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

(19) World Intellectual Property

21 November 2019 (21.11.2019)

- (21) International Application Number: **KM, ML, MR, NE, SN, TD, TG**).
- **17** May **2019 (17.05.2019)**
- **(25) Filing Language:** English
- **(26)** Publication Language: English
- *(30)* **Priority Data: 62/673,516 18** May **2018 (18.05.2018) US 62/795,900 23** January **2019 (23.01.2019) US**
- **(71) Applicant: THE JOHNS HOPKINS UNIVERSITY** *[US/US];* 3400 **N.** Charles Street, Baltimore, **MD 21218 (US).**
- **(72) Inventors: VELCULESCU, Victor E.;** 14064 Big Branch Drive, Dayton, Maryland **21036 (US). CRISTIANO,** Stephen; 3400 North Charles Street, Baltimore, Mary land **21218 (US). LEAL,** Alessandro; 3400 North Charles Street, Baltimore, Maryland **21218 (US). PHALLEN, Jil lian A.; 2123** Moyer St., Baltimore, Maryland **21231 (US). FIKSEL, Jacob;** 3400 North Charles Street, Bal timore, Maryland **21218 (US). ADLEFF, Vilmos;** 3400 North Charles Street, Baltimore, Maryland **21218 (US). SCHARPF, Robert** B.; 3400 North Charles Street, Balti more, Maryland **21218 (US).**
- **(74) Agent: WILLIS, Margaret S.** et al.; Fish **&** Richardson **P.C.,** P.O. Box 1022, Minneapolis, Minnesota 55440-1022 **(US).**
- **(81) Designated** States (unless otherwise indicated, for every kind of national protection available): **AE, AG, AL,** AM, **AO, AT, AU,** AZ, BA, BB, BG, BH, **BN,** BR, BW, BY, BZ, **CA, CH, CL, CN, CO,** CR, **CU,** CZ, **DE, DJ,** DK, DM, **DO, DZ, EC, EE, EG, ES,** FI, GB, **GD, GE, GH, GM, GT, HN,** HR, **HU, ID,** IL, **IN,** IR, **IS, JO, JP,** KE, KG, KH, **KN,** KP, KR, KW, KZ, **LA, LC,** LK, LR, **LS, LU,** LY, MA, MD, ME, **MG,** MK, **MN,** MW, MX, MY, MZ, **NA, NG, NI, NO, NZ,** OM, PA, PE, **PG,** PH, PL, PT, **QA,** RO, RS, RU, RW, **SA, SC, SD, SE, SG,** SK, **SL, SM, ST, SV,** SY, TH, **TJ,** TM, **TN,** TR, TT, TZ, **UA, UG, US, UZ, VC, VN,** ZA, ZM, ZW.
- **(84) Designated States** (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, **GH, GM,** *KE,* LR, **LS,** MW, MZ, **NA,** RW, **SD, SL, ST,** SZ, TZ, **UG,** ZM, ZW), Eurasian (AM, *AZ,* BY, KG, KZ, RU, **TJ, TM),** European **(AL, AT,** BE, BG, **CH,** CY, CZ, **DE,** DK,

(54) **Title:** CELL-FREE **DNA** FOR **ASSESSING** AND/OR **TREATING CANCER**

(57) Abstract: This document relates to methods and materials for assessed, monitored, and/or treated mammals (e.g., humans) having cancer. For example, methods and materials for identifying a mammal as having cancer (e.g., a localized cancer) are provided. For example, methods and materials for assessing, monitoring, and/or treating a mammal having cancer are provided.

(1) Organization11111111111111111111111I1111111111111ii111liiiii

International Bureau **(10)** International Publication Number (43) International Publication Date

21 November 2019 (21.11.2019) WIPO | PCT

WO 2019/222657 A1

(51) International Patent Classification: EE, ES, FI, FR, GB, GR, HR, **HU, IE, IS,** IT, LT, **LU,** LV, *G01N33/50* **(2006.01)** *C12Q 1/6869* **(2018.01) MC,** MK, MT, **NL, NO,** PL, PT, RO, RS, **SE, SI,** SK, **SM,** *A61P35/00* **(2006.01)** *C12N15/07(2006.01)* TR), OAPI (BF, **BJ, CF, CG, CI, CM, GA, GN, GQ,** GW,

PCT/US2019/032914 Published:

(22) **International Filing Date:** $\qquad \qquad \qquad \qquad \qquad \qquad \qquad$ with international search report (Art. 21(3))

CELL-FREE DNA FOR ASSESSING AND/OR TREATING CANCER

CROSS-REFERENCE To RELATED APPLICATIONS

This application claims the benefit of **U.S.** Patent Application Serial No. **62/673,516,** filed on May **18, 2018,** and claims the benefit of **U.S.** Patent Application Serial No.

5 62/795,900, filed on January **23, 2019.** The disclosure of the prior applications are considered part of (and are incorporated **by** reference in) the disclosure of this application.

STATEMENT REGARDING FEDERAL FUNDING

This invention was made with **U.S.** government support under grant No. **CA121113** from the National Institutes of Health. The **U.S.** government has certain rights in the **10** invention.

BACKGROUND

1. Technical Field

This document relates to methods and materials for assessing and/or treating mammals (e.g., humans) having cancer. For example, this document provides methods and **¹⁵**materials for identifying a mammal as having cancer (e.g., a localized cancer). For example, this document provides methods and materials for monitoring and/or treating a mammal having cancer.

2. Background Information

Much of the morbidity and mortality of human cancers world-wide is a result of the **20** late diagnosis of these diseases, where treatments are less effective (Torre et al., **2015** *CA Cancer J Clin* **65:87;** and World Health Organization, **2017** *Guide to Cancer Early Diagnosis).* Unfortunately, clinically proven biomarkers that can be used to broadly diagnose and treat patients are not widely available (Mazzucchelli, 2000 *Advances in clinical pathology* **4:111;** Ruibal Morell, **1992** *The Internationaljournal of biological markers 7:160;*

²⁵Galli et al., **2013** *Clinical chemistry and laboratory medicine* **51:1369;** Sikaris, 2011 *Heart, lung & circulation* 20:634; Lin et al., **2016** in *Screeningfor Colorectal Cancer: A Systematic*

Review for the U.S. Preventive Services Task Force. (Rockville, MD); Wanebo et al., 1978 N Engl **J** Med 299:448; and Zauber, **2015** *Dig Dis Sci* **60:681).**

SUMMARY

- Recent analyses of cell-free **DNA** suggests that such approaches may provide new **⁵**avenues for early diagnosis (Phallen et al., **2017** *Sci TranslMed* **9;** Cohen et al., **2018** *Science* **359:926;** Alix-Panabieres et al., **2016** *Cancer discovery* **6:479;** Siravegna et al., **2017** *Nature reviews. Clinical oncology 14:531;* Haber et al., 2014 *Cancer discovery* 4:650; Husain et al., **2017 JAMA 318:1272;** and Wan et al., **2017** *Nat Rev* Cancer *17:223).*
- This document provides methods and materials for determining a cell free **DNA 10** (cfDNA) fragmentation profile in a mammal (e.g., in a sample obtained from a mammal). In some cases, determining a cfDNA fragmentation profile in a mammal can be used for identifying a mammal as having cancer. For example, cfDNA fragments obtained from a mammal (e.g., from a sample obtained from a mammal) can be subjected to low coverage whole-genome sequencing, and the sequenced fragments can be mapped to the genome (e.g.,
- **¹⁵**in non-overlapping windows) and assessed to determine a cfDNA fragmentation profile. This document also provides methods and materials for assessing and/or treating mammals (e.g., humans) having, or suspected of having, cancer. In some cases, this document provides methods and materials for identifying a mammal as having cancer. For example, a sample (e.g., a blood sample) obtained from a mammal can be assessed to determine if the mammal
- **20** has cancer based, at least in part, on the cfDNA fragmentation profile. In some cases, this document provides methods and materials for monitoring and/or treating a mammal having cancer. For example, one or more cancer treatments can be administered to a mammal identified as having cancer (e.g., based, at least in part, on a cfDNA fragmentation profile) to treat the mammal.
- **25** Described herein is a non-invasive method for the early detection and localization of cancer. cfDNA in the blood can provide a non-invasive diagnostic avenue for patients with cancer. As demonstrated herein, **DNA** Evaluation of Fragments for early Interception **(DELFI)** was developed and used to evaluate genome-wide fragmentation patterns of cfDNA of **236** patients with breast, colorectal, lung, ovarian, pancreatic, gastric, or bile duct cancers **30** as well as 245 healthy individuals. These analyses revealed that cfDNA profiles of healthy

individuals reflected nucleosomal fragmentation patterns of white blood cells, while patients with cancer had altered fragmentation profiles. **DELFI** had sensitivities of detection ranging from **57%** to **>99%** among the seven cancer types at **98%** specificity and identified the tissue of origin of the cancers to a limited number of sites in **75%** of cases. Assessing cfDNA (e.g., **⁵**using **DELFI)** can provide a screening approach for early detection of cancer, which can increase the chance for successful treatment of a patient having cancer. Assessing cfDNA (e.g., using **DELFI)** can also provide an approach for monitoring cancer, which can increase

the chance for successful treatment and improved outcome of a patient having cancer. In addition, a cfDNA fragmentation profile can be obtained from limited amounts of cfDNA and **10** using inexpensive reagents and/or instruments.

In general, one aspect of this document features methods for determining a cfDNA fragmentation profile of a mammal. The methods can include, or consist essentially of, processing cfDNA fragments obtained from a sample obtained from the mammal into sequencing libraries, subjecting the sequencing libraries to whole genome sequencing (e.g., **15** low-coverage whole genome sequencing) to obtain sequenced fragments, mapping the sequenced fragments to a genome to obtain windows of mapped sequences, and analyzing the windows of mapped sequences to determine cfDNA fragment lengths. The mapped sequences can include tens to thousands of windows. The windows of mapped sequences can be non-overlapping windows. The windows of mapped sequences can each include **20** about **5** million base pairs. The cfDNA fragmentation profile can be determined within each window. The cfDNA fragmentation profile can include a median fragment size. The cfDNA fragmentation profile can include a fragment size distribution. The cfDNA fragmentation profile can include a ratio of small cfDNA fragments to large cfDNA fragments in the windows of mapped sequences. The cfDNA fragmentation profile can be over the whole **25** genome. The cfDNA fragmentation profile can be over a subgenomic interval (e.g., an interval in a portion of a chromosome).

In another aspect, this document features methods for identifying a mammal as having cancer. The methods can include, or consist essentially of, determining a cfDNA

fragmentation profile in a sample obtained from a mammal, comparing the cfDNA **30** fragmentation profile to a reference cfDNA fragmentation profile, and identifying the mammal as having cancer when the cfDNA fragmentation profile in the sample obtained

from the mammal is different from the reference cfDNA fragmentation profile. The reference cfDNA fragmentation profile can be a cfDNA fragmentation profile of a healthy mammal. The reference cfDNA fragmentation profile can be generated **by** determining a cfDNA fragmentation profile in a sample obtained from the healthy mammal. The reference

- **5 DNA** fragmentation pattern can be a reference nucleosome cfDNA fragmentation profile. The cfDNA fragmentation profiles can include a median fragment size, and a median fragment size of the cfDNA fragmentation profile can be shorter than a median fragment size of the reference cfDNA fragmentation profile. The cfDNA fragmentation profiles can include a fragment size distribution, and a fragment size distribution of the cfDNA
- **10** fragmentation profile can differ **by** at least **10** nucleotides as compared to a fragment size distribution of the reference cfDNA fragmentation profile. The cfDNA fragmentation profiles can include position dependent differences in fragmentation patterns, including a ratio of small cfDNA fragments to large cfDNA fragments, where a small cfDNA fragment can be **100** base pairs **(bp)** to **150 bp** in length and a large cfDNA fragments can be **151 bp** to
- **¹⁵**220 **bp** in length, and where a correlation of fragment ratios in the cfDNA fragmentation profile can be lower than a correlation of fragment ratios of the reference cfDNA fragmentation profile. The cfDNA fragmentation profiles can include sequence coverage of small cfDNA fragments, large cfDNA fragments, or of both small and large cfDNA fragments, across the genome. The cancer can be colorectal cancer, lung cancer, breast
- **20** cancer, bile duct cancer, pancreatic cancer, gastric cancer, or ovarian cancer. The step of comparing can include comparing the cfDNA fragmentation profile to a reference cfDNA fragmentation profile in windows across the whole genome. The step of comparing can include comparing the cfDNA fragmentation profile to a reference cfDNA fragmentation profile over a subgenomic interval (e.g., an interval in a portion of a chromosome). The
- **25** mammal can have been previously administered a cancer treatment to treat the cancer. The cancer treatment can be surgery, adjuvant chemotherapy, neoadjuvant chemotherapy, radiation therapy, hormone therapy, cytotoxic therapy, immunotherapy, adoptive T cell therapy, targeted therapy, or any combinations thereof. The method also can include administering to the mammal a cancer treatment (e.g., surgery, adjuvant chemotherapy,
- **³⁰**neoadjuvant chemotherapy, radiation therapy, hormone therapy, cytotoxic therapy, immunotherapy, adoptive T cell therapy, targeted therapy, or any combinations thereof). The

mammal can be monitored for the presence of cancer after administration of the cancer treatment.

In another aspect, this document features methods for treating a mammal having cancer. The methods can include, or consist essentially of, identifying the mammal as having **⁵**cancer, where the identifying includes determining a cfDNA fragmentation profile in a sample obtained from the mammal, comparing the cfDNA fragmentation profile to a reference cfDNA fragmentation profile, and identifying the mammal as having cancer when the cfDNA fragmentation profile obtained from the mammal is different from the reference cfDNA fragmentation profile; and administering a cancer treatment to the mammal. The **¹⁰**mammal can be a human. The cancer can be colorectal cancer, lung cancer, breast cancer, gastric cancers, pancreatic cancers, bile duct cancers, or ovarian cancer. The cancer treatment can be surgery, adjuvant chemotherapy, neoadjuvant chemotherapy, radiation therapy, hormone therapy, cytotoxic therapy, immunotherapy, adoptive T cell therapy, targeted therapy, or combinations thereof. The reference cfDNA fragmentation profile can be

- **¹⁵**a cfDNA fragmentation profile of a healthy mammal. The reference cfDNA fragmentation profile can be generated **by** determining a cfDNA fragmentation profile in a sample obtained from a healthy mammal. The reference **DNA** fragmentation pattern can be a reference nucleosome cfDNA fragmentation profile. The cfDNA fragmentation profile can include a median fragment size, where a median fragment size of the cfDNA fragmentation profile is
- **20** shorter than a median fragment size of the reference cfDNA fragmentation profile. The cfDNA fragmentation profile can include a fragment size distribution, where a fragment size distribution of the cfDNA fragmentation profile differs **by** at least **10** nucleotides as compared to a fragment size distribution of the reference cfDNA fragmentation profile. The cfDNA fragmentation profile can include a ratio of small cfDNA fragments to large cfDNA
- **25** fragments in the windows of mapped sequences, where a small cfDNA fragment is **100 bp** to **150 bp** in length, where a large cfDNA fragments is **151 bp** to 220 **bp** in length, and where a correlation of fragment ratios in the cfDNA fragmentation profile is lower than a correlation of fragment ratios of the reference cfDNA fragmentation profile. The cfDNA fragmentation profile can include the sequence coverage of small cfDNA fragments in windows across the

30 genome. The cfDNA fragmentation profile can include the sequence coverage of large cfDNA fragments in windows across the genome. The cfDNA fragmentation profile can

include the sequence coverage of small and large cfDNA fragments in windows across the genome. The step of comparing can include comparing the cfDNA fragmentation profile to a reference cfDNA fragmentation profile over the whole genome. The step of comparing can include comparing the cfDNA fragmentation profile to a reference cfDNA fragmentation

⁵profile over a subgenomic interval. The mammal can have previously been administered a cancer treatment to treat the cancer. The cancer treatment can be surgery, adjuvant chemotherapy, neoadjuvant chemotherapy, radiation therapy, hormone therapy, cytotoxic therapy, immunotherapy, adoptive T cell therapy, targeted therapy, or combinations thereof The method also can include monitoring the mammal for the presence of cancer after **¹⁰**administration of the cancer treatment.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood **by** one of ordinary skill in the art to which this invention pertains. Although methods and materials similar or equivalent to those described herein can be used to practice the invention, suitable methods and materials are described below. **All 15** publications, patent applications, patents, and other references mentioned herein are incorporated **by** reference in their entirety. In case of conflict, the present specification,

including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

The details of one or more embodiments of the invention are set forth in the **20** accompanying drawings and the description below. Other features, objects, and advantages of the invention will be apparent from the description and drawings, and from the claims.

