



- (51) International Patent Classification:
C12Q 1/6886 (2018.01) G16H 50/20 (2018.01)
- (21) International Application Number:
PCT/US2024/015444
- (22) International Filing Date:
12 February 2024 (12.02.2024)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
63/445,284 13 February 2023 (13.02.2023) US
63/470,101 31 May 2023 (31.05.2023) US
63/528,237 21 July 2023 (21.07.2023) US
- (71) Applicant: DELFI DIAGNOSTICS, INC. [US/US]; 2809 Boston Street, Suite 503, Baltimore, Maryland 21224 (US).
- (72) Inventors: SKIDMORE, Zachary L.; c/o DELFI DIAGNOSTICS, INC., 2809 Boston Street, Suite 503, BALTIMORE, Maryland 21224 (US). RINALDI, Lorenzo; c/o DELFI DIAGNOSTICS, INC., 2809 Boston Street, Suite 503, Baltimore, Maryland 21224 (US).
- (74) Agent: HAILE, Lisa A. et al.; DLA PIPER LLP US, 4365 Executive Drive, Suite 1100, SAN DIEGO, California 92121-2133 (US).

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

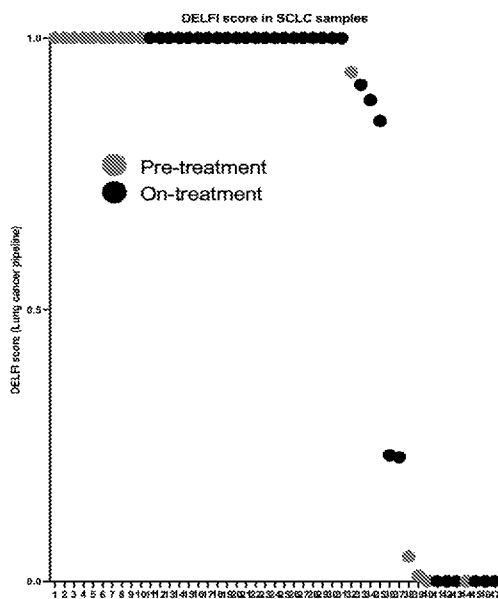
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR,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

Published:

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

(54) Title: DELFI-DERIVED CELL-FREE DNA FRAGMENTATION PATTERNS DIFFERENTIATE HISTOLOGIC SUBTYPES OF LUNG CANCERS IN A NON-INVASIVE MANNER

FIGURE 1



(57) Abstract: The present disclosure provides methods of uses there-of for improved diagnostic applications using genome-wide patterns of fragmented cell-free DNA (cfDNA) from plasma, derived by low-coverage whole-genome sequencing. In particular, the present invention provides new and effective methods for subtyping non-small cell lung cancer in a subject as low neuroendocrine or high neuroendocrine non-small cell lung cancer.



**DELFI-DERIVED CELL-FREE DNA FRAGMENTATION PATTERNS
DIFFERENTIATE HISTOLOGIC SUBTYPES OF LUNG CANCERS IN A NON-
INVASIVE MANNER**

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of priority under 35 U.S.C. § 119(e) of U.S. Provisional Patent Application Serial No. 63/445,284 filed on February 13, 2023, U.S. Provisional Patent Application Serial No. 63/470,101 filed on May 31, 2023, and U.S. Provisional Patent Application Serial No. 63/528,237 filed on July 21, 2023. The disclosures of the prior applications are considered part of and are herein incorporated by reference in the disclosure of this application in their entirety.

FIELD OF THE INVENTION

[0002] The present invention relates generally to the non-invasive diagnosis and subtyping of small cell lung cancer (SCLC) and more specifically to the analysis of genome-wide patterns of fragmented cell-free DNA (cfDNA) in conjunction with clinical and demographic features of individual patients. The present invention is also related to the of Cell-free DNA fragmentation profiling as a method for tumor fraction assessment and treatment monitoring in non-small cell lung cancer (NSCLC). Given the logistical difficulties to perform SCLC and NSCLC tumor biopsies, new approaches are needed for viable and quick methods of subtyping SCLC and NSCLC in a non-invasive manner.

BACKGROUND

[0003] Small cell lung cancer is an aggressive malignancy with a poor prognosis. Although Small cell lung cancers are clinically managed as a single cancer type, new evidence supports these subtypes (high- neuroendocrine vs low-neuroendocrine) of small cell lung cancer acquire diverse transcriptional and epigenetic states. Furthermore, distinct small cell lung cancer subtypes respond to specific treatment such as immunotherapy in the case of the low-neuroendocrine highly inflamed small cell lung cancer subtype. non-small cell lung cancer is any type of epithelial lung cancer other than small cell lung cancer. The most common types of non-small cell lung cancer are squamous cell carcinoma, large cell carcinoma, and adenocarcinoma, but there are several other types that occur less frequently, and all types can

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

occur in unusual histological variants. As a class, non-small cell lung cancer is usually less sensitive to chemotherapy and radiation therapy than small cell lung cancer. Patients with resectable disease may be cured by surgery or surgery followed by chemotherapy, as well as chemotherapy followed by surgery. Local control can be achieved with radiation therapy in many patients with unresectable disease, but cure is seen in relatively few patients. Patients with locally advanced unresectable disease may achieve long-term survival with radiation therapy combined with chemotherapy. Patients with advanced metastatic disease may achieve improved survival and palliation of symptoms with chemotherapy, targeted agents, and other supportive measures. Novel, non-invasive methods for identifying and subtyping non-small cell lung cancer and small cell lung cancer are needed to improve patient outcomes.

SUMMARY OF THE INVENTION

[0004] The present invention is based on the seminal discovery that the characterizing genome-wide patterns of fragmentation of cell-free DNA (cfDNA) in plasma using low-coverage whole-genome sequencing can improve cancer diagnosis when analyzed in conjunction with certain clinical and demographic features of individual patients.

[0005] In one embodiment, the present invention provides methods for subtyping small cell lung cancer in a subject as low neuroendocrine or high neuroendocrine small cell lung cancer comprising: processing cfDNA fragments from a sample obtained from the subject and generating sequencing libraries; subjecting the sequencing libraries to whole genome sequencing to obtain sequenced fragments, wherein genome coverage is about 30× to 0.1×; mapping the sequenced fragments to a genome to obtain genomic intervals of mapped sequences at specified transcription factor binding sites; analyzing the genomic intervals of mapped sequences to determine cfDNA fragment lengths and amounts to establish a cfDNA fragment coverage score at specified transcription factor binding sites using the cfDNA fragment lengths and amounts; and subtyping the small cell lung cancer in the subject based on transcription factor activation; wherein a decrease in an aggregate cfDNA fragment coverage scores at the specified transcription factor binding sites is indicative of a high-neuroendocrine small cell lung cancer subtype.

[0006] In one embodiment, the present invention provides methods for subtyping small cell lung cancer in a subject as low neuroendocrine or high neuroendocrine small cell lung cancer comprising: processing cfDNA fragments from a sample obtained from the subject and

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

generating sequencing libraries; subjecting the sequencing libraries to whole genome sequencing to obtain sequenced fragments, wherein genome coverage is about $10\times$ to $0.1\times$; mapping the sequenced fragments to a genome to obtain genomic intervals of mapped sequences at specified transcription factor binding sites; analyzing the genomic intervals of mapped sequences to determine cfDNA fragment lengths and amounts to establish a cfDNA fragment coverage score at specified transcription factor binding sites using the cfDNA fragment lengths and amounts; and subtyping the small cell lung cancer in the subject based on transcription factor activation; wherein a decrease in an aggregate cfDNA fragment coverage scores at the specified transcription factor binding sites is indicative of a high-neuroendocrine small cell lung cancer subtype.

[0007] In one embodiment, the present invention provides methods for subtyping small cell lung cancer in a subject as low neuroendocrine or high neuroendocrine small cell lung cancer comprising: processing cfDNA fragments from a sample obtained from the subject and generating sequencing libraries; subjecting the sequencing libraries to whole genome sequencing to obtain sequenced fragments, wherein genome coverage is about $9\times$ to $0.1\times$; mapping the sequenced fragments to a genome to obtain genomic intervals of mapped sequences at specified transcription factor binding sites; analyzing the genomic intervals of mapped sequences to determine cfDNA fragment lengths and amounts to establish a cfDNA fragment coverage score at specified transcription factor binding sites using the cfDNA fragment lengths and amounts; and subtyping the small cell lung cancer in the subject based on transcription factor activation; wherein a decrease in an aggregate cfDNA fragment coverage scores at the specified transcription factor binding sites is indicative of a high-neuroendocrine small cell lung cancer subtype. In some aspects, the method uses the specified transcription factor is ASCL1, NEUROD1, POUF23, YAP1, or any combination thereof.

[0008] In some aspects, the method uses machine learning to subtype small cell lung cancer in a subject.

[0009] In some aspects, subtyping is performed by calculating a \log_2 (Read Depth Ratio) across 20bp windows starting 2kb from the specified transcription factor binding sites.

[0010] In some aspects, the \log_2 (Read Depth Ratio) is the total coverage over a coverage correction factor calculated using the median coverage of 5' and 3' 500bp anchors at the ends of the 2kb window (i.e. $\log_2(\text{Read Depth}/\text{Correction Factor})$).

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

[0011] In some aspects, the method uses all coverage calculations are shifted by 1 to avoid divisions by 0. In some aspects this adjustment is expressed as $f(x) = \log_2((\text{window depth} + 1)/(\text{median}(\text{anchors}) + 1))$.

[0012] In some aspects, the \log_2 (Read Depth Ratio) is calculated for each transcription factor binding site and then aggregated by calculating the median of each window to obtain a fragment coverage score.

[0013] Some aspects further comprise the use of a general additive model to smooth coverage and correct for GC bias.

[0014] In some aspects, the genomic intervals are non-overlapping.

[0015] In some aspects, the genomic intervals each comprise thousands to millions of base pairs.

[0016] In some aspects, a cfDNA fragmentation profile is determined within each genomic intervals.

[0017] In some aspects, the cfDNA fragmentation profile comprises a median fragment size.

[0018] In some aspects, the cfDNA fragmentation profile comprises a fragment size distribution.

[0019] Some aspects further comprise administering to the subject identified as having a high-neuroendocrine small cell lung cancer subtype, a therapeutic agent suitable for the treatment of the type of cancer.

[0020] In some aspects, the therapeutic agent is an immunotherapy.

[0021] In one embodiment, the present invention provides methods for the non-invasive subtyping of non-small cell lung cancer in a subject as adenocarcinoma or squamous carcinoma comprising: processing cfDNA fragments from a sample obtained from the subject and generating sequencing libraries; subjecting the sequencing libraries to whole genome sequencing to obtain sequenced fragments, wherein genome coverage is about $30\times$ to $0.1\times$; mapping the sequenced fragments to a genome to obtain genomic intervals of mapped sequences at specified transcription factor binding sites; analyzing the genomic intervals of mapped sequences to determine cfDNA fragment lengths and amounts to establish a cfDNA fragment coverage score at specified transcription factor binding sites using the cfDNA fragment lengths and amounts; and subtyping the small cell lung cancer in the subject based on transcription factor activation; wherein a decrease in an aggregate cfDNA fragment

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

coverage scores at the specified transcription factor binding sites is indicative of an adenocarcinoma or squamous carcinoma subtype.

[0022] In one embodiment, the present invention provides methods for the non-invasive subtyping of non-small cell lung cancer in a subject as adenocarcinoma or squamous carcinoma comprising: processing cfDNA fragments from a sample obtained from the subject and generating sequencing libraries; subjecting the sequencing libraries to whole genome sequencing to obtain sequenced fragments, wherein genome coverage is about $10\times$ to $0.1\times$; mapping the sequenced fragments to a genome to obtain genomic intervals of mapped sequences at specified transcription factor binding sites; analyzing the genomic intervals of mapped sequences to determine cfDNA fragment lengths and amounts to establish a cfDNA fragment coverage score at specified transcription factor binding sites using the cfDNA fragment lengths and amounts; and subtyping the small cell lung cancer in the subject based on transcription factor activation; wherein a decrease in an aggregate cfDNA fragment coverage scores at the specified transcription factor binding sites is indicative of an adenocarcinoma or squamous carcinoma subtype.

[0023] In one embodiment, the present invention provides methods for the non-invasive subtyping of non-small cell lung cancer in a subject as adenocarcinoma or squamous carcinoma comprising: processing cfDNA fragments from a sample obtained from the subject and generating sequencing libraries; subjecting the sequencing libraries to whole genome sequencing to obtain sequenced fragments, wherein genome coverage is about $9\times$ to $0.1\times$; mapping the sequenced fragments to a genome to obtain genomic intervals of mapped sequences at specified transcription factor binding sites; analyzing the genomic intervals of mapped sequences to determine cfDNA fragment lengths and amounts to establish a cfDNA fragment coverage score at specified transcription factor binding sites using the cfDNA fragment lengths and amounts; and subtyping the small cell lung cancer in the subject based on transcription factor activation; wherein a decrease in an aggregate cfDNA fragment coverage scores at the specified transcription factor binding sites is indicative of an adenocarcinoma or squamous carcinoma subtype.

[0024] In some aspects, the specified transcription factor is ASCL1, NEUROD1, POUF23, YAP1, or any combination thereof.

[0025] In some aspects, the method uses machine learning to subtype small cell lung cancer in the subject.

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

[0026] In some aspects, subtyping is performed by calculating a log₂ (Read Depth Ratio) across 20bp windows starting 2kb from the specified transcription factor binding sites.

[0027] In some aspects, the log₂ (Read Depth Ratio) is the total coverage over a coverage correction factor calculated using the median coverage of 5' and 3' 500bp anchors at the ends of the 2kb window (i.e. log₂(Read Depth/Correction Factor)).

[0028] In some aspects, all coverage calculations are shifted by 1 to avoid divisions by 0. In some aspects this adjustment is expressed as $f(x) = \log_2((\text{window depth} + 1)/(\text{median}(\text{anchors}) + 1))$.

[0029] In some aspects, the log₂ (Read Depth Ratio) is calculated for each transcription factor binding site and then aggregated by calculating the median of each window to obtain a fragment coverage score.

[0030] Some aspects further comprise the use of a general additive model to smooth coverage and correct for GC bias.

[0031] In some aspects, the genomic intervals are non-overlapping.

[0032] In some aspects, the genomic intervals each comprise thousands to millions of base pairs.

[0033] In some aspects, a cfDNA fragmentation profile is determined within each genomic intervals.

[0034] In some aspects, the cfDNA fragmentation profile comprises a median fragment size.

[0035] In some aspects, the cfDNA fragmentation profile comprises a fragment size distribution.

[0036] Some aspects further comprise administering to the subject identified as having an adenocarcinoma or squamous carcinoma subtype, a therapeutic agent suitable for the treatment of the type of cancer.

[0037] In some aspects, the therapeutic agent is an immunotherapy.

[0038] Some aspects further comprise quantifying a circulating tumor fraction, mirroring major allele fraction (MAF) performance in relation to Response Evaluation Criteria in Solid Tumors (RECIST) evaluation in both Adenocarcinoma and Squamous carcinoma.

BRIEF DESCRIPTION OF THE DRAWINGS

[0039] **Figure 1** illustrates DELFI-scores of the patients diagnosed with SCLC in the study of Example 1.

[0040] **Figure 2** illustrates that genome-wide fragmentation profiles were significantly consistent among pre-treatment, post-treatment, progression and response time points suggesting genome-wide circulating tumor DNA fragment sizes is a powerful method to detect changes during monitoring of immunotherapy treatment of SCLC cancer treatment.

[0041] **Figure 3** illustrates genome-wide fragmentation profiles (and correspondent DELFI-scores below) showing partial differences between the two main subtypes of SCLC cases: High Neuroendocrine and Low Neuroendocrine; suggesting genome-wide circulating tumor DNA fragment sizes may be a powerful method to detect subtyping during monitoring of immunotherapy treatment of SCLC cancer treatment.

[0042] **Figure 4** illustrates differential genes between High and Low neuroendocrine SCLC cases, using publicly available data (Lissa et al. 2022) and a principal Component Analysis (on DELFI 30X plasma WGS data) to cluster the two subtypes. PCA analysis data revealed distinguishable clusters between SCLC High-NE vs Low-NE pre-treatment samples.

