nucleotide

triphosphates (NTPs) and a polymerase or functional fragment thereof. In some embodiments, the steps of disassociating a plurality of double stranded nucleic acid sequences comprising at least one nullomer by exposing the double-stranded nucleic acid sequences to a predetermined melting temperature for a period of time sufficient to create single stranded nullomer, annealing at least one primer to the nullomer, and allowing a sufficient period of time to extend the primer in the presence of NTPs and the polymerase are repeated multiple times such that copies of the at least one nullomer are produced.

The disclosure further provides a method of identifying one or plurality of nullomers in a sample comprising: (a) isolating a plurality of nucleic acids from the sample; (b) contacting the nucleic acids to one or a plurality of probes specific for one or a plurality of nullomers; (c) detecting the presence of the probes associated with the one or plurality of neomers; (d) correlating the presence or quantity of probes with the likelihood or the presence or quantity of neomers in the sample; and (e) comparing the sequence of the neomer with the sequence of a library of known neomer sequences. In some embodiments, the probe or plurality of probes comprise a complementary nucleic acid sequence bound to or associated with a fluorescent molecule, radioactive isotope or chemiluminescent molecule. In some embodiments, the method further comprises a step of performing polymerase chain reaction (PCR) with one or a plurality of primers specific for the one or plurality of neomers.

The disclosure relates to a method of diagnosing a subject with a cancer comprising:

(a) isolating a plurality of nucleic acids from a sample; (b) contacting the nucleic acids to one or a plurality of probes specific for one or a plurality of neomers; (c) detecting the presence of the probes associated with the one or plurality of neomers; and (d) correlating the presence or quantity of probes with the likelihood of the presence or quantity of neomer in the sample; and (e) diagnosing the subject with cancer if the likelihood of the presence or the quantity of a neomer in a sample is over a threshold value.

In some embodiments, the sample in any of the disclosed methods is a cell free nucleic acid sample.

The disclosure also provides a computer-implemented method of identifying a mutation associated with a comprising: (a) isolating one or a plurality of nucleic acid molecules from a sample associated with the hyperproliferative disorder; (b) contacting the nucleic acids to one or a plurality of probes specific for one or a plurality of neomers; (c) in a system configured to compile

data and detect the presence or quantify the presence of a nucleic acid sequence, detecting the presence of the probes associated with the one or plurality of neomers; (d) correlating the presence or quantity of the neomer to the likelihood of a specific mutation serving as a biomarker for a hyperproliferative disorder. In some embodiments, the method further comprises, prior to step (a), in a system configured to compile data and detect the presence or quantity of nucleic acids in a sample: compiling genetic data about a population of subjects including the subject that has a mutation candidate that is a biomarker for a hyperproliferative disorder. In some embodiments, the method further comprises, after step (d), a step of: (e) selecting a cancer treatment for the subject based upon identification of the hyperproliferative disorder. In some embodiments, the hyperproliferative disorder is breast cancer, ovarian cancer, lung cancer, pancreatic cancer, or liver cancer. In some embodiments, the hyperproliferative disorder is breast cancer, pancreatic cancer, esophagus cancer, lymphoid cancer, kidney cancer, ovarian cancer, head and neck cancer, lung cancer, stomach cancer, CNS cancer, uterus cancer, skin cancer, colorectal cancer, prostate cancer, bladder cancer, bone and soft tissue cancer, biliary cancer, cervix cancer, thyroid cancer, myeloid cancer, or liver cancer. In some embodiments, the hyperproliferative disorder is a malignant tumor. In some embodiments, the sample is a brush biopsy, puncture biopsy, fluid from a needle biopsy, blood, blood cells, cells from a hair sample, nucleic acids from a hair sample, saliva, or spit. In some embodiments, the probe or plurality of probes comprise a complementary nucleic acid sequence bound to or associated with a fluorescent molecule, radioactive isotope or chemiluminescent molecule. In some embodiments, the method further comprises a step of performing PCR with one or a plurality of primers specific for the one or plurality of nullomers.

The disclosure additionally provides a method of treating a hyperproliferative disorder in a subject in need thereof comprising: (a) exposing a sample from the subject to a probe specific for at least one neomer chosen from Table 1; (b) detecting the presence, absence or quantity of the at least one neomer in the sample; (c) normalizing the presence, absence, or quantity of the at least one neomer in the sample against the presence, absence or quantity of the at least one neomer in a sample of a healthy subject or a sample of a subject known to have the hyperproliferative disorder; (d) correlating the presence, absence, or quantity of the at least one neomer in the sample to the subject having the hyperproliferative disorder; and (e) administering a therapeutically effective amount of one or a plurality of active agents to the subject. In some embodiments, the method further comprises obtaining the sample from the subject prior to the step of exposing. In some

embodiments, the one or plurality of active agents is chosen from one or a combination of the agents identified in Table 3. In some embodiments, the sample is plasma, serum, whole blood, respiratory tissue, respiratory mucosal sample, saliva, urine, blood cells, cells from a hair sample, nucleic acids from a hair sample, or spit. In some embodiments, the sample is a blood sample comprising cell-free genomic DNA or RNA from a solid tumor or cell-free genomic DNA or RNA from a circulating tumor cell from the subject. In some embodiments, the sample is a blood sample comprising a circulating tumor cell from the subject.

In some embodiments, step (b) further comprises calculating one or more scores based upon the presence, absence, or quantity of the at least one neomer, and step (d) further comprises correlating the one or more scores to the presence, absence, or quantity of the at least one nullomer such that, if the amount of the at least one neomer is greater than the quantity of the at least one neomer in a control sample; or, if the amount of the at least one neomer is substantially equal to the quantity of the at least one neomer in a sample taken from a subject known to have a hyperproliferative disorder, then the subject is diagnosed as having a hyperproliferative disorder. In some embodiments, the probe is a radioactive probe, a chemiluminescent probe, or a fluorescent probe. In some embodiments, the sample is free of cells.

In some embodiments, the at least one nullomer is detected by next generation sequencing, quantitative real-time reverse transcription-PCR (qRT-PCR), isothermal amplification, microarray, multiplex nullomer profiling assay, RNA in situ hybridization (RNA-ish), or northern blotting. In some embodiments, the at least one nullomer is detected by qRT-PCR. In some embodiments, the step of quantifying at least one quantity of the at least one nullomer in the sample comprises using a fluorescence and/or digital imaging.

In some embodiments, the step of analyzing comprises detecting a presence, absence, or quantity of at least 2 different neomers. In some embodiments, the step of analyzing comprises detecting the presence, absence, or quantity of the at least one neomer by PCR amplification using one or a plurality of primers specific for the at least one neomer chosen from Table 1. In some embodiments, the step of analyzing comprises detecting presence, absence, or quantity of the at least one neomer by a probe comprising a nucleic acid sequence complementary to the nucleic acid sequence of the at least one nullomer.

The disclosure further provide a method of diagnosing a subject with cancer comprising: (a) contacting a plurality of nucleic acids from a sample to a system comprising a probe specific

for one or a plurality of neomers; and (b) detecting the presence of or quantifying the amount of one or more nucleic acids from the sample. In some embodiments, the method comprises detecting the presence, absence or quantity of one or a plurality of the neomers provided in Table 1. In some embodiments, the method comprises detecting the presence, absence or quantity of nullomers that comprise at least 93% sequence identify to one or a plurality of the nullomers provided in Table 1. In some embodiments, the at least one nullomer is detected by qRT-PCR. In some embodiments, the at least one nullomer is detected by CRISPR diagnosis. In some embodiments, the at least one nullomer is detected by CRISPR diagnosis and Cas9, Cas12 or Cas13 protein is used.

In some embodiments, the method further comprises, after the step of detecting, normalizing the quantity of the probe as compared to a quantity of signal from a negative control. In some embodiments, the method further comprises, after the step of detecting, correlating the one or more scores to the presence, absence, or quantity of the at least one neomer such that, if the amount of the at least one neomer is greater than the quantity of the at least one neomer in a control sample; or, if the amount of the at least one neomer is substantially equal to the quantity of the at least one neomer in a sample taken from a subject known to have a hyperproliferative disorder, then the subject is diagnosed as having a hyperproliferative disorder. In some embodiments, the hyperproliferative disorder is solid tumor of the breast, pancreas, ovary, lung or liver. In some embodiments, the hyperproliferative disorder is breast cancer, pancreatic cancer, esophagus cancer, lymphoid cancer, kidney cancer, ovary cancer, head and neck cancer, lung cancer, stomach cancer, CNS cancer, uterus cancer, skin cancer, colorectal cancer, prostate cancer, bladder cancer, bone and soft tissue cancer, biliary cancer, cervix cancer, thyroid cancer, myeloid cancer, or liver cancer. In some embodiments, the hyperproliferative disorder is a metastatic tumor if the presence or quantity of the neomer corresponds to the presence of a circulating tumor cell in a blood sample.

Also provided is a kit comprising one or more probes or primers for detecting the presence, absence or quantity of one or a plurality of the neomers provided in Table 1 or neomers that comprise at least 93% sequence identify to one or a plurality of the neomers provided in Table 1. In some embodiments, the one or more probes comprised in the disclosed kit comprise one or a combination of the neomer sequences of Table 1 or complementary thereof.

Further provided is a computer program product encoded on a computer-readable storage medium, wherein the computer program product comprises instructions for: a) detecting the presence, absence or quantity of at least one neomer in a sample of a subject; b) normalizing the presence, absence, or quantity of the at least one neomer in the sample against the presence, absence or quantity of the at least one neomer in a control sample; and c) correlating the presence, absence, or quantity of the at least one neomer in the sample to a likelihood that the subject having a hyperproliferative disorder. In some embodiments, the computer program product further comprises instructions for calculating a score associated with the presence, absence or quantity of the at least one neomer in the sample and correlating the score to a likelihood that the subject has a hyperproliferative disorder. In some embodiments, the computer program product further comprises instructions for: a) detecting and normalizing the presence, absence or quantity of a second neomer in the sample; b) calculating a combined score associated with the presence, absence or quantity of the at least one neomer and the second neomer in the sample; and c) correlating the combined score to a likelihood that the subject having a hyperproliferative disorder. In some embodiments, at least 2 different neomers in the sample are detected, normalized and correlated by the computer program product. In some embodiments, the computer program product detects the presence, absence, or quantity of the at least one neomer by qRT-PCR amplification. In some embodiments, the control sample used in the computer program product is obtained from a subject free of a hyperproliferative disorder.

The disclosure also provides a system comprising: a) the computer program product of any one of claims 54 to 59; and b) a processor operable to execute programs; and/or a memory associated with the processor.

The disclosure further provides a system for detecting the presence or quantity of neomer in a sample of a subject comprising: a processor operable to execute programs, a memory associated with the processor, a database associated with said processor and said memory, and a program stored in the memory and executable by the processor, the program being operable for: a) detecting the presence, absence or quantity of at least one neomer in a sample of a subject; b) normalizing the presence, absence, or quantity of the at least one neomer in the sample against the presence, absence or quantity of the at least one neomer in a control sample; and c) correlating the presence, absence, or quantity of the at least one neomer in the sample to a likelihood that the subject having a hyperproliferative disorder. In some embodiments, the program is further

operable for calculating a score associated with the presence, absence or quantity of the at least one neomer in the sample and correlating the score to a likelihood that the subject has a hyperproliferative disorder. In some embodiments, the program is further operable for detecting and normalizing the presence, absence or quantity of a second neomer in the sample.

In some embodiments, the one or plurality of probes used in any of the disclosed methods, systems, or computer program product, or comprised in any of the disclosed kits comprise a nucleic acid sequence chosen from Table 1. In some embodiments, the one or plurality of probes used in any of the disclosed methods, systems, or computer program product, or comprised in any of the disclosed kits comprise a nucleic acid sequence comprising at least about 93% sequence identity to any of the sequences in Table 1.

In some embodiments, the one or plurality of probes used in any of the disclosed methods, systems, or computer program product, or comprised in any of the disclosed kits comprise a nucleic acid sequence that is complementary to any of the neomer sequences provided in Table 1, or a fragment thereof. In some embodiments, the one or plurality of probes used in any of the disclosed methods, systems, or computer program product, or comprised in any of the disclosed kits comprise a nucleic acid sequence that is complementary to a neomer comprising at least about 93% sequence identity to any of the neomer sequences provided in Table 1, or a fragment thereof.

In some embodiments, the one or plurality of primers specific for the one or plurality of neomers used in any of the disclosed methods, systems, or computer program product, or comprised in any of the disclosed kits comprise a nucleic acid sequence that is complementary to any of the neomer sequences provided in Table 1, or a fragment thereof. In some embodiments, the one or plurality of primers specific for the one or plurality of neomers used in any of the disclosed methods, systems, or computer program product, or comprised in any of the disclosed kits comprise a nucleic acid sequence that is complementary to a neomer comprising at least about 93% sequence identity to any of the neomer sequences provided in Table 1, or a fragment thereof.

BRIEF DESCRIPTION OF DRAWINGS

FIGS. 1A-1E Neomers can detect cancer tissue of origin. **Fig. 1A.** Schematic overview of neomer cancer diagnostic pipeline. **Fig. 1B.** Number of neomers per patient sample across tissues. Each dot represents a patient sample. **Fig. 1C.** Number of neomers and the number of substitutions for 2,577 patients (Spearman's rho = 0.75). **Fig. 1D.** Heatmap showing the Jaccard index for the

overlap of neomer sets associated with different cancer types. **Fig. 1E.** Heatmap showing the occurrence of neomers across patients for each cancer type. Each row represents a cancer type and each column a patient. The intensity of the heatmap (log2-scale) shows the number of neomers for each tissue set.

Figs. 2A-2D. Neomers can distinguish cancer features. **Fig. 2A-B.** Classifier accuracy **(A)** using an unsupervised classifier and F1 **(B)** score for the same classifier for each of the twenty-one cancer types. **Fig. 2C.** Separation of MSI and MSS samples using a supervised selection of neomers. **Fig. 2D.** Separation of POLE proficient and deficient samples using nullomers. In **Figs. 2C-D,** the vertical line displays the harmonic mean.

Figs. 3A-3G. Identification of cancer in liquid biopsy samples using neomers. **Fig. 3A.** Cancer status detection in lung patients and healthy controls from whole-genome sequencing of liquid biopsy samples (***p-value<0.0005, Mann-Whitney U). Number of neomers detected in cfDNA from healthy controls, lung cancer patients, and matching tumors. **Fig. 3B.** Number of neomers for lung cancer stratified by tumor stage (p-value<0.006, Kruskal-Wallis test). **Fig. 3C.** Number of neomers observed in ovarian samples and healthy controls **Fig. 3D.** Cancer status detection in ovarian cancer and controls using rare 13mers. (*p-value<0.03, Mann-Whitney U). **Fig. 3E.** Number of first order nullomers detected in ovarian samples and healthy controls (*p-value<0.01, Mann-Whitney U), **Fig. 3F.** Number of neomers for prostate and control. **Fig. 3G.** Jaccard similarity between prostate neomers found in cfDNA and tumor samples.

Figs. 4A-C. Neomers alter the activity of gene regulatory elements. **Figs. 4A-B.** Integrative Genomics Viewer track snapshot of the location of neomer resurfacing mutation in prostate cancer RPS2-SNHG9 (**A**) and TMEM127-CIAO1 (**B**). Presented are genebody location (GENCODE V36), regulatory element according to ENCODE cCRE (yellow = enhancer, red = promoter), and the ENCODE layered epigenetic enhancer mark H3K27ac. **Fig. 4C.** Relative luciferase units from a luciferase reporter assay for promoters containing either the reference or nullomer variant. Transfection efficiency was normalized using renilla luciferase and significance is calculated using a two-way ANOVA with multiple testing and Šidák correction.

Figs. 5A-E. Fig. 5A. Number of patients per cancer tissue. **Figs. 5B-C.** Number of nullomers resurfaced due to indels or substitutions observed for each tumor sample (per patient). **Fig. 5D.** Association between number of mutations and number of nullomers observed. **Fig. 5E.** Number of nullomers of different lengths observed per patient.

DETAILED DESCRIPTION OF EMBODIMENTS

Before the present methods and systems are described, it is to be understood that the present disclosure is not limited to the particular processes, compositions, or methodologies described, as these may vary. It is also to be understood that the terminology used in the description is for the purposes of describing the particular versions or embodiments only, and is not intended to limit the scope of the present disclosure. Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of embodiments of the present disclosure, the methods, devices, and materials in some embodiments are now described. All publications mentioned herein are incorporated by reference in their entireties. Nothing herein is to be construed as an admission that the present disclosure is not entitled to antedate such disclosure by virtue of prior invention.

Definitions

Unless specifically defined otherwise, all technical and scientific terms used herein shall be taken to have the same meaning as commonly understood by one of ordinary skill in the art (e.g., in cell culture, molecular genetics, microRNA and detection thereof, immunology, immunohistochemistry, protein chemistry, and biochemistry). The meaning and scope of the terms should be clear, however, in the event of any latent ambiguity, definitions provided herein take precedent over any dictionary or extrinsic definition. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular.

The indefinite articles "a" and "an," as used herein in the specification and in the claims, unless clearly indicated to the contrary, should be understood to mean "at least one."

The phrase "and/or," as used herein in the specification and in the claims, should be understood to mean "either or both" of the elements so conjoined, i.e., elements that are conjunctively present in some cases and disjunctively present in other cases. Other elements may optionally be present other than the elements specifically identified by the "and/or" clause, whether related or unrelated to those elements specifically identified unless clearly indicated to the contrary. Thus, as a non-limiting example, a reference to "A and/or B," when used in conjunction with open-ended language such as "comprising" can refer, in one embodiment, to A without B (optionally including elements other than B); in another embodiment, to B without A (optionally

including elements other than A); in yet another embodiment, to both A and B (optionally including other elements); etc.

As used herein in the specification and in the claims, "or" should be understood to have the same meaning as "and/or" as defined above. For example, when separating items in a list, "or" or "and/or" shall be interpreted as being inclusive, i.e., the inclusion of at least one, but also including more than one, of a number or list of elements, and, optionally, additional unlisted items. Only terms clearly indicated to the contrary, such as "only one of" or "exactly one of," or, when used in the claims, "consisting of," will refer to the inclusion of exactly one element of a number or list of elements. In general, the term "or" as used herein shall only be interpreted as indicating exclusive alternatives (i.e. "one or the other but not both") when preceded by terms of exclusivity, "either," "one of," "only one of," or "exactly one of." "Consisting essentially of," when used in the claims, shall have its ordinary meaning as used in the field of patent law.

The term "about" is used herein to mean within the typical ranges of tolerances in the art. For example, "about" can be understood as about 2 standard deviations from the mean. According to certain embodiments, when referring to a measurable value such as an amount and the like, "about" is meant to encompass variations of ±20%, ±10%, ±5%, ±1%, ±0.9%, ±0.8%, ±0.7%, ±0.6%, ±0.5%, ±0.4%, ±0.3%, ±0.2% or ±0.1% from the specified value as such variations are appropriate to perform the disclosed methods. When "about" is present before a series of numbers or a range, it is understood that "about" can modify each of the numbers in the series or range.

As used herein, the term "animal" includes, but is not limited to, humans and non-human vertebrates such as wild animals, rodents, such as rats, ferrets, and domesticated animals, and farm animals, such as dogs, cats, horses, pigs, cows, sheep, and goats. In some embodiments, the animal is a mammal. In some embodiments, the animal is a human. In some embodiments, the animal is a non-human mammal.

An "algorithm," "formula," or "model" is any mathematical equation, algorithmic, analytical or programmed process, or statistical technique that takes one or more continuous or categorical inputs (herein called "parameters") and calculates an output value, sometimes referred to as an "index" or "index value." Non-limiting examples of "formulas" include sums, ratios, and regression operators, such as coefficients or exponents, biomarker (e.g., nullomers disclosed herein) value transformations and normalizations (including, without limitation, those normalization schemes based on clinical parameters, such as gender, age, or ethnicity), rules and

guidelines, statistical classification models, and neural networks trained on historical populations. Of particular use in combining markers are linear and non-linear equations and statistical classification analyses to determine the relationship between levels of the biomarkers detected in a subject sample and the subject's risk of disease (for example). In panel and combination construction, of particular interest are structural and syntactic statistical classification algorithms, and methods of risk index construction, utilizing pattern recognition features, including established techniques such as cross correlation, Principal Components Analysis (PCA), factor rotation, Logistic Regression (LogReg), Linear Discriminant Analysis (LDA), Eigengene Linear Discriminant Analysis (ELDA), Support Vector Machines (SVM), Random Forest (RF), Recursive Partitioning Tree (RPART), as well as other related decision tree classification techniques, Shruken Centroids (SC), StepAIC, Kth-Nearest Neighbor, Boosting, Decision Trees, Neural Networks, Bayesion Networks, Support Vector Machines, and Hidden Markov Models, among others. Many of these techniques are useful either combined with a biomarker selection technique, such as forward selection, backwards selection, or stepwise selection, complete enumeration of all potential panels of a given size, genetic algorithms, or they may themselves include biomarker selection methodologies in their own technique. These may be coupled with information criteria, such as Akaike's Information Criterion (AIC) or Bayes Information Criterion (BIC), in order to quantify the trade-off between additional biomarkers and model improvement, and to aid in minimizing overfit. The resulting predictive models may be validated in other studies, or cross-validated in the study they were originally trained in, using such techniques as Leave-One-Out (LOO) and 10-Fold cross-validation (10-Fold-CV).

The term "at least" prior to a number or series of numbers (*e.g.* "at least two") is understood to include the number adjacent to the term "at least," and all subsequent numbers or integers that could logically be included, as clear from context. When "at least" is present before a series of numbers or a range, it is understood that "at least" can modify each of the numbers in the series or range.

The term "biomarker" as used herein refers to a biological molecule present in an individual at varying concentrations useful in predicting the cancer status of an individual. A biomarker may include but is not limited to, nucleic acids, proteins and variants and fragments thereof. A biomarker may be DNA comprising the entire or partial nucleic acid sequence encoding the biomarker, or the complement of such a sequence. Biomarker nucleic acids useful in the disclosure

are considered to include both DNA and RNA comprising the entire or partial sequence of any of the nucleic acid sequences of interest. In some embodiments, the biomarker of the disclosure is any of the nullomers disclosed herein.

The term "bodily fluid" as used herein refers to a bodily fluid including blood (or a fraction of blood such as plasma or serum), lymph, mucus, tears, saliva, sweat, sputum, urine, semen, stool, cerebrospinal fluid (CSF), breast milk, and, ascites fluid. In some embodiments, the bodily fluid is blood. In some embodiments, the bodily fluid is a fraction of blood. In some embodiments, the bodily fluid is plasma. In some embodiments, the bodily fluid is serum. In some embodiments, the bodily fluid is urine. In some embodiments the bodily fluid is free of cells. In some embodiments, the bodily fluid comprises a circulating tumor cell. In some embodiments, the sample comprises cell-free nucleic acids.

The terms "cancer" and "cancerous" as used herein refer to or describe a physiological condition in mammals in which a population of cells are characterized by unregulated cell growth. Thus, the term "cancer" refers to a group of diseases involving abnormal cell growth with the potential to invade or spread to other parts of the body. Examples of cancer include, but not limited to, lung cancer, bone cancer, blood cancer, chronic myelomonocytic leukemia (CMML), bile duct cancer, cervical cancer, liver cancer, pancreatic cancer, skin cancer, cancer of the head and neck, cancer of the eye, cutaneous or intraocular melanoma, uterine cancer, ovarian cancer, rectal cancer, cancer of the anal region, stomach cancer, colon cancer, breast cancer, testicular cancer, gynecologic tumors (e.g., uterine sarcomas, carcinoma of the fallopian tubes, carcinoma of the endometrium, carcinoma of the cervix, carcinoma of the vagina or carcinoma of the vulva), Hodgkin's disease, cancer of the esophagus, cancer of the small intestine, cancer of the endocrine system (e.g., cancer of the thyroid, parathyroid or adrenal glands), sarcomas of soft tissues, cancer of the urethra, cancer of the penis, prostate cancer, chronic or acute leukemia, solid tumors of childhood, lymphocytic lymphomas, cancer of the bladder, cancer of the kidney or ureter (e.g., renal cell carcinoma, carcinoma of the renal pelvis), or neoplasms of the central nervous system (e.g., primary CNS lymphoma, spinal axis tumors, brain stem gliomas or pituitary adenomas).

As used herein, the term "characterizing cancer in a subject" refers to the identification of one or more properties of a cancer sample in a subject, including but not limited to, the presence of benign, pre-cancerous or cancerous tissue, the stage of the cancer, the type of the cancer, the tissue of origin of the cancer, and the subject's prognosis. Cancers may be characterized by the

identification of the expression of one or more cancer marker genes, including but not limited to, the nullomers disclosed herein. As used herein, the term "stage of cancer" refers to a qualitative or quantitative assessment of the level of advancement of a cancer. Criteria used to determine the stage of a cancer include, but are not limited to, the size of the tumor and the extent of metastases (e.g., localized or distant). In some embodiments, the subject has been previously diagnosed with having a cancer and received, or is currently receiving, cancer treatment, including but not limited to surgical intervention and cancer therapy, and in such embodiments, the term "characterizing cancer in a subject" refers to monitoring the progress of the cancer treatment.

The terms "complementary" or "complementarity" refer to polynucleotides (i.e., a sequence of nucleotides) related by base-pairing rules, for example, the sequence "5'-AGT-3'," is complementary to the sequence "5'-ACT-3'." Complementarity may be "partial," in which only some of the nucleic acids' bases are matched according to the base pairing rules, or there may be "complete" or "total" complementarity between the nucleic acids. The degree of complementarity between nucleic acid strands can have significant effects on the efficiency and strength of hybridization between nucleic acid strands under defined conditions. This is of particular importance for methods that depend upon binding between nucleic acid bases.

As used herein, the terms "comprising" (and any form of comprising, such as "comprise," "comprises," and "comprised"), "having" (and any form of having, such as "have" and "has"), "including" (and any form of including, such as "includes" and "include"), or "containing" (and any form of containing, such as "contains" and "contain"), are inclusive or open-ended and do not exclude additional, unrecited elements or method steps.

The term "correlate" or "correlating" as used herein refers to a statistical association between instances of two events, where events may include numbers, data sets, and the like. For example, when the events involve numbers, a positive correlation (also referred to herein as a "direct correlation") means that as one increases, the other increases as well. A negative correlation (also referred to herein as an "inverse correlation") means that as one increases, the other decreases. The disclosure provides nullomers, the levels of which are correlated with a particular outcome measure, such as between the presence of a particular nullomer and the likelihood of developing a particular type of cancer. For example, the increased level of a nullomer may be negatively correlated with a likelihood of good clinical outcome for the patient. In this case, for example, the patient may have a decreased likelihood of long-term survival without

recurrence of the cancer and/or a positive response to a chemotherapy, and the like. Such a negative correlation indicates that the patient likely has a poor prognosis or will respond poorly to a chemotherapy, and this may be demonstrated statistically in various ways, e.g., by a high hazard ratio.

As used herein, the terms "detect," "detecting" or "detection" refer to either the general act of discovering or discerning or the specific observation of a composition. Detecting a composition may comprise determining the presence or absence of a composition. Detecting may comprise quantifying a composition. For example, detecting comprises determining the expression level of a composition. The composition may comprise a nucleic acid molecule. For example, the composition may comprise one or a plurality of the nullomers disclosed herein. Alternatively, or additionally, the composition may be a detectably labeled composition.

The term "diagnosis" or "prognosis" as used herein refers to the use of information (e.g., genetic information or data from other molecular tests on biological samples, signs and symptoms, physical exam findings, cognitive performance results, etc.) to anticipate the most likely outcomes, timeframes, and/or response to a particular treatment for a given disease, disorder, or condition, based on comparisons with a plurality of individuals sharing common nucleotide sequences, symptoms, signs, family histories, or other data relevant to consideration of a patient's health status.

The terms "functional fragment" means any portion of a polypeptide or nucleic acid sequence from which the respective full-length polypeptide or nucleic acid relates that is of a sufficient length and has a sufficient structure to confer a biological affect that is similar or substantially similar to the full-length polypeptide or nucleic acid upon which the fragment is based. In some embodiments, a functional fragment is a portion of a full-length or wild-type nucleic acid sequence that encodes any one of the nucleic acid sequences disclosed herein, and said portion encodes a polypeptide of a certain length and/or structure that is less than full-length but encodes a domain that still biologically functional as compared to the full-length or wild-type protein. In some embodiments, the functional fragment may have a reduced biological activity, about equivalent biological activity, or an enhanced biological activity as compared to the wild-type or full-length polypeptide sequence upon which the fragment is based. In some embodiments, the functional fragment is derived from the sequence of an organism, such as a human. In such embodiments, the functional fragment may retain about 99%, 98%, 97%, 96%, 95%, 94%, 93%,

92%, 91%, or 90% sequence identity to the wild-type or given sequence upon which the sequence is derived. In some embodiments, the functional fragment may retain about 85%, 80%, 75%, 70%, 65%, or 60% sequence identity to the wild-type sequence upon which the sequence is derived. In some embodiments, the given sequence is a nullomer sequence of Table 1 or Table B. In other embodiments, the given sequence is a complementary sequence of any of the nullomer sequences of Table 1 or Table B.

The term "hyperproliferation" as used herein is defined as clonal expansion, in which daughter cells share a set of somatic mutations that were not originally present in the germline and which could include but are not limited to driver mutations. Clonal expansion could include but is not limited to resistance to cell death, evasion of growth suppressors, sustaining proliferate signaling, enabling replicative immortality, activating invasion and metastasis or inducing angiogenesis.

The term "hyperproliferative cell" refers to a cell located in a tissue or organ having a "hyperproliferative disorder," a disease or disorder characterized by abnormal proliferation, abnormal growth, abnormal senescence, abnormal quiescence, or abnormal removal of cells in an organism, and includes all forms of hyperplasias, neoplasias, and cancer. In some embodiments, the "hyperproliferative cell" is a precancerous cell in form of hyperplasias. In some embodiments, the "hyperproliferative cell" is precancerous cell in form of neoplasias. In some embodiments, the "hyperproliferative cell" is a cancerous cell. In some embodiments, the hyperproliferative disorder or disease is a cancer derived from the gastrointestinal tract or urinary system. In some embodiments, a hyperproliferative disorder or disease is a cancer of the adrenal gland, bile ducts, bladder, blood, bone, bone marrow, brain, breast, cervix, colon, esophagus, eye, gall bladder, ganglia, gastrointestinal tract, heart, lymphatic system, liver, lung, kidney, muscle, ovary, pancreas, parathyroid, penis, prostate, prostate glands, rectum, salivary glands, skin, spine, stomach, spleen, testis, thymus, thyroid, or uterus. In some embodiments, the term hyperproliferative disorder or disease is a cancer chosen from: lung cancer, bone cancer, blood cancer, chronic myelomonocytic leukemia (CMML), bile duct cancer, cervical cancer, liver cancer, pancreatic cancer, skin cancer, cancer of the head and neck, cancer of the eye, cutaneous or intraocular melanoma, uterine cancer, ovarian cancer, rectal cancer, cancer of the anal region, stomach cancer, colon cancer, breast cancer, testicular cancer, gynecologic tumors (e.g., uterine sarcomas, carcinoma of the fallopian tubes, carcinoma of the endometrium, carcinoma of the

cervix, carcinoma of the vagina or carcinoma of the vulva), Hodgkin's disease, cancer of the esophagus, cancer of the small intestine, cancer of the endocrine system (e.g., cancer of the thyroid, parathyroid or adrenal glands), sarcomas of soft tissues, cancer of the urethra, cancer of the penis, prostate cancer, chronic or acute leukemia, solid tumors of childhood, lymphocytic lymphomas, cancer of the bladder, cancer of the kidney or ureter (e.g., renal cell carcinoma, carcinoma of the renal pelvis), or neoplasms of the central nervous system (e.g., primary CNS lymphoma, spinal axis tumors, brain stem gliomas or pituitary adenomas). In some embodiments, the hyperproliferative disorder or disease is a breast cancer, pancreatic cancer, esophagus cancer, lymphoid cancer, kidney cancer, ovary cancer, head and neck cancer, lung cancer, stomach cancer, CNS cancer, uterus cancer, skin cancer, colorectal cancer, prostate cancer, bladder cancer, bone and soft tissue cancer, biliary cancer, cervix cancer, thyroid cancer, myeloid cancer, or liver cancer. In some embodiments, the hyperproliferative disorder or disease comprises one or a plurality of mutations in one or a plurality of genes selected from Table A.

Table A. Cancer-related genes and their corresponding GenBank accession numbers.

Gene	**GenBank**	**Gene**	**GenBank**
TP53	P04637.4	MUC17	Q685J3.2
PIK3CA	P42336.2	SPOP	O43791.1
APC	P25054.2	NAV3	Q8IVL0.3
KRAS	P01116.1	FLG	P20930.3
ARID1A	O14497.3	TTN	Q8WZ42.4
BRAF	P15056.4	LRP1B	Q9NZR2.2
PTEN	P60484.1	HMCN1	Q96RW7.2
IDH1	O75874.2	KMT2C	Q8NEZ4.3
CDKN2A	Q8N726.2	CDKN2B	P42772.1
MSH6	P52701.2	MYC	P01106.1
RB1	P06400.2	SOX2	P48431.1
CTNNB1	P35222.1	MCL1	Q07820.3
ATM	Q13315.4	HRAS	P01112.1
FAT1	Q14517.2	BRCA1	P38398.2
TP63	Q9H3D4.1	BRCA2	P51587.3

FBXW7	Q969H0.1	PALB2	Q86YC2.1
FAT4	Q6V0I7.2	RAD51D	O75771.1
BAP1	Q92560.2	TERT	O14746.1
RPL22	P35268.2	RECQL	P46063.3
TSHZ3	Q63HK5.2	CHEK2	O96017.1
RUNX1	Q01196.3	SMAD4	Q13485.1
NRAS	P01111.1	MSH2	P43246.1
EGFR	P00533.2	MUTYH	Q9UIF7.1
PIK3R1	P27986.2	HNF1A	P20823.2
NPM1	P06748.2	KIT	P10721.1

As used herein, the phrase "in need thereof" means that the animal or mammal has been identified or suspected as having a need for the particular method or treatment. In some embodiments, the identification can be by any means of diagnosis or observation. In any of the methods and treatments described herein, the animal or mammal can be in need thereof.

The term "label" as used herein refers to any atom or molecule that can be used to provide a detectable (preferably quantifiable) effect, and that can be attached to a nucleic acid or protein. Labels include but are not limited to dyes; radiolabels such as ^{2}P; binding moieties such as biotin; haptens such as digoxgenin; luminogenic, phosphorescent or fluorogenic moieties; and fluorescent dyes alone or in combination with moieties that can suppress or shift emission spectra by fluorescence resonance energy transfer (FRET). Labels may provide signals detectable by fluorescence, radioactivity, colorimetry, gravimetry, X-ray diffraction or absorption, magnetism, enzymatic activity, and the like. A label may be a charged moiety (positive or negative charge) or alternatively, may be charge neutral. Labels can include or consist of nucleic acid or protein sequence, so long as the sequence comprising the label is detectable. In some embodiments, nucleic acids are detected directly without a label (e.g., directly reading a sequence).

The term "level" as used herein refers to qualitative or quantitative amount of the number of copies of a nullomer. A nullomer exhibits an "increased level" when the level of the nullomer is higher in a first sample, such as in a clinically relevant subpopulation of patients (e.g., patients who have cancer), than in a second control sample, such as in a related subpopulation (e.g., patients who do not have cancer). In the context of an analysis of a level of a nullomer in a tumor sample

obtained from an individual patient, a nullomer exhibits "increased level" when the level of the nullomer in the subject trends toward, or more closely approximates, the level characteristic of a clinically relevant subpopulation of patients.

The term "measuring" or "measurement" means assessing the presence, absence, quantity or amount (which can be an effective amount) of either a given substance within a clinical or subject-derived sample, including the derivation of qualitative or quantitative concentration levels of such substances, or otherwise evaluating the values or categorization of a subject's clinical parameters. Alternatively, the term "detecting" or "detection" may be used and is understood to cover all measuring or measurement as described herein.

The term "metastasis" as used herein refers to the process by which a cancer spreads or transfers from the site of origin to other regions of the body. A "metastatic" or "metastasizing" cell is one that loses adhesive contacts with neighboring cells and migrates (e.g., via the bloodstream or lymph) from the primary site of disease to secondary sites.

The particular use of terms "nucleic acid," "oligonucleotide," and "polynucleotide" should in no way be considered limiting and may be used interchangeably herein. "Oligonucleotide" is used when the relevant nucleic acid molecules typically comprise less than about 100 bases. "Polynucleotide" is used when the relevant nucleic acid molecules typically comprise more than about 100 bases. Both terms are used to denote a DNA, RNA, modified or synthetic DNA or RNA sequence (including, but not limited to nucleic acids comprising synthetic and naturally-occurring base analogs, dideoxy or other sugars, thiols or other non-natural or natural polymer backbones), or other nucleobase containing polymers capable of hybridizing to DNA and/or RNA. Accordingly, the terms should not be construed to define or limit the length of the nucleic acids referred to and used herein, nor should the terms be used to limit the nature of the polymer backbone to which the nucleobases are attached.

The term "nucleic acid sequence" or "polynucleotide sequence" refers to a contiguous string of nucleotide bases and in particular contexts also refers to the particular placement of nucleotide bases in relation to each other as they appear in a polynucleotide.

"Nucleobase" means a heterocyclic moiety capable of non-covalently pairing with another nucleobase.

"Nucleoside" means a nucleobase linked to a sugar moiety.

"Nucleotide" means a nucleoside having a phosphate group covalently linked to the sugar

portion of a nucleoside. In some embodiments, the nucleotide is characterized as being modified if the 3' phosphate group is covalently linked to a contiguous nucleotide by any linkage other than a phosphodiester bond.

"Compound comprising a modified oligonucleotide consisting of a number of linked nucleosides" means a compound that includes a modified oligonucleotide having the specified number of linked nucleosides. Thus, the compound may include additional substituents or conjugates. Unless otherwise indicated, the compound does not include any additional nucleosides beyond those of the modified oligonucleotide.

"Modified oligonucleotide" means an oligonucleotide having one or more modifications relative to a naturally occurring terminus, sugar, nucleobase, and/or internucleoside linkage. A modified oligonucleotide may comprise unmodified nucleosides.

"Single-stranded modified oligonucleotide" means a modified oligonucleotide which is not hybridized to a complementary nucleic acid strand.

"Modified nucleoside" means a nucleoside having any change from a naturally occurring nucleoside. A modified nucleoside may have a modified sugar, and an unmodified nucleobase. A modified nucleoside may have a modified sugar and a modified nucleobase. A modified nucleoside may have a natural sugar and a modified nucleobase. In some embodiments, a modified nucleoside is a bicyclic nucleoside. In some embodiments, a modified nucleoside is a non-bicyclic nucleoside.

The term "nullomers" as used herein refers to expressed oligonucleotide sequences in a species, the genetic templates of which are congenitally absent in the species. In some embodiments, the nullomers of the disclosure are nullomers not present in the published human genome sequences. In some embodiments, the nullomers of the disclosure are nullomers not present in the published human genome sequences and associated with one or a plurality of cancers.

As used herein "one or more of" includes at least one of the recited components, or 2, 3, 4, 5, or 5 etc. of the recited components. In some embodiments, the phase includes all of the recited components.

Ranges provided herein are understood to include all individual integer values and all subranges within the ranges.

As used herein, the term "sample" refers to a biological sample obtained or derived from a source of interest, as described herein. In some embodiments, a source of interest comprises an

organism, such as an animal or human. In some embodiments, a biological sample comprises biological tissue or fluid. In some embodiments, a biological sample may be or comprise bone marrow, blood, blood cells, cells from a hair sample, ascites, tissue or fine needle biopsy samples, cell-containing body fluids, free floating nucleic acids, sputum, saliva or spit, urine, cerebrospinal fluid, peritoneal fluid, pleural fluid, feces, lymph, gynecological fluids, skin swabs, vaginal swabs, oral swabs, nasal swabs, washings or lavages such as a ductal lavages or broncheoalveolar lavages, aspirates, scrapings, bone marrow specimens, tissue biopsy specimens, surgical specimens, feces, other body fluids, secretions and/or excretions, and/or cells therefrom, etc. In some embodiments, the sample is a brush biopsy, puncture biopsy, or fluid from a needle biopsy. In some embodiments, the sample is blood or blood cells. In some embodiments, the sample is cells from a hair sample or nucleic acids from a hair sample. In some embodiments, the sample is sputum, saliva or spit. In some embodiments, a biological sample is or comprises cells obtained from an individual. In some embodiments, a sample is a "primary sample" obtained directly from a source of interest by any appropriate means. For example, in some embodiments, a primary biological sample is obtained by methods selected from the group consisting of biopsy (e.g., fine needle aspiration or tissue biopsy), surgery, collection of body fluid (e.g., blood, lymph, feces etc.), etc. In some embodiments, as will be clear from context, the term "sample" refers to a preparation that is obtained by processing (e.g., by removing one or more components of and/or by adding one or more agents to) a primary sample. For example, filtering using a semi-permeable membrane. Such a "processed sample" may comprise, for example nucleic acids or proteins extracted from a sample or obtained by subjecting a primary sample to techniques such as amplification or reverse transcription of mRNA, isolation and/or purification of certain components, etc.

As used herein, the term "minimal residual disease" refers to a small number of cancer cells remaining in the body after treatment or surgical intervention. These cells cannot usually be detected by standard scans or tests, due to lower abundance than detection sensitivity thresholds.

A "score" is a value or set of values selected so as to provide a normalized quantitative measure of a variable or characteristic of a subject's condition, and/or to discriminate, differentiate or otherwise characterize a subject's condition. The value(s) comprising the score can be based on, for example, quantitative data resulting in a measured amount of one or more sample constituents obtained from the subject, or from clinical parameters, or from clinical assessments, or any combination thereof. In certain embodiments, the score can be derived from a single

constituent, parameter or assessment, while in other embodiments the score is derived from multiple constituents, parameters and/or assessments. The score can be based upon or derived from an interpretation function; e.g., an interpretation function derived from a particular predictive model using any of various statistical algorithms known in the art. A "change in score" can refer to the absolute change in score, e.g. from one time point to the next, or the percent change in score, or the change in the score per unit time (i.e., the rate of score change). In some embodiments, the score is calculated through an interpretation function or algorithm. In some embodiments, the subject is suspected of having expression of a gene that promotes or contributes to the likelihood of acquiring a disease state or whose expression is correlative to the presence of a pathogen. Calculation of score can be accomplished using known algorithms executable in computer program products within equipment used in sequencing or analyzing samples. In some embodiments, the methods disclosed herein comprise substeps of detecting the presence, absence or quantity of a given biomarker by calculating the quantity of a probe in a control sample, calculating the quantity of a probe in the subject sample, and normalizing the signal obtained from the subject sample by subtracting the signal obtained from the control sample.

As used herein, "sequence identity" is determined by using the stand-alone executable BLAST engine program for blasting two sequences (bl2seq), which can be retrieved from the National Center for Biotechnology Information (NCBI) ftp site, using the default parameters (Tatusova and Madden, FEMS Microbiol Lett., 1999, 174, 247-250; which is incorporated herein by reference in its entirety). Alternatively, "% sequence identity" can be determined using the EMBOSS Pairwise Alignment Algorithms tool available from The European Bioinformatics Institute (EMBL-EBI), which is part of the European Molecular Biology Laboratory (EMBL). This tool is accessible at the website ebi.ac.uk/Tools/emboss/align/. This tool utilizes the Needleman-Wunsch global alignment algorithm (Needleman, S. B. and Wunsch, C. D. (1970) J. Mol. Biol. 48, 443-453; Kruskal, J. B. (1983) An overview of sequence comparison, In D. Sankoff and B. Kruskal, (ed.), Time warps, string edits and macromolecules: the theory and practice of sequence comparison, pp. 1-44, Addison Wesley). Default settings are utilized which include Gap Open: 10.0 and Gap Extend 0.5. The default matrix "Blosum62" is utilized for amino acid sequences and the default matrix "DNAfull" is utilized for nucleic acid sequences.

As used herein, the term "statistically significant" means an observed alteration is greater than what would be expected to occur by chance alone (e.g., a "false positive"). Statistical

significance can be determined by any of various methods well-known in the art. An example of a commonly used measure of statistical significance is the p-value. The p-value represents the probability of obtaining a given result equivalent to a particular datapoint, where the datapoint is the result of random chance alone. A result is often considered highly significant (not random chance) at a p-value less than or equal to about 0.05.

The terms "subject," "individual," and "patient" are used interchangeably herein to refer to a vertebrate, preferably a mammal, more preferably a human. Mammals include, but are not limited to, murine, simians, humans, farm animals, cows, pigs, goats, sheep, horses, dogs, sport animals, and pets. Tissues, cells and their progeny obtained in vivo or cultured in vitro are also encompassed by the definition of the term "subject." In some embodiments, the subject is a human. For treatment of those conditions which are specific for a specific subject, such as a human being, the term "patient" may be interchangeably used. In some instances in the description of the present disclosure, the term "patient" will refer to human patients suffering from a particular disease or disorder. In some embodiments, the subject may be a non-human animal. The term "mammal" encompasses both humans and non-humans and includes but is not limited to humans, non-human primates, canines, felines, murine, bovines, equines, caprine, and porcines.

By "substantially identical" is meant a nucleic acid molecule (or polypeptide) comprises at least about 50% sequence identity to a reference nucleic acid sequence (for example, any one of the nucleic acid sequences described herein) or amino acid sequence. In some embodiments, such a sequence is at least about 60%, 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, or even 99% identical at the nucleic acid level or amino acid level to the reference sequence used for comparison.

As used herein, the term "therapeutic" means an agent utilized to treat, combat, ameliorate, prevent or improve an unwanted condition or disease of a patient.

The term "therapeutically effective amount" means a quantity sufficient to achieve a desired therapeutic effect, for example, an amount which results in the prevention or amelioration of or a decrease in the symptoms associated with a disease that is being treated, e.g., disorders associated with cancer growth or a hyperproliferative disorder. The amount of compound administered to the subject will depend on the type and severity of the disease and on the characteristics of the individual, such as general health, age, sex, body weight and tolerance to drugs. It will also depend on the degree, severity and type of disease. The skilled artisan will be

able to determine appropriate dosages depending on these and other factors. The regimen of administration can affect what constitutes an effective amount. Further, several divided dosages, as well as staggered dosages, can be administered daily or sequentially, or the dose can be continuously infused, or can be a bolus injection. Further, the dosages of the compound(s) of the disclosure can be proportionally increased or decreased as indicated by the exigencies of the therapeutic or prophylactic situation. Typically, an effective amount of the compounds of the present disclosure, sufficient for achieving a therapeutic effect, range from about 0.000001 mg per kilogram body weight per day to about 10,000 mg per kilogram body weight per day. Preferably, the dosage ranges are from about 0.0001 mg per kilogram body weight per day to about 100 mg per kilogram body weight per day. The compounds disclosed herein can also be administered in combination with each other, or with one or more additional therapeutic compounds.

The terms "treatment" or "treating" as used herein is an approach for obtaining beneficial or desired results including clinical results for the subject. For purposes herein, beneficial or desired clinical results include, but are not limited to, one or more of the following: (1) preventing or delaying the appearance of clinical symptoms of the state, disorder, or condition developing in a person who may be afflicted with or predisposed to the state, disorder or condition but does not yet experience or display clinical symptoms of the state, disorder or condition; (2) inhibiting the state, disorder or condition, i.e., arresting, reducing or delaying the development of the disease or a relapse thereof (in case of maintenance treatment) or at least one clinical symptom, sign, or test, thereof; or (3) relieving the disease, i.e., causing regression of the state, disorder or condition or at least one of its clinical or sub-clinical symptoms or signs. In some embodiments, a subject is successfully "treated" according to the methods of the present disclosure if the patient shows one or more of the following: a reduction in the number of and/or complete absence of cancer cells; a reduction in the tumor size; an inhibition of tumor growth; inhibition of and/or an absence of cancer cell infiltration into peripheral organs including the spread of cancer cells into soft tissue and bone; inhibition of and/or an absence of tumor or cancer cell metastasis; inhibition and/or an absence of cancer growth; relief of one or more symptoms associated with the specific cancer; reduced morbidity and mortality; improvement in quality of life; reduction in tumorigenicity; reduction in the number or frequency of cancer stem cells; or some combination of such effects.

The term "tumor" as used herein, refers to all neoplastic cell growth and proliferation, whether malignant or benign, and all pre-cancerous and cancerous cells and tissues. A "benign"

tumor is not cancerous and it does not invade nearby tissue or spread to other parts of the body. A "premalignant" tumor is a tumor which is not yet cancerous but has the potential to become malignant. A "malignant" tumor, on the other hand, is cancerous and can grow and spread to other parts of the body.

The term "tumor sample" as used herein refers to a sample comprising tumor material obtained from a cancer patient. The term encompasses tumor tissue samples, for example, tissue obtained by surgical resection and tissue obtained by biopsy, such as for example, a core biopsy or a fine needle biopsy. In some embodiments, the tumor sample is a fixed, wax-embedded tissue sample, such as a formalin-fixed, paraffin-embedded tissue sample. Additionally, the term "tumor sample" encompasses a sample comprising tumor cells obtained from sites other than the primary tumor, e.g., circulating tumor cells. The term also encompasses cells that are the progeny of the patient's tumor cells, e.g. cell culture samples derived from primary tumor cells or circulating tumor cells. The term further encompasses samples that may comprise protein or nucleic acid material shed from tumor cells in vivo, e.g., bone marrow, blood, plasma, serum, and the like. The term also encompasses samples that have been enriched for tumor cells or otherwise manipulated after their procurement and samples comprising polynucleotides and/or polypeptides that are obtained from a patient's tumor material.

Identification of Nullomers

The identification of nullomers can be performed using any methods known in the art. In some embodiments, the identification of nullomers of the disclosure is performed as previously described in Georgakopoulos-Soares et al., published in bioRxiv, available at biorxiv.org/content/10.1101/2020.03.02.972422v1, incorporated by reference herein. As a first step, a dataset is obtained. In some embodiments, the dataset is obtained from WGS cancers from ICGC under the project PanCancer Analysis of Whole Genomes (ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes, Nature, 2020, 578:82-93), which includes 46 cancer projects from 21 organs. WGS patients were analyzed using the GRCh37 (hg19) reference assembly of the human genome.

In some embodiments, somatic indel calls are performed using three pipelines from four somatic variant callers. These are the Wellcome Sanger Institute pipeline, the DKFZ/ EMBL pipeline and the Broad Institute pipeline, with somatic variant false discovery rate of about 2.5%. In some embodiments, indel calling is performed by those algorithms and only indels called by at

least two of the callers were analyzed, therefore generating a conservative dataset. As a result, the false negative rate of indel detection can be higher than that of other methods, and of each pipeline separately, which implies that many indels present in the samples were not identified successfully. For a small subset of indels, in some embodiments, the indel calls are visually examined using JBrowse Genome Browser32, to inspect the number of reads reporting the indel, if the indel calls are biased towards the end of the sequencing reads or if there were other systematic biases between the normal and tumor sequencing reads; such biases could not be identified.

In some embodiments, Bedtools intersect utility is used to measure overlap between indels and polyN tracts. The term overlap in this context refers to deleted bases occurring at any position across the entire length of the repeat or inserted bases occurring at any position across the length of the repeat and immediately before or after the repeat. Indel density is defined as the number of indel mutations for a given number of bases.

In some embodiments, the distance between each pair of consecutive indels is calculated per patient. In some embodiments, indels in different chromosomes are excluded because their pairwise distance cannot be defined. In some embodiments, the same analysis is performed separately for insertions and deletions.

In some embodiments, substitution calling is performed using four somatic mutation-calling algorithms, with mutation calls being shared by at least two algorithms. In the embodiments for lung cancers, C > A substitutions can be examined with respect to transcriptional strand asymmetries at polyG tracts and replication timing.

In some embodiments, the numbers of indels overlapping motifs found in the template or non-template strands are obtained using the bedtools intersect command. In some embodiments, strand bias is calculated for the vector of genes, reporting the number of polyN motif occurrences and the number of overlapping motifs as:

A = (indels overlapping motif at non-template)/(motif occurrences at non-template)

B = (indels overlapping motif at template)/(motif occurrences at template)

Strand bias = A/(A + B)

with motifs representing polyN repeat tracts of size 2-10 bp and dinucleotide repeat tracts of 1-5 repeated units, at genic regions.

In some embodiments, bootstrapping with replacement, randomly selecting the indels overlapping motifs at template and non-template strands from each randomly selected gene are

performed for equal number of genes in multiple iterations, from which the standard deviation for the strand bias can be calculated.

The nullomers can be of any length. In some embodiments, the nullomers are in a length of from about 8 to about 50 nucleotides. In some embodiments, the nullomers are in a length of from about 10 to about 45 nucleotides. In some embodiments, the nullomers are in a length of from about 12 to about 40 nucleotides. In some embodiments, the nullomers are in a length of from about 14 to about 30 nucleotides. In some embodiments, the nullomers are in a length of from about 16 to about 20 nucleotides. In some embodiments, the nullomers are in a length of from about 8 nucleotides. In some embodiments, the nullomers are in a length of about 10 nucleotides. In some embodiments, the nullomers are in a length of about 11 nucleotides. In some embodiments, the nullomers are in a length of about 12 nucleotides. In some embodiments, the nullomers are in a length of about 13 nucleotides. In some embodiments, the nullomers are in a length of about 14 nucleotides. In some embodiments, the nullomers are in a length of about 15 nucleotides. In some embodiments, the nullomers are in a length of about 16 nucleotides. In some embodiments, the nullomers are in a length of about 17 nucleotides. In some embodiments, the nullomers are in a length of about 18 nucleotides. In some embodiments, the nullomers are in a length of about 19 nucleotides. In some embodiments, the nullomers are in a length of about 20 nucleotides. In some embodiments, the nullomers are in a length of about 25 nucleotides. In some embodiments, the nullomers are in a length of about 30 nucleotides. In some embodiments, the nullomers are in a length of about 35 nucleotides. In some embodiments, the nullomers are in a length of about 40 nucleotides. In some embodiments, the nullomers are in a length of about 45 nucleotides. In some embodiments, the nullomers are in a length of about 50 nucleotides. In some embodiments, the nullomers are in a length of more than about 50 nucleotides.

Neomers as Biomarkers for Cancer

The disclosure provides nullomers identified in cancers of numerous organs or tissues, including pancreas, esophagus, lymphoid, kidney, ovary, head and neck, lung, stomach, liver, CNS, uterus, skin, colorectal, prostate, bladder, bone and soft tissue, breast, biliary, cervix, thyroid and myeloid. The neomers of the disclosure are provided in Table 1 or Table B.

In some embodiments, the disclosure relates to a nullomer comprising at least about 60%, 65%, 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89% 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 97%, 98, 99% or 100% sequence identity to any of the sequences provided in Table 1 or Table B.

In some embodiments, the disclosure relates to a neomer comprising any of the sequences provided in Table 1 or Table B. In some embodiments, the disclosure relates to a nucleic acid sequence that is complementary to any of the sequences provided in Table 1 or Table B. In some embodiments, the disclosure relates to relates to a system comprising a solid support and one or a plurality of nucleic acid sequences that is complementary to any of the sequences provided in Table 1. In some embodiments, the disclosure relates to relates to a system comprising a solid support and one or a plurality of nucleic acid sequences that is complementary to any of the sequences provided in Table 1 immobilized to a surface of the solid support.

List: Nullomer Sequences for Lung and Ovary

PCT/US2022/027536 describes nullomer detection steps and is incorporated by reference in its entirety herein.

Legend: Correspondence between DNA dinucleotides and encoded representation in the list of nullomer sequences. Each character represents a combination of two consecutive nucleotides in the nullomer sequence, e.g., AA is represented by "b", AC is represented by "d," etc.