DESCRIPTION OF THE DRAWINGS

Figure **1.** Schematic of an exemplary **DELFI** approach. Blood is collected from a cohort of healthy individuals and patients with cancer. Nucleosome protected cfDNA is **25** extracted from the plasma fraction, processed into sequencing libraries, examined through whole genome sequencing, mapped to the genome, and analyzed to determine cfDNA fragment profiles in different windows across the genome. Machine learning approaches are used to categorize individuals as healthy or as having cancer and to identify the tumor tissue of origin using genome-wide cfDNA fragmentation patterns.

Figure 2. Simulations of non-invasive cancer detection based on number of alterations analyzed and tumor-derived cfDNA fragment distributions. Monte Carlo simulations were performed using different numbers of tumor-specific alterations to evaluate the probability of detecting cancer alterations in cfDNA at the indicated fraction of tumor

7

- **⁵**derived molecules. The simulations were performed assuming an average of 2000 genome equivalents of cfDNA and the requirement of five or more observations of any alteration. These analyses indicate that increasing the number of tumor-specific alterations improves the sensitivity of detection of circulating tumor **DNA.**
- Figure **3.** Tumor-derived cfDNA fragment distributions. Cumulative density **¹⁰**functions of cfDNA fragment lengths of 42 loci containing tumor-specific alterations from **30** patients with breast, colorectal, lung, or ovarian cancer are shown with **95%** confidence bands (blue). Lengths of mutant cfDNA fragments were significantly different in size compared to wild-type cfDNA fragments (red) at these loci.

Figures 4A and 4B. Tumor-derived cfDNA **GC** content and fragment length. **A, GC ¹⁵**content was similar for mutated and non-mutated fragments. B, **GC** content was not correlated to fragment length.

Figure **5.** Germline cfDNA fragment distributions. Cumulative density functions of fragment lengths of 44 loci containing germline alterations (non-tumor derived) from **38** patients with breast, colorectal, lung, or ovarian cancer are shown with **95%** confidence **²⁰**bands. Fragments with germline mutations (blue) were comparable in length to wild-type

cfDNA fragment lengths (red).

Figure **6.** Hematopoietic cfDNA fragment distributions. Cumulative density functions of fragment lengths of 41 loci containing hematopoietic alterations (non-tumor derived) from **28** patients with breast, colorectal, lung, or ovarian cancer are shown with **95%** 25 confidence bands. After correction for multiple testing, there were no significant differences $(\alpha=0.05)$ in the size distributions of mutated hematopoietic cfDNA fragments (blue) and wild-type cfDNA fragments (red).

Figures **7A - 7F.** cfDNA fragmentation profiles in healthy individuals and patients with cancer. A, Genome-wide cfDNA fragmentation profiles (defined as the ratio of short to **³⁰**long fragments) from -9x whole genome sequencing are shown in **5 Mb** bins for **30** healthy individuals (top) and **8** lung cancer patients (bottom). B, An analysis of healthy cfDNA

(top), lung cancer cfDNA (middle), and healthy lymphocyte (bottom) fragmentation profiles and lymphocyte profiles from chromosome 1 at 1 **Mb** resolution. The healthy lymphocyte profiles were scaled with a standard deviation equal to that of the median healthy cfDNA profiles. Healthy cfDNA patterns closely mirrored those in healthy lymphocytes while lung

- **⁵**cancer cfDNA profiles were more varied and differed from both healthy and lymphocyte profiles. **C,** Smoothed median distances between adjacent nucleosome centered at zero using **100 kb** bins from healthy cfDNA (top) and nuclease-digested healthy lymphocytes (middle) are depicted together with the first eigenvector for the genome contact matrix obtained through previously reported Hi-C analyses of lymphoblastoid cells (bottom). Healthy
- **10** cfDNA nucleosome distances closely mirrored those in nuclease-digested lymphocytes as well as those from lymphoblastoid Hi-C analyses. cfDNA fragmentation profiles from healthy individuals (n=30) had high correlations while patients with lung cancer had lower correlations to median fragmentation profiles of lymphocytes **(D),** healthy cfDNA **(E),** and lymphocyte nucleosome (F) distances.
-

¹⁵Figure **8.** Density of cfDNA fragment lengths in healthy individuals and patients with lung cancer. cfDNA fragments lengths are shown for healthy individuals (n=30, gray) and patients with lung cancer (n=8, blue).

Figures **9A** and 9B. Subsampling of whole genome sequence data for analysis of cfDNA fragmentation profiles. **A,** High coverage (9x) whole-genome sequencing data were **20** subsampled to 2x, 1x, 0.5x, 0.2x, and **0.1x** fold coverage. Mean centered genome-wide fragmentation profiles in **5 Mb** bins for **30** healthy individuals and **8** patients with lung cancer are depicted for each subsampled fold coverage with median profiles shown in blue. B, Pearson correlation of subsampled profiles to initial profile at 9x coverage for healthy individuals and patients with lung cancer.

25 Figure **10.** cfDNA fragmentation profiles and sequence alterations during therapy. Detection and monitoring of cancer in serial blood draws from **NSCLC** patients **(n=19)** undergoing treatment with targeted tyrosine kinase inhibitors (black arrows) was performed using targeted sequencing (top) and genome-wide fragmentation profiles (bottom). For each case, the vertical axis of the lower panel displays **-1** times the correlation of each sample to

30 the median healthy cfDNA fragmentation profile. Error bars depict confidence intervals from binomial tests for mutant allele fractions and confidence intervals calculated using

Fisher transformation for genome-wide fragmentation profiles. Although the approaches analyze different aspects of cfDNA (whole genome compared to specific alterations) the targeted sequencing and fragmentation profiles were similar for patients responding to therapy as well as those with stable or progressive disease. As fragmentation profiles reflect

⁵both genomic and epigenomic alterations, while mutant allele fractions only reflect individual mutations, mutant allele fractions alone may not reflect the absolute level of correlation of fragmentation profiles to healthy individuals.

Figures **11A - 1IC.** cfDNA fragmentation profiles in healthy individuals and patients with cancer. A, Fragmentation profiles (bottom) in the context of tumor copy number **¹⁰**changes (top) in a colorectal cancer patient where parallel analyses of tumor tissue were performed. The distribution of segment means and integer copy numbers are shown at top right in the indicated colors. Altered fragmentation profiles were present in regions of the genome that were copy neutral and were further affected in regions with copy number changes. B, **GC** adjusted fragmentation profiles from 1-2x whole genome sequencing for **15** healthy individuals and patients with cancer are depicted per cancer type using **5 Mb**

windows. The median healthy profile is indicated in black and the **98%** confidence band is shown in gray. For patients with cancer, individual profiles are colored based on their correlation to the healthy median. **C,** Windows are indicated in orange if more than **10%** of the cancer samples had a fragment ratio more than three standard deviations from the median **20** healthy fragment ratio. These analyses highlight the multitude of position dependent alterations across the genome in cfDNA of individuals with cancer.

Figures **12A** and 12B. Profiles of cfDNA fragment lengths in copy neutral regions in healthy individuals and one patient with colorectal cancer. **A,** The fragmentation profile in 211 copy neutral windows in chromosomes **1-6** for **25** randomly selected healthy individuals

25 (gray). For a patient with colorectal cancer **(CGCRC291)** with an estimated mutant allele fraction of 20%, the cancer fragment length profile was diluted to an approximate **10%** tumor contribution (blue). **A** and B, While the marginal densities of the fragment profiles for the healthy samples and cancer patient show substantial overlap **(A,** right), the fragmentation profiles are different as can be seen visualization of the fragmentation profiles **(A,** left) and

30 by the separation of the colorectal cancer patient from the healthy samples in a principal component analysis (B).

Figures **13A** and 13B. Genome-wide **GC** correction of cfDNA fragments. To estimate and control for the effects of **GC** content on sequencing coverage, coverage in non overlapping **100kb** genomic windows was calculated across the autosomes. For each window, the average **GC** of the aligned fragments was calculated. **A,** Loess smoothing of

⁵raw coverage (top row) for two randomly selected healthy subjects **(CGPLH189** and **CGPLH380)** and two cancer patients **(CGPLLU161** and CGPLBR24) with undetectable aneuploidy (PA score \leq 2.35). After subtracting the average coverage predicted by the loess model, the residuals were rescaled to the median autosomal coverage (bottom row). As fragment length may also result in coverage biases, this **GC** correction procedure was

10 performed separately for short $(\leq 150 \text{ bp})$ and $\log (\geq 151 \text{ bp})$ fragments. While the 100 kb bins on chromosome **19** (blue points) consistently have less coverage than predicted **by** the loess model, we did not implement a chromosome-specific correction as such an approach would remove the effects of chromosomal copy number on coverage. B, Overall, a limited correlation was found between short or long fragment coverage and **GC** content after **15** correction among healthy subjects and cancer patients with a PA score **<3.**

Figure 14. Schematic of machine learning model. Gradient tree boosting machine learning was used to examine whether cfDNA can be categorized as having characteristics of a cancer patient or healthy individual. The machine learning model included fragmentation size and coverage characteristics in windows throughout the genome, as well as

- **20** chromosomal arm and mitochondrial **DNA** copy numbers. **A** 10-fold cross validation approach was employed in which each sample is randomly assigned to a fold and **9** of the folds **(90%** of the data) are used for training and one fold **(10%** of the data) is used for testing. The prediction accuracy from a single cross validation is an average over the **10** possible combinations of test and training sets. As this prediction accuracy can reflect bias
- **25** from the initial randomization of patients, the entire procedure was repeat, including the randomization of patients to folds, **10** times. For all cases, feature selection and model estimation were performed on training data and were validated on test data and the test data were never used for feature selection. Ultimately, a **DELFI** score was obtained that could be used to classify individuals as likely healthy or having cancer.

Figure **15.** Distribution of AUCs across the repeated 10-fold cross-validation. The **² ⁵ ', 50 ',** and **⁷ ⁵ '** percentiles of the **100** AUCs for the cohort of **215** healthy individuals and **208** patients with cancer are indicated **by** dashed lines.

Figures **16A** and 16B. Whole-genome analyses of chromosomal arm copy number **⁵**changes and mitochondrial genome representation. **A,** Z scores for each autosome arm are depicted for healthy individuals ($n=215$) and patients with cancer ($n=208$). The vertical axis depicts normal copy at zero with positive and negative values indicating arm gains and losses, respectively. Z scores greater than **50** or less than **-50** are thresholded at the indicated values. B, The fraction of reads mapping to the mitochondrial genome is depicted for **¹⁰**healthy individuals and patients with cancer.

Figures **17A** and **17B.** Detection of cancer using **DELFI. A,** Receiver operator characteristics for detection of cancer using cfDNA fragmentation profiles and other genome-wide features in a machine learning approach are depicted for a cohort of **215** healthy individuals and 208 patients with cancer **(DELFI, AUC** = 0.94), with \geq 95% **¹⁵**specificity shaded in blue. Machine learning analyses of chromosomal arm copy number (Chr copy number (ML)), and mitochondrial genome copy number (mtDNA), are shown in the indicated colors. B, Analyses of individual cancers types using the DELFI-combined approach had AUCs ranging from **0.86** to **>0.99.**

Figure **18. DELFI** detection of cancer **by** stage. Receiver operator characteristics for **²⁰**detection of cancer using cfDNA fragmentation profiles and other genome-wide features in a machine learning approach are depicted for a cohort of **215** healthy individuals and each stage of 208 patients with cancer with \geq 95% specificity shaded in blue.

Figure **19. DELFI** tissue of origin prediction. Receiver operator characteristics for **DELFI** tissue prediction of bile duct, breast, colorectal, gastric, lung, ovarian, and pancreatic **²⁵**cancers are depicted. In order to increase sample sizes within cancer type classes, cases detected with a **90%** specificity were included, and the lung cancer cohort was supplemented with the addition of baseline cfDNA data from **18** lung cancer patients with prior treatment (see, e.g., Shen *et al., 2018 Nature,* **563:579-583).**

Figure 20. Detection of cancer using **DELFI** and mutation-based cfDNA approaches. **30 DELFI** (green) and targeted sequencing for mutation identification (blue) were performed independently in a cohort of **126** patients with breast, bile duct, colorectal, gastric, lung, or

ovarian cancers. The number of individuals detected **by** each approach and in combination are indicated for **DELFI** detection with a specificity of **98%,** targeted sequencing specificity at **>99%,** and a combined specificity of **98%. ND** indicates not detected.

DETAILED DESCRIPTION

- **⁵**This document provides methods and materials for determining a cfDNA fragmentation profile in a mammal (e.g., in a sample obtained from a mammal). As used herein, the terms "fragmentation profile," "position dependent differences in fragmentation patterns," and "differences in fragment size and coverage in a position dependent manner across the genome" are equivalent and can be used interchangeably. In some cases,
- **10** determining a cfDNA fragmentation profile in a mammal can be used for identifying a mammal as having cancer. For example, cfDNA fragments obtained from a mammal (e.g., from a sample obtained from a mammal) can be subjected to low coverage whole-genome sequencing, and the sequenced fragments can be mapped to the genome (e.g., in non overlapping windows) and assessed to determine a cfDNA fragmentation profile. As
- **¹⁵**described herein, a cfDNA fragmentation profile of a mammal having cancer is more heterogeneous (e.g., in fragment lengths) than a cfDNA fragmentation profile of a healthy mammal (e.g., a mammal not having cancer). As such, this document also provides methods and materials for assessing, monitoring, and/or treating mammals (e.g., humans) having, or suspected of having, cancer. In some cases, this document provides methods and materials
- **20** for identifying a mammal as having cancer. For example, a sample (e.g., a blood sample) obtained from a mammal can be assessed to determine the presence and, optionally, the tissue of origin of the cancer in the mammal based, at least in part, on the cfDNA fragmentation profile of the mammal. In some cases, this document provides methods and materials for monitoring a mammal as having cancer. For example, a sample (e.g., a blood sample)
- **25** obtained from a mammal can be assessed to determine the presence of the cancer in the mammal based, at least in part, on the cfDNA fragmentation profile of the mammal. In some cases, this document provides methods and materials for identifying a mammal as having cancer, and administering one or more cancer treatments to the mammal to treat the mammal. For example, a sample (e.g., a blood sample) obtained from a mammal can be assessed to

30 determine if the mammal has cancer based, at least in part, on the cfDNA fragmentation

profile of the mammal, and one or more cancer treatments can be administered to the mammal.