[0043] **Figures 5A-5B** illustrates an Eigencor analysis revealing a significant correlation between Principal Component 1 and 2 with the NE status of the samples.

[0044] **Figures 6A-6B** illustrates the potential of the DELFI assay to subtype SCLC using TFBS. A genome-wide cfDNA fragmentation analyses were performed at ASCL1 binding sites in the pre-treatment NCI samples. Figure 5A illustrates that distinct clusters of SCLC samples were observed corresponding to different levels of ASCL1 activation. Importantly the two clusters corresponded perfectly to the different SCLC subtypes. The main driver of differentiation between samples was in fact differences in the fragment coverage identified at the center of the ASCL1 binding sites. Figure 6B illustrates that other clinical or sample characteristics have a negligible impact in the clustering of these samples. Overall, these data show how DELFI analysis can perfectly distinguish between SCLC subtypes using fragment coverage at TFBS.

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

[0045] **Figure 7** illustrates a patient example supporting the hypothesis that the detected signal comes from tumor infiltrating lymphocytes. Patient NCI-0422, diagnosed with SCLC with inflammation subtype was treated with Durvalumab plus Olaparib. NCI-0422 responded well to treatment but was later diagnosed with progressive disease. Variable genomic binding is detectable at pre-treatment and at progression timepoints. Peaks are centered at a mix of TFBS and TSS.

[0046] **Figures 8A-8B** illustrates ASCL1 binding site between resp vs non-responders. Figure 8A shows 500 cell type specific bins from the PMD data to determine fraction of white blood cells.

[0047] **Figure 9** illustrates a machine learning model to predict high vs low neuroendocrine SCLC.

[0048] **Figures 10A-10B** illustrates that the method described herein is able to differentiate neuroendocrine vs non-neuroendocrine by calculating Read Depth Ratio at SCLC specific genomic coordinates (Figure 10A.). Figure 10B shows that the method described herein appears to subtype better SCLC cases.

[0049] **Figure 11** illustrates that Read Depth Ratio at SCLC specific genomic coordinates are transformed into a 2-dimensional PCA to separate samples in Neuroendocrine vs non-neuroendocrine groups. Most of samples are clearly divided into different clusters.

[0050] **Figure 12** illustrates PCA derived from Read Depth Ratio at SCLC specific genomic coordinates is able to separate pre-treatment samples into specific SCLC subtypes (A,N,P,Y). The ground truth of the SCLC subtypes was calculated using tissue RNA-seq gene expression differential analysis.

[0051] **Figure 13** illustrates depicts how fragmentation profiles inform on lung cancer status and treatment response patterns.

[0052] **Figure 14** illustrates that model derived DELFI-TF exhibits a strong correlation with Mutant Allele Frequency across samples.

[0053] **Figure 15** illustrates that DELFI-TF accurately quantify circulating tumor fraction, mirroring MAF performance in relation to RECIST evaluation.

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

- [0054] **Figure 16** illustrates that tumor derived cfDNA has altered fragmentation.
- [0055] **Figures 17A-17C** illustrates that DELFI-TF accurately detects circulating tumor fraction without being confounded by clonal hematopoiesis.
- [0056] **Figure 18** depicts cfDNA fragmentation patterns accurately differentiate NSCLC subtypes.
- [0057] **Figure 19** depicts proof-of-concept DELFI-TF model development.
- [0058] **Figure 20** is an example computer 800 that may be used to implement the methods described herein.

DETAILED DESCRIPTION

[0059] The present invention is based on the seminal discovery that the characterizing genome-wide patterns of fragmentation of cell-free DNA (cfDNA) in plasma using low-coverage whole-genome sequencing improves cancer diagnosis when analyzed in conjunction with certain clinical and demographic features of individual patients.

[0060] Described herein are non-invasive methods for subtyping small cell lung cancer in a subject as low neuroendocrine or high neuroendocrine small cell lung cancer comprising: processing cfDNA fragments from a sample obtained from the subject and generating sequencing libraries; subjecting the sequencing libraries to whole genome sequencing to obtain sequenced fragments, wherein genome coverage is about $30\times$ to $0.1\times$; mapping the sequenced fragments to a genome to obtain genomic intervals of mapped sequences at specified transcription factor binding sites; analyzing the genomic intervals of mapped sequences to determine cfDNA fragment lengths and amounts to establish a cfDNA fragment coverage score at specified transcription factor binding sites using the cfDNA fragment lengths and amounts; and subtyping the small cell lung cancer in the subject based on transcription factor activation; wherein a decrease in an aggregate cfDNA fragment coverage scores at the specified transcription factor binding is indicative of a high-neuroendocrine small cell lung cancer subtype.

[0061] Described herein are non-invasive methods for subtyping small cell lung cancer in a subject as low neuroendocrine or high neuroendocrine small cell lung cancer comprising: processing cfDNA fragments from a sample obtained from the subject and generating

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

sequencing libraries; subjecting the sequencing libraries to whole genome sequencing to obtain sequenced fragments, wherein genome coverage is about $10\times$ to $0.1\times$; mapping the sequenced fragments to a genome to obtain genomic intervals of mapped sequences at specified transcription factor binding sites; analyzing the genomic intervals of mapped sequences to determine cfDNA fragment lengths and amounts to establish a cfDNA fragment coverage score at specified transcription factor binding sites using the cfDNA fragment lengths and amounts; and subtyping the small cell lung cancer in the subject based on transcription factor activation; wherein a decrease in an aggregate cfDNA fragment coverage scores at the specified transcription factor binding sites is indicative of a high-neuroendocrine small cell lung cancer subtype.

[0062] Described herein are non-invasive methods for subtyping small cell lung cancer in a subject as low neuroendocrine or high neuroendocrine small cell lung cancer comprising: processing cfDNA fragments from a sample obtained from the subject and generating sequencing libraries; subjecting the sequencing libraries to whole genome sequencing to obtain sequenced fragments, wherein genome coverage is about $9\times$ to $0.1\times$; mapping the sequenced fragments to a genome to obtain genomic intervals of mapped sequences at specified transcription factor binding sites; analyzing the genomic intervals of mapped sequences to determine cfDNA fragment lengths and amounts to establish a cfDNA fragment coverage score at specified transcription factor binding sites using the cfDNA fragment lengths and amounts; and subtyping the small cell lung cancer in the subject based on transcription factor activation; wherein a decrease in an aggregate cfDNA fragment coverage scores at the specified transcription factor binding sites is indicative of a high-neuroendocrine small cell lung cancer subtype.

[0063] In some aspects, the method uses the specified transcription factor is ASCL1, NEUROD1, POUF23, YAP1, or any combination thereof.

[0064] In some aspects, the method uses machine learning to subtype small cell lung cancer in the subject.

[0065] In some aspects, subtyping is performed by calculating a \log_2 (Read Depth Ratio) across 20bp windows starting 2kb from the specified transcription factor binding sites.

[0066] In some aspects, the \log_2 (Read Depth Ratio) is the total coverage over a coverage correction factor calculated using the median coverage of 5' and 3' 500bp anchors at the ends of the 2kb window (i.e. $\log_2(\text{Read Depth}/\text{Correction Factor})$).

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

[0067] In some aspects, the method uses all coverage calculations are shifted by 1 to avoid divisions by 0. In some aspects, this adjustment is expressed as $f(x) = \log_2((\text{window depth} + 1)/(\text{median}(\text{anchors}) + 1))$.

[0068] In some aspects, the \log_2 (Read Depth Ratio) is calculated for each transcription factor binding site and then aggregated by calculating the median of each window to obtain a fragment coverage score.

[0069] Some aspects further comprise the use of a general additive model to smooth coverage and correct for GC bias.

[0070] In some aspects, the genomic intervals are non-overlapping.

[0071] In some aspects, the genomic intervals each comprise thousands to millions of base pairs.

[0072] In some aspects, a cfDNA fragmentation profile is determined within each genomic intervals.

[0073] In some aspects, the cfDNA fragmentation profile comprises a median fragment size. In some aspects, the cfDNA fragmentation profile comprises a fragment size distribution.

[0074] Some aspects further comprise administering to the subject identified as having a high-neuroendocrine small cell lung cancer subtype, a therapeutic agent suitable for the treatment of the type of cancer.

[0075] In some aspects, the therapeutic agent is an immunotherapy.

[0076] Described herein are non-invasive methods for subtyping small cell lung cancer in a subject as low neuroendocrine or high neuroendocrine small cell lung cancer comprising: processing cfDNA fragments from a sample obtained from the subject and generating sequencing libraries; subjecting the sequencing libraries to whole genome sequencing to obtain sequenced fragments, wherein genome coverage is about $30\times$ to $0.1\times$; mapping the sequenced fragments to a genome to obtain genomic intervals of mapped sequences at specified transcription factor binding sites;

[0077] analyzing the genomic intervals of mapped sequences to determine cfDNA fragment lengths and amounts to establish a cfDNA fragment coverage score at specified transcription factor binding sites using the cfDNA fragment lengths and amounts; and subtyping the small cell lung cancer in the subject based on transcription factor activation; wherein a decrease in an aggregate cfDNA fragment coverage scores at the specified transcription factor binding sites is indicative of a high-neuroendocrine small cell lung cancer

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

subtype. In some aspects, the method uses the specified transcription factor is ASCL1, NEUROD1, POUF23, YAP1, or any combination thereof. In some aspects, the method uses machine learning to subtype small cell lung cancer in the subject. In some aspects, subtyping is performed by calculating a \log_2 (Read Depth Ratio) across 20bp windows starting 2kb from the specified transcription factor binding sites. In some aspects, the \log_2 (Read Depth Ratio) is the total coverage over a coverage correction factor calculated using the median coverage of 5' and 3' 500bp anchors at the ends of the 2kb window (i.e. $\log_2(\text{Read Depth}/\text{Correction Factor})$). In some aspects, the method uses all coverage calculations are shifted by 1 to avoid divisions by 0. In some aspects this adjustment is expressed as $f(x) = \log_2((\text{window depth} + 1)/(\text{median}(\text{anchors}) + 1))$. In some aspects, the \log_2 (Read Depth Ratio) is calculated for each transcription factor binding site and then aggregated by calculating the median of each window to obtain a fragment coverage score. Some aspects further comprise the use of a general additive model to smooth coverage and correct for GC bias. In some aspects, the genomic intervals are non-overlapping. In some aspects, the genomic intervals each comprise thousands to millions of base pairs. In some aspects, a cfDNA fragmentation profile is determined within each genomic intervals. In some aspects, the cfDNA fragmentation profile comprises a median fragment size. In some aspects, the cfDNA fragmentation profile comprises a fragment size distribution. Some aspects further comprise administering to the subject identified as having a high-neuroendocrine small cell lung cancer subtype, a therapeutic agent suitable for the treatment of the type of cancer. In some aspects, the therapeutic agent is an immunotherapy.

[0078] Described herein are non-invasive methods for subtyping small cell lung cancer in a subject as low neuroendocrine or high neuroendocrine small cell lung cancer comprising: processing cfDNA fragments from a sample obtained from the subject and generating sequencing libraries; subjecting the sequencing libraries to whole genome sequencing to obtain sequenced fragments, wherein genome coverage is about $10\times$ to $0.1\times$; mapping the sequenced fragments to a genome to obtain genomic intervals of mapped sequences at specified transcription factor binding sites;

[0079] analyzing the genomic intervals of mapped sequences to determine cfDNA fragment lengths and amounts to establish a cfDNA fragment coverage score at specified transcription factor binding sites using the cfDNA fragment lengths and amounts; and subtyping the small cell lung cancer in the subject based on transcription factor activation; wherein a decrease in an aggregate cfDNA fragment coverage scores at the specified

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

transcription factor binding sites indicative of a high-neuroendocrine small cell lung cancer subtype. In some aspects, the method uses the specified transcription factor is ASCL1, NEUROD1, POUF23, YAP1, or any combination thereof. In some aspects, the method uses machine learning to subtype small cell lung cancer in the subject. In some aspects, subtyping is performed by calculating a log₂ (Read Depth Ratio) across 20bp windows starting 2kb from the specified transcription factor binding sites. In some aspects, the log₂ (Read Depth Ratio) is the total coverage over a coverage correction factor calculated using the median coverage of 5' and 3' 500bp anchors at the ends of the 2kb window (i.e. log₂(Read Depth/Correction Factor)). In some aspects, the method uses all coverage calculations are shifted by 1 to avoid divisions by 0. In some aspects this adjustment is expressed as $f(x) = \log_2((\text{window depth} + 1)/(\text{median}(\text{anchors}) + 1))$. In some aspects, the log₂ (Read Depth Ratio) is calculated for each transcription factor binding site and then aggregated by calculating the median of each window to obtain a fragment coverage score. Some aspects further comprise the use of a general additive model to smooth coverage and correct for GC bias. In some aspects, the genomic intervals are non-overlapping. In some aspects, the genomic intervals each comprise thousands to millions of base pairs. In some aspects, a cfDNA fragmentation profile is determined within each genomic intervals. In some aspects, the cfDNA fragmentation profile comprises a median fragment size. In some aspects, the cfDNA fragmentation profile comprises a fragment size distribution. Some aspects further comprise administering to the subject identified as having a high-neuroendocrine small cell lung cancer subtype, a therapeutic agent suitable for the treatment of the type of cancer. In some aspects, the therapeutic agent is an immunotherapy.

[0080] Described herein are non-invasive methods for subtyping small cell lung cancer in a subject as low neuroendocrine or high neuroendocrine small cell lung cancer comprising: processing cfDNA fragments from a sample obtained from the subject and generating sequencing libraries; subjecting the sequencing libraries to whole genome sequencing to obtain sequenced fragments, wherein genome coverage is about 9× to 0.1×; mapping the sequenced fragments to a genome to obtain genomic intervals of mapped sequences at specified transcription factor binding sites;

[0081] analyzing the genomic intervals of mapped sequences to determine cfDNA fragment lengths and amounts to establish a cfDNA fragment coverage score at specified transcription factor binding sites using the cfDNA fragment lengths and amounts; and subtyping the small cell lung cancer in the subject based on transcription factor activation;

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

wherein a decrease in an aggregate cfDNA fragment coverage scores at the specified transcription factor binding sites is indicative of a high-neuroendocrine small cell lung cancer subtype. In some aspects, the method uses the specified transcription factor is ASCL1, NEUROD1, POUF23, YAP1, or any combination thereof. In some aspects, the method uses machine learning to subtype small cell lung cancer in the subject. In some aspects, subtyping is performed by calculating a \log_2 (Read Depth Ratio) across 20bp windows starting 2kb from the specified transcription factor binding sites. In some aspects, the \log_2 (Read Depth Ratio) is the total coverage over a coverage correction factor calculated using the median coverage of 5' and 3' 500bp anchors at the ends of the 2kb window (i.e. $\log_2(\text{Read Depth}/\text{Correction Factor})$). In some aspects, the method uses all coverage calculations are shifted by 1 to avoid divisions by 0. In some aspects this adjustment is expressed as $f(x) = \log_2((\text{window depth} + 1)/(\text{median}(\text{anchors}) + 1))$. In some aspects, the \log_2 (Read Depth Ratio) is calculated for each transcription factor binding site and then aggregated by calculating the median of each window to obtain a fragment coverage score. Some aspects further comprise the use of a general additive model to smooth coverage and correct for GC bias. In some aspects, the genomic intervals are non-overlapping. In some aspects, the genomic intervals each comprise thousands to millions of base pairs. In some aspects, a cfDNA fragmentation profile is determined within each genomic intervals. In some aspects, the cfDNA fragmentation profile comprises a median fragment size. In some aspects, the cfDNA fragmentation profile comprises a fragment size distribution. Some aspects further comprise administering to the subject identified as having a high-neuroendocrine small cell lung cancer subtype, a therapeutic agent suitable for the treatment of the type of cancer. In some aspects, the therapeutic agent is an immunotherapy.

