2718
AACR April 2026

Liquid biopsy cfDNA methylation predicts lung tumor size and metastatic potential in a single assay

Section 2 | Board 11
Abstract 6208

Kade Pettie[1*], Shiva Farashahi[1], Jackson Killian[1], Andrew Wong[1], Yifan Wu[1], Dorna Kashef[1], Franziska Michor[1], Jocelyn Charlton[1], and Kieran Chacko[1]
[1]Harbinger Health, Cambridge, MA * Corresponding author: kpettie@harbinger-health.com

BACKGROUND

Liquid biopsy enables noninvasive assessment of cancer severity and prognosis, informing clinical decisions across the cancer care spectrum. Estimates of the fraction of tumor-derived DNA shed into circulation (tumor content; TC) reflect disease severity, with higher TC more frequently observed in advanced stages and linked to poorer outcomes. However, TC is a product of a complex pathology and is impacted by many factors including tumor type, size, shedding rate, aggressiveness, vascularization, genotype, and metastatic state. Here we leverage a targeted cell-free DNA (cfDNA) methylation assay (HHx)[1] to predict two of these TC-influencing factors: size and metastatic state.

METHODS & MATERIALS

We enriched for 18.6 Mb bisulfite-converted DNA and sequenced samples to ~300X depth. We developed a lung-specific TC estimator from differential analysis of synthetic cfDNA data[2] (lung vs non-cancer). To assess metastatic potential, we trained models for two binary prediction tasks: distant metastasis (stage I/II vs IV) and late stage (stage I/II vs III/IV). A methylation metric quantifying per-region methylation levels was residualized with respect to sample-level TC to capture signals orthogonal to size. TC was evaluated as a single predictor in parallel. We trained on a dataset of 54 true and 324 synthetic lung cancer cfDNA samples to further emphasize size-independent, stage-related features. All models were assessed at 90% target specificity on a test set of 30 cancer samples[3] classified as lung cancer by a multiclass tissue-of-origin model (TOO)[4], including 6 non-lung-cancer samples incorrectly classified as lung, to represent the end-to-end performance of the assay (**Fig 1**).

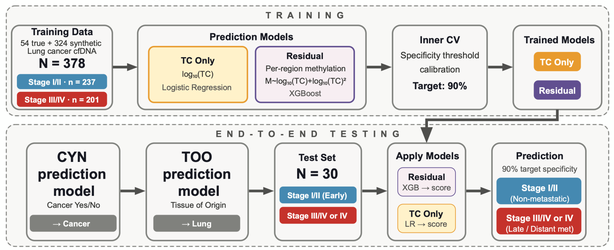

Figure 1. Schematic of model development and evaluation for metastatic disease detection. CYN = Cancer Yes/No, XGB = Extreme Gradient Boosting, LR = Logistic Regression.

For size prediction, we used a custom optical character recognition-large language model pipeline to extract tumor attributes from radiology reports (**Fig 2**). We then fit log-linear models linking TC to size metrics, including maximum standardized uptake value (SUVmax) from PET scan reports as a size proxy, and generated predictions for 57 held-out test samples. All samples were from the CORE-HH clinical study (NCT05435066).

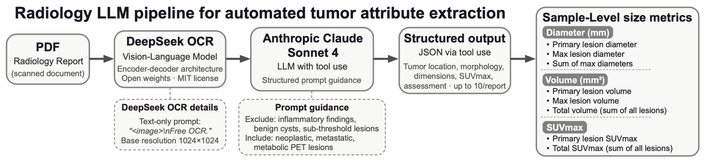

Figure 2. Flowchart of OCR-LLM tumor size extraction pipeline and evaluated size metrics.

RESULTS

Tumor content estimation

Our lung-specific TC estimator showed increasing TC by stage relative to non-cancer. TC levels were similar between adenocarcinoma and squamous cell carcinoma and elevated in small cell lung cancer in our true cfDNA data (**Fig 3**).