A cfDNA fragmentation profile can include one or more cfDNA fragmentation patterns. **A** cfDNA fragmentation pattern can include any appropriate cfDNA fragmentation **⁵**pattern. Examples of cfDNA fragmentation patterns include, without limitation, median fragment size, fragment size distribution, ratio of small cfDNA fragments to large cfDNA fragments, and the coverage of cfDNA fragments. In some cases, a cfDNA fragmentation pattern includes two or more (e.g., two, three, or four) of median fragment size, fragment size distribution, ratio of small cfDNA fragments to large cfDNA fragments, and the coverage of

- **¹⁰**cfDNA fragments. In some cases, cfDNA fragmentation profile can be a genome-wide cfDNA profile (e.g., a genome-wide cfDNA profile in windows across the genome). In some cases, cfDNA fragmentation profile can be a targeted region profile. **A** targeted region can be any appropriate portion of the genome (e.g., a chromosomal region). Examples of chromosomal regions for which a cfDNA fragmentation profile can be determined as
- **¹⁵**described herein include, without limitation, a portion of a chromosome (e.g., a portion of **2q,** 4p, **5p, 6q, 7p, 8q, 9q, l0q, 11q, 12q,** and/or 14q) and a chromosomal arm (e.g., a chromosomal arm of **8q,13q, 11q,** and/or **3p).** In some cases, a cfDNA fragmentation profile can include two or more targeted region profiles.
- In some cases, a cfDNA fragmentation profile can be used to identify changes (e.g., **20** alterations) in cfDNA fragment lengths. An alteration can be a genome-wide alteration or an alteration in one or more targeted regions/loci. **A** target region can be any region containing one or more cancer-specific alterations. Examples of cancer-specific alterations, and their chromosomal locations, include, without limitation, those shown in Table **3** (Appendix C) and those shown in Table **6** (Appendix F). In some cases, a cfDNA fragmentation profile can
- **25** be used to identify (e.g., simultaneously identify) from about **10** alterations to about **500** alterations (e.g., from about **25** to about **500,** from about **50** to about **500,** from about **100** to about **500,** from about 200 to about **500,** from about **300** to about **500,** from about **10** to about 400, from about **10** to about **300,** from about **10** to about 200, from about **10** to about **100,** from about **10** to about **50,** from about 20 to about 400, from about **30** to about **300,** from
- **30** about 40 to about 200, from about **50** to about **100,** from about 20 to about **100,** from about **25** to about **75,** from about **50** to about **250,** or from about **100** to about 200, alterations).

In some cases, a cfDNA fragmentation profile can be used to detect tumor-derived **DNA.** For example, a cfDNA fragmentation profile can be used to detect tumor-derived **DNA by** comparing a cfDNA fragmentation profile of a mammal having, or suspected of having, cancer to a reference cfDNA fragmentation profile (e.g., a cfDNA fragmentation

- **⁵**profile of a healthy mammal and/or a nucleosomal **DNA** fragmentation profile of healthy cells from the mammal having, or suspected of having, cancer). In some cases, a reference cfDNA fragmentation profile is a previously generated profile from a healthy mammal. For example, methods provided herein can be used to determine a reference cfDNA fragmentation profile in a healthy mammal, and that reference cfDNA fragmentation profile
- **10** can be stored (e.g., in a computer or other electronic storage medium) for future comparison to a test cfDNA fragmentation profile in mammal having, or suspected of having, cancer. In some cases, a reference cfDNA fragmentation profile (e.g., a stored cfDNA fragmentation profile) of a healthy mammal is determined over the whole genome. In some cases, a reference cfDNA fragmentation profile (e.g., a stored cfDNA fragmentation profile) of a **15** healthy mammal is determined over a subgenomic interval.

In some cases, a cfDNA fragmentation profile can be used to identify a mammal (e.g., a human) as having cancer (e.g., a colorectal cancer, a lung cancer, a breast cancer, a gastric cancer, a pancreatic cancer, a bile duct cancer, and/or an ovarian cancer).

- **A** cfDNA fragmentation profile can include a cfDNA fragment size pattern. cfDNA **20** fragments can be any appropriate size. For example, cfDNA fragment can be from about **50** base pairs **(bp)** to about 400 **bp** in length. As described herein, a mammal having cancer can have a cfDNA fragment size pattern that contains a shorter median cfDNA fragment size than the median cfDNA fragment size in a healthy mammal. **A** healthy mammal (e.g., a mammal not having cancer) can have cfDNA fragment sizes having a median cfDNA fragment size
- **25** from about **166.6 bp** to about **167.2 bp** (e.g., about **166.9 bp).** In some cases, a mammal having cancer can have cfDNA fragment sizes that are, on average, about **1.28 bp** to about 2.49 **bp** (e.g., about **1.88 bp)** shorter than cfDNA fragment sizes in a healthy mammal. For example, a mammal having cancer can have cfDNA fragment sizes having a median cfDNA fragment size of about 164.11 **bp** to about **165.92 bp** (e.g., about **165.02 bp).**

30 A cfDNA fragmentation profile can include a cfDNA fragment size distribution. As described herein, a mammal having cancer can have a cfDNA size distribution that is more

variable than a cfDNA fragment size distribution in a healthy mammal. In some case, a size distribution can be within a targeted region. **A** healthy mammal (e.g., a mammal not having cancer) can have a targeted region cfDNA fragment size distribution of about 1 or less than about **1.** In some cases, a mammal having cancer can have a targeted region cfDNA

- **⁵**fragment size distribution that is longer (e.g., **10, 15,** 20, **25, 30, 35,** 40, 45, **50** or more **bp** longer, or any number of base pairs between these numbers) than a targeted region cfDNA fragment size distribution in a healthy mammal. In some cases, a mammal having cancer can have a targeted region cfDNA fragment size distribution that is shorter (e.g., **10, 15,** 20, **25, 30, 35,** 40, 45, **50** or more **bp** shorter, or any number of base pairs between these numbers)
- 10 than a targeted region cfDNA fragment size distribution in a healthy mammal. In some cases, a mammal having cancer can have a targeted region cfDNA fragment size distribution that is about 47 **bp** smaller to about **30 bp** longer than a targeted region cfDNA fragment size distribution in a healthy mammal. In some cases, a mammal having cancer can have a targeted region cfDNA fragment size distribution of, on average, a **10, 11,** 12, **13,** 14, **15, 15,**
- **15 17, 18, 19,** 20 or more **bp** difference in lengths of cfDNA fragments. For example, a mammal having cancer can have a targeted region cfDNA fragment size distribution of, on average, about a **13 bp** difference in lengths of cfDNA fragments. In some case, a size distribution can be a genome-wide size distribution. **A** healthy mammal (e.g., a mammal not having cancer) can have very similar distributions of short and long cfDNA fragments
- **20** genome-wide. In some cases, a mammal having cancer can have, genome-wide, one or more alterations (e.g., increases and decreases) in cfDNA fragment sizes. The one or more alterations can be any appropriate chromosomal region of the genome. For example, an alteration can be in a portion of a chromosome. Examples of portions of chromosomes that can contain one or more alterations in cfDNA fragment sizes include, without limitation,
-

²⁵portions of **2q,** 4p, **5p, 6q, 7p, 8q, 9q, l0q, 11q, 12q,** and 14q. For example, an alteration can be across a chromosome arm (e.g., an entire chromosome arm).

A cfDNA fragmentation profile can include a ratio of small cfDNA fragments to large cfDNA fragments and a correlation of fragment ratios to reference fragment ratios. As used herein, with respect to ratios of small cfDNA fragments to large cfDNA fragments, a small **30** cfDNA fragment can be from about **100 bp** in length to about **150 bp** in length. As used

herein, with respect to ratios of small cfDNA fragments to large cfDNA fragments, a large

cfDNA fragment can be from about **151 bp** in length to 220 **bp** in length. As described herein, a mammal having cancer can have a correlation of fragment ratios (e.g., a correlation of cfDNA fragment ratios to reference **DNA** fragment ratios such as **DNA** fragment ratios from one or more healthy mammals) that is lower (e.g., 2-fold lower, 3-fold lower, 4-fold

⁵lower, 5-fold lower, 6-fold lower, 7-fold lower, 8-fold lower, 9-fold lower, 10-fold lower, or more) than in a healthy mammal. **A** healthy mammal (e.g., a mammal not having cancer) can have a correlation of fragment ratios (e.g., a correlation of cfDNA fragment ratios to reference **DNA** fragment ratios such as **DNA** fragment ratios from one or more healthy mammals) of about 1 (e.g., about **0.96).** In some cases, a mammal having cancer can have a 10 correlation of fragment ratios (e.g., a correlation of cfDNA fragment ratios to reference DNA fragment ratios such as **DNA** fragment ratios from one or more healthy mammals) that is, on average, about **0.19** to about **0.30** (e.g., about **0.25)** lower than a correlation of fragment ratios (e.g., a correlation of cfDNA fragment ratios to reference **DNA** fragment ratios such as **DNA** fragment ratios from one or more healthy mammals) in a healthy mammal.

15 A cfDNA fragmentation profile can include coverage of all fragments. Coverage of all fragments can include windows (e.g., non-overlapping windows) of coverage. In some cases, coverage of all fragments can include windows of small fragments (e.g., fragments from about **100 bp** to about **150 bp** in length). In some cases, coverage of all fragments can include windows of large fragments (e.g., fragments from about **151 bp** to about 220 **bp** in

20 length).

In some cases, a cfDNA fragmentation profile can be used to identify the tissue of origin of a cancer (e.g., a colorectal cancer, a lung cancer, a breast cancer, a gastric cancer, a pancreatic cancer, a bile duct cancer, or an ovarian cancer). For example, a cfDNA fragmentation profile can be used to identify a localized cancer. When a cfDNA

25 fragmentation profile includes a targeted region profile, one or more alterations described herein (e.g., in Table **3** (Appendix C) and/or in Table **6** (Appendix F)) can be used to identify the tissue of origin of a cancer. In some cases, one or more alterations in chromosomal regions can be used to identify the tissue of origin of a cancer.

A cfDNA fragmentation profile can be obtained using any appropriate method. In **30** some cases, cfDNA from a mammal (e.g., a mammal having, or suspected of having, cancer) can be processed into sequencing libraries which can be subjected to whole genome

sequencing (e.g., low-coverage whole genome sequencing), mapped to the genome, and analyzed to determine cfDNA fragment lengths. Mapped sequences can be analyzed in non overlapping windows covering the genome. Windows can be any appropriate size. For example, windows can be from thousands to millions of bases in length. As one non-limiting

⁵example, a window can be about **5** megabases **(Mb)** long. Any appropriate number of windows can be mapped. For example, tens to thousands of windows can be mapped in the genome. For example, hundreds to thousands of windows can be mapped in the genome. **A** cfDNA fragmentation profile can be determined within each window. In some cases, a cfDNA fragmentation profile can be obtained as described in Example **1.** In some cases, a **10** cfDNA fragmentation profile can be obtained as shown in Figure **1.**

In some cases, methods and materials described herein also can include machine learning. For example, machine learning can be used for identifying an altered fragmentation profile (e.g., using coverage of cfDNA fragments, fragment size of cfDNA fragments, coverage of chromosomes, and mtDNA).

15 In some cases, methods and materials described herein can be the sole method used to identify a mammal (e.g., a human) as having cancer (e.g., a colorectal cancer, a lung cancer, a breast cancer, a gastric cancer, a pancreatic cancer, a bile duct cancer, and/or an ovarian cancer). For example, determining a cfDNA fragmentation profile can be the sole method used to identify a mammal as having cancer.

20 In some cases, methods and materials described herein can be used together with one or more additional methods used to identify a mammal (e.g., a human) as having cancer (e.g., a colorectal cancer, a lung cancer, a breast cancer, a gastric cancer, a pancreatic cancer, a bile duct cancer, and/or an ovarian cancer). Examples of methods used to identify a mammal as having cancer include, without limitation, identifying one or more cancer-specific sequence

25 alterations, identifying one or more chromosomal alterations (e.g., aneuploidies and rearrangements), and identifying other cfDNA alterations. For example, determining a cfDNA fragmentation profile can be used together with identifying one or more cancer specific mutations in a mammal's genome to identify a mammal as having cancer. For example, determining a cfDNA fragmentation profile can be used together with identifying

30 one or more aneuploidies in a mammal's genome to identify a mammal as having cancer.

In some aspects, this document also provides methods and materials for assessing, monitoring, and/or treating mammals (e.g., humans) having, or suspected of having, cancer. In some cases, this document provides methods and materials for identifying a mammal as having cancer. For example, a sample (e.g., a blood sample) obtained from a mammal can be **⁵**assessed to determine if the mammal has cancer based, at least in part, on the cfDNA fragmentation profile of the mammal. In some cases, this document provides methods and materials for identifying the location (e.g., the anatomic site or tissue of origin) of a cancer in a mammal. For example, a sample (e.g., a blood sample) obtained from a mammal can be assessed to determine the tissue of origin of the cancer in the mammal based, at least in part, 10 on the cfDNA fragmentation profile of the mammal. In some cases, this document provides methods and materials for identifying a mammal as having cancer, and administering one or more cancer treatments to the mammal to treat the mammal. For example, a sample (e.g., a blood sample) obtained from a mammal can be assessed to determine if the mammal has cancer based, at least in part, on the cfDNA fragmentation profile of the mammal, and

15 administering one or more cancer treatments to the mammal. In some cases, this document provides methods and materials for treating a mammal having cancer. For example, one or more cancer treatments can be administered to a mammal identified as having cancer (e.g., based, at least in part, on the cfDNA fragmentation profile of the mammal) to treat the mammal. In some cases, during or after the course of a cancer treatment (e.g., any of the **20** cancer treatments described herein), a mammal can undergo monitoring (or be selected for

increased monitoring) and/or further diagnostic testing. In some cases, monitoring can include assessing mammals having, or suspected of having, cancer **by,** for example, assessing a sample (e.g., a blood sample) obtained from the mammal to determine the cfDNA fragmentation profile of the mammal as described herein, and changes in the cfDNA

25 fragmentation profiles over time can be used to identify response to treatment and/or identify the mammal as having cancer (e.g., a residual cancer).

Any appropriate mammal can be assessed, monitored, and/or treated as described herein. **A** mammal can be a mammal having cancer. **A** mammal can be a mammal suspected of having cancer. Examples of mammals that can be assessed, monitored, and/or treated as **³⁰**described herein include, without limitation, humans, primates such as monkeys, dogs, cats, horses, cows, pigs, sheep, mice, and rats. For example, a human having, or suspected of

having, cancer can be assessed to determine a cfDNA fragmentation profiled as described herein and, optionally, can be treated with one or more cancer treatments as described herein.