[0082] In one embodiment, the present invention provides methods for the non-invasive subtyping of non-small cell lung cancer in a subject as adenocarcinoma or squamous carcinoma comprising: processing cfDNA fragments from a sample obtained from the subject and generating sequencing libraries; subjecting the sequencing libraries to whole genome sequencing to obtain sequenced fragments, wherein genome coverage is about $30\times$ to $0.1\times$; mapping the sequenced fragments to a genome to obtain genomic intervals of mapped sequences at specified transcription factor binding sites; analyzing the genomic intervals of mapped sequences to determine cfDNA fragment lengths and amounts to establish a cfDNA fragment coverage score at specified transcription factor binding sites using the cfDNA fragment lengths and amounts; and subtyping the small cell lung cancer in the subject based

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

on transcription factor activation; wherein a decrease in an aggregate cfDNA fragment coverage scores at the specified transcription factor binding sites is indicative of an adenocarcinoma or squamous carcinoma subtype.

[0083] In one embodiment, the present invention provides methods for the non-invasive subtyping of non-small cell lung cancer in a subject as adenocarcinoma or squamous carcinoma comprising: processing cfDNA fragments from a sample obtained from the subject and generating sequencing libraries; subjecting the sequencing libraries to whole genome sequencing to obtain sequenced fragments, wherein genome coverage is about $10\times$ to $0.1\times$; mapping the sequenced fragments to a genome to obtain genomic intervals of mapped sequences at specified transcription factor binding sites; analyzing the genomic intervals of mapped sequences to determine cfDNA fragment lengths and amounts to establish a cfDNA fragment coverage score at specified transcription factor binding sites using the cfDNA fragment lengths and amounts; and subtyping the small cell lung cancer in the subject based on transcription factor activation; wherein a decrease in an aggregate cfDNA fragment coverage scores at the specified transcription factor binding sites is indicative of an adenocarcinoma or squamous carcinoma subtype.

[0084] In one embodiment, the present invention provides methods for the non-invasive subtyping of non-small cell lung cancer in a subject as adenocarcinoma or squamous carcinoma comprising: processing cfDNA fragments from a sample obtained from the subject and generating sequencing libraries; subjecting the sequencing libraries to whole genome sequencing to obtain sequenced fragments, wherein genome coverage is about $9\times$ to $0.1\times$; mapping the sequenced fragments to a genome to obtain genomic intervals of mapped sequences at specified transcription factor binding sites; analyzing the genomic intervals of mapped sequences to determine cfDNA fragment lengths and amounts to establish a cfDNA fragment coverage score at specified transcription factor binding sites using the cfDNA fragment lengths and amounts; and subtyping the small cell lung cancer in the subject based on transcription factor activation; wherein a decrease in an aggregate cfDNA fragment coverage scores at the specified transcription factor binding sites is indicative of an adenocarcinoma or squamous carcinoma subtype.

[0085] In some aspects, the specified transcription factor is ASCL1, NEUROD1, POUF23, YAP1, or any combination thereof.

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

[0086] In some aspects, the method uses machine learning to subtype small cell lung cancer in the subject.

[0087] In some aspects, subtyping is performed by calculating a log₂ (Read Depth Ratio) across 20bp windows starting 2kb from the specified transcription factor binding sites.

[0088] In some aspects, the log₂ (Read Depth Ratio) is the total coverage over a coverage correction factor calculated using the median coverage of 5' and 3' 500bp anchors at the ends of the 2kb window (i.e. log₂(Read Depth/Correction Factor)).

[0089] In some aspects, all coverage calculations are shifted by 1 to avoid divisions by 0. In some aspects this adjustment is expressed as $f(x) = \log_2((\text{window depth} + 1)/(\text{median}(\text{anchors}) + 1))$.

[0090] In some aspects, the log₂ (Read Depth Ratio) is calculated for each transcription factor binding site and then aggregated by calculating the median of each window to obtain a fragment coverage score.

[0091] Some aspects further comprise the use of a general additive model to smooth coverage and correct for GC bias.

[0092] In some aspects, the genomic intervals are non-overlapping.

[0093] In some aspects, the genomic intervals each comprise thousands to millions of base pairs.

[0094] In some aspects, a cfDNA fragmentation profile is determined within each genomic intervals.

[0095] In some aspects, the cfDNA fragmentation profile comprises a median fragment size.

[0096] In some aspects, the cfDNA fragmentation profile comprises a fragment size distribution.

[0097] Some aspects further comprise administering to the subject identified as having an adenocarcinoma or squamous carcinoma subtype, a therapeutic agent suitable for the treatment of the type of cancer.

[0098] In some aspects, the therapeutic agent is an immunotherapy.

[0099] Some aspects further comprise quantifying a circulating tumor fraction, mirroring major allele frequency (MAF) performance in relation to Response Evaluation Criteria in Solid Tumors (RECIST) evaluation in both Adenocarcinoma and Squamous carcinoma.

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

[0100] In one embodiment, the present invention provides methods for the non-invasive subtyping of non-small cell lung cancer in a subject as adenocarcinoma or squamous carcinoma comprising: processing cfDNA fragments from a sample obtained from the subject and generating sequencing libraries; subjecting the sequencing libraries to whole genome sequencing to obtain sequenced fragments, wherein genome coverage is about 30× to 0.1×; mapping the sequenced fragments to a genome to obtain genomic intervals of mapped sequences at specified transcription factor binding sites; analyzing the genomic intervals of mapped sequences to determine cfDNA fragment lengths and amounts to establish a cfDNA fragment coverage score at specified transcription factor binding sites using the cfDNA fragment lengths and amounts; and subtyping the small cell lung cancer in the subject based on transcription factor activation; wherein a decrease in an aggregate cfDNA fragment coverage scores at the specified transcription factor binding sites is indicative of an adenocarcinoma or squamous carcinoma subtype. In some aspects, the specified transcription factor is ASCL1, NEUROD1, POUF23, YAP1, or any combination thereof. In some aspects, the method uses machine learning to subtype small cell lung cancer in the subject. In some aspects, subtyping is performed by calculating a \log_2 (Read Depth Ratio) across 20bp windows starting 2kb from the specified transcription factor binding sites. In some aspects, the \log_2 (Read Depth Ratio) is the total coverage over a coverage correction factor calculated using the median coverage of 5' and 3' 500bp anchors at the ends of the 2kb window (i.e. $\log_2(\text{Read Depth}/\text{Correction Factor})$). In some aspects, all coverage calculations are shifted by 1 to avoid divisions by 0. In some aspects this adjustment is expressed as $f(x) = \log_2((\text{window depth} + 1)/(\text{median}(\text{anchors}) + 1))$. In some aspects, the \log_2 (Read Depth Ratio) is calculated for each transcription factor binding site and then aggregated by calculating the median of each window to obtain a fragment coverage score. Some aspects further comprise the use of a general additive model to smooth coverage and correct for GC bias. In some aspects, the genomic intervals are non-overlapping. In some aspects, the genomic intervals each comprise thousands to millions of base pairs. In some aspects, a cfDNA fragmentation profile is determined within each genomic intervals. In some aspects, the cfDNA fragmentation profile comprises a median fragment size. In some aspects, the cfDNA fragmentation profile comprises a fragment size distribution. Some aspects further comprise administering to the subject identified as having an adenocarcinoma or squamous carcinoma subtype, a therapeutic agent suitable for the treatment of the type of cancer. In some aspects, the therapeutic agent is an immunotherapy.

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

Some aspects further comprise quantifying a circulating tumor fraction, mirroring major allele frequency (MAF) performance in relation to Response Evaluation Criteria in Solid Tumors (RECIST) evaluation in both Adenocarcinoma and Squamous carcinoma.

[0101] In one embodiment, the present invention provides methods for the non-invasive subtyping of non-small cell lung cancer in a subject as adenocarcinoma or squamous carcinoma comprising: processing cfDNA fragments from a sample obtained from the subject and generating sequencing libraries; subjecting the sequencing libraries to whole genome sequencing to obtain sequenced fragments, wherein genome coverage is about $10\times$ to $0.1\times$; mapping the sequenced fragments to a genome to obtain genomic intervals of mapped sequences at specified transcription factor binding sites; analyzing the genomic intervals of mapped sequences to determine cfDNA fragment lengths and amounts to establish a cfDNA fragment coverage score at specified transcription factor binding sites using the cfDNA fragment lengths and amounts; and subtyping the small cell lung cancer in the subject based on transcription factor activation; wherein a decrease in an aggregate cfDNA fragment coverage scores at the specified transcription factor binding sites is indicative of an adenocarcinoma or squamous carcinoma subtype. In some aspects, the specified transcription factor is ASCL1, NEUROD1, POUF23, YAP1, or any combination thereof. In some aspects, the method uses machine learning to subtype small cell lung cancer in the subject. In some aspects, subtyping is performed by calculating a \log_2 (Read Depth Ratio) across 20bp windows starting 2kb from the specified transcription factor binding sites. In some aspects, the \log_2 (Read Depth Ratio) is the total coverage over a coverage correction factor calculated using the median coverage of 5' and 3' 500bp anchors at the ends of the 2kb window (i.e. $\log_2(\text{Read Depth}/\text{Correction Factor})$). In some aspects, all coverage calculations are shifted by 1 to avoid divisions by 0. In some aspects this adjustment is expressed as $f(x) = \log_2((\text{window depth} + 1)/(\text{median}(\text{anchors}) + 1))$. In some aspects, the \log_2 (Read Depth Ratio) is calculated for each transcription factor binding site and then aggregated by calculating the median of each window to obtain a fragment coverage score. Some aspects further comprise the use of a general additive model to smooth coverage and correct for GC bias. In some aspects, the genomic intervals are non-overlapping. In some aspects, the genomic intervals each comprise thousands to millions of base pairs. In some aspects, a cfDNA fragmentation profile is determined within each genomic intervals. In some aspects, the cfDNA fragmentation profile comprises a median fragment size. In some aspects, the cfDNA fragmentation profile comprises a fragment size

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

distribution. Some aspects further comprise administering to the subject identified as having an adenocarcinoma or squamous carcinoma subtype, a therapeutic agent suitable for the treatment of the type of cancer. In some aspects, the therapeutic agent is an immunotherapy. Some aspects further comprise quantifying a circulating tumor fraction, mirroring major allele frequency (MAF) performance in relation to Response Evaluation Criteria in Solid Tumors (RECIST) evaluation in both Adenocarcinoma and Squamous carcinoma.

[0102] In one embodiment, the present invention provides methods for the non-invasive subtyping of non-small cell lung cancer in a subject as adenocarcinoma or squamous carcinoma comprising: processing cfDNA fragments from a sample obtained from the subject and generating sequencing libraries; subjecting the sequencing libraries to whole genome sequencing to obtain sequenced fragments, wherein genome coverage is about $9\times$ to $0.1\times$; mapping the sequenced fragments to a genome to obtain genomic intervals of mapped sequences at specified transcription factor binding sites; analyzing the genomic intervals of mapped sequences to determine cfDNA fragment lengths and amounts to establish a cfDNA fragment coverage score at specified transcription factor binding sites using the cfDNA fragment lengths and amounts; and subtyping the small cell lung cancer in the subject based on transcription factor activation; wherein a decrease in an aggregate cfDNA fragment coverage scores at the specified transcription factor binding sites is indicative of an adenocarcinoma or squamous carcinoma subtype. In some aspects, the specified transcription factor is ASCL1, NEUROD1, POUF23, YAP1, or any combination thereof. In some aspects, the method uses machine learning to subtype small cell lung cancer in the subject. In some aspects, subtyping is performed by calculating a \log_2 (Read Depth Ratio) across 20bp windows starting 2kb from the specified transcription factor binding sites. In some aspects, the \log_2 (Read Depth Ratio) is the total coverage over a coverage correction factor calculated using the median coverage of 5' and 3' 500bp anchors at the ends of the 2kb window (i.e. $\log_2(\text{Read Depth}/\text{Correction Factor})$). In some aspects, all coverage calculations are shifted by 1 to avoid divisions by 0. In some aspects this adjustment is expressed as $f(x) = \log_2((\text{window depth} + 1)/(\text{median}(\text{anchors}) + 1))$. In some aspects, the \log_2 (Read Depth Ratio) is calculated for each transcription factor binding site and then aggregated by calculating the median of each window to obtain a fragment coverage score. Some aspects further comprise the use of a general additive model to smooth coverage and correct for GC bias. In some aspects, the genomic intervals are non-overlapping. In some aspects, the genomic intervals each comprise thousands

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

to millions of base pairs. In some aspects, a cfDNA fragmentation profile is determined within each genomic intervals. In some aspects, the cfDNA fragmentation profile comprises a median fragment size. In some aspects, the cfDNA fragmentation profile comprises a fragment size distribution. Some aspects further comprise administering to the subject identified as having an adenocarcinoma or squamous carcinoma subtype, a therapeutic agent suitable for the treatment of the type of cancer. In some aspects, the therapeutic agent is an immunotherapy. Some aspects further comprise quantifying a circulating tumor fraction, mirroring major allele frequency (MAF) performance in relation to Response Evaluation Criteria in Solid Tumors (RECIST) evaluation in both Adenocarcinoma and Squamous carcinoma.

[0103] Before the present compositions and methods are described, it is to be understood that this invention is not limited to the particular methods and systems described, as such methods and systems may vary. It is also to be understood that the terminology used herein is for purposes of describing particular embodiments only, and is not intended to be limiting, since the scope of the present invention will be limited only in the appended claims.

[0104] As used in this specification and the appended claims, the singular forms “a”, “an”, and “the” include plural references unless the context clearly dictates otherwise. Thus, for example, references to “the method” includes one or more methods, and/or steps of the type described herein which will become apparent to those persons skilled in the art upon reading this disclosure and so forth.

[0105] Unless defined otherwise, 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 belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the invention, the preferred methods and materials are now described.

[0106] Described herein are methods to subtype non-small cell lung cancer (NSCLC) in a non-invasive manner using cell-free DNA (cfDNA) fragmentomes. NSCLC is often presented in different subtypes such as adenocarcinoma and squamous cell carcinoma. This novel non-invasive DELFI-based approach distinguishes NSCLC adenocarcinoma versus Squamous carcinoma subtypes investigating cell-free fragments that reflect the unique epigenetic states of NSCLCs. Other cfDNA-based liquid biopsies are not known to differentiate NSCLC subtypes without having access to tissue clinical data.

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

[0107] NSCLC subtypes can be predicted from the DELFI-based approach using cfDNA LC-WGS data. First, the DELFI machine learning classifier detects the presence of cancer in patients with NSCLC. Both genome-wide fragmentation profiles and correspondent DELFI-Tumor Fraction (TF) scores are investigated to identify preliminary differences between the two main clinical subtypes of NSCLC cases: adenocarcinoma and squamous cell carcinoma. DELFI-TF accurately detects circulating tumor fraction without being confused by clonal hematopoiesis. Furthermore, DELFI-TF accurately quantifies circulating tumor fraction, mirroring MAF performance in relation to RECIST evaluation in both Adenocarcinoma and Squamous carcinoma. Publicly available data was used for running a principal component analysis to cluster the two subtypes using Whole Genome fragmentation data. **Figure 19** depicts proof-of-concept DELFI-TF model development. Second, a list of differentially accessible transcription start sites (TSS) is applied and is capable of differentiating between adenocarcinoma versus squamous. A list of differentially accessible transcription start sites can be obtained from public databases such as, but not limited to UCSC or Ensembl. The creation of such lists is similar to what would be done for RNA-Seq experiments except instead of transcript per million the read depth ratio is used as a proxy for expression. In some aspects, sites with the largest and most significant log fold changes from 20% of a test cohort are selected and applied to the remainder of the cohort at the same loci followed by performed hierarchical clustering to determine if subtypes remained together. In another aspect, the most highly expressed transcripts from TCGA for a given cancer type are selected and those transcripts which are also expressed at any level in AML (blood cancer as an imperfect proxy for normal blood) are filtered out from the list. Third, if the most differential DELFI-TSSs exhibit short/long changes was investigated, further confirming these cfDNA molecules are tumor-derived. Finally, a receiver operating characteristic (ROC) representing sensitivity and specificity of the DELFI-fragmentome approach to identify NSCLC subtypes is used to identify the diagnostic accuracy of our approach between the identified cluster samples. Given the logistical difficulties to perform NSCLC tumor biopsies, we believe this approach could be a viable and quick method of subtyping NSCLC in a non-invasive manner.