First base/second base	A	C	G	T
A	b	d	E	f
C	h	i	J	k
G	l	o	P	q
T	r	s	V	x

Lung

kxppqdse;dqxrojpf;ijorbdqv;ipkbbjxl;jorbdqvb;pkbbjxlb;sipkbbjx;efdfbqlp;drllrhre;keefd fbq;lrhreveo;vdkphrlb;hffposhf;vhhsxkbk;khvsrvlf;rdbvxvxh;dxflrokh;jdidiixx;bfhbdoxx;vidssiij ;ohkkiiio;xexoebvq;ifsvhdrd;hejvqpbp;iboqpvle;hkdfivev;foeihkix;xkexlrdb;jvkrbbdp;bodqvlbe; qxxrehjx;xbodqvlb;xrehjxpb;xxbodqvl;oierkefe;srsrqdvp;lqpkorvi;kvofehoh;ofehohkh;bbhvvdkq ;dvvhbsko;elxvhhqq;frobpjvl;eheoqpeq;obpjvlpv;robpjvlp;xeheoqpe;hvlhhvxf;hrdfqqif;sfbhvvsh; livbqbsr;pqefrkse;prlxdxhp;vspllsbl;fqjpefhe;jfqjpefh;lvspllsb;xldiddkh;fvbihhis;dksxdhxk;eokfb vhe;pirrfobq;vidqkxkp;eiebedqp;sivhekrp;dsirekvh;hkikboxo;idsirekv;loidsire;oidsirek;dxrqekhx ;khjidssd;hkbvrhhb;lrbkoxhv;ihvbokos;kbhpbbei;erffhrxl;qbrsffvb;fvqliseh;hlbrhroe;kkllkkfv;vie qsbhr;hrxbefrj;dqbsxrff;kpverkve;bkpverkv;dvqlqdxl;bexfosxd;qxvldkqo;bxdhqrfp;iqxxrehj;bod

qvlbe;qxxrehjx;xbodqvlb;xrehjxpb;xxbodqvl;rifdbvqf;fprvlpde;fvrvqfpe;qfprvlpd;vrvqfpel;xqfpr
vlp;flpfpfql;vsdfifis;xvrboxfp;djjhjoke;irpiqoop;qkqeibvq;fvorhlif;foxrkekq;opovhpsd;ephposql;j
vdkoeij;sqlivhoi;vvdkoeij;ivkiifqf;dejrhxse;drdejrhx;fdrdejrh;fhrkbhoq;hrkbhoqd;jrhxserl;kbhoq
dfx;rkbhoqdf;sfdrdejr;exflsxrx;bxlqlfhr;prxierxr;fxjqxqsh;dbdjbfoi;dxpbhbil;hbilbvih;kvldbdjb;ld
bdjbfo;lkvldbdj;pkksibve;hposxihf;posxihfl;oebsfshq;epihekko;sllifixv;sbelvqkj;lbfodxqh;vbbvh
kvs;qfvsbbrv;oxbeirqk;pkrbokes;jkhsxkqq;fjkhsxkq;odshhfpq;phlodvvq;hlodvvqp;rrpddfbd;vprh
evlh;fqidxqdi;vpelfqxv;hihkqkvr;rbfkbkfs;xpexhfhv;qlfvlqsb;sflxldsh;ebxphfxf;bhlfjevx;ddslxkq
x;pkoevros;vovhqqek;rerrkrpl;fikbqfdx;selvqvbl;vlbxhqbr;vqeqbxis;fprprfsk;xdkxkosi;kpeexlvx;
kievessh;shqlqkie;kpvpebvd;korbkqrq;vksvxhke;drojevbx;hkeilqlf;kqsdroje;ksvxhkei;oxkqsdro;q
sdrojev;xhkeilql;sdksklxo;hfhpbpvd;hjxxoxxs;rkofoebi;dvhvhldh;sxhdhskl;prvisxie;pqfokksh;pv
prvisx;qfokkshq;qpqfokks;rvisxies;lqepboik;ielpxhdd;pvxldkiv;qvbiskpv;vqqvbisk;vxldkivq;xfid
rhxb;qevlrfff;vrqpfbrr;qsifoddo;xihvhhjk;hlqkxqbx;vfxvkxok;phbobbse;hroxvpqi;kxhrlobi;sfebe
bqr;hreofkpe;sdffbqqd;kehfrffo;ehqbevxl;vbeqsbeh;fppkqhll;epdkxbsi;rlxpksdk;pvbohfkb;isedsjl
f;rxipesve;bbxprlpr;hsxivovb;kxlohddv;pskllpbv;qiveelfq;vqiveelf;lxhlkrfl;hxbqklfh;prhbfxhr;qp
xvkpfi;bdhkkpdd;fiehbids;hbidsvlh;lshobdhk;plshobdh;shobdhkk;qjeqqopf;lsjhhisl;sjhhisli;rideh
vrh;dkkfxofd;kkfxofdh;qdfovvqf;xfkdsdfb;hjodfxfd;eqfbfqoj;rbvviqpv;khhjrodx;evkdqqlo;vkdqq
loh;xvbokqpk;oedvrboh;xrefbpro;phvlisfk;qereffpe;vrqblrvl;xqereffp;ppexjeod;eqokjbki;isldsikx
;kjbkiixs;okjbkiix;visldsik;xjeodkqq;xkfdebdi;qipispjd;esjokjol;hqipispj;ioesjokj;jhqipisp;sjhqipi
s;jvxpjpki;loiohddo;ebqedxxe;lerlkxrl;ofofxorf;ehbvhvki;eqxxvrfq;djsixxxx;isixxrqq;fqkirflr;fee
xsdbv;xxfhkdjb;fxpvoreq;pislqlrh;liqlhdvr;bfvxvpdj;djvdbkqf;iqlhdvrx;jvdbkqfv;pdjvdbkq;pliqlh
dv;qlhdvrxq;qxpliqlh;vpdjvdbk;vxvpdjvd;xpliqlhd;xvpdjvdb;obseleql;hfhpbpvd;fsjlqklp;rxipesve
;vkrpxlfo;xkqorkfk;qbifsxfr;ierlqksd;bksqqsbk;bdxjvxhd;dxjvxhdx;dblqksxp;qokjbkii;eqokjbki;i
sldsikx;kjbkiixs;okjbkiix;visldsik;xjeodkqq;eqrsxbks;eferodeb;qpjsvifv;ldfiokiv;pbhsjisk;rjvpblb
b;ddesielx;ibbiofqr;bbiofqrk;bdjhvxdv;biofqrkp;djhvxdvq;hbdjhvxd;iofqrkpr;lobolqov;lbsidxfl;h
revpfxh;qplxsbhf;dfoddrld;evdboxef;hsrekvsd;xxxrvsid;kpepkqev;qibvqlbe;edsdbhhf;hxqqvves;h
hokeohf;hxpirpkq;ffxvjqoe;djhbffvf;ervoxovh;hiobbrxl;kvkodjhb;odjhbffv;ovhiobbr;oxovhiob;qf
pkvkod;vhiobbrx;vkodjhbf;lhkveoeq;vddxlphp;rhbhqihb;pidibhvh;ohihdfoe;fsshvekl;kfkerkpo;sr
srqdvp;xqvbplih;kpqiksbh;pqiksbhd;sfvfxxsp;sjbbbsbv;jrfkxbrv;dqbsxrff;seqepqrl;hprppxed;pfb
vfxvd;bbsfrsjs;bfhxfiqi;bsfrsjsh;fhxfiqie;kvvfifve;irhbeosr;vfhvfrxv;proflrds;edbeqkkf;peijibhj;lo
iohddo;kdphfbib;kxpxfviq;vqrxojrq;xvqrxojr;llxqpvle;oedbpklr;ieqqiikr;ihpvsiix;iixfhokv;qiikrse
i;siixfhok;fjkihokf;lqhlkkqj;rbhrdrfs;elxvrprv;hpfxqeqf;odesxjhd;jblkqoee;oledvvhl;vjblkqoe;xol

edvvh;bovlrvsx;hsdvpvrx;eddhhplf;qrpldfhh;irdkqqib;kbivvshe;lvxikbiv;qlvxikbi;vfqsirdk;rhvee deq;svxikkij;vvobhhxf;rrfxqphd;hdelpvlr;fidkivxo;fbfvfref;erfiblee;rdhprkvh;dbvheriv;fobqdkpx; obqdkpxb;rhfobqdk;rrdsxbdr;lkobloel;dvhkrhdd;qvrevhqv;rrxrvjoh;ebrrxrvj;lfffxfop;bhosvqfd;ps feeldd;ixshfxvd;bhqvokrf;dexpixbr;erbdexpi;hqbbhqvo;ihqbbhqv;qvokrfbo;vokrfbod;lbevedhh;v ixvkvii;bebhoxix;ioodpilq;qledsrkp;eoqlvovq;fxdsxiik;sdkjxfid;hisqrshi;khisqrsh;oeihpexl;edsrof fe;xsvkbidr;ddfjbflx;hhslbvfx;pdrhqeij;sbrhxhhq;khlhqoir;vsxvqrpp;hpvxeddv;eqqrlhko;vshxbqd d;qifrerhi;fbxlfjvv;ddlxhfrs;srlflkxe;oqkrkqqv;xqokxlvl;llhrdori;xfpivlop;rvokpejq;oevvqevl;hpe opfqs;pbhsjisk;irhksxfb;xiehbphd;oxvhfsie;rsrfsvxi;qhbhxphr;brqflbhr;fqxhrkfs;xpdxsere;vlkxhq bp;slsevfss;kjfhqlxk;ifbrheqx;kherkhie;bkpverkv;dvqlqdxl;qkxoeovl;eksrserl;erlqhhkl;hqeesddv; bskpfrdv;lplljosv;lrehdsfl;idedhekx;rboxpsvv;ddkikirs;feovpejf;rehhfqxs;drhbhfxp;pxbvxlqk;qve orlle;xhhbhlvo;qvjblkqo;jblkqoee;oledvvhl;vjblkqoe;xoledvvh;pddseflx;ksbhlpro;iirdhblh;ovskie ls;rrvihbvo;ffoibfpi;qffoibfp;xfkdfkll;khlvrlre;ihobvohb;xfhrjxfx;srxisvxo;ehdeelfe;hdeelfeb;eeoe opfb;dbeehver;bvoexxxf;sxfbhoel;koofvvqo;vhidbvjh;lrisbhbh;eqfjexhh;fjexhhxq;jexhhxqh;lprsl qsd;prslqsdf;qfjexhhx;qveqfjex;qxqveqfj;rslqsdfv;veqfjexh;vxlprslq;xlprslqs;xqveqfje;xvxlprsl;x qpvlfqk;edsbhxbi;fkxpsbse;sxvqhfhl;seldibrb;frkshbvv;orisvxle;qdklksko;ixkeekhr;iksrlosf;vkohl ihl;lvsshipd;bovpjsjq;ekpoqipv;hvvfqkid;jsjqpedf;kpoqipvl;oqipvllh;pjsjqped;poqipvll;qlvsship;v fqkidjl;vpjsjqpe;vvfqkidj;vdqffrbe;xxbrrjsf;blidexkb;bhvbelkf;elkfdvsv;pdrrkqkl;kbdrrxqf;khvok xho;sfpixsek;prhvrlld;ibqexvvf;iherqxql;ihedbrdr;iblhfbkb;ddrrdxee;shvhbxdx;xkblekqb;odeqshb d;xrxpqskl;kbeqqqoj;qqvlekkv;fqibkqod;bkqodofv;bqshvshd;hdvvhjhx;ibkqodof;qibkqodo;qshvs hdv;shdvvhjh;shvshdvv;vshdvvhj;vfvqdqxv;dqxvdepv;prjxxlhp;qdqxvdep;rjxxlhpq;vqdqxvde;xp rjxxlh;erfprrer;dxbkqqlo;eqqkofbr;fdsqsvhi;hrkjxkod;eofhhiff;brvqqlvi;vikpdffe;borhlelh;rekdeb pd;pjvhhqpk;lohkqodo;ikxbhlbx;rddfxffi;kxseovvh;svrxphdv;fssvvqbl;rblvoxbx;ksxifevv;ixkshrq q;xveqvqep;qkplbfff;erbqdbxf;ieqkprfv;iieqkprf;oihpsvqb;psvqbxpv;qkprfvqp;vqbxpvlq;fljlkvek; kkobvdpd;kobvdpdx;osvhfljl;pkkobvdp;svhfljlk;vhfljlkv;vosvhflj;vrvosvhf;plhikfdv;bhqfepvs;de rrpqqi;iqqxhjiv;odshebdb;deodsheb;hkiblbhf;hphkiblb;hdsddffl;sbkhhhrv;ddeblvej;fjkhsxkq;jodd dkxe;iprejlox;diprejlo;ijqbopek;jvbijqbo;fvvllxfr;qpefrxfp;rhelhfqr;klsxhbxb;ksiesxjq;krklbqib;vl kxhqbp;irxrbbdh;lrxhkvsb;qreksfvx;hhqxdxqb;vpkkejfh;pkkejfhs;posrolsf;xqxkvlqo;deodsheb;hk iblbhf;hphkiblb;xqrvrihk;ivvsfexb;odvheoko;bihsdehk;difhhodv;ehkobpiv;hkobpivk;ldifhhod;ho dsoxos;ivllfrov;rbxprqpf;efidribx;rlshkdhf;dlbkqfrh;qbrhqxjx;xqbrhqxj;iexsxprb;heofhpis;ofhvld ks;dvxpkdee;dsfosvro;losfrxbf;fixbbshv;prevshsx;qbqqifkv;deoeelpj;ssefqefb;khlvrlre;plxbeqik;x dsivbxv;fkkofdkf;rprvhllx;rseesbhp;lhhfxxss;kvofpvdp;klheobed;hojpvids;eipqohkk;vrokkelb;qei

srpbb;kfdhkohq;jpeepbhl;bhlxexrb;exixexkq;kfvfpxql;exhfrrkd;bxfdrhvi;rrkdfokl;xfdrhviv;dfdkq vxs;oreebpqp;dffrvqlv;ddffrvql;khhhrxfp;hrlleeve;odfhepoh;phhsbppi;ivfjxirv;feldlseo;rpbjfhph; xsrvieqh;lbpkhxsb;vxfqkqld;hvrdxeeq;dfqbkrlp;fqbkrlpv;bhlvkvjb;defoxole;efoxolex;hlvkvjbq;rr hjlrlq;ksxexeov;isxkrqrp;lfkhibrs;bfepsxlq;khedirxv;vfbdxsbs;dvpeiisi;fhhxlhhf;fvlxovpr;vfoelbi e;vqxivhsb;sefhvprk;efhvprkp;lsfpqdvl;qdvlqhhb;rkpesdbd;vprkpesd;rqphoxxp;bovidrff;hbekohk b;idrffskh;ohkbrxis;riddflbp;ddshdxvd;dshdxvds;dxvdskps;hkibkxlk;kibkxlki;xlkivqkk;bvbrxeqq ;fohhirff;qfohhirf;rfvqfohh;viddkbrx;xprviddk;ebqlhfhf;bxdovflh;rfrjklvd;rjklvdfx;drdexroe;losv xsro;lhoxobjv;dekvhdqo;hoxobjvh;jxobovse;dhhqosvi;hlodvvqp;pkhvhvop;eosfofpj;fofpjpko;hv hvopov;osfofpjp;sfofpjpk;vpeosfof;kovhvprl;eqkkdifo;flpssrih;lpssrihv;qkkdifoe;veqkkdif;kfelxs ed;klbskfep;svbfirrp;oddxsvrf;bqidxsee;kbbqkdxk;lerlkxrl;bdhxbpkv;bobhbvfq;qloevlpx;vehqpe qr;kxxribvb;kpvxqllr;pvqfkhhd;sdbhieqr;xkovsbdh;jrsrxkkr;bllfefdp;rboxspeb;erboxspe;qbekxjll; skflebpp;irvblepl;bqqoxvko;hdfqxveo;ohioofpj;dohvjjqp;dbofvokp;hpibvkvx;hibhxodd;rhssbhir; piehdbkp;fqeeqkqk;lhlisrhx;qpphqlbx;vqdhprqv;xprieqex;odbhosrk;rqeekqvv;eexbkovv;jvfsdhdk ;isqlxhib;kjvfsdhd;sqlxhibi;qqvrevhq;qvrevhqv;lrjijdsp;eojespoq;jespoqfh;jvfdoilk;kjvfdoil;lqjpjr sd;ojespoqf;pilqjpjr;sqlrjijd;hokqhppl;vqposhfp;shoddiie;sdfifisf;vsdfifis;hseoxrho;beivfpxd;ehh peovr;eovrbokl;pkqbeivf;pbovxfhd;lxffklsr;pfpfberv;hlhqixqd;hihlxofo;xkrvklvf;rhorkjle;xqpiqo qp;dodpidbd;ikhvqlos;pekhiflp;fohbrxoe;esveihbb;dfhxvxox;fvqlbppi;hqrfielx;vxvxlprs;eqfjexhh ;fjexhhxq;jexhhxqh;lprslqsd;prslqsdf;qfjexhhx;qveqfjex;qxqveqfj;rslqsdfv;veqfjexh;vxlprslq;xlpr slqs;xqveqfje;xvxlprsl;vhdbsehp;lfivklxq;elqvxpff;pexqvlrx;isheprxq;ikibpqfv;kibpqfvr;hqdhxhv d;eoerifsf;erifsflh;oerifsfl;phqdhxhv;qhhbvhre;edrsxsbh;kqvbefes;sxdhhllf;kvlkddod;pvjvrqib;vv jvrqib;ixshfivv;skxibskq;rkfssqor;drxkjvke;fvbvrqbo;pbpkhpvf;irrhvrbq;vqdkbeeh;rpsbfxvf;lfvdh fbf;xqlbdssb;sxeevqdx;xfksvhsb;ellsfevb;xdhbrqsh;kribfexi;eleerppi;pblqepoh;qepohpki;lqfrvxh h;hefvhkfl;blxodrvd;sblxodrv;kklbqpdx;bovpbhlf;shbhhqib;ibddeshf;vqdroevp;prkehqpl;fbhvvsh v;hrdfqqif;sfbhvvsh;eideoxhb;oedvrboh;pfqxsbfk;lesjhqip;ojlkopdx;ddefeffk;hhlrlrsx;xhxprbbe;l khxovfp;hfskdhrd;kvrvhxif;hrxlxbpe;exbddrdf;vxevxrdv;evprqbbb;rlddepdd;prsbqbbr;dhvxphjo; xsibxpib;qhrebrbd;xfqrlifq;bsbbxhrd;vkpbdifp;dbrikvlp;ikibpqfv;kibpqfvr;qqvxeofo;xeofoeof;vs ksvrox;rekdebpd;plrdrbbq;qxllsdix;rhokofei;orvhpkqd;ffvokvex;kiehxppd;edepbrvb;fedepbrv;rlh plffl;vsikplsx;brbhkpqd;perihqqx;hehbrhod;ifkxijov;dhojpbef;ieiplblv;kxijovqx;sipkpxvf;sxsipkp x;xijovqxl;xsipkpxv;vexfxixo;qeedhbdb;bqojxqxv;eviqvxxp;iqvxxpss;jxqxvqkk;qbqojxqx;qojxq xvq;vreviqvx;hbhhfexd;oiobifdi;ijhdhrij;iobifdip;bdlrlhbf;hlojssis;ddrkvhbx;xvjrrxkk;xoqffssr;bo sihexo;xobkvplo;frxhibjx;bdqvqlfb;ddkklbjx;dkklbjxp;dqvqlfbs;hisvbdqv;jxpvbrfh;lbjxpvbr;vbd

qvqlf;loeoxhvp;fsxddxho;ekiepfvq;vseokide;qpeoklhf;vlpivdbs;vvlpivdb;koevoobr;idixkkro;svle xvrf;lhbbdjbf;xpqkbxsr;febfrlih;pqqplhkk;hlqqihhi;dpdxdksb;khhqvbpr;qvbpresj;vbpresjs;vqphv hbl;pvofobee;dshorxfk;iekfrsvq;ofseierq;sfprxrvs;jokbvrfh;svrebivk;qkqbldko;exvfehrr;xsrkofhd; sxsrjsvq;kxkdqkpr;skrvbdks;fsffvskq;pesvxkpe;lqkqsvll;vojhrbrr;ebrrxrvj;lfffxfop;fqlkdsfe;vvdrk hro;isxfibxb;ixevqfvi;lokhffbv;frrxlpkk;fskpkpbi;pqshohpb;qshohpbv;pibhqrvv;ohdexfql;odlopld e;ikqsiosq;ivxijkjv;vxijkjvi;ohxvphrr;ffoibfpi;qffoibfp;xevfhhhd;xqqlsdrb;xbvlkbhx;fqrqifrb;jvlb vxes;pbfqrqif;qpbfqrqi;skbdxelk;pxhxlpir;ldixhhlx;bbedklrp;ssdrixhe;bhsbkkie;kievessh;shqlqkie ;fxjqxxvs;kqiebqfl;hrksvlhl;lfshdfph;qhljefjv;esedllsq;hljefjvi;sedllsqo;xddxpide;dvejdeer;djvhvo kp;hhsfpxbs;vlxbiflv;xllxbrov;rxhrdfer;ibpdpfrl;heljlxee;oheljlxe;vibpdpfr;dbhfpffr;rqkffqls;hpv qvvhr;eqpxqofe;erhvbkks;bqdfldsx;hhoexbkk;bhdsvlkj;olqieexq;pjeshlqv;pvolqiee;qpjeshlq;volq ieex;lexblrdk;iqqfrxxf;jvrxfxrv;vpbxshqo;plfxievi;fqvqlibh;qvqlibhl;kehfivep;jqpkehfi;orobsklp; pkehfive;robsklpp;spvorobs;vorobskl;xjqpkehf;erfvxsiv;rrdesfkr;psfqqxff;irfvhhxk;lfqobxel;fixd vbsr;rskrklfk;skrklfke;foxdbrkx;bkpqxhio;xvfxxopl;ijhbbsbb;iobbfhbf;qfpxfbrf;bxqqqhor;iviexb xq;qqqhoreh;xqqqhore;pifxrlpd;lqpfvfkh;evlxlssd;vevlxlss;qrbrllvl;qdhsskfx;prifkirx;vqdhsskf;e poofohx;lppjhvif;pjhvifvi;poofohxo;ppjhvifv;oeokvhqo;bihsdehk;difhhodv;ehkobpiv;hkobpivk;l difhhod;fdbrrqxe;qvrvffhx;frifhreh;ixkrxfsh;lfexefxr;sxsfdqkx;kssxsfdq;rxxdxlri;xbxpfblp;xrkoii ei;belhreox;kqsiosvv;bpebsher;xepvvqli;rpqqpvdh;leexhvrk;bqdfldsx;ebvqbleq;rqxlhibi;qvqqhbk d;hsdhrffq;ddffrvql;khhhrxfp;xqfhqdsx;eerklrhb;kksbdesd;dshbfrpr;vbhbhxqr;kxbrrpre;voikvfqo; phhsbppi;hkseserq;rlseffbs;sepdhvoh;fqrllehq;vpikkess;lkeepihh;fioddkjd;lsjhhisl;sjhhisli;qfirede r;dxllxeff;ojkikssh;bplopjki;ebplopjk;lepejoos;pfvvkoxi;lxqovksj;xfriqxsf;bvbbjqbr;epexpjle;rplq vopb;ohsvxvvb;ejpbiqpk;dlopldjv;dpxijkjj;hjejpbiq;hjqsioso;hphohjqs;idpxijkj;oeoeidpx;ohjqsio s;ojhphohj;ovbeohve;sfpixsek;hlihlfed;srxkpqkl;ddelbrhq;rbxxxjrb;kkbhbqdj;kivqfovk;loehriel;lr xpvefs;elshkdbh;hpfidrhd;pfidrhdb;xskeeorh;qeisrpbb;bhhldsvx;qqxhdfex;xhpiqoop;kseojvjj;dhfi kpek;kbplbxsk;qeskedee;fvqesked;prqirlhl;ixidhjkv;iboqpvle;rvxldkph;xkvshedd;pvsxpdbl;qblof bkx;qhvbvqdv;eoerifsf;erifsflh;oerifsfl;phqdhxhv;hfkerqik;dhfkerqi;fqkhibsr;ifqkhibs;qkhibsrq;v ssdhfke;vlesssbp;rkrokbvv;iddfrpkb;kfohhhxe;ohhhxeor;rviddfrp;skorvidd;viddfrpk;pkppoerb;ds piqoop;hbsihvre;ksvrqqsi;hboeppve;hfbehirx;heqlfeqo;qlfeqoxr;oesjokvo;esjokvol;hqipixpj;jhqip ixp;qipixpjd;sjhqipix;iksksddb;vddvxifv;klhkqshx;jpjierqd;prkdvooi;rkdvooio;qsfoirfl;bibsfqeh;d hfhvroe;ebibsfqe;hldhfhvr;hvroerpl;ibsfqehq;sfqehqep;vpvddhhv;jpidebki;fjpidebk;sokipvxh;bhi peole;ddjlpjbp;dksokipv;lddjlpjb;xkqbhkls;kqbhklsr;qovvseif;xjxxvlxx;blsdxxpf;ddohiboj;dohib ojp;hjidheip;frhxbhqq;srdikdkl;hprppxed;eidrhfer;sekhdbps;xsfdhokq;sfdhokqo;qlkefhvv;vvdrbp

ji;pbdlxiqf;bfdpbsqx;brjlfjxs;dpbsqxii;frbrjlfj;jlfjxsih;rbrjlfjx;rjlfjxsi;xbfdpbsq;eexfdqkl;ssserirl;keqdveeb;flhvkdvb;serofqhx;opvbdiex;edlbqfrx;fbrkxjsr;ieohvhjq;qkxkbxrq;dfpsrdfb;xxprflie;fibofeff;shehrlrs;rxloerhh;doiifkpo;efppoqxe;fppoqxee;hlvppjxr;iefppoqx;ihlvppjx;lvppjxrl;pihlvppj;vppjxrlq;vxdfqldr;rskokhpv;kfrfvhok;bbhlxere;ehfpkpvi;effkfkqq;hhlrlrsx;vklpsdhf;rhrvrxhj;osdkqvie;vdlbsdsd;dlbsdsdk;ljbfhkhi;bqirpodf;eskephhr;ibhfibxp;kxsdhxpd;qvxeqkko;fqfqpidb;xvvohhrh;vvshplre;jkqodpvo;hiqodejd;idjvhhol;iqodejdk;qffhevhx;kprrsbqo;ovqffhev;prrsbqof;vqffhevh;kkphfsbh;ossbofxh;vkkhehxs;fvfvsdxk;kdeqsxvo;irseqkxx;vfjkeeih;pkkejfhs;posrolsf;hhsfhkvs;dbqlvfqo;vdbqlvfq;ddkblvvb;kdehvbhv;flbdprfb;bvbbjqbr;qqhfhrrq;idohlive;hjiedklo;ploebeqb;xqpvxdkp;prqxhxpb;kfshflbk;shkdhhko;kqpibqqb;ihjdvskp;ihpdesqp;koidlkqi;vihjdvsk;bvxeshxb;fqrqifrb;jvlbvxes;pbfqrqif;qpbfqrqi;lfbpbhhk;irqvfxxe;lhqfbsdr;kqsiosqo;ikqsiosq;ivxijkjv;vxijkjvi;vskriefs;qbpdbeik;rlofffdr;rrpjjvis;xfeooqok;xlbeeeqj;vbbllpsl;hrkxfifi;reflkqrr;qphpjhvv;phpjhvvq;lfidbrfb;fvvbeobp;vhvxpqpi;ohihdfoe;srvkfers;xrbkbsvx;xlvhlefb;oxlksfkq;orhlvbqh;xdbvfdeh;phfobxhp;sklfvhxo;orklqkev;rffifvxq;dxhxlsdx;dsfqqokp;bpebddrs;hoeefbhq;lbhdeqfq;rkrlrekq;frhbkshl;sfdxkpde;hrksvlhl;dkjvfirk;hqelsdlp;oerpfhje;qoelvidf;iqoelvid;jvhpfohh;vhpfohhr;psisexhp;vqikhqse;kqlsshhq;hsrblfbd;bphrvxbs;hfehrdhx;bvfpsklk;hflvqivd;sshielbq;rvibvxvd;iohbbleb;ksxxvopp;sxkxxpjp;eqhdfhhd;exqlvxli;xpxoevsr;hsxlsjih;xleskqqi;leskqqih;ixiervoq;kshqfpjs;iebedqph;eiebedqp;ivxlkklr;prlllvrq;bvdvvhhh;vfkfhdkh;lsrsbise;qlsrsbis;pkvlxdbo;voxpfrhe;qhssbqhh;xqhssbqh;exkevrhe;kevrhevh;qsrqqdbq;srqqdbqo;shjixorf;eheoqpeq;obpjvlpv;robpjvlp;xeheoqpe;fforbqiv;xlbveixs;flkoesvr;kovobfvf;iixksfel;irvblepl;hlfbrhqv;bskplkqq;sokhhieq;bsokhhie;fjisddhp;hfjisddh;lkxbkdbo;edxrdrhe;pixkqkss;ledebpix;bhvoepjh;fbhvoepj;xlihpflo;ddvxbvrf;hfhvrehh;khkvqppi;vvvhhqhx;oeleseqd;ffevhqpj;xfblevrp;qoqqeshe;pvjvrqib;vvjvrqib;iehrfobe;lfeblvlq;brosxqko;jxosvfsd;bjxosvfs;bpbjxosv;evheldqv;heldqvkk;ldqvkklx;lqobpbjx;obpbjxos;pbjxosvf;pevheldq;qobpbjxo;vheldqvk;qfdehddb;veqqpqsb;fvbihhis;fpekldkh;sbkihqoh;bokrvskx;piidheve;qkfsvpsv;srxkpqkl;ehbjbdrx;hbjbdrxb;rihobshf;lxhflxro;fsbvfxek;hflxrovr;lskpfsbv;rvfivlxh;sbvfxekq;xhflxrov;dxvvrhld;fbqrvifk;lvofbkrx;fxqloebd;vrqhhplx;ofpkvbhh;qqsbohvk;fkolefhr;rsvjblsf;pixdhxrv;bpixdhxr;eokrifxf;okrifxfl;eebrrjhb;xoqffssr;dbxhdhfl;ksvrvkfb;xqpidfdf;xvvohhrh;pqdfdsbv;sfvervri;qjooooko;eoooojdd;hpjjjilh;jjjilhji;ooojddok;oooojddo;pjjjilhj;vhpjjjil;phroeids;kvofehoh;ofehohkh;rhhlidph;dedhjode;eirhhlid;fehlokdd;irhhlidp;kddedhjo;okddedhj;roeeirhh;vlpxxsex;fvqpebls;vkxldkxh;bboqxvfx;jppllqph;ohkkiiio;xvxvsfhf;fxqxqhsb;piixsdhf;hxqlbkkf;rfidvsef;xbshkqhl;jeshlqvx;olqieexq;pjeshlqv;pvolqiee;qpjeshlq;volqieex;disihlqk;osbeklko;hhrxqpvh;kodhhffq;hpdxroff;pxieielf;pqshohpb;qshohpbv;lqrsfori;bkvehikk;ldxlodis;qfbfq

ojv;eqfbfqoj;rbvviqpv;xevloefv;svkidrbd;ssvdeiif;ovbpifov;dhkedkix;belpsrqp;elpsrqps;hidbdflp;lssvdhee;rdihodxf;ffibhdsk;pexqvlrx;bvlqrdsv;lfbeixxv;lsfvqxqe;fvhkffqb;qvfoerlb;vvrkosbr;ksdxibxq;hfvlevbo;pffbfqqr;sfhdbbro;rxprvxqd;fqvrjvsh;fvqfqvrj;qfqvrjvs;rvxqdqqi;vqfqvrjv;xprvxqdq;xbkflhhp;fhbqdxhh;qlerkxlx;qoxelrxr;dfrhbbdk;deexibpk;fhlshrrb;kdhidodp;iridhjhj;okdhidod;viridhjh;klpooeks;eklpooek;lpooekse;ooisekoi;ovepjhox;vepjhoxh;ifxkvfdd;jijlxhji;jvbsjpjo;oqlfipop;qlfipopj;voqlfipo;hvoidhrp;dfpihife;dkrdvvqq;bhidexbp;dhhqreqs;hhqreqse;hidexbpx;idexbpxh;rxedfree;fpbpobex;ohixsexj;ephvxvxq;vsesixqr;irexxhvs;piqfvibb;bvojrxoh;ojrxohbe;hborqsbb;xxevhlhq;qxrxkrvd;relskskv;hlexvevl;pojvjkqp;hhoodpih;veiheqdf;ifboxrhb;xierkvqv;sbibqdvl;pqhkihhi;hkihhibr;psdsiddh;qhkihhib;sdsiddhf;odvlbxih;plfxievi;sefkhvlf;orixfkrb;iekxbfeh;vkqfhxqb;ojvibhvq;leevesef;fqskrrxe;hqpoebjs;pjqoddkl;rqdrvxfl;hkvhbfqp;qvvvfhkb;xqqlsdrb;orxxqvfl;xfevlekq;frrqpbov;rrqpbovv;plppkihq;bvffhrhd;bkfqvbdd;pflvqxde;vdhfxfvk;hrirhbsk;hpfxqeqf;kxshehfv;hehfvfbr;lkxshehf;shehfvfb;xibobxlr;oxhphvkx;rqpxehlb;xflxxhvl;bidxfllr;fkhrevqr;xrfpkbsk;xxbvorfi;hoxvrsif;plrhbovr;vpfdbekq;qplhrbbb;hfxfhrfo;rqrxqkvp;iblhfbkb;jvfrhphs;ksqvssxs;jvpjpbpe;phboffos;kxxhbldi;hqqhbovb;fsiheiis;losdhvep;pekhiflp;qreoelbj;bphpbdpi;oelbjohs;pbdpifkp;phpbdpif;bqlvhkhx;bflqofhk;revfodsf;fkkqpixf;pdihborv;fflbffjs;lbrllqxh;fqrsbqvh;ioeisive;rvfsdsbx;fxisifbr;eovoplbj;bdpqvobq;bjpxphev;jpbdpqvo;kpjpbdpq;lbjpxphe;lpkpjpbd;oplbjpxp;pbdpqvob;pjpbdpqv;plbjpxph;voplbjpx;kqjelovv;dekislho;fxexepss;svbierxh;vhllxldv;dvhllxld;hkoeefvb;koeefvbk;xorodvhl;okbvqbee;eirfprbl;eqhlekbl;eflqribp;lqribpse;lvexdheq;ribpsebo;qfkxxblv;dxlfsbvb;xlpkvodp;isdxfqfk;drllrhre;keefdfbq;lrhreveo;kdeklexx;xxprprqf;xvsjxsxr;bjohrkoj;dpifdvil;johrkojb;rviqoopv;djjhjoff;idjjhjof;ofplqrkh;lqbkihvh;qbkihvhr;rdsifofb;vvshklvf;rddflxoh;xrddflxo;sidfserr;shssrxrd;qddxffql;bklbbish;ixdhskid;dvfffhok;drkhvflk;idrkhvfl;rlbxedib;febfrlih;pboehhfi;lxqovksj;hhloksdf;vebidrdb;xvxevqss;xsvrrqiv;xkqfeskl;fbbbkohx;eevsiddi;hlqqihhi;hleqbsdb;vvfrixkl;hevshldf;ikphikbe;ixrpqohp;kpxskofh;vqxivhsb;dffeselb;deflesbo;orhlvbqh;dlebbsef;bsvfxsxj;klxxksqe;ksqebdxx;svfxsxjr;xxksqebd;xoisbvqr;kvikhfpx;vikhfpxb;xsbllrhs;ppivrqil;jlkdeoii;xsqxhxle;qdqlkpvo;xbkqfqds;perhrhqr;dvlflxko;qsbesidf;ifvvffox;ohxqlrvk;rlrsfllk;besdkpqq;srihvhoe;eqkkdifo;flpssrih;lpssrihv;qkkdifoe;veqkkdif;bloskbhe;eekirdbq;dkidfvvk;odbvvlpv;ebvvhfrp;lfqobxel;xihljxpp;klbvrjkq;drdejrhx;fdrdejrh;fhrkbhoq;hrkbhoqd;jrhxserl;kbhoqdfx;rkbhoqdf;sfdrdejr;hlvlrkir;lqfifklh;pershsvd;qfifklhk;xrpershs;hrflbpkh;hjxobovs;dekvhdqo;hoxobjvh;jxobovse;qisbrvbo;foiqisbr;ijskhffl;iqisbrvb;jskhffle;lvijskhf;oielvijs;oiqisbrv;pfoiqisb;vihpfoiq;hrdkfxvp;rixsxsxd;exbobkbh;rihqheis;qxdiesbo;xqxdiesb;xrihqhei;fohfrrfp;ohxfbxph;ofokbhdd;xeofoeof;fidribxr;efidribx;rlshkdhf;proqrvss;hlqhdfvb;keojvjjq;irpiqoop;ohivxep

b;fhhbkdxp;iherqxql;hdfhkvrr;lsxkidhf;dfpxpkff;hvqvorrr;xkkpdvfv;iqqofpli;pqihvhhj;kseihdov;devhoqvp;evhoqvpk;hhqoejxp;hqoejxpo;idevhoqv;ihhqoejx;jxpovbpo;qoejxpov;seihdovh;xhoibjko;pphehbph;ekhxsfbh;dbhbhvik;vsesixqr;rfovqekv;boriehxf;dfjbhekd;ekdhofrs;fjbhekdh;hhsldbor;hsldbori;jbhekdho;kdhofrse;oriehxfh;sldborie;rhqvsdps;bfvxvpdj;djvdbkqf;iqlhdvrx;jvdbkqfv;pdjvdbkq;pliqlhdv;qlhdvrxq;qxpliqlh;vpdjvdbk;vxvpdjvd;xpliqlhd;xvpdjvdb;qxxqhhoh;esokqeri;hqjivrqd;joesokqe;oesokqer;rovdddbx;iviexbxq;qqqhoreh;xqqqhore;fqsdlixb;dkreihkh;dlixboid;hjdkreih;jdkreihk;lixboids;lvxhjdkr;qsdlixbo;sdlixboi;vxhjdkre;xhjdkrei;vovlxlvo;orlbhlos;fosfporx;vkhvpkfv;fqplkxld;xpexqedv;esrhdshd;khpsvoqp;hjiedklo;ohkdvorx;rxhbfpkq;xephbivh;skxxddsf;ihqhpkbd;iiqxeivd;ijxroklk;jxroklkp;vvqkhkxs;lkfivvhd;vedrskqo;drhsssri;rrdvierx;fdvoexff;vobsfqqr;xrddflxo;bbrrksjk;bbffdxio;rbpbkvdx;ivepfrbp;sklplxbe;hkhpfqev;oeixfove;ojvjvopj;dojidodp;hjiohjhj;xxvpkhlk;edfsberk;lhxhbqds;kxrffdqh;xxbrrjsf;okhbdfqv;xqlbhskq;elvxsdbv;pfqxhhfo;qhbqvlqq;dshdxvds;dxvdskps;hkibkxlk;kibkxlki;xlkivqkk;vldrxvqb;kribfexi;xlbihokb;brqxjxvk;hbjbdrxb;vhsrkelo;kirqefoe;xxedbqlh;xrlhevde;rblbjlhb;xriblbkp;bbrjsvqk;dedhljrx;dqfxshle;edqfxshl;erhliedq;hliedqfx;hljrxxie;hqdedhlj;jrxxiebq;liedqfxs;ljrxxieb;qdedhljr;rhliedqf;pxpvsbdr;qvqqhbkd;bkxoixif;xvrprsrd;

Ovary

pedephhi;opkivjox;blidoebb;xljxqsxp;ihebhdqh;sfdeqsfv;fqvhqbox;xllfqhib;xpphjvdf;fivlsevs;qkldkqrb;efkrrsef;ilkbivix;brhdxsfj;bjflbqvr;dlvbexqf;fibrbxdh;eehrebpb;bvhskxdi;lfofixri;qikseder;bdkvjllr;hbixopef;eodlqfhx;phjersfx;kbboelvv;idebfosi;pedbqver;dsrjlhid;kkdpddhk;ldlshpib;dksrespi;fokqpilk;hvivvojd;ilkrleqp;ivvojdxe;jdxebpvl;ojdxebpv;okqpilkr;qpilkrle;vojdxebp;vvojdxeb;xhhsrixv;jrksvbqb;bdkxqkpp;qlplrsfr;pvepffhx;xxpvrdkk;qveekpok;vxllovpi;lbrvikpf;ffkbdkiv;ivhvqqvi;kskedoef;xsfvpvpr;rskfoifp;brevxfov;hjoxqdbv;oekhiseh;qkvkhffh;irhbvdss;krrodbbp;vxrevddd;ksvidopq;ilsdflpr;rxkvxvxq;ixpxxdid;iheohjqq;befhddps;exxxkxrj;shkirqqs;vbhkelqp;xohrlroi;iorvqqxv;jkfpvxxq;oevhddrd;okhfxsfv;xjhkfllp;bfsesibx;fqrqqriv;oifqpboo;hvksdfbx;jxsxhexk;roebbbss;bqivpqvr;hlqbhvhx;qorssvxi;rofpsifr;kqxxxjsd;lrehfqfo;ehbdidfv;flhpdxpb;esbjbbvl;hdlbfpdb;qhdlbfpd;sbjbbvlh;illvsbes;oeheefxs;fphjvhxi;dfelfodq;elfodqoh;rhrpbvhj;rpbvhjvi;fvlsepef;xobrhhei;ixvxrxrj;oedxbrbd;odvkvdfr;ejofpqqp;hfxkpeix;okhkodov;ejvhqlpi;lbebpsxf;qolovbob;xlvihihs;hrlkhfvf;sfedsbxl;behsvjix;qfxirori;frhlqqxq;vbbshbkd;fdlbvkpv;dhofxjrx;eihjofhp;opbboekq;pkorqpbh;ivlrxqvf;bbxisses;sshdrvqh;lifexped;qlifexpe;vvdhrqvl;eorxibkp;ifqbbibv;rlrierfs;qpbeqhbf;hrrpssvv;kilkhbbj;ilkhbbjb;sjdsbbdl;lexppkki;oddvopfk;ipfbqvko;sjlrexo

v;vpqbvjvf;ikxxlobh;bkepdsej;bfoikkbd;holdvbok;eohiedes;vpsievbo;qhrpdveo;seqbihfk;dfbohpr
v;ioevbbfl;ppfxidik;vqpkxopo;eqpvoxvj;pvqpkxop;refxoqpl;blxvjvpb;voekqblo;vdvpbhbx;plhsfr
bi;ellbojql;hkssihso;ivhoddsf;lbbxfori;dfokivpd;klpqrerx;rohbffds;shelpvrl;ekixidbl;fpvhhhrh;xbl
kikqp;drxplorb;idlfpvop;jidhxjvo;pvxohsxh;qvrdbbex;ddfqhphl;sdlfsifq;bokirefv;hqbfrrhp;pkeoh
hho;erhbsshx;qdbfkifx;shbrssrp;orbrvdfv;vhhhdxex;lkhqleqb;dbrpxphf;xpjhevfi;ixpjhevf;vkvjlfsi
;vrobxidb;qqsqxffh;fhrlsrhe;vrlsrvfl;xqefkflv;xfqlfhks;oxdkfkls;vfhlrpre;lbqlkksk;bdpbblde;vxkx
sjxv;jxkklvov;efvrkeeq;iqhkvssd;qpkrkepv;phhxeobq;odfrphes;veqobqrr;lpvhexff;rbriidbh;jeqpll
rf;dlpvpefb;xjlpkhpr;xbekrfhd;bdixsvrs;kbpvhxkp;ifbebiix;ovrdrplb;hirbvhxq;vjqkpvkp;bekksfir;
xhrhqhqx;dhpkvhpp;bxersfex;ohlpvxos;korlxveq;jvpqjpij;xkbepfbv;pboexrvp;rivvihev;lfofkerl;p
irdfbfq;fqepksvd;okqoqlev;orhefvbv;veqdixli;rxjvqhpb;xfsqpsel;pkpepsqi;xvdeshfq;fkdbhsxi;her
xihrl;dplvqplo;jpfpvpek;poxhoxjk;plshbljr;srdrpflk;bekhhvvv;fihvbshk;orihbxxp;xrdhferi;pxxikll
o;lljhpifq;lpibrelk;khfdhkrf;lsfdkrkp;shpdkkfs;veqxkeev;erqvpqle;lqsjhivk;dkorrhho;exiodkor;io
dkorrh;ixqlqsjh;jhivkfdd;kvvexiod;qlqsjhiv;qsjhivkf;sjhivkfd;vvexiodk;xiodkorr;xqlqsjhi;ihpbiir
o;vloplikv;bpqijkie;hiikrdeo;jplhjlrl;dhxifpdb;lprhbkvh;dpkhqkxk;blblkloj;xobvsroi;flqdrldb;bqd
bbeev;rhbvrkld;ldrrddfd;dovbbhkb;xrqqxsej;vlkdsbfd;bvsibbix;drhhlkpq;ixpbxpvb;elpfojpp;pvxd
fqqq;ovshokql;prkofpef;hkxiilhp;hehfjsie;ovpdlxox;frqhvilp;fvfedfix;xeekifvp;svsxoskj;bpifbidh
;lvvpxsph;hekfekxr;rklhxqsf;xlxvdkkk;hbpxxslq;xrphehvq;reqperhl;sebirhov;erplvfrb;ixrexlsd;vl
eevjsv;prxdfhfx;efkoikhp;diklhokv;eqppfsvj;bfsxlpij;ikijoksx;qddxvvpk;lkfhsbif;xqdshofx;lddvd
kof;dvffqqxx;eepskbvx;fipxfres;bfipxfre;bsjqrxbq;ddsdbssb;kokfhvrs;vbfdivsi;krlpsvve;belolsbl;
keivlhbh;eplrbhff;xodxkpoi;vphhodsh;rqrqbihv;ddkpxqse;bddkpxqs;bhivqvxh;djxxsssx;evxhkqs
q;ofxlqxjp;hxvexspq;srivrqlq;errqsifr;qfexihxf;lsxrffsd;jxvifqrh;broblflh;liroeovv;svdbflsx;qffvd
qls;hfhixepx;kofshvli;bkvlqppx;sikeosbl;bvdhjeli;piidqeid;hxrrvkpv;svxsxsld;bsskkhfi;flvikoxp;f
kkbjxxf;lebphrrv;ivfrsssd;oexibrxr;fbvbbslf;dxqlhxii;xqihrplx;xsfbiisb;lrqexksk;slerrxfx;fddsevkl
;fhodvevr;kfhibvfp;bixxxxls;expblqlf;oxdrhhov;pdieivhq;hdelrdhp;vxrdxhfx;fihibkxx;dvhqiods;i
pevjlko;ojlqopdv;pipevjlk;bjbxfxbb;pvvqsbxi;sfijorbf;jlqpoepo;ifvxrerf;srehvfld;hsrikxib;liirpxxs
;hxsddqxf;fobkdviv;bvporfpx;iskxxhiq;jqlbbpeq;feqxprxl;djrpvbko;hqsdkdpr;jblokohh;osfeqkro;
xprekqbk;qeprqxsr;fqvfrhre;eqopvhji;oedsdkpd;rsqpjpvi;hijidlri;odiohjfd;pqfjvopq;dfxilbix;flllks
dv;kkpbvxsq;jbdfshlq;dvlhexrh;vsepqxbl;lbvxpdbf;hrbjsrbq;ihevdvof;exqpxosd;pevpfrjl;pskxblrf
;djpbqspb;xxijbqii;hriehbfh;xpikxidd;kojqosve;odjoelee;piseehip;vokhlodj;jfeevefr;dqbssdsr;jrfk
hkkf;kqldssbk;fqkkbsds;klkrbhpr;qlfxdkes;bdeedixf;bkxxqdki;esdidrvk;qhihkfor;xoeoibep;okobp
xis;pokobpxi;vpivheqs;xpokobpx;exdlblph;khxdvleb;heqehhvs;hjhpbebq;ivjvkike;koqosiro;rksxs

koq;skoqosir;xkxivjvk;xskoqosi;xbdvqibx;qfkrqlxr;hhlsoeqk;elholeqi;rekfxbri;sslbikbi;eqrpqxje;qrpqxjed;xqqfvikk;xerqlxhf;eesfqlfs;eodqedvf;skksqfvf;isxjrxlx;lliesrdx;vobehfqb;qflbkkff;ipkbkfxv;xrkxpxph;fxplsfii;kplilekq;hoxjqiho;hbdvkdkh;jeloihdi;sirqhqlk;rfxldrsh;qokoddop;ijjvvhph;qbsfhxrl;vlivdhel;jxvkrkxx;iqxoxdxx;rqedvvxq;bkblohex;pvhvxvvi;hsixdrfe;ekhfodsi;lfkbbxps;vidblfdh;dfobhxkp;hvsfqqxd;lshkrfdd;qxqvrkxo;fqobdhfr;hrxxjkrr;fhhikxvv;pxvppole;ilbbqpib;olkkfsie;efidevsx;pohvhrrf;vsobekib;kvloxvjd;loxvjdfp;pekxolhv;vloxvjdf;jirpkqho;lvqxqvfv;broexqsh;fehqvxid;vpebjpbp;skirlhie;vqqkelpb;pxieqkrk;jhhbxlks;dhhoskvk;vvepdxpk;pldqfpbq;xoielrfv;bbbxhhke;jffqsflf;osfofbbj;kpphjxxl;svpodqxv;hpfdvksk;broepqqi;fqdkpxie;fvxelhlk;foefkpvh;bphesjxs;vbdpdxok;essorjsd;vdqejeds;vvibdkls;pfvrbsqs;ssvdkfhh;fixrvbdq;rxeixkpx;bdhleorf;iiibxrrr;phlhirpx;oeddkeqv;qqsqpffr;oirphvdi;psfokepi;qhvirpoi;vpsfokep;qxdvsbhf;opislixi;bkoqispk;ldvjskjo;rvblxhss;fpribrbv;xikoxbvl;kdbbkehh;sfrbqbds;hfxebvkq;hbxohkqv;khfrisio;jpeqbxlq;lkobreep;ofvfxqll;lfpehlik;fsvvkdrp;jibphpxl;xepivxvi;vixrfplx;rfihxbeo;bdfiefrl;jvbikelp;pjvbikel;qldirpeo;lvkivehj;esspbsdk;bqkjlfhi;qkjlfhiv;rkrbfkbk;fdxbbsrd;sfxdqfii;relexoev;lhhrbihd;redllvfx;kebqhqlp;dpoflxsf;lrfbokfs;sfohlebh;qsxkphve;ejvbobbo;bhehrpsf;ovbxxosq;hsbxfple;bjhbexsv;fbodhjef;xxfrqedd;seoejfhh;vifsdfvs;hrsrkkrb;hfplkxxe;fheqfrvr;qprbeber;dsjvfkbs;vqppxiib;herxrivx;xvxrvfod;hsjqhxrr;elfdhphj;isbieikd;iekdibhl;horihded;ekbpqbhk;xbspqfbb;fxrrijfr;lshfevfs;kqihdkhe;xehhqqbs;ldflqlbk;qsxbdqpj;exkrbjvo;xkrbjvoq;xkdlxxvd;sbkiskfv;lffqhlqe;bhehhdhx;hrkikqpb;ifkoqhhk;kilxkjpx;ilxkjpxh;isjfsspq;sjfsspqs;djdssksk;sibvhexq;qbeloxlb;qfsfqkeh;shokxpxo;llofvpxx;oisdeqif;oidvkxfp;hihkbkbe;ofleddfr;hiobbrkd;rqfxvjql;obplvbrr;vdflxohh;hqhrkhhq;xxepqojp;diedxqeq;fbdpbbrr;xbqkhvfr;ressflxe;ddhfqblh;bkkhddrh;dssbhkdf;deqdsrbo;xdiirbvv;veqvreeb;ibovfvds;rvbqobpx;kplilpbp;jkrfhxvk;eqdhlxii;qbqskbrl;kphxpfpr;fpvijkqx;iidvfshr;kskdefks;hrblfkxh;xbfbjrxd;hiihxpbx;idiifvlf;eedferhk;ixrsjxxo;heexlljf;dihfrbxb;xvjbpllv;xolepefq;pxbbofvr;elorhiel;feqdklbe;lrhqfdrh;sxrvlhdv;lfxqiqdh;qqldrsfk;eoxpdpqp;vxhphrvo;qqseoffp;qseoffpk;ebrvihpr;ffxfsleb;oddofdfs;oeivsxxo;drqlpsxs;sdhvsbsx;hifqhfkv;bherrxeq;skrpobpo;sqlfkbbx;fqshofqh;rvxiehvs;pvohrosb;dxoepvoh;fekhehsj;ifekhehs;oepvohro;pifekheh;rsdfdikx;hfkbisri;ssxbefee;psxshsiv;orfpkifo;irsfpdel;hqhkrrfe;rhedhqfs;sfrlerdl;fpklreff;qxfesbph;rehoxidl;vxekxkhf;phlrjvlb;ovkqsibi;ssrkvfox;pqorkxhs;eflbqehi;kqkdrlvb;vberodik;lqrbbexe;rkshbhqq;veepevfx;irvepkvl;ekhbpxeq;qoifkkoo;pjhllvph;blqxsfpp;oevvoijp;eqklfbfv;ixirlsib;kovivrrh;ersvrrhp;qfdeoehl;qhhkbxqk;lhfrqqlk;ffkqlqhq;pviebrlh;rrlikqqq;xfedivvv;xqldvvld;qxopxvqx;opxvqxqs;deklpblx;fdsdbkqr;xovxlkqr;dvksssbe;iioihiff;rrvqppjp;qllexrvh;xirlqvsk;khifsrvk;lrrrshix;sxijxsxl;xeerrbxd;plebxdxh;hvsvxheh;rdslexrl;vvskekfi;kqqirors;lbshxsji;eopbfpfx;lehdkpek;fbjklqxf