Any appropriate sample from a mammal can be assessed as described herein (e.g., assessed for a **DNA** fragmentation pattern). In some cases, a sample can include **DNA** (e.g., **⁵**genomic **DNA).** In some cases, a sample can include cfDNA (e.g., circulating tumor **DNA** (ctDNA)). In some cases, a sample can be fluid sample (e.g., a liquid biopsy). Examples of samples that can contain **DNA** and/or polypeptides include, without limitation, blood (e.g., whole blood, serum, or plasma), amnion, tissue, urine, cerebrospinal fluid, saliva, sputum, broncho-alveolar lavage, bile, lymphatic fluid, cyst fluid, stool, ascites, pap smears, breast **¹⁰**milk, and exhaled breath condensate. For example, a plasma sample can be assessed to determine a cfDNA fragmentation profiled as described herein.

A sample from a mammal to be assessed as described herein (e.g., assessed for a **DNA** fragmentation pattern) can include any appropriate amount of cfDNA. In some cases, a sample can include a limited amount of **DNA.** For example, a cfDNA fragmentation profile **¹⁵**can be obtained from a sample that includes less **DNA** than is typically required for other cfDNA analysis methods, such as those described in, for example, Phallen et al., **2017** Sci *TranslMed* **9;** Cohen et al., **2018** *Science* **359:926;** Newman et al., 2014 *Nat Med 20:548;* and Newman et al., **2016** *Nat Biotechnol* 34:547).

In some cases, a sample can be processed (e.g., to isolate and/or purify **DNA** and/or **20** polypeptides from the sample). For example, **DNA** isolation and/or purification can include cell lysis (e.g., using detergents and/or surfactants), protein removal (e.g., using a protease), and/or RNA removal (e.g., using an RNase). As another example, polypeptide isolation and/or purification can include cell lysis (e.g., using detergents and/or surfactants), **DNA** removal (e.g., using a DNase), and/or RNA removal (e.g., using an RNase).

25 A mammal having, or suspected of having, any appropriate type of cancer can be assessed (e.g., to determine a cfDNA fragmentation profile) and/or treated (e.g., **by** administering one or more cancer treatments to the mammal) using the methods and materials described herein. **A** cancer can be any stage cancer. In some cases, a cancer can be an early stage cancer. In some cases, a cancer can be an asymptomatic cancer. In some **30** cases, a cancer can be a residual disease and/or a recurrence (e.g., after surgical resection

and/or after cancer therapy). **A** cancer can be any type of cancer. Examples of types of

cancers that can be assessed, monitored, and/or treated as described herein include, without limitation, colorectal cancers, lung cancers, breast cancers, gastric cancers, pancreatic cancers, bile duct cancers, and ovarian cancers.

When treating a mammal having, or suspected of having, cancer as described herein, **⁵**the mammal can be administered one or more cancer treatments. **A** cancer treatment can be any appropriate cancer treatment. One or more cancer treatments described herein can be administered to a mammal at any appropriate frequency (e.g., once or multiple times over a period of time ranging from days to weeks). Examples of cancer treatments include, without limitation adjuvant chemotherapy, neoadjuvant chemotherapy, radiation therapy, hormone

- 10 therapy, cytotoxic therapy, immunotherapy, adoptive T cell therapy (e.g., chimeric antigen receptors and/or T cells having wild-type or modified T cell receptors), targeted therapy such as administration of kinase inhibitors (e.g., kinase inhibitors that target a particular genetic lesion, such as a translocation or mutation), (e.g. a kinase inhibitor, an antibody, a bispecific antibody), signal transduction inhibitors, bispecific antibodies or antibody fragments (e.g.,
- **15** BiTEs), monoclonal antibodies, immune checkpoint inhibitors, surgery (e.g., surgical resection), or any combination of the above. In some cases, a cancer treatment can reduce the severity of the cancer, reduce a symptom of the cancer, and/or to reduce the number of cancer cells present within the mammal.
- In some cases, a cancer treatment can include an immune checkpoint inhibitor. Non **²⁰**limiting examples of immune checkpoint inhibitors include nivolumab (Opdivo), pembrolizumab (Keytruda), atezolizumab (tecentriq), avelumab (bavencio), durvalumab (imfinzi), ipilimumab (yervoy). See, e.g., Pardoll (2012) Nat. Rev Cancer 12: 252-264; Sun et al. **(2017)** Eur Rev Med Pharmacol Sci **21(6): 1198-1205;** Hamanishi et al. **(2015) J.** Clin. Oncol. 33(34): 4015-22; Brahmer et al. (2012) **N** Engl **J** Med **366(26): 2455-65;** Ricciuti et
- **25** al. **(2017) J.** Thorac Oncol. **12(5):** e5I-e55; Ellis et al. **(2017)** Clin Lung Cancer pii: **S1525 7304(17)30043-8;** Zou and Awad **(2017)** Ann Oncol 28(4): **685-687;** Sorscher **(2017) N** Engl **J** Med **376(10: 996-7;** Hui et al. **(2017)** Ann Oncol 28(4): **874-881;** Vansteenkiste et al. **(2017)** Expert Opin Biol Ther **17(6): 781-789;** Hellmann et al. **(2017)** Lancet Oncol. **18(1):** 31-41; Chen **(2017) J.** Chin Med Assoc **80(1):** 7-14.

30 In some cases, a cancer treatment can be an adoptive T cell therapy (e.g., chimeric antigen receptors and/or T cells having wild-type or modified T cell receptors). See, e.g.,

Rosenberg and Restifo **(2015)** Science **348(6230): 62-68;** Chang and Chen **(2017)** Trends Mol Med **23(5):** 430-450; Yee and Lizee **(2016)** Cancer **J. 23(2):** 144-148; Chen et al. **(2016)** Oncoimmunology **6(2):** e1273302; **US** 2016/0194404; **US 2014/0050788; US 2014/0271635; US 9,233,125;** incorporated **by** reference in their entirety herein.

⁵In some cases, a cancer treatment can be a chemotherapeutic agent. Non-limiting examples of chemotherapeutic agents include: amsacrine, azacitidine, axathioprine, bevacizumab (or an antigen-binding fragment thereof), bleomycin, busulfan, carboplatin capecitabine, chlorambucil, cisplatin, cyclophosphamide, cytarabine, dacarbazine, daunorubicin, docetaxel, doxifluridine, doxorubicin, epirubicin, erlotinib hydrochlorides,

- **¹⁰**etoposide, fiudarabine, floxuridine, fludarabine, fluorouracil, gemcitabine, hydroxyurea, idarubicin, ifosfamide, irinotecan, lomustine, mechlorethamine, melphalan, mercaptopurine, methotrxate, mitomycin, mitoxantrone, oxaliplatin, paclitaxel, pemetrexed, procarbazine, all trans retinoic acid, streptozocin, tafluposide, temozolomide, teniposide, tioguanine, topotecan, uramustine, valrubicin, vinblastine, vincristine, vindesine, vinorelbine, and
- 15 combinations thereof. Additional examples of anti-cancer therapies are known in the art; see, e.g. the guidelines for therapy from the American Society of Clinical Oncology **(ASCO),** European Society for Medical Oncology **(ESMO),** or National Comprehensive Cancer Network **(NCCN).**
- When monitoring a mammal having, or suspected of having, cancer as described **20** herein (e.g., based, at least in part, on the cfDNA fragmentation profile of the mammal), the monitoring can be before, during, and/or after the course of a cancer treatment. Methods of monitoring provided herein can be used to determine the efficacy of one or more cancer treatments and/or to select a mammal for increased monitoring. In some cases, the monitoring can include identifying a cfDNA fragmentation profile as described herein. For
- **25** example, a cfDNA fragmentation profile can be obtained before administering one or more cancer treatments to a mammal having, or suspected or having, cancer, one or more cancer treatments can be administered to the mammal, and one or more cfDNA fragmentation profiles can be obtained during the course of the cancer treatment. In some cases, a cfDNA fragmentation profile can change during the course of cancer treatment (e.g., any of the
- **³⁰**cancer treatments described herein). For example, a cfDNA fragmentation profile indicative that the mammal has cancer can change to a cfDNA fragmentation profile indicative that the

mammal does not have cancer. Such a cfDNA fragmentation profile change can indicate that the cancer treatment is working. Conversely, a cfDNA fragmentation profile can remain static (e.g., the same or approximately the same) during the course of cancer treatment (e.g., any of the cancer treatments described herein). Such a static cfDNA fragmentation profile **⁵**can indicate that the cancer treatment is not working. In some cases, the monitoring can include conventional techniques capable of monitoring one or more cancer treatments (e.g., the efficacy of one or more cancer treatments). In some cases, a mammal selected for increased monitoring can be administered a diagnostic test (e.g., any of the diagnostic tests disclosed herein) at an increased frequency compared to a mammal that has not been selected 10 for increased monitoring. For example, a mammal selected for increased monitoring can be administered a diagnostic test at a frequency of twice daily, daily, bi-weekly, weekly, bi monthly, monthly, quarterly, semi-annually, annually, or any at frequency therein. In some cases, a mammal selected for increased monitoring can be administered a one or more additional diagnostic tests compared to a mammal that has not been selected for increased

- **¹⁵**monitoring. For example, a mammal selected for increased monitoring can be administered two diagnostic tests, whereas a mammal that has not been selected for increased monitoring is administered only a single diagnostic test (or no diagnostic tests). In some cases, a mammal that has been selected for increased monitoring can also be selected for further diagnostic testing. Once the presence of a tumor or a cancer (e.g., a cancer cell) has been **20** identified (e.g., **by** any of the variety of methods disclosed herein), it may be beneficial for the mammal to undergo both increased monitoring (e.g., to assess the progression of the tumor or cancer in the mammal and/or to assess the development of one or more cancer
	- biomarkers such as mutations), and further diagnostic testing (e.g., to determine the size and/or exact location (e.g., tissue of origin) of the tumor or the cancer). In some cases, one
- **25** or more cancer treatments can be administered to the mammal that is selected for increased monitoring after a cancer biomarker is detected and/or after the cfDNA fragmentation profile of the mammal has not improved or deteriorated. Any of the cancer treatments disclosed herein or known in the art can be administered. For example, a mammal that has been selected for increased monitoring can be further monitored, and a cancer treatment can be
- **³⁰**administered if the presence of the cancer cell is maintained throughout the increased monitoring period. Additionally or alternatively, a mammal that has been selected for

increased monitoring can be administered a cancer treatment, and further monitored as the cancer treatment progresses. In some cases, after a mammal that has been selected for increased monitoring has been administered a cancer treatment, the increased monitoring will reveal one or more cancer biomarkers (e.g., mutations). In some cases, such one or more **⁵**cancer biomarkers will provide cause to administer a different cancer treatment (e.g., a resistance mutation may arise in a cancer cell during the cancer treatment, which cancer cell

harboring the resistance mutation is resistant to the original cancer treatment).

When a mammal is identified as having cancer as described herein (e.g., based, at least in part, on the cfDNA fragmentation profile of the mammal), the identifying can be **¹⁰**before and/or during the course of a cancer treatment. Methods of identifying a mammal as having cancer provided herein can be used as a first diagnosis to identify the mammal (e.g., as having cancer before any course of treatment) and/or to select the mammal for further diagnostic testing. In some cases, once a mammal has been determined to have cancer, the mammal may be administered further tests and/or selected for further diagnostic testing. In **15** some cases, methods provided herein can be used to select a mammal for further diagnostic testing at a time period prior to the time period when conventional techniques are capable of diagnosing the mammal with an early-stage cancer. For example, methods provided herein for selecting a mammal for further diagnostic testing can be used when a mammal has not been diagnosed with cancer **by** conventional methods and/or when a mammal is not known to **20** harbor a cancer. In some cases, a mammal selected for further diagnostic testing can be administered a diagnostic test (e.g., any of the diagnostic tests disclosed herein) at an increased frequency compared to a mammal that has not been selected for further diagnostic testing. For example, a mammal selected for further diagnostic testing can be administered a diagnostic test at a frequency of twice daily, daily, bi-weekly, weekly, bi-monthly, monthly,

- **25** quarterly, semi-annually, annually, or any at frequency therein. In some cases, a mammal selected for further diagnostic testing can be administered a one or more additional diagnostic tests compared to a mammal that has not been selected for further diagnostic testing. For example, a mammal selected for further diagnostic testing can be administered two diagnostic tests, whereas a mammal that has not been selected for further diagnostic
- **³⁰**testing is administered only a single diagnostic test (or no diagnostic tests). In some cases, the diagnostic testing method can determine the presence of the same type of cancer (e.g.,

having the same tissue or origin) as the cancer that was originally detected (e.g., based, at least in part, on the cfDNA fragmentation profile of the mammal). Additionally or alternatively, the diagnostic testing method can determine the presence of a different type of cancer as the cancer that was original detected. In some cases, the diagnostic testing method

- **⁵**is a scan. In some cases, the scan is a computed tomography **(CT),** a **CT** angiography **(CTA),** a esophagram (a Barium swallom), a Barium enema, a magnetic resonance imaging (MRI), a PET scan, an ultrasound (e.g., an endobronchial ultrasound, an endoscopic ultrasound), an X ray, a **DEXA** scan. In some cases, the diagnostic testing method is a physical examination, such as an anoscopy, a bronchoscopy (e.g., an autofluorescence bronchoscopy, a white-light
- 10 bronchoscopy, a navigational bronchoscopy), a colonoscopy, a digital breast tomosynthesis, an endoscopic retrograde cholangiopancreatography (ERCP), an ensophagogastroduodenoscopy, a mammography, a Pap smear, a pelvic exam, a positron emission tomography and computed tomography (PET-CT) scan. In some cases, a mammal that has been selected for further diagnostic testing can also be selected for increased
- **¹⁵**monitoring. Once the presence of a tumor or a cancer (e.g., a cancer cell) has been identified (e.g., **by** any of the variety of methods disclosed herein), it may be beneficial for the mammal to undergo both increased monitoring (e.g., to assess the progression of the tumor or cancer in the mammal and/or to assess the development of one or more cancer biomarkers such as mutations), and further diagnostic testing (e.g., to determine the size and/or exact location of **20** the tumor or the cancer). In some cases, a cancer treatment is administered to the mammal
- that is selected for further diagnostic testing after a cancer biomarker is detected and/or after the cfDNA fragmentation profile of the mammal has not improved or deteriorated. Any of the cancer treatments disclosed herein or known in the art can be administered. For example, a mammal that has been selected for further diagnostic testing can be administered a further
- **²⁵**diagnostic test, and a cancer treatment can be administered if the presence of the tumor or the cancer is confirmed. Additionally or alternatively, a mammal that has been selected for further diagnostic testing can be administered a cancer treatment, and can be further monitored as the cancer treatment progresses. In some cases, after a mammal that has been selected for further diagnostic testing has been administered a cancer treatment, the
- **³⁰**additional testing will reveal one or more cancer biomarkers (e.g., mutations). In some cases, such one or more cancer biomarkers (e.g., mutations) will provide cause to administer a

different cancer treatment (e.g., a resistance mutation may arise in a cancer cell during the cancer treatment, which cancer cell harboring the resistance mutation is resistant to the original cancer treatment).

The invention will be further described in the following examples, which do not limit 5 the scope of the invention described in the claims.