[0108] 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

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

genome” are equivalent and can be used interchangeably. 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., in non-overlapping windows) and assessed to determine a cfDNA fragmentation profile. As 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 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) 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 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.

[0109] A cfDNA fragmentation profile can include one or more cfDNA fragmentation patterns. A cfDNA fragmentation pattern can include any appropriate cfDNA fragmentation 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 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,

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

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 described herein include, without limitation, a portion of a chromosome (e.g., a portion of 2q, 4p, 5p, 6q, 7p, 8q, 9q, 10q, 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.

[0110] In some cases, a cfDNA fragmentation profile can be used to identify changes (e.g., 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 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 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).

[0111] 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 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 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

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

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 healthy mammal is determined over a subgenomic interval.

[0112] 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).

[0113] A cfDNA fragmentation profile can include a cfDNA fragment size pattern. cfDNA 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 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).

[0114] 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 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) 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

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

size distribution of, on average, a 10, 11, 12, 13, 14, 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 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, portions of 2q, 4p, 5p, 6q, 7p, 8q, 9q, 10q, 11q, 12q, and 14q. For example, an alteration can be across a chromosome arm (e.g., an entire chromosome arm).

[0115] 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 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 lower, 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 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.

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

[0116] 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 length).

[0117] A cfDNA fragmentation profile can be obtained using any appropriate method. In 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 example, 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 cfDNA fragmentation profile can be obtained as shown in FIG. 1.

[0118] 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).

[0119] 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.

[0120] 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

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

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

[0121] 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 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, 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 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 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 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).

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

[0122] 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 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.

[0123] 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., 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 milk, and exhaled breath condensate. For example, a plasma sample can be assessed to determine a cfDNA fragmentation profiled as described herein.

[0124] 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 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 Transl Med 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).

[0125] In some cases, a sample can be processed (e.g., to isolate and/or purify DNA and/or 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).

[0126] Additional methods are described in U.S. Patents No. 10,982,279 and No. 10,975,431, the disclosure of which is considered part of and is herein incorporated by reference in the disclosure of this application in its entirety.

Example Hardware Implementation

[0127] **Figure 20** illustrates an example computer 800 that may be used to implement the methods described herein. For example, the computer 800 may include a machine learning system that trains a machine learning model to subtype small cell lung cancer or non-small cell lung cancer as described above or a portion or combination thereof in some embodiments. The computer 800 may be any electronic device that runs software applications derived from compiled instructions, including without limitation personal computers, servers, smart phones, media players, electronic tablets, game consoles, email devices, etc. In some implementations, the computer 800 may include one or more processors 802, one or more input devices 804, one or more display devices 806, one or more network interfaces 808, and one or more computer-readable mediums 812. Each of these components may be coupled by bus 810, and in some embodiments, these components may be distributed among multiple physical locations and coupled by a network.

[0128] Display device 806 may be any known display technology, including but not limited to display devices using Liquid Crystal Display (LCD) or Light Emitting Diode (LED) technology. Processor(s) 802 may use any known processor technology, including but not limited to graphics processors and multi-core processors. Input device 804 may be any known input device technology, including but not limited to a keyboard (including a virtual keyboard), mouse, track ball, camera, and touch-sensitive pad or display. Bus 810 may be any known internal or external bus technology, including but not limited to ISA, EISA, PCI, PCI Express, USB, Serial ATA or FireWire. Computer-readable medium 812 may be any non-transitory medium that participates in providing instructions to processor(s) 804 for execution, including without limitation, non-volatile storage media (e.g., optical disks, magnetic disks, flash drives, etc.), or volatile media (e.g., SDRAM, ROM, etc.).

[0129] Computer-readable medium 812 may include various instructions 814 for implementing an operating system (e.g., Mac OS®, Windows®, Linux). The operating system may be multi-user, multiprocessing, multitasking, multithreading, real-time, and the like. The operating system may perform basic tasks, including but not limited to: recognizing input from input device 804; sending output to display device 806; keeping track of files and directories on computer-readable medium 812; controlling peripheral devices (e.g., disk drives, printers, etc.) which can be controlled directly or through an I/O controller; and managing traffic on bus 810. Network communications instructions 816 may establish and maintain network

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

connections (e.g., software for implementing communication protocols, such as TCP/IP, HTTP, Ethernet, telephony, etc.).

[0130] Machine learning instructions 818 may include instructions that enable computer 800 to function as a machine learning system and/or to training machine learning models to generate DMS values as described herein. Application(s) 820 may be an application that uses or implements the processes described herein and/or other processes. The processes may also be implemented in operating system 814. For example, application 820 and/or operating system may create tasks in applications as described herein.

[0131] The described features may be implemented in one or more computer programs that may be executable on a programmable system including at least one programmable processor coupled to receive data and instructions from, and to transmit data and instructions to, a data storage system, at least one input device, and at least one output device. A computer program is a set of instructions that can be used, directly or indirectly, in a computer to perform a certain activity or bring about a certain result. A computer program may be written in any form of programming language (e.g., Objective-C, Java), including compiled or interpreted languages, and it may be deployed in any form, including as a stand-alone program or as a module, component, subroutine, or other unit suitable for use in a computing environment.

[0132] Suitable processors for the execution of a program of instructions may include, by way of example, both general and special purpose microprocessors, and the sole processor or one of multiple processors or cores, of any kind of computer. Generally, a processor may receive instructions and data from a read-only memory or a random-access memory or both. The essential elements of a computer may include a processor for executing instructions and one or more memories for storing instructions and data. Generally, a computer may also include, or be operatively coupled to communicate with, one or more mass storage devices for storing data files; such devices include magnetic disks, such as internal hard disks and removable disks; magneto-optical disks; and optical disks. Storage devices suitable for tangibly embodying computer program instructions and data may include all forms of non-volatile memory, including by way of example semiconductor memory devices, such as EPROM, EEPROM, and flash memory devices; magnetic disks such as internal hard disks and removable disks; magneto-optical disks; and CD-ROM and DVD-ROM disks. The processor and the memory may be supplemented by, or incorporated in, ASICs (application-specific integrated circuits).

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

[0133] To provide for interaction with a user, the features may be implemented on a computer having a display device such as an LED or LCD monitor for displaying information to the user and a keyboard and a pointing device such as a mouse or a trackball by which the user can provide input to the computer.

[0134] The features may be implemented in a computer system that includes a back-end component, such as a data server, or that includes a middleware component, such as an application server or an Internet server, or that includes a front-end component, such as a client computer having a graphical user interface or an Internet browser, or any combination thereof. The components of the system may be connected by any form or medium of digital data communication such as a communication network. Examples of communication networks include, e.g., a telephone network, a LAN, a WAN, and the computers and networks forming the Internet.

[0135] The computer system may include clients and servers. A client and server may generally be remote from each other and may typically interact through a network. The relationship of client and server may arise by virtue of computer programs running on the respective computers and having a client-server relationship to each other.

[0136] One or more features or steps of the disclosed embodiments may be implemented using an Application Programming Interface (API). An API may define one or more parameters that are passed between a calling application and other software code (e.g., an operating system, library routine, function) that provides a service, that provides data, or that performs an operation or a computation.

[0137] The API may be implemented as one or more calls in program code that send or receive one or more parameters through a parameter list or other structure based on a call convention defined in an API specification document. A parameter may be a constant, a key, a data structure, an object, an object class, a variable, a data type, a pointer, an array, a list, or another call. API calls and parameters may be implemented in any programming language. The programming language may define the vocabulary and calling convention that a programmer will employ to access functions supporting the API.

[0138] In some implementations, an API call may report to an application the capabilities of a device running the application, such as input capability, output capability, processing capability, power capability, communications capability, etc.

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

[0139] While various embodiments have been described above, it should be understood that they have been presented by way of example and not limitation. It will be apparent to persons skilled in the relevant art(s) that various changes in form and detail can be made therein without departing from the spirit and scope. In fact, after reading the above description, it will be apparent to one skilled in the relevant art(s) how to implement alternative embodiments. For example, other steps may be provided, or steps may be eliminated, from the described flows, and other components may be added to, or removed from, the described systems. Accordingly, other implementations are within the scope of the following claims.

[0140] In addition, it should be understood that any figures which highlight the functionality and advantages are presented for example purposes only. The disclosed methodology and system are each sufficiently flexible and configurable such that they may be utilized in ways other than that shown.

[0141] Although the term “at least one” may often be used in the specification, claims and drawings, the terms “a”, “an”, “the”, “said”, etc. also signify “at least one” or “the at least one” in the specification, claims and drawings.

[0142] Finally, it is the applicant's intent that only claims that include the express language "means for" or "step for" be interpreted under 35 U.S.C. 112(f). Claims that do not expressly include the phrase "means for" or "step for" are not to be interpreted under 35 U.S.C. 112(f).

[0143] The presently described methods and systems are useful for subtyping non-small cell lung cancer or small cell lung cancer in a subject and optionally treating the cancer subtype in the subject. Any appropriate subject, such as a mammal can be assessed, and/or treated as described herein. Examples of some mammals that can be assessed, and/or treated as 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 using a method described herein and, optionally, can be treated with one or more cancer treatments as described herein. The methods disclosed herein may include administering to the subject identified as having the type of cancer, a therapeutic agent suitable for the treatment of the type of cancer.

[0144] When treating a subject having, or suspected of having, cancer as described herein, the subject 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 subject at any appropriate frequency (e.g., once or multiple times over a

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

period of time ranging from days to weeks). Examples of cancer treatments include, without limitation, surgical intervention, adjuvant chemotherapy, neoadjuvant chemotherapy, radiation therapy, hormone 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., BiTEs), monoclonal antibodies, immune checkpoint inhibitors, surgery (e.g., surgical resection), or any combination of the above. In some aspects, 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 subject.

[0145] In some aspects, 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, etoposide, fludarabine, floxuridine, fludarabine, fluorouracil, gemcitabine, hydroxyurea, idarubicin, ifosfamide, irinotecan, lomustine, mechlorethamine, melphalan, mercaptopurine, methotrexate, mitomycin, mitoxantrone, oxaliplatin, paclitaxel, pemetrexed, procarbazine, all-trans retinoic acid, streptozocin, tafluposide, temozolomide, teniposide, tioguanine, topotecan, uramustine, valrubicin, vinblastine, vincristine, vindesine, vinorelbine, and 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).

[0146] In various aspects, DNA is present in a biological sample taken from a subject and used in the methodology of the invention. The biological sample can be virtually any type of biological sample that includes DNA. The biological sample is typically a fluid, such as whole blood or a portion thereof with circulating cfDNA. In embodiments, the sample includes DNA from a tumor or a liquid biopsy, such as, but not limited to amniotic fluid, aqueous humor, vitreous humor, blood, whole blood, fractionated blood, plasma, serum, breast milk, cerebrospinal fluid (CSF), cerumen (earwax), chyle, chime, endolymph, perilymph, feces, breath, gastric acid, gastric juice, lymph, mucus (including nasal drainage and phlegm),

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

pericardial fluid, peritoneal fluid, pleural fluid, pus, rheum, saliva, exhaled breath condensates, sebum, semen, sputum, sweat, synovial fluid, tears, vomit, prostatic fluid, nipple aspirate fluid, lachrymal fluid, perspiration, cheek swabs, cell lysate, gastrointestinal fluid, biopsy tissue and urine or other biological fluid. In one aspect, the sample includes DNA from a circulating tumor cell.

[0147] As disclosed above, the biological sample can be a blood sample. The blood sample can be obtained using methods known in the art, such as finger prick or phlebotomy. Suitably, the blood sample is approximately 0.1 to 20 ml, or alternatively approximately 1 to 15 ml with the volume of blood being approximately 10 ml. Smaller amounts may also be used, as well as circulating free DNA in blood. Microsampling and sampling by needle biopsy, catheter, excretion or production of bodily fluids containing DNA are also potential biological sample sources.

[0148] The methods and systems of the disclosure utilize nucleic acid sequence information and can therefore include any method or sequencing device for performing nucleic acid sequencing including nucleic acid amplification, polymerase chain reaction (PCR), nanopore sequencing, 454 sequencing, insertion tagged sequencing. In some aspects, the methodology or systems of the disclosure utilize systems such as those provided by Illumina, Inc. (including but not limited to HiSeq™ X10, HiSeq™ 1000, HiSeq™ 2000, HiSeq™ 2500, Genome Analyzers™, MiSeq™, NextSeq, NovaSeq 6000 systems), Applied Biosystems Life Technologies (SOLiD™ System, Ion PGM™ Sequencer, ion Proton™ Sequencer) or Genapsys or BGI MGI and other systems. Nucleic acid analysis can also be carried out by systems provided by Oxford Nanopore Technologies (GridiON™, MiniON™) or Pacific Biosciences (Pacbio™ RS II or Sequel I or II).

[0149] The present invention includes systems for performing steps of the disclosed methods and is described partly in terms of functional components and various processing steps. Such functional components and processing steps may be realized by any number of components, operations and techniques configured to perform the specified functions and achieve the various results. For example, the present invention may employ various biological samples, biomarkers, elements, materials, computers, data sources, storage systems and media, information gathering techniques and processes, data processing criteria, statistical analyses, regression analyses and the like, which may carry out a variety of functions.

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

[0150] Accordingly, the invention further provides a non-invasive system for subtyping small cell lung cancer or non-small cell lung cancer. In various aspects, the system includes: (a) a sequencer configured to generate a low-coverage whole genome sequencing data set for a sample; and (b) a computer system and/or processor with functionality to perform a method of the invention.

[0151] In some aspects, the computer system further includes one or more additional modules. For example, the system may include one or more of an extraction and/or isolation unit operable to select suitable genetic components analysis, e.g., cfDNA fragments of a particular size.

[0152] In some aspects, the computer system further includes a visual display device. The visual display device may be operable to display a curve fit line, a reference curve fit line, and/or a comparison of both.

[0153] Methods for the non-invasive subtyping of small cell lung cancer or non-small cell lung cancer according to various aspects of the present invention may be implemented in any suitable manner, for example using a computer program operating on the computer system. As discussed herein, an exemplary system, according to various aspects of the present invention, may be implemented in conjunction with a computer system, for example a conventional computer system comprising a processor and a random access memory, such as a remotely-accessible application server, network server, personal computer or workstation. The computer system also suitably includes additional memory devices or information storage systems, such as a mass storage system and a user interface, for example a conventional monitor, keyboard and tracking device. The computer system may, however, include any suitable computer system and associated equipment and may be configured in any suitable manner. In one embodiment, the computer system comprises a stand-alone system. In another embodiment, the computer system is part of a network of computers including a server and a database.

[0154] The software required for receiving, processing, and analyzing information may be implemented in a single device or implemented in a plurality of devices. The software may be accessible via a network such that storage and processing of information takes place remotely with respect to users. The system according to various aspects of the present invention and its various elements provide functions and operations to facilitate detection and/or analysis, such as data gathering, processing, analysis, reporting and/or diagnosis. For example, in the present aspect, the computer system executes the computer program, which may receive, store, search,

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

analyze, and report information relating to the human genome or region thereof. The computer program may comprise multiple modules performing various functions or operations, such as a processing module for processing raw data and generating supplemental data and an analysis module for analyzing raw data and supplemental data to generate quantitative assessments of a disease status model and/or diagnosis information.

[0155] The procedures performed by the system may comprise any suitable processes to facilitate analysis and/or subtyping of small cell lung cancer or non-small cell lung cancer. In one embodiment, the system is configured to establish a disease subtype model and/or determine disease subtype in a patient. Determining or identifying disease subtype may include generating any useful information regarding the condition of the patient relative to the disease, such as performing a diagnosis, providing information helpful to a diagnosis, assessing the stage or progress of a disease, identifying a condition that may indicate a susceptibility to the disease, identify whether further tests may be recommended, predicting and/or assessing the efficacy of one or more treatment programs, or otherwise assessing the disease status, likelihood of disease, or other health aspect of the patient.

[0156] The following examples are provided to further illustrate the advantages and features of the present invention, but it is not intended to limit the scope of the invention. While this example is typical of those that might be used, other procedures, methodologies, or techniques known to those skilled in the art may alternatively be used.