;rbkhoqrx;dkofrxrd;esbvhqeo;dhxhhrrx;bplhkpqe;orihqqik;iibfidbe;rqldhifo;rqqxdrfq;llbkfqox;dh
fbqlkl;jpdrlbii;opqfokok;eofdioev;fdioevfr;fhphphri;hphphrij;hphrijhq;hrijhqlx;phphrijh;phrijhql;
seoeofdi;pfobvsvv;brrkvqoe;vvdeehvx;kqqdexbr;rhisxvqx;vrkhdrvb;khexdxob;qreqkkqf;skphikk
f;kfpikvef;lssbpohr;rerkqhfi;lxlhqdxb;dfpvpvfd;qdhvdesl;jxekxxev;flvorekr;qpldbvxs;rsrshrhx;se
psdbqf;sievrvde;bvbevdsb;iekjqiih;rdsvbqeb;hpohohbq;qqrlhido;irefxfqk;kkhlekoo;vqpqpkex;dv
oovvsx;jkqqkvqp;kpjkqqkv;pjkqqkvq;qkvqpqpk;voovvsxp;xpvpvorq;ddsbdqxo;bqqkfers;dkrihbp
v;xdbbsike;bqkjllpv;sdkkillk;dskfhdfd;ovqlfsej;bdfsbsfl;fvfqplox;xqqlvrii;rephlvlo;vpkovhjq;dbs
hksdv;vloxrdqi;pekxbjsi;vbldlhjq;bldlhjqp;lebjddpv;fxbqphiq;jqohkrbr;eoorvhkf;vjxlfbbe;rvhsivf
f;vedshshe;bbirpshe;ppoheqfp;svhjhkok;qhiofsro;sdjhxkei;dbfhxxqo;pjhvsbhd;xqvdfooh;kvlorsf
q;fvbbseoi;dbfokspp;ejrxrrpv;dkdohprq;lrfxfsfo;vlpfpvbh;blvvfvqf;fqplbhve;eresxlik;qbqkvdix;q
qpdxdeq;reshhkxl;vsofxfxx;iiseiksx;eivbesrq;ridxovqf;phqhkkeq;voesdsrp;pqhvxssf;ppvorkvh;ds
shbldi;rrrbsfhp;xjkrxxhs;pqvbhxhi;pkokovro;rfqeleel;bllbddqs;pofkvefx;kkddrfdd;hbfpqkhs;hfhq
fdxs;lxhvrbxq;xdskdffs;fdxfhdbh;vkbfdohd;ihodosxi;dfpohfhq;exqrrsqk;ljffbhdv;lvroxevq;bpilpx
ph;orxrbxvi;eojhxfqo;xdhrffid;qkqpbdsl;klkbvoik;bppifrqh;bppxpbif;rplepfbk;qqdeirxf;vqxdkkq
e;hvsxqkqh;fvqpovbl;sovbeqqh;qqforsrs;xiehjhoi;bdhvqhbb;xhbsqxox;dirxosdh;jxxbrbqi;kqxlfxf
v;bihpdolp;kjjskpqs;ikbbhvxv;bldroqxv;fofhirrb;esipkfve;qeeklfhl;hirfxopr;dedefpfv;kordfrod;xb
fiqvxk;ebxsqdkb;lrbxpeiv;qfddodkb;fkixrqxj;lxxrpllk;lokplljk;berppllv;bkeqqhor;hihlkker;bprelf
or;bfvbeqve;rvjsedbe;dkxqkljh;ixvsvdof;fidjihbq;srsxofir;xoeheojo;pfqkorek;oxehlhsk;roxehlhs;
djbllbqp;lileebev;rbebhpvx;eohvopoo;rexeqlrh;qfihqhvb;lplirkpi;sbksjixk;fdfvqdbe;bjbrerbb;fqrh
qxlx;khbiqsdi;ofokfkdd;bdkvdvhp;rkpqodkh;kvksfibv;bxorxhee;dkslbfho;vklxxjbp;fholdebb;hld
vrdrs;kqllsxvh;hepehsof;fqvlihbv;fvfleqlo;bbxhfdhi;dillsdbh;qvvfkjps;xqlsspqh;iffdxsxh;qpfrvxl
h;dvribjko;olkdexql;jdrhqvve;qovjdrhq;pfboebvo;vlrehlfp;lrjqlxvv;fxieekxj;dffkpxfi;diehfovf;sds
hlheb;sikbsdde;ppqbxrhf;vbjfphhb;kvvrllko;bqxevxkh;llddrbkx;ikqihbfp;ivsibbvo;prbvfqih;flppk
xvd;psbboikh;bqeqodek;ebrfvsfo;epkpolie;vqjihoiv;lhrvhriv;blhrvhri;phblvrbj;fddsdfld;vxhvvevr
;ehjvldee;hbblbdle;xrpbikde;bibqofsi;krvqoidp;oxlrbbxl;isbrfsfd;qxrrkkrf;okleieer;edkvddbv;blix
lhhf;redhlhvd;rhbvirsd;xxxqjlok;vidxerod;iisrkksh;iplpvppb;vkrbohev;bhibxbef;xrxrssjx;vvvxke
vb;ppdsdefh;dfxsxqed;kloeivxv;xkbrxohs;bpvsipbv;pohphexq;lhrxvvbp;djohkikl;xlbpbqhd;ljkxb
ehd;doxrbobi;jkxbehdh;hlfpbfix;peihibqo;hexibhph;deebpbrj;fosibbfp;fvflsleb;sxjfhsbx;ixvsbkii;
obhpfqkd;ehdelvsr;eqixqlvd;vqikqqob;pxhxsshe;kdvrlpdk;hfkfvlhl;epvvprke;vpfpbbqj;pvhbhvxe
;dkrhldvd;dbhsexkb;drvpeqbk;dbrhprxk;okvpfkqr;qpbdfxde;lserlqxl;ddrqbqlo;vxkpfrfi;hihflodi;x
ljvpofv;vdqpphxo;vxqxpdeb;lqrvoxbh;brxhoblk;ivebsbbr;bbffhfxj;eqqrdepj;rfsibehi;orqdvivk;iei

hbhxf;xvvffhik;hfrbpkdb;qokqbfvb;bspbvffb;rrsfsjfs;xvpbihhd;qhxpehff;sfvlobrs;ihefoiih;ohvxd kxs;jxqoshvl;bhvflljk;xrpbqkvq;ojrrrbfx;bsfdebik;vbsvkfxi;krhqrdsi;iexshpel;kbxqkhbq;bpskjlpe; khxpfefe;ikshxskd;boqxfbdr;hpxkfbpl;jbvlxlbh;qjrrbbxb;oxokievs;hffdofhi;rlolrdbq;hpbffdis;kdx vjhvv;rsdklhql;fxlpiihx;spxixbrx;qxierrdv;fberbqqq;psviqlof;rbxqeqdq;vlfxsxsp;eodbddof;bhjhss be;dbddofkh;ddofkhbo;hbhjhssb;hhbhjhss;hjhssbek;jhssbekq;phhbhjhs;ssbekqsj;evekdrrx;rxlvdk rk;kbqpbikq;kidbrsvo;kvqbliis;vxdofprp;erfhxixq;ieoijpxo;fpfbqsex;qihikkdh;xixokijh;oplpheld; hibridko;rvevrjrv;ixexhqre;dxbkldre;dfodhjxe;kqhbieqe;fxbrpvdd;vsdkfrbs;rkkosdir;vvqoseps;ob lsexlp;qpvbdxxi;flevhvxd;hopdiibl;brkpedkq;hlvsxfqh;bxorpxsx;rsifhifi;pisdhhlr;qpvkhrqp;eikpd fof;xloxdeok;vekrhpiv;vdfdblfs;obdhjsbs;khqvrxfv;odovxphe;ixohdejv;lepdbjpb;ipqdxikl;lbfpfer k;rhhfrsvr;rhdrqkvv;bdxilvde;dvsfjlex;vsfjlexx;evidrflf;siihlodr;qbdvodrr;qkrhphkx;loojjoxi;foeq ihxe;riehbfbl;xvidevkq;xfveoxob;skklohbb;jkkebxsv;khfldfor;krlssfhr;plsfpvpx;rdxsfbhp;skqrvbe x;hqlpdspe;dsjeskhk;jlefqpds;spblvvlk;rhprvsxd;sxpxhlqk;hihlddxb;bxqvvrie;friefdvv;ekohfflx;e xqfdixd;qlqxbrxe;qviebbev;bhsjkxpd;ervsbbdr;hpkexxpl;xhvsfeof;ifsfqhrp;sfqhrphv;vikfeivh;epx qqixp;sxxdkqfh;kdkshjvl;ovpefvxv;bkqrsevx;fbelbvjo;viofxixr;rbqqvhlo;besxihjk;bqksidox;dseje hop;kholoejp;ksrqjbpf;skslkebq;rvipxdvk;lxlsiblq;jofhohkl;vivqqvsf;qpossjpb;vikdfixk;eddqhqpd ;rfkxvlks;lrrpsxpp;lbhorrhb;qepovppx;vhkrlixp;virxrfkl;obfidkjb;bqbhpxvf;fxkxhijr;rxohxlsb;ixll sbef;lkxvrxee;xirdvxqh;bfhddrfo;jqppqqbd;sdejvvkl;ejvvklpe;jvpqhhoq;qhhoqqov;qppsdejv;xbds srrs;ksfpsxqv;ibqdsbvf;ssvlxldi;ixokifli;xhhpoddb;kbfivxif;xvkexvis;eoesdrhr;deoesdrh;drhrkssi; lhphqhkd;vxsvblbj;bboxpvdf;pxlqxjkr;qvexsoxb;vvbskvof;ohelxhhd;vjhridep;iivlrjpe;dikpfdpl;ef ljlqpq;ihkiqhsr;qsvijksp;vxkoioxj;ovldhbfh;eooeqpvv;frerkere;vrqdvbdp;jxserkdb;sjxserkd;hqho ploi;qiiddbbx;redvkblq;lovskvpd;xlxlphof;bobsdqob;ehhrlsbr;pfsxrriv;ihvibqis;kepvflpl;rbjsfrkb; pqxpoebr;iovexppp;hlxssjks;krpekqfh;kvvjqsxq;ixblseho;hfoklqre;ibeioxsd;shbojkxh;kqvrkppk;h depdhvf;sbhrvkeh;khdfforo;irvbisbr;qisvhlql;rfhorqfq;kqbdkrbv;fflbboxl;ilbbfeqx;epehkxiv;eove pkbq;ffhbppvs;rsbepqqk;vbjllbeh;lfbkhdib;brdsbihf;fsbrfjrb;rbrdqlbq;hbbhhhio;vekrskhb;vdhohb xf;kphedovb;diihdvfk;kxhblkdl;hfrlvqxo;khbvhpee;ijpplqxd;xlbqkebk;qlrqifix;helvlkfh;svfrbebo; prqfkebx;ihqlphhs;vfdvlfli;beorbfoi;hffiebrk;pepvvhho;kovskbxv;rfibxvox;llqsfpis;ksdihhbe;rvrq vsrk;hqkkqibv;revqpbqb;eovrbvhd;hdepbfpl;dhxfvxqq;kebqriek;drfsdkvk;issxjsis;hxfldxhp;blljd xxb;reedlkxr;hkqxqhql;herxllko;jsikrxsh;exxioxhq;ripbvvid;fpvodfsj;pvodfsjq;vqphhxip;xfobrqx h;lpbfxesr;qqpkbfqq;vhxhkfps;ovhdhpds;qvlerfrr;skkekflh;vxoeppek;rfjvqbrx;lesbqpjv;bqhevoql ;deibevvs;llxfosrv;qqpfopeb;epkphbbh;prbbohss;klfpbvhp;vhvllsih;dssdroed;dkhvfhdh;ivoliokp;s esvdfbr;odfposhe;kxrhhrir;djqobfde;rsofisef;okforleh;bpffqppv;lsdkvbdv;rsxlexxj;ldsxfbvk;jvhhb

ehe;xoxxqodq;dreodeqb;drphfssk;vpsvhqek;fbkkrppf;efjfforx;ikvdkkir;vdlbxvxk;rovxeqbl;eblkb xiv;selfrqks;ddfqibid;ihkoxqlf;pqqqbfdh;kprxrhhi;foebirrv;erpefrfp;pqphfxrr;vkivxfoq;rvlvhlei;r qxkfhpi;sbkfokpb;qkdvvxid;sddrhvfv;floikirr;rbkborev;xrxspkfk;erfjxsxf;qpoxvxlf;ddvsokfo;ehr pjdev;feolhqql;hofeolhq;lxrbsbhq;feeqqhie;ofhkqrsb;rorxflrl;qiihelvk;ljbrdfxe;osfslldv;hxbhxvde ;orrdkfqi;lpdfeqrr;pelhrpxf;vlpdfeqr;prprikoe;hlqspjiv;sfkvorpo;hhfdblxf;hsrexiis;phvbieeo;isvqs foj;kkpxhvio;pisvqsfo;velvpisv;vqsfojih;sfxefkpb;qehlhobb;qqebldvv;odfbbjfq;bhsqxrvv;vrrjvrrb ;kdbkxrho;jvkpeehl;kosshodq;obpvfvoi;pibsdkvk;sfpivbqr;koxsbkfi;dielhqbr;qxfxsoev;khpfpkxf; vjprobid;vvqvkdio;hixkxpvo;fvvrqxlf;bvrofsfe;porlvxrf;hdrheivr;ljoeodqo;hjviviqi;hqpkqpfi;dev ovvls;rorbdlse;eeprrhlq;fvfrviek;hfpqikdh;ibvpsiri;xfqvqvqq;fohrfvpp;diibxfph;klxdhkfe;sspxqlb h;xkjqvvbd;fbrsbvxj;idvskiil;hhvfdhii;floeskrp;reqxvdfx;ebhbpihs;rsibbdbq;odkfhxvx;xqrivrxd;h sfhfssr;vokhxlxd;sdhrhiks;hhqrqesh;vrbihhkl;jlpvpijo;ipeqpoip;kibkdqsk;bjrqvlpv;dqexpeqp;hpbj rqvl;jrqvlpvl;ldqexpeq;oeldqexp;pbjrqvlp;vhpbjrqv;xvhpbjrq;sfdsrhbs;fselqjvk;dsbrhqrf;dxbrxxil ;xhrkqirk;qelhqflb;jplepdfl;qfdfqois;dreeedfo;bsderlqq;kovleibo;hifrislb;likhoxki;diseksk;shvrfp pr;vroephlk;qxoipixi;qkxbjhxx;klqllhep;hdfqrffi;qfrfkvve;kifeqpxx;sexklboi;rdedeorv;klsvxdfv;i shrdvvj;kifbkqol;shrdvvje;hphbvqor;ridkbvxd;vrsifosi;bbfovxdq;fbxjqxse;hdbxvbdq;fpkddqbq;s hhexssr;jxkxlsfx;qkqrqrhf;dqbbebbe;vxpsrfbk;dhkkhibl;vbbkpkfi;efxvxidp;hlklsfbf;sxsxiobe;qrd xpepd;kkvvfqls;qbshvbpf;rqsvovex;bhvpsrbo;kbjxeklv;bjxeklve;ddrdqrov;dqrovflo;drdqrovf;hkb jxekl;jxeklvek;klhqedse;rqqfxixl;efpjpbeh;ihlvphhe;keishpqe;fqsdflid;vblkpiib;fqhpqifk;bphhkbr f;pvovkffp;qprdelve;sksijsds;xkqsfdkb;eqoqxxrb;lsesrhbp;hodobevk;dfbhelep;qvlqfdvl;lppjlebp; eepopble;pxqsxeed;elpfqfrl;fboerjxl;boerjxlb;rehqdqvb;pkrrrxrl;eeikxlob;obhlejbh;boskopdp;sxi qvled;qxrbbqjk;khlxslfp;xpoelvvh;kvphpfqo;lrlrkxkr;jerohlhp;soxphdvs;fesbkvsd;xrllqbde;hixlp qpo;jvodbdrf;bbkilpii;khvhkkfx;ejloeqpj;qhsrqxxl;dqbbovvs;ehxfbqee;vobvkobi;rvpoepbk;sxpqk dkp;srofdbvo;rospflhl;odqodqhh;lpsderve;vqbdkhpd;bridexbq;hevkbjbf;bdiihhqs;pbfhofkf;lbsehs rs;fdsellrv;hovhrflx;lihbposb;bjbxxvlv;sijjvxxh;llkqqoho;pvfkhbdr;lrkbkllx;xhbojflx;sbeilvfr;blv hlifl;fhfkshdh;rllkhrrx;hfxdffee;fhlqbide;bdbshkpi;pkiefqfs;rvvldhhv;ofievdds;pbposvbh;elprblbs ;qrfqllbj;fkeverlv;xvverdbf;dhlrobdd;kkkqqqsh;eesrsvlo;dvlxlihr;fidsrkrf;spfrlbfl;xeekkhlr;elblob oo;iosxrbdf;spfoehbe;ekhrbrsb;qbfhsbre;pfheeffi;ovepefvh;bixpqvfr;vribxvqf;xkprkfbk;flvkxpfb; sxrosxdk;fkxekkri;sskhdjrd;sdeehxph;fqbdxheh;oxlexrhs;loixphkx;lifloixp;pdhveikv;vlifloix;boe pxqqf;kqqvxdsi;eikfhxxb;kerdrkeq;lkqflfbv;rhqvjlev;hhvkrdhd;hrrrfoqf;froxqffi;vlrrhekb;didohb hb;bkdrrrjb;odehfxio;lproxkii;ikhhehhb;ibqhkhxf;fsvlvoxs;sjvqxbhe;qkqbqddd;eqskbddr;qkjkpks v;xblxpeor;refvlpkb;svhqdfbd;qppllqdb;sjpbllh;vdebbroq;eqhebxrr;rvhpxbof;irpljxqh;vobpsxxl;j

vqdbeeb;kkjsvlkv;leikebsi;dldvlbdq;jxxiexjs;xfqldkvo;xkjeikie;bqbhpbvs;dfkrvqlx;irfxxelx;qkks hxlp;bhvebffj;hvxqkhks;fxbxxdqe;sbidrdhe;bdifekxe;xfxdvvih;ekxfepsk;ifkoppvr;vxlkqfex;hkhh kbir;khfvxbfb;jeovexlo;ojkxlxfr;fsvlhoir;sxqqhfbr;hedfvfqq;bpvsfevb;flsdkvvq;idbpvfhr;fdrdhbd r;islkhbob;dfrbxhjb;jvbxrfqo;dbepfqee;fpqdvxvp;obvveosv;phfqlpkk;vbhhvfsr;kqoedjpl;ivvhlipp ;qekpvovh;pveoxrie;psvfxfrh;xqbrbsed;xflqkqpf;rleifxhi;ovieiddb;xoqxxvxo;eqokifeo;lhibfodi;ps hebdhs;rvdohhlf;hdsvpkvv;plefrpof;dxbhkfhq;xkljhhhh;lhfpqkko;bkbfhdho;kehosiex;kbplxfxe;hf dbqkib;xfpvxirr;rvqqskfb;hqorxhxf;xefdrxpx;hvsbvqbh;qlvbprhh;vvfleqdd;deorevxl;xhqvffrd;ex brsbkl;sesfsrer;dhxvbkbr;kirefrvf;kxofxsip;dsbqqeqs;xssdrfeq;klkhxpep;pkfdkpbv;lksvjxos;qehb efkf;qpppporo;vppppkeh;obfhbkdk;fphhxhes;vsishjri;lpfhhieb;lxlsxoxk;vxkriiox;ojkhispp;bvbihp fo;sevplhhs;oxepieoh;oxsdvffs;rbphfexd;xrehlbqj;xxkifbqd;bisffrkp;ixqseehv;pxexdhsx;ihvsvsdd ;forvqohp;psvkoeri;kpsvkoer;vqkovhqd;dhojiidk;vdsvxvrx;lhoqxbob;vkrbjkqj;qfbrxpbs;bohbjflv; xffhdjbv;fspxlrrf;vboikxis;bsibqfvr;psvppidl;khobdeje;ixeeifhx;drdbvssf;ebdpbelx;xijssrqb;kfoxd dse;vevrehro;bbhxxpoi;vpihbfqx;ppevqhjo;bkhbkqkp;xpiesvfb;vlexkerr;jisfxsxh;bjbpkqqb;sidpk pev;fvsfobkd;exhkrldh;qqkspqee;qfkqpqev;fprpsvbp;lrqvqpih;seeqlxfb;evpvlrbe;fvdfblfv;okekhf kb;xlfldbpv;okdkrevx;rlsvkpis;ffqbjhvv;diselhrx;bvhdfskh;vosdhldv;qevxbler;fkxdxlpp;rsxrkvep ;sxrkvepo;hqxvbpdl;fpkqexlh;vhdhxlib;evqkrdfq;qpsxbhvx;jphlhvif;lqvxqbdk;dxfohvkq;hfxhpvh s;okxbefod;jebrkdhk;eollfdri;lfdridvb;oeollfdr;ollfdrid;phpjebrk;pjebrkdh;voeollfd;oflbhxrq;hpxo fxlh;seqvhxvd;vfqdkdhi;sxxfhxdo;rhlhpivb;dfvkbiff;obdebrev;eklflqvo;rhqlboxf;khdxfdlf;qshdxf ix;jqppoeji;ehoosdpv;ejkhjqpp;hoosdpvp;jkhjqppo;kloejkhj;loejkhjq;oejkhjqp;sdpvppho;svehoos d;vehoosdp;rxeqsrhh;hiskrllr;siskfvbq;oedefqqv;bxxxfipe;ldbxdjbb;rdqhqrxr;oshiijde;ielxrvks;xq xloiie;dikbdsep;fvdxhbef;xrsqoxvf;hfsrfqks;vblhxebx;hvoebsxb;qpvbbhjs;drfpxkbs;hvqbrfqq;hh xfrifo;exoxirri;srldvxlf;fideifrq;rhjrbdre;ivfkispv;oklsskjq;xqxkikor;heffskhi;ivllvrsd;lsedkqpv;er vriibp;hqbirxho;erdkfseh;iefpfhoh;idkobvbb;exlhdfes;hbrqepsf;hbssdsee;xokklbsr;qklvxhbb;kljkl els;vkkkkjex;xfehprlb;xkdkorrf;ledkoeed;desxdfqs;bhxrvqpb;qixpsvpe;jpihpqpe;povxsivr;fffvfeq o;qpjelxrr;irqqkixv;sookpirq;pskxxrkh;qxxvdixh;hvhvehbh;qbofovvv;lfxdqrsx;fvhqpeer;eqovbbp p;oqeeofob;hbjxvldf;hvesekll;ikxffvsk;hxkrfrqv;svreqkvk;rlfrdvii;exjsxiss;lpbljbkq;plhkodji;ikhv eqks;rrpkiskr;sisbdeql;ovdskhod;lxsefxfi;pehsxjxl;ledxqvsx;fsdpvesh;dpveshrh;hjqlqifd;jqlqifde; sdpveshr;xhjqlqif;bdxdbxfk;jqdbplbv;ikxekepq;iirorbpl;lbvieoif;kxdqlfrk;qpsisehr;vqqxfqve;ieqr vbxr;hxliifsb;xbxerbxi;diphxoxi;ijofvksh;qhxibqpx;ehlflrsd;eebdjlbb;hifexrvr;qrqrqphh;hhvfrepb ;fbiqloef;xrvpxvhf;jvrbpelq;sbxikkvr;bfdebvsq;jdfsvrxs;lxrlisde;exepbhvh;blpvbqhd;qdvfsfox;bb kqesko;lkkdbfrv;ehvbhivv;pqvbkxqx;iiorexhq;posvphkb;rqihdfvi;ekojvixi;refersxo;fxsxrjbf;xsqb

ebbv;hhpxrdsr;qhxhqfkf;fvsfsers;vbphfldd;bjrfrhqb;pkxqeeer;sepfpief;bqhbefks;vfhqxkjp;ksepfp
rh;lrlderfe;lvhkobdd;feplshvl;qrhbbxxo;pjojjhlp;veekkeps;dhheedek;kopkpqde;jkkkvxos;lvbxfhh
o;bhdhsxbp;oxfxdfdb;xfxxsrjb;lxxxrokf;lsixkrrd;exfblepf;oifqflsv;xrkplvhb;vxqkhjir;jvihpkfx;qo
fvvkrk;qpbffxlq;psvqlvxx;xdfpxoeq;ehdlxkfl;evpsvkxj;eeeerivf;oeqlprqh;oqovsepf;rprsforb;edrr
hhfh;psxbblld;bjbfffsv;idjhkhik;seopkhhp;dkhpjosd;hpjosdep;opkhhpod;pjosdeph;qfxrfxsq;ferhfl
fq;dflvihxx;hbvhlkqr;ldhbdlbb;jdishlpe;efbbisqk;brfibqds;bqdshixj;fibqdshi;bhoqphkl;eqskpofv;r
bqkkikh;johkppif;llbisbrl;hfvrrpff;ribxxxri;eqksvpqv;bfvssefp;bhfblqlp;kbqfrkhq;kpbvhihd;rbdlv
ler;qqlklqbk;vshobevd;sxkiqqik;qrvhfflq;qleeiesi;xvqflskb;rfivvfed;srsdelxb;qhevvfpx;ldhsddxl;
vqxhidqp;pbxqobqk;kkqekobj;eelsieqh;dorblhhs;kivsihfk;lrkbxoxf;qppokohr;pvvlierq;ieodhhij;er
fqfkkb;rsbxhpxk;kohqvxll;kdxpvfbv;hdhdevvp;rdqkfled;pvvrvoxv;bxoobhvb;kqbvphld;frlsdvqd;
bpqfesfe;obvkivpv;rpphhjvb;dlbqsree;xdksdrdf;kfirxoff;ikliiffq;eqlirqbh;xpeqbbps;rxejqqof;xfro
pvvh;hiffefxd;bhkosrhs;isxqdepf;disxqdep;bxsqxrbx;iqsdfxrh;qbbvvbjl;ixxfjxxq;dovbdjss;jklbiq
kk;lqvpxdxi;qexxhkkr;ekxriqeb;oxxdjrlb;forhklpq;xrlkvloe;lelpdxss;khqriksi;isbrlbhr;dhrhkosr;ri
fddvkk;lxoihvdk;xkihvhfk;prerffdi;eqbqbrri;qqflflvr;risrvsxp;pivsbqkd;fpdfifff;ebfkkjob;jpqhpef
s;fxovdkoh;ffidfshv;sedkebvs;rbdbfpqp;hflqxxve;oxhprbbi;qkrobfob;bkfrvlsi;lxbrvkrh;bhfvhefl;r
eerrobs;sfqppxhs;vkdhlffp;vijkxkov;kqlhihrx;ffpqqhhp;kxlixpol;firpebqx;kdvivrbp;sbdkdxsr;rxii
xbfi;fxppshhp;hblrqqpf;idbvxqqi;dikbehdp;vlpskjde;vfrodfro;lkqkxosj;hhhvxvpx;xikdrrvp;jxxeh
qxs;vvrrpfqp;rxxssqer;hxovrefx;hribikex;xksrdjdr;ksxkbilk;sxksrdjd;xksxkbil;kehhoihd;lvivreif;s
qexrxff;qpvbdrkl;flprkbhl;qpqfxorv;sohfsbfd;erxvbxpj;fhkvsssk;rhfdsxvv;jlbffkdq;bbjlbffk;bbrsr
jvo;bdpbbrsr;bffkdqph;bjlbffkd;brsrjvod;dpbbrsrj;ffkdqphh;fkdqphhb;lbffkdqp;pbbrsrjv;rsrjvodb
;rvkoibhs;fqvphofd;sfxfpvqd;dfvifeos;lvhrvxek;hdeivxqr;shhohvsl;kideifqj;fsxidver;qbdkdkbf;be
klvxeh;iibpqqpx;pbiksdeo;bkfhxexq;ejrqpppq;hiiihkdo;iiihkdox;rooxvjeo;isobejke;kjhboorp;sobe
jkel;xhokjhbo;fhdfofro;hkqlqbhk;fxlqfbsq;vhkphfdf;qhlfohke;lixbfqvk;bvqbkqox;kpfbxqeb;ivlsf
xhv;fhijvsdk;jseervkp;sqpodkqh;qjvphivs;pekqqopv;lihflfhv;vipeokiv;divoeess;ibvvkplb;kofxfhq
q;elrplfvv;bephhvrv;eqoxklkp;pfoxfkxh;bppphivb;bldxxhxq;rosvedhl;lhbfqdqx;erehvbho;oxqhif
bl;bppohoii;devffrhq;xevhfvpv;bjpvxirh;bdpqqskd;lbjpvxir;xdfefekh;hvqxkvqb;beehqofi;plvhkos
x;xpdbvrph;siefxvix;jkbblbxf;pdervxkp;rfkbkfrv;ehhbvqll;eflffkjx;qhxvbbdq;vsfxlbbj;lrqrvohx;s
rvlfbvh;dfbprrbx;lqplivde;xpkdkser;sepixeei;hvofpxdb;lvbkbvqe;rifrqsfh;hfvpfsvq;kpxqrrko;hrp
xervx;ipovkjvo;hkxslhlf;sfsvslbe;ihefxdxb;vrjledbl;qkvsxiqd;obbovoip;sefifrrd;jxxhbqlb;kphrexi
d;idhbivlk;pqfqpifi;hieqsivi;vijlhfxs;dixlqhxs;pbflkhep;kvhfqrhp;psfqbbdx;xhpshlfl;bvfbqqep;ieb
vfbqq;pshlflre;hdsveofi;evhqdhvd;expjbjkp;pvpbffqf;dsxjkpex;seeqprvb;blvdlbkq;exrorokq;llisb

hvf;evdvhbiv;prrsdjvx;vliikqld;drklrvdf;bdxbfdld;xqsxqxbj;iqrbhebh;sjxbdbld;qbddvlqq;kkoeifrs;ffkovvfq;fvvfsvqv;lesleqke;eohbsjvq;orkqrbkd;xiohhhhq;lhepbbeo;hlbdkxlv;kbqrsfbp;hvobejrr;ofphboqf;rqsrdevk;qrsdeere;fjelfobb;pevffplk;isxrklqi;hxqxxfld;xkxbhqqe;ribkrixk;hsrhrodd;feqbovrr;hlkoxsvo;jxrhrrxf;iepevhqb;qfrvklhs;oedrbpvr;vxdxeixh;jikkfxxk;bddxdebd;fxdjhprx;xrioeqfr;xefbfbdp;xehlrqdk;odvkkvee;ddebhvil;lxvvflvk;evxpblql;xvpehpvs;lflkdbhh;dibvhbsr;efxofvpx;eqiddhqq;kqkrpvvv;kkkfplpr;kikhhdvx;blkbiexf;sffkkhfd;lbovplfk;ixrfvfqb;vbqklqoe;okprxpxv;joipsjii;pjlipojo;vhdbvbqs;kvedxlvf;hebbdkbp;bdiplfvv;ovrfphdr;fbexlpfr;elrervxe;bqxqvoxj;jrsvvfbb;eepbkqvq;bibhqsis;lbersisb;eifkvjvh;sfrldhqb;vlexsbhk;rsvfrdhd;isjpdklp;fhrqppkq;srlelxvx;qpkxkpdd;ldsvsexb;epiisboq;hjkvepoi;pbfpifkr;bvvkqqes;kleehrke;isxxpphx;vvokxxer;loibfkqe;eihbsvrp;ivrbvsbr;ddesroef;svkedvvx;hrpvblbv;xoevsxll;oexohelh;pessivbk;vrpplebf;vfoeebhi;bhdvfqkq;brqqsfji;olvbhkfx;oshvphrj;bikdejix;vkeovoko;qeloexph;leevhohh;qpeedikb;ofrlixql;bofrlixq;ehxedkvv;xedkvved;rsbhplkx;oeqlkbfh;dlpphsfp;esferoxe;qeqsfqhb;kddrfkef;dhvivsof;jvokjqoi;hrielojv;iiixjhbb;kfrxlhlq;obdqoffq;jlfhhxsi;bfilbihl;bsjbdieb;ohhkhjrb;hhkhjrbb;iddsdqbb;iklkfxhe;qjhlxxkp;hlbfkolh;divvideh;sdrhisib;oebelfrx;bfbsxifj;iebhokpj;sohpkqsv;eedjrlbf;bddsdier;qbdhidbk;lbidiqqs;bdhijvxk;vlsshdse;qpblrqkv;qvlboiff;ijpskiok;dsieqqqq;xllqhppx;vvovbexr;rfxqokro;vpldoebh;sshkvkbl;vkhfoidl;kdhkqxhr;pbqhkqoh;vebbbiie;jokvhqxx;xqxxqrdq;fkfdhrsk;lrpddrfi;lixrssfb;beqhebjl;efpvvlrr;ffiddhsx;disshors;kpebierk;ehlodvrq;jpjjoppv;keefvvlp;dkpqbqqb;hxsqlpix;kvdivhqs;kibqbvif;poxjbbob;isbfrdix;xfblkrfb;prijloke;dvvefpvf;hvfodxsb;sheifdki;kibohris;rdkhvkdv;hlqllohr;xjxboxvb;xrqlliid;jebxbdbe;fprkppfv;irfiedfq;plffpiid;xvosiifl;eqvvfpss;edhsdbiv;pedhsdbi;pllifhhd;vfdlllbr;dvfdlllb;fdlllbrh;klrjeebf;lrjeebfd;qkbqsrrb;qqxlkllp;kissesbd;dsshvqep;iibrpxxk;hxllhhpe;ddirxisr;qxqrqskr;bhhxisfo;frkelfok;xqdeedsd;ivqxhoxb;lssdrbfp;hqdxdxii;qhrkdqbk;ixpqikrh;dkvpsixd;kqelekix;eloxirhp;vjxxkxds;pfhephve;kefvffrl;ofelvevl;qiiffkv;rhesffpp;ixievofe;oierlkqb;xrvskbdi;ojedsjho;sxlfpikl;epifsblb;hpohxheb;pbljfiki;rvxkkljd;vfkbqhfo;fkbqhfol;lsresbvj;hrvosrib;eslqpbfp;hqjevlbv;qkhlrbei;sxdxfosv;lqffpflh;verrvlvd;psdrxrqr;qxeqvprr;bfrrsibh;leivferk;lrqrkpxs;xhlfelrk;xhhqhiif;lrvidbdx;hqqfhpds;hkpibdle;qdvbqfph;vlxvlpis;hhfrrihv;vqirlsed;bqeevifl;xsfphkkd;hqedvbor;eqvveodv;evisdbis;oevisdbi;qokhhdki;qppoevis;visdbisj;vqxdjxkv;hqhvqqrp;dfqsvvll;prldbffv;ihplhbbf;ovredvei;vqoerpfq;iqdfsskh;xlqebvhs;verlfofh;pxrvfdfp;qdrfklxh;koxqpkkv;pxhprokr;obeffqid;vehjqrxq;qiiesxrs;orlqkxkf;djxovlrd;oxpbdkbo;fhqffpvk;rhljrvfd;bfdjsvhl;ivhfqhli;qxqqeovl;hkvbsvde;xdiillls;kefbrerh;krlpfpfo;xeelvlvi;rbxploie;ljxsfbbh;rhfhiiel;oxxdvsvl;llqqpfii;epxhqxxh;vebesjie;fshxossx;sbqsxphj;rlvbpvsi;hoshdivl;rdekfklo;iifqqfvq;bjshpdhp;xlqiqpdd;kkrrhvqh;rldvoxvh;rxkflohv;qohphlrb;iibqvfif;drhsvebk;

blpvirpb;lqrpbhdf;ssrsvhsv;vvhbidvp;sdfviofd;fvorqeqp;qxdfprpf;hlefrhpd;rsxxpfhl;lesfhobl;exhs bhel;fvffovkq;vxexllob;ikhofqlf;dxdxrlqr;bkiibrrr;bdsihfff;ebdbskqs;iedivork;xksxsxjb;dxefkoiv; lsephlsr;sephlsre;rvsfbjxr;kirhfhex;hohhhpop;fekphifr;vpbkxexv;ibvpbkxe;pldxrqxq;dsdxqxbk;e pfqxrdk;ovhvidji;qbvspkrd;kpeesxse;vqekddko;sxppxopx;djhdihed;ebiobiib;iobiiblh;ldjhdihe;lldj hdih;brrsshpe;eqkvbeqj;ilixsbli;svbvxexl;lbejseee;oldehfof;jbhobvhs;rfjdshbe;dxodhikr;xfissvqd; jxpxfxpp;hfrlekxs;evjhpqhp;derofkle;rivovlqi;irdhqdbo;kxodqlhb;kxqhjvhb;dffqhbqj;lhjbebfs;jsv ellbk;jkpsxpdx;blbvbxjp;brplobhi;kpodhsef;rqxhibke;rqvqlbke;xpkvvqpf;rshidboh;plbqhvhs;bhp vloeo;xklobhxq;hblsqksi;hpxxhrli;ovxvkxlh;vvbrqrsk;qlfexfir;dsrkxdfs;rkvhvfdv;dosrkkli;hjkkds vd;xvbksbhd;dvvrebsh;jokxikov;vpjbbfxv;fvhbfxxj;obbxxsoi;vhbfxxji;xobbxxso;epreffbk;bfkixe jb;ffxjkbpe;fxjkbpef;rxsorell;xsorellx;hhshqrdh;povhohfj;kvehxebe;okphdfhq;ddbdedke;dhqxobr h;xbieffre;diixsjkk;hiikxios;iikxiosx;qlbilsbl;ehkkpkbo;qkvvikke;qlhprhqq;ifssoesi;kpdexpvp;shb eshei;evlejeqx;ffsrqboe;lhhhsexi;qrpevlre;fqlpifvx;dbvokhhr;ljloljph;kbivkljk;siqxfxxi;rfkeholk;x jxxkrll;bjrxxkox;rhhvbbvi;ldjbrbld;xkhxdkpf;xbfhivqk;xhvlibeb;hplrrrvf;frvdsfqf;qexvpikd;fdve drbl;fqjpposx;qhklxvdd;xqbdkdvq;rvfblrol;vrlpvevh;pqqlqvib;xbkfkerr;fxierpqf;hkdhdpbb;selrkb ri;dkhjrsfb;jbsvhbkx;bqxoefsp;eevpekis;kfsrbjfx;bdqphxbb;bxsioxfh;bfxijkrs;fxijkrsd;bfeirjbd;xq xjrpkf;ebedqhsr;hrqxkpkv;doiilbkk;qxdeedfe;jhbvvbkf;dobfqldr;dhrhlqhi;blhedife;esidfbxh;rfxso fbd;sxrlxprl;lbrskhkq;xdvbpqbl;kriksere;lbvpoxjk;qkxpsxkj;hdbpovbh;irlqxppf;kjpbbbob;dhfbve pv;pqppbrdq;dhjxfsih;fsihihpd;jxfsihih;xqrssfik;flbhbshd;expfxqxh;rpivvvkf;fiivhlxr;qisrrelf;bqp ffele;djddrfbf;bilhkbrb;rdjddrfb;xbilhkbr;vbxexjxr;rsxrbxlo;vkhfxbef;helporqf;dbpepker;vxlkqbf v;pepkdsfk;lqsfppis;skdeokpi;xxevedkr;vlqedfex;ihorpfof;vpxfrvvs;lpibbfvk;qldhxkrr;ffhlqxie;eq pflxrp;pidhhqes;ivfkobxv;bbofresp;frbribih;pqvqfxbs;bhlisied;bexvkelx;qdiebeff;xfefvipx;xvbrq vpe;heplbdih;bvkrsdbd;bdxxqqdr;ivkidflf;skoshhvb;epbvosex;vverelii;krkhhlos;ppqixdsd;pskrkh hl;skrkhhlo;fsxjhlxi;qrqxfrde;ivblkejd;bdxqplxi;qsxhfdrd;bixkbqsq;brhdksre;dbixkbqs;dksrexjx;h dksrexj;ixkbqsqx;kbqsqxrr;ksrexjxx;rexjxxfb;srexjxxf;xkbqsqxr;pfkhrkhl;vlssfdse;ksvhkoej;lfddf fbo;kvlxvsie;ixirfqlk;hsfxfidl;qqqrisid;bpkksvfi;lseblokx;ehfkffrs;eikxiffs;xkkekvir;dvqqrklp;rsef fsde;xrjbdoxp;vlvxbklb;orbqsdeh;pbhqqxkf;hkkbfovd;qbvvbfkp;sxhpxqoh;ivierlxs;pkxijllb;hredl hbp;fksbosdp;fhlljxie;frhhqvqp;vejxxlsf;elhskfib;ssvskopk;qkkqivjo;fihidkqv;osdlfhxh;fdrfrhvv; vdidqxbr;flihjxrf;pbhprksx;hejxibxx;dhkvxkpf;dixpirbl;ehifbxos;xlxsbird;ksxldvvr;evblkved;vol brdvv;vlreqqxp;xpfbpvxv;hlfxokho;kxsrhloi;hkpvhpsi;vqlshplf;iseqofkh;pvshfrxl;hphfdxpl;sxbq hskf;ejbdijks;ejpqxjkf;fdeorbbd;vrxvsiss;ddhsdlif;rqerbssf;dvphoelq;bxbllsvr;jpjijdsj;hbiipxxk;fp pxklpf;xfokehvx;svifhhrs;vhodvxfx;vihsbrho;bjxpvssh;dbjxpvss;dqvqqkid;hdqvqqki;hhdqvqqk;j

xpvsshk;psdbjxpv;qvqqkids;poebphks;epqlokio;opeivekk;shifvksk;blklodkq;ixkhiijd;qbhhlqxp;h lvlvxbk;blpkifxe;drbvlokk;qssfixrv;hkfkdrhr;xvflpkvx;qidvklhx;qqosksjv;fxlqqdfd;lpbqdxsk;lko kbrvf;rbfeoejl;bpfhisfb;jxbfqbfr;ihfrdfhx;xloforvo;hrikqvvb;qsdbhpqf;fxfflrql;kkfbxfpv;rksxefpd ;jxqbhreh;fihfdehe;oseesvsf;khlqkqhv;pxxqqvep;lhplhibh;xpkrxvrp;qkevrrrv;lkepbdef;klllrksf;hb klqkfr;hfehvhbe;bdikpqhh;kljbdfqf;bvdobbpl;xeqvkqfq;jpselqeb;lerfkqff;hkxxeppr;dfibdekr;pper iqel;djrpblpb;ephpperi;eplqdjrp;eriqelee;hpperiqe;periqele;phpperiq;plqdjrpb;poeplqdj;qdjrpblp;r iqeleel;xivfqdfl;oivofrbi;vskjpfpl;hxirkkvo;oxvxqdsd;lpdflqie;oehflojk;kojhixdk;qdedfkoh;bfbqk fih;lbresrsi;vkelopbx;ieblqfdd;pvdribfs;xbhlkbfv;ehhlerrf;ssdifkrl;shvqprho;jbblkirl;vsershqe;brb hxeep;sfseevfx;rridbhri;bohfqpbo;qkovxlko;qerfobpx;svvkkdxl;veflkrlx;vpiehlqk;veedlfoe;fhefb ses;vxqvlrle;ihjrkxbb;expekbex;lqvlorbq;bbrbxlpj;kpeqkivh;ovfqpqkp;ipokxdki;xfkkpxdb;kklkos bs;xxrrpvdl;lijskise;kjxxkxbr;ovesxbbx;dsxsevfk;xrkoshlv;hkrfkhle;xlrfsriq;ofxrxorj;rpexiesv;bh peodfk;xriierqs;disfsrkb;ikhxkdrb;fbhxiqkx;ipkrfohx;hiivlxhv;dsvixxpd;vdkhxqre;svvsxllf;xsrfkh dd;xdhhrblx;vqlpqrrb;jeeflqrl;ixkqvlql;xelreifp;xirijvfx;hhovsbqk;rdvqrdvd;sephexsp;forhvxdk;l veexovj;pveikvlp;kxxkokqd;dhirikbs;hirikbsf;qvshqbbe;ddoekbho;hixdsbek;roxvereq;xekxlqbp;o llfqskb;jebvxirb;dprflflb;beshrhhl;beqpqpvr;fbrqrqhe;beqplkxq;ddpxhbbp;errfbqrs;fhhsvbsq;rbxv kkxj;xeeoeqpf;hqlbhjss;qfsvvqlx;forpqxxq;vrkfoqfe;borefoko;hvekeboo;lfoqxkhq;ohohsrhs;hisx kijl;ddkkssip;dkkssipe;kssipeps;lhisxkij;xkijlpqk;plopldib;hlfddvxd;fddhlhre;iiihkkho;pkkpeopk; hrifkxff;hikrvxel;eikqkddx;dfkfplde;ovveqvop;rbddkfpi;pxiqfkxp;pbplsvhs;qsxhdvrd;xvpxprkb;o xqproir;eifqfskq;exseishe;xpeovbdv;sehllvpr;sdflrpbr;rrrkqsrq;ivqvexre;kblfvedx;hdbxrhdh;prisr bll;qvqsidkd;vvfsbeoh;lexvixbq;oefrxrkd;evbbxqxp;edflbjqr;vksdjsib;xsfpddre;ixevshvb;xevshvb e;bfoshejx;odhqejsd;vhieroqh;xvdrqqqd;vdrqqqdf;ddfflseh;evfkerdx;blxvihoe;fvehirpx;sisoxief;p xvlssrl;eqbhxpbo;dfkpikse;ihfkoeqk;ibsvhpsv;dvqxoxbv;lporobbr;lldbvffk;hfkxrixs;hxbsrikd;bbi offkv;dblrvjqx;hebelqfv;llqsexrf;ofixhpes;vhskselq;odhpoesk;evroehqb;jpqpikhe;oexbkkkq;qvblr fii;ibksiixo;frexvbds;blvliksb;brqldsdv;jrfqfoeb;iblejpkp;rbpsfvhr;fvflisex;rexfoxjf;pfhqlllh;flfxb vpo;vdseqbvb;pphvbxoo;kxfiqsdx;dkqrkfrr;sevfofsr;rqrssfbv;fqqdpkhq;fedvxvbe;qbbvqlrj;vvooe flp;kqpjhlve;qpjhlvel;fhbivvlf;rsvsxvbr;siiikpki;kidhefds;hsblxhbv;diehrdxq;pveqkksh;qbdhfeov; xqpqbijk;esikexbb;pfssedrs;kfbbqkel;ijfvhllo;qxssfoqh;ifvkjvdk;psdlobvl;rselxrfs;hhslxxxk;lbbfjf ql;xsoexqfr;dihbxqvs;bqbfxorl;qqhxfbqq;isxlkljk;sdxvqpvv;ppsfrsfl;bobohxqv;lvivpqkf;lxdiiioi;i sdqxoie;xxqssfpd;oksdirdr;ixhikbkb;bpidvsrd;exedevok;bleivobi;oherhqdj;fsxlehop;dbdhfvik;pjq eblbe;eoprlebb;lpjqeblb;bdesdhrl;qfohrrfh;vrviffbs;esrvrefp;brlqfbqi;fvvbdkdx;fshsfxdd;hhkbhev b;hhpevihf;vhbhfkbs;kofevxqo;phhddrvh;lqhleevr;hkfbvxph;iipihrhv;xeffvixd;rxeriixx;fxhqihpq;

ehbrpxxh;qepoksbp;kvbpiirk;xleoikds;dddbqveb;oivqjski;lpdliepi;plpdliep;vbqehbss;sbfvflff;hos vesjp;vqqhdkek;xpvsbiro;fxivhirl;bpdxdhxk;hhihibkf;vsjhhvob;qrblkovb;jirpsisk;xhkikqif;fkfore he;lxpqvvhr;eqpffosk;fkrfhoff;lsivdfbe;pdfvvhof;kfoevvbo;sepiklrp;irbqekxx;vvbpsdkv;vedsoed d;qkqflffp;dkqihibv;iibehklj;kqhxxssl;jebbflhl;xrqxvsid;lrlfvrer;jhbqphov;hksjkkkh;vfpqfopq;oxi kosvd;kodxpobo;vsqxprxd;hxohxdkf;rkpfxbqo;elqrbxjs;lrxxkvhk;qvipphle;xvoihipp;dijqppib;hb eqplko;ipvpohbe;jqppibbp;riipvpoh;jpjbsfhp;opolfhse;pjpjbsfh;polfhseq;qbkqxble;rpvppolp;xhbii vrk;kqephlff;qhbfdqrv;xxoripks;xhpdblre;kdibbldi;qkshvsbs;sxifqhfk;ikdllkbk;ekhvlfef;dvvlrvhx ;dkdvflpk;irklveox;sqxsrksb;drebbfkv;xxxllkli;plvkblpv;bqifeddr;ihfklbos;bqvkhqih;exosesie;ke bqkvsi;epdblksr;vxqlepjh;kqvvbpoo;dvdvhkeb;lfvdedsf;oeforbld;ldideojb;sskqlplb;lobevheh;fiid dkvl;sihhixpb;djpoexvo;fihfkixp;srlxbklp;ibpodfeq;obpqxrql;fplpqkos;exsvveor;hedssdsr;oebjhp ve;ehldoeql;xervxede;kbqshblp;ephlerll;rxpibehr;vxvxpssb;vfixdvee;hbexrris;efhrefxi;qsbxlxfj;b xlxfjvi;fvfrsqok;hfvfrsqo;sbxlxfjv;xhfvfrsq;xlxfjvis;kdddxvoi;bvdfkirh;vqsbfelh;xqirxvbi;xkrfkiv p;xoepkhof;hfksbhel;qfxdpqsb;rpkkrkxi;bdbjvxhv;bbhdqqsf;dvjbbfrp;hlpqqbqq;vjxrxkkx;lqbrorr k;rflpdbhx;qffbkkeb;peiidpos;kklokvvj;klokvvjp;lokvvjpf;veixqopl;hioixihl;kplepjqq;jrfxssix;iqb xxkik;viieekes;kdpfxpph;qqhrfveb;bibdexhv;bdhbhqsf;obbxbrkp;xhdxrorx;hfoqbqfx;hdvhkqfi;dv lhbhfl;ebjdiifl;ofikqbsb;srksibpi;xhebxvrl;lfefvqrv;ekllxdvs;fieovqoo;hshpkpvj;hsbepxrq;qkhvhf pr;dhxoflli;xivqqqpx;xibphbli;hbrlqeoh;xjhbsexe;didherxv;dksjiflb;kxkvvske;olbbvrhl;ixbfvosi;l dvvxrkd;rqbddexh;bxihehho;pllhsxoe;rerbhivp;jpoepops;beepqeke;evevvlkd;kelrjffl;bhshoqrk;df qleqpl;sspsdhpk;ojshdfph;sfheqkjo;bpsspkpv;fheqkjov;heqkjovq;hsbpsspk;psspkpvh;qhsbpssp;vs fheqkj;fpbsserf;psfxeovf;qhxrpklx;eersibov;rhfrhdse;qvkohhsj;hfplqqhb;pqxbdkpq;bejishox;xpd djdev;slskvvis;bvlhqvrh;ikrxxqbh;oebdklrv;vdrhrvxi;ddeohkkr;pkibierl;fxlkoeei;eodiodhk;hpevq opq;lqpvjpvi;pevqopqo;fsreierd;bdfvqkdq;hxfqrlfp;hhppfexi;lxpxebds;hbfhivps;rvbxphlh;dbdbhs ob;fhsdhsjh;okqxxdxd;klhpvxbb;hbrsivos;ibffikpk;elxxjxlf;vrexlkpr;dskvhrox;sdxevkqi;qrokhxo q;bjhflpkb;ddqkvixx;llfkbpih;hsddexfp;lkikqvlv;qdxhhlqx;xposirir;vlrdrdbq;rxrklxis;jrbkbdbd;rlf pqqqp;ebvpvvvq;vxofksko;rebkskii;qkxkiqjk;pxxvprsr;ipfpxsbs;efvbdhsj;brfldrev;oeeeiddd;hibk qsbi;odhdvxhd;krqdfvve;bqvvfohr;lxfqklkb;eqridbrx;qporkxsv;poshrhqv;krssxqrl;qrlxsbpf;xkdxs rhi;rrxrkkkj;rvrjsrbd;lrorvxfb;rifsessf;eobkfhrq;hphdrsfe;dkpeqdfx;ebeodbbj;iepphqbf;ovsbddbh; xoxiqxhd;dflqosrs;hhlqqkps;bkklifbv;kfhpvlbf;pjvhpkpj;fhrlliek;obfpbhqo;ifkqflel;ifhxhpbl;xerlb qpq;rqebevpx;rkxrehpv;rxphxdsk;bhxdblql;hifqlbdr;ddbbkpdr;jkqxphok;sovxvoei;xjkqxpho;bxso befq;xhshffhr;vehrrxrq;hhepqhlh;qkliixqo;klbfprpb;hvfvordx;qlqkrkbh;hielxrff;khvxebde;viberd ko;xelrfihe;vifhkihp;ifqddhfk;kphrldsr;jebbbfbi;rvssfpfe;pvfvhlpx;kxbxoblq;sxrfvhee;sqbxdbkb;r

qvrfrjb;hbsbfifj;fdhdvlfe;qshjrhpv;llkpjepb;evkiidrb;eoprbffb;dbbdqqrb;vesdxobk;hbrihipk;ifhrrp px;vobvqrde;jeebeflf;qdhjrbhh;xqqxdqpr;xplqbidb;elkeffoe;xldsfseq;fkkrsssl;bqbhdddf;hqxprhrf; lbvdrfbo;xddhqfvk;lkekvofh;klviborq;hqjokxsv;dsespixx;khqjokxs;hkxkvhde;ieejflfk;ihveqxqv;j bfbfele;boevrjxs;brprrhev;lxdxfohq;lvdbdder;vkshjhkx;fiekekkx;kphlxlls;sjiijbib;edkklhlh;jobisf kk;lkqxqshe;bxodrhvb;fvhkdflb;qdefqlex;bppfllqh;vliidbho;rfphhhkq;ofvlskof;vhpfibvh;hxrpkvq h;kddsrdvb;hqvpkklq;rplxffbh;qrrrfikf;qdflpllq;qxihrhhp;lvvvlxqp;xljflbbv;rjdeddbb;rbdplleb;elq vvhll;epbdbifq;qeqfplxb;qffldhro;efoqihib;llvjsidh;khplxhpj;fodfvfpl;fqqpkkiq;qqsvekrb;qhpbifo b;pxpvdbqe;pfbofppo;viihvkrs;ixiiixvb;bovpxsqs;vldffvbs;kkbleedd;hivperoe;hsxkoevp;skdvpieo ;lfblsvsd;khihqviv;pkssirke;bhkxeepl;lpvrqvkb;pobeixkv;drfpkkqv;vhplhhvf;oepddflv;xehxdexq; vhljqpxp;qoedpvqv;peekpfke;hqqkvskv;shxiqeqx;ookkhifb;frvqllpj;dirrssfj;lrbdffes;bpkoidbo;xr pdxprv;qpsvebkk;hrhjhovi;biksbrpq;idpklbep;ohjovbbp;hhkifofr;kohhfkeo;khdqhbfb;bvosvdkx;f pdelxid;bofqbpxi;phvqqrrp;krffsvdx;qlxkkpsd;vsvxqvpr;dibhdedf;khvfdxfq;dibbbbph;qibxpkvs;e shfvoxq;fiibskvs;bbqfefes;sxevehop;jkosdref;hbdjqiko;sisfoess;qexkrllr;vpdebixv;idpdhdhl;lbpee dje;fbhlelhs;khrbvhfi;vhqbrsdq;iokrdhih;lfrxbjbr;fkqdksfl;dxrisbex;bhxeqefo;ibskkhkq;qlbbdxql; evbbbkvv;kkqolovh;oekjhhlq;kqlhxrqh;hrkkxird;oskqdedv;fqlvrepx;oveqhrse;vhxlvprx;vvxxslke ;hijpoesj;ipdviipq;vkqldbbs;xpsvjveb;hpbeqfvq;lkrkfdvh;exhpefof;hijdspiv;lxlbhxbp;lphjdbxi;eo dlhfsk;hjdbxixl;odlhfskv;phjdbxix;efexbqok;llqhxjkf;hkheodpp;xkbfkjrb;exkesedv;kdxrhllo;dvpr qlpk;oblsxehe;hovxdeqo;redvlexi;qrqvkvhr;pqlbpoxf;plpvhsfk;bplfldsi;phkfkpvl;oejeilpk;ejeilpk o;holojeov;vholojeo;bkqdbfvs;polbkosh;ebrekqro;ekbhoxfs;dqxxffdh;plxbesrh;qvrhvhde;pfrhibq f;hpeqkofx;xdkeqokx;rirpvixv;hqirlesx;qxxxvqlj;hobhbdlv;ohehrbed;hqfkfbed;bdxfjvqr;xbidlrex; vjhqxrxf;lrfbbkoo;hishbehx;iflvvkhp;virordfh;vlhbkvvv;rpqelkxi;rieoxbxq;rderisvr;sqsbepqh;qlii xvbj;hxpdkphb;fixhqhrv;dkvhhbkq;fxbfdlhx;dlhxrbrv;rjdfxbff;xbfdlhxr;xrbrjdfx;qldvqohb;xhxob xbs;bhqbvkql;qvixifje;qrkivpri;kojefjvi;eoxjkskv;qbqieblo;oxxlsrel;rbskkhqr;xepbvers;exjxqsxr; pvhpbhpp

Biological Samples

The expression level of one or more disclosed nullomers can be determined in a biological sample obtained from a subject. A sample of a subject is one that originates from a subject. Such a sample may be further processed after it is obtained from the subject. For example, DNA or RNA may be isolated from a sample. In this example, the DNA or RNA isolated from the sample is also a sample obtained from the subject. A biological sample useful for determining the level

of one or more disclosed nullomers may be obtained from essentially any source, including cells, blood, hair, tissues, and fluids throughout the body.

In some embodiments, the biological sample used for determining the level of one or more disclosed nullomers is a sample. In some embodiments the sample comprises circulating nullomers, e.g., extracellular nullomers. Extracellular nullomers freely circulate in a wide range of biological material, including bodily fluids, such as fluids from the circulatory system, e.g., a blood sample or a lymph sample, or from another bodily fluid such as urine or saliva or serum. Accordingly, in some embodiments, the biological sample used for determining the level of one or more disclosed nullomers is a bodily fluid, for example, blood, fractions thereof, serum, plasma, urine, saliva, tears, sweat, semen, vaginal secretions, lymph, bronchial secretions, CSF, whole blood, etc. In some embodiments, the sample is a sample that is obtained non-invasively. In some embodiments, the sample is whole blood or blood cells. In some embodiments, the sample is cells from a hair sample or nucleic acids from a hair sample. In some embodiments, the sample is sputum, saliva or spit. In some embodiments, the sample is a serum sample from a human. In some embodiments, the sample is a bodily fluid from a human. In some embodiments, the sample is a liquid biopsy from a human. In some embodiments, the sample is free of cells but comprises cell free DNA or RNA.

In some embodiments, any of the methods disclosed herein comprise using a small volume of sample for detection and/or diagnosis. In some embodiments, the sample used in any of the disclosed methods has a volume of no more than about 100 microliters of fluid. In some embodiments, the sample has a volume of no more than about 90 microliters of fluid. In some embodiments, the sample has a volume of no more than about 80 microliters of fluid. In some embodiments, the sample has a volume of no more than about 70 microliters of fluid. In some embodiments, the sample has a volume of no more than about 60 microliters of fluid. In some embodiments, the sample has a volume of no more than about 50 microliters of fluid. In some embodiments, the sample has a volume of no more than about 40 microliters of fluid. In some embodiments, the sample has a volume of no more than about 30 microliters of fluid. In some embodiments, the sample has a volume of no more than about 20 microliters of fluid. In some embodiments, the sample has a volume of no more than about 10 microliters of fluid. In some embodiments, the sample has a volume of no more than about 5 microliters of fluid. In some embodiments, the sample has a volume of no more than about 1 microliters of fluid.