EXAMPLES

Example 1: Cell-free DNA fragmentation inpatients with cancer

Analyses of cell free **DNA** have largely focused on targeted sequencing of specific genes. Such studies permit detection of a small number of tumor-specific alterations in **10** patients with cancer and not all patients, especially those with early stage disease, have detectable changes. Whole genome sequencing of cell-free **DNA** can identify chromosomal abnormalities and rearrangements in cancer patients but detection of such alterations has been challenging in part due to the difficulty in distinguishing a small number of abnormal from normal chromosomal changes (Leary et al., 2010 *Sci TranslMed 2:20ra14;* and Leary et

¹⁵al., 2012 *Sci TranslMed* 4:162ra154). Other efforts have suggested nucleosome patterns and chromatin structure may be different between cancer and normal tissues, and that cfDNA in patients with cancer may result in abnormal cfDNA fragment size as well as position (Snyder et al., **2016** *Cell* **164:57;** Jahr et al., 2001 *Cancer Res* **61:1659;** Ivanov et al., **2015** *BMC Genomics* 16(Suppl **13):S1).** However, the amount of sequencing needed for nucleosome **20** footprint analyses of cfDNA is impractical for routine analyses.

The sensitivity of any cell-free **DNA** approach depends on the number of potential alterations examined as well as the technical and biological limitations of detecting such changes. As a typical blood sample contains -2000 genome equivalents of cfDNA per milliliter of plasma (Phallen et al., **2017** *Sci TranslMed* **9),** the theoretical limit of detection

25 of a single alteration can be no better than one in a few thousand mutant to wild-type molecules. An approach that detects a larger number of alterations in the same number of genome equivalents would be more sensitive for detecting cancer in the circulation. Monte Carlo simulations show that increasing the number of potential abnormalities detected from only a few to tens or hundreds can potentially improve the limit of detection **by** orders of

magnitude, similar to recent probability analyses of multiple methylation changes in cfDNA (Figure 2).

This study presents a novel method called **DELFI** for detection of cancer and further identification of tissue of origin using whole genome sequencing (Figure **1).** The approach **⁵**uses cfDNA fragmentation profiles and machine learning to distinguish patterns of healthy blood cell **DNA** from tumor-derived **DNA** and to identify the primary tumor tissue. **DELFI** was used for a retrospective analysis of cfDNA from 245 healthy individuals and **236** patients with breast, colorectal, lung, ovarian, pancreatic, gastric, or bile duct cancers, with most patients exhibiting localized disease. Assuming this approach had sensitivity ≥ 0.80 for **¹⁰**discriminating cancer patients from healthy individuals while maintaining a specificity of **0.95,** a study of at least 200 cancer patients would enable estimation of the true sensitivity with a margin of error of **0.06** at the desired specificity of **0.95** or greater.

Materials and Methods

Patient and sample characteristics

15 Plasma samples from healthy individuals and plasma and tissue samples from patients with breast, lung, ovarian, colorectal, bile duct, or gastric cancer were obtained from ILSBio/Bioreclamation, Aarhus University, Herlev Hospital of the University of Copenhagen, Hvidovre Hospital, the University Medical Center of the University of Utrecht, the Academic Medical Center of the University of Amsterdam, the Netherlands Cancer **20** Institute, and the University of California, San Diego. **All** samples were obtained under Institutional Review Board approved protocols with informed consent for research use at participating institutions. Plasma samples from healthy individuals were obtained at the time of routine screening, including for colonoscopies or Pap smears. Individuals were considered healthy if they had no previous history of cancer and negative screening results.

²⁵Plasma samples from individuals with breast, colorectal, gastric, lung, ovarian, pancreatic, and bile duct cancer were obtained at the time of diagnosis, prior to tumor resection or therapy. Nineteen lung cancer patients analyzed for change in cfDNA fragmentation profiles across multiple time points were undergoing treatment with anti EGFR or anti-ERBB2 therapy (see, e.g., Phallen *et al., 2019 Cancer Research* **15,** 1204

30 1213). Clinical data for all patients included in this study are listed in Table 1 (Appendix **A).**

Gender was confirmed through genomic analyses of X and Y chromosome representation. Pathologic staging of gastric cancer patients was performed after neoadjuvant therapy. Samples where the tumor stage was unknown were indicated as stage X or unknown.

Nucleosomal DNA purification

⁵Viably frozen lymphocytes were elutriated from leukocytes obtained from a healthy male **(C0618)** and female **(D0808-L)** (Advanced Biotechnologies Inc., Eldersburg, MD). Aliquots of 1 x **106** cells were used for nucleosomal **DNA** purification using EZ Nucleosomal **DNA Prep Kit (Zymo Research, Irvine, CA). Cells were initially treated with 100** μ **l of** Nuclei Prep Buffer and incubated on ice for **5** minutes. After centrifugation at **200g** for **5** 10 minutes, supernatant was discarded and pelleted nuclei were treated twice with 100 μ l of Atlantis Digestion Buffer or with 100 µl of micrococcal nuclease (MN) Digestion Buffer. Finally, cellular nucleic **DNA** was fragmented with **0.5U** of Atlantis dsDNase at 42°C for 20 minutes or **1.5U** of MNase at **37°C** for 20 minutes. Reactions were stopped using 5X MN Stop Buffer and **DNA** was purified using Zymo-SpinTM **JIC** Columns. Concentration and **¹⁵**quality of eluted cellular nucleic **DNA** were analyzed using the Bioanalyzer **2100** (Agilent

Technologies, Santa Clara, **CA).**

Sample preparation and sequencing of cfDNA

Whole blood was collected in **EDTA** tubes and processed immediately or within one day after storage at 4°C, or was collected in Streck tubes and processed within two days of **20** collection for three cancer patients who were part of the monitoring analysis. Plasma and cellular components were separated **by** centrifugation at **800g** for **10** min at 4°C. Plasma was centrifuged a second time at **18,000g** at room temperature to remove any remaining cellular debris and stored at -80° C until the time of DNA extraction. DNA was isolated from plasma using the Qiagen Circulating Nucleic Acids Kit (Qiagen GmbH) and eluted in LoBind tubes **25** (Eppendorf **AG).** Concentration and quality of cfDNA were assessed using the Bioanalyzer 2100 (Agilent Technologies).

NGS cfDNA libraries were prepared for whole genome sequencing and targeted sequencing using **5** to **250** ng of cfDNA as described elsewhere (see, e.g., Phallen *et al.,* **2017** *Sci Transl Med* 9:eaan2415). Briefly, genomic libraries were prepared using the NEBNext

30 DNA Library Prep Kit for Illumina [New England Biolabs **(NEB)]** with four main

modifications to the manufacturer's guidelines: (i) The library purification steps used the on bead AMPure XP approach to minimize sample loss during elution and tube transfer steps (see, e.g., Fisher *et al., 2011 Genome Biol* 12:R1); (ii) NEBNext End Repair, A-tailing, and adapter ligation enzyme and buffer volumes were adjusted as appropriate to accommodate

5 the on-bead AMPure XP purification strategy; (iii) a pool of eight unique Illumina dual index adapters with 8-base pair **(bp)** barcodes was used in the ligation reaction instead of the standard Illumina single or dual index adapters with **6-** or **8-bp** barcodes, respectively; and (iv) cfDNA libraries were amplified with Phusion Hot Start Polymerase.

Whole genome libraries were sequenced directly. For targeted libraries, capture was 10 performed using Agilent SureSelect reagents and a custom set of hybridization probes targeting **58** genes (see, e.g., Phallen *et al., 2017 Sci TranslMed* 9:eaan2415) per the manufacturer's guidelines. The captured library was amplified with Phusion Hot Start Polymerase **(NEB).** Concentration and quality of captured cfDNA libraries were assessed on the Bioanalyzer 2100 using theDNAl000 Kit (Agilent Technologies). Targeted libraries **15** were sequenced using **100-bp** paired-end runs on the Illumina HiSeq **2000/2500** (Illumina).

Analyses of targeted sequencing datafrom cfDNA

Analyses of targeted **NGS** data for cfDNA samples was performed as described elsewhere (see, e.g., Phallen *et al., 2017 Sci TranslMed* 9:eaan2415). Briefly, primary processing was completed using Illumina **CASAVA** (Consensus Assessment of Sequence **20** and Variation) software (version **1.8),** including demultiplexing and masking of dual-index

- adapter sequences. Sequence reads were aligned against the human reference genome (version **hg18** or **hg19)** using NovoAlign with additional realignment of select regions using the Needleman-Wunsch method (see, e.g., Jones *et al., 2015 Sci TranslMed* 7:283ra53). The positions of the sequence alterations have not been affected **by** the different genome builds.
- **25** Candidate mutations, consisting of point mutations, small insertions, and deletions, were identified using VariantDx (see, e.g., Jones *et al., 2015 Sci TranslMed* 7:283ra53) (Personal Genome Diagnostics, Baltimore, MD) across the targeted regions of interest.

To analyze the fragment lengths of cfDNA molecules, each read pair from a cfDNA molecule was required to have a Phred quality score \geq 30. All duplicate ctDNA fragments, **30** defined as having the same start, end, and index barcode were removed. For each mutation,

only fragments for which one or both of the read pairs contained the mutated (or wild-type) base at the given position were included. This analysis was done using the R packages Rsamtools and GenomicAlignments.

For each genomic locus where a somatic mutation was identified, the lengths of **⁵**fragments containing the mutant allele were compared to the lengths of fragments of the wild-type allele. **If** more than **100** mutant fragments were identified, Welch's two-sample **t** test was used to compare the mean fragment lengths. For loci with fewer than **100** mutant fragments, a bootstrap procedure was implemented. Specifically, replacement **N** fragments containing the wild-type allele, where **N** denotes the number of fragments with the mutation, **¹⁰**were sampled. For each bootstrap replicate of wild type fragments their median length was computed. The p-value was estimated as the fraction of bootstrap replicates with a median wild-type fragment length as or more extreme than the observed median mutant fragment length.

Analyses ofwhole genome sequencing datafrom cfDNA

¹⁵Primary processing of whole genome **NGS** data for cfDNA samples was performed using Illumina **CASAVA** (Consensus Assessment of Sequence and Variation) software (version **1.8.2),** including demultiplexing and masking of dual-index adapter sequences. Sequence reads were aligned against the human reference genome (version **hgl9)** using **ELAND.**

- **20** Read pairs with a **MAPQ** score below **30** for either read and PCR duplicates were removed. **hgl9** autosomes were tiled into **26,236** adjacent, non-overlapping **100 kb** bins. Regions of low mappability, indicated **by** the **10%** of bins with the lowest coverage, were removed (see, e.g., Fortin *et al., 2015 Genome Biol* **16:180),** as were reads falling in the Duke blacklisted regions (see, e.g.,
- **25** hgdownload.cse.ucsc.edu/goldenpath/hg19/encodeDCC/wgEncodeMapability/). Usingthis approach, **361 Mb (13%)** of the **hgl9** reference genome was excluded, including centromeric and telomeric regions. Short fragments were defined as having a length between **100** and **150 bp** and long fragments were defined has having a length between **151** and 220 **bp.**

To account for biases in coverage attributable to **GC** content of the genome, the **30** locally weighted smoother loess with span ³/₄ was applied to the scatterplot of average

fragment GC versus coverage calculated for each 100kb bin. This loess regression was performed separately for short and long fragments to account for possible differences in **GC** effects on coverage in plasma **by** fragment length (see, e.g., Benjamini *et al., 2012 Nucleic Acids Res* 40:e72). The predictions for short and long coverage explained **by GC** from the

5 loess model were subtracted, obtaining residuals for short and long that were uncorrelated with **GC.** The residuals were returned to the original scale **by** adding back the genome-wide median short and long estimates of coverage. This procedure was repeated for each sample to account for possible differences in **GC** effects on coverage between samples. To further reduce the feature space and noise, the total GC-adjusted coverage in **5 Mb** bins was

10 calculated.

To compare the variability of fragment lengths from healthy subjects to fragments in patients with cancer, the standard deviation of the short to long fragmentation profiles for each individual was calculated. The standard deviations in the two groups were compared **by** a Wilcoxon rank sum test.

¹⁵*Analyses of chromosome arm copy number changes*

To develop arm-level statistics for copy number changes, an approach for aneuploidy detection in plasma as described elsewhere (see, e.g., Leary *et al., 2012 Sci TranslMed* 4:162ra154) was adopted. This approach divides the genome into non-overlapping 50KB bins for which GC-corrected log2 read depth was obtained after correction by loess with span

20 3/4. This loess-based correction is comparable to the approach outlined above, but is evaluated on a log2 scale to increase robustness to outliers in the smaller bins and does not stratify **by** fragment length. To obtain an arm-specific Z-score for copy number changes, the mean GC-adjusted read depth for each arm (GR) was centered and scaled **by** the average and standard deviation, respectively, of GR scores obtained from an independent set of **50**

25 healthy samples.

Analyses of mitochondrial-aligned reads from cfDNA

Whole genome sequence reads that initially mapped to the mitochondrial genome were extracted from bam files and realigned to the **hg19** reference genome in end-to-end mode with Bowtie2 as described elsewhere (see, e.g., Langmead *et al., 2012 Nat Methods* **30 9:357-359).** The resulting aligned reads were filtered such that both mates aligned to the

mitochondrial genome with **MAPQ >= 30.** The number of fragments mapping to the mitochondrial genome was counted and converted to a percentage of the total number of fragments in the original bam files.

Prediction model for cancer classification

⁵To distinguish healthy from cancer patients using fragmentation profiles, a stochastic gradient boosting model was used (gbm; see, e.g., Friedman *et al., 2001 Ann Stat* **29:1189 1232;** and Friedman *et al., 2002 Comput Stat Data An* **38:367-378).** GC-corrected total and short fragment coverage for all 504 bins were centered and scaled for each sample to have mean **0** and unit standard deviation. Additional features included Z-scores for each of the **39 ¹⁰**autosomal arms and mitochondrial representation (log10-transformed proportion of reads mapped to the mitochondria). To estimate the prediction error of this approach, 10-fold cross-validation was used as described elsewhere (see, e.g., Efron et al., **1997 J** Am Stat Assoc **92, 548-560).** Feature selection, performed only on the training data in each cross validation run, removed bins that were **highly** correlated (correlation **> 0.9)** or had near zero

¹⁵variance. Stochastic gradient boosted machine learning was implemented using the R package gbm package with parameters n.trees=150, interaction.depth=3, shrinkage=0.1, and n.minobsinside=10. To average over the prediction error from the randomization of patients to folds, the 10-fold cross validation procedure was repeated **10** times. Confidence intervals for sensitivity fixed at **98%** and **95%** specificity were obtained from 2000 bootstrap

20 replicates.

Prediction model for tumor tissue of origin classification

For samples correctly classified as cancer patients at **90%** specificity (n **=** 174), a separate stochastic gradient boosting model was trained to classify the tissue of origin. To account for the small number of lung samples used for prediction, **18** cfDNA baseline

25 samples from late stage lung cancer patients were included from the monitoring analyses. Performance characteristics of the model were evaluated **by** 10-fold cross-validation repeated **10** times. This gbm model was trained using the same features as in the cancer classification model. As previously described, features that displayed correlation above **0.9** to each other or had near zero variance were removed within each training dataset during cross-validation.

The tissue class probabilities were averaged across the **10** replicates for each patient and the class with the highest probability was taken as the predicted tissue.