EXAMPLE 1

[0157] Dissecting Small Cell Lung Cancer Subtypes by cell-free DNA fragmentomes

[0158] BACKGROUND: Small Cell Lung Cancer (SCLC) is an aggressive form of lung cancer, strongly associated with smoking and exposure to other environmental chemicals. Patients with SCLC suffer from a 5-year survival rate of less than 8% (Gay, C. M. *et al.* Patterns of transcription factor programs and immune pathway activation define four major subtypes of SCLC with distinct therapeutic vulnerabilities. *Cancer Cell* **39**, 346-360.e7 (2021)). SCLC is a heterogeneous tumor type consisting of tumor cells with neuroendocrine and non-neuroendocrine features. SCLC has two major subtypes: High-grade Neuroendocrine (High-NE) versus Low-grade Neuroendocrine (Low-NE) (Gay *et al.*). High-NE subtypes are characterized by the activation of lineage-specific transcription factors: *ASCL1* and *NEUROD1*. Low-NE is characterized by non-neuroendocrine factors such as *POU2F3* (SCLC-P) or the high presence of inflammatory T-cells (SCLC-I). Despite these molecular and clinical

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

heterogeneities, SCLC is treated as a single entity with predictably poor results. Recent results showed that the inflamed group (SCLC-I) has distinct biology and responds better to immunotherapy compared to the other SCLC subtypes (Gay *et al.* and Lissa, D. *et al.* Heterogeneity of neuroendocrine transcriptional states in metastatic small cell lung cancers and patient-derived models. *Nat Commun* **13**, 2023 (2022)). DELFI-based cell-free fragmentomes accurately differentiate High-NE vs Low-NE SCLC subtypes in a non-invasive manner. DELFI SCLC subtyping final goal is to guide optimal treatment selection for each patient diagnosed with advanced SCLC.

[0159] METHODS: Circulating cell-free DNA (cfDNA) was isolated from plasma samples of patients diagnosed with relapsed SCLC and treated with durvalumab plus olaparib in a phase II trial (NCT02484404). Pre-treatment tissue biopsies immunohistochemistry and genomics data were used to classify SCLC subtypes into High-NE (n=10) vs Low-NE (n=5). To infer tumor gene expression profiles in cfDNA, we investigated genome-wide signals of tissue-specific transcription factors differentially regulated in SCLC and applied a novel DELFI-based approach to inform SCLC molecular subtypes. In detail, DELFI calculates fragment distribution for each TFBS within a 2kb window and scales between 0-1 independently for each sample. A principal component analysis was then performed in R and clusters were defined based on ASCL1 TFBS sites (Mathios, D. *et al.* Detection and characterization of lung cancer using cell-free DNA fragmentomes. *Nat Commun* **12**, 5060 (2021)). Clinical information was examined in orthogonal analyses.

[0160] RESULTS: The DELFI's Proprietary Fragmentomics Platform detects SCLC cases (N=47) with high sensitivity, with a median DELFI score of 1.0 (95% CI 0.99-1), matching previous data from internal DELFI SCLC samples. Genome-wide cfDNA fragmentation analyses at ASCL1 binding sites (~12,000 genomic coordinates) in the SCLC patients reveal a decrease in coverage near transcription factor binding sites of SCLC patients compared to non-cancer and NSCLC samples. The difference at the center of the ASCL1 binding sites was the main driver of the differentiation between also pre-treatment SCLC samples. Investigation of fragments distribution at ASCL1 binding sites was capable of differentiating High-NE vs Low-NE samples. Patients (treated with I/O) classified as LOW-NE by the DELFI subtyping classifier had a better treatment response than DELFI-classified HIGH-NE. Differential pseudogene expression identified thousands of highly enriched Transcription Start Sites (TSS) in LOW-NE cases compared to HIGH-NE. LOW-NE samples showed a strong enrichment for

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

the T-lymphocytes specific factor KLF4, mirroring the neutrophils to lymphocytes ratio of this SCLC subtype. These data reflect the inflamed phenotype, possibly explaining why these cancer subtypes are more likely to respond to I/O treatment.

[0161] Table 1. Patient characteristics and associated clinical information

Characteristic	Validation set, n=20
Age, median y (range)	63 (41, 75)
Sex, n (%)	
Male	9 (45)
Female	11 (55)
TIL, n (%)	
Desert	2 (10)
Excluded	9 (45)
Inflamed	3 (15)
No performed	6 (30)
Metastasis, n (%)	
0	7 (35)
1	13 (65)
Unknown	0 (0)
PFS	
No progression or death, n (%)	0 (0)
Progression or death, n (%)	20 (100)
Median PFS, months (range)	16.7 (7.4, 46.2)
Median treatment duration, days (range)	55 (20, 494)

EXAMPLE 2

[0162] Use of DELFI circulating cfDNA-based approach to subtype small cell lung cancer (SCLC) in a non-invasive manner using cell-free DNA fragmentomes.

[0163] SCLC is an aggressive malignancy with a poor prognosis. Although SCLCs are clinically managed as a single cancer type, new evidence supports that these subtypes (high- vs low-neuroendocrine) of SCLC acquire diverse transcriptional and epigenetic states. Furthermore, distinct SCLC subtypes respond to specific treatment such as immunotherapy in the case of the low-neuroendocrine highly inflamed SCLC subtype.

[0164] The novel non-invasive DELFI-based approach described herein distinguishes between high- and low-neuroendocrine SCLC subtypes investigating cell-free fragments that

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

reflect the unique epigenetic states of SCLCs. This novel method is capable of differentiating the patients that will likely respond to immunotherapy, such as immune checkpoint blockade.

[0165] Using WGS data from plasma-derived cfDNA, fragment coverage at specific genomic coordinates was investigated to reveal the specific subtype of SCLC cases. The DELFI-based targeted analysis is capable to differentiate between high-neuroendocrine vs low-neuroendocrine SCLC cases.

[0166] SCLC subtypes can be predicted from the DELFI-based approach using cfDNA LC-WGS data.

[0167] First, the DELFI machine learning classifier detects the presence of cancer in patients with SCLC. Both genome-wide fragmentation profiles and correspondent DELFI scores were investigated to identify preliminary differences between the two main clinical subtypes of SCLC cases: high-neuroendocrine and low-neuroendocrine. Later publicly available data was used for running a principal component analysis to cluster the two subtypes.

[0168] Second, using publicly available data, it was determined that these clusters of SCLC samples exhibit a decrease in aggregate fragment coverage at ASCL1 transcription factor binding sites classifying these cases as high-neuroendocrine SCLCs.

[0169] Third, if one of these clusters of SCLC samples exhibit reduction of fragment coverage at genomic binding sites regulated by hematopoietic transcription factors (low-neuroendocrine SCLC) was also investigated.

[0170] Fourth, High vs Low Neuroendocrine cases were distinguished based on fragment length distribution at T-cells-specific partially methylated domains.

[0171] Finally, a receiver operating characteristic (ROC) representing sensitivity and specificity of the DELFI-fragmentome approach to identify SCLC subtypes was used to identify the diagnostic accuracy of our approach between the identified cluster samples.

[0172] Given the limited access to SCLC tumor biopsies it is believed that this approach could be a viable method of subtyping SCLC in a non-invasive manner. A promising assay for both pharmaceutical companies during clinical trials to evaluate novel immunotherapy drugs and for clinicians looking to select which SCLC diagnosed patients have the best chances to respond to immunotherapy.

[0173] The DELFI-scores of the patients diagnosed with SCLC in this study are shown in Figure 1.

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

[0174] **Figure 2** illustrates that genome-wide fragmentation profiles were significantly consistent among pre-treatment, post-treatment, progression and response time points suggesting genome-wide circulating tumor DNA fragment sizes is a powerful method to detect changes during monitoring of immunotherapy treatment of SCLC cancer treatment.

[0175] **Figure 3** illustrates genome-wide fragmentation profiles (and correspondent DELFI-scores below) showing partial differences between the two main subtypes of SCLC cases: High Neuroendocrine and Low Neuroendocrine; suggesting genome-wide circulating tumor DNA fragment sizes may be a powerful method to detect subtyping during monitoring of immunotherapy treatment of SCLC cancer treatment.

[0176] **Figure 4** illustrates differential genes between High and Low neuroendocrine SCLC cases, using publicly available data (Lissa et al. 2022) and a principal Component Analysis (on DELFI 30X plasma WGS data) to cluster the two subtypes. PCA analysis data revealed distinguishable clusters between SCLC High-NE vs Low-NE pre-treatment samples.

[0177] **Figure 5** illustrates an Eigencor analysis revealing a significant correlation between Principal Component 1 and 2 with the NE status of the samples.

[0178] **Figure 6 A-B** illustrates the potential of the DELFI assay to subtype SCLC using TFBS. A genome-wide cfDNA fragmentation analyses were performed at ASCL1 binding sites in the pre-treatment NCI samples. Figure 5A illustrates that distinct clusters of SCLC samples were observed corresponding to different levels of ASCL1 activation. Importantly the two clusters corresponded perfectly to the different SCLC subtypes. The main driver of differentiation between samples was in fact differences in the fragment coverage identified at the center of the ASCL1 binding sites. Figure 6B illustrates that other clinical or sample characteristics have a negligible impact in the clustering of these samples. Overall, these data show how DELFI analysis can perfectly distinguish between SCLC subtypes using fragment coverage at TFBS.

[0179] **Figure 7** illustrates a patient example supporting the hypothesis that the detected signal comes from tumor infiltrating lymphocytes. Patient NCI-0422, diagnosed with SCLC with inflammation subtype was treated with Durvalumab plus Olaparib. NCI-0422 responded well to treatment but was later diagnosed with progressive disease. Variable genomic binding is detectable at pre-treatment and at progression timepoints. Peaks are centered at a mix of TFBS and TSS.

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

[0180] **Figure 8 A-B** illustrates ASCL1 binding site between resp vs non-responders. Figure 8A shows 500 cell type specific bins from the PMD data to determine fraction of white blood cells. Figure 8B shows lymphocyte tissue specific bins.

[0181] **Figure 9** illustrates a machine learning model to predict high vs low neuroendocrine SCLC.

[0182] Subtyping is performed by calculating the \log_2 (Read Depth Ratio) across 20bp windows starting 2kb from ASCL1 TFBS sites. Read Depth Ratio is the total coverage over a coverage correction factor calculated using the median coverage of 5' and 3' 500bp anchors at the ends of the 2kb window (i.e. $\log_2(\text{Read Depth}/\text{Correction Factor})$). A General Additive Model is then used to smooth coverage and correct for GC bias. To validate the approach described herein, neuroendocrine status is derived from mean NE50 score per subject/treatment status, a positive value is classified as NE+. Subtype status is derived from aggregating the mean for each of the 4 subtypes (ASCL1, NEUROD1, POUF23, YAP1) such that each subject/treatment status has one value for each subtype. The subtype with the highest score is then selected as the truth. Fragmentomics are calculated as previously described (Cristiano et al.), Scores are for the latest available models.

[0183] **Figure 10 A-B** illustrates that the method described herein is able to differentiate neuroendocrine vs non-neuroendocrine by calculating Read Depth Ratio at SCLC specific genomic coordinates (Figure 10A.). Figure 10B shows that the method described herein appears to subtype better SCLC cases.

[0184] **Figure 11** illustrates that Read Depth Ratio at SCLC specific genomic coordinates are transformed into a 2-dimensional PCA to separate samples in Neuroendocrine vs non-neuroendocrine groups. Most of samples are clearly divided into different clusters.

[0185] **Figure 12** illustrates PCA derived from Read Depth Ratio at SCLC specific genomic coordinates is able to separate pre-treatment samples into specific SCLC subtypes (A,N,P,Y). The ground truth of the SCLC subtypes was calculated using tissue RNA-seq gene expression differential analysis.

[0186] The fragmentomics platform described herein detects small cell lung cancer (SCLC) with high sensitivity. The DELFI SCLC Subtyping Assay capable of subtyping SCLC samples, without the need of clinical knowledge. The DELFI SCLC Subtyping Assay identified four groups of SCLC samples belonging to different levels of ASCL1 activation. Samples with high activation of ASCL1 were confirmed to belong to Neuroendocrine subtypes (SCLC-A

and SCLC-N), while samples with low ASCL1 signal were confirmed to belong to non-neuroendocrine subtypes (SCLC-P and SCLC-Y).

EXAMPLE 3

[0187] Cell-free DNA fragmentation profiling as a method for tumor fraction assessment and treatment monitoring in NSCLC

[0188] cfDNA Fragmentomics Background: Traditionally monitoring is done with imaging, many individuals do not have ready access to hospitals with appropriate imaging equipment and expertise and need to travel significant distances for such procedures making continual monitoring problematic. There is no *a priori* need to know where a tumor is located. There is additionally no need to know beforehand the somatic mutations a tumor harbors. Costs associated with liquid biopsies are cheap when compared to imaging which should improve as sequencing costs continue to decline.

[0189] cfDNA Mutations have Limited Signal and can be Confounded by Clonal Hematopoiesis. DELFI captures many features of the cfDNA universe. cfDNA fragmentation patterns are determined by basal chromatin organization. cfDNA fragmentation profiles are highly consistent in healthy people and altered in patients with cancer.

[0190] DELFI-TF Model Overview and Application to Monitoring NSCLC Patients during Therapy: cfDNA aliquots from plasma samples of CRC patients were analyzed using ddPCR for RAS mutation status as well as low-pass WGS sequencing. WGS data was aligned and fragment size distributions obtained for 504 5Mb bins across the genome. A Bayesian regression model was trained and cross-validated against RAS MT samples using fragmentomics features. Figure 20 depicts proof-of-concept DELFI-TF model development.

[0191] A CRC trained DELFI-TF model was applied to a real-world NSCLC cohort.

Table 2 - NSCLC Cohort Clinical Features

Clinical features	
Median age (range)	71.5 (53-88)
Sex	
Male	11
Female	10
Median BMI (range)	24.5 (17-37)

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

Stage	7
Stage III	14
Stage IV	
PD-L1 expression	5
Strong positive	7
Weak positive	4
Negative	5
NA	
Treatment type	3
Immunotherapy	9
Immunotherapy + chemotherapy	9
Chemotherapy	
Best overall response	1
Complete response	2
Partial response	1
Stable disease	10
Progressive disease	7
Ongoing treatment	

[0192] Results: **Figure 13** illustrates depicts how fragmentation profiles inform on lung cancer status and treatment response patterns.

[0193] **Figure 14** illustrates that model derived DELFI-TF exhibits a strong correlation with Mutant Allele Frequency across samples.

[0194] **Figure 15** illustrates that DELFI-TF accurately quantify circulating tumor fraction, mirroring MAF performance in relation to RECIST evaluation.

[0195] **Figure 16** illustrates that tumor derived cfDNA has altered fragmentation.

[0196] **Figure 17 A-C** illustrates that DELFI-TF accurately detects circulating tumor fraction without being confounded by clonal hematopoiesis.

[0197] **Figure 18** depicts cfDNA fragmentation patterns accurately differentiate NSCLC subtypes.

[0198] Conclusions: Genome-wide cfDNA fragment profiles are abnormal in patients with cancer

[0199] cfDNA fragmentation scores (DELFI-TF) are highly correlated with known mutation allele frequencies. cfDNA fragmentation predicts RECIST status. cfDNA fragmentation is not confounded by clonal hematopoiesis. cfDNA fragmentation features can noninvasively distinguish histologic subtypes of lung cancers.

