Lung Cancer Subtyping Using Cell-free DNA and Protein Biomarkers

Stephen Cristiano[1], Milou Schuurbiers[2], Laurel K. Millberg[1], Jamie E. Medina[1], Zachary L. Skidmore[1], Paul van der Leest[2], Garrett Graham[1], Alessandro Leal[3], Bryan Chesnick[1], Kim Monkhorst[2], Nicholas C. Dracopoli[1], Robert Scharpf[4], Peter Bach[1], Daan van den Broek[2], Amoolya Singh[1], Victor Velculescu[4], Sian Jones[1], Michel Van den Heuvel[5], Lorenzo Rinaldi[1]*

[1] Delfi Diagnostics, Inc., Palo Alto, CA, USA; [2] Netherlands Cancer Institute, Amsterdam, The Netherlands; [3] NYU Langone Health Perlmutter Comprehensive Cancer Center, New York, USA [4] The Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, MD, USA; [5] Department of Pulmonary Diseases, Radboud University Medical Center, The Netherlands

#1135

BACKGROUND

Lung cancer is the leading cause of cancer-related mortality worldwide. Accurate histological subtyping to differentiate between lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), and small cell lung cancer (SCLC) is critical for guiding optimal therapeutic strategies; however, up to 20% of patients lack sufficient tissue for conventional histopathological classification. Liquid biopsies using cell-free DNA (cfDNA) fragmentomics offer a promising non-invasive alternative for cancer characterization when tissue is not available.

METHODS

We examined 761 patients with newly diagnosed, treatment-naive lung cancer of all stages, including lung adenocarcinoma (n=459), squamous cell carcinoma (n=152), small cell carcinoma (n=42), and other lung cancer subtypes (n=92) from the prospective Lung Cancer Early Molecular Assessment trial (LEMA, NCT02894853). Low-coverage whole-genome sequencing of plasma samples was performed to derive fragmentation features used to subtype lung cancer. Circulating tumor DNA (ctDNA) burden was estimated from fragmentation using the DELFI-TF method. We developed a machine learning classifier trained exclusively on the tissue-based copy number signatures from the Clinical Lung Cancer Genome Project (CLCGP) and applied it to patient plasma samples.

RESULTS

This tissue-trained, plasma-validated subtyping algorithm achieved an AUC of 0.99 (95% CI = 0.98-1.00) for distinguishing NSCLC from SCLC and an AUC of 0.91 (95% CI = 0.87-0.95) for differentiating LUAD from LUSC. The model correctly classified 88% of SCLC, 80% of LUAD, and 87% of LUSC cases where the tumor fraction was ≥0.3% (n=276).

Among a subset of 361 NSCLC patients, integration of five serum protein biomarkers resulted in a multimodal model that differentiated LUAD from LUSC across all tumor fractions with high performance (AUC=0.85, CI = 0.80-0.90), an improvement over cfDNA (p<0.01; AUC=0.78, CI = 0.74-0.82) or protein-only classifiers (p<0.001; AUC=0.70, CI = 0.62-0.78).

CONCLUSIONS

These findings establish cfDNA fragmentation and protein biomarkers as a viable non-invasive approach for lung cancer subtyping when tissue is unavailable, with potential to expedite subtype-specific treatment selection and improve clinical outcomes.

Table 1.
Patient characteristics

Participants/samples, n	
745	

Median Age	71 (39-94)
Male	398 (53%)
Female	347 (47%)

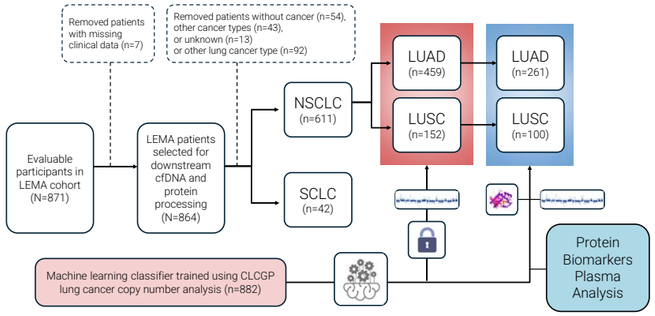
Stage, n (%)

Stage I	188 (25.2%)
Stage II	69 (9.2%)
Stage III	186 (24.9%)
Stage IV	301 (40.4%)

