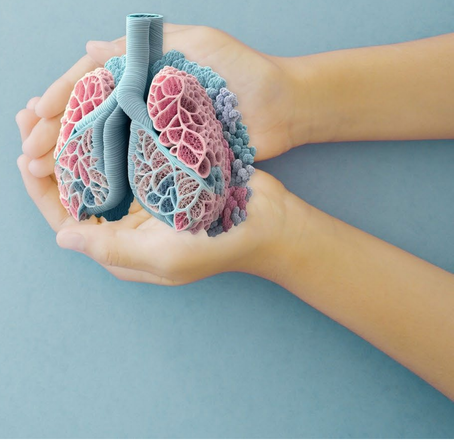
Point-of-Care Breath Volatile Organic Compound Analysis as a Tool for Lung Cancer Screening: A Pilot Feasibility Study

Yakov Pichkar[1], **Spiros Manolakos**[1], Kaela M. Phillips[2], Matthew B. Schabath[2,3], Ashish Chaudhary[1]

[1] Detect-ION, Tampa, Florida
[2] Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center & Research Institute, Tampa, Florida
[3] Department of Thoracic Oncology, H. Lee Moffitt Cancer Center & Research Institute, Tampa, Florida

Presented at the Dr. Robert Gillies Machine Learning Workshop in Cancer, Tampa, FL



1 The screening access gap

Low-dose computed tomography (**LDCT**) screening reduces lung cancer mortality, but its reach is limited on **two fronts**:

35.1% of lung cancer patients meet USPSTF screening criteria[2]

18.7% of screening-eligible individuals are up to date with LDCT[1]

Breath diagnostics can address both. Volatile organic compounds (VOCs) in exhaled breath, measured by GC-MS, carry a non-invasive lung cancer signal[3,4]. A point-of-care breath test could **reach people outside current guidelines** and make screening **easy enough that eligible people follow through**.

Can a compact, point-of-care GC-MS and machine learning detect lung cancer from breath, taking screening beyond the imaging center?

2 Study design & breath acquisition

Breath from **65 Moffitt patients** (lung cancer cases and lung cancer screening controls) was collected in parallel on two platforms: **CLARION**, Detect-ION's portable GC-MS, and sorbent tubes for laboratory **high-resolution mass spectrometry (HRMS)**.

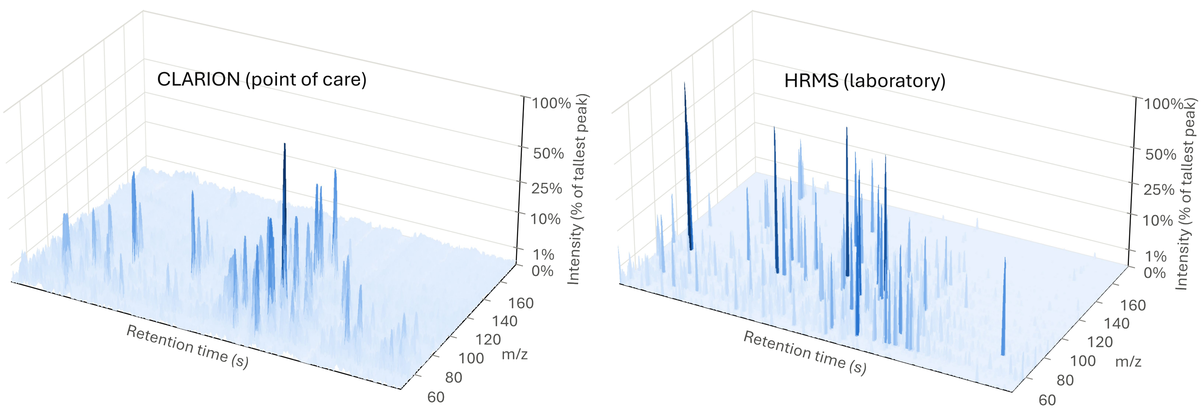
40 treatment-naïve NSCLC, confirmed by pathology