CLAIMS

1. A non-invasive method for subtyping small cell lung cancer in a subject as low neuroendocrine or high neuroendocrine small cell lung cancer comprising:
processing cfDNA fragments from a sample obtained from the subject and generating sequencing libraries;
subjecting the sequencing libraries to whole genome sequencing to obtain sequenced fragments, wherein genome coverage is about $30\times$ to $0.1\times$;
mapping the sequenced fragments to a genome to obtain genomic intervals of mapped sequences at specified transcription factor binding sites;
analyzing the genomic intervals of mapped sequences to determine cfDNA fragment lengths and amounts to establish a cfDNA fragment coverage score at specified transcription factor binding sites using the cfDNA fragment lengths and amounts; and
subtyping the small cell lung cancer in the subject based on transcription factor activation; wherein a decrease in an aggregate cfDNA fragment coverage scores at the specified transcription factor binding sites is indicative of a high-neuroendocrine small cell lung cancer subtype.
2. The method of claim 1, wherein the specified transcription factor is ASCL1, NEUROD1, POUF23, YAP1, or any combination thereof.
3. The method of claim 1, wherein the method uses machine learning to subtype small cell lung cancer in the subject.
4. The method of claim 1, wherein subtyping is performed by calculating a \log_2 (Read Depth Ratio) across 20bp windows starting 2kb from the specified transcription factor binding sites.
5. The method of claim 4, wherein the \log_2 (Read Depth Ratio) is the total coverage over a coverage correction factor calculated using the median coverage of 5' and 3' 500bp anchors at the ends of the 2kb window (i.e. $\log_2(\text{Read Depth}/\text{Correction Factor})$).
6. The method of claim 5, wherein all coverage calculations are shifted by 1 to avoid divisions by 0.

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

7. The method of claim 5, wherein the \log_2 (Read Depth Ratio) is calculated for each transcription factor binding site and then aggregated by calculating the median of each window to obtain a fragment coverage score.
8. The method of claim 1, further comprising the use of a general additive model to smooth coverage and correct for GC bias.
9. The method of claim 1, wherein the genomic intervals are non-overlapping.
10. The method of claim 1, wherein the genomic intervals each comprise thousands to millions of base pairs.
11. The method of claim 1, wherein a cfDNA fragmentation profile is determined within each genomic intervals.
12. The method of claim 11, wherein the cfDNA fragmentation profile comprises a median fragment size.
13. The method of claim 11, wherein the cfDNA fragmentation profile comprises a fragment size distribution.
14. The method claim 1, further comprising administering to the subject identified as having a high-neuroendocrine small cell lung cancer subtype, a therapeutic agent suitable for the treatment of the subtype of cancer.
15. The method of claim 14 wherein the therapeutic agent is an immunotherapy.
16. A non-invasive method for subtyping non-small cell lung cancer in a subject as adenocarcinoma or squamous carcinoma comprising:
 - processing cfDNA fragments from a sample obtained from the subject and generating sequencing libraries;
 - subjecting the sequencing libraries to whole genome sequencing to obtain sequenced fragments, wherein genome coverage is about $30\times$ to $0.1\times$;
 - mapping the sequenced fragments to a genome to obtain genomic intervals of mapped sequences at specified transcription factor binding sites;
 - analyzing the genomic intervals of mapped sequences to determine cfDNA fragment lengths and amounts to establish a cfDNA fragment coverage score at specified transcription factor binding sites using the cfDNA fragment lengths and amounts; and

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

subtyping the small cell lung cancer in the subject based on transcription factor activation; wherein a decrease in an aggregate cfDNA fragment coverage scores at the specified transcription factor binding sites is indicative of an adenocarcinoma or squamous carcinoma subtype.

17. The method of claim 16, wherein the specified transcription factor is ASCL1, NEUROD1, POUF23, YAP1, or any combination thereof.
18. The method of claim 16, wherein the method uses machine learning to subtype small cell lung cancer in the subject.
19. The method of claim 16, wherein subtyping is performed by calculating a log₂ (Read Depth Ratio) across 20bp windows starting 2kb from the specified transcription factor binding sites.
20. The method of claim 19, wherein the log₂ (Read Depth Ratio) is the total coverage over a coverage correction factor calculated using the median coverage of 5' and 3' 500bp anchors at the ends of the 2kb window (i.e. log₂(Read Depth/Correction Factor)).
21. The method of claim 19, wherein all coverage calculations are shifted by 1 to avoid divisions by 0.
22. The method of claim 19, wherein the log₂ (Read Depth Ratio) is calculated for each transcription factor binding site and then aggregated by calculating the median of each window to obtain a fragment coverage score.
23. The method of claim 16, further comprising the use of a general additive model to smooth coverage and correct for GC bias.
24. The method of claim 16, wherein the genomic intervals are non-overlapping.
25. The method of claim 16, wherein the genomic intervals each comprise thousands to millions of base pairs.
26. The method of claim 16, wherein a cfDNA fragmentation profile is determined within each genomic intervals.
27. The method of claim 26, wherein the cfDNA fragmentation profile comprises a median fragment size.
28. The method of claim 26, wherein the cfDNA fragmentation profile comprises a fragment size distribution.

PATENT
ATTORNEY DOCKET NO. DELFI2150-3WO

29. The method claim 16, further comprising administering to the subject identified as having an adenocarcinoma or squamous carcinoma subtype, a therapeutic agent suitable for the treatment of the type of cancer.
30. The method of claim 29, wherein the therapeutic agent is an immunotherapy.
31. The method of claim 16, further comprising quantifying a circulating tumor fraction, mirroring major allele frequency (MAF) performance in relation to Response Evaluation Criteria in Solid Tumors (RECIST) evaluation in both adenocarcinoma and squamous carcinoma.

FIGURE 1

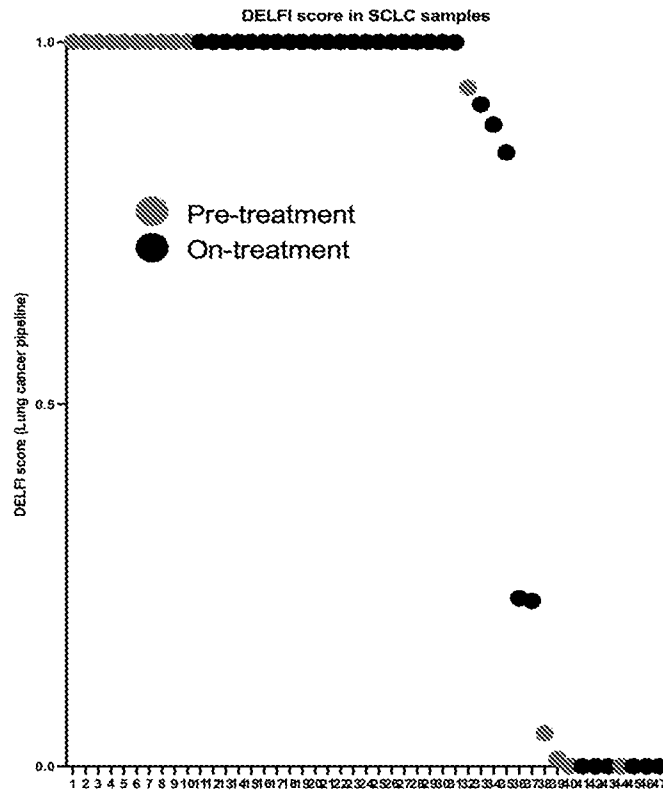


FIGURE 2

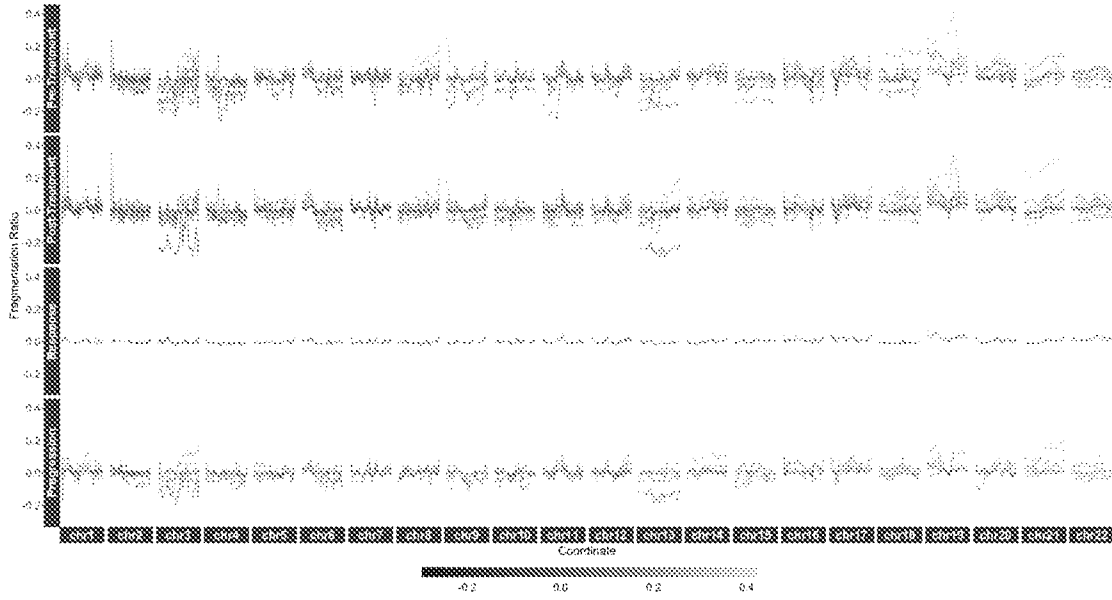


FIGURE 3

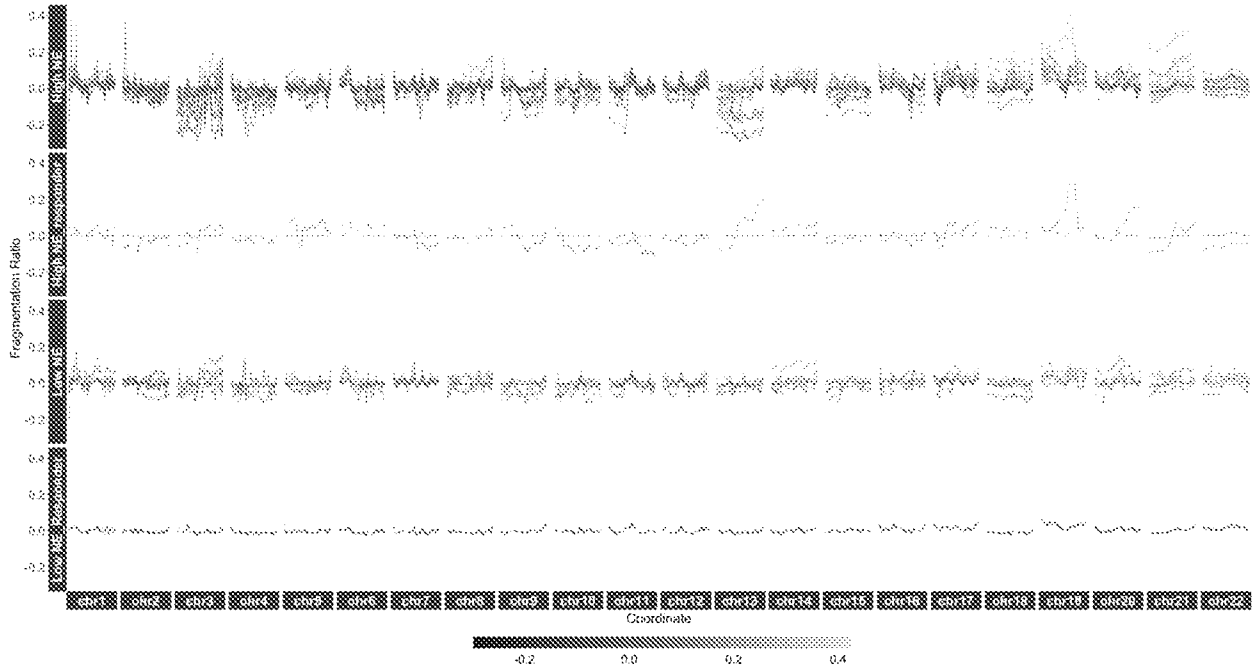


FIGURE 5B

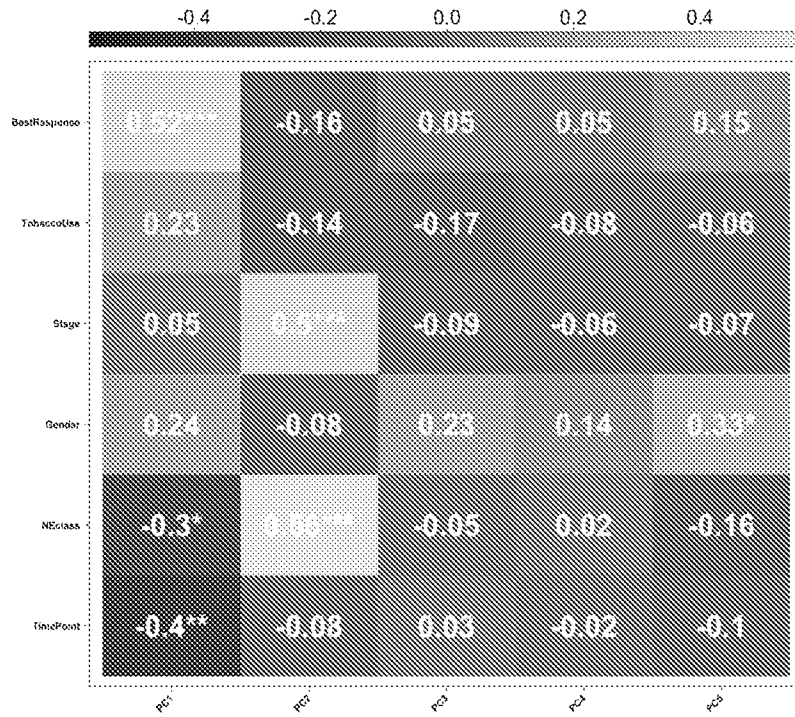
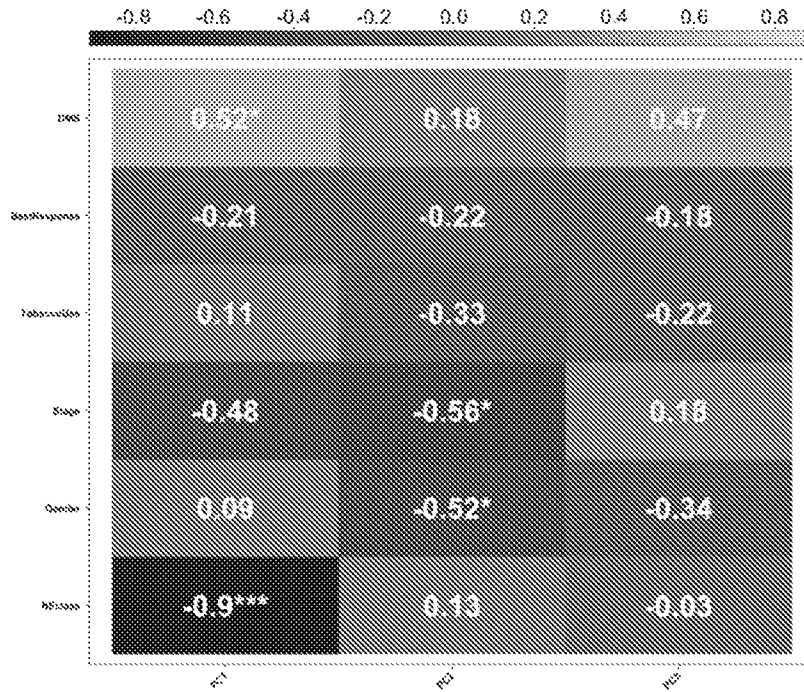
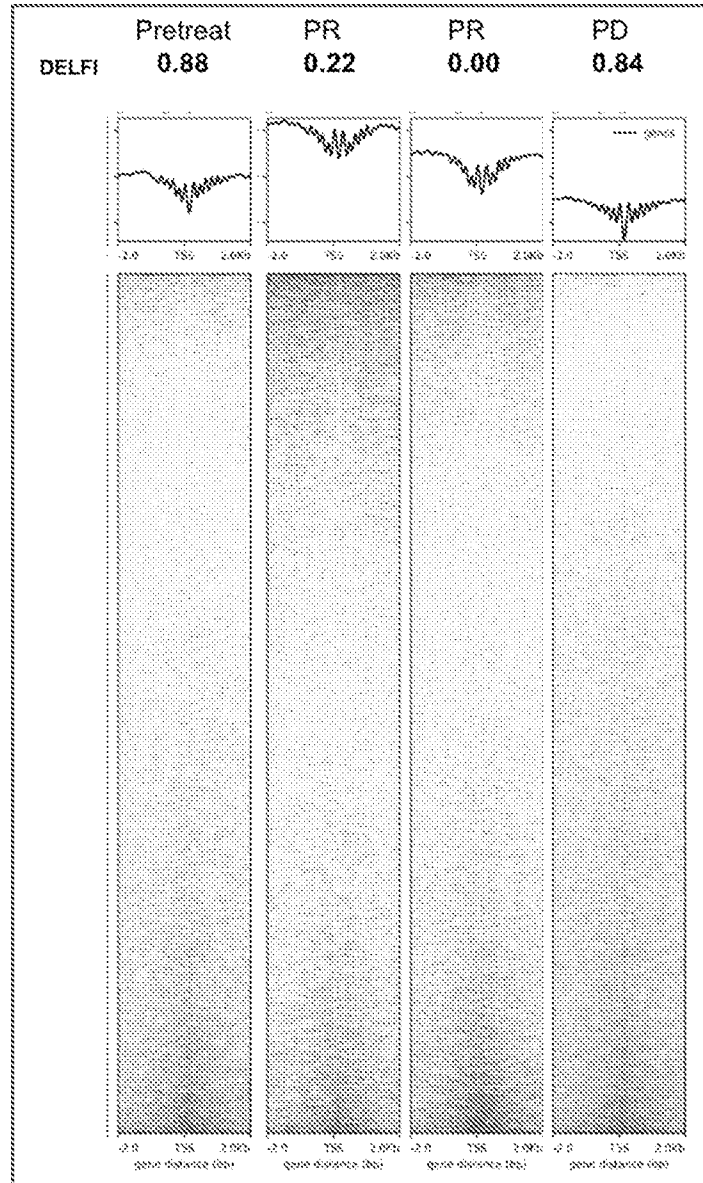