Histology, n (%)

Adenocarcinoma (ADC)	459 (61.6%)
Squamous cell carcinoma (SCC)	152 (20.4%)
Large cell carcinoma (LCC)	15 (2%)
Mixed phenotype (ASC)	13 (1.7%)
Small Cell Lung cancer (SCLC)	42 (5.5%)
Unknown histology (Other)	64 (8.6%)

Figure 1:
Approach for developing a lung cancer subtype classifier using cell-free DNA and protein biomarkers



- The flow diagram illustrates the processing and analysis of the DELFI pWGFrag-Lung assay for the 653 patients selected for subtyping analysis
- A machine learning classifier was trained using copy number data from CLCGP lung cancer tissue samples
- This classifier was then used to analyze cfDNA WGS data to categorize patients to SCLC, LUAD and LUSC subtypes
- The model performance was further improved by combining fragmentation with protein biomarkers into a multimodal model

Figure 2:
Genome-wide fragmentation profiles across different lung cancer subtypes

- The ratio of short to long genome-wide cfDNA fragmentation profiles in 5Mb bins across LUAD, LUSC, and SCLC lung cancer subtypes in the LEMA cohort
- Consistent aberrations in fragmentation patterns are observed in the subtypes, with SCLC showing the most divergent profile, characterized by widespread genomic alterations

Figure 7:
Implementation of cfDNA-fragmentome-based approaches for lung cancer subtyping in clinical practice

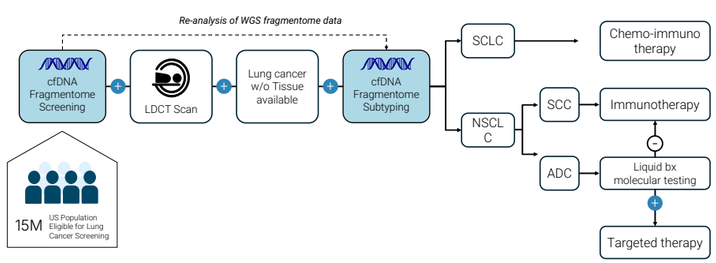
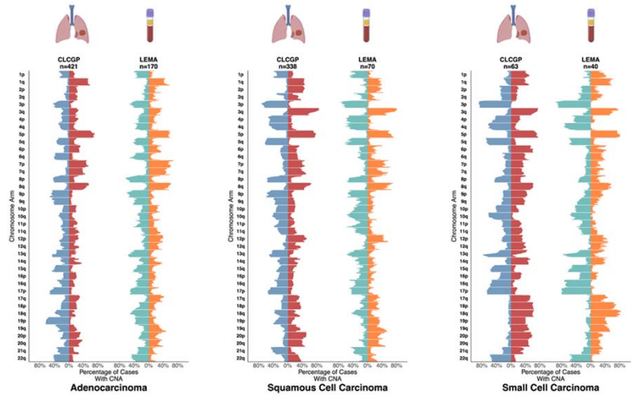
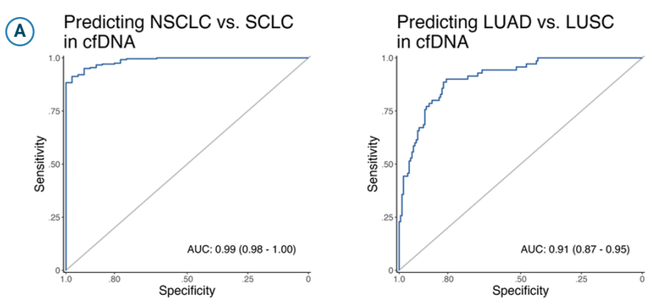


Figure 3:
Copy number signatures distinguish subtypes in tissue and plasma



- Copy number analysis using WGS data generated from lung cancer tissue samples obtained from the CLCGP database and cfDNA collected from patients from the LEMA cohort with a DELFI-Tumor Fraction ≥0.3%
- Segmentation scores were derived for each sample for adjacent non-overlapping 1 Mb windows
- The right-hand bars indicate proportion of samples with amplifications and left-hand bars indicate proportion of samples with deletions