In some embodiments, the disclosed methods comprise isolating total DNA or RNA and/or amplifying nullomers in a sample of no more than about 5 microliters, no more than about 10 microliters, no more than about 20 microliters, no more than about 40 microliters, no more than about 80 microliters, no more than about 100 microliters, no more than about 200 microliters, no more than about 300 microliters, no more than about 400 microliters, no more than about 500 microliters, no more than about 600 microliters, no more than about 700 microliters, no more than about 800 microliters, no more than about 900 microliters, no more than about 1 milliliter, no more than about 1.1 milliliters, no more than about 1.2 milliliters, no more than about 1.3 milliliters, no more than about 1.4 milliliters, no more than about 1.5 milliliters, no more than about 1.6 milliliters, no more than about 1.7 milliliters, no more than about 1.8 milliliters, no more than about 1.9 milliliters, or no more than about 2.0 milliliters. In some embodiments, the sample size is from about 1 microliters to about 2 milliliters, from about 20 microliters to about 2 milliliters, from about 5 microliters to about 1.5 milliliters, from about 10 microliters to about 500 microliters, from about 15 microliters to about 300 microliters, from about 20 microliters to about 200 microliters, from about 30 microliters to about 100 microliters, from about 1 microliters to about 100 microliters, from about 5 microliters to about 75 microliters, or from about 10 microliters to about 50 microliters of liquid sample in the form of subject plasma, whole blood, blood cells, cells from a hair sample, saliva or spit, or serum.

In some embodiments, the methods disclosed herein comprise isolating total DNA or RNA and/or amplifying nullomers in a sample of no more than about 5 microliters of serum, no more than about 10 microliters of serum, no more than about 20 microliters of serum, no more than about 40 microliters of serum, no more than about 80 microliters of serum, no more than about 100 microliters of serum, no more than about 200 microliters of serum, no more than about 300 microliters of serum, no more than about 400 microliters of serum, no more than about 500 microliters of serum, no more than about 600 microliters of serum, no more than about 700 microliters of serum, no more than about 800 microliters of serum, no more than about 900 microliters of serum, no more than about 1 milliliter of serum, no more than about 1.1 milliliters of serum, no more than about 1.2 milliliters of serum, no more than about 1.3 milliliters of serum, no more than about 1.4 milliliters of serum, no more than about 1.5 milliliters of serum, no more than about 1.6 milliliters of serum, no more than about 1.7 milliliters of serum, no more than about 1.8 milliliters of serum, no more than about 1.9 milliliters of serum, or no more than about 2.0

milliliters of serum.

Circulating nullomers include nullomers in cells, extracellular nullomers in microvesicles, in exosomes and extracellular nullomers that are not associated with cells or microvesicles (extracellular, non-vesicular nullomers). In some embodiments, the biological sample used for determining the level of one or more nullomers (e.g., a sample containing circulating nullomers) may contain cells. In other embodiments, the biological sample may be free or substantially free of cells (e.g., a serum sample). In some embodiments, a sample containing circulating nullomers, e.g., extracellular nullomers, is a blood-derived sample. Exemplary blood-derived sample types include, e.g., a plasma sample, a serum sample, a blood sample, etc. In other embodiments, a sample containing circulating nullomers is a lymph sample. Circulating nullomers are also found in urine and saliva, and biological samples derived from these sources are likewise suitable for determining the level of one or more disclosed nullomers.

In some embodiments, any of the methods of the disclosure comprises a step of isolating total DNA or RNA from a sample or cell or exosome or microvesicle. Methods of isolating DNA or RNA for expression analysis from blood, plasma and/or serum (see for example, Tsui NB et al. (2002) Clin. Chem. 48,1647-53, incorporated by reference in its entirety herein) and from urine (see for example, Boom R et al. (1990) J Clin Microbiol. 28, 495-503, incorporated by reference in its entirety herein) have been described and routinely used by the skilled person.

Determining the Level of Nullomers in a Sample

The level of one or more disclosed nullomers in a biological sample can be determined by any suitable method. Any reliable method for measuring the level or amount of a nullomer in a sample can be used. Generally, nullomers can be detected and quantified from a sample (including fractions thereof), such as samples of isolated DNA or RNA by various methods known for DNA or mRNA, including, for example, amplification-based methods (e.g., Polymerase Chain Reaction (PCR), Real-Time Polymerase Chain Reaction (RT-PCR), Quantitative Polymerase Chain Reaction (qPCR), rolling circle amplification, etc.), hybridization-based methods (e.g., hybridization arrays (e.g., microarrays), NanoString analysis, Northern Blot analysis, branched DNA (bDNA) signal amplification, in situ hybridization, etc.), and sequencing-based methods (e.g., next-generation sequencing methods, for example, using the Illumina or IonTorrent platforms). Other exemplary techniques include ribonuclease protection assay (RPA) and mass

spectroscopy.

In some embodiments where RNA is used as samples, RNA is converted to DNA (cDNA) prior to analysis. cDNA can be generated by reverse transcription of isolated RNA using conventional techniques. In some embodiments, nullomer is amplified prior to measurement. In other embodiments, the level of nullomer is measured during the amplification process. In still other embodiments, the level of nullomer is not amplified prior to measurement. Some exemplary methods suitable for determining the level of nullomer in a sample are described in greater detail below. These methods are provided by way of illustration only, and it will be apparent to a skilled person that other suitable methods may likewise be used.

A. Amplification-Based Methods

Many amplification-based methods exist for detecting the level of nullomers, including, but not limited to, PCR, RT-PCR, qPCR, and rolling circle amplification. Other amplification-based techniques include, for example, ligase chain reaction, multiplex ligatable probe amplification, in vitro transcription (IVT), strand displacement amplification, transcription-mediated amplification, RNA (Eberwine) amplification, and other methods that are known to persons skilled in the art.

A typical PCR reaction includes multiple steps, or cycles, that selectively amplify target nucleic acid species: a denaturing step, in which a target nucleic acid is denatured; an annealing step, in which a set of PCR primers (i.e., forward and reverse primers) anneal to complementary DNA strands, and an elongation step, in which a thermostable DNA polymerase elongates the primers. By repeating these steps multiple times, a DNA fragment is amplified to produce an amplicon, corresponding to the target sequence. Typical PCR reactions include 20 or more cycles of denaturation, annealing, and elongation. In many cases, the annealing and elongation steps can be performed concurrently, in which case the cycle contains only two steps. A reverse transcription reaction (which produces a cDNA sequence having complementarity to a RNA) may be performed prior to PCR amplification. Reverse transcription reactions include the use of, e.g., a RNA-based DNA polymerase (reverse transcriptase) and a primer.

Kits for quantitative real time PCR of nullomers are known, and are commercially available. Examples of suitable kits include, but are not limited to, the TaqMan mRNA Assay (Applied Biosystems) and the mirVana qRT-PCR nullomer detection kit (Ambion). The RNA can be ligated to a single stranded oligonucleotide containing universal primer sequences, a

polyadenylated sequence, or adaptor sequence prior to reverse transcriptase and amplified using a primer complementary to the universal primer sequence, poly(T) primer, or primer comprising a sequence that is complementary to the adaptor sequence.

In some instances, custom qRT-PCR assays can be developed for determination of nullomer levels. Custom qRT-PCR assays to measure nullomers in a biological sample, e.g., a body fluid, can be developed using, for example, methods that involve an extended reverse transcription primer and locked nucleic acid modified PCR. Custom nullomer assays can be tested by running the assay on a dilution series of chemically synthesized nullomer corresponding to the target sequence. This permits determination of the limit of detection and linear range of quantitation of each assay. Furthermore, when used as a standard curve, these data permit an estimate of the absolute abundance of nullomers measured in biological samples.

Amplification curves may optionally be checked to verify that Ct values are assessed in the linear range of each amplification plot. Typically, the linear range spans several orders of magnitude. For each candidate nullomer assayed, a chemically synthesized version of the nullomer can be obtained and analyzed in a dilution series to determine the limit of sensitivity of the assay, and the linear range of quantitation. Relative expression levels may be determined, for example, as described by Livak et al., Methods (2001) December; 25(4):402-8.

In some embodiments, two or more nullomers are amplified in a single reaction volume. For example, multiplex q-PCR, such as qRT-PCR, enables simultaneous amplification and quantification of at least two nullomers of interest in one reaction volume by using more than one pair of primers and/or more than one probe. The primer pairs comprise at least one amplification primer that specifically binds each nullomer, and the probes are labeled such that they are distinguishable from one another, thus allowing simultaneous quantification of multiple nullomers.

Rolling circle amplification is a DNA-polymerase driven reaction that can replicate circularized oligonucleotide probes with either linear or geometric kinetics under isothermal conditions (see, for example, Lizardi et al., Nat. Gen. (1998) 19(3):225-232; Gusev et al., Am. J. Pathol. (2001) 159(1):63-69; Nallur et al., Nucleic Acids Res. (2001) 29(23):E118). In the presence of two primers, one hybridizing to the (+) strand of DNA, and the other hybridizing to the (-) strand, a complex pattern of strand displacement results in the generation of over 10^9 copies of each DNA molecule in 90 minutes or less. Tandemly linked copies of a closed circle DNA molecule may be formed by using a single primer. The process can also be performed using a

matrix-associated DNA. The template used for rolling circle amplification may be reverse transcribed. This method can be used as a highly sensitive indicator of nullomer sequence and expression level at very low nullomer concentrations (see, for example, Cheng et al., Angew Chem. Int. Ed. Engl. (2009) 48(18):3268-72; Neubacher et al., Chembiochem. (2009) 10(8):1289-91).

In some embodiments, the disclosure provide a method for identifying the presence, absence, or quantity of one or a plurality of the disclosed nullomers comprising: a) isolating nucleic acids from a sample; and b) mixing the nucleic acids with one or a plurality of primers under conditions and for a period of time sufficient to allow amplification of the one or plurality nullomers, wherein the one or plurality of primers comprises sequences that are complementary to any of the nullomers provided in Table 1. In some embodiments, the nucleic acid from a sample is cell-free (cfDNA). In some embodiments, the nucleic acid from a sample is circulating tumor (ctDNA). In some embodiments, the primer used in the disclosed method comprises from about 6 to about 16 nucleotides. In some embodiments, the primer used in the disclosed method comprises from about 7 to about 15 nucleotides. In some embodiments, the primer used in the disclosed method comprises from about 8 to about 14 nucleotides. In some embodiments, the primer used in the disclosed method comprises about 6 nucleotides. In some embodiments, the primer used in the disclosed method comprises about 7 nucleotides. In some embodiments, the primer used in the disclosed method comprises about 8 nucleotides. In some embodiments, the primer used in the disclosed method comprises about 9 nucleotides. In some embodiments, the primer used in the disclosed method comprises about 10 nucleotides. In some embodiments, the primer used in the disclosed method comprises about 11 nucleotides. In some embodiments, the primer used in the disclosed method comprises about 12 nucleotides. In some embodiments, the primer used in the disclosed method comprises about 13 nucleotides. In some embodiments, the primer used in the disclosed method comprises about 14 nucleotides. In some embodiments, the primer used in the disclosed method comprises about 15 nucleotides. In some embodiments, the primer used in the disclosed method comprises about 16 nucleotides.

In some embodiments, the identification of the presence or quantity of one or a plurality of the disclosed nullomers is indicative that the subject from which the sample is obtained has the cancer type corresponding to the particular nullomer identified in Table 1.

B. Hybridization-Based Methods

Nullomers may be detected using hybridization-based methods, including but not limited to hybridization arrays (e.g., microarrays), NanoString analysis, Southern Blot analysis, Northern Blot analysis, branched DNA (bDNA) signal amplification, and in situ hybridization.

Microarrays can be used to measure the levels of large numbers of nullomers simultaneously. Microarrays can be fabricated using a variety of technologies, including printing with fine-pointed pins onto glass slides, photolithography using pre-made masks, photolithography using dynamic micromirror devices, ink-jet printing, or electrochemistry on microelectrode arrays. Also useful are microfluidic TaqMan Low-Density Arrays, which are based on an array of microfluidic qRT-PCR reactions, as well as related microfluidic qRT-PCR based methods.

Axon B-4000 scanner and Gene-Pix Pro 4.0 software or other suitable software can be used to scan images. Non-positive spots after background subtraction, and outliers detected by the ESD procedure, are removed. The resulting signal intensity values are normalized to per-chip median values and then used to obtain geometric means and standard errors for each nullomer. Each signal can be transformed to log base 2, and a one-sample t test can be conducted. Independent hybridizations for each sample can be performed on chips with each nullomer spotted multiple times to increase the robustness of the data.

Microarrays can be used for the expression profiling of nullomers in diseases. For example, DNA or RNA can be extracted from a sample and, optionally, the nullomers are size-selected from total DNA or RNA. Oligonucleotide linkers can be attached to the 5' and 3' ends of the nullomers and the resulting ligation products are used as templates for an RT-PCR reaction. The sense strand PCR primer can have a fluorophore attached to its 5' end, thereby labeling the sense strand of the PCR product. The PCR product is denatured and then hybridized to the microarray. A PCR product, referred to as the target nucleic acid that is complementary to the corresponding nullomer capture probe sequence on the array will hybridize, via base pairing, to the spot at which the, capture probes are affixed. The spot will then fluoresce when excited using a microarray laser scanner. In some embodiments, probes of the disclosure are nucleic acid sequences comprising from about 10 to about 20 nucleotides in length and are DNA or RNA or NDA/RNA hybrid seqeunces complementary to a nullomer of Table 1, Table 5, Table 6 or Table B. In some embodiments, the disclosure relate to composition comprising one or a plurality f such probes. And in some embodiments, those probes comprise a fluorescent probe detectable when exposed to light emitted onto the probe.

The fluorescence intensity of each spot is then evaluated in terms of the number of copies of a particular nullomer, using a number of positive and negative controls and array data normalization methods, which will result in assessment of the level of expression of a particular nullomer.

Total RNA containing the nullomers extracted from a body fluid sample can also be used directly without size-selection of the nullomers. For example, the RNA can be 3' end labeled using T4 RNA ligase and a fluorophore-labeled short RNA linker. Fluorophore-labeled nullomers complementary to the corresponding nullomer capture probe sequences on the array hybridize, via base pairing, to the spot at which the capture probes are affixed. The fluorescence intensity of each spot is then evaluated in terms of the number of copies of a particular nullomer, using a number of positive and negative controls and array data normalization methods, which will result in assessment of the level of expression of a particular nullomer.

Several types of microarrays can be employed including, but not limited to, spotted oligonucleotide microarrays, pre-fabricated oligonucleotide microarrays or spotted long oligonucleotide arrays.

Nullomers can also be detected without amplification using the nCounter Analysis System (NanoString Technologies, Seattle, Wash.). This technology employs two nucleic acid-based probes that hybridize in solution (e.g., a reporter probe and a capture probe). After hybridization to a nullomers disclosed herein, excess probes are removed, and probe/target complexes are analyzed in accordance with the manufacturer's protocol. nCounter nullomer assay kits are available from NanoString Technologies, which are capable of distinguishing between highly similar nullomers with great specificity.

Nullomers can also be detected using branched DNA (bDNA) signal amplification (see, for example, Urdea, Nature Biotechnology (1994), 12:926-928). RNA assays based on bDNA signal amplification are commercially available. One such assay is the QuantiGene.RTM. 2.0 nullomer Assay (Affymetrix, Santa Clara, Calif.). Southern Blot, Northern Blot and in situ hybridization may also be used to detect nullomers. Suitable methods for performing Southern Blot, Northern Blot and in situ hybridization are known in the art.

In some embodiments, biomarker expression is determined by an assay known to those of skill in the art, including but not limited to, multi-analyte profile test, enzyme-linked immunosorbent assay (ELISA), radioimmunoassay, Western blot assay, immunofluorescent assay,

enzyme immunoassay, immunoprecipitation assay, chemiluminescent assay, immunohistochemical assay, dot blot assay, or slot blot assay. In some embodiments, wherein an antibody is used in the assay the antibody is detectably labeled. The antibody labels may include, but are not limited to, immunofluorescent label, chemiluminescent label, phosphorescent label, enzyme label, radiolabel, avidin/biotin, colloidal gold particles, colored particles, and magnetic particles. In some embodiments, biomarker expression is determined by an IHC assay.

In some embodiments, biomarker expression is determined using an agent that specifically binds the biomarker. Any molecular entity that displays specific binding to a biomarker can be employed to determine the level of that biomarker protein in a sample. Specific binding agents include, but are not limited to, antibodies, antibody fragments, antibody mimetics, and polynucleotides (e.g., aptamers). One of skill understands that the degree of specificity required is determined by the particular assay used to detect the biomarker protein. In some embodiments, the disclosure relates to a system comprising a solid support (such as an ELISA plate, gel, bead or column comprising an antibody, antibody fragment, antibody mimetic, and/or polynucleotides capable of binding to T3p or a salt thereof.

C. Sequencing-Based Methods

Advanced sequencing methods can likewise be used as available. For example, nullomers can be detected using Illumina. Next Generation Sequencing (e.g., Sequencing-By-Synthesis or TruSeq methods, using, for example, the HiSeq, HiScan, GenomeAnalyzer, or MiSeq systems (Illumina, Inc., San Diego, Calif.)). Nullomers can also be detected using Ion Torrent Sequencing (Ion Torrent Systems, Inc., Gulliford, Conn.), or other suitable methods of semiconductor sequencing.

D. Additional Nullomers Detection Tools

Mass spectroscopy can be used to quantify nullomers using RNase mapping. Isolated RNAs can be enzymatically digested with RNA endonucleases (RNases) having high specificity (e.g., RNase Tl, which cleaves at the 3'-side of all unmodified guanosine residues) prior to their analysis by MS or tandem MS (MS/MS) approaches. The first approach developed utilized the on-line chromatographic separation of endonuclease digests by reversed phase HPLC coupled directly to ESI-MS. The presence of posttranscriptional modifications can be revealed by mass shifts from those expected based upon the RNA sequence. Ions of anomalous mass/charge values can then be isolated for tandem MS sequencing to locate the sequence placement of the

posttranscriptionally modified nucleoside.

Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) has also been used as an analytical approach for obtaining information about posttranscriptionally modified nucleosides. MALDI-based approaches can be differentiated from ESI-based approaches by the separation step. In MALDI-MS, the mass spectrometer is used to separate the nullomers.

To analyze a limited quantity of intact nullomers, a system of capillary LC coupled with nanoESI-MS can be employed, by using a linear ion trap-orbitrap hybrid mass spectrometer (LTQ Orbitrap XL, Thermo Fisher Scientific) or a tandem-quadrupole time-of-flight mass spectrometer (QSTAR XL, Applied Biosystems) equipped with a custom-made nanospray ion source, a Nanovolume Valve (Valco Instruments), and a splitless nano HPLC system (DiNa, KYA Technologies). Analyte/TEAA is loaded onto a nano-LC trap column, desalted, and then concentrated. Intact nullomers are eluted from the trap column and directly injected into a Cl 8 capillary column, and chromatographed by RP-HPLC using a gradient of solvents of increasing polarity. The chromatographic eluent is sprayed from a sprayer tip attached to the capillary column, using an ionization voltage that allows ions to be scanned in the negative polarity mode.

Additional methods for nullomer detection and measurement include, for example, strand invasion assay (Third Wave Technologies, Inc.), surface plasmon resonance (SPR), cDNA, MTDNA (metallic DNA; Advance Technologies, Saskatoon, SK), and single-molecule methods such as the one developed by US Genomics. Multiple nullomers can be detected in a microarray format using a novel approach that combines a surface enzyme reaction with nanoparticle-amplified SPR imaging (SPRI). The surface reaction of poly(A) polymerase creates poly(A) tails on nullomers hybridized onto locked nucleic acid (LNA) microarrays. DNA-modified nanoparticles are then adsorbed onto the poly(A) tails and detected with SPRI. This ultrasensitive nanoparticle-amplified SPRI methodology can be used for nullomers profiling at attomole levels. IN some embodiments, CRISPR-Cas9 complexes can be used to detect the presence of nullomers in vitro based upon exposure of a sample from a patient to sgRNA-Cas protein complex, wherein the sgRNA is complementary to at least a portion of the nullomer sequence. In some embodiments, the exposure is to genomic DNA within a cancer cell.

In some embodiments, the disclosure relates to a composition or system comprising one or a plurality of sgRNAs that comprise about 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% to the sequences of Table 6 or Table B. In some embodiments, the

disclosure relates to a composition or system comprising one or a plurality of sgRNAs that comprise from about 98 to about 110 nucleotides in length with at least one portion of the sgRNA complementary to a nucleic sequence from about 8 to about 18 nucleotides of any nullomer disclosed in Table 1, Table 5, Table 6 or Table B.

As used herein, the term "mutagen" means any molecule, a nucleic acid sequence, amino acid sequence, or hybrid amino acid or nucleic acid sequence that causes a mutation or modification in one or more regions of endogenous nucleic acid when exposed for a time period sufficient to cause the mutation. In some embodiments, the mutation is a point mutation, frameshift mutation, deletion, truncation, or addition. In some embodiments, the mutagen is a vector or a gene-modifying enzyme.

The term "vector" as used herein refers to any genetic element, such as a plasmid, phage, transposon, cosmid, chromosome, artificial chromosome, virus, virion, etc., which is capable of replication when associated with the proper control elements and which can transfer gene sequences between cells. Thus, the term includes cloning and expression vehicles, as well as viral vectors.

The term "gene-modifying enzyme" as used herein refers to an enzyme that is capable of modifying a gene by introducing a mutation (e.g., point mutation, frameshift mutation, deletion, or truncation) causing gene inactivation or introducing heterologous nucleotides (e.g., genes) through non-homologous end joining or homologous recombination. Exemplary gene-modifying enzymes, include but not limited to, a Cas protein, a meganuclease, a transcription activator-like effector nucleases (TALEN), a transposon, a zinc-finger nuclease (ZFN), or a recombinase. In some embodiments, the gene-modifying enzyme suitable for the methods disclosed herein is a Cas protein, a meganuclease, a TALEN, a ZFN, or a recombinase. In some embodiments, the gene-modifying enzyme suitable for the methods disclosed herein is a Cas protein. In some preferred embodiments, the gene-modifying enzyme suitable for the methods disclosed herein is a Cas9 protein.

The term "Cas9 protein" refers to the "clustered, regularly interspaced, short palindromic repeats (CRISPR)-associated protein 9." This term is well known in the art and has been described, e.g. in Makarova et al. (2011) Nat. Rev. Microbiol., 9:467-477, and in Makarova et al. (2011) Biol. Direct., 6:38. Cas proteins are endonuclease that form part of an adaptive defense mechanism evolved by bacteria and archaea to protect them from invading viruses and plasmids. Cas9 protein

or gene information can be obtained from a known database such as the GenBank of NCBI (National Center for Biotechnology Information), but is not limited thereto. Moreover, the Cas9 protein may comprise not only wild-type Cas9, but also deactivated Cas9 (dCas9), or Cas9 variants such as Cas9 nickase. The deactivated Cas9 may be RFN (RNA-guided FokI nuclease) comprising a FokI nuclease domain bound to dCas9, or may be dCas9 to which a transcription activator or repressor domain is bound. In addition, the Cas9 protein is not limited in its origin. For example, the Cas9 protein may be derived from Streptococcus pyogenes, Francisella novicida, Streptococcus thermophilus, Legionella pneumophila, Listeria innocua, or Streptococcus mutans.

Cas9 protein is the major protein element of the CRISPR/Cas9 system, which forms a complex with crRNA (CRISPR RNA) and tracrRNA (trans-activating crRNA) to form activated endonuclease or nickase. "CRISPR system" refers collectively to transcripts or synthetically produced transcripts and other elements involved in the expression of or directing the activity of CRISPR-associated ("Cas") genes, including sequences encoding a Cas gene, a tracr (trans-activating CRISPR) sequence (e.g. tracrRNA or an active partial tracrRNA), a tracr-mate sequence (encompassing a "direct repeat" and a tracrRNA-processed partial direct repeat in the context of an endogenous CRISPR system), a guide sequence (also referred to as a "spacer" in the context of an endogenous CRISPR system), or other sequences and transcripts from a CRISPR locus. In some embodiments, one or more elements of a CRISPR system is derived from a type I, type II, or type III CRISPR system. In some embodiments, one or more elements of a CRISPR system is derived from a particular organism comprising an endogenous CRISPR system, such as Streptococcus pyogenes. In general, a CRISPR system is characterized by elements that promote the formation of a CRISPR complex at the site of a target sequence (also referred to as a protospacer in the context of an endogenous CRISPR system). In the context of formation of a CRISPR complex, "target sequence" refers to a nucleic acid sequence to which a guide sequence is designed to have complementarity, where hybridization between a target sequence and a guide sequence promotes the formation of a CRISPR complex. Full complementarity is not necessarily required, provided there is sufficient complementarity to cause hybridization and promote formation of a CRISPR complex. A target sequence may comprise any polynucleotide, such as DNA or RNA polynucleotides, but in some embodiments, the target sequence is a nullomer or a region of a nullomer that is from about 10 to about 35 nucleotides of the nullomer sequence of any nullomer from Table 1. In some embodiments, the target sequence is a DNA polynucleotide and

is referred to a DNA target sequence. In some embodiments, a target sequence comprises at least three nucleic acid sequences that are recognized by a Cas-protein when the Cas protein is associated with a CRISPR complex or system which comprises at least one sgRNA or one tracrRNA/crRNA duplex at a concentration and within an microenvironment suitable for association of such a system. In some embodiments, the target DNA comprises at least one or more proto-spacer adjacent motifs which sequences are known in the art and are dependent upon the Cas protein system being used in conjunction with the sgRNA or crRNA/tracrRNAs employed by this work. In some embodiments, the target DNA comprises NNG, where G is a guanine and N is any naturally occurring nucleic acid. In some embodiments the target DNA comprises any one or combination of NNG, NNA, GAA, NNAGAAW and NGGNG, where G is an guanine, A is adenine, and N is any naturally occurring nucleic acid from one nullomer in Table 1.

Typically, in the context of an endogenous CRISPR system, formation of a CRISPR complex (comprising a guide sequence hybridized to a target sequence and complexed with one or more Cas proteins) results in cleavage of one or both strands in or near (e.g. within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, or more base pairs from) the target sequence. without wishing to be bound by theory, the tracr sequence, which may comprise or consist of all or a portion of a wild-type tracr sequence (e.g. about or more than about 20, 26, 32, 45, 48, 54, 63, 67, 85, or more nucleotides of a wild-type tracr sequence), may also form part of a CRISPR complex, such as by hybridization along at least a portion of the tracr sequence to all or a portion of a tracr mate sequence that is operably linked to the guide sequence. In some embodiments, the tracr sequence has sufficient complementarity to a tracr mate sequence to hybridize and participate in formation of a CRISPR complex. As with the target sequence, it is believed that complete complementarity is not needed, provided there is sufficient to be functional (bind the Cas protein or functional fragment thereof). In some embodiments, the tracr sequence has at least 50%, 60%, 70%, 80%, 90%, 95% or 99% of sequence complementarity along the length of the tracr mate sequence when optimally aligned. In some embodiments, one or more vectors driving expression of one or more elements of a CRISPR system are introduced into a host cell such that the presence and/or expression of the elements of the CRISPR system direct formation of a CRISPR complex at one or more target sites. For example, a Cas enzyme, a guide sequence linked to a tracr-mate sequence, and a tracr sequence could each be operably linked to separate regulatory elements on separate vectors. Alternatively, two or more of the elements expressed from the same or different regulatory elements, may be

combined in a single vector, with one or more additional vectors providing any components of the CRISPR system not included in the first vector. In some embodiments, the target site is a genomic DNA of a cancer cell within the host or a cancer cell isolated from the subject in a sample or within a system independent of a tumor.

With at least some of the modification contemplated by this disclosure, in some embodiments, the guide sequence or RNA or DNA sequences that form a CRISPR complex are at least partially synthetic. The CRISPR system elements that are combined in a single vector may be arranged in any suitable orientation, such as one element located 5' with respect to ("upstream" of) or 3' with respect to ("downstream" of) a second element. In some embodiments, the disclosure relates to a composition comprising a chemically synthesized guide sequence. In some embodiments, the chemically synthesized guide sequence is used in conjunction with a vector comprising a coding sequence that encodes a CRISPR enzyme, such as a type II Cas9 protein. In some embodiments, the chemically synthesized guide sequence is used in conjunction with one or more vectors, wherein each vector comprises a coding sequence that encodes a CRISPR enzyme, such as a type II Cas9 protein. The coding sequence of one element may be located on the same or opposite strand of the coding sequence of a second element, and oriented in the same or opposite direction. In some embodiments, a single promoter drives expression of a transcript encoding a CRISPR enzyme and one or more additional (second, third, fourth, etc.) guide sequences, tracr mate sequence (optionally operably linked to the guide sequence), and a tracr sequence embedded within one or more intron sequences (e.g. each in a different intron, two or more in at least one intron, or all in a single intron). In some embodiments, the CRISPR enzyme, one or more additional guide sequence, tracr mate sequence, and tracr sequence are each a component of different nucleic acid sequences. For instance, in the case of a tracr and tracr mate sequences and in some embodiments, the disclosure relates to a composition comprising at least a first and second nucleic acid sequence, wherein the first nucleic acid sequence comprises a tracr sequence and the second nucleic acid sequence comprises a tracr mate sequence, wherein the first nucleic acid sequence is at least partially complementary to the second nucleic acid sequence such that the first and second nucleic acid for a duplex and wherein the first nucleic acid and the second nucleic acid either individually or collectively comprise a DNA-targeting domain, a Cas protein binding domain, and a transcription terminator domain. In some embodiments, the CRISPR enzyme, one or more additional guide sequence, tracr mate sequence, and tracr sequence are operably linked to

and expressed from the same promoter. In some embodiments, the disclosure relates to compositions comprising any one or combination of the disclosed domains on one guide sequence or two separate tracrRNA/crRNA sequences with or without any of the disclosed modifications. Any methods disclosed herein also relate to the use of tracrRNA/crRNA sequence interchangeably with the use of a guide sequence, such that a composition may comprise a single synthetic guide sequence and/or a synthetic tracrRNA/crRNA with any one or combination of modified domains disclosed herein.

The CRISPR system suitable for the present disclosure can also comprise a modified CRISPR enzyme (or "Cas protein") or a nucleotide sequence encoding one or more Cas proteins. Any protein capable of enzymatic activity in cooperation with a guide sequence is a Cas protein. In some embodiments, the disclosure relates to a system comprises a vector comprising a regulatory element operably linked to an enzyme-coding sequence encoding a CRISPR enzyme, such as a Cas protein from the Cas family of enzymes. In some embodiments, the disclosure relates to a system, composition, or pharmaceutical composition comprising any one or plurality of Cas proteins either individually or in combination with one or a plurality of guide sequences. Compositions of one or a plurality of Cas proteins may be administered to a subject with any of the disclosed guide sequences sequentially or contemporaneously. Non-limiting examples of Cas proteins include Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (also known as Csn1 and Csx12), Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, type V CRISPR-Cas systems (e.g., Cas12), and Type VI CRISPR-Cas systems (e.g., Cas13), and variants and fragments thereof, or modified versions thereof having at least 70% sequence identity to any of the above Cas proteins. These enzymes are known, for example, the amino acid sequence of *S. pyogenes* Cas9 protein may be found in the SwissProt database under accession number Q99ZW2. In some embodiments, the unmodified CRISPR enzyme has DNA cleavage activity, such as Cas9. In some embodiments the CRISPR enzyme is Cas9, and may be Cas9 from *S. pyogenes* or *S. pneumoniae*. In some embodiments, the CRISPR enzyme directs cleavage of one or both strands at the location of a target sequence, such as within the target sequence and/or within the complement of the target sequence. In some embodiments, the CRISPR enzyme directs cleavage of one or both strands within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, 200, 500, or more base pairs from the

first or last nucleotide of a target sequence. In some embodiments, a vector encodes a CRISPR enzyme or Cas protein that is mutated to with respect to a corresponding wild-type enzyme such that the mutated CRISPR enzyme lacks the ability to cleave one or both strands of a target polynucleotide containing a target sequence. For example, an aspartate-to-alanine substitution (D10A) in the RuvC I catalytic domain of Cas9 from S. pyogenes converts Cas9 from a nuclease that cleaves both strands to a nickase (cleaves a single strand). Other examples of mutations that render Cas9 a nickase include, without limitation, H840A, N854A, and N863A. In some embodiments, a Cas9 nickase may be used in combination with guide sequence(s), e.g., two guide sequences, which target respectively sense and antisense strands of the DNA target. This combination allows both strands to be nicked and used to induce NHEJ.

As a further example, two or more catalytic domains of Cas9 (RuvC I, RuvC II, and RuvC III) may be mutated to produce a mutated Cas9 substantially lacking all DNA cleavage activity. In some embodiments, a D10A mutation is combined with one or more of H840A, N854A, or N863A mutations to produce a Cas9 enzyme substantially lacking all DNA cleavage activity. In some embodiments, a CRISPR enzyme is considered to substantially lack all DNA cleavage activity when the DNA cleavage activity of the mutated enzyme is less than about 25%, 10%, 5%, 1%, 0.1%, 0.01%, or lower with respect to its non-mutated form. Other mutations may be useful; where the Cas9 or other CRISPR enzyme is from a species other than S. pyogenes, mutations in corresponding amino acids may be made to achieve similar effects.

The disclosure relates to a method of detecting the presence of a nullomer by exposing a Cas protein and sgRNA specific to a target nullomer sequence to a nullomer target sequence. In some embodiments, the nullomer target sequence is any nullomer from Table 1 and the sgRNA sequence specific for the nullomer is any RNA molecule that comprises from about 10 to about 35 nucleotides complementary to a nullomer in Table 1. In some embodiments, the method further comprises allowing a time period sufficient for the sgRNA to associate with the nullomer and the Cas protein to excise the nullomer from the genomic DNA of a host cell or cell within a sample. Detection of the nullomer can further comprise identifying the nullomer sequence excised from the cell by amplification through PCR or a non-amplification event such as those disclosed herein.

In certain embodiments, labels, dyes, or labeled probes and/or primers are used to detect amplified or unamplified nullomers. The skilled artisan will recognize which detection methods are appropriate based on the sensitivity of the detection method and the abundance of the target.

Depending on the sensitivity of the detection method and the abundance of the target, amplification may or may not be required prior to detection. One skilled in the art will recognize the detection methods where nullomer amplification is preferred.

A probe or primer may include standard (A, T or U, G and C) bases, or modified bases. Modified bases include, but are not limited to, the AEGIS bases (from Eragen Biosciences), which have been described, e.g., in U.S. Pat. Nos. 5,432,272, 5,965,364, and 6,001,983. In certain aspects, bases are joined by a natural phosphodiester bond or a different chemical linkage. Different chemical linkages include, but are not limited to, a peptide bond or a Locked Nucleic Acid (LNA) linkage, which is described, e.g., in U.S. Pat. No. 7,060,809.

In a further aspect, oligonucleotide probes or primers present in an amplification reaction are suitable for monitoring the amount of amplification product produced as a function of time. In certain aspects, probes having different single stranded versus double stranded character are used to detect the nucleic acid. Probes include, but are not limited to, the 5'-exonuclease assay (e.g., TAQMAN) probes (see U.S. Pat. No. 5,538,848), stem-loop molecular beacons (see, e.g., U.S. Pat. Nos. 6,103,476 and 5,925,517), stemless or linear beacons (see, e.g., WO 9921881, U.S. Pat. Nos. 6,485,901 and 6,649,349), peptide nucleic acid (PNA) Molecular Beacons (see, e.g., U.S. Pat. Nos. 6,355,421 and 6,593,091), linear PNA beacons (see, e.g. U.S. Pat. No. 6,329,144), non-FRET probes (see, e.g., U.S. Pat. No. 6,150,097), Sunrise.TM./AmplifluorB.TM. probes (see, e.g., U.S. Pat. No. 6,548,250), stem-loop and duplex SCORPION probes (see, e.g., U.S. Pat. No. 6,589,743), bulge loop probes (see, e.g., U.S. Pat. No. 6,590,091), pseudo knot probes (see, e.g., U.S. Pat. No. 6,548,250), cyclicons (see, e.g., U.S. Pat. No. 6,383,752), MGB Eclipse™ probe (Epoch Biosciences), hairpin probes (see, e.g., U.S. Pat. No. 6,596,490), PNA light-up probes, antiprimer quench probes (Li et al., Clin. Chem. 53:624-633 (2006)), self-assembled nanoparticle probes, and ferrocene-modified probes described, for example, in U.S. Pat. No. 6,485,901.

In certain embodiments, one or more of the primers in an amplification reaction can include a label. In yet further embodiments, different probes or primers comprise detectable labels that are distinguishable from one another. In some embodiments, a nucleic acid, such as the probe or primer, may be labeled with two or more distinguishable labels.

In some aspects, a label is attached to one or more probes and has one or more of the following properties: (i) provides a detectable signal; (ii) interacts with a second label to modify the detectable signal provided by the second label, e.g., FRET (Fluorescent Resonance Energy

Transfer); (iii) stabilizes hybridization, e.g., duplex formation; and (iv) provides a member of a binding complex or affinity set, e.g., affinity, antibody-antigen, ionic complexes, hapten-ligand (e.g., biotin-avidin). In still other aspects, use of labels can be accomplished using any one of a large number of known techniques employing known labels, linkages, linking groups, reagents, reaction conditions, and analysis and purification methods.

Nullomers can be detected by direct or indirect methods. In a direct detection method, one or more nullomers are detected by a detectable label that is linked to a nucleic acid molecule. In such methods, the nullomers may be labeled prior to binding to the probe. Therefore, binding is detected by screening for the labeled nullomer that is bound to the probe. The probe is optionally linked to a bead in the reaction volume.

In certain embodiments, nucleic acids are detected by direct binding with a labeled probe, and the probe is subsequently detected. In some embodiments, the nucleic acids, such as amplified nullomers, are detected using FIexMAP Microspheres (Luminex) conjugated with probes to capture the desired nucleic acids. Some methods may involve detection with polynucleotide probes modified with fluorescent labels or branched DNA (bDNA) detection, for example.

In some embodiments, biomarker expression is determined using a PCR-based assay comprising specific primers and/or probes for each biomarker. As used herein, the term "probe" refers to any molecule that is capable of selectively binding a specifically intended target biomolecule. In some embodiments, as used herein, the term "probe" refers to any molecule that may bind or associate, indirectly or directly, covalently or non-covalently, to any of the substrates and/or reaction products and/or proteases disclosed herein and whose association or binding is detectable using the methods disclosed herein. In some embodiments, the term "probe" refers to any molecule comprising a nucleic acid sequence that is complementary to any of the nucleic acid sequences disclosed in TABLE 1 or one comprising at least about 70%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% sequence identity to any of the nucleic acid sequences disclosed in TABLE 1. In some embodiments, the term "probe" refers to any molecule comprising a nucleic acid sequence that is complementary to a fragment of any of the nucleic acid sequences disclosed in TABLE 1 or one comprising at least about 70%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% sequence identity to a fragment of any of the nucleic acid sequences disclosed in TABLE 1. In some embodiments, the term "probe" refers to

a sgRNA molecule comprising a nucleic acid sequence that is complementary to a fragment of any of the nucleic acid sequences disclosed in TABLE 1 or one comprising at least about 70%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% sequence identity to a fragment of any of the nucleic acid sequences disclosed in TABLE 1. In some embodiments, the probe is a fluorogenic probe, antibody or absorbance-based probes. If an absorbance-based probe, the chromophore pNA (para-nitroanaline) may be used as a probe for detection and/or quantification of a target nucleic acid sequence disclosed herein. In some embodiments, the probe may comprise a nucleic acid sequence labeled with a fluorogenic molecule or a substrate that when exposed to an enzyme becomes fluorogenic and the nucleic acid sequence is complementary to any of the nucleic acid sequences disclosed in TABLE 1 or one comprising at least about 70%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% sequence identity to any of the nucleic acid sequences disclosed in TABLE 1 or Table B. Probes can be synthesized by one of skill in the art using known techniques, or derived from biological preparations. Probes may include but are not limited to, RNA, DNA, proteins, peptides, aptamers, antibodies, and organic molecules. The term "primer" or "probe" encompasses oligonucleotides that have a specific sequence or oligoribonucleotides that have a specific sequence. In some embodiments, the probe are from about 5 to about 20 nucleotides in length and are complementary to the nucleic acid sequences in TABLE 1 or in Table B and comprise at least about 70%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or about 100% sequence identity to any one or combination of nucleic acid sequences complementary to those provided in TABLE 1 or Table B. In some embodiments, the probe are from about 5 to about 20 nucleotides in length and are complementary to the nucleic acid sequences in TABLE 1 and comprise at least about 70%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or about 100% sequence identity to any one or combination of nucleic acid sequences complementary to those provided in TABLE 7. In some embodiments, the probe are from about 5 to about 20 nucleotides in length and are complementary to the nucleic acid sequences in TABLE 1 and comprise at least about 70%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or about 100% sequence identity to any one or combination of nucleic acid sequences complementary to those provided in TABLE 8.

The target molecule could be any one or a combination of nucleic acid sequences identified in TABLE 1. In some embodiments, the target molecule is a nucleic acid sequence comprising at least about 70%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or about 99% sequence identity to any one or combination of nucleic acid sequences provided in TABLE 1. In some embodiments, the target molecule is any amplified fragment of any one or combination of nucleic acid sequences identified in TABLE 1, and/or any one or combination of nucleic acid sequence comprising at least about 70%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or about 99% sequence identity to any one or combination of nucleic acid sequences in TABLE 1.

In other embodiments, nucleic acids are detected by indirect detection methods. For example, a biotinylated probe may be combined with a streptavidin-conjugated dye to detect the bound nucleic acid. The streptavidin molecule binds a biotin label on amplified nullomer, and the bound nullomer is detected by detecting the dye molecule attached to the streptavidin molecule. In some embodiments, the streptavidin-conjugated dye molecule comprises PHYCOLINK. Streptavidin R-Phycoerythrin (PROzyme). Other conjugated dye molecules are known to persons skilled in the art.

Labels include, but are not limited to, light-emitting, light-scattering, and light-absorbing compounds which generate or quench a detectable fluorescent, chemiluminescent, or bioluminescent signal (see, e.g., Kricka, L., Nonisotopic DNA Probe Techniques, Academic Press, San Diego (1992) and Garman A., Non-Radioactive Labeling, Academic Press (1997).). A dual labeled fluorescent probe that includes a reporter fluorophore and a quencher fluorophore is used in some embodiments. It will be appreciated that pairs of fluorophores are chosen that have distinct emission spectra so that they can be easily distinguished.

In certain embodiments, labels are hybridization-stabilizing moieties which serve to enhance, stabilize, or influence hybridization of duplexes, e.g., intercalators and intercalating dyes (including, but not limited to, ethidium bromide and SYBR-Green), minor-groove binders, and cross-linking functional groups (see, e.g., Blackburn et al., eds. “DNA and RNA Structure” in Nucleic Acids in Chemistry and Biology (1996)).

In other embodiments, methods relying on hybridization and/or ligation to quantify nullomers may be used, including oligonucleotide ligation (OLA) methods and methods that allow a distinguishable probe that hybridizes to the target nucleic acid sequence to be separated from an

unbound probe. As an example, HARP-like probes, as disclosed in U.S. Publication No. 2006/0078894 may be used to measure the quantity of nullomers. In such methods, after hybridization between a probe and the targeted nucleic acid, the probe is modified to distinguish the hybridized probe from the unhybridized probe. Thereafter, the probe may be amplified and/or detected. In general, a probe inactivation region comprises a subset of nucleotides within the target hybridization region of the probe. To reduce or prevent amplification or detection of a HARP probe that is not hybridized to its target nucleic acid, and thus allow detection of the target nucleic acid, a post-hybridization probe inactivation step is carried out using an agent which is able to distinguish between a HARP probe that is hybridized to its targeted nucleic acid sequence and the corresponding unhybridized HARP probe. The agent is able to inactivate or modify the unhybridized HARP probe such that it cannot be amplified. A probe ligation reaction may also be used to quantify nullomers. In a Multiplex Ligation-dependent Probe Amplification (MLPA) technique (Schouten et al., Nucleic Acids Research 30:e57 (2002)), pairs of probes which hybridize immediately adjacent to each other on the target nucleic acid are ligated to each other driven by the presence of the target nucleic acid. In some aspects, MLPA probes have flanking PCR primer binding sites. MLPA probes are specifically amplified when ligated, thus allowing for detection and quantification of nullomer biomarkers.

Detecting a Level of Nullomers

The nullomers described herein can be used individually or in combination in diagnostic tests to assess the type of cancer, tissue of origin, and status or stage of the cancer in a subject. Cancer status or stage includes the presence or absence of the cancer. Cancer status or stage may also include monitoring the course of the cancer, for example, monitoring disease progression. Based on the cancer status or stage of a subject, additional procedures may be indicated, including, for example, additional diagnostic tests or therapeutic procedures.

The power of a diagnostic test to correctly predict disease status is commonly measured in terms of the accuracy of the assay, the sensitivity of the assay, the specificity of the assay, or the "Area Under a Curve" (AUC), for example, the area under a Receiver Operating Characteristic (ROC) curve. As used herein, accuracy is a measure of the fraction of misclassified samples. Accuracy may be calculated as the total number of correctly classified samples divided by the total number of samples, e.g., in a test population. Sensitivity is a measure of the "true positives" that are predicted by a test to be positive, and may be calculated as the number of correctly identified

cancer samples divided by the total number of cancer samples. Specificity is a measure of the "true negatives" that are predicted by a test to be negative, and may be calculated as the number of correctly identified normal samples divided by the total number of normal samples. AUC is a measure of the area under a Receiver Operating Characteristic curve, which is a plot of sensitivity vs. the false positive rate (1-specificity). The greater the AUC, the more powerful the predictive value of the test. Other useful measures of the utility of a test include the "positive predictive value," which is the percentage of actual positives who test as positives, and the "negative predictive value," which is the percentage of actual negatives who test as negatives. In some embodiments, the level of one or more nullomers in samples obtained from subjects having different cancer statuses show a statistically significant difference of at least about 0.05 ($p = 0.05$) relative to normal subjects, as determined relative to a suitable control. In some embodiments, the level of one or more nullomers in samples obtained from subjects having different cancer statuses show a statistically significant difference of at least about 0.01 ($p = 0.01$) relative to normal subjects, as determined relative to a suitable control. In some embodiments, the level of one or more nullomers in samples obtained from subjects having different cancer statuses show a statistically significant difference of at least about 0.005 ($p = 0.005$) relative to normal subjects, as determined relative to a suitable control. In some embodiments, the level of one or more nullomers in samples obtained from subjects having different cancer statuses show a statistically significant difference of at least about 0.001 ($p = 0.001$) relative to normal subjects, as determined relative to a suitable control.

In other embodiments, diagnostic tests that use nullomers described herein individually or in combination show an accuracy of at least about 75%, e.g., an accuracy of at least about 75%, about 80%, about 85%, about 90%, about 95%, about 97%, about 99% or about 100%. In other embodiments, diagnostic tests that use nullomers described herein individually or in combination show a specificity of at least about 75%, e.g., a specificity of at least about 75%, about 80%, about 85%, about 90%, about 95%, about 97%, about 99% or about 100%. In other embodiments, diagnostic tests that use nullomers described herein individually or in combination show a sensitivity of at least about 75%, e.g., a sensitivity of at least about 75%, about 80%, about 85%, about 90%, about 95%, about 97%, about 99% or about 100%. In other embodiments, diagnostic tests that use nullomers described herein individually or in combination show a specificity and sensitivity of at least about 75% each, e.g., a specificity and sensitivity of at least about 75%, about

80%, about 85%, about 90%, about 95%, about 97%, about 99% or about 100% (for example, a specificity of at least about 80% and sensitivity of at least about 80%, or for example, a specificity of at least about 80% and sensitivity of at least about 95%).

Each nullomer listed in TABLE 1 is identified as being associated with certain type(s) of cancer as provided. In some instances, one particular nullomer may be associated with more than one types of cancers. In other instances, one particular nullomer may be associated with only one type of cancer.

Each nullomer listed in TABLE 1 is differentially present in biological samples derived from subjects having certain types of cancers as compared with normal subjects, and thus each is individually useful in facilitating the determination of those types of cancer in a test subject. Such a method involves determining the level of the nullomer in a sample obtained from the subject. Determining the level of the nullomer in a sample may include measuring, detecting, or assaying the level of the nullomer in the sample using any suitable method, for example, the methods set forth herein. Determining the level of the nullomer in a sample may also include examining the results of an assay that measured, detected, or assayed the level of the nullomer in the sample. The method may also involve comparing the level of the nullomer in a sample with a suitable control. A change in the level of the nullomer relative to that in a normal subject as assessed using a suitable control is indicative of the cancer status or stage of the subject. A diagnostic amount of a nullomer that represents an amount of the nullomer above or below which a subject is classified as having a particular cancer status or stage can be used. For example, if the nullomer is upregulated in samples from an individual having cancer as compared to a normal individual, a measured amount above the diagnostic cutoff provides a diagnosis of the type of cancer that individual has. Generally, the nullomers in TABLE 1 and Table 7 are upregulated in cancer samples relative to samples obtained from normal individuals. As is well-understood in the art, adjusting the particular diagnostic cut-off used in an assay allows one to adjust the sensitivity and/or specificity of the diagnostic assay as desired. The particular diagnostic cut-off can be determined, for example, by measuring the amount of the nullomer in a statistically significant number of samples from subjects with different cancer statuses, and drawing the cut-off at the desired level of accuracy, sensitivity, and/or specificity. In certain embodiments, the diagnostic cut-off can be determined with the assistance of a classification algorithm, as described elsewhere herein.

Accordingly, methods are provided for diagnosing cancer in a subject, by determining the

level of at least one nullomer in a sample from the subject, wherein a difference in the level of the at least one nullomer versus that in a normal subject (as determined relative to a suitable control) is indicative of cancer in the subject. In some embodiments, the at least one nullomer includes one or more nullomers from TABLE 1. In some embodiments, a difference in the level of the at least one nullomer versus that in a normal subject (as determined relative to a suitable control) is indicative of the type(s) of cancer identified as being associated with the detected at least one nullomer in the subject. For example, the disclosed method of determining the level of at least one nullomer in a sample from a subject, wherein an increase in the level of the at least one nullomer relative to a control is indicative of cancer in the subject, particularly of the type(s) of cancer identified as being associated with the at least one nullomer detected. In some embodiments, the subject is diagnosed with having breast cancer, pancreatic cancer, esophagus cancer, lymphoid cancer, kidney cancer, ovary cancer, head and neck cancer, lung cancer, stomach cancer, CNS cancer, uterus cancer, skin cancer, colorectal cancer, prostate cancer, bladder cancer, bone and soft tissue cancer, biliary cancer, cervix cancer, thyroid cancer, myeloid cancer, or liver cancer by the disclosed method.

Optionally, the method may further comprise providing a diagnosis that the subject has or does not have cancer based on the level of at least one nullomer in the sample. In addition or alternatively, the method may further comprise correlating a difference in the level or levels of at least one nullomer relative to a suitable control with a diagnosis of cancer in the subject. In some embodiments, such a diagnosis may be provided directly to the subject, or it may be provided to another party involved in the subject's care.

While individual nullomers are useful in diagnostic applications for various types of cancer, as shown herein, a combination of nullomers may provide greater predictive value of cancer status or stage than the nullomers when used alone. Specifically, the detection of a plurality of nullomers can increase the accuracy, sensitivity, and/or specificity of a diagnostic test. The detection of a plurality of nullomers can also assist in narrowing down the type of cancer and/or status or stage thereof in a subject. This is particular useful when a given nullomer is identified as being associated with more than one type of cancer. For instance, if nullomer A is identified as being associated with cancers X, Y and Z, nullomer B is identified as being associated with cancers X and Y, and nullomer C is identified as being associated with cancers X and Z, by a process of elimination, a detection of the presence of nullomers A, B and C in a subject is indicative that the

subject has cancer X. The disclosure thus includes the individual nullomer provided in TABLE 1 and nullomer combinations as set forth herein, and their use in methods and kits described herein.

Accordingly, methods are provided for diagnosing cancer in a subject, by determining the level of two or more nullomers in a sample from the subject, wherein a difference in the level of the nullomers versus that in a normal subject (as determined relative to a suitable control) is indicative of cancer in the subject. In some embodiments, the nullomers include one or more of nullomers provided in TABLE1. In some embodiments, the type(s) of cancer thus diagnosed is/are the one(s) provided in TABLE 1 as being associated with each individual nullomer provided in TABLE 1.

Also provided is a method of diagnosing cancer in a subject by determining the levels of two or more nullomers in a sample from the subject, comparing the levels of the two or more nullomers in the sample to a set of data representing levels of the nullomers present in normal subjects and subjects having a particular type of cancer, and diagnosing the subject as having or not having that particular type of cancer based on the comparison. In such a method, the set of data serves as a suitable control or reference standard for comparison with the sample from the subject.

Comparison of the sample from the subject with the set of data may be assisted by a classification algorithm, which computes whether or not a statistically significant difference exists between the collective levels of the two or more nullomers in the sample, and the levels of the same nullomers present in normal subjects or subjects having cancer.

Generation of Classification Algorithms for Qualifying Cancer Type and Status

In some embodiments, data that are generated using samples such as "known samples" can then be used to "train" a classification model. A "known sample" is a sample that has been pre-classified, e.g., classified as being derived from a normal subject or from a subject having a particular type of cancer. The data that are derived from the spectra and are used to form the classification model can be referred to as a "training data set." Once trained, the classification model can recognize patterns in data derived from spectra generated using unknown samples. The classification model can then be used to classify the unknown samples into classes. This can be useful, for example, in predicting whether or not a particular biological sample is associated with a certain biological condition (e.g., diseased versus non-diseased).

In some embodiments, data for the training data set that is used to form the classification

model can be obtained directly from quantitative PCR (for example, Ct values obtained using the double delta Ct method), or from high-throughput expression profiling, such as microarray analysis (for example, total counts or normalized counts from a nullomer or neomer expression assay).

Classification models can be formed using any suitable statistical classification (or "learning") method that attempts to segregate bodies of data into classes based on objective parameters present in the data. Classification methods may be either supervised or unsupervised. Examples of supervised and unsupervised classification processes are described in Jain, "Statistical Pattern Recognition: A Review," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 1, January 2000, the teachings of which are incorporated by reference in its entirety.

In supervised classification, training data containing examples of known categories are presented to a learning mechanism, which learns one or more sets of relationships that define each of the known classes. New data may then be applied to the learning mechanism, which then classifies the new data using the learned relationships. Examples of supervised classification processes include linear regression processes (e.g., multiple linear regression (MLR), partial least squares (PLS) regression and principal components regression (PCR)), binary decision trees (e.g., recursive partitioning processes such as CART – classification and regression trees), artificial neural networks such as back propagation networks, discriminant analyses (e.g., Bayesian classifier or Fischer analysis), logistic classifiers, and support vector classifiers (support vector machines).

In other embodiments, the classification models that are created can be formed using unsupervised learning methods. Unsupervised classification attempts to learn classifications based on similarities in the training data set, without pre-classifying the spectra from which the training data set was derived. Unsupervised learning methods include cluster analyses. A cluster analysis attempts to divide the data into "clusters" or groups that ideally should have members that are very similar to each other, and very dissimilar to members of other clusters. Similarity is then measured using some distance metric, which measures the distance between data items, and clusters together data items that are closer to each other. Clustering techniques include the MacQueen's K-means algorithm and the Kohonen's Self-Organizing Map algorithm. Learning algorithms asserted for use in classifying biological information are described, for example, in PCT International

Publication No. WO 01/31580 (Barnhill et al., "Methods and devices for identifying patterns in biological systems and methods of use thereof"), U.S. application publication No. 2002/0193950 A1 (Gavin et al, "Method or analyzing mass spectra"), U.S. application publication No. 2003/0004402 A1 (Hitt et al., "Process for discriminating between biological states based on hidden patterns from biological data"), and U.S. application publication No. 2003/0055615 A1 (Zhang and Zhang, "Systems and methods for processing biological expression data"). The contents of the foregoing patent applications are incorporated herein by reference in their entireties.

The classification models can be formed on and used on any suitable digital computer. Suitable digital computers include micro, mini, or large computers using any standard or specialized operating system, such as a Unix, WINDOWS or LINUX based operating system.

The training data set(s) and the classification models can be embodied by computer code that is executed or used by a digital computer. The computer code can be stored on any suitable computer readable media including optical or magnetic disks, sticks, tapes, etc., and can be written in any suitable computer programming language including C, C++, visual basic, etc.