FIGURE 6B



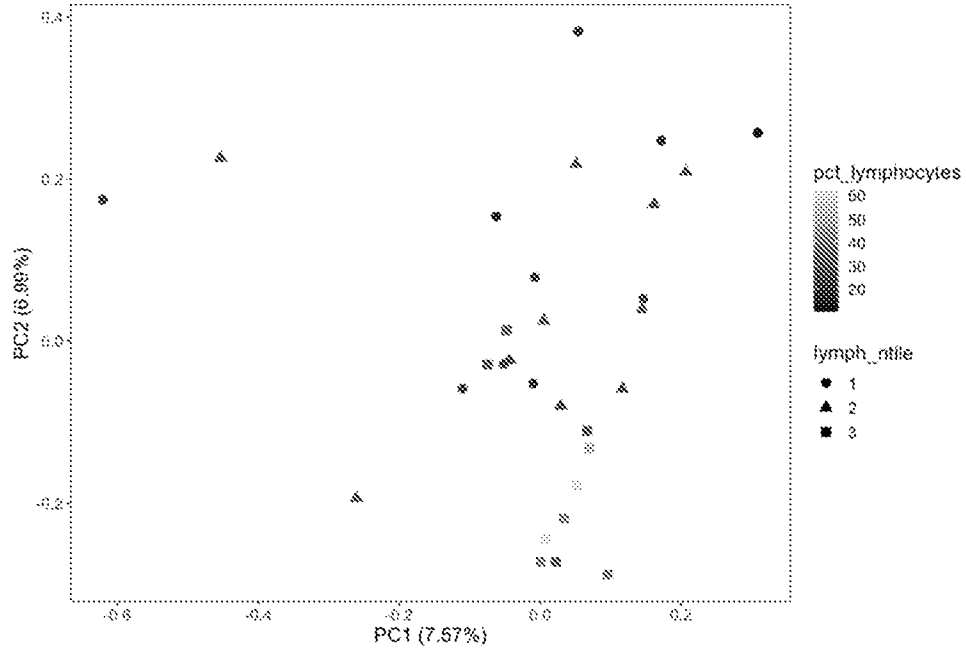
9/26

FIGURE 7



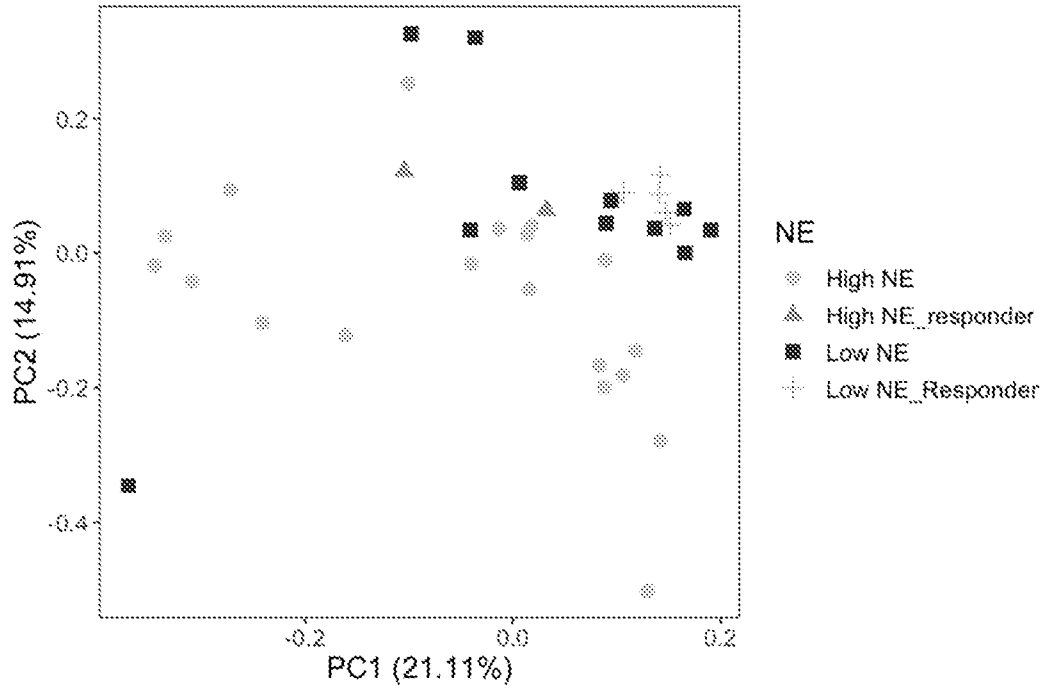
10/26

FIGURE 8A



11/26

FIGURE 8B



12/26

FIGURE 9

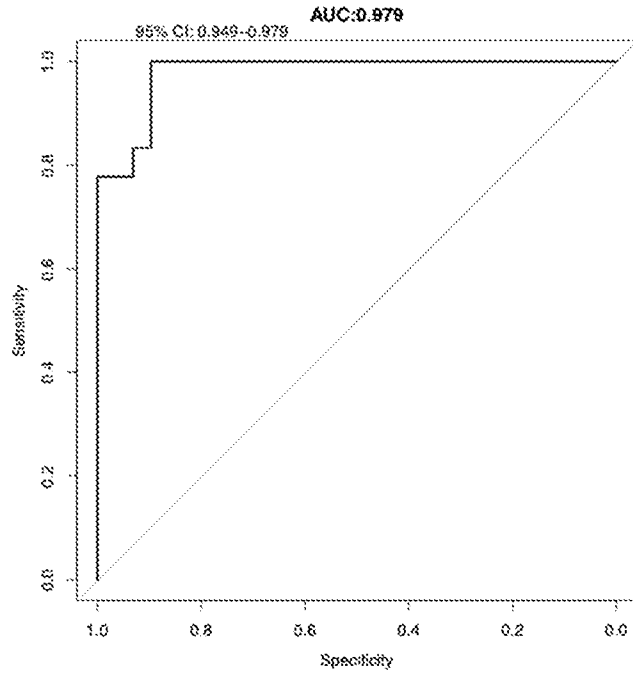


FIGURE 10A

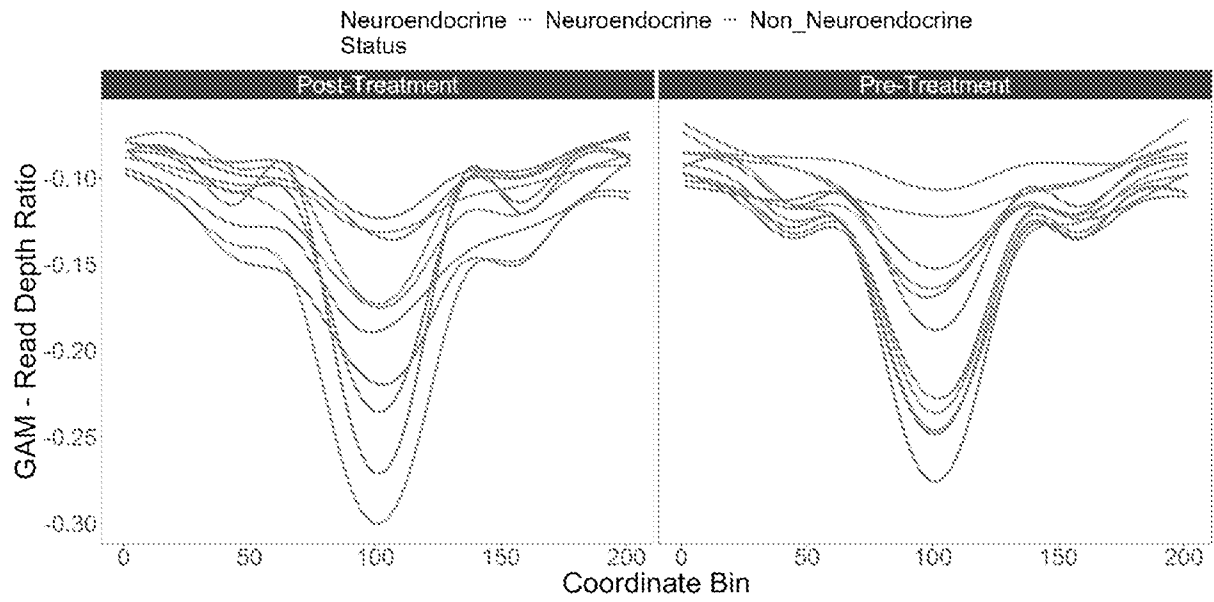


FIGURE 10B

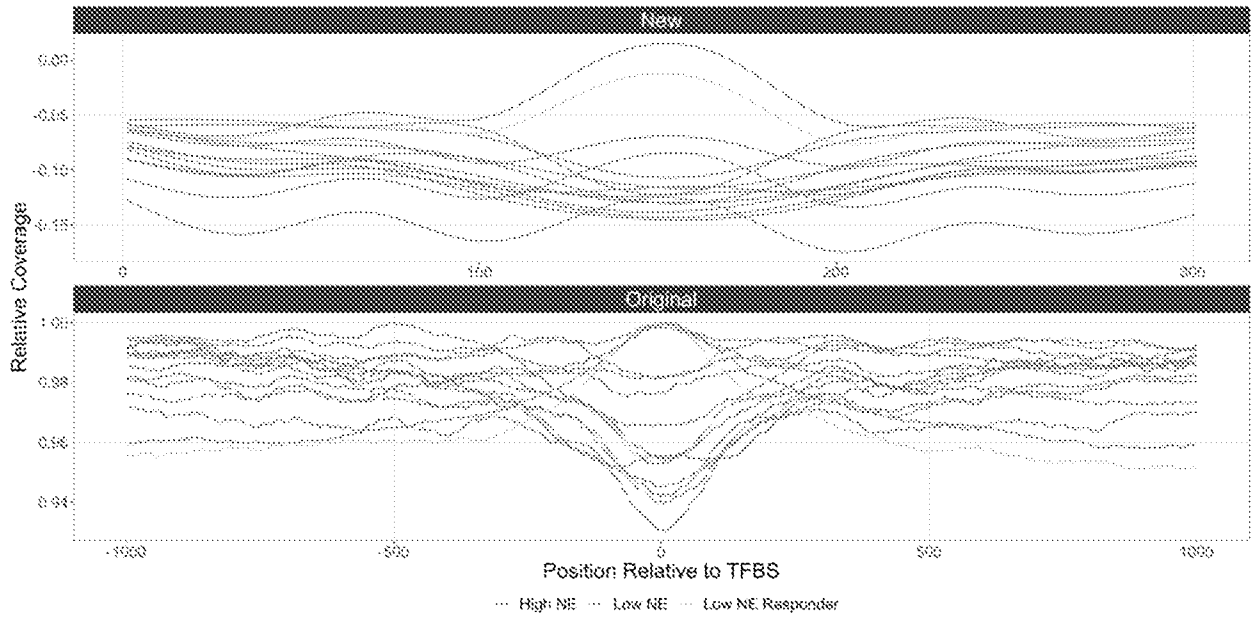
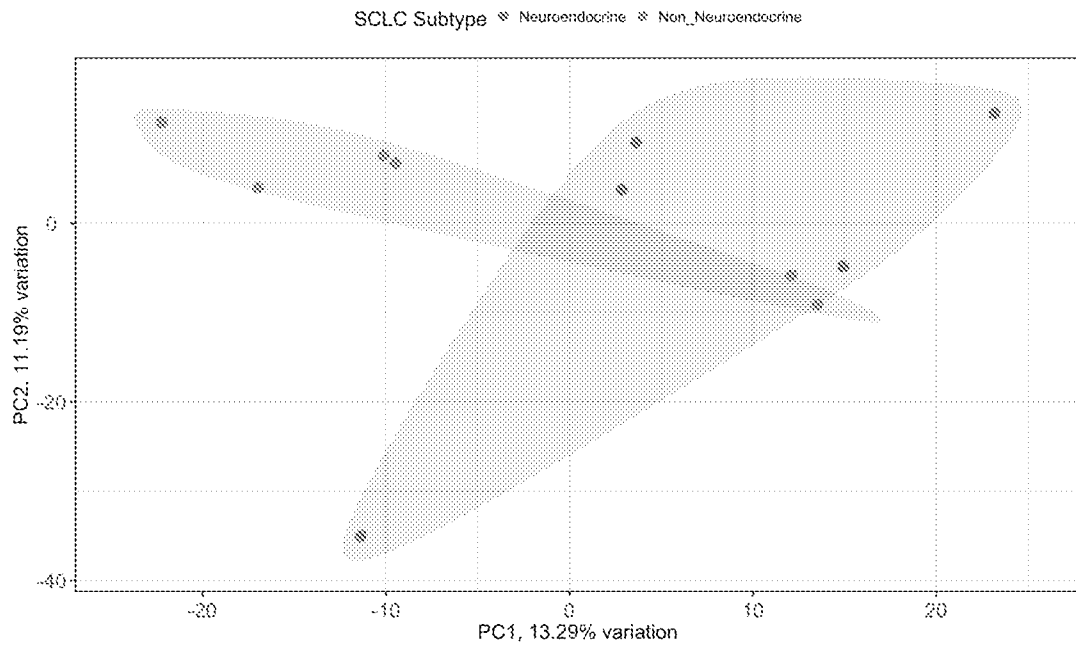