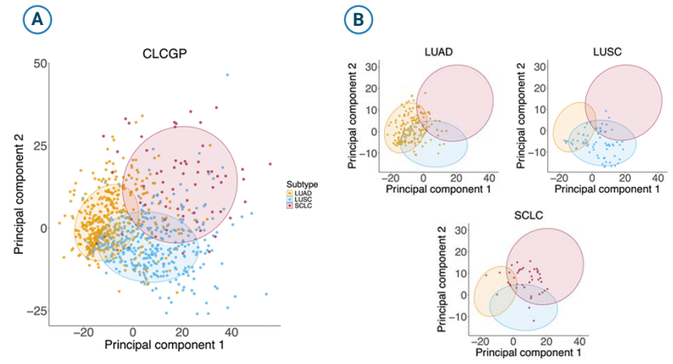
Figure 5:
Tumor tissue-trained model applied to cfDNA fragmentomes with tumor fraction ≥0.3%



A)Left, validation of a CLCGP tumor tissue-trained model for distinguishing NSCLC from SCLC using plasma samples with a DELFI-TF ≥0.3% (n=280). Receiver operating characteristic (ROC) curves show that SCLC and NSCLC are distinguished with AUC=0.99 (95% CI=0.98-1.0). Right, validation of a CLCGP tumor tissue-trained model for distinguishing LUAD from LUSC using plasma samples with a DELFI-TF ≥0.3% (n=240). ROC curves show that LUAD and LUSC are distinguished with AUC=0.91 (95% CI=0.87-0.95)

B)Alluvial diagram displaying the accuracy of subtyping of LEMA patients with LUAD, LUSC, and SCLC lung cancers

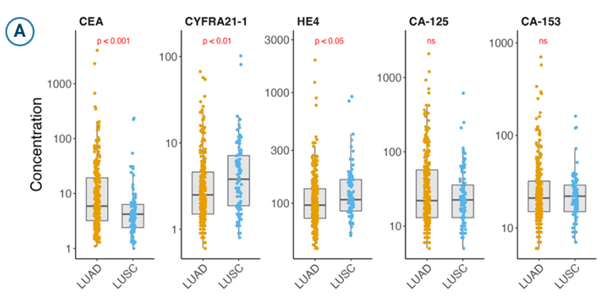
Figure 4:
Machine learning classifier using tissue-based copy number signatures accurately predicts lung cancer subtypes



A)PCA of copy number signatures from lung cancer tissue samples obtained from the CLCGP database, showing separation of lung cancer subtypes. Each point represents an individual sample colored by subtype: LUAD (orange), LUSC (blue), and SCLC (red). Colored ellipses contain 75% of the respective subtype in the CLCGP cohort: LUAD (yellow), LUSC (blue), and SCLC (red)

B)Plasma cfDNA copy number signatures from the LEMA cohort projected onto ellipses representing the lung cancer subtypes from CLCGP for LUAD (top left), LUSC (top right), and SCLC (bottom). The classification identified in tumor tissue was recapitulated in cfDNA from the LEMA cohort

Figure 6:
Multimodal fragmentome-protein model distinguishes LUAD from LUSC in all LEMA samples with both biomarkers available



A)Box plots comparing indicated protein biomarker values between LUAD and LUSC cases

B)ROC curve showing that LUAD and LUSC are distinguished using cross-validated cfDNA copy number signatures, protein biomarkers, or a multimodal model encompassing both approaches using all plasma samples from the LEMA cohort with protein measurements available (n=361).

Presented at AACR Annual Meeting 2026; 17-22 April 2026, San Diego, USA

References:
[1] Mazzone et al. Cancer Discov. 2024 Nov 1;14(11):2224-2242. [2] Mathios et al. Nat Commun. 2021 Aug 20;12(1):5060. [3] Cristiano S, et al. Nature. 2019;570:385-9. [4] Schouten et al PLoS One. 2024 Jul 31;19(7) [5] van't Erve et al. Nat Commun. 2024 Oct 21;15(1):8801

- **Tumor subtyping is necessary for therapy selection in cancer patients, but not feasible in some individuals.**
- **Cell-free DNA fragmentomes and protein measurements can distinguish lung cancer subtypes with high performance.**
- **This approach provides an avenue for non-invasive lung cancer subtyping to guide treatment selection that could lead to improve clinical outcomes.**