The learning algorithms described herein can be used for developing classification algorithms for nullomers or meomers for various types of tumors. The classification algorithms can, in turn, be used in diagnostic tests by providing diagnostic values (e.g., cut-off points) for neomers used singly or in combination. The algorithms can also be used to correlate the presence or absence of a neomer in a sample to a presence of a mutation or presence of a functional error in a particular genetic element of the cell in a subject. In some cases the presence of the neomer can indicate the liklehood of the presence of a mutation at a particular locus within the genome of the cancer cell or the likelihood of the presence of a dysfunction of a particular regulatory element within the genome of the cancer cell. Moreover, such a calculaoin or determination of the likelihood can establish a recommendation of therapy for the subject, such that there is a greater likelihood the subject is responsive to the therapy.

Table C lists the type of cancer associated with the Genes and Neomers identified in Table B.

Table C also lists the types of treatments recommended for the particular cancer types lists.

Therefore methods of the disclosure relate to characterizing a mutation in a gene, a cancer type and a proposed therapy to the neomers of Table B.

Table B:

RELATED GENE	hrom.	pos	ref	alt.	seq	coordinates	list of nullomers
EGFR1	7	55181378	C	T	CCGTGCAGCTCATCATGCAGCTCATGCCCTT	55181363:55181393	AGGGCATGAGCTGCAT,ATGCAGCTCATGCCCT (SEQ ID NO: 1 – SEQ ID NO: 2)
EGFR1	7	55191822	T	G	TCACAGATTTTGGGCGGGCCAAACTGCTGGG	55191807:55191837	CACAGATTTTGGGCGG,CCAGCAGTTTGGCCCG,ACAGATTTTGGGCGGG,CAGCAGTTTGGCCCGC,AGCAGTTTGGCCCGCC,AGATTTTGGGCGGGCC,GCAGTTTGGCCCGCCC,GATTTTGGGCGGGCCA,CAGTTTGGCCCGCCCA,ATTTTGGGCGGGCCAA,AGTTTGGCCCGCCCAA,TTTTGGGCGGGCCAAA,GTTTGGCCCGCCCAAA,TTGGGCGGGCCAAAC,TTTGGCCCGCCCAAAA,TTGGGCGGGCCAAACT,TTGGCCCGCCCAAAAT,TGGGCGGGCCAAACTG,TGGCCCGCCCAAAATC,GGGCGGGCCAAACTGC,GGCCCGCCCAAAATCT,GGCGGGCCAAACTGCT,GCGGGCCAAACTGCTG,CCCGCCCAAAATCTGT,CGGGCCAAACTGCTGG,CCGCCCAAAATCTGTG (SEQ ID NO: 3 – SEQ ID NO: 28)
BRCA2	13	32316529	T	C	CTGCAACAAAGCAGGCATTGACAAATTTTAT	32316514:32316544	GCAACAAAGCAGGCAT,ATGCCTGCTTTGTTGC (SEQ ID NO: 29 – SEQ ID NO: 30)
BRCA2	13	32316528	G	C	GCTGCAACAAAGCAGCTATTGACAAATTTTA	32316513:32316543	AATTTGTCAATAGCTG,ATTTGTCAATAGCTGC,GCAGCTATTGACAAAT,CAGCTATTGACAAATT (SEQ ID NO: 31 – SEQ ID NO: 34)
BRCA2	13	32319283	C	T	ACTCTGCCGCTGTACTAATCTCCTGTAAAAG	32319268:32319298	CTTTTACAGGAGATTA,TTTTACAGGAGATTAG,TCTGCCGCTGTACTAA,CTGCCGCTGTACTAAT,TTACAGGAGATTAGTA,TGCCGCTGTACTAATC,TACAGGAGATTAGTAC,GCCGCTGTACTAATCT,ACAGGAGATTAGTACA,CCGCTGTACTAATCTC,CGCTGTACTAATCTCC,AGGAGATTAGTACAGC,GCTGTACTAATCTCCT,GGAGATTAGTACAGCG,GAGATTAGTACAGCGG,TGTACTAATCTCCTGT,AGATTAGTACAGCGGC,GTACTAATCTCCTGTA,GATTAGTACAGCGGCA,TACTAATCTCCTGTAA,ATTAGTACAGCGGCAG,TTAGTACAGCGGCAGA,CTAATCTCCTGTAAAA,TAATCTCCTGTAAAAG (SEQ ID NO: 35 – SEQ ID NO: 58)
BRCA2	13	32330949	G	T	TTTTCCAATCATGATTAAAGTCTGAAGAAAA	32330934:32330964	TTTCCAATCATGATTA,TTCCAATCATGATTAA,CTTCAGACTTTAATCA,CAGACTTTAATCATGA,TCATGATTAAAGTCTG,TGATTAAAGTCTGAAG,TTAATCATGATTGGAA,TAATCATGATTGGAAA (SEQ ID NO: 59 – SEQ ID NO: 66)
EGFR	7	55201767	C	A	TTCCACCGTGGCTTGAATTGATA	55201752:55201782	TTCCACCGTGGCTTGA,CCACCGTGGCTTGAAT,CACCGTGGCTTGAATT,ACCGTGGCTTGAATTG,CCGTGGCTTGAATTGA,CGTGGCTTGAATTGAT,ATCAATTCAAGCCACG,TCAATTCA

RELATED GENE	hrom.	pos	ref	alt.	seq	coordinates	list of nullomers
					GAAATGGG		AGCCACGG,CAATTCAAGCCACGGT,AATTCAAGCCACGGTG,ATTCAAGCCACGGTGG,TCAAGCCACGGTGGAA (SEQ ID NO: 67 – SEQ ID NO: 78)
EGFR	7	55201302	C	T	TACCTCATCCCACAGTAGGGCTTCTTCAGCA	55201287:55201317	TCATCCCACAGTAGGG,TCCCACAGTAGGGCTT,AAGCCCTACTGTGGGA,CCCTACTGTGGGATGA (SEQ ID NO: 79 – SEQ ID NO: 82)
EGFR	7	55200384	C	T	TTCTCCAAAATGGCCTGAGACCCCCAGCGCT	55200369:55200399	AGCGCTGGGGGTCTCA,CAAAATGGCCTGAGAC,GTCTCAGGCCATTTTG,TGAGACCCCCAGCGCT (SEQ ID NO: 83 – SEQ ID NO: 86)
EGFR	7	55174014	G	T	AAGATCAAAGTGCTGTGCTCCGGTGCGTTCG	55173999:55174029	CGAACGCACCGGAGCA,GAACGCACCGGAGCAC,GATCAAAGTGCTGTGC,AACGCACCGGAGCACA,ACGCACCGGAGCACAG,CGCACCGGAGCACAGC,GCACCGGAGCACAGCA,AAAGTGCTGTGCTCCG,AAGTGCTGTGCTCCGG,ACCGGAGCACAGCACT,AGTGCTGTGCTCCGGT,CCGGAGCACAGCACTT,CGGAGCACAGCACTTT,TGCTGTGCTCCGGTGC,GCTGTGCTCCGGTGCG,CTGTGCTCCGGTGCGT,TGTGCTCCGGTGCGTT,GCACAGCACTTTGATC,GTGCTCCGGTGCGTTC,TGCTCCGGTGCGTTCG (SEQ ID NO: 87 – SEQ ID NO: 106)
EGFR	7	55191831	T	A	TTGGGCTGGCCAAACAGCTGGGTGCGGAAGA	55191816:55191846	TTGGGCTGGCCAAACA,TCTTCCGCACCCAGCT,CTTCCGCACCCAGCTG,TTCCGCACCCAGCTGT,CACCCAGCTGTTTGGC,GCCAAACAGCTGGGTG,ACAGCTGGGTGCGGAA,CAGCTGGGTGCGGAAG,AGCTGGGTGCGGAAGA,TGTTTGGCCAGCCCAA (SEQ ID NO: 107 – SEQ ID NO: 116)
EGFR	7	55198760	T	A	TGGATCCAAGCCATAAGACGGAATCCCTGCC	55198745:55198775	TGGATCCAAGCCATAA,GCAGGGATTCCGTCTT,CAGGGATTCCGTCTTA,AGGGATTCCGTCTTAT,TCCAAGCCATAAGACG,GGGATTCCGTCTTATG,CCAAGCCATAAGACGG,GGATTCCGTCTTATGG,CAAGCCATAAGACGGA,GATTCCGTCTTATGGC,ATTCCGTCTTATGGCT,AGCCATAAGACGGAAT,GCCATAAGACGGAATC,TCCGTCTTATGGCTTG,CCATAAGACGGAATCC,CCGTCTTATGGCTTGG,CATAAGACGGAATCCC,CGTCTTATGGCTTGGA,ATAAGACGGAATCCCT,TAAGACGGAATCCCTG,AAGACGGAATCCCTGC,TTATGGCTTGGATCCA (SEQ ID NO: 117 – 138)
TP53	17	7673767	G	A	GGAGATTCTCTTCCTATGTGCGCCGGTCTCT	7673752:7673782	AGAGACCGGCGCACAT,GAGACCGGCGCACATA,AGACCGGCGCACATAG,GACCGGCGCACATAGG,ACCGGCGCACATAGGA,CCGGCGCACATAGGAA,CGGCGCACATAGGAAG,CTCTTCCTATGTGCGC,GGCGCACATAGGAAGA,TCTTCCTATGTGCGCC,GCGCACATAGGAAGAG,CTTCCTATGTGCGCCG,TTCCTATGTGCGCCGG,TCCTATGTGCGCCGGT,CCTATGTGCGCCG

RELATED GENE	hro m.	pos	ref	alt.	seq	coordinates	list of nullomers
							GTC,CTATGTGCGCCGGTCT,TATGTGCGCCG GTCTC,ATGTGCGCCGGTCTCT (SEQ ID NO: 139 – SEQ ID NO: 156)

Table C

BRCA2		
Indication - Sample Type	Drug Trade Name (Generic) NDA / BLA	Biomarker(s)
Ovarian Cancer - Whole Blood	Lynparza (olaparib) NDA 208558	BRCA1 and BRCA2
Breast Cancer - Whole Blood	Lynparza (olaparib) NDA 208558	BRCA1 and BRCA2
Breast Cancer - Whole Blood	Talzenna (talazoparib) NDA 211651	BRCA1 and BRCA2
Ovarian Cancer - Whole Blood	Rubraca (rucaparib) NDA 209115	BRCA1 and BRCA2
Pancreatic Cancer - Whole Blood	Lynparza (olaparib) NDA 208558	BRCA1 and BRCA2
Metastatic Castrate Resistant Prostate Cancer (mCRPC) - Whole Blood	Lynparza (olaparib) NDA 208558	BRCA1 and BRCA2
Ovarian Cancer - Tissue	Rubraca (rucaparib) NDA 209115	BRCA1 and BRCA2
Ovarian Cancer - Tissue	Rubraca (rucaparib) NDA 209115	BRCA1 and BRCA2
Ovarian Cancer - Tissue	Lynparza (olaparib) NDA 208558	BRCA1 and BRCA2
Metastatic Castrate Resistant Prostate Cancer (mCRPC) - Plasma	Rubraca (rucaparib) NDA 209115	BRCA1 and BRCA2
Ovarian Cancer - Plasma	Rubraca (rucaparib) NDA 209115	BRCA1 and BRCA2
Metastatic Castrate Resistant Prostate Cancer (mCRPC) - Plasma	Lynparza (olaparib) NDA 208558	BRCA1, BRCA2 and ATM
EGFR1		
Indication - Sample Type	Drug Trade Name (Generic) NDA / BLA	
Non-Small Cell Lung Cancer (NSCLC) - Tissue	Tarceva (erlotinib) NDA 021743	
Non-Small Cell Lung Cancer (NSCLC) - Tissue	Tagrisso (osimertinib) NDA 208065	
Non-Small Cell Lung Cancer (NSCLC) - Tissue	Tagrisso (osimertinib) NDA 208065	

Non-Small Cell Lung Cancer (NSCLC) - Plasma	Tagrisso (osimertinib) NDA 208065
Non-Small Cell Lung Cancer (NSCLC) - Tissue or Plasma	Iressa (gefitinib) NDA 206995
Non-Small Cell Lung Cancer (NSCLC) - Tissue or Plasma	Iressa (gefitinib) NDA 206995
Non-Small Cell Lung Cancer (NSCLC) - Tissue or Plasma	Tarceva (erlotinib) NDA 021743
Non-Small Cell Lung Cancer (NSCLC) - Tissue or Plasma	Gilotrif (afatinib) NDA 201292
Non-Small Cell Lung Cancer (NSCLC) - Tissue or Plasma	Tagrisso (osimertinib) NDA 208065
Non-Small Cell Lung Cancer (NSCLC) - Plasma	Tagrisso (osimertinib) NDA 208065
Non-Small Cell Lung Cancer (NSCLC) - Plasma	Tarceva (erlotinib) NDA 021743
Colorectal Cancer - Tissue	Erbitux (cetuximab) BLA 125084
Colorectal Cancer - Tissue	Vectibix (panitumumab) BLA 125147
Non-Small Cell Lung Cancer (NSCLC) - Tissue	Gilotrif (afatinib) NDA 201292
Non-Small Cell Lung Cancer (NSCLC) - Tissue	Iressa (gefitinib) NDA 206995
Non-Small Cell Lung Cancer (NSCLC) - Tissue	Tarceva (erlotinib) NDA 021743
Non-Small Cell Lung Cancer (NSCLC) - Tissue	Tagrisso (osimertinib) NDA 208065
Non-Small Cell Lung Cancer (NSCLC) - Tissue	Tagrisso (osimertinib) NDA 208065
Non-Small Cell Lung Cancer (NSCLC) - Plasma	Iressa (gefitinib) NDA 206995
Non-Small Cell Lung Cancer (NSCLC) - Plasma	Tagrisso (osimertinib) NDA 208065
Non-Small Cell Lung Cancer (NSCLC) - Plasma	Tarceva (erlotinib) NDA 021743
Non-Small Cell Lung Cancer (NSCLC) - Plasma	Tagrisso (osimertinib) NDA 208065
Non-Small Cell Lung Cancer (NSCLC) - Plasma	Rybrevant (amivantamb) BLA 761210
Non-Small Cell Lung Cancer (NSCLC) - Tissue	Iressa (gefitinib) NDA 206995
Non-Small Cell Lung Cancer (NSCLC) - Tissue	Gilotrif (afatinib) NDA 201292
Non-Small Cell Lung Cancer (NSCLC) - Tissue	Tarceva (erlotinib) NDA 021743

Non-Small Cell Lung Cancer (NSCLC) - Tissue	Vizimpro (dacomitinib) NDA 211288
Non-Small Cell Lung Cancer (NSCLC) - Tissue	Iressa (gefitinib) NDA 206995
Non-Small Cell Lung Cancer (NSCLC) - Tissue	Rybrevant (amivantamb) BLA 761210
Non-Small Cell Lung Cancer (NSCLC) - Tissue	Exkivity (mobocertinib) NDA 215310
Non-Small Cell Lung Cancer (NSCLC) - Tissue	Gilotrif (afatinib) NDA 201292
Non-Small Cell Lung Cancer (NSCLC) - Tissue	Iressa (gefitinib) NDA 206995
Non-Small Cell Lung Cancer (NSCLC) - Tissue	Gilotrif (afatinib) NDA 201292
Non-Small Cell Lung Cancer (NSCLC) - Tissue	Vizimpro (dacomitinib) NDA 211288
TP53	
Indication - Sample Type	**Drug Trade Name (Generic) NDA / BLA**
B-cell Chronic Lymphocytic Leukemia - Peripheral Blood	Venclexta (venetoclax) NDA 208573

In some embodiments, the methods comprise a step of correlating the presence of the neomer to a mutation or dysfunctional phenotype of the cancer cell. In some embodiments, the methods further comprise pairing the mutation or dysfunctional phenotype to a therapy recommendation for the subject or a likelihood that the subject would be responsive to a certain therapy.

In some embodiments, the presence of a neomer that comprises at least about 80% sequence identity to any one of SEQ ID NO: 1 through SEQ ID NO: 2, is correlated to a mutation at sequence CCGTGCAGCTCATCATGCAGCTCATGCCCTT of the EGFR1 locus or regulatory element within the cancer cell of the subject. Further still, methods comprise a step of recommending or determining a therapy for the subject based upon the presence of the neomer. In some embodiments, the subject has non-small cell lung cancer. Therefore in some embodiments, methods further comprise a step of recommending therapy of one or a combination of erlotinib, osimertinib, gefitinib, afatinib, amivantamb, dacomitinib or mobocertinib if the sample of the subject comprises the presence of the neomer that comprises at least about 80% sequence identity to any one of SEQ ID NO: 1 through SEQ ID NO: 2.

In some embodiments, the presence of a neomer that comprises at least about 80% sequence identity to any one of SEQ ID NO: 1 through SEQ ID NO: 2, is correlated to a mutation at sequence CCGTGCAGCTCATCATGCAGCTCATGCCCTT of the EGFR1 locus or regulatory element within the cancer cell of the subject. Further still, methods comprise a step of recommending or determining a therapy for the subject based upon the presence of the neomer. In some embodiments, the subject has colorectal cancer. Therefore in some embodiments, methods further comprise a step of recommending therapy of one or a combination of cetuximab and/or panitumumab if the sample of the subject comprises the presence of the neomer that comprises at least about 80% sequence identity to any one of SEQ ID NO: 1 through SEQ ID NO: 2.

In some embodiments, the presence of a neomer that comprises at least about 80% sequence identity to any one of SEQ ID NO: 3 through SEQ ID NO: 28, is correlated to a mutation at sequence TCACAGATTTTGGGCGGGCCAAACTGCTGGG of the EGFR1 locus or regulatory element within the cancer cell of the subject. Further still, methods comprise a step of recommending or determining a therapy for the subject based upon the presence of the neomer. In some embodiments, the subject has non-small cell lung cancer. Therefore in some embodiments, methods further comprise a step of recommending therapy of one or a combination of erlotinib, osimertinib, gefitinib, afatinib, amivantamb, dacomitinib or mobocertinib if the sample of the subject comprises the presence of the neomer that comprises at least about 80% sequence identity to any one of SEQ ID NO: 3 through SEQ ID NO: 28.

In some embodiments, the presence of a neomer that comprises at least about 80% sequence identity to any one of SEQ ID NO: 3 through SEQ ID NO: 28, is correlated to a mutation at sequence TCACAGATTTTGGGCGGGCCAAACTGCTGGG of the EGFR1 locus or regulatory element within the cancer cell of the subject. Further still, methods comprise a step of recommending or determining a therapy for the subject based upon the presence of the neomer. In some embodiments, the subject has colorectal cancer. Therefore in some embodiments, methods further comprise a step of recommending therapy of one or a combination of cetuximab and/or panitumumab if the sample of the subject comprises the presence of the neomer that comprises at least about 80% sequence identity to any one of SEQ ID NO: 3 through SEQ ID NO: 28.

In some embodiments, the presence of a neomer that comprises at least about 80% sequence identity to any one of SEQ ID NO: 29 through SEQ ID NO: 66, is correlated to a mutation at sequence identified in Table B of the BRCA2 locus or regulatory element within the cancer cell

of the subject. Further still, methods comprise a step of recommending or determining a therapy for the subject based upon the presence of the neomer. Therefore in some embodiments, methods further comprise a step of recommending therapy of one or a combination of the therapies listed in Table C if the sample of the subject comprises the presence of the neomer that comprises at least about 80% sequence identity to any one of SEQ ID NO: 29 through SEQ ID NO: 66.

In some embodiments, the presence of a neomer that comprises at least about 80% sequence identity to any one of SEQ ID NO: 67 through SEQ ID NO: 138, is correlated to a mutation at sequence identified in Table B of the EGFR locus or regulatory element within the cancer cell of the subject. Further still, methods comprise a step of recommending or determining a therapy for the subject based upon the presence of the neomer. Therefore in some embodiments, methods further comprise a step of recommending therapy of one or a combination of the therapies listed in Table C if the sample of the subject comprises the presence of the neomer that comprises at least about 80% sequence identity to any one of SEQ ID NO: 67 through SEQ ID NO: 138.

In some embodiments, the presence of a neomer that comprises at least about 80% sequence identity to any one of SEQ ID NO: 139 through SEQ ID NO: 156, is correlated to a mutation at sequence identified in Table B of the TP53 locus or regulatory element within the cancer cell of the subject. Further still, methods comprise a step of recommending or determining a therapy for the subject based upon the presence of the neomer. Therefore in some embodiments, methods further comprise a step of recommending therapy of one or a combination of the therapies listed in Table C if the sample of the subject comprises the presence of the neomer that comprises at least about 80% sequence identity to any one of SEQ ID NO: 139 through SEQ ID NO: 156.

Additional Diagnostic Tests

In some embodiments, the quantity of neomers is indicative of various types of cancer may be used as a stand-alone diagnostic indicator of cancer in a subject. Optionally, the methods may include the performance of at least one additional test to facilitate the diagnosis of cancer. For example, other tests in addition to determining the level of one or more nullomers and/or neomers in order to facilitate a diagnosis of cancer may be performed. Any other test or combination of tests used in clinical practice to facilitate a diagnosis of cancer may be used in conjunction with the neomers. In some embodiments, the method of diagnosis comprise identifying the presence or quantity of the amount of neomer in a sample from a subject.

Methods of Treatment

In some embodiments, where a subject is diagnosed with a particular type of cancer by the methods described herein, the disclosure further provides methods of treating the subject identified as having a cancer or a method of propsing a therapy for better responsiveness of the subject to cancer therapy. Accordingly, in some embodiments, the disclosure relates to a method of treating cancer in a subject, comprising determining the level of at least one neomer in a sample from the subject, wherein a difference in the level of at least one neomer versus that in a normal subject as determined relative to a suitable control is indicative of cancer in the subject, and administering a therapeutically effective amount of a cancer therapeutic to the subject. In another embodiments, the disclosure relates to a method of treating a subject having cancer, comprising identifying a subject having cancer in which the level of at least one neomer in a sample from the subject is different (e.g., increased) versus that in a normal subject as determined relative to a suitable control, and administering a therapeutically effective amount of a cancer therapeutic to the subject.

The term "cancer therapeutic" includes, for example, substances approved by the U.S. Food and Drug Administration for the treatment of cancer. For instance, drugs approved to treat breast cancer include, but are not limited to, Abemaciclib, Abitrexate (Methotrexate), Abraxane (Paclitaxel Albumin-stabilized Nanoparticle Formulation), Ado-Trastuzumab Emtansine, Afinitor (Everolimus), Anastrozole, Aredia (Pamidronate Disodium), Arimidex (Anastrozole), Aromasin (Exemestane), Capecitabine, Clafen (Cyclophosphamide), Cyclophosphamide, Cytoxan (Cyclophosphamide), Docetaxel, Doxorubicin Hydrochloride, Ellence (Epirubicin Hydrochloride), Epirubicin Hydrochloride, Eribulin Mesylate, Everolimus, Exemestane, 5-FU (Fluorouracil Injection), Fareston (Toremifene), Faslodex (Fulvestrant), Femara (Letrozole), Fluorouracil Injection, Folex (Methotrexate), Folex PFS (Methotrexate), Fulvestrant, Gemcitabine Hydrochloride, Gemzar (Gemcitabine Hydrochloride), Goserelin Acetate, Halaven (Eribulin Mesylate), Herceptin (Trastuzumab), Ibrance (Palbociclib), Ixabepilone, Ixempra (Ixabepilone), Kadcyla (Ado-Trastuzumab Emtansine), Kisqali (Ribociclib), Lapatinib, Ditosylate, Letrozole, Megestrol Acetate, Methotrexate, Methotrexate LPF (Methotrexate), Mexate (Methotrexate), Mexate-AQ (Methotrexate), Neosar (Cyclophosphamide), Neratinib Maleate, Nerlynx (Neratinib Maleate), Nolvadex (Tamoxifen Citrate), Paclitaxel, Paclitaxel Albumin-stabilized Nanoparticle Formulation, Palbociclib, Pamidronate Disodium, Perjeta (Pertuzumab), Pertuzumab, Ribociclib,

Tamoxifen Citrate, Taxol (Paclitaxel), Taxotere (Docetaxel), Thiotepa, Toremifene, Trastuzumab, Tykerb (Lapatinib Ditosylate), Velban (Vinblastine Sulfate), Velsar (Vinblastine Sulfate), Verzenio (Abemaciclib), Vinblastine Sulfate, Xeloda (Capecitabine), Zoladex (Goserelin Acetate).

The cancer therapeutics may be administered to a subject using a pharmaceutical composition. Suitable pharmaceutical compositions comprise a pharmaceutically effective amount of a cancer therapeutic (or a pharmaceutically acceptable salt or ester thereof), and optionally comprise a pharmaceutically acceptable carrier. In certain embodiments, these compositions optionally further comprise one or more additional therapeutic agents.

As used herein, the term "pharmaceutically acceptable salt" refers to those salts which are, within the scope of sound medical judgment, suitable for use in contact with the tissues of humans and lower animals without undue toxicity, irritation, allergic response and the like, and are commensurate with a reasonable benefit/risk ratio. Pharmaceutically acceptable salts of amines, carboxylic acids, and other types of compounds, are well known in the art. For example, S. M. Berge, et al. describe pharmaceutically acceptable salts in detail in J. Pharmaceutical Sciences, 66: 1-19 (1977), incorporated herein by reference. The salts can be prepared in situ during the final isolation and purification of the compounds, or separately by reacting a free base or free acid function with a suitable reagent. For example, a free base function can be reacted with a suitable acid. Furthermore, where the compounds carry an acidic moiety, suitable pharmaceutically acceptable salts thereof may, include metal salts such as alkali metal salts, e.g., sodium or potassium salts, and alkaline earth metal salts, e.g., calcium or magnesium salts. In some embodiments, the cancer therapeutic is a pharmaceutically acceptable salt.

The term "pharmaceutically acceptable ester," as used herein, refers to esters that hydrolyze in vivo and include those that break down readily in the human body to leave the parent compound or a salt thereof. Suitable ester groups include, for example, those derived from pharmaceutically acceptable aliphatic carboxylic acids, particularly alkanoic, alkenoic, cycloalkanoic and alkanedioic acids, in which each alkyl or alkenyl moiety advantageously has not more than 6 carbon atoms. In some embodiments, the cancer therapeutic is a pharmaceutically acceptable ester.

As described above, the pharmaceutical compositions may additionally comprise a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" includes any

and all solvents, diluents, or other liquid vehicle, dispersion or suspension aids, surface active agents, isotonic agents, thickening or emulsifying agents, preservatives, solid binders, lubricants and the like, suitable for preparing the particular dosage form desired. Remington's Pharmaceutical Sciences, Sixteenth Edition, E. W. Martin (Mack Publishing Co., Easton, Pa., 1980) discloses various carriers used in formulating pharmaceutical compositions and known techniques for the preparation thereof. Some examples of materials which can serve as pharmaceutically acceptable carriers include, but are not limited to, sugars such as lactose, glucose and sucrose; starches such as corn starch and potato starch; cellulose and its derivatives such as sodium carboxymethyl cellulose, ethyl cellulose and cellulose acetate; powdered tragacanth; malt; gelatine; talc; excipients such as cocoa butter and suppository waxes; oils such as peanut oil, cottonseed oil; safflower oil, sesame oil; olive oil; corn oil and soybean oil; glycols; such as propylene glycol; esters such as ethyl oleate and ethyl laurate; agar; buffering agents such as magnesium hydroxide and aluminum hydroxide; alginic acid; pyrogenfree water; isotonic saline; Ringer's solution; ethyl alcohol, and phosphate buffer solutions, as well as other non-toxic compatible lubricants such as sodium lauryl sulfate and magnesium stearate, as well as coloring agents, releasing agents, coating agents, sweetening, flavoring and perfuming agents, preservatives and antioxidants can also be present in the composition, according to the judgment of the formulator.

Compositions for use in the present disclosure may be formulated to have any concentration of the cancer therapeutic desired. In some embodiments, the composition is formulated such that it comprises a therapeutically effective amount of the cancer therapeutic.

The disclosure generally relates to a method of diagnosing a subject with a benign, pre-malignant, or malignant hyperproliferative cell comprising: detecting the presence, absence, and/or quantity of at least one neomer in a sample. In some embodiments, the step of detecting comprise exposing a sample from a subject (e.g., a human subject) to one or a plurality of probes, each probe capable of binding one or a plurality of neomers in the sample. In some embodiments, the probe is a labeled nucleic acid molecule (DNA, RNA or hybrid thereof) that comprises at least about 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to the complement of any nucleic acid sequences of Table 1, Table 5, Table 6 or Table B. In some embodiments, the probe is a labeled nucleic acid molecule (DNA, RNA or hybrid thereof) that is an RNA sequence comprising at least about 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%,

95%, 96%, 97%, 98%, 99% or 100% sequence identity to the complement of any nucleic acid sequences of TABLE 1, where each thymine is replaced with a uracil. In some embodiments, the plurality of probes are one or a combination of labeled nucleic acid sequences that are an RNA complementary to a nucleic acid sequence comprising at least about 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to any nucleic acid sequences of TABLE 1 or Table B. In some embodiments, the plurality of probes are one or a combination of labeled nucleic acid sequences chosen from any nucleic acid sequences of TABLE 1. In some embodiments, the plurality of probes comprise one or a combination of nucleic acid sequences complementary to the nucleic acid sequences chosen from any nucleic acid sequences of TABLE 1. In some embodiments, the probe is a labeled nucleic acid molecule (DNA, RNA or hybrid thereof) that comprises at least about 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to the complement of any nucleic acid sequences of TABLE 1. In some embodiments, the probe is a labeled nucleic acid molecule (DNA, RNA or hybrid thereof) that is an RNA sequence comprising at least about 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to the complement of any nucleic acid sequences of TABLE 1, where each thymine is replaced with a uracil. In some embodiments, the plurality of probes are one or a combination of labeled nucleic acid sequences that are an RNA complementary to a nucleic acid sequence comprising at least about 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to any nucleic acid sequences of TABLE 1. In some embodiments, the plurality of probes are one or a combination of labeled nucleic acid sequences chosen from any nucleic acid sequences of TABLE 1. In some embodiments, the plurality of probes comprise one or a combination of nucleic acid sequences complementary to the nucleic acid sequences chosen from any nucleic acid sequences of TABLE 1.

In some embodiments, the probe is a labeled nucleic acid molecule (DNA, RNA or hybrid thereof) that comprises at least about 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to the complement of any nucleic acid sequences of TABLE 1.

In some embodiments, the probe is a labeled nucleic acid molecule (DNA, RNA or hybrid thereof) that is an RNA sequence comprising at least about 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to the complement of any

nucleic acid sequences of TABLE 5, where each thymine is replaced with a uracil. In some embodiments, the plurality of probes are one or a combination of labeled nucleic acid sequences that are an RNA complementary to a nucleic acid sequence comprising at least about 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to any nucleic acid sequences of TABLE 5. In some embodiments, the plurality of probes are one or a combination of labeled nucleic acid sequences chosen from any nucleic acid sequences of TABLE 5. In some embodiments, the plurality of probes comprise one or a combination of nucleic acid sequences complementary to the nucleic acid sequences chosen from any nucleic acid sequences of TABLE 5.

In some embodiments, the probe is a labeled nucleic acid molecule (DNA, RNA or hybrid thereof) that comprises at least about 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to the complement of any nucleic acid sequences of TABLE 6. In some embodiments, the probe is a labeled nucleic acid molecule (DNA, RNA or hybrid thereof) that is an RNA sequence comprising at least about 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to the complement of any nucleic acid sequences of TABLE 6, where each thymine is replaced with a uracil. In some embodiments, the plurality of probes are one or a combination of labeled nucleic acid sequences that are an RNA complementary to a nucleic acid sequence comprising at least about 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to any nucleic acid sequences of TABLE 6. In some embodiments, the plurality of probes are one or a combination of labeled nucleic acid sequences chosen from any nucleic acid sequences of TABLE 6. In some embodiments, the plurality of probes comprise one or a combination of nucleic acid sequences complementary to the nucleic acid sequences chosen from any nucleic acid sequences of TABLE 6.

In some embodiments, the probe is a labeled nucleic acid molecule (DNA, RNA or hybrid thereof) that comprises at least about 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to the complement of any nucleic acid sequences of TABLE B. In some embodiments, the probe is a labeled nucleic acid molecule (DNA, RNA or hybrid thereof) that is an RNA sequence comprising at least about 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to the complement of any nucleic acid sequences of TABLE B, where each thymine is replaced with a uracil. In some

embodiments, the plurality of probes are one or a combination of labeled nucleic acid sequences that are an RNA complementary to a nucleic acid sequence comprising at least about 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to any nucleic acid sequences of TABLE B. In some embodiments, the plurality of probes are one or a combination of labeled nucleic acid sequences chosen from any nucleic acid sequences of TABLE B. In some embodiments, the plurality of probes comprise one or a combination of nucleic acid sequences complementary to the nucleic acid sequences chosen from any nucleic acid sequences of TABLE B.

In any of the disclosed method embodiments, the subject may be a human diagnosed with or suspected as having cancer. In any of the disclosed method embodiments, wherein the step of detecting is preceded by a step of acquiring a sample from the subject.

In some embodiments, the probe or plurality of probes are one or a plurality of antibodies or antibody fragments comprising a CDR that binds to a nucleic acid molecule (DNA, RNA or hybrid thereof) that comprises at least 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to any nucleic acid sequences of TABLE 1. In some embodiments, the probe or plurality of probes are one or a plurality of antibodies or antibody fragments comprising a CDR that binds to a nucleic acid molecule (DNA, RNA or hybrid thereof) that comprises at least about 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to any nucleic acid sequences of TABLE 1, wherein each of sequences are modified such that the thymines in each sequence are replaced with a uracil. In some of the embodiments, the methods further comprise isolating RNA from the sample before exposing the sample to one or a plurality of probes. In some embodiments, the method comprises detecting or quantifying an amount of neomers in a sample by performing semiquantitative or quantitative PCR or sequencing analysis of the neomers in a sample. Probes may be immobilized to a solid support such as an ELISA plate, plastic, slide, microarray, silica chip or other surface such that the single-strand nucleotide sequences are exposed to a sample comprising neomers from a subject. The probes may comprise, in some embodiments, from about 5 to about 100 nucleotides in length and comprise any of the sequences provided in TABLE 1 or any complementary sequence in RNA or DNA form of the sequences set forth in TABLE 1. In any of the disclosed method embodiments, the step of detecting the presence, absence, and/or quantity of at least one neomer having at least about 70% sequence identity to one of the neomers in a sample comprises using a

chemoluminescent probe, fluorescent probe, and/or fluorescence microscopy, calculating the presence or quantity by correlating the signal of the detectable probe to the presence of the neomer.

In some embodiments, any of the methods disclosed herein further comprise a step of correlating the presence or quantity of one or more neomers, such as those disclosed in TABLE 1 or any combination thereof, to the likelihood that the subject has cancer. In some embodiments, the disclosure relates to a method of preparing, isolating or assessing a nucleic acid or ribonucleic acid fraction from a subject useful for analyzing a neomer involved in cancer comprising: extracting DNA or RNA from a substantially cell-free sample of blood plasma or blood serum of a subject to obtain DNA or RNA pools; (b) producing a fraction of the DNA or RNA extracted in (a) by: (i) sequence discrimination of the DNA or RNA; and (ii) selectively removing neomers by exposing one or a plurality of probes to the neomers, wherein the neomers after (b) comprises one or a plurality of neomers disclosed in TABLE 1; and (c) analyzing the neomers in the fraction of DNA or RNA produced in (b). In some embodiments, the step of analyzing comprises normalizing the amount of neomers in the sample as compared to a control amount of neomers from a control sample and determining whether the subject has cancer by comparing the normalized presence, absence or quantity of neomers in the sample to the presence, absence or quantity of neomers in a control sample.

Kits for Detection of Neomers

The disclosure also provides kits for diagnosing type of cancer, tissue of origin, and status or stage of the cancer in a subject, which kits are useful for determining the level of one or more neomers from TABLE 1 or Table B, wherein the sequences optionally comprise uracils in place of one, more than one, or all of the disclosed thymines), and combinations thereof. In some embodiments, the one or more neomers are selected from the neomers listed in TABLE 1 or Table B. Kits may include materials and reagents adapted to selectively detect the presence of a neomer or group of neomers diagnostic for cancer in a sample of a subject. For example, in some embodiments, the kit may include a reagent that specifically hybridizes to a neomer. Such a reagent may be a nucleic acid molecule in a form suitable for detecting the neomer, for example, a probe or a primer. The kit may include reagents useful for performing an assay to detect one or more neomers, for example, reagents which may be used to detect one or more neomers in a qPCR reaction. The kit may likewise include a microarray useful for detecting one or more neomers.

In some embodiments, the kit may contain instructions for suitable operational parameters

in the form of a label or product insert. For example, the instructions may include information or directions regarding how to collect a sample, how to determine the level of one or more neomers in a sample, and/or how to correlate the level of one or more neomers in a sample with the type of cancer, tissue of origin, and status or stage of the cancer of a subject.

In some embodiments, the kit can contain one or more containers with neomer samples, to be used as reference standards, suitable controls, or for calibration of an assay to detect the neomers in a test sample.

TABLE 2. Radioisotopes that may be incorporated into pharmaceutical compositions or used as probes or labels with neomers.

^{2}H, ^{3}H, ^{13}C, ^{14}C, ^{15}N, $^{16}0$, $^{17}0$, ^{31}P, ^{32}P, ^{35}S, ^{18}F, $^{36}C1$, ^{225}Ac, ^{227}Ac, ^{212}Bi, ^{213}Bi, ^{109}Cd, ^{60}Co, ^{64}Cu, ^{67}Cu, ^{166}Dy, ^{169}Er, ^{152}Eu, ^{154}Eu, ^{153}Gd, ^{198}Au, ^{166}Ho, ^{125}I, ^{131}I, ^{192}Ir, ^{177}Lu, ^{99}Mo, $^{194}0s$, ^{103}Pd, ^{195m}Pt, ^{32}P, ^{33}P, ^{223}Ra, ^{186}Re, ^{188}Re, ^{105}Rh, ^{145}Sm, ^{153}Sm, ^{47}Sc, ^{75}Se, ^{85}Sr, ^{89}Sr, ^{99m}Tc, ^{228}Th, ^{229}Th, ^{170}Tm, ^{117m}Sn, ^{188}W, ^{127}Xe, ^{175}Yb, ^{90}Y, ^{91}Y

TABLE 3. Table of chemotherapeutic agents.

Alkylating agents

- Cyclophosphamide
- Mechlorethamine
- Chlorambucil
- Melphalan

Anthracyclines

- Daunorubicin
- Doxorubicin
- Epirubicin
- Idarubicin
- Mitoxantrone
- Valrubicin

Cytoskeletal disruptors (Taxanes)

- Paclitaxel
- Docetaxel

Epothilones

Histone Deacetylase Inhibitors

- Vorinostat
- Romidepsin

Inhibitors of Topoisomerase I

- Irinotecan
- Topotecan

Inhibitors of Topoisomerase II

- Etoposide
- Teniposide
- Tafluposide

Kinase inhibitors

- Afatinib
- Bortezomib
- Erlotinib
- Gefitinib
- Imatinib
- Osimertinib

Vemurafenib

- Vismodegib

Monoclonal antibodies

- Bevacizumab
- Cetuximab
- Ipilimumab
- Ofatumumab
- Ocrelizumab
- Panitumab
- Rituximab

Nucleotide analogs and precursor analogs

- Azacitidine
- Azathioprine

- Capecitabine
- Cytarabine
- Doxifluridine
- Fluorouracil
- Gemcitabine
- Hydroxyurea
- Mercaptopurine
- Methotrexate
- Tioguanine (formerly Thioguanine)

PARP inhibitors

- Olaparib
- Rucaparib

Peptide antibiotics

- Bleomycin
- Actinomycin

Platinum-based agents

- Carboplatin
- Cisplatin
- Oxaliplatin

Retinoids

- Tretinoin
- Alitretinoin
- Bexarotene

Vinca alkaloids and derivatives

- Vinblastine
- Vincristine
- Vindesine
- Vinorelbine
- Actinomycin
- All-trans retinoic acid
- Azacitidine

- Azathioprine
- Bleomycin
- Bortezomib
- Carboplatin
- Capecitabine
- Cisplatin
- Chlorambucil
- Cyclophosphamide
- Cytarabine
- Daunorubicin
- Docetaxel
- Doxifluridine
- Doxorubicin
- Epirubicin
- Epothilone
- Etoposide
- Fluorouracil
- Gemcitabine
- Hydroxyurea
- Idarubicin
- Imatinib
- Irinotecan
- Mechlorethamine
- Mercaptopurine
- Methotrexate
- Mitoxantrone
- Oxaliplatin
- Paclitaxel
- Pemetrexed
- Teniposide
- Tioguanine

- Topotecan
- Valrubicin
- Vinblastine
- Vincristine
- Vindesine
- Vinorelbine

Systems

In some methods of treatment disclosed herein, the agent in selected from one or a plurality of agents chosen from Table 3.

The above-described methods can be implemented in any of numerous ways. For example, the embodiments may be implemented using a computer program product (i.e. software), hardware, software or a combination thereof. When implemented in software, the software code can be executed on any suitable processor or collection of processors, whether provided in a single computer or distributed among multiple computers.

Further, it should be appreciated that a computer may be embodied in any of a number of forms, such as a rack-mounted computer, a desktop computer, a laptop computer, or a tablet computer. Additionally, a computer may be embedded in a device not generally regarded as a computer but with suitable processing capabilities, including a Personal Digital Assistant (PDA), a smart phone or any other suitable portable or fixed electronic device.

Also, a computer may have one or more input and output devices. These devices can be used, among other things, to present a user interface. Examples of output devices that can be used to provide a user interface include printers or display screens for visual presentation of output and speakers or other sound generating devices for audible presentation of output. Examples of input devices that can be used for a user interface include keyboards, and pointing devices, such as mice, touch pads, and digitizing tablets. As another example, a computer may receive input information through speech recognition or in other audible format.

Such computers may be interconnected by one or more networks in any suitable form, including a local area network or a wide area network, such as an enterprise network, and intelligent network (IN) or the Internet. Such networks may be based on any suitable technology and may operate according to any suitable protocol and may include wireless networks, wired networks or fiber optic networks.

A computer employed to implement at least a portion of the functionality described herein may include a memory, coupled to one or more processing units (also referred to herein simply as "processors"), one or more communication interfaces, one or more display units, and one or more user input devices. The memory may include any computer-readable media, and may store computer instructions (also referred to herein as "processor-executable instructions") for implementing the various functionalities described herein. The processing unit(s) may be used to execute the instructions. The communication interface(s) may be coupled to a wired or wireless network, bus, or other communication means and may therefore allow the computer to transmit communications to and/or receive communications from other devices. The display unit(s) may be provided, for example, to allow a user to view various information in connection with execution of the instructions. The user input device(s) may be provided, for example, to allow the user to make manual adjustments, make selections, enter data or various other information, and/or interact in any of a variety of manners with the processor during execution of the instructions.

The various methods or processes outlined herein may be coded as software that is executable on one or more processors that employ any one of a variety of operating systems or platforms. Additionally, such software may be written using any of a number of suitable programming languages and/or programming or scripting tools, and also may be compiled as executable machine language code or intermediate code that is executed on a framework or virtual machine.

In this respect, various inventive concepts may be embodied as a computer readable storage medium (or multiple computer readable storage media) (e.g., a computer memory, one or more floppy discs, compact discs, optical discs, magnetic tapes, flash memories, circuit configurations in Field Programmable Gate Arrays or other semiconductor devices, or other non-transitory medium or tangible computer storage medium) encoded with one or more programs that, when executed on one or more computers or other processors, perform methods that implement the various embodiments of the invention disclosed herein. The computer readable medium or media can be transportable, such that the program or programs stored thereon can be loaded onto one or more different computers or other processors to implement various aspects of the present invention as discussed above. In some embodiments, the system comprises cloud-based software that executes one or all of the steps of each disclosed method instruction.

The terms "program" or "software" are used herein in a generic sense to refer to any type

of computer code or set of computer-executable instructions that can be employed to program a computer or other processor to implement various aspects of embodiments as discussed above. Additionally, it should be appreciated that according to one aspect, one or more computer programs that when executed perform methods of the present disclosure need not reside on a single computer or processor, but may be distributed in a modular fashion amongst a number of different computers or processors to implement various aspects of the present invention.

Computer-executable instructions may be in many forms, such as program modules, executed by one or more computers or other devices. Generally, program modules include routines, programs, objects, components, data structures, etc. that perform particular tasks or implement particular abstract data types. Typically, the functionality of the program modules may be combined or distributed as desired in various embodiments.

Also, data structures may be stored in computer-readable media in any suitable form. For simplicity of illustration, data structures may be shown to have fields that are related through location in the data structure. Such relationships may likewise be achieved by assigning storage for the fields with locations in a computer-readable medium that convey relationship between the fields. However, any suitable mechanism may be used to establish a relationship between information in fields of a data structure, including through the use of pointers, tags or other mechanisms that establish relationship between data elements.

Also, the disclosure relates to various embodiments in which one or more methods. The acts performed as part of the method may be ordered in any suitable way. Accordingly, embodiments may be constructed in which acts are performed in an order different than illustrated, which may include performing some acts simultaneously, even though shown as sequential acts in illustrative embodiments.

In some embodiments, the disclosure relates to a system that comprises at least one processor, a program storage, such as memory, for storing program code executable on the processor, and one or more input/output devices and/or interfaces, such as data communication and/or peripheral devices and/or interfaces. In some embodiments, the user device and computer system or systems are communicably connected by a data communication network, such as a Local Area Network (LAN), the Internet, or the like, which may also be connected to a number of other client and/or server computer systems. The user device and client and/or server computer systems may further include appropriate operating system software.

In some embodiments, components and/or units of the devices described herein may be able to interact through one or more communication channels or mediums or links, for example, a shared access medium, a global communication network, the Internet, the World Wide Web, a wired network, a wireless network, a combination of one or more wired networks and/or one or more wireless networks, one or more communication networks, an a-synchronic or asynchronous wireless network, a synchronic wireless network, a managed wireless network, a non-managed wireless network, a burstable wireless network, a non-burstable wireless network, a scheduled wireless network, a non-scheduled wireless network, or the like.

Discussions herein utilizing terms such as, for example, "processing," "computing," "calculating," "determining," or the like, may refer to operation(s) and/or process(es) of a computer, a computing platform, a computing system, or other electronic computing device, that manipulate and/or transform data represented as physical (e.g., electronic) quantities within the computer's registers and/or memories into other data similarly represented as physical quantities within the computer's registers and/or memories or other information storage medium that may store instructions to perform operations and/or processes.

Some embodiments may take the form of an entirely hardware embodiment, an entirely software embodiment, or an embodiment including both hardware and software elements. Some embodiments may be implemented in software, which includes but is not limited to firmware, resident software, microcode, or the like.

Furthermore, some embodiments may take the form of a computer program product accessible from a computer-usable or computer-readable medium providing program code for use by or in connection with a computer or any instruction execution system. For example, a computer-usable or computer-readable medium may be or may include any apparatus that can contain, store, communicate, propagate, or transport the program for use by or in connection with the instruction execution system, apparatus, or device. If a computer program product is used within a system, the system performs a computer-implemented method of selecting a neomer sequence in some embodiments. In some embodiments, the methods are computer-implemented methods of selecting a therapy, analyzing data from a sample or diagnosing a subject comprising:

(a) isolating RNA or DNA from a sample; and

(b) quantifying the presence of neomer in the sample. In some embodiments, the therapy is chosen or the diagnosis is made based opon the presence of the neomer. In some embodiments,

the sample is cell free. In some embodiments, the neomer are chosen from one or a plurality of neomers identified in Table disclosed herein or one or a plurality of sequences that comprise at least about 85%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% sequence identity to any of the sequences identified in Table 1.

In some embodiments, the medium may be or may include an electronic, magnetic, optical, electromagnetic, InfraRed (IR), or semiconductor system (or apparatus or device) or a propagation medium. Some demonstrative examples of a computer-readable medium may include a semiconductor or solid state memory, magnetic tape, a removable computer diskette, a Random Access Memory (RAM), a Read-Only Memory (ROM), a rigid magnetic disk, an optical disk, or the like. Some demonstrative examples of optical disks include Compact Disk-Read-Only Memory (CD-ROM), Compact Disk-Read/Write (CD-R/W), DVD, or the like.

In some embodiments, a data processing system suitable for storing and/or executing program code may include at least one processor coupled directly or indirectly to memory elements, for example, through a system bus. The memory elements may include, for example, local memory employed during actual execution of the program code, bulk storage, and cache memories which may provide temporary storage of at least some program code in order to reduce the number of times code must be retrieved from bulk storage during execution.

In some embodiments, input/output or I/O devices (including but not limited to keyboards, displays, pointing devices, etc.) may be coupled to the system either directly or through intervening I/O controllers. In some embodiments, network adapters may be coupled to the system to enable the data processing system to become coupled to other data processing systems or remote printers or storage devices, for example, through intervening private or public networks. In some embodiments, modems, cable modems and Ethernet cards are demonstrative examples of types of network adapters. Other suitable components may be used.

Some embodiments may be implemented by software, by hardware, or by any combination of software and/or hardware as may be suitable for specific applications or in accordance with specific design requirements. Some embodiments may include units and/or sub-units, which may be separate of each other or combined together, in whole or in part, and may be implemented using specific, multi-purpose or general processors or controllers. Some embodiments may include buffers, registers, stacks, storage units and/or memory units, for temporary or long-term storage of data or in order to facilitate the operation of particular implementations.

Some embodiments may be implemented, for example, using a machine-readable medium or article which may store an instruction or a set of instructions that, if executed by a machine, cause the machine to perform a method steps and/or operations described herein. Such machine may include, for example, any suitable processing platform, computing platform, computing device, processing device, electronic device, electronic system, computing system, processing system, computer, processor, or the like, and may be implemented using any suitable combination of hardware and/or software. The machine-readable medium or article may include, for example, any suitable type of memory unit, memory device, memory article, memory medium, storage device, storage article, storage medium and/or storage unit; for example, memory, removable or non-removable media, erasable or non-erasable media, writeable or re-writeable media, digital or analog media, hard disk drive, floppy disk, Compact Disk Read Only Memory (CD-ROM), Compact Disk Recordable (CD-R), Compact Disk Re-Writeable (CD-RW), optical disk, magnetic media, various types of Digital Versatile Disks (DVDs), a tape, a cassette, or the like. The instructions may include any suitable type of code, for example, source code, compiled code, interpreted code, executable code, static code, dynamic code, or the like, and may be implemented using any suitable high-level, low-level, object-oriented, visual, compiled and/or interpreted programming language, e.g., C, C++, Java™, BASIC, Pascal, Fortran, Cobol, assembly language, machine code, or the like.

Many of the functional units described in this specification have been labeled as circuits, in order to more particularly emphasize their implementation independence. For example, a circuit may be implemented as a hardware circuit comprising custom very-large-scale integration (VLSI) circuits or gate arrays, off-the-shelf semiconductors such as logic chips, transistors, or other discrete components. A circuit may also be implemented in programmable hardware devices such as field programmable gate arrays, programmable array logic, programmable logic devices or the like.

In some embodiment, the circuits may also be implemented in machine-readable medium for execution by various types of processors. An identified circuit of executable code may, for instance, comprise one or more physical or logical blocks of computer instructions, which may, for instance, be organized as an object, procedure, or function. Nevertheless, the executables of an identified circuit need not be physically located together, but may comprise disparate instructions stored in different locations which, when joined logically together, comprise the

circuit and achieve the stated purpose for the circuit. Indeed, a circuit of computer readable program code may be a single instruction, or many instructions, and may even be distributed over several different code segments, among different programs, and across several memory devices. Similarly, operational data may be identified and illustrated herein within circuits, and may be embodied in any suitable form and organized within any suitable type of data structure. The operational data may be collected as a single data set, or may be distributed over different locations including over different storage devices, and may exist, at least partially, merely as electronic signals on a system or network.

The computer readable medium (also referred to herein as machine-readable media or machine-readable content) may be a tangible computer readable storage medium storing the computer readable program code. The computer readable storage medium may be, for example, but not limited to, an electronic, magnetic, optical, electromagnetic, infrared, holographic, micromechanical, or semiconductor system, apparatus, or device, or any suitable combination of the foregoing. As alluded to above, examples of the computer readable storage medium may include but are not limited to a portable computer diskette, a hard disk, a random access memory (RAM), a read-only memory (ROM), an erasable programmable read-only memory (EPROM or Flash memory), a portable compact disc read-only memory (CD-ROM), a digital versatile disc (DVD), an optical storage device, a magnetic storage device, a holographic storage medium, a micromechanical storage device, or any suitable combination of the foregoing. In the context of this document, a computer readable storage medium may be any tangible medium that can contain, and/or store computer readable program code for use by and/or in connection with an instruction execution system, apparatus, or device.

The computer readable medium may also be a computer readable signal medium. A computer readable signal medium may include a propagated data signal with computer readable program code embodied therein, for example, in baseband or as part of a carrier wave. Such a propagated signal may take any of a variety of forms, including, but not limited to, electrical, electro-magnetic, magnetic, optical, or any suitable combination thereof. A computer readable signal medium may be any computer readable medium that is not a computer readable storage medium and that can communicate, propagate, or transport computer readable program code for use by or in connection with an instruction execution system, apparatus, or device. As also alluded to above, computer readable program code embodied on a computer readable signal medium may

be transmitted using any appropriate medium, including but not limited to wireless, wireline, optical fiber cable, Radio Frequency (RF), or the like, or any suitable combination of the foregoing. In one embodiment, the computer readable medium may comprise a combination of one or more computer readable storage mediums and one or more computer readable signal mediums. For example, computer readable program code may be both propagated as an electro-magnetic signal through a fiber optic cable for execution by a processor and stored on RAM storage device for execution by the processor.

Computer readable program code for carrying out operations for aspects of the present invention may be written in any combination of one or more programming languages, including an object oriented programming language such as Java, Smalltalk, C++ or the like and conventional procedural programming languages, such as the "C" programming language or similar programming languages. The computer readable program code may execute entirely on a user's computer, partly on the user's computer, as a stand-alone computer-readable package, partly on the user's computer and partly on a remote computer or entirely on the remote computer or server. In the latter scenario, the remote computer may be connected to the user's computer through any type of network, including a local area network (LAN) or a wide area network (WAN), or the connection may be made to an external computer (for example, through the Internet using an Internet Service Provider).

The program code may also be stored in a computer readable medium that can direct a computer, other programmable data processing apparatus, or other devices to function in a particular manner, such that the instructions stored in the computer readable medium produce an article of manufacture including instructions which implement the function/act specified in the schematic flowchart diagrams and/or schematic block diagrams block or blocks.

Functions, operations, components and/or features described herein with reference to one or more embodiments, may be combined with, or may be utilized in combination with, one or more other functions, operations, components and/or features described herein with reference to one or more other embodiments, or vice versa.

Other embodiments are described in the following non-limiting Examples. Various publications, including patents, published applications, technical articles and scholarly articles are cited throughout the specification. Each of these cited publications is incorporated by reference herein in its entirety.