Analyses ofnucleosomal DNA from human lymphocytes and cfDNA

- From the nuclease treated lymphocytes, fragment sizes were analyzed in **5 Mb** bins as **⁵**described for whole genome cfDNA analyses. **A** genome-wide map of nucleosome positions was constructed from the nuclease treated lymphocyte cell-lines. This approach identified local biases in the coverage of circulating fragments, indicating a region protected from degradation. **A** "Window positioning score" (WPS) was used to score each base pair in the genome (see, e.g., Snyder et al., **2016** *Cell* **164:57).** Using a sliding window of **60bp**
- 10 centered around each base, the WPS was calculated as the number of fragments completely spanning the window minus the number of fragments with only one end in the window. Since fragments arising from nucleosomes have a median length of **167 bp,** a high WPS indicated a possible nucleosomic position. WPS scores were centered at zero using a running median and smoothed using a Kolmogorov-Zurbenko filter (see, e.g., Zurbenko, *The spectral*
- **¹⁵***analysis of time series.* North-Holland series in statistics and probability; Elsevier, New York, NY, **1986).** For spans of positive WPS between **50** and 450 **bp,** a nucleosome peak was defined as the set of base pairs with a WPS above the median in that window. The calculation of nucleosome positions for cfDNA from **30** healthy individuals with sequence coverage of 9x was determined in the same manner as for lymphocyte **DNA.** To ensure that
- **20** nucleosomes in healthy cfDNA were representative, a consensus track of nucleosomes was defined consisting only of nucleosomes identified in two or more individuals. Median distances between adjacent nucleosomes were calculated from the consensus track.

Monte Carlo simulation of detection sensitivity

A Monte Carlo simulation was used to estimate the probability of detecting a **25** molecule with a tumor-derived alteration. Briefly, 1 million molecules were generated from a multinomial distribution. For a simulation with m alterations, wild-type molecules were simulated with probability p and each of the *m* tumor alterations were simulated with probability $(1-p)/m$. Next, $g * m$ molecules were sampled randomly with replacement, where **g** denotes the number of genome equivalents in 1 ml of plasma. **If** a tumor alteration was **30** sampled s or more times, the sample was classified as cancer-derived. The simulation was

repeated **1000** times, estimating the probability that the *in silico* sample would be correctly classified as cancer by the mean of the cancer indicator. Setting $g = 2000$ and $s = 5$, the number of tumor alterations was varied **by** powers of 2 from **I** to **256** and the fraction of tumor-derived molecules from **0.0001%** to **1%.**

⁵*Statistical analyses*

All statistical analyses were performed using R version 3.4.3. The R packages caret (version **6.0-79)** and gbm (version 2.1-4) were used to implement the classification of healthy versus cancer and tissue of origin. Confidence intervals from the model output were obtained with the pROC (version **1.13)** R package (see, e.g., Robin *et al., 2011 BMC*

¹⁰*bioinformatics* **12:77).** Assuming the prevalence of undiagnosed cancer cases in this population is high (1 or 2 cases per **100** healthy), a genomic assay with a specificity of **0.95** and sensitivity of **0.8** would have useful operating characteristics (positive predictive value of **0.25** and negative predictive value near **1).** Power calculations suggest that an analysis of more than 200 cancer patients and an approximately equal number of healthy controls, enable

¹⁵an estimation of the sensitivity with a margin of error of **0.06** at the desired specificity of **0.95** or greater.

Data and Code Availability

Sequence data utilized in this study have been deposited at the European Genome phenome Archive under study accession nos. **EGASOOOO1003611** and **EGAS00001002577.** 20 Code for analyses is available at github.com/Cancer-Genomics/delfi scripts.

Results

DELFI allows simultaneous analysis of a large number of abnormalities in cfDNA through genome-wide analysis of fragmentation patterns. The method is based on low coverage whole genome sequencing and analysis of isolated cfDNA. Mapped sequences are **25** analyzed in non-overlapping windows covering the genome. Conceptually, windows may range in size from thousands to millions of bases, resulting in hundreds to thousands of windows in the genome. **5 Mb** windows were used for evaluating cfDNA fragmentation patterns as these would provide over 20,000 reads per window even at a limited amount of 1 2x genome coverage. Within each window, the coverage and size distribution of cfDNA

fragments was examined. This approach was used to evaluate the variation of genome-wide fragmentation profiles in healthy and cancer populations (Table **1;** Appendix **A).** The genome-wide pattern from an individual can be compared to reference populations to determine if the pattern is likely healthy or cancer-derived. As genome-wide profiles reveal **⁵**positional differences associated with specific tissues that may be missed in overall fragment size distributions, these patterns may also indicate the tissue source of cfDNA.

The fragmentation size of cfDNA was focused on as it was found that cancer-derived cfDNA molecules may be more variable in size than cfDNA derived from non-cancer cells. cfDNA fragments from targeted regions that were captured and sequenced at high coverage

10 (43,706 total coverage, 8,044 distinct coverage) from patients with breast, colorectal, lung or ovarian cancer (Table 1 (Appendix **A),** Table 2 (Appendix B), and Table **3** (Appendix C)) were initially examined. Analyses of loci containing **165** tumor-specific alterations from **81** patients (range of **1-7** alterations per patient) revealed an average absolute difference of **6.5 bp (95% CI, 5.4-7.6 bp)** between lengths of median mutant and wild-type cfDNA fragments

- **¹⁵**(Fig. **3,** Table **3** (Appendix **C)).** The median size of mutant cfDNA fragments ranged from **30** bases smaller at chromosome **3** position 41,266,124 to 47 bases larger at chromosome 11 position **108,117,753** than the wild-type sequences at these regions (Table **3;** Appendix C). **GC** content was similar for mutated and non-mutated fragments (Fig. 4a), and there was no correlation between **GC** content and fragment length (Fig. 4b). Similar analyses of 44
- **20** germline alterations from **38** patients identified median cfDNA size differences of less than 1 **bp** between fragment lengths of different alleles (Fig. **5,** Table **3** (Appendix **C)).** Additionally, 41 alterations related to clonal hematopoiesis were identified through a previous sequence comparison of **DNA** from plasma, buffy coat, and tumors of the same individuals. Unlike tumor-derived fragments, there were no significant differences between
- **25** fragments with hematopoietic alterations and wild type fragments (Fig. **6,** Table **3** (Appendix **C)).** Overall, cancer-derived cfDNA fragment lengths were significantly more variable compared to non-cancer cfDNA fragments at certain genomic regions **(p<0.001,** variance ratio test). It was hypothesized that these differences may be due to changes in higher-order chromatin structure as well as other genomic and epigenomic abnormalities in cancer and
- **30** that cfDNA fragmentation in a position-specific mannercould therefore serve as a unique biomarker for cancer detection.

As targeted sequencing only analyzes a limited number of loci, larger-scale genome wide analyses to detect additional abnormalities in cfDNA fragmentation were investigated. cfDNA was isolated from -4 ml of plasma from **8** lung cancer patients with stage I-III disease, as well as from **30** healthy individuals (Table 1 (Appendix **A),** Table 4 (Appendix

5 D), and Table **5** (Appendix **E)). A** high efficiency approach was used to convert cfDNA to next generation sequencing libraries and performed whole genome sequencing at $\sim 9x$ coverage (Table 4; Appendix **D).** Overall cfDNA fragment lengths of healthy individuals were larger, with a median fragment size of **167.3 bp,** while patients with cancer had median fragment sizes of **163.8 (p<0.01,** Welch's t-test) (Table **5;** Appendix **E).** To examine

¹⁰differences in fragment size and coverage in a position dependent manner across the genome, sequenced fragments were mapped to their genomic origin and fragment lengths were evaluated in 504 windows that were **5 Mb** in size, covering **-2.6 Gb** of the genome. For each window, the fraction of small cfDNA fragments **(100** to **150 bp** in length) to larger cfDNA fragments **(151** to 220 **bp)** as well as overall coverage were determined and used to obtain **15** genome-wide fragmentation profiles for each sample.

Healthy individuals had very similar fragmentation profiles throughout the genome (Fig. **7** and Fig. **8).** To examine the origins of fragmentation patterns normally observed in cfDNA, nuclei were isolated from elutriated lymphocytes of two healthy individuals and treated with **DNA** nucleases to obtain nucleosomal **DNA** fragments. Analyses of cfDNA **20** patterns in observed healthy individuals revealed a high correlation to lymphocyte

nucleosomal **DNA** fragmentation profiles (Fig. **7b** and **7d)** and nucleosome distances (Fig. 7c and **7f).** Median distances between nucleosomes in lymphocytes were correlated to open **(A)** and closed (B) compartments of lymphoblastoid cells as revealed using the Hi-C method (see, e.g., Lieberman-Aiden *et al., 2009 Science* **326:289-293;** and Fortin *et al.,* **2015**

²⁵*Genome Biol* **16:180)** for examining the three-dimensional architecture of genomes (Fig. 7c). These analyses suggest that the fragmentation patterns of normal cfDNA are the result of nucleosomal **DNA** patterns that largely reflect the chromatin structure of normal blood cells.

In contrast to healthy cfDNA, patients with cancer had multiple distinct genomic differences with increases and decreases in fragment sizes at different regions (Fig. 7a and

30 7b). Similar to our observations from targeted analyses, there was also greater variation in fragment lengths genome-wide for patients with cancer compared to healthy individuals.
To determine whether cfDNA fragment length patterns could be used to distinguish patients with cancer from healthy individuals, genome-wide correlation analyses were performed of the fraction of short to long cfDNA fragments for each sample compared to the median fragment length profile calculated from healthy individuals (Fig. 7a, **7b,** and 7e).

⁵While the profiles of cfDNA fragments were remarkably consistent among healthy individuals (median correlation of **0.99),** the median correlation of genome-wide fragment ratios among cancer patients was 0.84 **(0.15** lower, **95% CI 0.07-0.50, p<0.001,** Wilcoxon rank sum test; Table **5** (Appendix **E)).** Similar differences were observed when comparing fragmentation profiles of cancer patients to fragmentation profiles or nucleosome distances in

¹⁰healthy lymphocytes (Fig. 7c, **7d,** and **7f).** To account for potential biases in the fragmentation profiles attributable to **GC** content, a locally weighted smoother was applied independently to each sample and found that differences in fragmentation profiles between healthy individuals and cancer patients remained after this adjustment (median correlation of cancer patients to healthy $= 0.83$) (Table 5; Appendix E).

¹⁵Subsampling analyses of whole genome sequence data was performed at 9x coverage from cfDNA of patients with cancer at \sim 2x, \sim 1x, \sim 0.5x, \sim 0.2x, and \sim 0.1x genome coverage, and it was determined that altered fragmentation profiles were readily identified even at 0.5x genome coverage (Fig. **9).** Based on these observations, whole genome sequencing was performed with coverage of 1-2x to evaluate whether fragmentation profiles may change **²⁰**during the course of targeted therapy in a manner similar to monitoring of sequence

- alterations. cfDNA from **19** non-small cell lung cancer patients including **5** with partial radiographic response, **8** with stable disease, 4 with progressive disease, and 2 with unmeasurable disease, during the course of anti-EGFR or anti-ERBB2 therapy was evaluated (Table **6;** Appendix F). As shown in Fig. **10,** the degree of abnormality in the fragmentation
- 25 profiles during therapy closely matched levels of EGFR or ERBB2 mutant allele fractions as determined using targeted sequencing (Spearman correlation of mutant allele fractions to fragmentation profiles **=** 0.74). This correlation is remarkable as genome-wide and mutation-based methods are orthogonal and examine different cfDNA alterations that may be suppressed in these patients due to prior therapy. Notably all cases that had progression free
- **³⁰**survival of six or more months displayed a drop of or had extremely low levels of ctDNA after initiation of therapy as determined **by** fragmentation profiles, while cases with poor

clinical outcome had increases in ctDNA. These results demonstrate the feasibility of fragmentation analyses for detecting the presence of tumor-derived cfDNA, and suggests that such analyses may also be useful for quantitative monitoring of cancer patients during treatment.

⁵The fragmentation profiles were examined in the context of known copy number changes in a patient where parallel analyses of tumor tissue were obtained. These analyses demonstrated that altered fragmentation profiles were present in regions of the genome that were copy neutral and that these may be further affected in regions with copy number changes (Fig. 11a and Fig. 12a). Position dependent differences in fragmentation patterns 10 could be used to distinguish cancer-derived cfDNA from healthy cfDNA in these regions (Fig. 12a, **b),** while overall cfDNA fragment size measurements would have missed such differences (Fig. 12a).

These analyses were extended to an independent cohort of cancer patients and healthy individuals. Whole genome sequencing of cfDNA at 1-2x coverage from a total of **208** 15 patients with cancer, including breast $(n=54)$, colorectal $(n=27)$, lung $(n=12)$, ovarian $(n=28)$, pancreatic (n=34), gastric (n=27), or bile duct cancers (n=26), as well as **215** individuals without cancer was performed (Table 1 (Appendix **A)** and Table 4 (Appendix **D)). All** cancer patients were treatment naïve and the majority had resectable disease $(n=183)$. After **GC** adjustment of short and long cfDNA fragment coverage (Fig. 13a), coverage and size **²⁰**characteristics of fragments in windows throughout the genome were examined (Fig. **1Ib,** Table 4 (Appendix **D)** and Table **7** (Appendix **G)).** Genome-wide correlations of coverage to **GC** content were limited and no differences in these correlations between cancer patients and healthy individuals were observed (Fig. **13b).** Healthy individuals had **highly** concordant

25 correlation to the median healthy profile (Table **7;** Appendix **G).** An analysis of the most commonly altered fragmentation windows in the genome among cancer patients revealed a median of **60** affected windows across the cancer types analyzed, highlighting the multitude of position dependent alterations in fragmentation of cfDNA in individuals with cancer (Fig. **1Ic).**

fragmentation profiles, while patients with cancer had high variability with decreased

³⁰To determine if position dependent fragmentation changes can be used to detect individuals with cancer, a gradient tree boosting machine learning model was implemented to

examine whether cfDNA can be categorized as having characteristics of a cancer patient or healthy individual and estimated performance characteristics of this approach **by** ten-fold cross validation repeated ten times (Figs. 14 and **15).** The machine learning model included GC-adjusted short and long fragment coverage characteristics in windows throughout the

- **⁵**genome. **A** machine learning classifier for copy number changes from chromosomal arm dependent features rather than a single score was also developed (Fig. 16a and Table **8** (Appendix H)) and mitochondrial copy number changes were also included (Fig. **16b)** as these could also help distinguish cancer from healthy individuals. Using this implementation of **DELFI,** a score was obtained that could be used to classify patients as healthy or having
- **¹⁰**cancer. **152** of the **208** cancer patients were detected **(73%** sensitivity, **95% CI 67%-79%)** while four of the **215** healthy individuals were misclassified **(98%** specificity) (Table **9).** At a threshold of **95%** specificity, **80%** of patients with cancer were detected **(95% CI, 74% 85%),** including **79%** of resectable (stage **I -** III) patients (145 of **183)** and **82%** of metastatic (stage IV) patients **(18** out of 22) (Table **9).** Receiver operator characteristic analyses for
- **¹⁵**detection of patients with cancer had an **AUC** of 0.94 **(95% CI 0.92 - 0.96),** ranged among cancer types from **0.86** for pancreatic cancer to **>0.99** for lung and ovarian cancers (Figs. 17a and **17b),** and had AUCs **>0.92** across all stages (Fig. **18).** The **DELFI** classifier score did not differ with age among either cancer patients or healthy individuals (Table **1;** Appendix **A).**

20

	Individuals			95% specificity		98% specificity			
		analyzed	Individuals detected	Sensitivity	95% CI	Individuals detected	Sensitivity	95% CI	
Healthy		215	10			4			
Cancer		208	166	80%	74%-85%	152	73%	67%-79%	
Type	Breast	54	38	70%	56%-82%	31	57%	43%-71%	
	Bile duct	26	23	88%	70%-98%	21	81%	61%-93%	
	Colorectal	27	22	81%	62%-94%	19	70%	50%-86%	
	Gastric	27	22	81%	62%-94%	22	81%	62%-94%	
	Lung	12	12	100%	74%-100%	12	100%	74%-100%	
	Ovarian	28	25	89%	72%-98%	25	89%	72%-98%	
	Pancreatic	34	24	71%	53%-85%	22	65%	46%-80%	
Stage		41	30	73%	53%-86%	28	68%	52%-82%	
	\mathbf{I}	109	85	78%	69%-85%	78	72%	62%-80%	
	\mathbf{III}	33	30	91%	76%-98%	26	79%	61%-91%	
	IV	22	18	82%	60%-95%	17	77%	55%-92%	
	0, X	3	3	100%	29%-100%	3	100%	29%-100%	

Table **9. DELFI** performance for cancer detection.