25 lung cancer screening controls, radiology-confirmed Lung-RADS 1 or 2

Daily room-air blanks for background subtraction



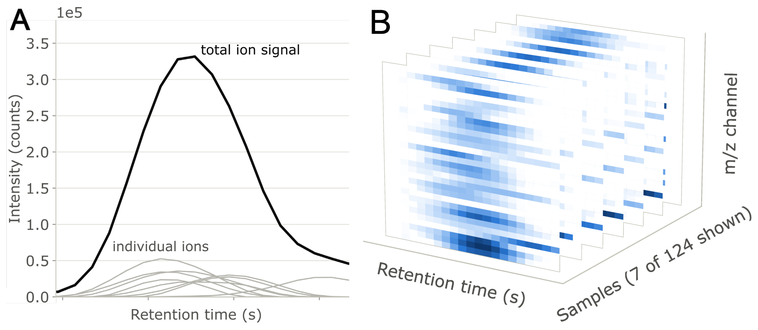




Ion intensity by retention time and m/z for one breath sample on CLARION (left) and HRMS (right). Despite different time and mass resolution, one quantification pipeline serves both.

3 From raw spectra to ML-ready features

Peak detection per sample → **Cross-sample alignment (time, m/z)** → **Consensus compounds** → **CP tensor decomposition**



Each retention window with signal (**A**) becomes a three-way tensor of ion intensities across m/z channels, retention times and samples (**B**).

Unsupervised **CP (PARAFAC) decomposition** writes the tensor as a sum of components (**C**): outer products of a compound's spectrum, elution profile and per-sample abundance. Abundances become VOC features (**D**); spectra are matched to the NIST library.

4 Machine-learning model development

VOC abundances from each platform were used as features to train interpretable classifiers for non-small cell lung cancer (**NSCLC**).

>100 VOCs · CLARION

>900 VOCs · HRMS

→ **Quality control**: Outlier removal (insufficient breath volume) → **Signal correction**: LOWESS detrending, batch-shift correction, background subtraction

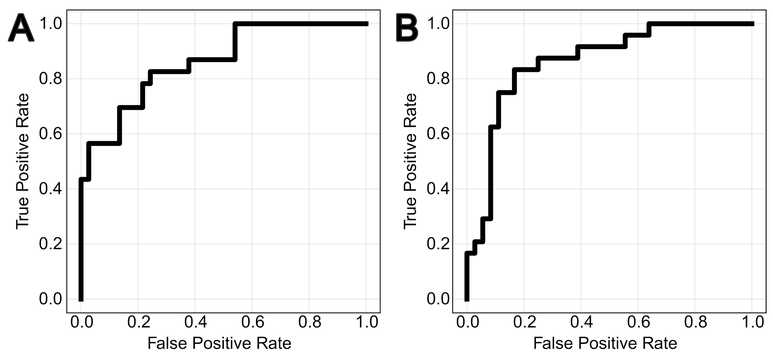
Elastic-net logistic regression with nested cross-validation

Fit within inner folds, evaluated on held-out outer folds (no data leakage):

- **AUC-based feature filter**
- **Centering & Pareto scaling**
- **Tune L1 ratio & regularization strength**

Models best separated **NSCLC from controls** (AUC ≈ 0.86 on both platforms; ~80% accuracy) and were weaker at separating early- from late-stage disease.

Comparison (n)	CLARION AUC (Sens/Spec/Acc %)	HRMS AUC (Sens/Spec/Acc %)
NSCLC (36) vs. controls (23)	**0.864** (82.6 / 75.7 / 78.3)	**0.863** (83.3 / 83.3 / 83.3)
Early NSCLC (23) vs. controls (23)	**0.877** (73.9 / 95.7 / 84.8)	**0.841** (73.9 / 87.5 / 80.9)
Late NSCLC (12) vs. controls (23)	**0.807** (71.4 / 82.6 / 78.4)	**0.677** (66.7 / 75.0 / 72.2)
Early (23) vs. late NSCLC (12)	**0.612** (64.3 / 69.6 / 67.6)	**0.696** (60.9 / 83.3 / 68.6)
Adenocarcinoma (27) vs. controls (23)	**0.768** (74.1 / 69.6 / 72.0)	**0.787** (70.4 / 87.5 / 78.4)
Squamous-cell carcinoma (7) vs. controls (23)	**0.754** (60.9 / 100.0 / 71.9)	**0.774** (57.1 / 91.7 / 77.4)