EXAMPLES

Example 1. Neomers for cancer detection

1. Methods

i. Computational characterization of nullomers

The GRCh38 reference assembly of the human genome was used throughout the study. Nullomer extraction was performed for kmer lengths up to 17 base pairs using the algorithm described in [19]. By definition, the reverse complement of a nullomer will also be a nullomer. Throughout this manuscript when counting nullomers, the reverse complement of nullomer i was also considered separately, unless i is a palindrome. Substitutions and indels identified from WGS of tumor samples from 2,577 individuals across 21 tissues were obtained from https://dcc.icgc.org/releases/PCAWG/ 22. Recurrent nullomers (neomers) (ri) were annotated as those that resulted from substitutions or indels across two or more patients within a cancer type. When possible, ri was chosen to get ~10,000 neomers from each tissue, otherwise it was set to 2 (Table 4). Driver mutation-derived neomers were defined as nullomers detected from driver mutations and were identified using the database (REF).

ii. Classification of tumor tissue of origin using neomers

We trained a classifier to distinguish tissue of origin for a cancer sample based on observed neomers using the libSVM package for Julia [84] with default parameters to train a support vector machine classifier with a linear kernel. We used 10-fold cross validation whereby the classifier was evaluated on a held out fraction of the data. The set of neomers for each cancer type was recalculated for each round to only include the patients in the training set.

iii. Supervised selection of nullomers

The MMR status of each biopsy sample was derived from [85]. The model was trained on neomers identified in MSI samples and the performance of the algorithm evaluated. For the MSI versus the MSS samples, we counted the number of neomers that contained either AAAAAAAA or TTTTTTTT repeats, since MSI cancers have been associated with mutations of polyA/T repeats [35]. The threshold for determining MSI or MSS was set as the harmonic mean of the maximum number of counts in the MSS set and the minimum number of counts in the MSI set. The POLE deficiency status of each biopsy sample was derived from [85] and we used a similar strategy to that of MMR status, but instead counted neomers created through either a TCT>TAT

or TCG>TTG mutation. Since the number of patients in each category was limited, we only used a 5-fold cross validation. For ovarian cancer analyses, two sets of nullomers were used. From the nullomers of length k=13 identified in [19], we retained only the ones that could not be created by any of the single base pair substitutions identified in gnomAD v2 43. Similarly, we selected all nullomer of length k=15 of order 1, i.e. nullomers that can only be created from the human reference genome through at least two insertions, deletions, or substitutions.

iv. cfDNA extraction and WGS

Prostate cancer samples, described in [86], were obtained from the Witte lab at UCSF as extracted cfDNA kept at -80C. Extracted dsDNA concentration was measured using the Qubit High-Sensitivity dsDNA kit. For ovarian and lung cancer samples, cfDNA was extracted from 1mL of plasma, following centrifugation at 4C at 600rpm for 3 minutes, to remove larger debris, using the QIAamp Circulating Nucleic Acid Kit (Qiagen). cfDNA was eluted in 50uL of elution buffer and measured using the Qubit High-Sensitivity dsDNA kit, and validated for size distribution (160-180bp) using an Agilent BioAnalyzer 2100 Sensitivity DNA chip. Up to 10ng of cfDNA was used for sequencing library construction using the library preparation enzymatic fragmentation kit 2.0 (Twist Bioscience) adjusted with IDT's xGen UDI-UMI 96 barcodes system (IDT) to replace the Twist universal adapter, and using KAPA HiFi polymerase instead of the polymerase provided by the kit. Libraries for lung, ovarian and their respective controls were sequenced as PE150 using a NovaSeq 6000 S4 system (Illumina) leveraging UMIs to remove PCR duplicates, aiming for x15-20 coverage per multiplexed library, via Novogene. Prostate cancer samples and their controls were sequenced on an Illumina HiSeq 4000 machine PE150, pooled across 3 lanes, at the UC Davis DNA core facility.

v. Genomic DNA extraction from solid tumor samples

Frozen solid tumor samples from both ovarian and prostate cohorts were received from the Chapman and Witte labs respectively. Tumor masses were kept in dry ice and excised on a frozen tray, to yield out pieces of 1-8 mm^3, and gDNA was extracted using the DNeasy Blood and Tissue kit (Qiagen), according to manufacturer instructions. gDNA concentrations were measured using NanoDrop.

vi. Neomer identification in cfDNA samples

To improve the detection of lung cancer mutations in cfDNA, we used neomers that are present in >=2 patients rather than >=3 as was used for the tissue-of-origin classifier. To filter

out common population variants, we obtained variant information from the gnomAD v2[43] and excluded all neomers that were generated due to variants with a frequency >0.05. Variants that were not single base pair substitutions were not considered. The FASTQ files were scanned for neomers by searching for exact matches to the 16-mers of interest. To reduce errors, only reads where the nullomer was found in both pairs were kept. The remaining reads were then mapped to the reference genome (hg38) using Bowtie2[87] with parameters "--mp 3". The aligned reads were then analyzed using samtools mpileup and only those loci that contained $>C/1.5$ read pairs, where C is the coverage for the sample, were counted as bona fide nullomers. For the analysis of the more deeply sequenced prostate samples (**Fig. 3G**), we adjusted the threshold to $C/10$. When processing the data, reads were not mapped to the genome, which means that the coverage estimates are based on the total length of the sequenced reads rather than on the mappable reads and should thus be considered an upper bound. To evaluate the ability to detect cancer from cfDNA, we first randomly split samples into training and validation sets. Prior to the split, samples were randomly removed from either the controls or patients to ensure that the two groups were of equal size. From the training set we determined a threshold, t = max (number of neomers in control samples from the training set) - 1. Samples in the validation set were then designated as having cancer if the number of neomers detected was $>=t$. To calculate F1 scores, this procedure was repeated 100 times. For the downsampling analysis, each read pair in sample i was retained with a probability of c/C_i, where c is the target coverage and C_i is the full coverage of the sample, i.e. sampling without replacement. For each value of c the downsampling was repeated 100 times.

vii. Promoter luciferase assays

Promoter sequences with and without the neomer (Table 7) were synthetically generated and cloned into the modified Promega promoter assay luciferase vector pGL4.11b (a gift from Dr. Rick Myers, HudsonAlpha) by BioMatik Inc and Sanger sequence verified. LNCaP cells were plated at an initial density of 2*10^5 cells/well in 24-well tissue culture plates and maintained in RPMI medium, 10% FBS supplemented with L-Glutamine and Penicillin/Streptomycin. Plasmids together with a renilla expressing plasmid, pGL4.74 (Promega), at a ratio of 10:1 luciferase:renilla were transfected using the X-tremeGENE™ HP DNA Transfection Reagent (Roche) using 1:4 ratio of DNA (ug) to reagent (ul). 72 hours post transfection luciferase and renilla levels were measured using the Dual-Luciferase Reporter

Assay System (Promega) following the manufacturer's protocol using a GloMax Explorer Multimode Microplate Reader (Promega). Luciferase activity was normalized to renilla levels and presented as Relative Luciferase Units (RLU). Statistical analysis was performed using Prism version 9.0.2 (GraphPad). All values were reported as means (AVG) and standard errors (SE). p values < 0.05 were considered statistically significant.

viii. Software availability

We generated an easy to use software package that enables performing nullomer cancer analyses from sequence-based datasets. The package is composed of six functions: 1) EnumerateNullomers, which extracts all nullomers of specified kmer lengths in a FASTA sample; 2) ExtractMutationNullomers, which finds all mutations that cause the resurfacing of a list of nullomers; 3) IdentifyRecurrentNullomers, which identifies nullomers that recur in a dataset through mutagenesis; 4) FindAlmostNullomers, which identifies the positions that can create a list of nullomers genome-wide for every possible substitution and single base-pair insertion and deletion; 5) FindNullomerVariants, which removes nullomers that are likely to result from common variants in a user specified variant VCF file; 6) FindDNANullomersFromReads, which performs the identification of nullomers in raw read samples. The package can be found at: https://github.com/Ahituv-lab/Nullomerator and a readthedocs tutorial provides in depth details on how to run the software functions.

ix. Comparison to validated neoantigens

We downloaded a list of 1,967 validated neoantigens from http://biopharm.zju.edu.cn/download.neoantigen/iedb_validated.zip. Requiring both predicted strong binding and a positive validation, provided 1,700 neoantigens. To evaluate the enrichment of neoantigens corresponding to neomers, we assumed a hypergeometric distribution with 1,700 draws from an urn with 188,659 white balls (total number of neomers) and 186,067,892 black balls (number of nullomers found with lower recurrency than what was required for neomers).

x. Massively parallel reporter assays

Lentiviral bound MPRA was done as described previously [60]. An oligonucleotide library of 230bp long fragments bearing 1) 4,609 loci of recurrent mutations across prostate cancer patients, which cause neomer resurfacing and their reference genome pair; 2) 64 fragments to tile the 350bp long TMEM127-CIAO1 and RPS2-SNHG9 promoters used in the luciferase assay; as

well as 3) 100 scrambled control sequences. A pool of 230-nt oligos containing each of these 200-nt sequences flanked by 15-nt primer recognition sequences was synthesized (SurePrint Oligonucleotide Libraries; Agilent Technologies, Santa Clara, CA), amplified, and cloned into the MPRA plasmid. Briefly, amplified inserts were cloned upstream of a minimally active promoter (mP) expressing a 5'UTR barcoded (BC) eGFP reporter transcript. We aimed for ~100 BCs per sequence when constructing and extracting the plasmid library. Each insert was associated with a set of BCs using paired-end (PE150) customized NGS (see Gordon et al. Step 83 [60]). Lentivirus was produced and titered, later to be used to infect prostate cancer cell line (LNCaP, DU-145 and PC-3) at an MOI of 50 virus particles per cell. DNA and RNA were extracted from the 3 replicates of the cell culture experiment used for library construction and multiplexed for NGS. All MPRA-related sequencing was performed using an Illumina NextSeq 500 (Novogene) with either PE150 for the CRS-BC association library or PE15 for the DNA/RNA BC count portion of the protocol.

2. Annotation of mutations that lead to nullomers

As cancer is associated with a large number of somatic DNA mutations, we investigated if they can result in the resurfacing of nullomers (**Fig. 1A**). Using our previously characterized human nullomers[19], we analyzed WGS results from 2,577 patients across 21 different cancer types from TCGA22 for resurfacing nullomers (**Fig. 5A**). We focused on 16bp nullomers, as it is the shortest length where we detect a sufficient number of nullomers per patient, with the human reference genome having only 37.24% of all possible 16mers. The majority of the 44,599,472 single nucleotide substitutions give rise to multiple nullomers, allowing us to identify 213,164,038 resurfacing nullomers across all cancer types. Furthermore, we identified 2,470,091 nullomers resulting from short insertions and deletions (1-100 base pairs). The median number of nullomers created by each substitution was two and for indels four (**Figs. 5B-C**). On average, 58.29% of substitutions in a patient resulted in one or more nullomers, with 2.1% of the nullomers residing in coding regions. The median number of nullomers found across cancer patients was 9,107 (**Figs. 1B-C**) and the number of nullomers was directly proportional to the number of mutations (**Fig. 5D**). As mutations were identified by comparing to healthy tissues, we did not filter for common variants that could result in nullomers[19].

To further prioritize nullomers that could be used as cancer biomarkers, we focused on the subset of nullomers that are recurrent, i.e. those found in more than one patient for a specific

cancer type, termed hereafter as neomers. The number of neomers was proportional to the total number of mutations (**Fig. 1C**, Table 4). As both the number of patients per cancer type and the mutational load varied, the median number of neomers for each tissue type ranged from 0-98. Analysis of the most frequent neomers revealed several previously known cancer-associated mutations (Table 1). For example, some of the most recurrent coding neomers were the result of either the Gly12Asp, Gly12Val or Gly12Cys missense mutation in the KRAS proto-oncogene GTPase (KRAS), which are known to make up 80% of cancer-associated KRAS mutations and lead to KRAS being constitutively active[23,24]. Although KRAS mutations have been associated with several cancer types, 190/215 (88%) of these mutations were found in pancreatic cancers. Several frequently occurring coding neomers were also found in other known cancer-associated genes, such as tumor protein p53 (*TP53*), B-Raf proto-oncogene serine/threonine kinase (*BRAF*) and phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha (*PIK3CA*). The most frequent neomer was located in a noncoding region, within the telomerase reverse transcriptase (*TERT*) promoter, which is known to be associated with numerous cancer types[25]. This mutation, called -124C>T or C228T, is extremely common in numerous cancer types[26] and is thought to disrupt a G-quadruplex[27] leading to the binding of GAPB[28], an ETS transcription factor, resulting in increased *TERT* expression. We found this mutation in 97 patients, with the highest incidence in glioblastoma (51%), fitting with its known high prevalence rate and diagnostic use for this cancer type[29].

Table 1. Common cancer-associated neomers. Six of the most common neomers created by a single mutation. The nucleotide in red is the neomer causing mutation.

Neomer	Locus	Coordinate (hg38)	Mutation	Name	Patient #	Cancer #
AGGGCCCGGAAGGGGC	*TERT*	chr5:1295113	G>A	-124C>T C228T	97	7
TCTTGCCTACGCCATC	*KRAS*	chr12:25245350	C>T	c.35G>AGly12Asp	89	6
CTGTTGGCGTAGGCAA	*KRAS*	chr12:25245350	C>A	c.35G>TGly12Val	79	6
ACGCCACGAGCTCCAA	*KRAS*	chr12:25245351	C>G	c.34G>T Gly12Cys	47	1
GGTGCATGTTTGTGCC	*TP53*	chr17:7673802	C>T	c.818G>A Arg273Pro	35	13
GTGGGGGCAGTGCCTC	*TP53*	chr17:7675088	C>T	c.524G>A Arg175His	28	11

Table 4. Minimal recurrency thresholds and associated number of neomers per tissue type.

Cancer tissue	Recurrency threshold	Number of neomers	Number of patients
Biliary	2	166	262
Bladder	2	3,556	23
Bone_SoftTissue	2	2,018	89
Breast	2	27,280	211
CNS	2	7,238	287
Cervix	2	172	20
Colorectal	4	4,526	52
Esophagus	5	16,278	97
Head_neck	2	10,194	56
Kidney	2	17,584	186
Liver	3	4,756	336
Lung	3	3,028	84
Lymphoid	4	12,441	197
Myeloid	2	28	38
Ovary	2	13,744	110
Pancreas	2	61,557	313
Prostate	2	5,270	199
Skin	7	1,534	107
Stomach	3	1,832	68
Thyroid	2	60	48
Uterus	2	8,008	44

We also identified several neomers that are frequently created by different mutations (Table 5). Interestingly, some of these frequently recurrent neomers are created by different mutations, yet are predominantly found in one cancer. For example, GTTTTTCTCCTAGACC is found 40 times in skin cancer at 31 different loci while CTGGCAGTGAGCCACG is found 21 times in liver cancer across 18 loci. The majority (98%) of these frequent neomers reside in noncoding regions, and many of them reside in intronic regions (35%). For example, CGACGTTCTGCCCACT is found in 32 loci, primarily in pancreatic and stomach cancer. Of those loci, 21/32 (65.6%) were found in noncoding regions nearby pancreatic cancer associated genes. These include, for example, the C-C motif chemokine ligand (CCL4)[30], the POM121 transmembrane nucleoporin like 12 (POM121L12), which is commonly mutated in gastrointestinal cancers[31] and the potassium voltage-gated channel modifier subfamily V member 1 (KCNV1) gene, where promoter hypermethylation has been associated with both pancreatic[32] and esophageal cancer[33].

Table 5. Frequent cancer-associated neomers. Five of the top frequently recurring neomers created by several different mutations.

Neomer	Loci #	Patient #	Cancer #	Cancer type
CGACGTTCTGCCCACT	32	74	7	Mainly pancreatic and stomach
GTTTTTCTCCTAGACC	32	41	2	All but one in skin
CTGCAGTGGCGCAATA	30	30	11	A third of the mutations in colorectal
TTATAGGGGTCCAGTG	25	26	1	Colorectal only
CTGGCAGTGAGCCACG	23	26	2	21 in liver, 5 esophagus

To further validate that our neomers identify cancer-associated mutations, we examined whether neomers are enriched in known cancer driver mutations. Since driver mutations are recurrent, we define as driver mutation-derived neomers the neomers detected at driver mutation loci. We annotated a cohort of driver mutations from 29 different cancer types for their overlap with neomers. In the pan-cancer analysis, we identified 19,594,212 neomers resulting from 50,167 putative driver mutations (**Fig. 1**). For specific cancer types, we found that on average, 81% of driver mutations resulted in one or more neomers, ranging between 63.88% and 86.50%

in pancreatic and lung squamous adenocarcinoma respectively. The number of neomers also varied by cancer type, ranging between 92 in pancreatic cancer and 9,434 in cutaneous melanoma (**Fig. 1D**). In general, driver mutations were 1.4-fold more likely to result in the creation of a neomer. Taken together our results suggest that a large subset of clinically relevant cancer mutations are associated with nullomers.

3. Generation of cancer subtype neomer classifier

We next set out to assess whether neomers can be used to distinguish between cancer types. We filtered neomers by keeping only those that appeared $>=r_i$ times in specific cancer type *i* (Table 4). Comparison of the set of neomers associated with each cancer type revealed a small overlap, as indicated by the Jaccard index which is <0.04, suggesting that each cancer type has a distinct neomer signature (**Fig. 1D**). We also counted the number of times neomers are found in each patient, finding that patients are strongly enriched for only one set of cancer specific neomers (**Fig. 1E**).

We then tested if our annotated cancer-specific neomers can be used to classify tumor samples. We trained a support vector machine classifier to identify tumor type. The classifier takes as input a 21-dimensional vector indicating the number of neomers found for each cancer specific set. Evaluation using 10-fold cross-validation, revealed that our classifier achieves both high sensitivity and specificity, with an F1 score of 0.92 and an accuracy of 0.99 (**Fig. 2A-B**). Performance was better than a recent deep learning model[21] and also requires less computational resources to train.

4. Neomers can distinguish additional cancer features

We next tested whether a hand crafted feature selection approach, i.e. using neomers that are thought to be informative based on prior biological knowledge, would improve performance. For this approach, we initially utilized microsatellite unstable (MSI) and microsatellite stable (MSS) cancers. MSI is associated with better cancer prognosis, increased benefits from surgery and higher sensitivity to immunotherapy, but with a lack of efficacy from adjuvant treatment[34]. Since MSI cancers are associated with mutations at polyA and polyT stretches[35], we hypothesized that neomers containing these motifs would be able to effectively distinguish these two cancer types. We identified ten MSI samples from a cohort of 560 breast cancers[36] and compared them to ten randomly selected MSS samples from the same cohort. We found that the polyA/T neomers were able to separate the two categories with an accuracy of 100% (**Fig. 2C**).

We next applied a similar strategy to distinguish patients with DNA polymerase epsilon catalytic subunit (POLE) deficiency, as these tumors are known to respond more favorably to immune checkpoint inhibitors[37–39]. We identified 25 patients from the TCGA dataset labeled as POLE deficient, and searched for neomers created through a TCT>TAT or TCG>TTG mutation, which are the most common types of mutations in this context[39]. Comparing against POLE proficient tumors, we found that the number of neomers identified for each group have very little overlap (**Fig. 2D**), and the classifier achieved an accuracy of 96%.

5. Neomers detect cancer in cfDNA

We next tested whether neomers could be used to diagnose cancer in cfDNA. We first used lung cancer as our test case, due to having an ample amount of cancer-associated neomers (Table 4), difficulty to diagnose using current techniques, with early stages being mostly asymptomatic[40,41], and being the leading cause of cancer-associated deaths in the world[40,42]. We initially excluded all neomers that could arise due to common germline variants[19]. For each lung cancer associated neomer, we characterized all possible single nucleotide substitutions in the reference genome that could give rise to this neomer. By intersecting this list of neomer creating substitutions with known germline variants identified by the gnomAD project[43], we calculated the probability that each neomer will be present in an individual. We excluded all neomers that are found in the population with $p>0.05$, providing 153,026 lung cancer associated neomers for subsequent cfDNA analyses.

We assembled a control cohort of cfDNA from the plasma of 381 individuals, of which, 180 are considered cancer free(e.g. “healthy”, any cancer detected >2 years after sampling), and 201 individuals were diagnosed with lung cancer. Lung cancer patients represented all stages - 7 stage I (36% of cancer cases), 36 stage II, 34 stage III and 53 stage IV). The majority of our samples were adenocarcinoma (50.2% of cancer samples) and squamous cell carcinoma (37.8%) types. As for risk factors among our cancer patient cohort, the median age was 64+/- 7.8 years old, with patients as young as 42 years old, while the healthy patients had a median age of 58+/- 8.1 years, and 90% are in the USPTF age interval for early screening for lung cancer. Of the healthy patients 12.7% were active smokers (37.2% were noted as unknown), while 64.2% of the cancer patients were noted as active or pass smokers (complying with USPTF guidelines for lung cancer screening)(Table 6). These samples were sequenced at a genomic depth of 10-20X coverage (C) and analyzed for neomers. To reduce the impact of sequencing errors, we only

considered neomers that were found on both read-pairs, resulting in 18,090-55,409 read-pairs per sample. To further avoid spurious neomers, reads were next mapped to the reference human genome, and only neomers that had an overlap of more than C/1.5 read-pairs from different sequence pairs were used for subsequent analyses. Of note, an important computational advantage compared to conventional mutation calling pipelines is that only a single pass is made across the reads to identify those containing neomers, and only those reads are aligned to the reference genome. We identified a median neomer number of 396 for the controls and 639 for the cancer patients (p-value<0.005, Mann-Whitney U) (**Fig. 3A**). Using these data, we trained a classifier that compares the number of detected neomers to a threshold, achieving an F1-score of 0.82 using 2-fold cross validation (**Fig. 3B**). Importantly, our classifier also performs well for early stages with only a slight drop in performance for stage I.

Table 6. Neomers resulting in neoantigens.

Gene	Mutation	Number of Patients	Validation	Neomers
EP300	D1399N	3	Positive	AACACTATTGAGGTAA, ACTATTGAGGTAAGAT,ATCTTACCTCAATAGT, GAACACTATTGAGGTA,TACCTCAATAGTGTTC, TTACCTCAATAGTGTT
BRAF	L597Q	2	Positive	AGGTGATTTTGGTCAA,ATTTTGGTCAAGCTAC,GTAGCTTG ACCAAAAT,TAGGTGATTTTGGTCA,TGACCAAAATCACCTA ,TGTAGCTTGACCAAAA,TTGACCAAAATCACCT,TTTTGGTC AAGCTACA
TP53	H193R	8	Positive	AAGACGCTGAGGAGGG,ACGCTGAGGAGGGGCC,ACTCGG ATAAGACGCT,AGACGCTGAGGAGGGG,AGCGTCTTATCCG AGT,CACTCGGATAAGACGC,CAGCGTCTTATCCGAG,CCAC TCGGATAAGACG,CCCCTCCTCAGCGTCT,CCCTCCTCAGCG TCTT,CCTCAGCGTCTTATCC,CGGATAAGACGCTGAG,CGTC TTATCCGAGTGG,CTCAGCGTCTTATCCG,CTCGGATAAGAC GCTG,GCGTCTTATCCGAGTG,GGATAAGACGCTGAGG,GGC CCCTCCTCAGCGT,GTCTTATCCGAGTGGA,TCAGCGTCTTA TCCGA,TCCACTCGGATAAGAC, TCGGATAAGACGCTGA
CTNNB1	G34R	3	Positive	AATGGATTCGAGAGTC,ACCAGAATGGATTCGA,ACCTGGA CTCTCGAAT,ATTCGAGAGTCCAGGT,CACCAGAATGGATTC G,CCAGAATGGATTCGAG,CCTGGACTCTCGAATC,CGAATC CATTCTGGTG,CGAGAGTCCAGGTAAG,CTCGAATCCATTCT GG,CTTACCTGGACTCTCG,GACTCTCGAATCCATT,GATTCG AGAGTCCAGG,TACCTGGACTCTCGAA,TCGAATCCATTCTG GT,TCGAGAGTCCAGGTAA,TTACCTGGACTCTCGA,TTCGA GAGTCCAGGTA
TP53	G266V	4	Positive-High	AAGCTGTTCCGTACC,AAGCTGTTCCGTACCA,AATCTACTG GTACGGA,ACCAGTAGATTACCAC,ACTGGTACGGAACAGC, AGCTGTTCCGTACCAG,ATCTACTGGTACGGAA,CAAAGCT

Gene	Mutation	Number of Patients	Validation	Ncomers
				GTTCCGTAC,CCGTACCAGTAGATTA,CGTACCAGTAGATTA C,CTACTGGTACGGAACA,CTGGTACGGAACAGCT,CTGTTC CGTACCAGTA,GCTGTTCCGTACCAGT,GGTAATCTACTGGT AC,GGTACGGAACAGCTTT,GTAATCTACTGGTACG,GTACC AGTAGATTACC,GTACGGAACAGCTTTG,GTGGTAATCTACT GGT,GTTCCGTACCAGTAGA,TAATCTACTGGTACGG,TACG GAACAGCTTTGA,TACTGGTACGGAACAG,TCAAAGCTGTT CCGTA,TCCGTACCAGTAGATT,TCTACTGGTACGGAAC,TG GTACGGAACAGCTT,TGTTCCGTACCAGTAG,TTCCGTACCA GTAGAT
TP53	R337C	7	Positive-High	AACATCTCGAAGCACT,AAGCACTCACGCCCAC,ACATCTC GAAGCACTC,ACTCACGCCCACGGAT,AGCACTCACGCCCA CG,AGTGCTTCGAGATGTT,ATCCGTGGGCGTGAGT,ATCTC GAAGCACTCAC,CACTCACGCCCACGGA,CATCTCGAAGCA CTCA,CCGTGGGCGTGAGTGC,CGTGAGTGCTTCGAGA,CGT GGGCGTGAGTGCT,CTCGAAGCACTCACGC,GAACATCTCG AAGCAC,GAAGCACTCACGCCCA,GAGTGCTTCGAGATGT,G CACTCACGCCCACGG,GCGTGAGTGCTTCGAG,GGAACATC TCGAAGCA,GGCGTGAGTGCTTCGA,GTGAGTGCTTCGAGA T,GTGCTTCGAGATGTTC,GTGGGCGTGAGTGCTT,TCCGTGG GCGTGAGTG,TCGAAGCACTCACGCC,TCTCGAAGCACTCA CG,TGAGTGCTTCGAGATG,TGCTTCGAGATGTTCC,TGGGC GTGAGTGCTTC
TP53	C141Y	4	Positive-Intermediate	AAGACCTACCCTGTGC,CAGGGTAGGTCTTGGC,CTGCACA GGGTAGGTC,GACCTACCCTGTGCAG,GCACAGGGTAGGTC TT,GCCAAGACCTACCCTG
TP53	C135F	4	Positive-High	GTCTTGGCCAGTTGGA,TCCAACTGGCCAAGAC
ERBB2	D769Y	3	Positive-Intermediate	AAATCTTATACGTAAG,AAGAAATCTTATACGT,AATCTTAT ACGTAAGC,ACGTATAAGATTTCTT,AGAAATCTTATACGTA ,AGGGGCTTACGTATAA,ATACGTAAGCCCCTCC,ATCTTAT ACGTAAGCC,CTTACGTATAAGATTT,CTTATACGTAAGCCC C,GAAATCTTATACGTAA,GAGGGGCTTACGTATA,GCTTAC GTATAAGATT,GGAGGGGCTTACGTAT, GGCTTACGTATAAGAT. GGGCTTACGTATAAGA, GGGGCTTACGTATAAG, TACGTAAGCCCCTCCA,TACGTATAAGATTTCT,TATACGTA AGCCCCTC,TCTTATACGTAAGCCC,TGGAGGGGCTTACGTA ,TTACGTATAAGATTTC,TTATACGTAAGCCCCT
TPR	S2155L	2	Positive	AACCTGCGGCAAACTA,ACCTGCGGCAAACTAA,AGCAACC TGCGGCAAA,ATTTAGTTTGCCGCAG,CATTTAGTTTGCCGC A,CCATTTAGTTTGCCGC,CCTGCGGCAAACTAAA,CGGCAA ACTAAATGGA,CTGCGGCAAACTAAAT,GCAACCTGCGGCA AAC,GCGGCAAACTAAATGG,GGCAAACTAAATGGAC,GTC CATTTAGTTTGCC,GTTTGCCGCAGGTTGC,TAGTTTGCCGC AGGTT,TCCATTTAGTTTGCCG,TGCGGCAAACTAAATG,TTA GTTTGCCGCAGGT,TTTAGTTTGCCGCAGG,TTTGCCGCAGG TTGCT
ASH2L	E617K	2	Positive	ATCACGTGGAGACAAA,TATCACGTGGAGACAA,TTGTCTC CACGTGATA,TTTGTCTCCACGTGAT
ANKLE	E103Q	2	Positive	CAGAAAAAATTGGCTC,GAGCCAATTTTTTCTG

Gene	Mutation	Number of Patients	Validation	Neomers
2				
DNMT1	E432K	2	Positive	ATCGAGAAGAATATCA,CGAGAAGAATATCAAA,GAAGAGTTTGATATTC,GAATATCAAACTCTTC,TCGAGAAGAATATCAA,TGATATTCTTCTCGAT,TTGATATTCTTCTCGA,TTTGATATTCTTCTCG

We next set out to test how our neomer classifier performs with a low amount of tumor-associated neomers. We focused on ovarian cancer which only has 12,788 associated neomers (amongst our analyzed TCGA WGS datasets that are unlikely to be present due to germline variants (Table 4). In addition, as ovarian cancer does not have an effective screening test44, most women are diagnosed at stages IIIC and IV, when 5-year survival rates are 39% and 17%, respectively45, urgently needing the development of novel early detection methods. We obtained plasma samples from X patients and Y controls, and sequenced them at a genomic depth of 10-20X coverage. We carried out similar analyses to our lung cancer patients, and found that the differences between cases and controls was modest, with a median of 10 neomers in the patients and 6 in the controls (p-value=0.11, Mann-Whitney U) (**Fig. 3C**). Consequently, our classifier achieved an F1 score of 0.57. We next used a supervised strategy, where we searched for 62,272 nullomers of length 13 that could not be created through any known germline variant obtained from [19]. We hypothesized that although these nullomers may not be specific to ovarian cancer, they would be helpful in distinguishing cancer from non-cancer. Encouragingly, we observed a larger difference between cases and controls with nullomer medians of 44 and 31, respectively (p-value<0.03, Mann-Whitney U) (**Fig. 3D**), and an F1 classifier score of 0.66. Next, we used a set of 2,502,376 15 base pair, first order nullomers (also obtained from [19]), k-mers that cannot be created through a single base pair substitution or indel, assuming that these would also be unlikely to be found in healthy cfDNA samples. We found only a small fraction of these nullomers in our cfDNA samples, with cases having a median of 30 compared to 17 in controls (p-value<0.01, Mann-Whitney U) (**Fig. 3E**), resulting in an F1 score of 0.69. Taken together, these results suggest that our strategy requires a large number of neomers to be successful, but the use of nullomers that are rare in healthy samples could increase the ability to distinguish cfDNA from cancer patients.

We next set out to assess the correlation between neomers detected from cfDNA and tumors from the same patient. Due to sample availability, we applied our approach to localized prostate cancer. Prostate cancer is the fifth leading cause of death worldwide46 and the current primary screen for this cancer, measuring blood levels of the prostate-specific antigen (PSA), has a high false negative and false positive rates47. In addition, localized prostate cancer has a low abundance of ctDNA making it difficult to detect by ultra low pass WGS or targeted cfDNA sequencing[48] or via methylation[49] compared to metastatic[50], providing a challenging test for our neomer approach. The number of neomers we observed for this cancer was on the low end (N=4,621; median per patient=29.5) from all 21 analyzed tissues (**Fig. 1B**, Table 4). We generated cfDNA WGS datasets from twelve controls and eight localized prostate cancer patients. Searching for 4,621 neomers, we identified a median of 4 in the controls and 8 in the patients (p-value=0.069, Mann-Whitney U) (**Fig. 3F**). Even though the difference is small, the neomers can serve as a sensitive and specific indicator for this type of cancer, as the classifier achieved an F1 score of 0.67. For the eight cases, we also had available tumor samples, and generated WGS for them. As these samples were sequenced more deeply, at 14-40X, we adjusted the threshold for the number of reads required to identify neomers (see Methods). With the relaxed threshold, we observed a similar number of neomers in both datasets (30-40). For 7/8 samples the majority of neomers detected in the tumor were also found in the cfDNA and for 4/8 samples the highest overlap was between the matching pairs (**Fig. 3G**). Combined, these results suggest that neomers could provide a powerful tool for cfDNA cancer diagnosis.

6. Neomers alter promoter activity

Only a small number of mutations in gene regulatory elements that affect gene expression have been found to be associated with cancer[22,51,52]. Since driver mutations are enriched for neomers, we hypothesized that neomer creating mutations in noncoding regions would be more likely to have functional consequences. Of note, the top patient recurrent neomer was in the TERT promoter (Table 1), which is associated with numerous cancers[25]. Focusing on prostate cancer, we selected five neomers for luciferase reporter assays using the following criteria: i) neomers that reside in a promoter based on ENCODE annotations[53]; ii) the gene regulated by the promoter is associated with prostate cancer. Our list included neomers in: 1) a promoter between two divergent genes, RPS2 and the lncRNA gene SNHG9 (**Fig. 4A**), both of which are overexpressed in prostate cancer[54]; 2) a promoter between two divergent genes, TMEM127 and CIAO1 (**Fig. 4B**), with the

former being downregulated in prostate cancer[55]; 3) a promoter between two divergent genes, TTC23 and LRRC28, with the former showing aberrant splicing that relates to therapy resistance in prostate cancer cells[56]; 4) The promoter of GNAI2, a protein that interacts with CXCR5, which positively correlates with prostate cancer progression[57]; 5) A promoter between two divergent genes, PRICKLE4 and FRS3, with the latter thought to affect malignant but not benign prostate cells[58].

We cloned the promoter sequences with and without the neomer into a luciferase promoter assay vector and compared their activity in androgen-sensitive human prostate adenocarcinoma cells (LNCaP). For two out of the five assayed promoters, we observed a significant effect on reporter activity (**Fig. 4C**). For the RPS2-SNHG9 promoter, the neomer led to significantly increased activity, in line with this gene being overexpressed in cancer[54]. For the TMEM127-CIAO1 promoter, the neomer completely abolished activity, fitting with its observed downregulation in prostate cancer[55]. Combined, our experimental results show that neomers could have a significant effect on promoter activity and could potentially be used to identify cancer associated gene regulatory mutations.

7. Neomers alter enhancer activity

We next set out to test whether neomers could be used to identify driver mutations in enhancers. We overlapped prostate cancer associated neomer generating mutations with annotated gene regulatory elements in a prostate cell line obtained from the ENCODE consortium[53,59], identifying 4,609 loci. To test their regulatory effect in a high throughput manner, we used a lentivirus-based MPRA (lentiMPRA) to test their regulatory activity with and without the neomer (**Fig. 5A**). Two hundred base pair sequences, where the position of the neomer is used as a center, were synthesized and cloned upstream of a minimal promoter followed by a GFP reporter gene. Lentivirus was generated and the library was infected using three technical replicates into LNCaP cells and RNA and DNA barcodes were sequenced to determine regulatory activity, as described in Gordon et. al[60]. We observed a good correlation between RNA (Pearson 0.72-0.75; **Fig. 5A**) and DNA barcode across all three technical replicates (Pearson = 0.89-0.91; **Fig. 5A**). Out of the 4,609 loci, 567 showed significant enhancer activity (RNA/DNA ratio >= 1.5). Amongst them, 122 NEO and WT sequence pairs showed a difference in enhancer activity, indicated by the ratio of NEO to WT activity levels (See Methods for more) due to the neomer, 59 pairs (1.28% of loci) with increased activity and 63 (1.36% of loci) with decreased activity.

We next characterized the potential target genes of these differentially active neomer sequences. We assigned target genes USING... and ran gene ontology (GO) enrichment analysis to these sequences using g:GOst tool from the g:Profiler suite61. For the neomer sequences leading to reduced regulatory activity we found...(**Fig. 5B**) GO analysis of sequences leading to increased activity found enrichment for terms related to cell-to-cell junctions and gamma-catenin binding (GO:0045295) (**Fig. 5C**), which is important in cell-cell adhesion and has been associated with prostate cancer progression through interaction with the beta-catenin Wnt signaling axis 62. The neomer that showed the highest level of increased activity compared to reference (4.09 fold) is located in an intron of the catenin alpha 1 (CTNNA1) gene. This gene is a core member of the cadherin/catenin complex and is involved in the regulation of the Wnt/beta-catenin pathway, which has been widely studied in cancer[63], including prostate cancer64, and is the target of several therapies[65,66]. TFBS analysis of the neomer found that it leads to a gain of a TCF7L1 motif and loss of a STAT2 TFBS (**Fig. 5D**), both of which were shown to play a role in prostate cancer malignancy[67,68]. A neomer residing in the 3rd intron of HERC3 gene resulted in 2.7 fold down-regulation of the reporter activity in our MPRA. HERC3 gene had been shown to inhibit metastasis of colorectal cancer and was indicated to be downregulated in colorectal cancer and its downregulation showed poor overall survival (OS) and disease-free survival (DFS) in colorectal patients (Zhang et al. 2022). Taken together, these results demonstrate that neomers could be utilized to identify gene regulatory driver mutations in cancer.

8. Discussion

Cancer is a DNA mutation associated disease. Here, we show that by analyzing cancer WGS, both from tumors and cfDNA, we can find cancer-associated DNA mutations that lead to the generation of neomers, short sequences that are predominantly absent from genomes of healthy individuals. Further analyses of these sets of neomers show that they can be used not only to classify cancer tissue of origin, but also additional cancer features, such as MSI or POLE deficiency with high accuracy. Analysis of cfDNA WGS datasets finds that neomers could be used to tease out patients from controls in several cancers, including those with a low mutational burden. Finally, using reporter assays, we show that neomers have a functional effect on regulatory sequences.

We utilized 2,577 patients from 21 different cancerous tissues to develop a cancer tissue of origin classifier. Overall, we observed almost no overlap between neomers found in different

cancer types. This allowed us to detect cancer tissue of origin with extremely high specificity and accuracy (F1 score of 0.92 and an accuracy of 0.99) performing better than recent deep learning models[21]. In general, the classifier has better performance for cancer types with more patients and high mutation burden (**Figs. 2A-B**). Our analyses also showed that other than tissue origin, nullomers can also be used to detect other cancer features. It would be interesting to test whether nullomers and neomers could diagnose additional tumor features and also detect other cancer characteristics such as chance of recurrence, drug response, mortality and others.

As for cfDNA, despite various known challenges, including fragmentation and low levels of ctDNA, we still managed to obtain high F1 scores. This was particularly evident for lung cancer, which has a high amount of associated neomers and despite having patients at both early and late stages, we obtained an F1 score of 0.82, which compares favorably with a recent study based on metabolite profiling[73]. Moreover, our application to ovarian cancer demonstrates that the use of multiple panels of neomers and nullomers allows us to accurately detect a disease that has hitherto been very challenging to diagnose at an early stage. Since the number of neomers is proportional to the mutational burden, this suggests that our classifier will be more accurate for some types of cancers, as observed in our data for lung compared to ovarian and prostate that have a lower number of mutations/neomers. Nevertheless, since we have most likely not discovered all neomers for the tissues in our study, obtaining additional WGS datasets from tumor, matched control and cfDNA would be extremely beneficial in improving the ability of our classifier to diagnose cancer.

Our cfDNA detection approach has several advantages over current methods: 1) Detecting short DNA sequences that are enriched in cancer samples provides an easy to use diagnostic that could allow detection from low amounts of ctDNA. In addition to sequencing-based assays, alternate techniques could potentially be used, such as CRISPR-based detection tools that utilize Cas12 or Cas13[74], that can also allow the testing of thousands of sequences in parallel[75]. In addition, with neomer-based diagnostics potentially not needing large amounts of starting material, cfDNA could be collected from urine, sputum, saliva or other bodily fluids, which were shown to be a viable but reduced source of cfDNA[76,77]. 2) Our WGS approach is applicable to more sparsely sequenced samples, as our downsampling analysis indicates that already at 3x coverage we obtain similar F1 scores for both lung and ovarian cancer (data not shown). 3) As our diagnostic is based on neomer detection followed by genome alignments of nullomer encompassing reads, it is extremely effective from a computational standpoint. 4) Neomers could easily be combined with

other sequence or analyte based cancer biomarkers and risk factors to improve the diagnostic positive predictive value. For example, it was recently shown that combining a blood test that detects both protein biomarkers and DNA mutations along with positron emission tomography - computed tomography (PET-CT) could detect multiple cancers[78]. In another example, the use of cfDNA fragmentation patterns combined with CT imaging, clinical risk factor and serum levels of carcinoembryonic antigen significantly increased the ability to diagnose lung cancer[79]. Adding neomers to known cancer-associated coding mutations in the screening of cfDNA could increase sensitivity and specificity. In summary, coupling neomer-based diagnostics to existing cancer biomarkers and risk factors could improve the power to detect various cancer subtypes.

As nullomers/neomers do not exist in the human genome they could also be exceptional candidates for neoantigens, to be targeted via immunotherapy. Previous work has shown that minimal absent words, short sequences that are absent from a genome or proteome, could be used to identify phosphorylation sites of high confidence, some of which could be associated with cancer[80]. Analysis of the Immune Epitope Database of validated antigens[81] found that 13 of the recurrent coding neomers can create neoantigens with predicted strong binding levels that were subsequently validated (Table 7). From the 1,700 neoantigens with strong binding levels, only 1.72 (p-value<1e-8, hypergeometric test) is expected to correspond to a neomer, suggesting that missense mutations also resulting in neomers are 7-fold more likely to also generate strongly binding neoantigens.

Table 7. Promoter sequences cloned for luciferase assays.

***FRS3-PRICKL4*, chr6:41779847-41780883 (hg38)**		
FRS3-PRICKL4 ***NEOMER*** ***(REF)*** ***(MUT)***	CGCTCCCGGGTCCCGCTGCAGCAGCCGCCGCCCGCCCGG AGCTAAGGCCTCCCCGCCCCGCCCCGCCCCGCGGCCGCG GGGTCGGAGGTCACCGGCAGCACCAACGAGACGCTGCG GGCTGGCGGACCCGAGCGGGGGCGACCGGCACGCAGCC TAGAAGGTCCCTTTGGAGGACTTTGCAGTCGGGAGCAGT GAGCATGCGCAGACGCCAGAGGGCGCTCCTGGCCAGGG GCTGAGCAAGCCCGGAGGGGACAGGGCGGAGCCTGCGG CGGAGGCGGGGCGCGGAGGCGGGGCTGAGGACACCAAA GTTGGGGGTGGTCCGGGGAGCGGAAAGGGGACGGAGGG TTGGGCGAGAAACGAAGGGAAGGGGAGAGGGTCTGAAG GGACCTGAAAGGAGAAGGGCAAGACGATGGAACCTGGA AGAGGGTTGAGGGGAGGGGAAGGGGCCAGGGGGACCA CTGGGGACAAGGGGGAGAGAACTGAAGGGAAACACGCG GAGGAGCGGGGCTGGGAGACCACAGGGAGAGTTTGGGG	1035

	AGACCTGGGAACTGCGGAGAGACAGGGCTGCGGGCCCG GCGAGGCGTTTCCATTCAGCGAATGGTTGAAAACCCAGC GTAAACGCGGCTCCCGGCGTCTGCTGGATGCCTGCCGGG CGCAGTCCGAGCCAGAGAGCCCTTGCACCTGCATGG[C> T]AGAGGCCATGGAGCCCGCACATTTCCACCCTGAGATC CCATAAATGTGGTTTTGTTTTCCTTTTTTTTTTCCTCCCCC CAATCTGCAGCTCCAGGCCCCTCCCTGAGCCGCCCCTCT CCACCAGCCGGGGAGGGGCGGGCGGGGGCCGCTGGGGC GTGGGAGAGGAAGGGCCGCGCTATTAGGGCGGGCGGGG GCGGACGGGGGCGGGCGCTGAAGCCTTGGGGTGGGGAC CGGGTGAGGGGCCTGACGGCAGAGTCGGTCCAGAACTG CCCGGACAGCGACGCACAGCGAGGGTGAGACCCCAGCG GAGTCCTGGTCCCCTCCCCTAAAATCCCAGCAGCCAGAC CGGG	
***TMEM127-CIAO1*, chr2:96265933-96266282 (hg38)**		
TMEM127-CIAO1 ***NEOMER*** ***(REF)*** ***(MUT)***	GGGGCCGAGCGTTCCCAGGCCCCAATCCCGCAACGCCCG GACAGACCCGGGGCCGATGCACTTCCGGCTTCCTCCTCC TCTGCCTCCGCCTCCTCCGGCTCCCCGCCCCGTCCCCCGC GCGCGTCACAGCCTGGGCGC[G>T]CCGCCGGCTCTGCGC AGGGCCTCACGGGACTGGTAGTTCGAGGCGGTCCCGCGA TGCTTTCCCTGCCGAGGTGAGGGGAGCGACTACAACTCC CAGAGGGTATGGCGGCCGCAGCAGTGGGCAGTTGAGTTT CCGGGGCGACCCAGGTCCTCCGGCGGAAGCGGGAGCCT CTGTCGGCCGCGGAAGCCTGGAGTGGGCGGTACGCAGA CGC	350
***LLRC28-TTC23*, chr15:99251142-99251491 (hg38)**		
LLRC28-TTC23 ***NEOMER*** ***(REF)*** ***(MUT)***	CCGACTAAATTACCCCTTCAAGACGCTAAGCAAAAGCCA CTCCGTCGACATCAAACCCCACGCCGGAGCCGGAAGCA AGAGTAGAGCCGCGGGCGGGGCCAGCATAGAGGAAGTA GCGCTCTCGGCCAATACACAGATCGCTTGGCAAACGAGC ATTTTCCCGCGCCTATCGTGAAAGCGCGAGAACACAGGG GATTATCGCGAGGTTCCGTGCTCCTCTCCCGCTCCCTCTT CCCCAGCTCCCTCTGGCGCGGGTTCCGCTCCACCCCCGG AGCCCCGCCCCCC[G>T]ACGCCCGGCGCTCCGCCCCGCCC TCGCCTGGCTCCCTGCCCCGCCCCGCCCGCCGTCCCCCGC TT	350
***RPS2-SNHG9*, chr16:1964728-1965077 (hg38)**		

RPS2-SNHG9 ***NEOMER*** ***(REF)*** ***(MUT)***	GGGAGCCTAGACCCGACCCGATGTCCGCGGATTCCCGCC GCCCACGCAGAGGCCCGCTGCAGC[G>A]ACCAACAGGA CTCACGTGTTTTGTCGGAAAAGAAGAACGAGACCTACTG GGAAGCAGCTTTTATAGCACGCCAAGCGCCGCGAGATCT CCGCAGCCCCGCCCCAAGCGGGAGCGGGCCGAGCTCCT ATAAGACAACCTGTGATTGGCTCCGCGGTGCCCCGCCCT CACCGGGCTCTGAGTGCTCTTGCCCGTCCGGCCCCAGCC GCGGCCCGGGAATCTACGTCACCCGAAAAGCGACTATA AACGCCGGCGCCTCCGTCCCCAGCCGCGGCTCGGGAATC CACCC GGGAGCCTAGACCCGACCCGATGTCCGCGGATTCCCGCC GCCCACGCAGAGGCCCGCTGCAGC**G**ACCAACAGGACTC ACGTGTTTTGTCGGAAAAGAAGAACGAGACCTACTGGG AAGCAGCTTTTATAGCACGCCAAGCGCCGCGAGATCTCC GCAGCCCCGCCCCAAGCGGGAGCGGGCCGAGCTCCTAT AAGACAACCTGTGATTGGCTCCGCGGTGCCCCGCCCTCA CCGGGCTCTGAGTGCTCTTGCCCGTCCGGCCCCAGCCGC GGCCCGGGAATCTACGTCACCCGAAAAGCGACTATAAA CGCCGGCGCCTCCGTCCCCAGCCGCGGCTCGGGAATCCA CCC GGGAGCCTAGACCCGACCCGATGTCCGCGGATTCCCGCC GCCCACGCAGAGGCCCGCTGCAGCAACCAACAGGACTC ACGTGTTTTGTCGGAAAAGAAGAACGAGACCTACTGGG AAGCAGCTTTTATAGCACGCCAAGCGCCGCGAGATCTCC GCAGCCCCGCCCCAAGCGGGAGCGGGCCGAGCTCCTAT AAGACAACCTGTGATTGGCTCCGCGGTGCCCCGCCCTCA CCGGGCTCTGAGTGCTCTTGCCCGTCCGGCCCCAGCCGC GGCCCGGGAATCTACGTCACCCGAAAAGCGACTATAAA CGCCGGCGCCTCCGTCCCCAGCCGCGGCTCGGGAATCCA CCC	350
***GNAI2*, chr3:50237923-50238272 (hg38)**		
GNAI2 ***NEOMER*** ***(REF)*** ***(MUT)***	CATGTATGTCCATTATGGGCCATCTGTTGACTGTCCCTGC GGACCCGGCCTCCGGGGCACAGCCTTGGAGGTTTCAAAA TGGCATCCGCTTCGCTGCTGCCTCAGGGAGGAAGAGAAGG G[G>C]GCGGGAAGGGGCGGCGGCCCAGCCTCCTGGGCT GCTTTGCGTTGGGCGCTTGCCTCTTCCTGTCTCTGCTGTG AGTGCGGCAGCGGCAGTGAGTGGGTGTCGACGCGGCGG AATGCCCGTCGCTGCTGCTGCTGCCCGACGGGCCTG GGGAGGGCACTTCCGGTACCGCTGCCTCCCAGTGCCCAG CCAGGAGTCTCTAGACAGCCTCTCCCCAGGCGATTGCAG TT	350

	CATGTATGTCCATTATGGGCCATCTGTTGACTGTCCCTGC GGACCCGGCCTCCGGGGCACAGCCTTGGAGGTTTCAAAA TGGCATCCGCTTCGCTGCTGCCTCAGGGAGGAAGAGAAG GGGCGGGAAGGGGCGGCGGCCCAGCCTCCTGGGCTGCT TTGCGTTGGGCGCTTGCCTCTTCCTGTCTCTGCTGTGAGT GCGGCAGCGGCAGTGAGTGGGTGTCGACGCGGCGGAAT GCCCGTCGCTGCTGCTGCTGCTGCCCGACGGGCCTGGGG AGGGCACTTCCGGTACCGCTGCCTCCCAGTGCCCAGCCA GGAGTCTCTAGACAGCCTCTCCCCAGGCGATTGCAGTT CATGTATGTCCATTATGGGCCATCTGTTGACTGTCCCTGC GGACCCGGCCTCCGGGGCACAGCCTTGGAGGTTTCAAAA TGGCATCCGCTTCGCTGCTGCCTCAGGGAGGAAGAGAAG GCGCGGGAAGGGGCGGCGGCCCAGCCTCCTGGGCTGCT TTGCGTTGGGCGCTTGCCTCTTCCTGTCTCTGCTGTGAGT GCGGCAGCGGCAGTGAGTGGGTGTCGACGCGGCGGAAT GCCCGTCGCTGCTGCTGCTGCTGCCCGACGGGCCTGGGG AGGGCACTTCCGGTACCGCTGCCTCCCAGTGCCCAGCCA GGAGTCTCTAGACAGCCTCTCCCCAGGCGATTGCAGTT	

Neomers can be used as a novel tool to identify cancer-associated gene regulatory mutations. Amongst the 210 prostate cancer promoter neomers, we selected five promoters and found that two of them significantly affected promoter activity due to the neomer. Their difference in activity was in line with the gene's expression change in prostate cancer, with RPS2-SNHG9 having increased activity fitting with RPS2 overexpression in prostate cancer[54] and TMEM127-CIAO1 abolishing activity, in line with TMEM127 observed downregulation in cancer[55,82]. Our MPRA library of 4,609 neomer causing mutations in enhancers revealed that 2.6% can change gene expression by >1.5-fold, suggesting that a subset of these mutations could have important functional consequences in cancer. Understanding tumor onset and its transition to a metastatic cancer has been limited mostly to coding genes or their pre-determined regulatory sequences. With the vast majority of mutations falling into non-coding regions, making them much more challenging to interpret, novel approaches are required for identifying mutations that have functional consequences. Here, we show that by prioritizing mutations that result in neomers, it is possible to shortlist candidates that are likely to have a phenotypic impact. This approach can be applied to any cancer type and thus holds the potential to expand our catalog of functionally relevant cancer driver non-coding mutations.

In summary, we show that neomers can provide a powerful tool for cancer diagnosis. As they can easily be detected via sequence or CRISPR-based tools, it should be straightforward to integrate them in current routine cancer diagnostic tests and their use could increase the sensitivity and specificity of these tests. Combining neomer-based screening with clinical characteristics and additional diagnostic tools/features could increase the positive predictive value. In addition, as cfDNA could also be isolated from urine and saliva, and detection of these sequences only requires a relatively small amount of DNA, neomer-based diagnosis could be carried out in a non-invasive manner. Our work also suggests that neomers could be used to highlight cancer-associated gene regulatory mutations which have been difficult to identify. Further high-throughput characterization of these mutations could allow the detection of bona fide cancer-associated functional regulatory mutations that could be used for diagnosis and treatment.

REFERENCES

1. Cancer. https://www.who.int/news-room/fact-sheets/detail/cancer.

2. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2020. CA Cancer J. Clin. 70, 7–30 (2020).

3. Hawkes, N. Cancer survival data emphasise importance of early diagnosis. BMJ 364, (2019).

4. Etzioni, R. et al. The case for early detection. Nat. Rev. Cancer 3, 243–252 (2003).

5. Cancer. https://www.who.int/cancer/detection/en/.

6. Bronkhorst, A. J., Ungerer, V. & Holdenrieder, S. The emerging role of cell-free DNA as a molecular marker for cancer management. Biomol Detect Quantif 17, 100087 (2019).

7. Heitzer, E., Auinger, L. & Speicher, M. R. Cell-Free DNA and Apoptosis: How Dead Cells Inform About the Living. Trends Mol. Med. 26, 519–528 (2020).

8. Brill MD FACP, J. Screening for Cancer: The Economic, Medical, and Psychosocial Issues. AJMC https://cdn.sanity.io/files/0vv8moc6/ajmc/6f98af549a672a707e7d20d211706afaf0380b7a.pdf/AJMC_ACE0199_EarlyCancerDetection__WEB.pdf (2020).

9. Zill, O. A., Banks, K. C., Fairclough, S. R. & Mortimer, S. A. The landscape of actionable genomic alterations in cell-free circulating tumor DNA from 21,807 advanced cancer patients. Clin. Cancer Res. (2018).

10. Saghafinia, S., Mina, M., Riggi, N., Hanahan, D. & Ciriello, G. Pan-Cancer Landscape of Aberrant DNA Methylation across Human Tumors. Cell Rep. 25, 1066–1080.e8 (2018).

11. Sadeh, R. et al. ChIP-seq of plasma cell-free nucleosomes identifies gene expression programs of the cells of origin. Nat. Biotechnol. (2021) doi:10.1038/s41587-020-00775-6.

12. Barbany, G. et al. Cell-free tumour DNA testing for early detection of cancer--a potential future tool. J. Intern. Med. 286, 118–136 (2019).

13. Razavi, P. et al. High-intensity sequencing reveals the sources of plasma circulating cell-free DNA variants. Nat. Med. 25, 1928–1937 (2019).

14. Ji, L. et al. Methylated DNA is over-represented in whole-genome bisulfite sequencing data. Front. Genet. 5, 341 (2014).

15. Worm Ørntoft, M.-B. Review of Blood-Based Colorectal Cancer Screening: How Far Are Circulating Cell-Free DNA Methylation Markers From Clinical Implementation? Clin. Colorectal Cancer 17, e415–e433 (2018).

16. Warton, K. & Samimi, G. Methylation of cell-free circulating DNA in the diagnosis of cancer. Front Mol Biosci 2, 13 (2015).

17. Hampikian, G. & Andersen, T. ABSENT SEQUENCES: NULLOMERS AND PRIMES. in Biocomputing 2007 355–366 (WORLD SCIENTIFIC, 2006).

18. Vergni, D. & Santoni, D. Nullomers and High Order Nullomers in Genomic Sequences. PLoS One 11, e0164540 (2016).

19. Georgakopoulos-Soares, I., Yizhar-Barnea, O., Mouratidis, I., Hemberg, M. & Ahituv, N. Absent from DNA and protein: genomic characterization of nullomers and nullpeptides across functional categories and evolution. Genome Biol. 22, 245 (2021).

20. The Cancer Genome Atlas Program. https://www.cancer.gov/tcga (2018).

21. Jiao, W. et al. A deep learning system accurately classifies primary and metastatic cancers using passenger mutation patterns. Nat. Commun. 11, 728 (2020).

22. ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. Nature 578, 82–93 (2020).

23. Prior, I. A., Lewis, P. D. & Mattos, C. A comprehensive survey of Ras mutations in cancer. Cancer Res. 72, 2457–2467 (2012).

24. Muñoz-Maldonado, C., Zimmer, Y. & Medová, M. A Comparative Analysis of Individual RAS Mutations in Cancer Biology. Front. Oncol. 9, 1088 (2019).

25. Vinagre, J. et al. Frequency of TERT promoter mutations in human cancers. Nat. Commun. 4, 2185 (2013).

26. Heidenreich, B., Rachakonda, P. S., Hemminki, K. & Kumar, R. TERT promoter mutations in cancer development. Curr. Opin. Genet. Dev. 24, 30–37 (2014).

27. Song, J. H. et al. Small-Molecule-Targeting Hairpin Loop of hTERT Promoter G-Quadruplex Induces Cancer Cell Death. Cell Chem Biol 26, 1110–1121.e4 (2019).