To assess the contribution of fragment size and coverage, chromosome arm copy **⁵**number, or mitochondrial mapping to the predictive accuracy of the model, the repeated **10** fold cross-validation procedure was implemented to assess performance characteristics of these features in isolation. It was observed that fragment coverage features alone **(AUC =** 0.94) were nearly identical to the classifier that combined all features **(AUC =** 0.94) (Fig. 17a). In contrast, analyses of chromosomal copy number changes had lower performance **10 (AUC = 0.88)** but were still more predictive than copy number changes based on individual scores $(AUC=0.78)$ or mitochondrial mapping $(AUC=0.72)$ (Fig. 17a). These results suggest that fragment coverage is the major contributor to our classifier. Includingall features in the prediction model may contribute in a complementary fashion for detection of patients with cancer as they can be obtained from the same genome sequence data.

¹⁵As fragmentation profiles reveal regional differences in fragmentation that may differ between tissues, a similar machine learning approach was used to examine whether cfDNA patterns could identify the tissue of origin of these tumors. It was found that this approach had a **61%** accuracy **(95% CI 53%-67%),** including **76%** for breast, **44%** for bile duct, **71%** for colorectal, **67%** for gastric, **53%** for lung, 48% for ovarian, and **50%** for pancreatic

cancers (Fig. **19,** Table **10).** The accuracy increased to **75% (95% CI 69%-81%)** when considering assigning patients with abnormal cfDNA to one of two sites of origin (Table **10).** For all tumor types, the classification of the tissue of origin **by DELFI** was significantly higher than determined **by** random assignment **(p<0.01,** binomial test, Table **10).**

5

Table 10. DELFI tissue of origin prediction

*Patients detected are based on DELFI detection at 90% specificity. Lung cohort includes additional lung cancer patients with prior therapy.

As cancer-specific sequence alterations can be used to identify patients with cancer, it was evaluated whether combining **DELFI** with this approach could increase the sensitivity of cancer detection (Fig. 20). An analysis of cfDNA from a subset of the treatment naive cancer

⁵patients using both **DELFI** and targeted sequencing revealed that **82% (103** of **126)** of patients had fragmentation profile alterations, while **66% (83** of **126)** had sequence alterations. Over **89%** of cases with mutant allele fractions **>1%** were detected **by DELFI** while for cases with mutant allele fractions **<1%** the fraction detected **by DELFI** was **80%,** including for cases that were undetectable using targeted sequencing (Table **7;** Appendix **G). ¹⁰**When these approaches were used together, the combined sensitivity of detection increased to **91% (115** of **126** patients) with a specificity of **98%** (Fig. 20).

Overall, genome-wide cfDNA fragmentation profiles are different between cancer patients and healthy individuals. The variability in fragment lengths and coverage in a position dependent manner throughout the genome may explain the apparently contradictory **¹⁵**observations of previous analyses of cfDNA at specific loci or of overall fragment sizes. In patients with cancer, heterogeneous fragmentation patterns in cfDNA appear to be a result of mixtures of nucleosomal **DNA** from both blood and neoplastic cells. These studies provide a method for simultaneous analysis of tens to potentially hundreds of tumor-specific abnormalities from minute amounts of cfDNA, overcoming a limitation that has precluded **20** the possibility of more sensitive analyses of cfDNA. **DELFI** analyses detected a higher fraction of cancer patients than previous cfDNA analysis methods that have focused on sequence or overall fragmentation sizes (see, e.g., Phallen *et al., 2017 Sci TranslMed* 9:eaan2415; Cohen *et al., 2018 Science* **359:926;** Newman *et al., 2014 Nat Med 20:548;* Bettegowda *et al., 2014 Sci TranslMed* 6:224ra24; Newman *et al., 2016 Nat Biotechnol*

25 34:547). As demonstrated in this Example, combining **DELFI** with analyses of other cfDNA alterations may further increase the sensitivity of detection. As fragmentation profiles appear related to nucleosomal **DNA** patterns, **DELFI** may be used for determining the primary source of tumor-derived cfDNA. The identification of the source of circulating tumor **DNA** in over half of patients analyzed may be further improved **by** including clinical

30 characteristics, other biomarkers, including methylation changes, and additional diagnostic approaches (Ruibal Morell, **1992** *The Internationaljournal of biological markers 7:160;*

Galli et al., **2013** *Clinical chemistry and laboratory medicine* **51:1369;** Sikaris, 2011 *Heart, lung & circulation* 20:634; Cohen *et al., 2018 Science* **359:926).** Finally, this approach requires only a small amount of whole genome sequencing, without the need for deep sequencing typical of approaches that focus on specific alterations. The performance **⁵**characteristics and limited amount of sequencing needed for **DELFI** suggests that our approach could be broadly applied for screening and management of patients with cancer.

These results demonstrate that genome-wide cfDNA fragmentation profiles are different between cancer patients and healthy individuals. As such, cfDNA fragmentation profiles can have important implications for future research and applications of non-invasive **¹⁰**approaches for detection of human cancer.

Other Embodiments

It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined **by** the scope of the appended claims. Other **15** aspects, advantages, and modifications are within the scope of the following claims.

APPENDIX A: Table 1. Summary of patients and samples analyzed

 ${\bf 45}$

 ${\bf 48}$

WO 2019/222657

PCT/US2019/032914

 $52\,$

WO 2019/222657

PCT/US2019/032914

PCT/US2019/032914

 $57\,$

PCT/US2019/032914

APPENDIX -D: Table 4. Summary of whole genome cfDNA analyses

Patient	Timepoint	Analysis type	Patient Type	Read Length	Total Bases	High Quality Bases	Coverage
					Sequenced	Analyzed	
CGCRC291	Preoperative treatment naive	WGS	Colorectal Cancer	100	7232125000	4695396600	1.86
CGCRC292	Preoperative treatment naïve	WGS	Colorectal Cancer	100	6794092800	4471065400	1.77
CGCRC293	Preoperative treatment naive	WGS	Colorectal Cancer	100	8373899600	5686176000	2.26
CGCRC294	Preoperative treatment naïve	WGS	Colorectal Cancer	100	8081312000	5347045800	2.12
CGCRC296	Preoperative treatment naïve	WGS	Colorectal Cancer	100	10072029200	6770998200	2.69
CGCRC299	Preoperative treatment naïve	WGS	Colorectal Cancer	100	10971591600	7632723200	3.03
CGCRC300	Preoperative treatment naïve	WGS	Colorectal Cancer	100	9894332600	6699951000	2.66
CGCRC301	Preoperative treatment naive	WGS	Colorectal Cancer	100	7857346200	5021002000	1.99
CGCRC302	Preoperative treatment naïve	WGS	Colorectal Cancer	100	11671913000	8335275800	3.31
CGCRC304	Preoperative treatment naïve	WGS	Colorectal Cancer	100	19011739200	12957614200	5.14
			Colorectal Cancer				
CGCRC305	Preoperative treatment naive	WGS		100	7177341400	4809957200	1.91
CGCRC306	Preoperative treatment naïve	WGS	Colorectal Cancer	100	8302233200	5608043600	2.23
CGCRC307	Preoperative treatment naïve	WGS	Colorectal Cancer	100	8034729400	5342620000	2.12
CGCRC308	Preoperative treatment naïve	WGS	Colorectal Cancer	100	8670084800	5934037200	2.35
CGCRC311	Preoperative treatment naïve	WGS	Colorectal Cancer	100	6947634400	4704601800	1.87
CGCRC315	Preoperative treatment naïve	WGS	Colorectal Cancer	100	5205544000	3419565400	1.36
CGCRC316	Preoperative treatment naïve	WGS	Colorectal Cancer	100	6405388600	4447534800	1.76
CGCRC317	Preoperative treatment naïve	WGS	Colorectal Cancer	100	6060390400	4104616600	1.63
CGCRC318	Preoperative treatment naïve	WGS	Colorectal Cancer	100	6848768600	4439404800	1.76
CGCRC319	Preoperative treatment naive	WGS	Colorectal Cancer	100	10545294400	7355181600	2.92
CGCRC320	Preoperative treatment naïve	WGS	Colorectal Cancer	100	5961999200	3945054000	1.57
CGCRC321	Preoperative treatment naïve	WGS	Colorectal Cancer	100	8248095400	5614355000	2.23
CGCRC333	Preoperative treatment naive	WGS	Colorectal Cancer	100	10540267600	6915490600	2.74
CGCRC336	Preoperative treatment naïve	WGS	Colorectal Cancer	100	10675581800	7087691800	2.81
CGCRC338	Preoperative treatment naïve	WGS	Colorectal Cancer	100	13788172600	8970308600	3.56
CGCRC341	Preoperative treatment naive	WGS	Colorectal Cancer	100	10753467600	7311539200	2.90
CGCRC342	Preoperative treatment naïve		Colorectal Cancer	100		7552793200	3.00
		WGS			11836966000		
CGH14	Human adult elutriated lymphocytes	WGS	Healthy	100	36525427600	24950300200	9.90
CGH ₁₅	Human adult elutriated lymphocytes	WGS	Healthy	100	29930855000	23754049400	9.43
CGLU316	Pre-treatment, Day -53	WGS	Lung Cancer	100	10354123200	6896471400	2.74
CGLU316	Pre-treatment, Day -4	WGS	Lung Cancer	100	7870039200	5254938800	2.09
CGLU316	Post-treatment, Day 18	WGS	Lung Cancer	100	8155322000	5416262400	2.15
CGLU316	Post-treatment, Day 87	WGS	Lung Cancer	100	9442310400	6087893400	2.42
CGLU344	Pre-treatment, Day -21	WGS	Lung Cancer	100	8728318600	5769097200	2.29
CGLU344	Pre-treatment, Day 0	WGS	Lung Cancer	100	11710249400	7826902600	3.11
CGLU344	Post-treatment, Day 0.1875	WGS	Lung Cancer	100	11569683000	7654701600	3.04
CGLU344	Post-treatment, Day 59	WGS	Lung Cancer	100	11042459200	6320138800	2.51
CGLU369	Pre-treatment, Day -2	WGS	Lung Cancer	100	8630932800	5779595800	2.29
CGLU369	Post-treatment, Day 12	WGS	Lung Cancer	100	9227709600	6136755200	2.44
CGLU369	Post-treatment, Day 68	WGS	Lung Cancer	100	7995282600	5239077200	2.08
CGLU369	Post-treatment, Day 110	WGS	Lung Cancer	100	8750541000	5626139000	2.23
CGLU373	Pre-treatment, Day -2	WGS	Lung Cancer	100	11746059600	7547485800	3.00
CGLU373	Post-treatment, Day 0.125	WGS	Lung Cancer	100	13801136800	9255579400	3.67
CGLU373		WGS		100	11537896800	7654111200	3.04
	Post-treatment, Day 7	WGS	Lung Cancer				
CGLU373	Post-treatment, Day 47		Lung Cancer	100	8046326400	5397702400	2.14
CGPLBR100	Preoperative treatment naïve	WGS	Breast Cancer	100	8440532400	5729474800	2.27
CGPLBR101	Preoperative treatment naïve	WGS	Breast Cancer	100	9786253600	6673495200	2.65
CGPLBR102	Preoperative treatment naïve	WGS	Breast Cancer	100	8664980400	5669781600	2.25
CGFLBR103	Preoperative treatment naive	WGS	Breast Cancer	100	9846936200	6662883400	2.64
CGPLBR104	Preoperative treatment naïve	WGS	Breast Cancer	100	9443375400	6497061000	2.58
CGPLBR12	Preoperative treatment naïve	WGS	Breast Cancer	100	7017577800	4823327400	1.91
CGPLBR18	Preoperative treatment naive	WGS	Breast Cancer	100	10309652800	7130386000	2.83
CGPLBR23	Preoperative treatment naïve	WGS	Breast Cancer	100	9034484800	6219625800	2.47
CGPLBR24	Preoperative treatment naïve	WGS	Breast Cancer	100	9891454200	6601857400	2.62
CGPLBR28	Preoperative treatment naive	WGS	Breast Cancer	100	7997607200	5400803200	2.14
CGPLBR30	Preoperative treatment naïve	WGS	Breast Cancer	100	8502597200	5885822400	2.34
CGPLBR31	Preoperative treatment naïve	WGS	Breast Cancer	100	12660085600	8551995600	3.39
CGPLBR32	Preoperative treatment naïve	WGS	Breast Cancer	100	8773498600	5839034600	2.32
CGPLBR33	Preoperative treatment naïve	WGS	Breast Cancer	100	10931742800	6967030600	2.76
CGPLBR34	Preoperative treatment naïve	WGS	Breast Cancer	100	10861398600	7453225800	2.96
CGPLBR35	Preoperative treatment naïve	WGS	Breast Cancer	100	9180193600	6158440200	2.44
CGPLBR36	Preoperative treatment naive	WGS	Breast Cancer	100	9159948400	6091817800	2.42
CGPLBR37	Preoperative treatment naïve	WGS	Breast Cancer	100	10307505800	6929530600	2.75
CGPLBR38	Preoperative treatment naive	WGS	Breast Cancer	100	9983824000	6841725400	2.71
CGPLBR40	Preoperative treatment naive	WGS	Breast Cancer	100	10148823800	7024345400	2.79
CGPLBR41	Preoperative treatment naïve	WGS	Breast Cancer	100	11168192000	7562945800	3.00
CGPLBR45	Preoperative treatment naïve	WGS	Breast Cancer	100	8793780600	6011109400	2.39
CGPLBR46	Preoperative treatment naïve	WGS	Breast Cancer	100	7228607600	4706130000	1.87
CGPLBR47	Preoperative treatment naive	WGS	Breast Cancer	100	7906911400	5341655000	2.12
CGPLBR48	Preoperative treatment naïve	WGS	Breast Cancer	100	6992032000	4428636200	1.76
CGPLBR49	Preoperative treatment naïve	WGS	Breast Cancer	100	7311195000	4559460200	1.81
CGPLBR50	Preoperative treatment naïve	WGS	Breast Cancer	100	11107960600	7582776600	3.01

APPENDIX E: Table 5. High coverage whole genome cfDNA analyses of healthy individuals and lung cancer patients

 ${\bf 72}$

WO 2019/222657

WO 2019/222657

 ${\bf 76}$

Patient Type $\frac{1}{2}\frac{1$ Patient

WHAT IS CLAIMED IS:

1. A method of determining a cell free **DNA** (cfDNA) fragmentation profile of a mammal, the method comprising:

processing cfDNA fragments obtained from a sample obtained from the mammal into sequencing libraries;

subjecting the sequencing libraries to low-coverage whole genome sequencing to obtain sequenced fragments;

mapping the sequenced fragments to a genome to obtain windows of mapped sequences; and

analyzing the windows of mapped sequences to determine cfDNA fragment lengths.