0.864 CLARION AUC — NSCLC vs. controls

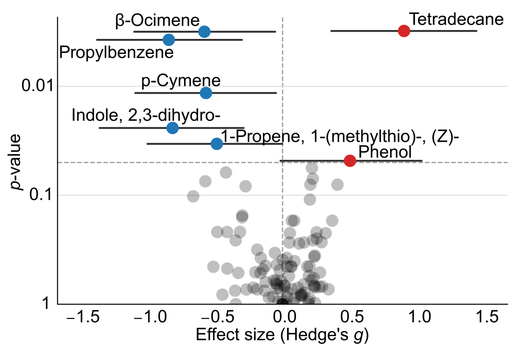
0.863 HRMS AUC — NSCLC vs. controls

*Cross-validated ROC curves, NSCLC vs. controls: CLARION (**A**) and HRMS (**B**). Point-of-care performance matches the laboratory reference.*

5 Interpretable AI: VOCs & transparency

Because model features are discrete, identified compounds, every prediction is explainable: we can (1) compare discriminative VOCs with prior literature and (2) trace each patient's model score to specific peaks in their chromatogram.

All **seven VOCs** most strongly associated with NSCLC on CLARION were also detected in paired HRMS data; **five** have been reported in prior cancer breath studies[5–7].

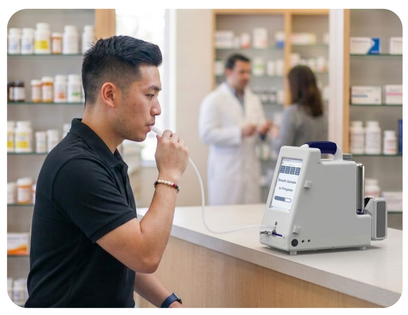


Mann–Whitney U p-values vs. effect size for CLARION VOCs (NSCLC vs. controls).

GC-MS data can support more complex models (e.g., deep learning on raw chromatograms), but explicitly quantifying VOC–disease relationships is scientifically informative and favors models that remain stable across study sites and patient conditions.

6 Conclusions & next steps

- **CLARION** and **HRMS** identify concordant breath VOC signatures predictive of NSCLC with comparable accuracy.
- Next: larger cohorts & subgroups, independent validation, & the key clinical gap, **indeterminate pulmonary nodules**, where a breath test could matter most.
- **A non-invasive, point-of-care breath test could reach people outside current guidelines & make screening easy enough that eligible people follow through.**



Funding: This research was funded by Detect-ION. KMP and MBS were supported in part by a Cancer Center Support Grant at the H. Lee Moffitt Cancer Center & Research Institute, an NCI-designated Comprehensive Cancer Center (P30 CA076292), and by an NCI Early Detection Research Network award (U01 CA200464 to MBS).

References

1. Bandi et al. Lung cancer deaths prevented and life-years gained from lung cancer screening. *JAMA* 2025;334(24):2225–2227.
2. Yang et al. Age-based screening for lung cancer surveillance in the US. *JAMA Netw Open* 2025;8(11):e2546222.
3. Hanna et al. Accuracy and methodologic challenges of volatile organic compound… *JAMA Oncol* 2019;5(1):e182815.
4. Saidi et al. Non-invasive prediction of lung cancer histological types through… *Sens Actuators B Chem* 2020;311:127932.
5. Monedeiro et al. Needle trap device-GC-MS for characterization of lung diseases based on breath… *Molecules* 2021;26(6):1789.
6. Smirnova et al. Predictive performance of selected breath volatile organic carbon… *Transl Lung Cancer Res* 2022;11(6):1009–1018.
7. Gashimova et al. Non-invasive exhaled breath and skin analysis to diagnose lung cancer… *ACS Omega* 2022;7(46):42613–42628.