FIGURE 11



17/26

FIGURE 13

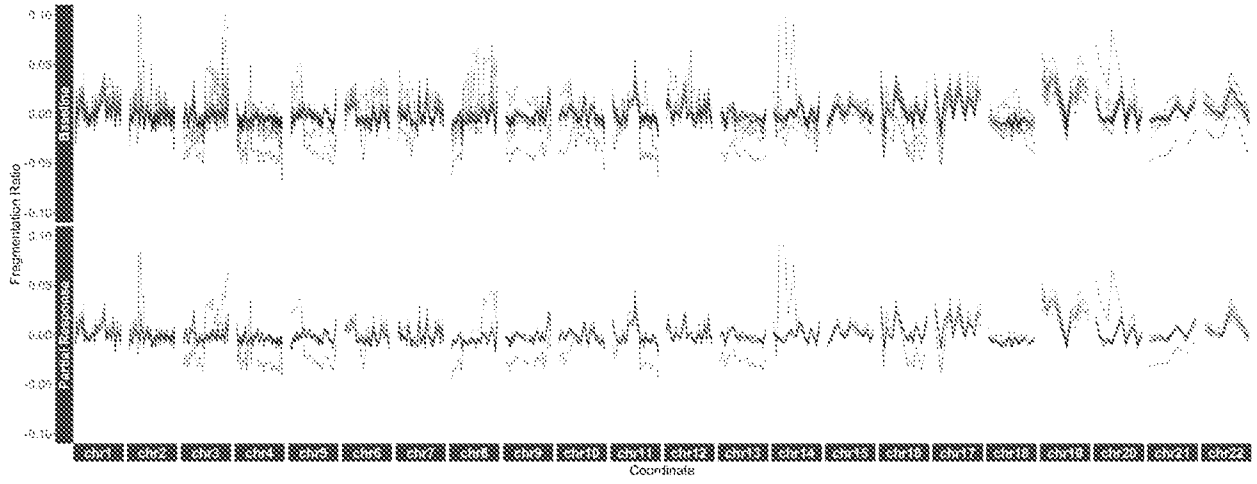


FIGURE 14

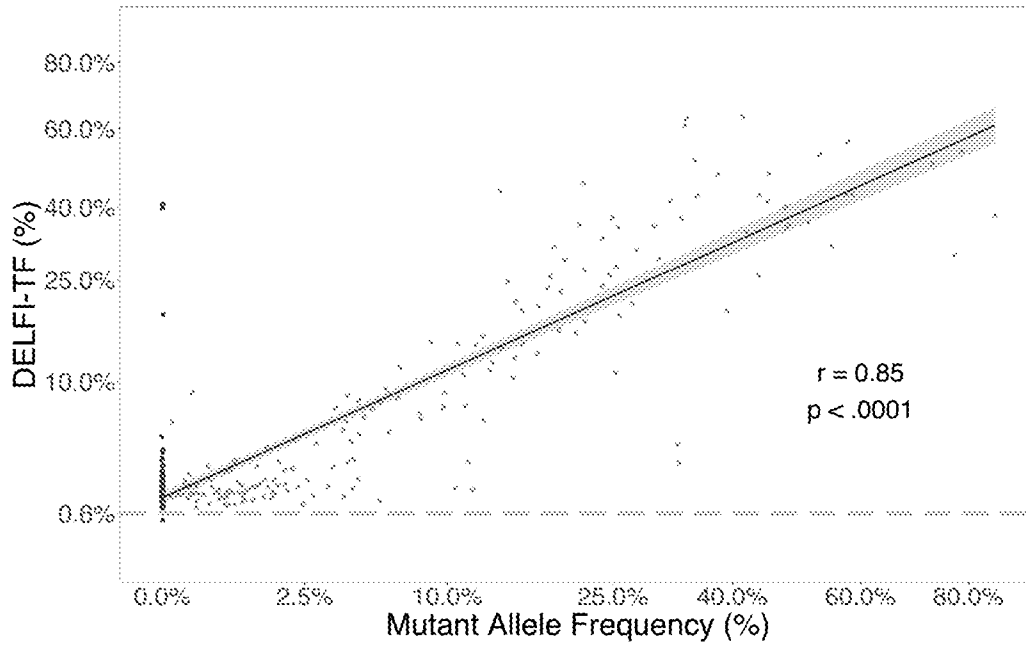


FIGURE 15

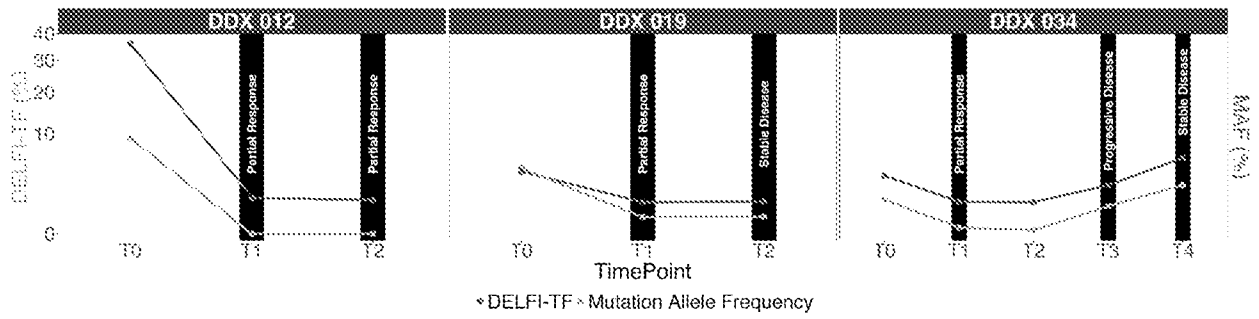


FIGURE 16

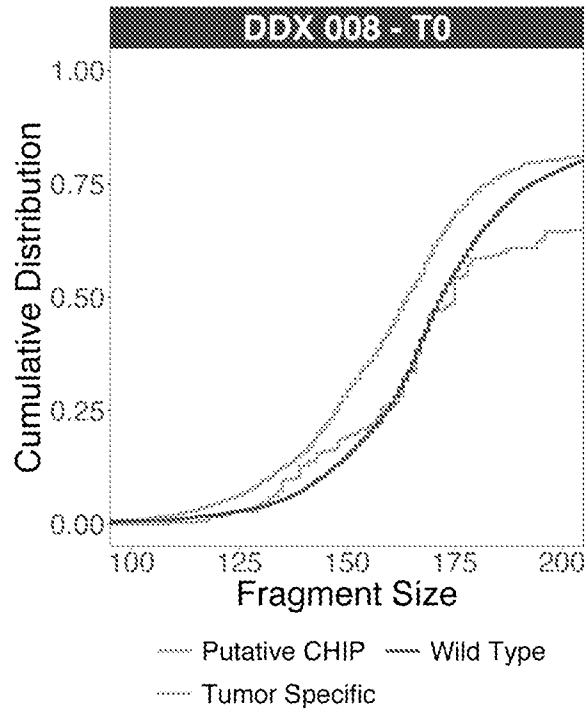


FIGURE 17A

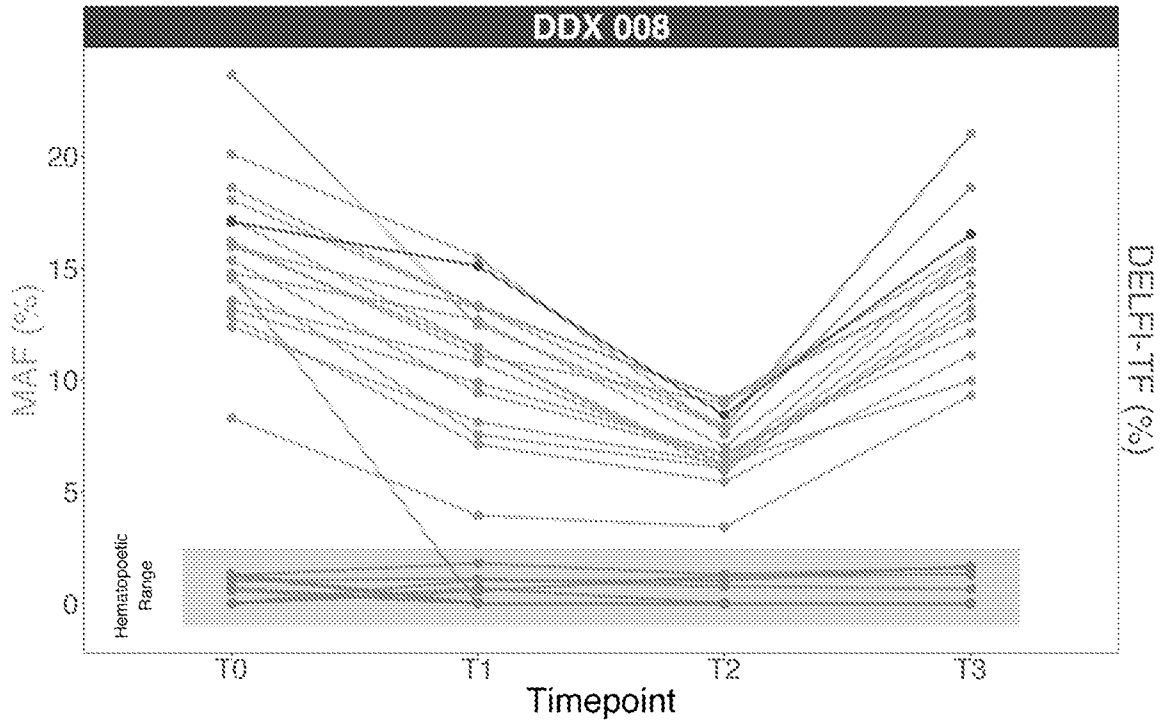


FIGURE 17B

Indivumed Cohort: DELFI-TF vs mean MAF

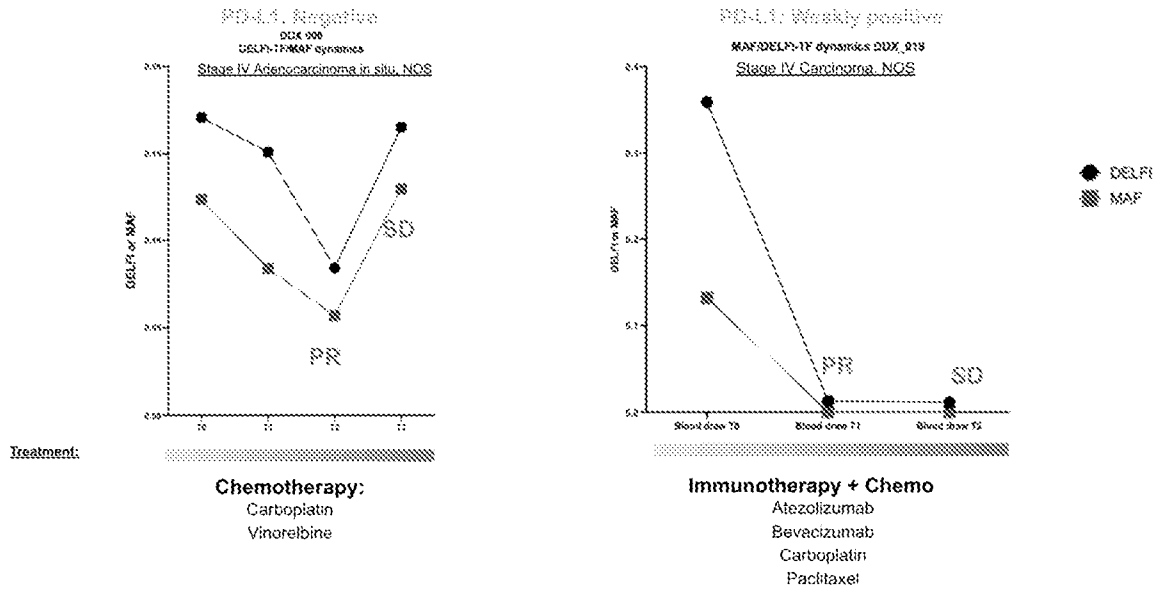


FIGURE 17C

DELFI-TF vs mean and max MAF

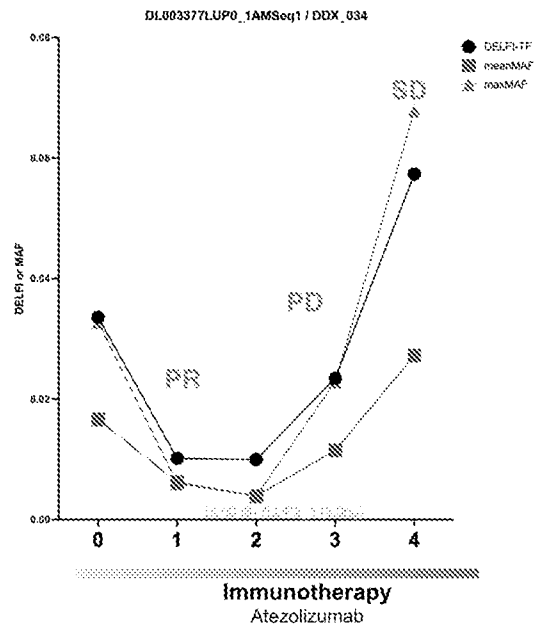


FIGURE 18

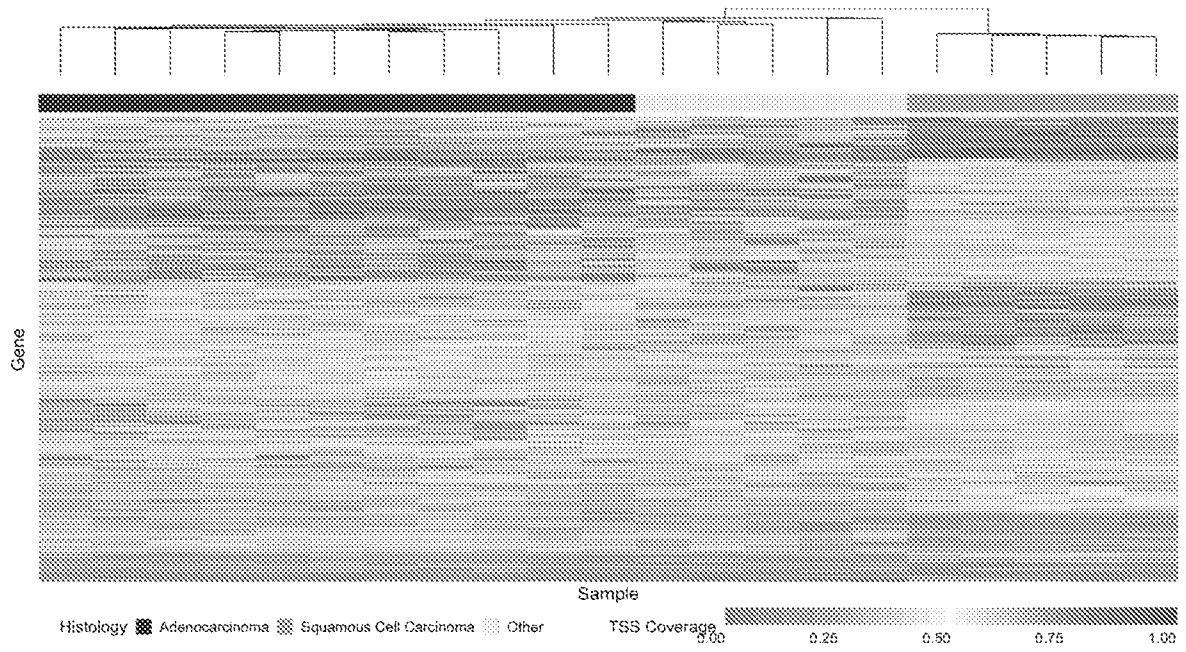


FIGURE 19

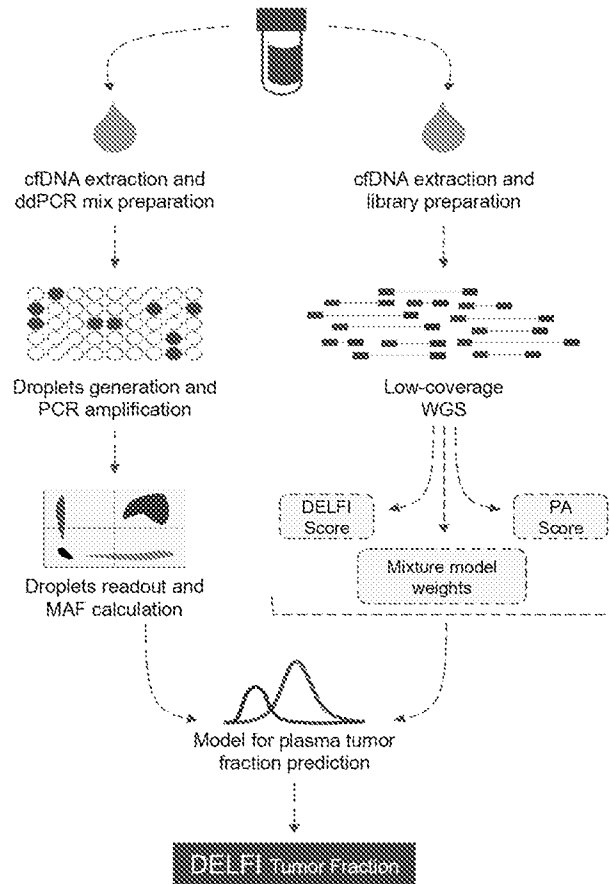


FIGURE 20