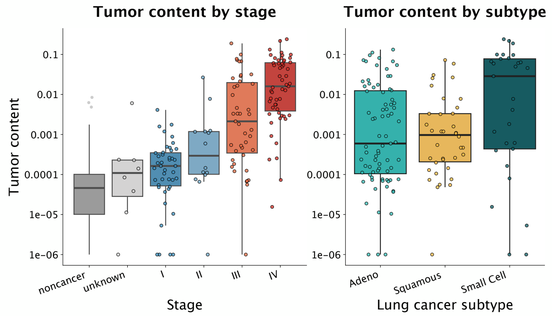

Figure 3. TC vs stage and subtype for lung cancer cfDNA samples. Adeno = adenocarcinoma, Squamous = squamous cell carcinoma.

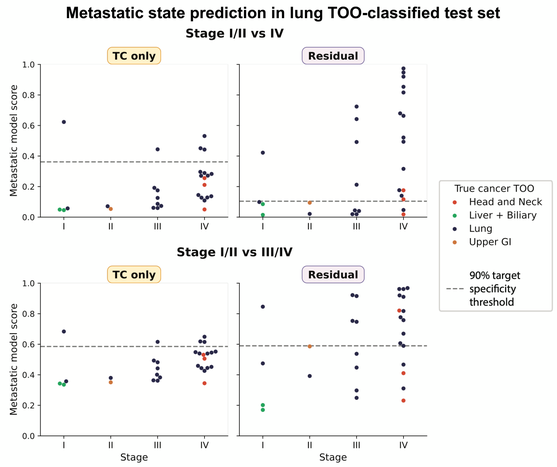

Figure 4. Test set true stage vs metastatic state score. (Top) Metastatic state scores from models trained to predict distant metastasis (stage I/II vs IV). (Bottom) Metastatic state scores from models trained to predict early vs late stage (stage I/II vs III/IV). Models were trained using TC only (left) or TC-residualized methylation features (right).

Metastasis prediction

The residual methylation models identified distant metastasis with 87.5% sensitivity (14/16 stage IV) at 83% specificity (5/6 stage I/II) and late stage with 67% sensitivity (16/24 stage III/IV) at 83% specificity (5/6 stage I/II). These results substantially improved on the TC-only model (19% and 17% sensitivity, respectively, at the same specificities), indicating methylation levels across HHx-targeted regions capture tumor progression-associated signal orthogonal to TC (**Fig 4**).

Top model features (Shapley importance-derived) were filtered through gene set enrichment analysis to identify lung- and metastasis-relevant genes for visualization in our independent test set. Notably, CDKN2A, a tumor suppressor frequently silenced by promoter hypermethylation in non-small cell lung cancer (NSCLC)[5], distinguishes a cluster of late-stage NSCLC samples, supporting the biological relevance of the model (**Fig 5**).

Test set methylation values of late stage-predictive features

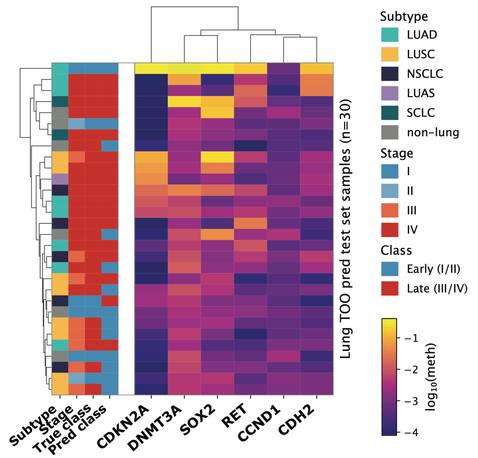

Figure 5. Methylation values of known lung- and/or metastasis-associated genes appearing in gene sets enriched among top features important for early vs late stage prediction. LUAD = Lung Adenocarcinoma, LUSC = Lung Squamous Cell Carcinoma, NSCLC = Non-Small Cell Lung Cancer (unspecified), SCLC = Small Cell Lung Cancer, LUAS = Lung Adenosquamous Carcinoma