28. Bell, R. J. A. et al. Cancer. The transcription factor GABP selectively binds and activates the mutant TERT promoter in cancer. Science 348, 1036–1039 (2015).

29. Powter, B. et al. Human TERT promoter mutations as a prognostic biomarker in glioma. J. Cancer Res. Clin. Oncol. 147, 1007–1017 (2021).

30. Romero, J. M. et al. A Four-Chemokine Signature Is Associated with a T-cell-Inflamed Phenotype in Primary and Metastatic Pancreatic Cancer. Clin. Cancer Res. 26, 1997–2010 (2020).

31. Antal, C. E. et al. Cancer-Associated Protein Kinase C Mutations Reveal Kinase's Role as Tumor Suppressor. Cell vol. 160 489–502 Preprint at https://doi.org/10.1016/j.cell.2015.01.001 (2015).

32. Vincent, A., Omura, N., Hong, S. M., Jaffe, A. & Eshleman, J. Genome-wide analysis of promoter methylation associated with gene expression profile in pancreatic adenocarcinoma. Clin. Cancer Res. (2011).

33. Xu, E. et al. Genome-wide methylation analysis shows similar patterns in Barrett's esophagus and esophageal adenocarcinoma. Carcinogenesis 34, 2750–2756 (2013).

34. Battaglin, F., Naseem, M., Lenz, H.-J. & Salem, M. E. Microsatellite instability in colorectal cancer: overview of its clinical significance and novel perspectives. Clin. Adv. Hematol. Oncol. 16, 735–745 (2018).

35. Strand, M., Earley, M. C., Crouse, G. F. & Petes, T. D. Mutations in the MSH3 gene preferentially lead to deletions within tracts of simple repetitive DNA in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. U. S. A. 92, 10418–10421 (1995).

36. Nik-Zainal, S. et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature 534, 47–54 (2016).

37. Garmezy, B. et al. Correlation of pathogenic POLE mutations with clinical benefit to immune checkpoint inhibitor therapy. J. Clin. Orthod. 38, 3008–3008 (2020).

38. Wang, F. et al. Evaluation of POLE and POLD1 Mutations as Biomarkers for

Immunotherapy Outcomes Across Multiple Cancer Types. JAMA Oncol 5, 1504–1506 (2019).

39. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. Nature 500, 415–421 (2013).

40. Thai, A. A., Solomon, B. J., Sequist, L. V., Gainor, J. F. & Heist, R. S. Lung cancer. Lancet 398, 535–554 (2021).

41. Blandin Knight, S. et al. Progress and prospects of early detection in lung cancer. Open Biol. 7, (2017).

42. de Groot, P. M., Wu, C. C., Carter, B. W. & Munden, R. F. The epidemiology of lung cancer. Transl Lung Cancer Res 7, 220–233 (2018).

43. Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 581, 434–443 (2020).

44. Henderson, J. T., Webber, E. M. & Sawaya, G. F. Screening for Ovarian Cancer: An Updated Evidence Review for the U.S. Preventive Services Task Force. (Agency for Healthcare Research and Quality (US), 2018).

45. Heintz, A. P. M. et al. Carcinoma of the ovary. International Journal of Gynecology & Obstetrics 95, S161–S192 (2006).

46. Rawla, P. Epidemiology of Prostate Cancer. World J. Oncol. 10, 63–89 (2019).

47. Barry, M. J. Prostate-Specific–Antigen Testing for Early Diagnosis of Prostate Cancer. New England Journal of Medicine vol. 344 1373–1377 Preprint at https://doi.org/10.1056/nejm200105033441806 (2001).

48. Hennigan, S. T. et al. Low Abundance of Circulating Tumor DNA in Localized Prostate Cancer. JCO Precis Oncol 3, (2019).

49. Bjerre, M. T. et al. Epigenetic Analysis of Circulating Tumor DNA in Localized and Metastatic Prostate Cancer: Evaluation of Clinical Biomarker Potential. Cells 9, (2020).

50. Maia, M. C., Salgia, M. & Pal, S. K. Harnessing cell-free DNA: plasma circulating tumour DNA for liquid biopsy in genitourinary cancers. Nat. Rev. Urol. 17, 271–291 (2020).

51. Poulos, R. C., Sloane, M. A., Hesson, L. B. & Wong, J. W. H. The search for cis-regulatory driver mutations in cancer genomes. Oncotarget 6, 32509–32525 (2015).

52. Elliott, K. & Larsson, E. Non-coding driver mutations in human cancer. Nat. Rev. Cancer (2021) doi:10.1038/s41568-021-00371-z.

53. Consortium, Encode Project et al. An integrated encyclopedia of DNA elements in the

human genome. Nature 489, 57–74 (2012).

54. Ohkia, A., Hu, Y., Wang, M., Garcia, F. U. & Stearns, M. E. Evidence for prostate cancer-associated diagnostic marker-1: immunohistochemistry and in situ hybridization studies. Clin. Cancer Res. 10, 2452–2458 (2004).

55. Qin, Y. et al. The tumor susceptibility gene TMEM127 is mutated in renal cell carcinomas and modulates endolysosomal function. Hum. Mol. Genet. 23, 2428–2439 (2014).

56. Bowler, E. et al. Hypoxia leads to significant changes in alternative splicing and elevated expression of CLK splice factor kinases in PC3 prostate cancer cells. BMC Cancer 18, 355 (2018).

57. El-Haibi, C. P. et al. Differential G protein subunit expression by prostate cancer cells and their interaction with CXCR5. Mol. Cancer 12, 64 (2013).

58. Valencia, T. et al. Role and expression of FRS2 and FRS3 in prostate cancer. BMC Cancer 11, 484 (2011).

59. ENCODE Project Consortium et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature 583, 699–710 (2020).

60. Gordon, M. G. et al. lentiMPRA and MPRAflow for high-throughput functional characterization of gene regulatory elements. Nat. Protoc. 15, 2387–2412 (2020).

61. Raudvere, U. et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res. 47, W191–W198 (2019).

62. Shiina, H. et al. Functional Loss of the gamma-catenin gene through epigenetic and genetic pathways in human prostate cancer. Cancer Res. 65, (2005).

63. Yu, F. et al. Wnt/β-catenin signaling in cancers and targeted therapies. Signal Transduct Target Ther 6, 307 (2021).

64. Kypta, R. M. & Waxman, J. Wnt/β-catenin signalling in prostate cancer. Nat. Rev. Urol. 9, 418–428 (2012).

65. Schneider, J. A. & Logan, S. K. Revisiting the role of Wnt/β-catenin signaling in prostate cancer. Mol. Cell. Endocrinol. 462, 3–8 (2018).

66. Wang, C., Chen, Q. & Xu, H. Wnt/β-catenin signal transduction pathway in prostate cancer and associated drug resistance. Discov Oncol 12, 40 (2021).

67. Wen, Y.-C. et al. TCF7L1 regulates cytokine response and neuroendocrine differentiation of prostate cancer. Oncogenesis 10, 81 (2021).

68. Shy, B. R. et al. Regulation of Tcf7l1 DNA Binding and Protein Stability as Principal

Mechanisms of Wnt/β-Catenin Signaling. Cell Rep. 4, 1 (2013).

69. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. Bioinformatics vol. 27 1017–1018 Preprint at https://doi.org/10.1093/bioinformatics/btr064 (2011).

70. Li, D. et al. FOXD3 is a novel tumor suppressor that affects growth, invasion, metastasis and angiogenesis of neuroblastoma. Oncotarget 4, 2021–2044 (2013).

71. Shen, L. et al. Role of PRDM1 in Tumor Immunity and Drug Response: A Pan-Cancer Analysis. Front. Pharmacol. 0, (2020).

72. Shiota, M., Fujimoto, N., Kashiwagi, E. & Eto, M. The Role of Nuclear Receptors in Prostate Cancer. Cells 8, (2019).

73. Schult, T. A. et al. Screening human lung cancer with predictive models of serum magnetic resonance spectroscopy metabolomics. Proc. Natl. Acad. Sci. U. S. A. 118, (2021).

74. Kellner, M. J., Koob, J. G., Gootenberg, J. S., Abudayyeh, O. O. & Zhang, F. SHERLOCK: nucleic acid detection with CRISPR nucleases. Nat. Protoc. 14, 2986–3012 (2019).

75. Ackerman, C. M. et al. Massively multiplexed nucleic acid detection with Cas13. Nature 582, 277–282 (2020).

76. Augustus, E. et al. The art of obtaining a high yield of cell-free DNA from urine. PLoS One 15, e0231058 (2020).

77. Ding, S. et al. Saliva-derived cfDNA is applicable for EGFR mutation detection but not for quantitation analysis in non-small cell lung cancer. Thorac Cancer 10, 1973–1983 (2019).

78. Lennon, A. M. et al. Feasibility of blood testing combined with PET-CT to screen for cancer and guide intervention. Science 369, (2020).

79. Mathios, D. et al. Detection and characterization of lung cancer using cell-free DNA fragmentomes. Nat. Commun. 12, 5060 (2021).

80. Koulouras, G. & Frith, M. C. Significant non-existence of sequences in genomes and proteomes. Nucleic Acids Res. (2021) doi:10.1093/nar/gkab139.

81. Vita, R. et al. The Immune Epitope Database (IEDB): 2018 update. Nucleic Acids Res. 47, D339–D343 (2019).

82. Edlind, M. P. & Hsieh, A. C. PI3K-AKT-mTOR signaling in prostate cancer progression and androgen deprivation therapy resistance. Asian J. Androl. 16, (2014).

83. Inoue, F. & Ahituv, N. Decoding enhancers using massively parallel reporter assays.

Genomics 106, 159–164 (2015).

84. Chang, C.-C. & Lin, C.-J. LIBSVM: A library for support vector machines. ACM Trans. Intell. Syst. Technol. 2, 1–27 (2011).

85. Zou, X. et al. A systematic CRISPR screen defines mutational mechanisms underpinning signatures caused by replication errors and endogenous DNA damage. Nat Cancer 2, 643–657 (2021).

86. Chen, E. et al. Cell-free DNA concentration and fragment size as a biomarker for prostate cancer. Sci. Rep. 11, 5040 (2021).

87. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359 (2012).

Ackerman et al. 2020. "Massively Multiplexed Nucleic Acid Detection with Cas13." Nature 582 (7811): 277–82.

Alileche et al. 2012. "Nullomer Derived Anticancer Peptides (NulloPs): Differential Lethal Effects on Normal and Cancer Cells in Vitro." Peptides 38 (2): 302–11.

Alileche et al. 2017. "The Effect of Nullomer-Derived Peptides 9R, 9S1R and 124R on the NCI-60 Panel and Normal Cell Lines." BMC Cancer 17 (1): 533.

Augustus et al. 2020. "The Art of Obtaining a High Yield of Cell-Free DNA from Urine." PloS One 15 (4): e0231058.

Barbany et al. 2019. "Cell-Free Tumour DNA Testing for Early Detection of Cancer--a Potential Future Tool." Journal of Internal Medicine 286 (2): 118–36.

Barry, Michael J. 2001. "Prostate-Specific–Antigen Testing for Early Diagnosis of Prostate Cancer." New England Journal of Medicine. https://doi.org/10.1056/nejm200105033441806.

Battaglin et al. 2018. "Microsatellite Instability in Colorectal Cancer: Overview of Its Clinical Significance and Novel Perspectives." Clinical Advances in Hematology & Oncology: H&O 16 (11): 735–45.

Bell et al. 2015. "Cancer. The Transcription Factor GABP Selectively Binds and Activates the Mutant TERT Promoter in Cancer." Science 348 (6238): 1036–39.

Bowler et al. 2018. "Hypoxia Leads to Significant Changes in Alternative Splicing and Elevated Expression of CLK Splice Factor Kinases in PC3 Prostate Cancer Cells." BMC Cancer 18 (1): 355.

Bronkhorst et al. 2019. "The Emerging Role of Cell-Free DNA as a Molecular Marker for Cancer Management." Biomolecular Detection and Quantification 17 (March): 100087.

Cackowski et al. 2018. "Minimal Residual Disease in Prostate Cancer." Advances in Experimental Medicine and Biology. https://doi.org/10.1007/978-3-319-97746-1_3.

"Cancer." n.d. Accessed February 6, 2021a. https://www.who.int/news-room/fact-sheets/detail/cancer. n.d. Accessed December 2, 2020b. https://www.who.int/cancer/detection/en/.

Chen et al. 2021. "Cell-Free DNA Concentration and Fragment Size as a Biomarker for Prostate Cancer." Scientific Reports 11 (1): 5040.

Consortium et al. 2012. "An Integrated Encyclopedia of DNA Elements in the Human Genome." Nature 489 (7414): 57–74.

Ding et al. 2019. "Saliva-Derived cfDNA Is Applicable for EGFR Mutation Detection but Not for Quantitation Analysis in Non-Small Cell Lung Cancer." Thoracic Cancer 10 (10): 1973– 83.

El-Haibi et al. 2013. "Differential G Protein Subunit Expression by Prostate Cancer Cells and Their Interaction with CXCR5." Molecular Cancer 12 (June): 64.

Etzioni et al. 2003. "The Case for Early Detection." Nature Reviews. Cancer 3 (4): 243–52.

Georgakopoulos-Soares et al. 2020. "Absent from DNA and Protein: Genomic Characterization of Nullomers and Nullpeptides across Functional Categories and Evolution." Cold Spring Harbor Laboratory. https://doi.org/10.1101/2020.03.02.972422.

Hampikian et al. 2006. "ABSENT SEQUENCES: NULLOMERS AND PRIMES." In Biocomputing 2007, 355–66. WORLD SCIENTIFIC.

Hawkes, Nigel. 2019. "Cancer Survival Data Emphasise Importance of Early Diagnosis." BMJ 364 (January). https://doi.org/10.1136/bmj.l408.

Heidenreich et al. 2014. "TERT Promoter Mutations in Cancer Development." Current Opinion in Genetics & Development 24 (February): 30–37.

Heitzer et al. 2020. "Cell-Free DNA and Apoptosis: How Dead Cells Inform About the Living." Trends in Molecular Medicine 26 (5): 519–28.

ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. 2020. "Pan-Cancer Analysis of Whole Genomes." Nature 578 (7793): 82–93.

Inoue et al. 2015. "Decoding Enhancers Using Massively Parallel Reporter Assays." Genomics 106 (3): 159–64.

Jiao et al. PCAWG Tumor Subtypes and Clinical Translation Working Group, Alexandra Danyi,

et al. 2020. "A Deep Learning System Accurately Classifies Primary and Metastatic Cancers Using Passenger Mutation Patterns." Nature Communications 11 (1): 728.

Ji et al. 2014. "Methylated DNA Is over-Represented in Whole-Genome Bisulfite Sequencing Data." Frontiers in Genetics 5 (October): 341.

Karczewski et al. 2020. "The Mutational Constraint Spectrum Quantified from Variation in 141,456 Humans." Nature 581 (7809): 434–43.

Kellner et al. 2019. "SHERLOCK: Nucleic Acid Detection with CRISPR Nucleases." Nature Protocols 14 (10): 2986–3012.

Koulouras et al. 2021. "Significant Non-Existence of Sequences in Genomes and Proteomes." Nucleic Acids Research, March. https://doi.org/10.1093/nar/gkab139.

Lee et al. 2020. "BRCA1/BRCA2 Pathogenic Variant Breast Cancer: Treatment and Prevention Strategies." Annals of Laboratory Medicine 40 (2): 114–21.

Lennon et al. 2020. "Feasibility of Blood Testing Combined with PET-CT to Screen for Cancer and Guide Intervention." Science 369 (6499). https://doi.org/10.1126/science.abb9601.

Muñoz-Maldonado et al. 2019. "A Comparative Analysis of Individual RAS Mutations in Cancer Biology." Frontiers in Oncology 9 (October): 1088.

Murray, Nigel P. 2018. "Minimal Residual Disease in Prostate Cancer Patients after Primary Treatment: Theoretical Considerations, Evidence and Possible Use in Clinical Management." Biological Research 51 (1): 32.

Nik-Zainal et al. 2016. "Landscape of Somatic Mutations in 560 Breast Cancer Whole-Genome Sequences." Nature 534 (7605): 47–54.

Ohkia et al. 2004. "Evidence for Prostate Cancer-Associated Diagnostic Marker-1: Immunohistochemistry and in Situ Hybridization Studies." Clinical Cancer Research: An Official Journal of the American Association for Cancer Research 10 (7): 2452–58.

Poulos et al. 2015. "The Search for Cis-Regulatory Driver Mutations in Cancer Genomes." Oncotarget 6 (32): 32509–25.

Powter et al. 2021. "Human TERT Promoter Mutations as a Prognostic Biomarker in Glioma." Journal of Cancer Research and Clinical Oncology 147 (4): 1007–17.

Prior et al. 2012. "A Comprehensive Survey of Ras Mutations in Cancer." Cancer Research 72 (10): 2457–67.

Qin et al. 2014. "The Tumor Susceptibility Gene TMEM127 Is Mutated in Renal Cell Carcinomas

and Modulates Endolysosomal Function." Human Molecular Genetics 23 (9): 2428–39.

Razavi et al. 2019. "High-Intensity Sequencing Reveals the Sources of Plasma Circulating Cell-Free DNA Variants." Nature Medicine 25 (12): 1928–37.

Sadeh et al. 2021. "ChIP-Seq of Plasma Cell-Free Nucleosomes Identifies Gene Expression Programs of the Cells of Origin." Nature Biotechnology, January. https://doi.org/10.1038/s41587-020-00775-6.

Saghafinia et al. 2018. "Pan-Cancer Landscape of Aberrant DNA Methylation across Human Tumors." Cell Reports 25 (4): 1066–80.e8.

Santoni et al. 2020. "In the Search of Potential Epitopes for Wuhan Seafood Market Pneumonia Virus Using High Order Nullomers." Journal of Immunological Methods 481¬482 (June): 112787.

Song et al. 2019. "Small-Molecule-Targeting Hairpin Loop of hTERT Promoter G-Quadruplex Induces Cancer Cell Death." Cell Chemical Biology 26 (8): 1110–21.e4.

"The Cancer Genome Atlas Program." 2018. 2018. https://www.cancer.gov/tcga.

Tung et al. 2018. "BRCA1/2 Testing: Therapeutic Implications for Breast Cancer Management." British Journal of Cancer 119 (2): 141–52.

Ulz et al. 2019. "Inference of Transcription Factor Binding from Cell-Free DNA Enables Tumor Subtype Prediction and Early Detection." Nature Communications 10 (1): 4666.

Valencia et al. 2011. "Role and Expression of FRS2 and FRS3 in Prostate Cancer." BMC Cancer 11 (November): 484.

Vergni et al. 2020. "The Farther the Better: Investigating How Distance from Human Self Affects the Propensity of a Peptide to Be Presented on Cell Surface by MHC Class I Molecules, the Case of Trypanosoma Cruzi." PloS One 15 (12): e0243285.

Vergni et al. 2016. "Nullomers and High Order Nullomers in Genomic Sequences." PloS One 11 (12): e0164540.

Vinagre et al. 2013. "Frequency of TERT Promoter Mutations in Human Cancers." Nature Communications 4: 2185.

Vita et al. 2019. "The Immune Epitope Database (IEDB): 2018 Update." Nucleic Acids Research 47 (D1): D339–43.

Warton et al. 2015. "Methylation of Cell-Free Circulating DNA in the Diagnosis of Cancer." Frontiers in Molecular Biosciences 2 (April): 13.

Worm et al. 2018. "Review of Blood-Based Colorectal Cancer Screening: How Far Are

Circulating Cell-Free DNA Methylation Markers From Clinical Implementation?" Clinical Colorectal Cancer 17 (2): e415–33.

Zill et al. 2018. "The Landscape of Actionable Genomic Alterations in Cell-Free Circulating Tumor DNA from 21,807 Advanced Cancer Patients." Clinical Cancer Research: An Official Journal of the American Association for Cancer Research.

CLAIMS:

1. A method of identifying the presence of a neomer in a sample from a subject comprising:
(a) identifying the presence of one or a plurality of nullomers in a dataset;
(b) identifying nullomers from a sample of a subject;
(c) annotating the nullomers from the subject as being specific to the subject by comparing the nullomers from the sample with nucleic acid sequences from a control subject;
(d) assigning a cancer type to a nullomer with a corresponding cancer type of the subject;
(e) creating a library of neomers that correspond to a cancer type by repeating steps (c) and (d).

2. The method of claim 1, wherein the step of identifying the presence of one or a plurality of nullomers in a dataset comprises:
(i) contacting isolated nucleic acids from a sample to one or a plurality of probes specific for one or a plurality of nullomers;
(ii) detecting the presence of the probes associated with the one or plurality of nullomers; and
(iii) correlating the presence or quantity of probes with the likelihood of the presence or quantity of nullomers in the sample.

3. The method of claim 2, further comprising the step of (iv) isolating a plurality of nucleic acids from the sample prior to performance of the step of contacting.

4. The method of any of claims 1 through 3, wherein the sample comprises cell free nucleic acid sequences.

5. The method of claim 3 or 4, wherein the step of isolating a plurality of nucleic acids from the sample comprises separating cell free nucleic acid sequence from protein in a sample.

6. The method of any of claims 3 through 5, wherein the step of isolating a plurality of nucleic acids from the sample comprises extracting nullomers from a sample comprising one or a combination of: siRNA, rRNA, circulating tumor DNA (ctDNA), cell-free mitochondrial DNA (ccf mtDNA), and cell-free fetal DNA (cffDNA).

7. A method of creating a library of neomers that correspond to a cancer type from a sample of a subject comprising:
(a) identifying the presence of one or a plurality of nullomers in a dataset;
(b) identifying nullomers from a sample of a subject;
(c) annotating the nullomers from the subject as being specific to the subject by comparing the nullomers from the sample with nucleic acid sequences from a control subject;
(d) assigning a cancer type to a nullomer and designating the nullomer a neomer with a corresponding cancer type of the patient if the frequency of the nullomer in the sample corresponds to the presence of a tumor from one or a plurality of subject from the dataset;
(e) creating a library of neomers that correspond to a cancer type by repeating steps (c) and (d).

8. The method of claim 7, wherein the step of identifying the presence of one or a plurality of nullomers in a dataset comprises:
(i) contacting isolated nucleic acids from a sample to one or a plurality of probes specific for one or a plurality of nullomers;
(ii) detecting the presence of the probes associated with the one or plurality of nullomers; and
(iii) correlating the presence or quantity of probes with the likelihood of the presence or quantity of nullomers in the sample.

9. The method of claim 8, further comprising the step of (iv) isolating a plurality of nucleic acids from the sample prior to performance of the step of contacting.

10. The method of any of claims 7 through 9, wherein the sample comprises cell free nucleic acid sequences.

11. The method of claim 9 or 10, wherein the step of isolating a plurality of nucleic acids from the sample comprises separating cell free nucleic acid sequence from protein in a sample.

12. The method of any of claims 9 through 11, wherein the step of isolating a plurality of nucleic acids from the sample comprises extracting nullomers from a sample comprising one or a

combination of: siRNA, rRNA, circulating tumor DNA (ctDNA), cell-free mitochondrial DNA (ccf mtDNA), and cell-free fetal DNA (cffDNA).

13. The method of any of claims 7 through 12, wherein the sample is blood, saliva, spit or plasma.

14. A method of identifying an effect of a presence of a neomer on a genetic element of genomic DNA of a cell comprising:

(a) cloning a neomer sequence into a plasmid comprising a regulatory element operably linked to an expressible nucleic acid encoding a reporter at or proximate to the regulatory element;

(b) transforming the plasmid into a cell; and

(c) monitoring the expression of the reporter relative to a plasmid that is free of the neomer sequence; and

(d) characterizing the effect of the neomer on the regulatory element as positive if expression of the reporter gene is normal relative to expression of the reporter gene in a cell comprising the plasmid free of the neomer; or characterizing the effect of the neomer on the regulatory element as negative if expression of the reporter gene is dysfunctional as compared to expression of the reporter in a plasmid that is free of the neomer sequence.

15. The method of claim 14, wherein the regulatory element is a mammalian promoter or enhancer element.

16. The method of claim 15, wherein the regulatory element is a human promoter or enhancer.

17. The method of any of claims 14 through 16, wherein expression of the reporter is performed in a cancer cell line.

18. The method of any of claims 14 through 17, wherein the reporter is luciferase or a functional fragment thereof.

19. The method of any of claims 14 through 18, wherein the effect of the neomer is characterized as negative if the promoter sequence is disrupted by the presence of the neomer resulting in a loss of function and/or decreased or abrogated expression of the reporter.

20. A method of identifying an altered genetic element of genomic DNA of a hyperproliferative cell within a subject:

(a) detecting the presence of a neomer in a sample;

(b) pairing the presence of a neomer to an altered activity of a genetic element in the cancer cell within the subject.

21. The method of claim 20 further comprising the step of (c) correlating the presence of the neomer to the position of a mutation on the genomic DNA of the hyperproliferative cell in the subject.

22. The method of claim 20, wherein the hyperproliferative cell is a cell from a solid tumor.

23. The method of any of claims 20 through 22, wherein the sample is bodily fluid comprising cell free DNA.

24. The method of any of claims 20 through 23, wherein the mutation is the presence of an indel within the genomic DNA of a hyperproliferative cell within a regulatory element.

25. The method of any claims 20 through 24 further comprising a step of (d) determining a therapy for the subject based upon the location of the mutation within the genomic DNA.

26. The method of any of claims 20 through 25, wherein the hyperproliferative cell is from the lung, ovary or breast.

27. A method of diagnosing a subject with a hyperproliferative disease or type of cancer comprising:

(a) isolating a plurality of nucleic acids from the sample;

(b) contacting the nucleic acids to one or a plurality of probes specific for one or a plurality of neomers;
(c) detecting the presence of the probes associated with the one or plurality of neomers; and
(d) correlating the presence or quantity of probes with the likelihood of the presence or quantity of neomer in the sample.

28. The method of claim 27 further comprising diagnosing the subject as having cancer based upon the presence or quantity of neomers in the sample.

29. The method of any of claims 26 and 27, wherein the cancer is a tumor exhibiting a mutation in the EGFR1, BRCA2 or TP53 protein or nucleic acid encoding EGFR1, BRCA2 or TP53.

30. The method of any of claims 26 and 27, wherein the cancer is derived from the breast, lung, or ovary of the subject and wherein the subject is diagnosed as having breast cancer, lung cancer or ovarian cancer.

31. The method of any of claims 27 through 30, wherein the sample comprises cell-free DNA.

32. The method of any of claims 27 through 31 further comprising isolating a sample from the subject.

33. A method of preparing cell free DNA from a sample comprising:
(a) isolating cell free DNA from a sample;
(b) extracting one or a plurality of nullomers from the cell free DNA by removing one or a combination of: siRNA, rRNA, circulating tumor DNA (ctDNA), cell-free mitochondrial DNA (ccf mtDNA), and cell-free fetal DNA (cffDNA) from the sample;
(c) analyzing the one or plurality of nullomers by comparing the sequence of the one or plurality of nullomers to a known library of nullomer sequences.

34. The method of claim 33 wherein the step of analyzing the one or plurality of nullomer sequences comprising quantifying and/or identifying the presence of the one or plurality of nullomers in the sequence by contacting one or a plurality of probes specific to one or plurality of nullomers; and characterizing the phenotype or position of a mutation within a cancer cell based upon the presence or quantity of the nullomer.

35. A method of treating cancer in a subject comprising:

(a) contacting a sample from the subject with one or a plurality of probes specific for one or a plurality of neomer sequences;

(b) detecting the presence, absence or quantity of neomer sequence in the sample by measuring the presence of the one or plurality of probes;

(c) characterizing a phenotype of the cancer or the presence of a mutation within the DNA of a hyperproliferative cell based upon the presence, absence or quantity of the neomer;

(d) treating the subject with a therapeutic agent based upon the phenotype of the cancer or the presence of a mutation within the DNA of the hyperproliferative cell in the subject.

36. The method of claim 35, wherein the hyperproliferative cell comprises a mutation in the gene encoding EGFR, EGFR1, TP53, or BRCA2.

37. The method of claim 35, wherein the therapeutic agent comprises any one or combination of therapeutic agents of Table 3.

38. The method of any of claims 35 through 37, wherein the sample comprises cell-free DNA.

39. The method of any of claims 35 through 38, wherein the sample is plasma, blood, or saliva.

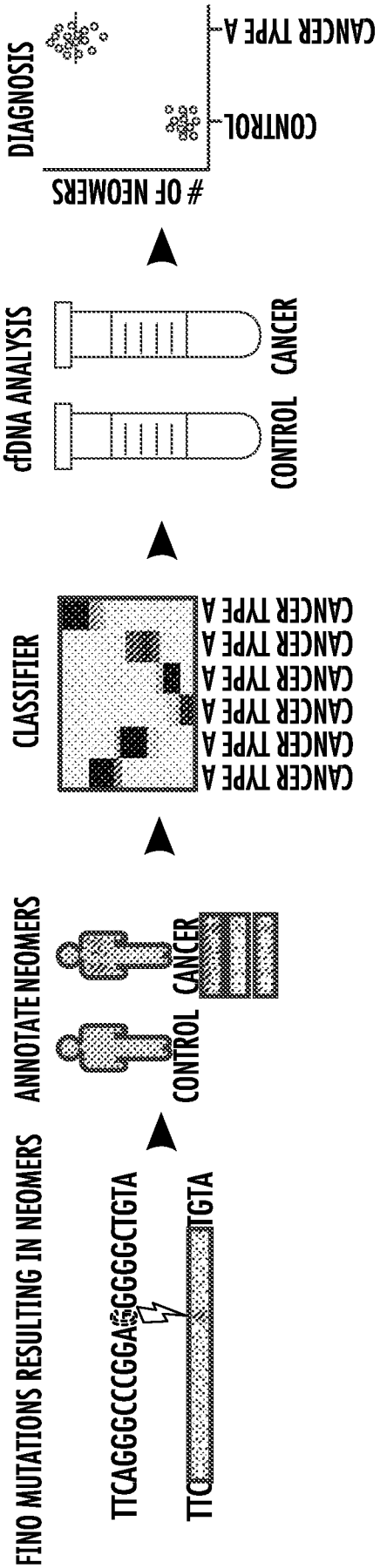

FIND MUTATIONS RESULTING IN NEOMERS
TTCAGGGCCCGGA GGGGCTGTA
TTC TGTA
ANNOTATE NEOMERS
CONTROL
CANCER
CLASSIFIER
CANCER TYPE A
CANCER TYPE A
CANCER TYPE A
CANCER TYPE A
CANCER TYPE A
CANCER TYPE A
cfDNA ANALYSIS
CONTROL
CANCER
DIAGNOSIS
OF NEOMERS
CONTROL
CANCER TYPE A

FIG. 1A

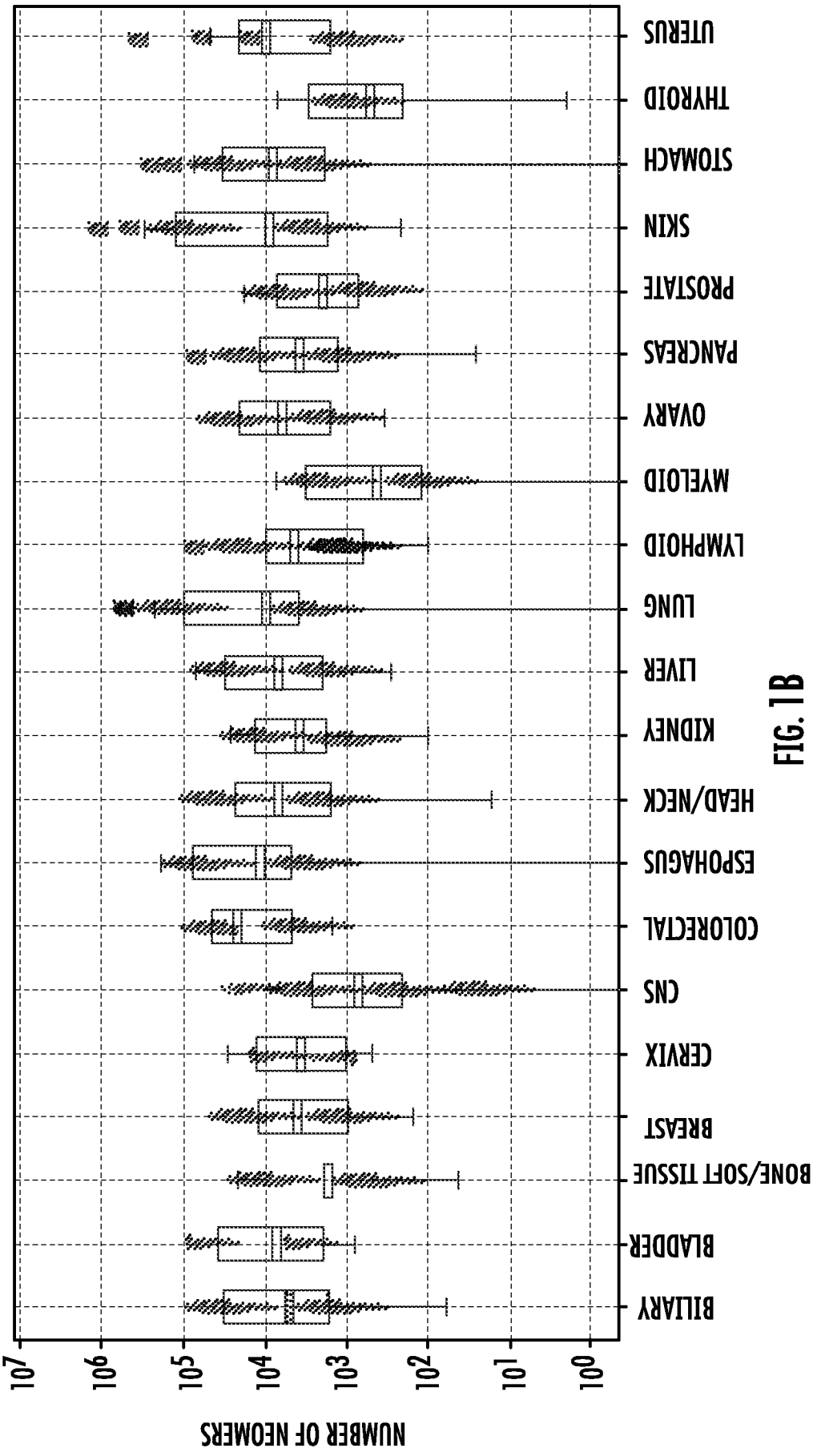

UTERUS
THYROID
STOMACH
SKIN
PROSTATE
PANCREAS
OVARY
MYELOID
LYMPHOID
LUNG
LIVER
KIDNEY
HEAD/NECK
ESPOHAGUS
COLORECTAL
CNS
CERVIX
BREAST
BONE/SOFT TISSUE
BLADDER
BILIARY
10^7
10^6
10^5
10^4
10^3
10^2
10^1
10^0
NUMBER OF NEOMERS

FIG. 1B

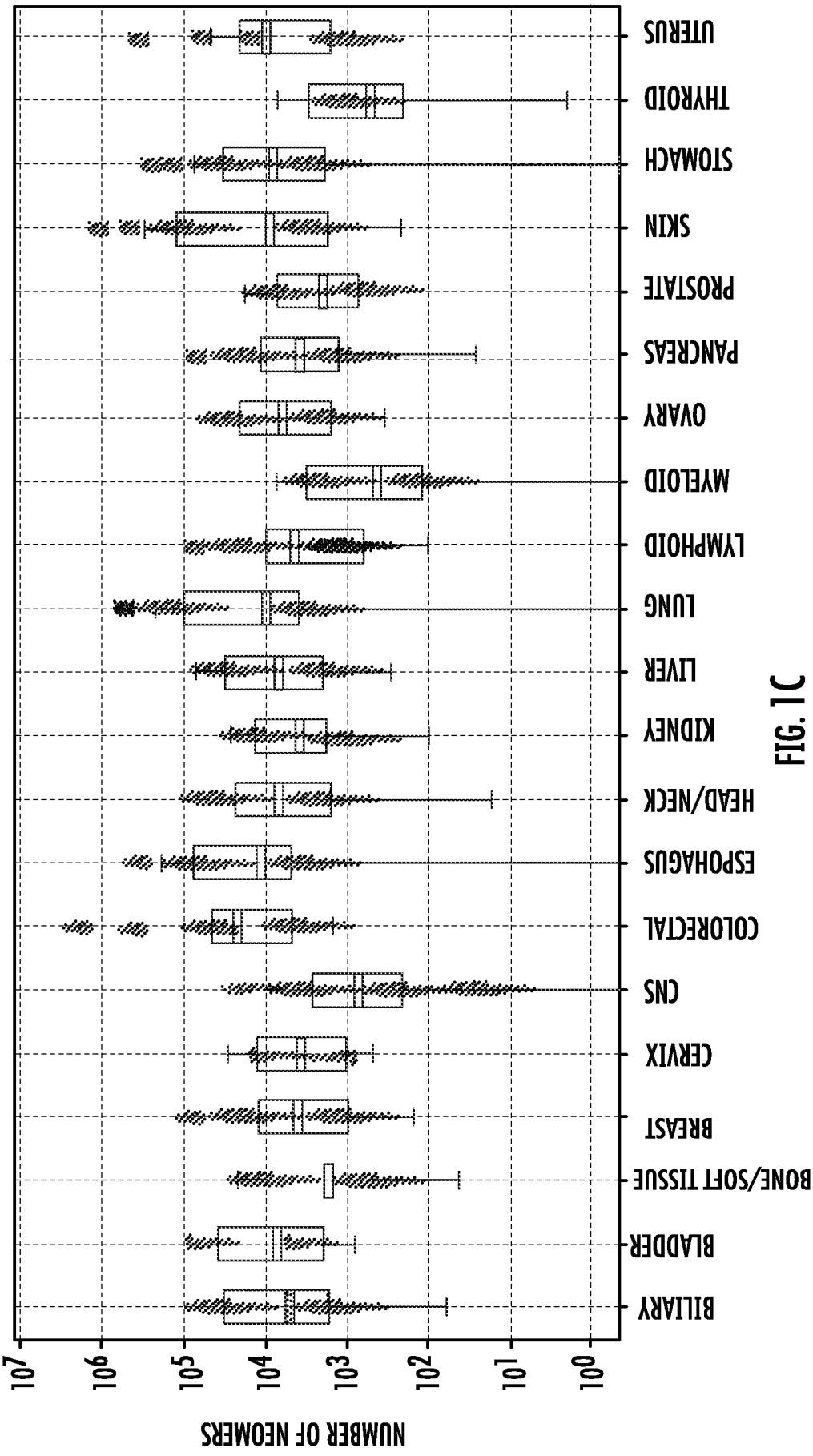

UTERUS
THYROID
STOMACH
SKIN
PROSTATE
PANCREAS
OVARY
MYELOID
LYMPHOID
LUNG
LIVER
KIDNEY
HEAD/NECK
ESPOHAGUS
COLORECTAL
CNS
CERVIX
BREAST
BONE/SOFT TISSUE
BLADDER
BILIARY
10^7
10^6
10^5
10^4
10^3
10^2
10^1
10^0
NUMBER OF NEOMERS

FIG. 1C

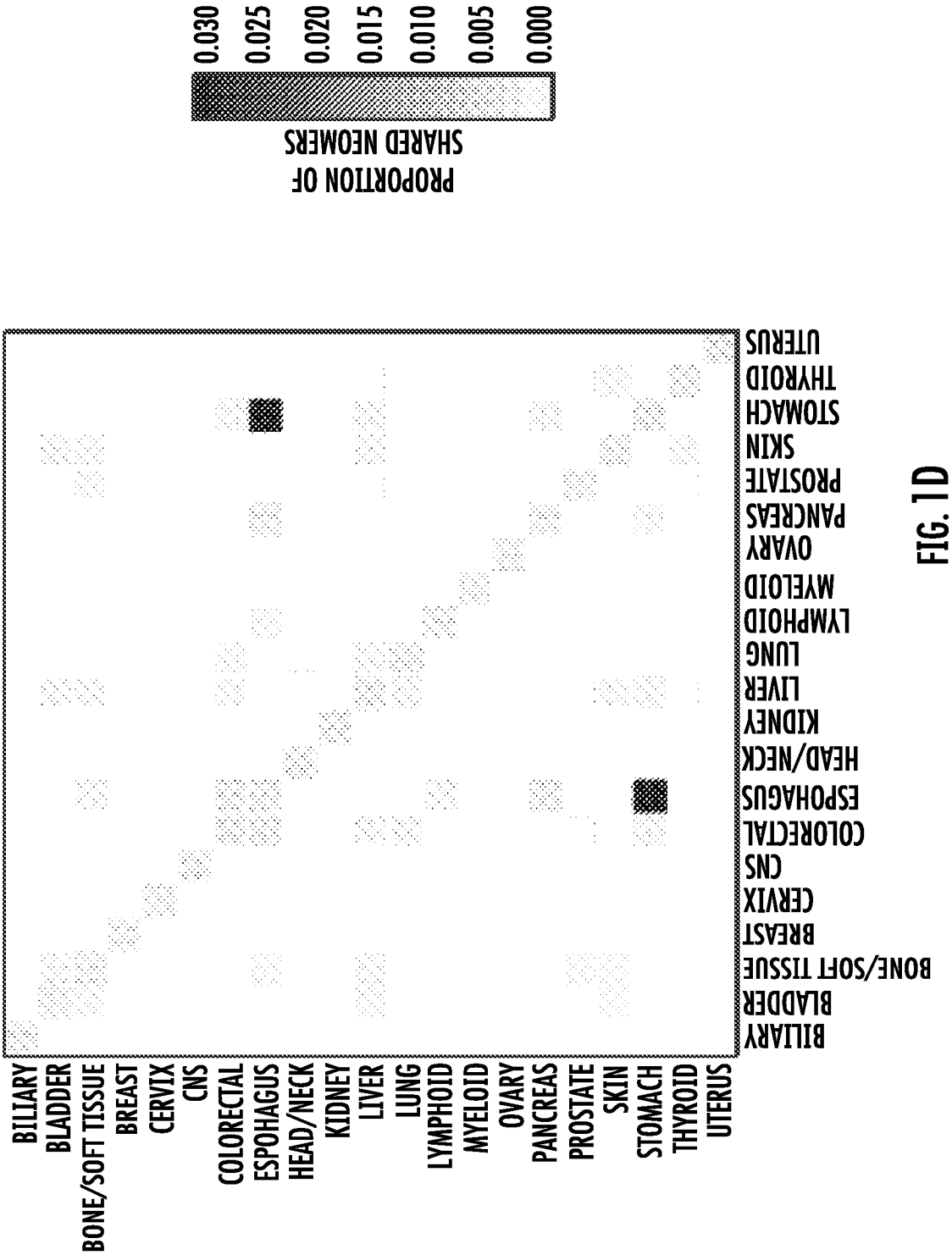

0.030
0.025
0.020
0.015
0.010
0.005
0.000
PROPORTION OF SHARED NEOMERS
BILIARY
BLADDER
BONE/SOFT TISSUE
BREAST
CERVIX
CNS
COLORECTAL
ESPOHAGUS
HEAD/NECK
KIDNEY
LIVER
LUNG
LYMPHOID
MYELOID
OVARY
PANCREAS
PROSTATE
SKIN
STOMACH
THYROID
UTERUS

FIG. 1D

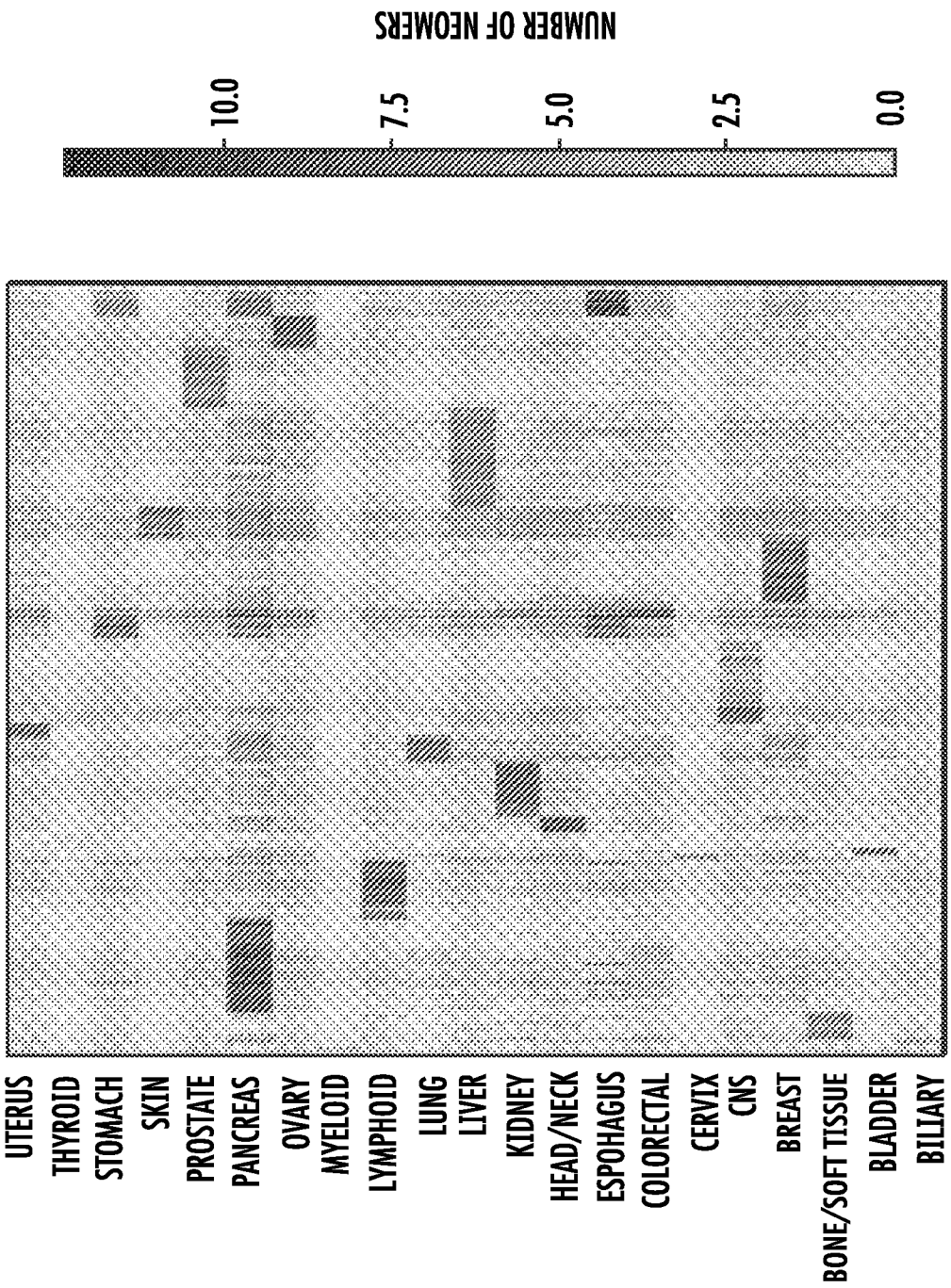

NUMBER OF NEOMERS
10.0
7.5
5.0
2.5
0.0
UTERUS
THYROID
STOMACH
SKIN
PROSTATE
PANCREAS
OVARY
MYELOID
LYMPHOID
LUNG
LIVER
KIDNEY
HEAD/NECK
ESPOHAGUS
COLORECTAL
CERVIX
CNS
BREAST
BONE/SOFT TISSUE
BLADDER
BILIARY
PATIENTS

FIG. 1E

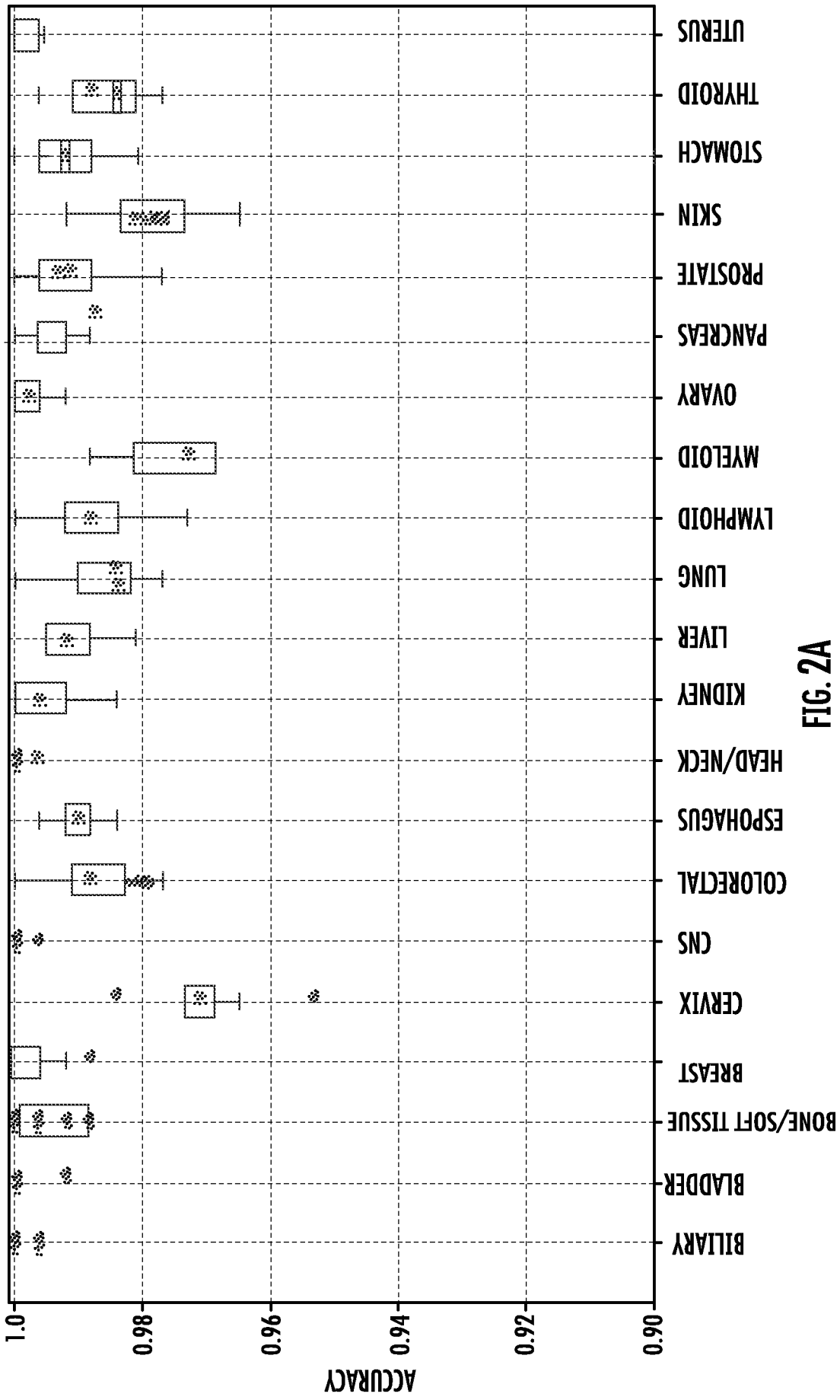

UTERUS
THYROID
STOMACH
SKIN
PROSTATE
PANCREAS
OVARY
MYELOID
LYMPHOID
LUNG
LIVER
KIDNEY
HEAD/NECK
ESPOHAGUS
COLORECTAL
CNS
CERVIX
BREAST
BONE/SOFT TISSUE
BLADDER
BILIARY
1.0
0.98
0.96
0.94
0.92
0.90
ACCURACY

FIG. 2A

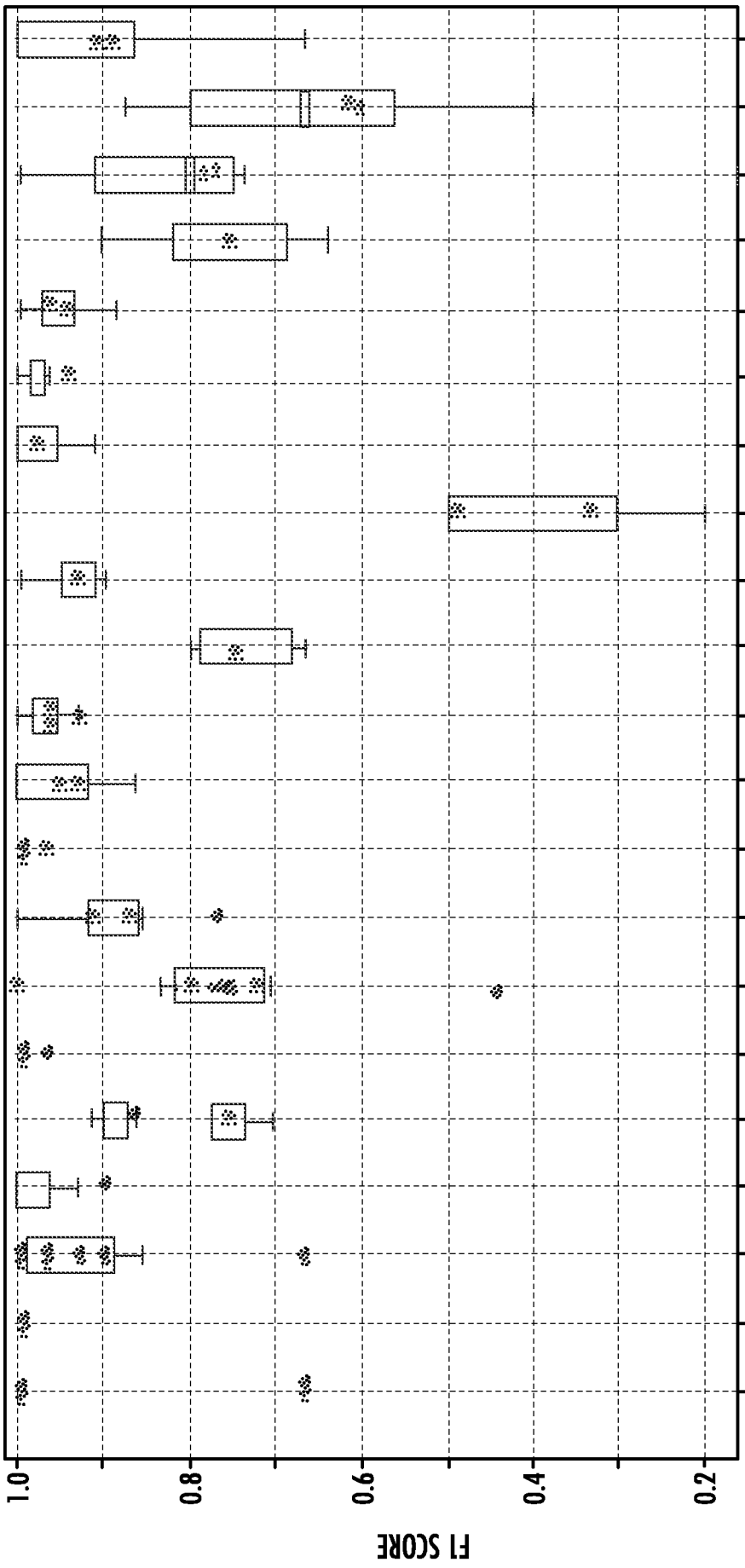

UTERUS
THYROID
STOMACH
SKIN
PROSTATE
PANCREAS
OVARY
MYELOID
LYMPHOID
LUNG
LIVER
KIDNEY
HEAD/NECK
ESPOHAGUS
COLORECTAL
CNS
CERVIX
BREAST
BONE/SOFT TISSUE
BLADDER
BILIARY
1.0
0.8
0.6
0.4
0.2
F1 SCORE