2. The method of claim **1,** wherein the mapped sequences comprise tens to thousands of windows.

3. The method of claims 1-2, wherein the windows are non-overlapping windows.

4. The method of any one of claims **1-3,** wherein the windows each comprise about **5** million base pairs.

5. The method of any one of claims 1-4, wherein a cfDNA fragmentation profile is determined within each window.

6. The method of any one of claims **1-5,** wherein cfDNA fragmentation profile comprises a median fragment size.

7. The method of any one of claims **1-5,** wherein cfDNA fragmentation profile comprises a fragment size distribution.

8. The method of any one of claims **1-5,** wherein the cfDNA fragmentation profile comprises a ratio of small cfDNA fragments to large cfDNA fragments in said windows of mapped sequences.

9. The method of any one of claims **1-5,** wherein the cfDNA fragmentation profile comprises the sequence coverage of small cfDNA fragments in windows across the genome.

10. The method of any one of claims **1-5,** wherein the cfDNA fragmentation profile comprises the sequence coverage of large cfDNA fragments in windows across the genome.

11. The method of any one of claims **1-5,** wherein the cfDNA fragmentation profile comprises the sequence coverage of small and large cfDNA fragments in windows across the genome.

12. The method of any one of claims **1-11,** wherein the cfDNA fragmentation profile is over the whole genome.

13. The method of any one of claims **1-11,** wherein the cfDNA fragmentation profile is over a subgenomic interval.

14. **A** method of identifying a mammal as having cancer, the method comprising:

determining a cell free **DNA** (cfDNA) fragmentation profile in a sample obtained from the mammal;

comparing the cfDNA fragmentation profile to a reference cfDNA fragmentation profile; and

identifying the mammal as having cancer when the cfDNA fragmentation profile obtained from the mammal is different from the reference cfDNA fragmentation profile.

15. The method of claim 14, wherein the reference cfDNA fragmentation profile is a cfDNA fragmentation profile of a healthy mammal.

16. The method of claim **15,** wherein the reference cfDNA fragmentation profile is generated **by** determining a cfDNA fragmentation profile in a sample obtained from the healthy mammal.

17. The method of claim 14, wherein the reference **DNA** fragmentation pattern is a reference nucleosome cfDNA fragmentation profile.

18. The method of any one of claims 14-17, wherein the cfDNA fragmentation profile comprises a median fragment size, and wherein a median fragment size of the cfDNA fragmentation profile is shorter than a median fragment size of the reference cfDNA fragmentation profile.

19. The method of any one of claims 14-17, wherein the cfDNA fragmentation profile comprises a fragment size distribution, and wherein a fragment size distribution of the cfDNA fragmentation profile differs **by** at least **10** nucleotides as compared to a fragment size distribution of the reference cfDNA fragmentation profile.

20. The method of any one of claims 14-17, wherein the cfDNA fragmentation profile comprises a ratio of small cfDNA fragments to large cfDNA fragments in said windows of mapped sequences, wherein a small cfDNA fragment is **100** base pairs **(bp)** to **150 bp** in length, wherein a large cfDNA fragments is **151 bp** to 220 **bp** in length, and wherein a correlation of fragment ratios in the cfDNA fragmentation profile is lower than a correlation of fragment ratios of the reference cfDNA fragmentation profile.

21. The method of any one of claims 14-17, wherein the cfDNA fragmentation profile comprises the sequence coverage of small cfDNA fragments in windows across the genome.

22. The method of any one of claims 14-17, wherein the cfDNA fragmentation profile comprises the sequence coverage of large cfDNA fragments in windows across the genome. **23.** The method of any one of claims 14-17, wherein the cfDNA fragmentation profile comprises the sequence coverage of small and large cfDNA fragments in windows across the genome.

24. The method of any one of claims 14-17, wherein the cancer is selected from the group consisting of: colorectal cancer, lung cancer, breast cancer, gastric cancers, pancreatic cancers, bile duct cancers, and ovarian cancer.

25. The method of claim 14, wherein the step of comparing comprises comparing the cfDNA fragmentation profile to a reference cfDNA fragmentation profile over the whole genome.

26. The method of claim 14, wherein the step of comparing comprises comparing the cfDNA fragmentation profile to a reference cfDNA fragmentation profile over a subgenomic interval.

27. The method of any one of claim 14-23, wherein the mammal has previously been administered a cancer treatment to treat the cancer.

28. The method of claim **27,** wherein the cancer treatment is selected from the group consisting of: surgery, adjuvant chemotherapy, neoadjuvant chemotherapy, radiation therapy, hormone therapy, cytotoxic therapy, immunotherapy, adoptive T cell therapy, targeted therapy, and combinations thereof.

29. The method of any one of claims 14-28, further comprising administering to the mammal a cancer treatment selected from the group consisting of: surgery, adjuvant chemotherapy, neoadjuvant chemotherapy, radiation therapy, hormone therapy, cytotoxic therapy, immunotherapy, adoptive T cell therapy, targeted therapy, and combinations thereof.

30. The method of claim **29,** wherein the mammal is monitored for the presence of cancer after administration of the cancer treatment.

31. The method of any one of claim 14 to claim **30,** the method further comprising identifying one or more cancer-specific sequence alterations in the sample.

32. The method of any one of claim 14 to claim **30,** the method further comprising identifying one or more chromosomal abnormalities in the sample.

33. The method of claim **32,** wherein the one or more chromosomal abnormalities comprises a copy number change in one or more chromosome arms.

34. **A** method of identifying the tissue of origin of a cancer in a mammal identified as having a cancer, the method comprising:

determining a cell free **DNA** (cfDNA) fragmentation profile in a sample obtained from the mammal;

comparing the cfDNA fragmentation profile to a reference cfDNA fragmentation profile; and

identifying the tissue of origin of the cancer in a mammal when the cfDNA fragmentation profile obtained from the mammal matches a reference cfDNA fragmentation profiles from a mammal identified as having a cancer with the same tissue of origin.

35. The method of claim 34, wherein the reference cfDNA fragmentation profile comprises reference cfDNA fragmentation profiles from mammals identified as having one or more of colorectal cancer, lung cancer, breast cancer, gastric cancers, pancreatic cancers, bile duct cancers, and ovarian cancer.

36. The method of claim **35,** wherein the reference cfDNA fragmentation profile is generated **by** determining a cfDNA fragmentation profile in a sample obtained from the mammals

identified as having one or more or colorectal cancer, lung cancer, breast cancer, gastric cancers, pancreatic cancers, bile duct cancers, and ovarian cancer.

37. The method of claim 34, wherein the reference **DNA** fragmentation pattern is a reference nucleosome cfDNA fragmentation profile.

38. The method of any one of claims **34-37,** wherein the cfDNA fragmentation profile comprises a median fragment size, and wherein a median fragment size of the cfDNA fragmentation profile is shorter than a median fragment size of the reference cfDNA fragmentation profile.

39. The method of any one of claims **34-37,** wherein the cfDNA fragmentation profile comprises a fragment size distribution, and wherein a fragment size distribution of the cfDNA fragmentation profile differs **by** at least **10** nucleotides as compared to a fragment size distribution of the reference cfDNA fragmentation profile.

40. The method of any one of claims **34-37,** wherein the cfDNA fragmentation profile comprises a ratio of small cfDNA fragments to large cfDNA fragments in said windows of mapped sequences, wherein a small cfDNA fragment is **100** base pairs **(bp)** to **150 bp** in length, wherein a large cfDNA fragments is **151 bp** to 220 **bp** in length, and wherein a correlation of fragment ratios in the cfDNA fragmentation profile is lower than a correlation of fragment ratios of the reference cfDNA fragmentation profile.

41. The method of any one of claims **34-37,** wherein the cfDNA fragmentation profile comprises the sequence coverage of small cfDNA fragments in windows across the genome.

42. The method of any one of claims **34-37,** wherein the cfDNA fragmentation profile comprises the sequence coverage of large cfDNA fragments in windows across the genome.

43. The method of any one of claims **34-37,** wherein the cfDNA fragmentation profile comprises the sequence coverage of small and large cfDNA fragments in windows across the genome.

44. The method of any one of claims **34-37,** wherein the cancer is selected from the group consisting of: colorectal cancer, lung cancer, breast cancer, gastric cancers, pancreatic cancers, bile duct cancers, and ovarian cancer.

45. The method of claim 34, wherein the step of comparing comprises comparing the cfDNA fragmentation profile to a reference cfDNA fragmentation profile over the whole genome.

46. The method of claim 34, wherein the step of comparing comprises comparing the cfDNA fragmentation profile to a reference cfDNA fragmentation profile over a subgenomic interval.

47. The method of any one of claims 34-46, the method further comprising identifying one or more cancer-specific sequence alterations in the sample.

48. The method of any one of claims 34-46, the method further comprising identifying one or more chromosomal abnormalities in the sample.

49. The method of claim 48, wherein the one or more chromosomal abnormalities comprises a copy number change in one or more chromosome arms.

50. A method treating a mammal having cancer, the method comprising: identifying said mammal as having cancer, wherein said identifying comprises: determining a cell free **DNA** (cfDNA) fragmentation profile in a sample obtained from the mammal;

comparing the cfDNA fragmentation profile to a reference cfDNA fragmentation profile; and

identifying the mammal as having cancer when the cfDNA fragmentation profile obtained from the mammal is different from the reference cfDNA fragmentation profile; and

administering a cancer treatment to said mammal.

51. The method of claim **50,** wherein said mammal is a human.

52. The method of any one of claims **50-51,** wherein the cancer is selected from the group consisting of: colorectal cancer, lung cancer, breast cancer, gastric cancers, pancreatic cancers, bile duct cancers, and ovarian cancer.

53. The method of any one of claims **50-52,** wherein said cancer treatment is selected from the group consisting of: surgery, adjuvant chemotherapy, neoadjuvant chemotherapy, radiation therapy, hormone therapy, cytotoxic therapy, immunotherapy, adoptive T cell therapy, targeted therapy, and combinations thereof.

54. The method of any one of claims **50-53,** wherein the reference cfDNA fragmentation profile is a cfDNA fragmentation profile of a healthy mammal.

55. The method of claim 54, wherein the reference cfDNA fragmentation profile is generated **by** determining a cfDNA fragmentation profile in a sample obtained from the healthy mammal.

56. The method of any one of claims **50-53,** wherein the reference **DNA** fragmentation pattern is a reference nucleosome cfDNA fragmentation profile.

57. The method of any one of claims **50-56,** wherein the cfDNA fragmentation profile comprises a median fragment size, and wherein a median fragment size of the cfDNA

fragmentation profile is shorter than a median fragment size of the reference cfDNA fragmentation profile.

58. The method of any one of claims **50-56,** wherein the cfDNA fragmentation profile comprises a fragment size distribution, and wherein a fragment size distribution of the cfDNA fragmentation profile differs **by** at least **10** nucleotides as compared to a fragment size distribution of the reference cfDNA fragmentation profile.

59. The method of any one of claims **50-56,** wherein the cfDNA fragmentation profile comprises a ratio of small cfDNA fragments to large cfDNA fragments in said windows of mapped sequences, wherein a small cfDNA fragment is **100** base pairs **(bp)** to **150 bp** in length, wherein a large cfDNA fragments is **151 bp** to 220 **bp** in length, and wherein a correlation of fragment ratios in the cfDNA fragmentation profile is lower than a correlation of fragment ratios of the reference cfDNA fragmentation profile.

60. The method of any one of claims **50-56,** wherein the cfDNA fragmentation profile comprises the sequence coverage of small cfDNA fragments in windows across the genome.

61. The method of any one of claims **50-56,** wherein the cfDNA fragmentation profile comprises the sequence coverage of large cfDNA fragments in windows across the genome.

62. The method of any one of claims **50-56,** wherein the cfDNA fragmentation profile comprises the sequence coverage of small and large cfDNA fragments in windows across the genome.

63. The method of any one of claims **50-62,** wherein the step of comparing comprises comparing the cfDNA fragmentation profile to a reference cfDNA fragmentation profile over the whole genome.

64. The method of any one of claims **50-62,** wherein the step of comparing comprises comparing the cfDNA fragmentation profile to a reference cfDNA fragmentation profile over a subgenomic interval.

65. The method of any one of claims 50-64, wherein the mammal has previously been administered a cancer treatment to treat the cancer.

66. The method of claim **65,** wherein the cancer treatment is selected from the group consisting of: surgery, adjuvant chemotherapy, neoadjuvant chemotherapy, radiation therapy, hormone therapy, cytotoxic therapy, immunotherapy, adoptive T cell therapy, targeted therapy, and combinations thereof.

67. The method of any one of claims **50-66,** wherein the mammal is monitored for the presence of cancer after administration of the cancer treatment.

<u>dda sababan an t-ainm anns an t-ainm anns an t-ai</u>

Noninvasive cancer screening (DELFI)

FIG. 1

FIG. 2

cfDNA fragments size

FIG. 3 (Cont.)

GC Content (%)

FIG. 4A (Cont.)

FIG.4B

 $FIG. 5$

 $8/33$

CGCRC297

CGCRC296

CGCRC292

CGCRC306

CGCRC306

PCT/US2019/032914

 $9/33$

cfDNA Fragment Size

FIG. 5(Cont.)

100 125 150 175 200

11/33

CfDNA fragment size

FIG. 6(Cont.)

FIG.7B

FIG. 8

SUBSTITUTE SHEET (RULE 26)

22/33

FIG. 13A

FIG. 13B

FIG. 14

Healthy Cancer Healthy Cancer Healthy Cancer \bullet \cdot 4_D 16p $11p$ $\overline{7}$ q $19q$ \blacksquare FIG. 16A \cdot : the \cdot g. 机公 3q **7p** $10q$ $15q$ $19p$ \bullet 1 1 r .: ŋ, နှ $6q$ $14q$ 10_p $18q$ \bullet \mathbf{r} 1 Healthy Cancer Healthy Cancer Healthy Cancer Healthy Cancer -2.6 $2q$ \overline{g} $13q$ 18p $22q$ 8 \mathbf{z} \blacksquare л в AS . tille. အြ $12q$ $17q$ $21q$ $\mathcal{Q}_\mathbf{D}$ கி \blacksquare я × 2 В \overline{q} $\overline{5}$ 8_q $\overline{12p}$ $17p$ $20q$ 8 L J. 1 1 t. 为.* $\frac{1}{4}$ $8₀$ \overline{a} $11q$ **16q** $20p$ 1 —————
အိမ္မွ တိုက္ပ္ <u>ន</u>់ដូកដូច់ នន់កនុំចំ င္ထဲမွာေမွ်ကို င်္ကဗံုဗံု នន់ុក់ទំន S score

26/33

FIG. 16B

FIG. 17A

FIG. 20