RESULTS

Tumor size prediction

To determine how our lung-specific TC estimator scales with tumor size, we evaluated fits and predictions for several size metrics (**Fig 2B**). Restricting to Stage I–III cases with PET metrics (N = 19) yielded the strongest fits ($\log_{10}$(total volume) $R^2 > 0.5$, $p < 5\times10^{-4}$) (**Fig 6**), suggesting that PET scan reports may have more TC-relevant size estimates.

Test set predictions (N = 57) showed moderate explanatory power ($R^2 = 0.24$, $p = 1.3\times10^{-4}$), suggesting that factors beyond size (e.g., visceral metastasis, genotype) may be influencing shedding (**Fig 7**). Despite complex shedding dynamics in metastasis, explanatory power was slightly improved when including late stage samples likely due to limited sample size.

Overall, the model fit suggests similar shedding rates and scaling factors for adeno- and squamous cell carcinomas with respect to our TC estimator, where a doubling in volume corresponds to a ~2.17x increase in TC.

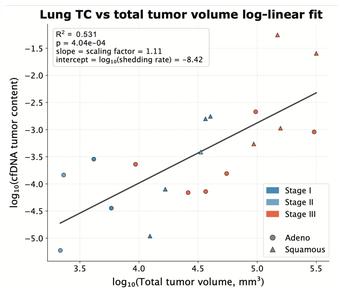

Figure 6. Fit of lung TC to total tumor volume. Log-linear model fit using training samples subset to stage I-III with PET scan metrics (SUVmax) available.

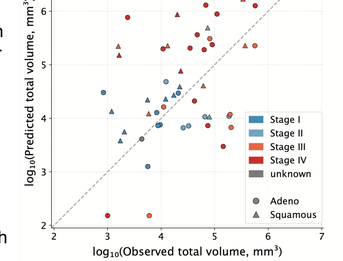

Figure 7. Test set true vs predicted tumor size. Predicted total tumor volume using log-linear fit on lung TC (**Fig 6**) vs observed total tumor volume from all test set radiology reports.

CONCLUSIONS

These findings highlight the potential for stage and size prediction models to deliver clinically actionable insights from a single blood draw. Late-stage prediction could guide workup prioritization, treatment intensity, and surveillance strategies, while size prediction may support prognosis and therapy selection. These capabilities motivate further model development for complementary prediction tasks (e.g., aggressiveness, genotype) toward a suite of tools for precision oncology.

REFERENCES
1. Gregg J, Michor F, et al. (2023) Novel blood-based assay for detection of early stage multi-cancer. J Clin Oncol 41, e15035-e15035.
2. Pettie K, et al. (2024) Fabrica™: A large-scale data simulation platform isolates tumor signal from cell-free DNA and improves tissue of origin prediction accuracy. AACR: Liquid Biopsy. Abstract nr B065.
3. DiRienzo G, et al. (2025) Tissue-specific predictive performance: A unified estimation and inference framework for multi category screening test. arxiv.org/abs/2505.21482.
4. Farashahi S, et al. (2025) Denoising Models Enhance Detection of Tumor-Derived cfDNA fragments and Cancer Tissue signal in Liquid Biopsy. AACR: Artificial Intelligence and Machine Learning. Abstract nr A035.
5. Gu J, Wen Y, et al. (2013) Association between P16INK4a Promoter Methylation and Non-Small Cell Lung Cancer: A Meta-Analysis. PLOS ONE 8(4): e60107.

DISCLOSURES
This study was sponsored by Harbinger Health. Cambridge, MA

ACKNOWLEDGEMENTS
We gratefully acknowledge all participants for their contributions, without whom this research would not have been possible. We also thank all cross-functional teams for their dedicated efforts that made this work and the resulting data possible.

www.harbinger-health.com