FIG. 2B

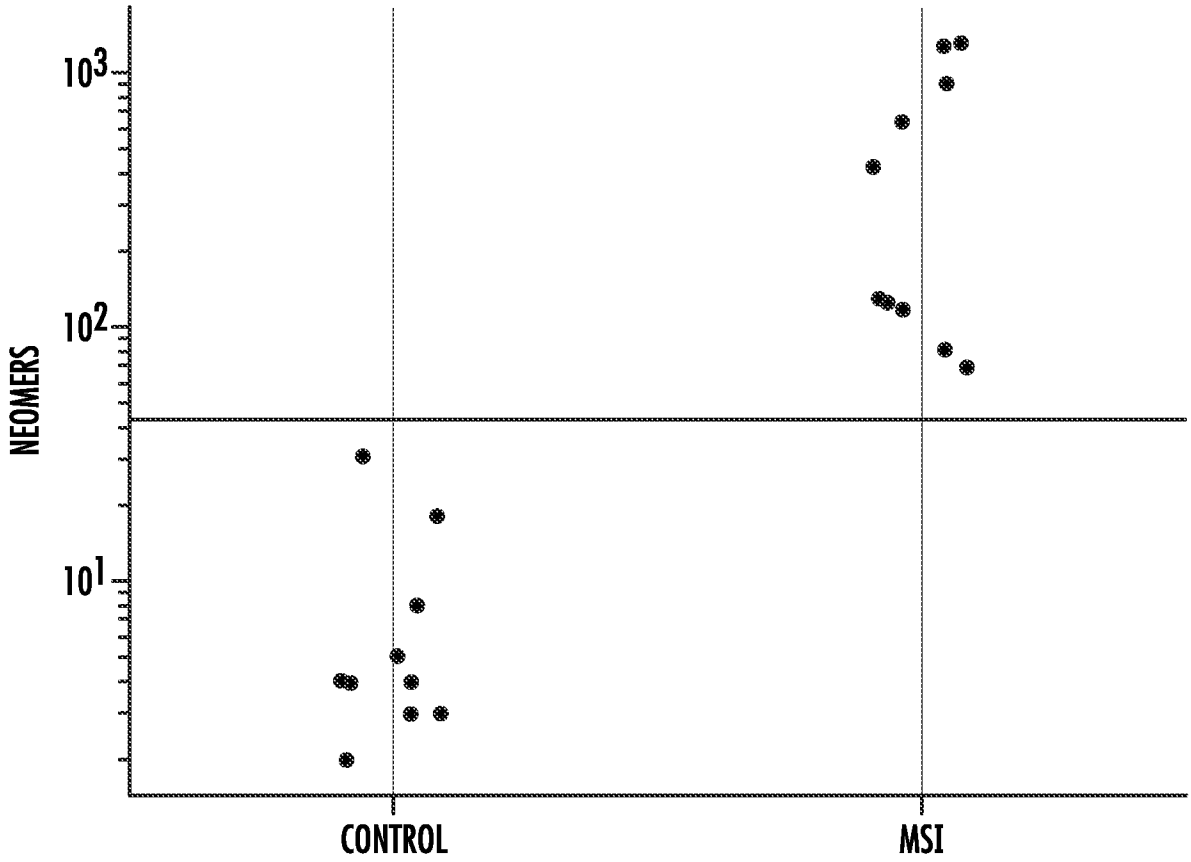


FIG. 2C

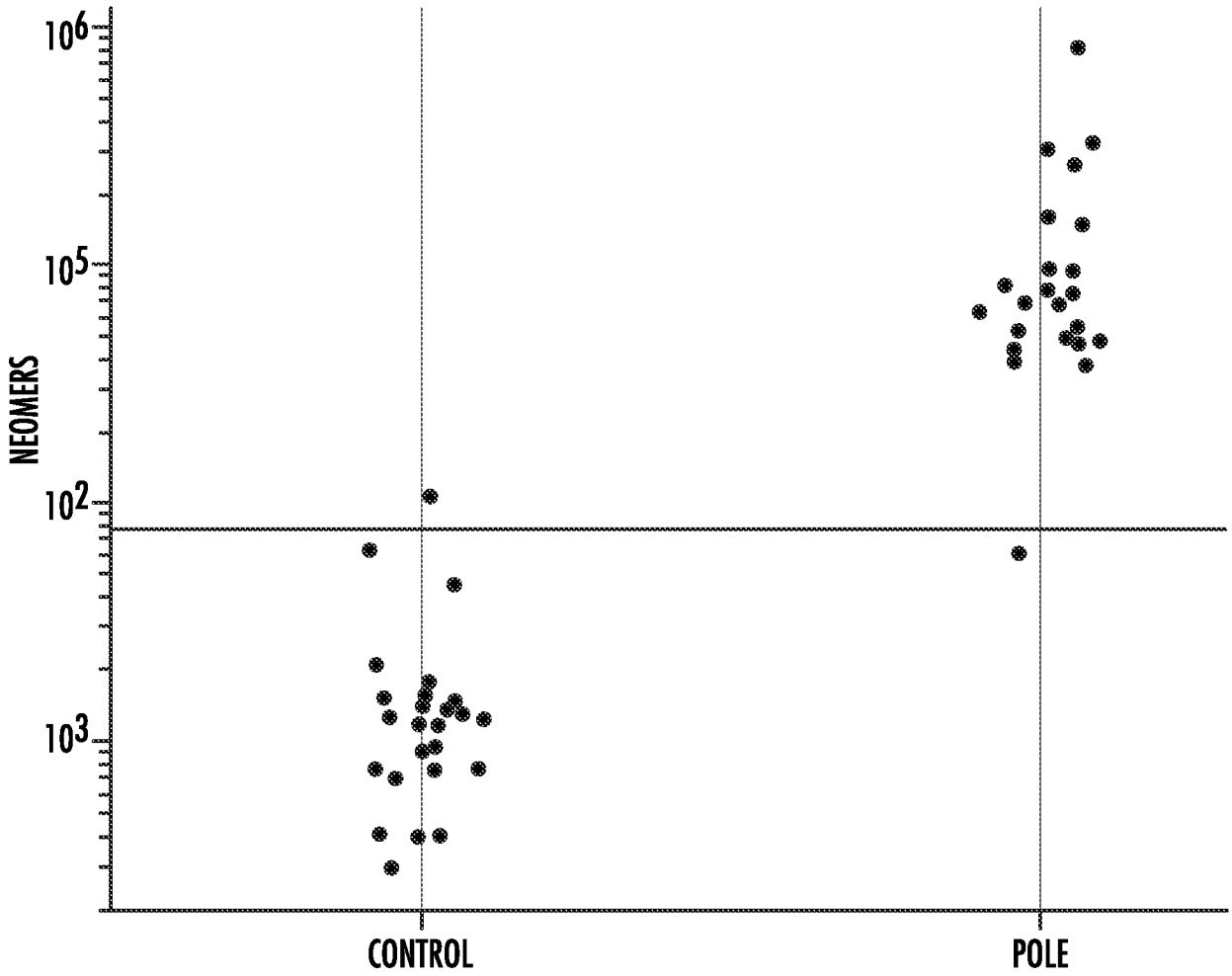


FIG. 2D

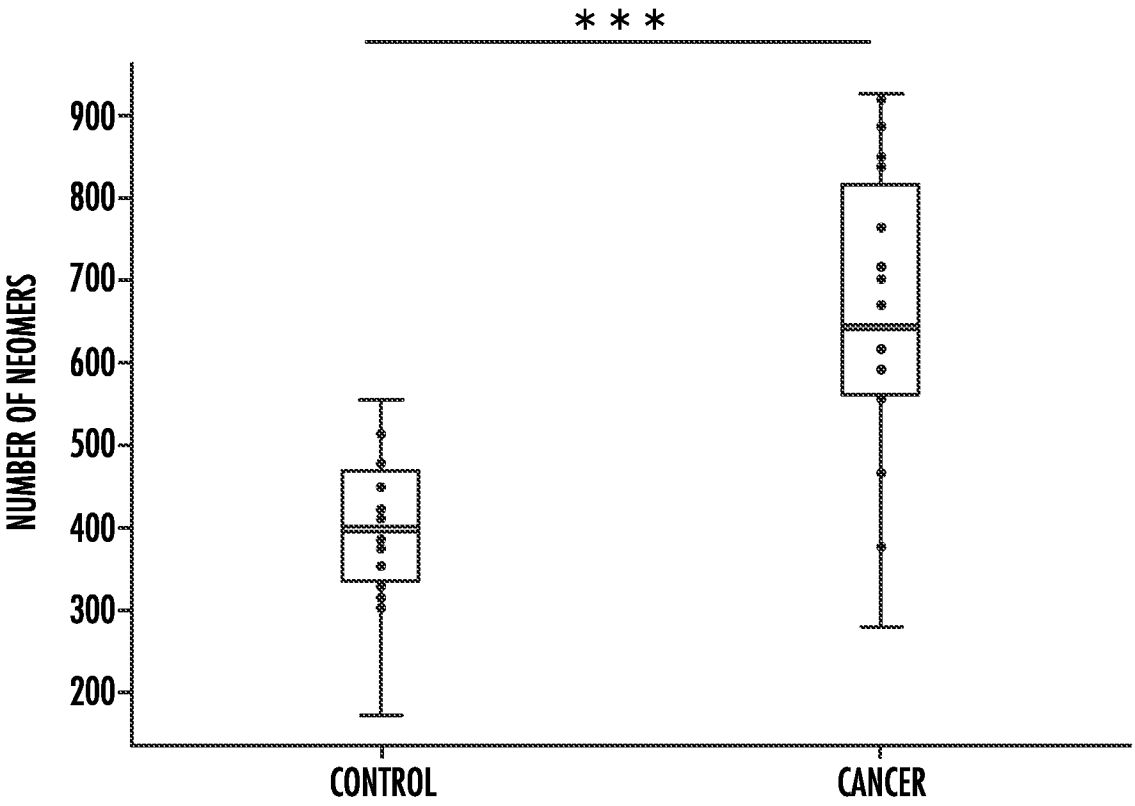

* * *
NUMBER OF NEOMERS
900
800
700
600
500
400
300
200
CONTROL
CANCER

FIG. 3A

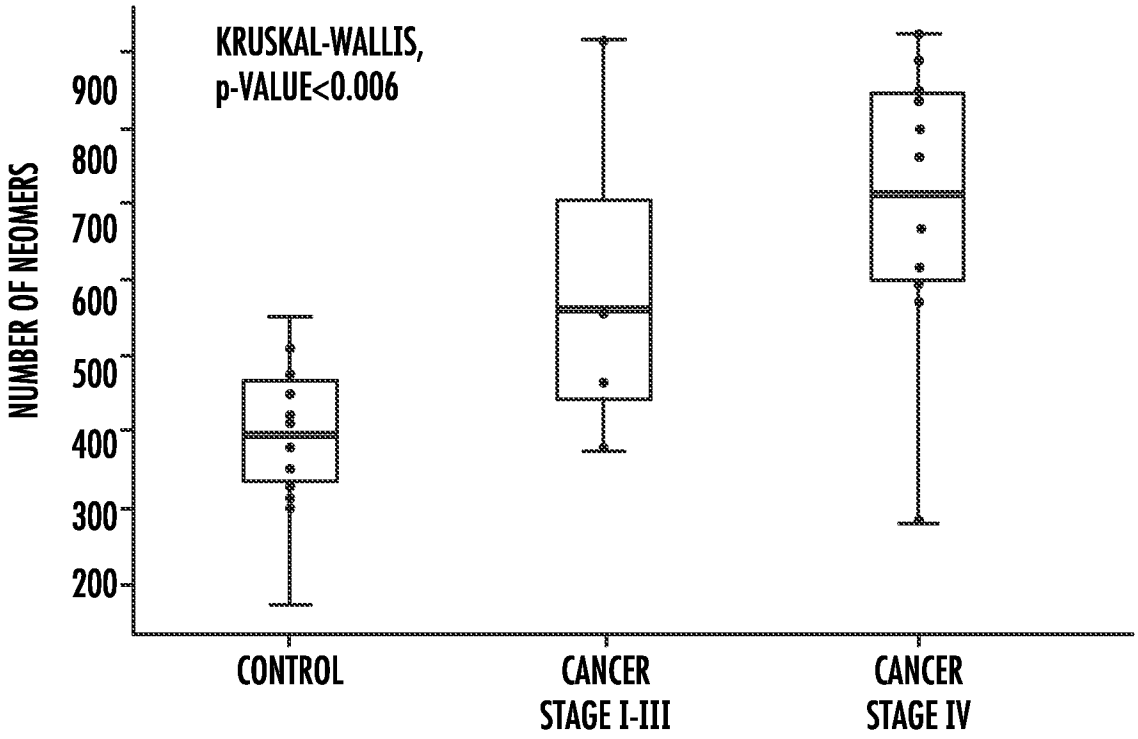

KRUSKAL-WALLIS,
p-VALUE<0.006
NUMBER OF NEOMERS
900
800
700
600
500
400
300
200
CONTROL
CANCER
STAGE I-III
CANCER
STAGE IV

FIG. 3B

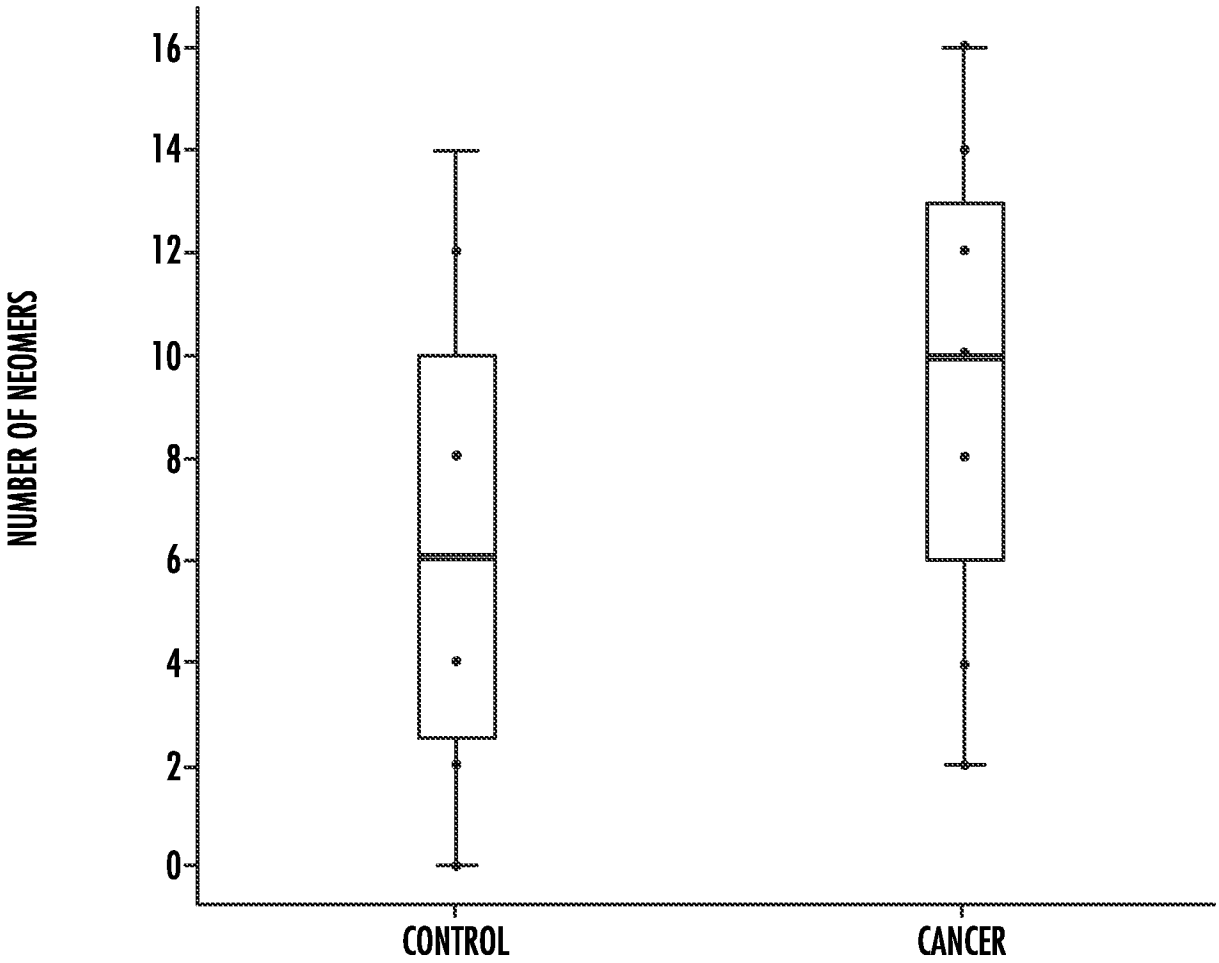


FIG. 3C

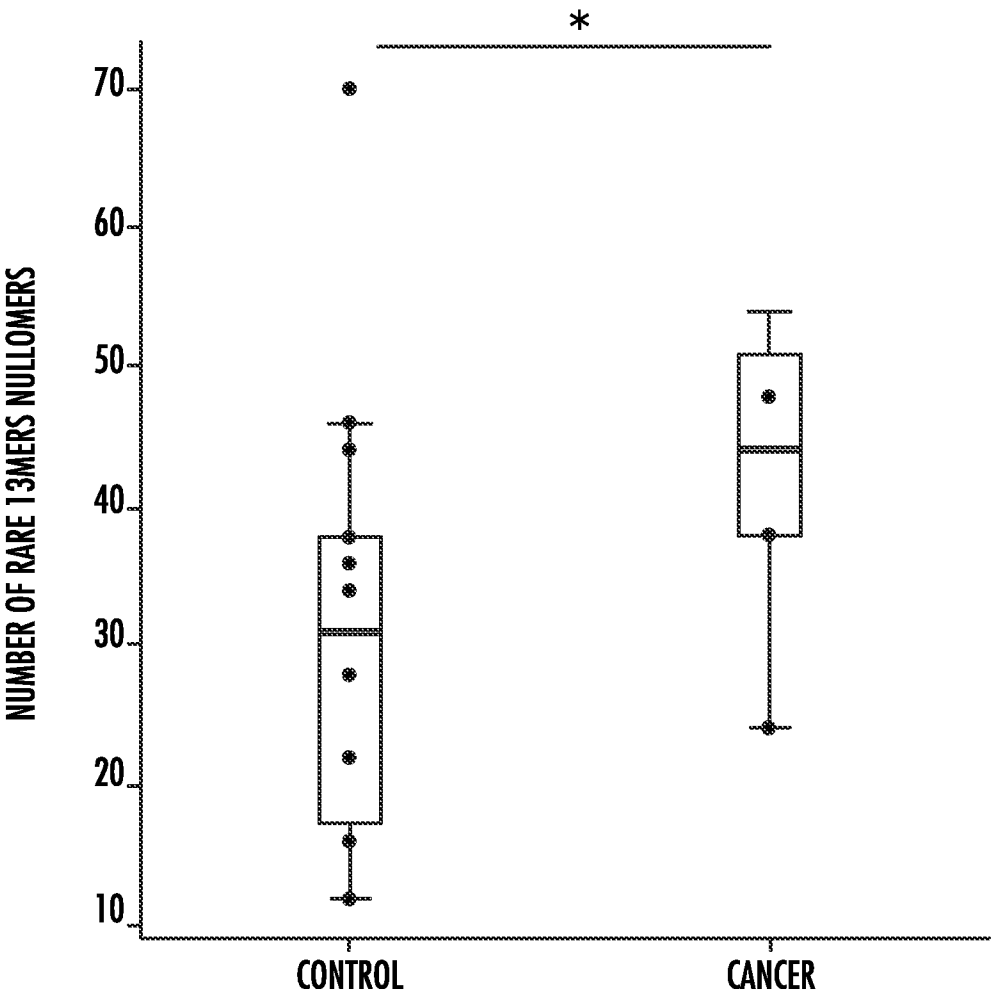

*
70
60
50
40
30
20
10
NUMBER OF RARE 13MERS NULLOMERS
CONTROL
CANCER

FIG. 3D

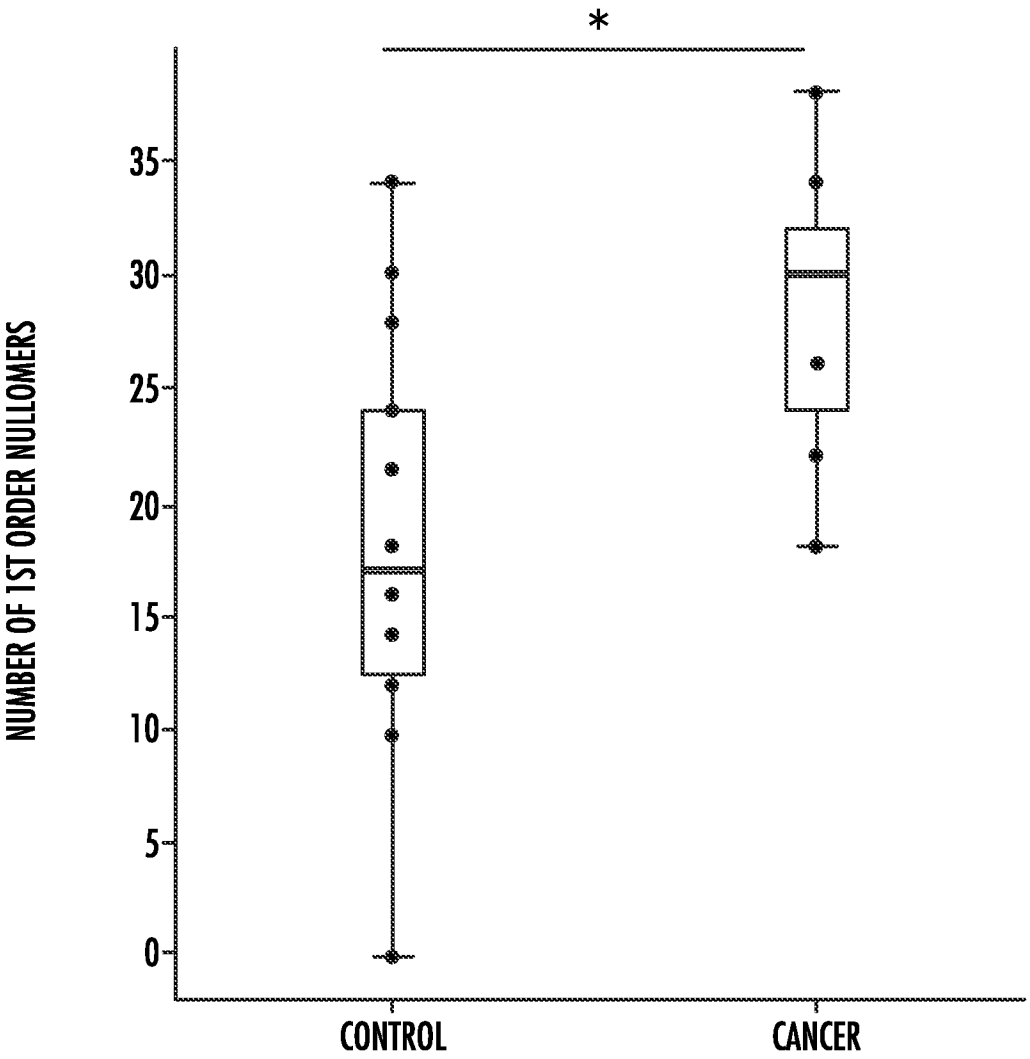

*
NUMBER OF 1ST ORDER NULLOMERS
35
30
25
20
15
10
5
0
CONTROL
CANCER

FIG. 3E

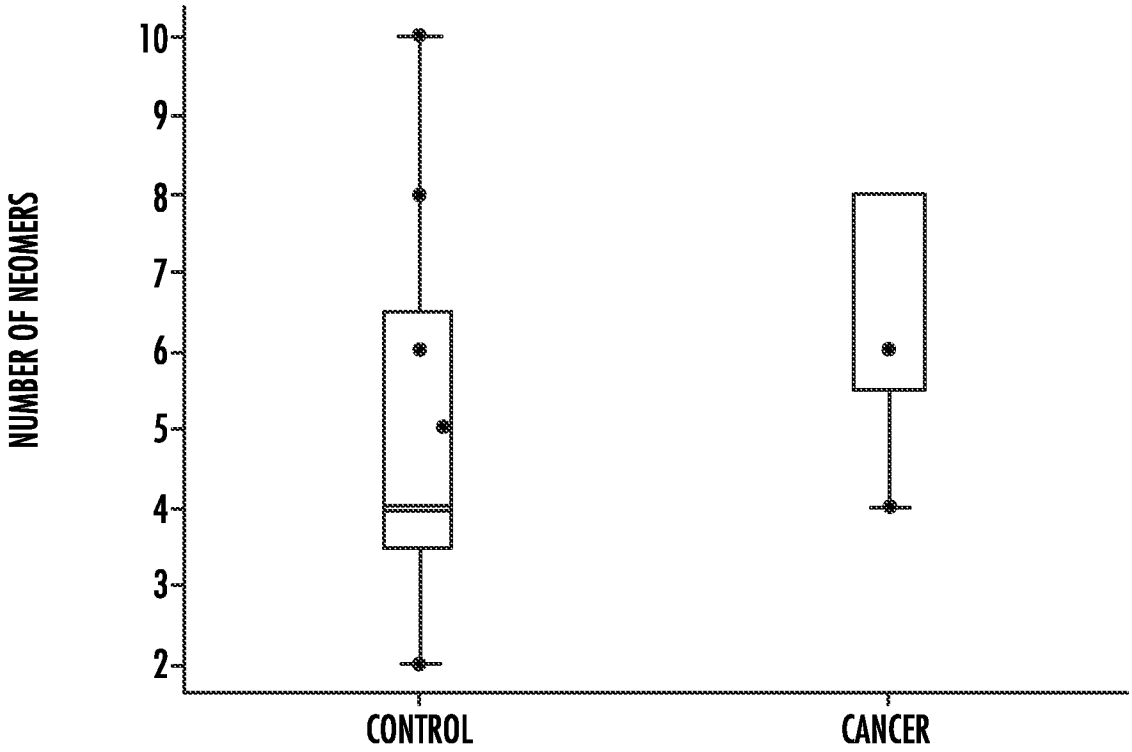


FIG. 3F

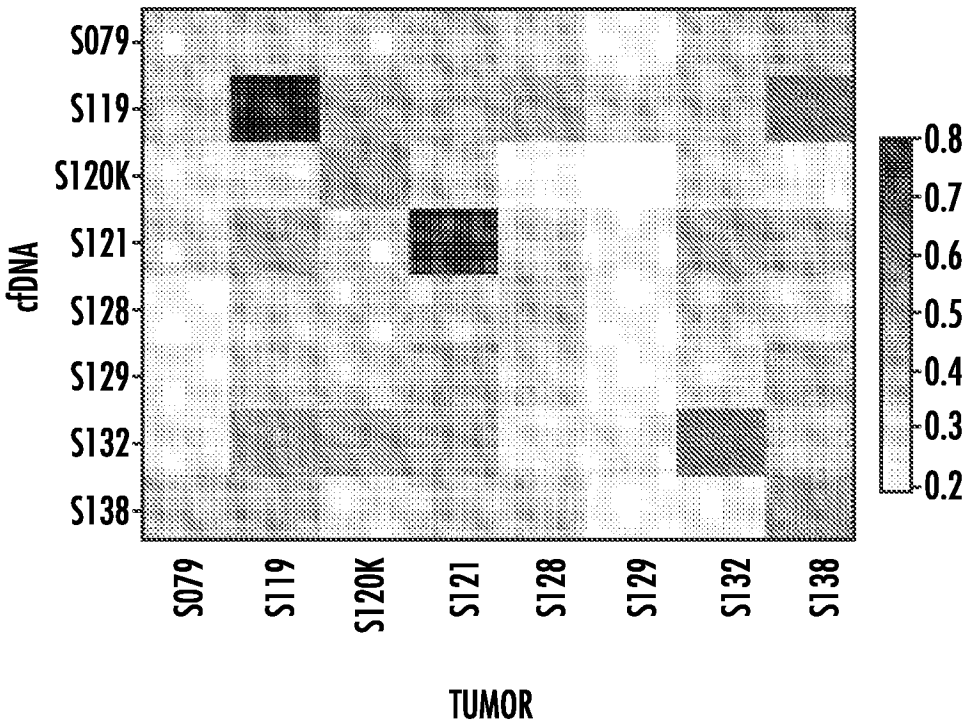

S079
S119
S120K
S121
S128
S129
S132
S138
cfDNA
S079
S119
S120K
S121
S128
S129
S132
S138
TUMOR
0.8
0.7
0.6
0.5
0.4
0.3
0.2

FIG. 3G

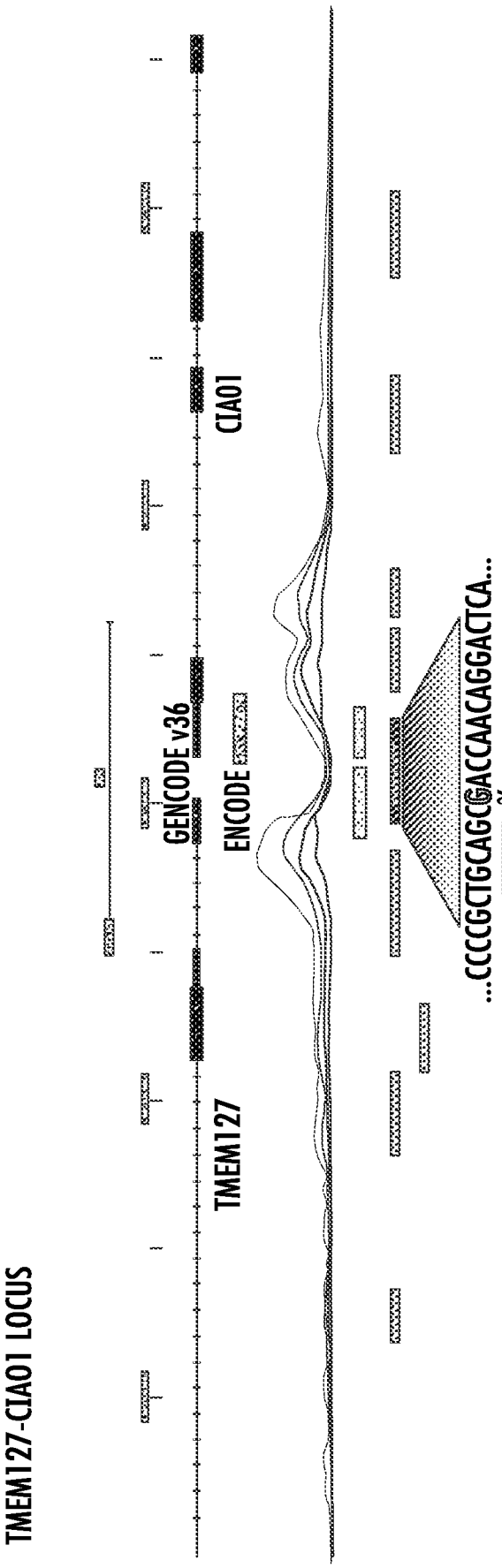

TMEM127-CIA01 LOCUS
TMEM127
ENCODE
GENCODE v36
CIA01
...CCCGCTGCAGCGACCAACAGGACTCA...
γ

FIG. 4A

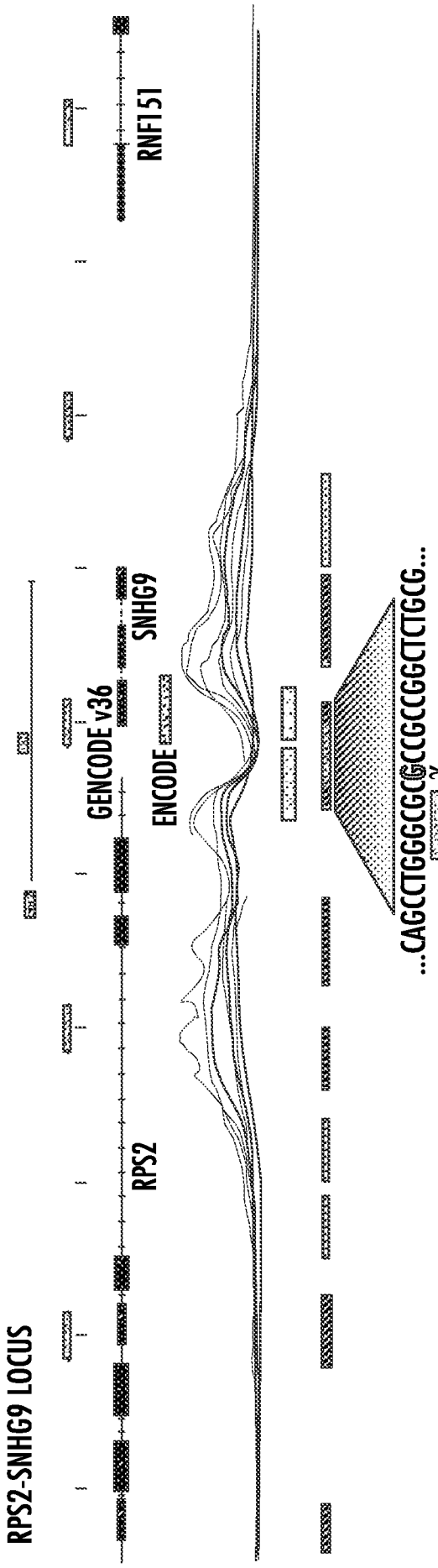

RPS2-SNHG9 LOCUS
RPS2
ENCODE
GENCODE v36
SNHG9
RNF151
...CAGCCTGGGCGCGCCGCCGGCTCTGCG...
γ

FIG. 4B

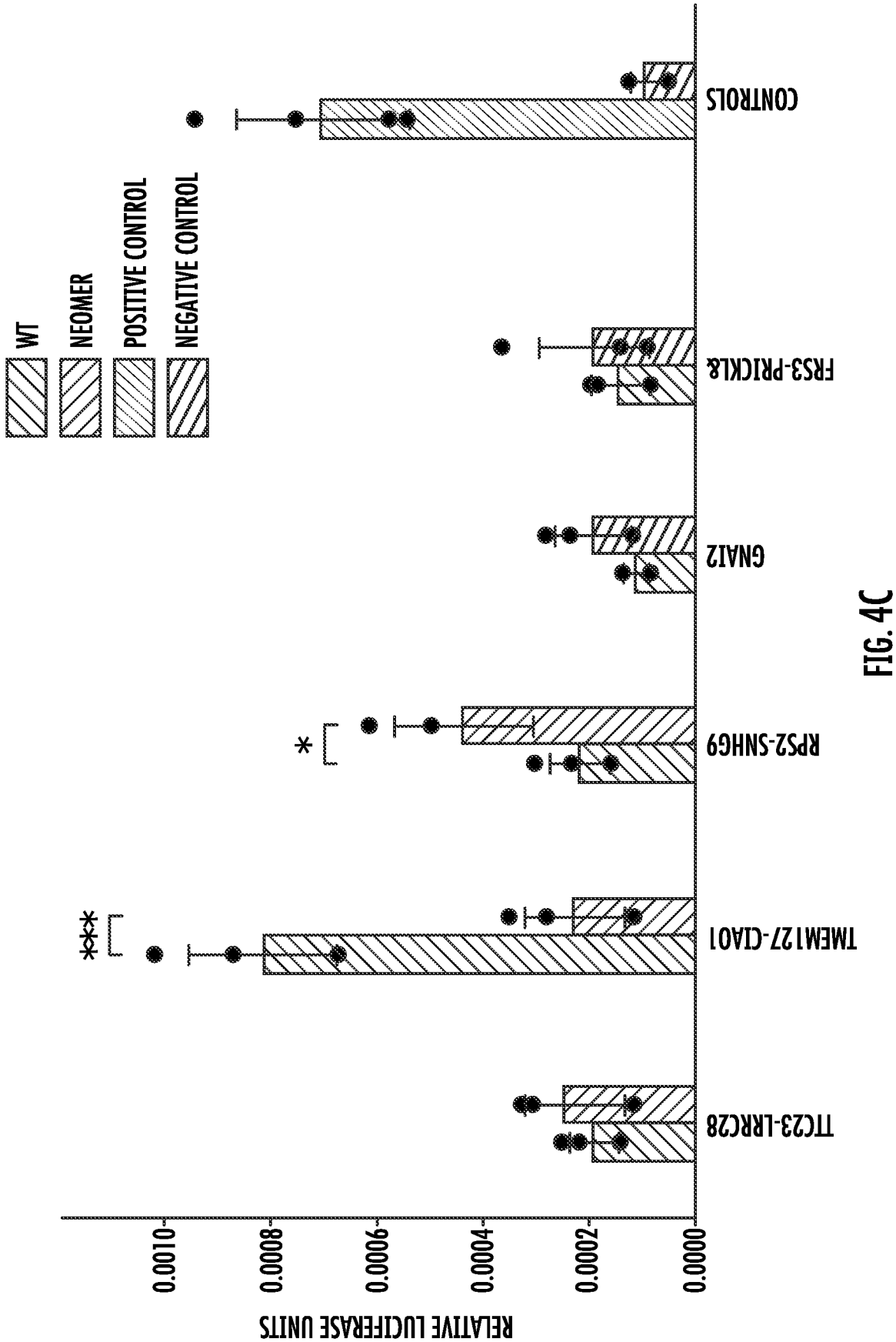

RELATIVE LUCIFERASE UNITS
0.0010
0.0008
0.0006
0.0004
0.0002
0.0000
TTC23-LRRC28
TMEM127-CIAO1
RPS2-SNHG9
GNAI2
FRS3-PRICKLE
CONTROLS
WT
NEOMER
POSITIVE CONTROL
NEGATIVE CONTROL
*
***

FIG. 4C

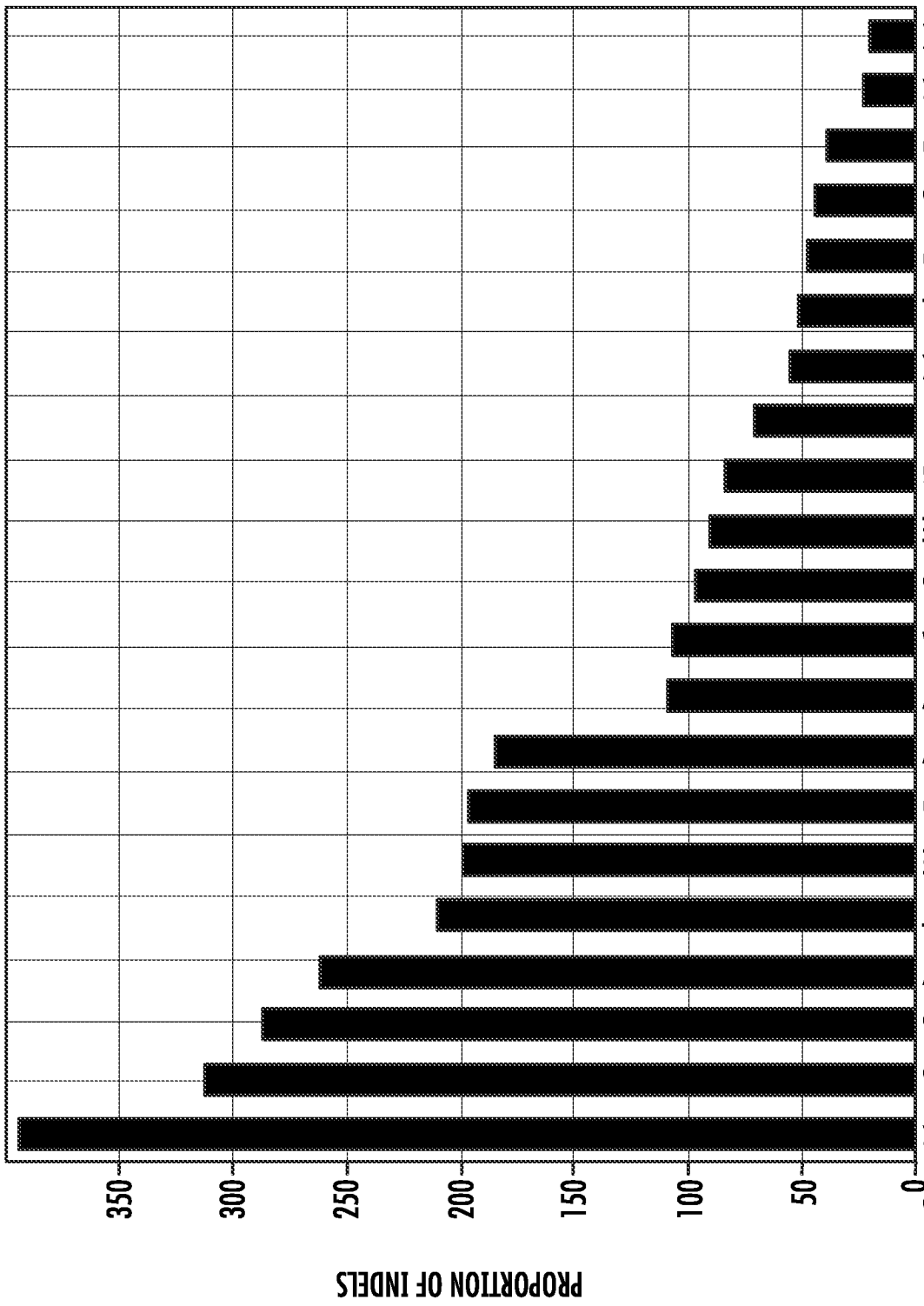

PROPORTION OF INDELS
0
50
100
150
200
250
300
350
CNS
LIVER
PANCREAS
CNS
BILIARY
BREAST
PROSTATE
LYMPHOID
KIDNEY
OVARY
SKIA
ESOPHAGUS
BONE_SOFTTISSUE
LUNG
STOMACH
HEAD_NECK
COLORECTAL
THYROID
UTERUS
MYELOID
BLADDER
CERVIX

FIG. 5A

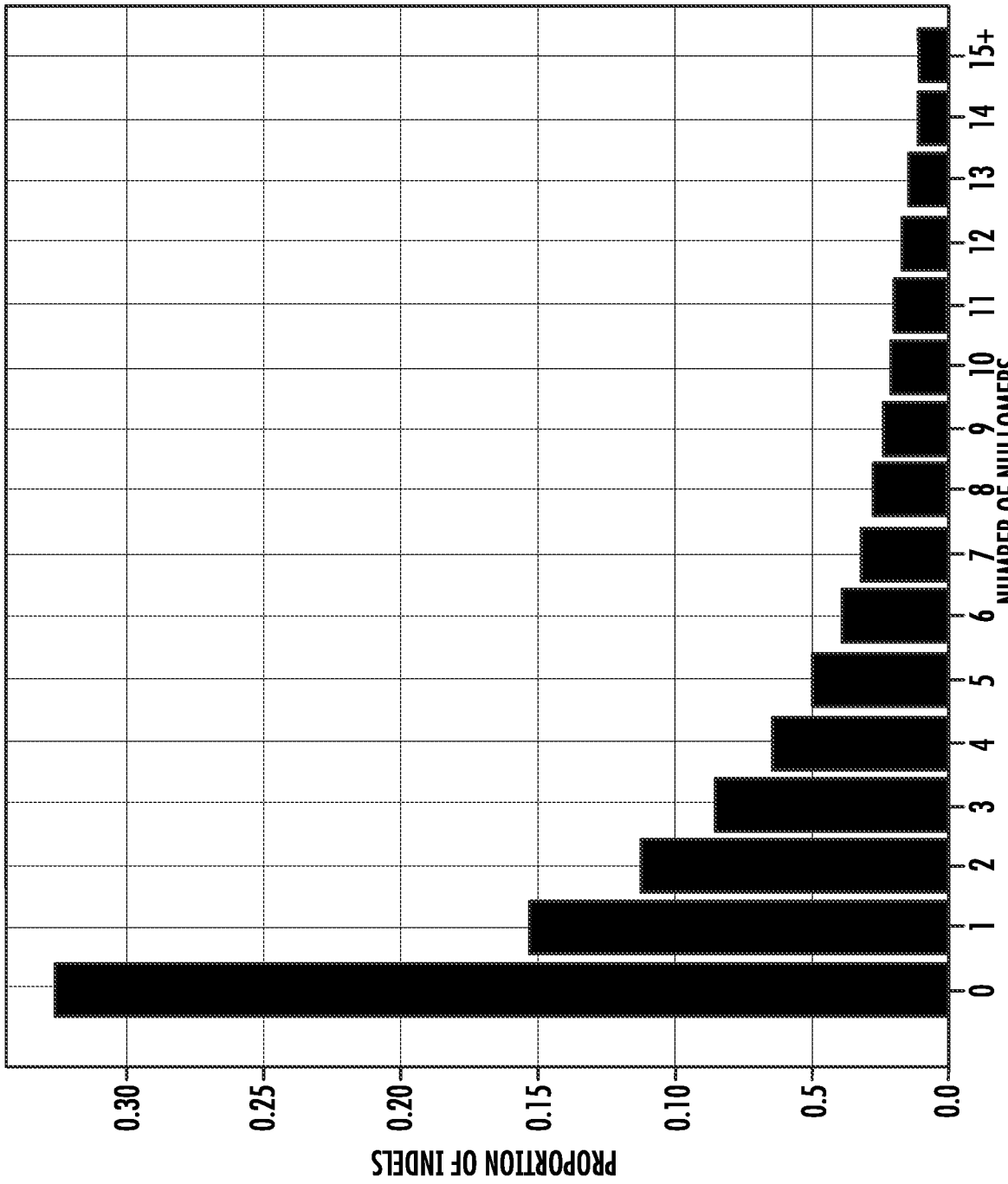

PROPORTION OF INDELS
0.30
0.25
0.20
0.15
0.10
0.5
0.0
NUMBER OF NULLOMERS
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15+

FIG. 5B

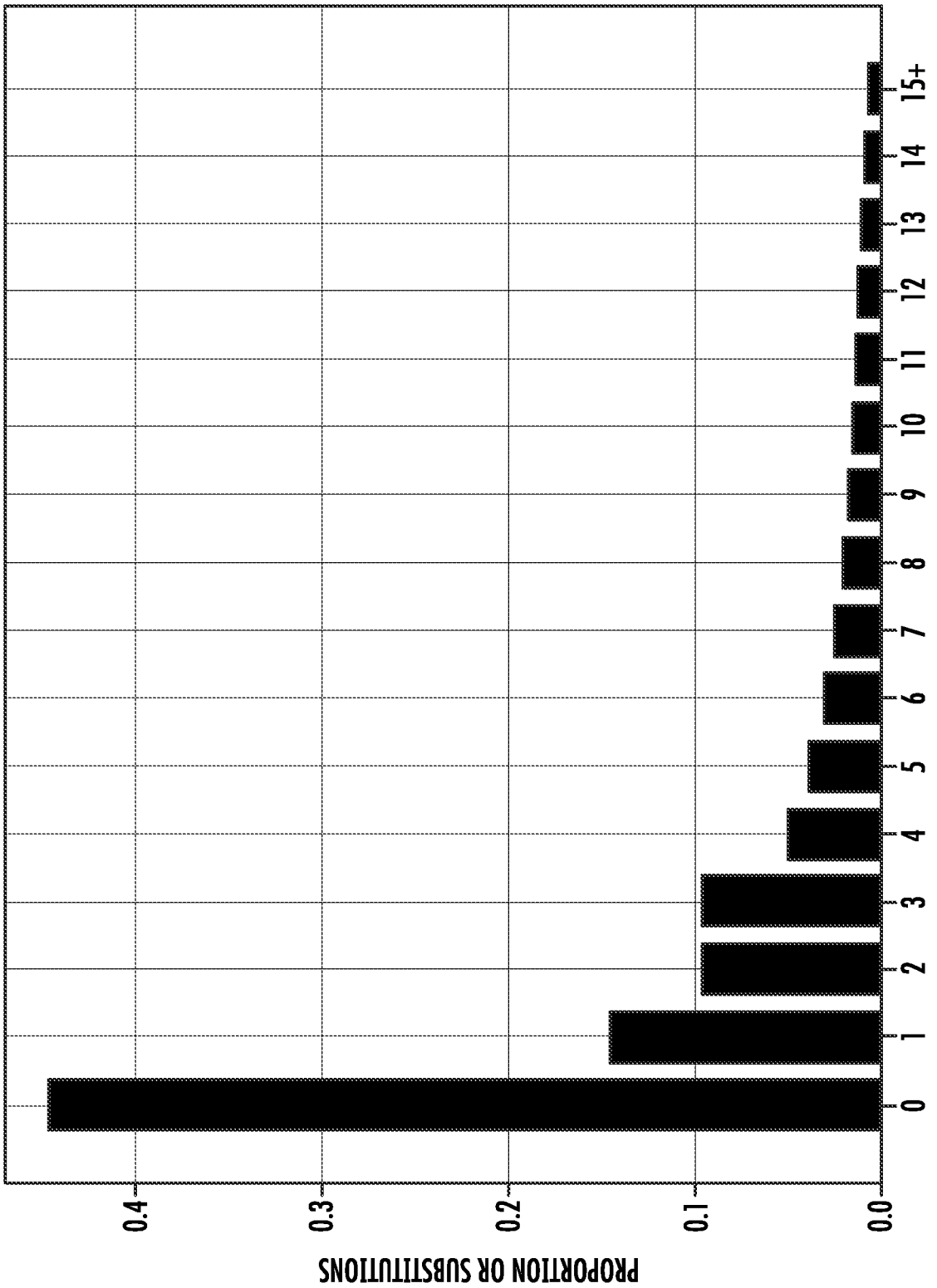

PROPORTION OR SUBSTITUTIONS
0.0
0.1
0.2
0.3
0.4
NUMBER OF NULLOMERS
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15+

FIG. 5C

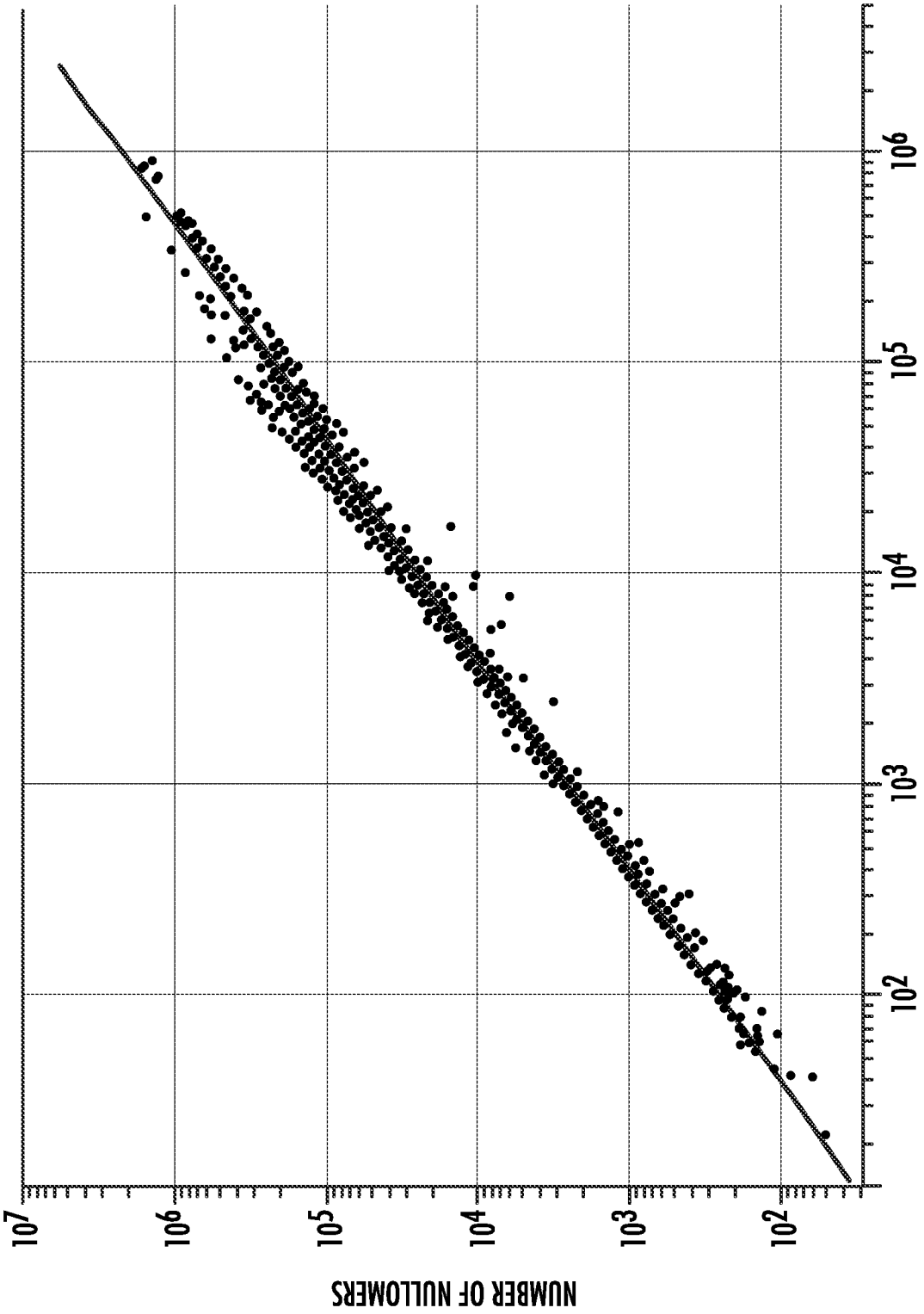

NUMBER OF NULLOMERS
10^7
10^6
10^5
10^4
10^3
10^2
NUMBER OF MUTATIONS
10^2
10^3
10^4
10^5
10^6

FIG. 5D

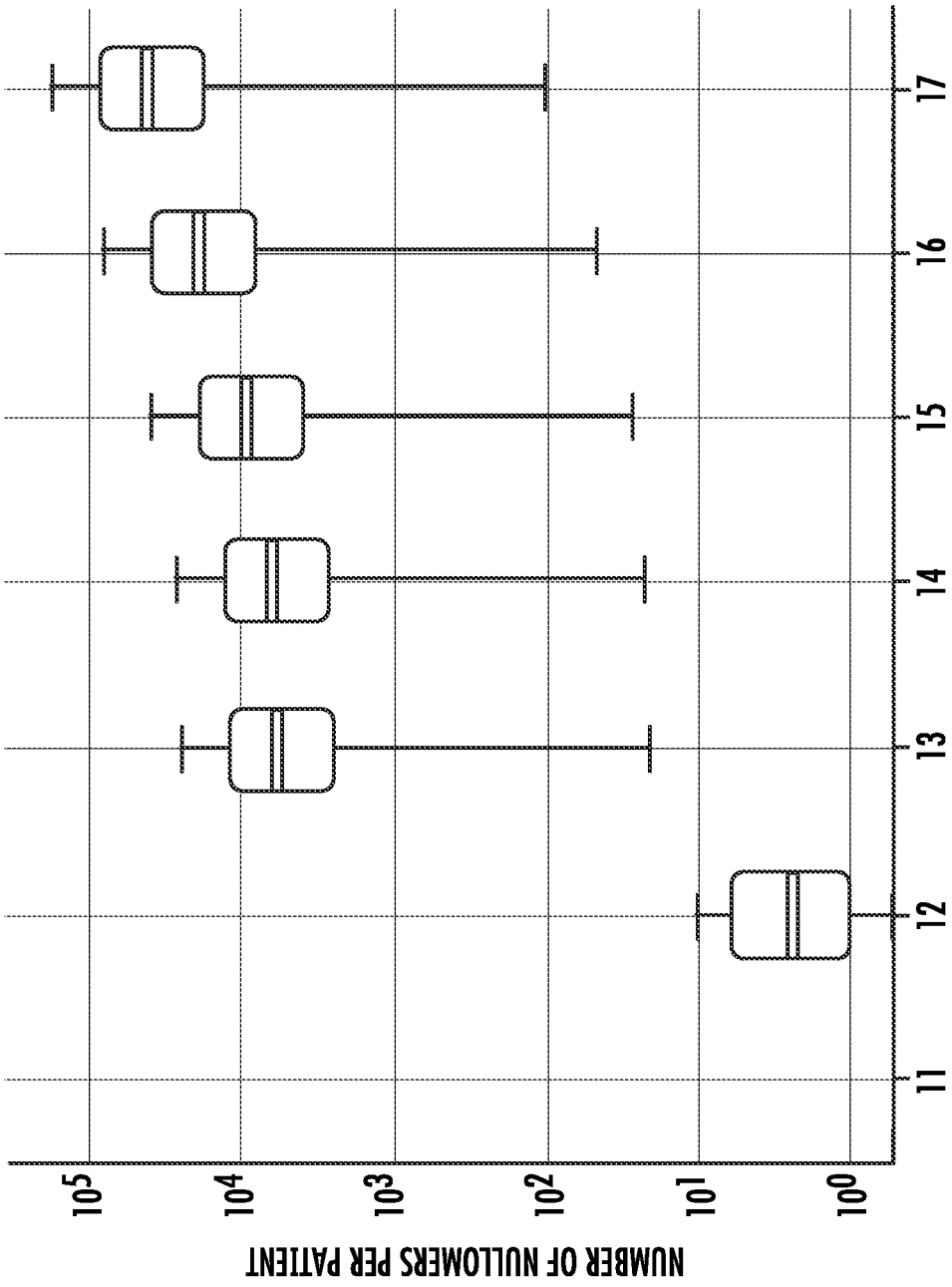

NUMBER OF NULLOMERS PER PATIENT
10^5
10^4
10^3
10^2
10^1
10^0
11
12
13
14
15
16
17
NULLOMER LENGTH

FIG. 5E