Identification of Lung Cancer Breath Biomarkers using Infrared Cavity Ring-Down Spectroscopy

Robyn Larracy[1], Angkoon Phinyomark[1], Tony Reiman[2], Luisa Galvis[3], Mahmoud Abdelsalam[4], Stephen Graham[5], Chris Purves[5], Gisia Beydaghyan[5], Erik Scheme[1]

1 University of New Brunswick, Fredericton, NB, Canada, 2 Saint John Regional Hospital, Saint John, NB, Canada
3 Dr. Everett Chalmers Regional Hospital, Fredericton, NB, Canada, 4 Moncton Hospital, Moncton, NB, Canada, 5 Picomole, Moncton, NB, Canada

Motivation and Background

Nearly 80% of lung cancer cases are discovered when the disease has already metastasized and survival rates are significantly reduced. Screening is critical to catch the disease in its early, treatable stages. In this study, we consider a screening system based on cavity ring-down spectroscopy (CRDS) analysis of exhaled breath samples. We use machine learning to identify potential biomarkers and develop a classification system for differentiating lung cancer and non-lung cancer individuals.

Sample Collection

Ten-litre **alveolar breath** samples were collected onto Tenax TA sorbent tubes for 96 control subjects and 62 pre-treatment, non-small cell lung cancer patients using Picomole's proprietary breath sampler. Participants with other lung conditions such as COPD, asthma, pneumonia and bronchitis were included in both groups (36 control subjects and 44 lung cancer subjects).

Picomole Exhaled Breath Sampler

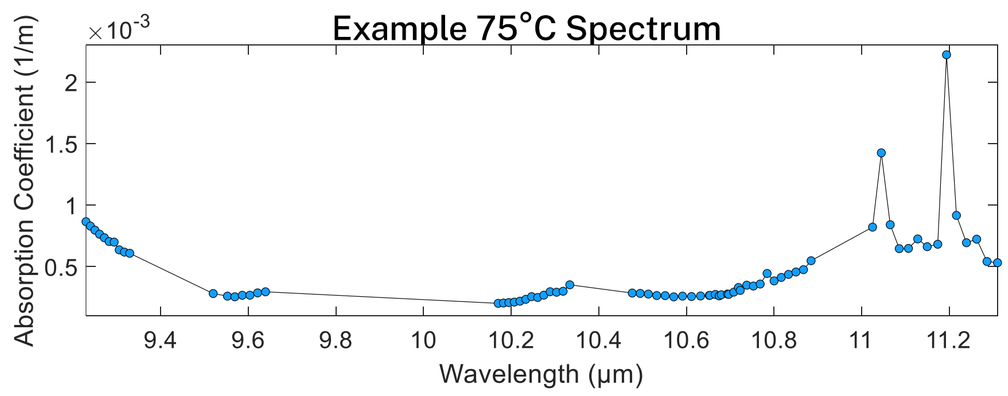
Group Demographics and Clinical Factors

Factor	Lung Cancer Cohort	Control Cohort
Sample Size	62	96
Sex		
Female	50%	53.1%
Male	50%	46.9%
Age (mean ± SD, years)		
Female (Range)	68.2 ± 9.1 (47-84)	61.0 ± 14.3 (22-83)
Male (Range)	71.3 ± 8.3 (57-83)	65.9 ± 12.1 (35-83)
Diagnosis		
Adenocarcinoma	58.1%	-
Squamous cell carcinoma	37.1%	
Unspecified NSCLC	4.8%	
Smoking		
Current Smokers	19.4%	6.7%
Former	75.8%	48.9%
Non-Smokers	4.8%	44.4%

CRDS Analysis and Spectral VOC Fitting

The **cavity ring-down spectroscopy analysis** was performed at four different desorption temperatures for each sample, 75°C, 150°C, 225°C and 300°C, yielding four spectra consisting of 73 CO2 (^{12}C and ^{13}C) wavelengths.

Stepwise spectral fitting was used to identify VOCs presented in each spectrum. Compound spectra from a reference library of **152 VOCs** were iteratively added and removed from each fit based on the Akaike Information Criterion.



Machine Learning Methods

Six different **data transformations** were considered for the VOC concentrations: $\log_e(X)$, $\log_e(X+1)$, $\log_{10}(X)$, $\log_2(X)$, $X^{1/2}$, and $X^{1/3}$.

For each data transformation, the top VOCs were selected using **minimum redundancy maximum relevance (mRMR) feature selection**. This algorithm ranks features sequentially based on a difference of Pearson correlations: the feature's correlation with the group membership labels minus its correlation with other features.

Linear support vector machines were trained and validated using **leave-one-subject-out cross-validation** to evaluate each set of top ranked mRMR features.

Results

The best leave-one-subject-out classification performance was achieved using **30 features (26 unique compounds)** with a **$\log_e(X)$ transform**: accuracy 85.44%, sensitivity 77.42%, specificity 90.63%, and AUC 0.858.

Untransformed Concentrations for Best Performing Feature Set

#	Feature	Lung Cancer (μ±σ ppbv)	Control (μ±σ ppbv)
1	Dimethyl sulfide (75°C)†	0.302 ± 0.747	0.024 ± 0.134
2	Octafluoropropane (300°C)	1.295 ± 4.091	5.879 ± 33.198
3	Isopropanol (150°C)*†	0.178 ± 0.555	0.023 ± 0.173
4	D-Limonene (150°C)	0.029 ± 0.162	0.353 ± 1.182
5	1,2-Dichlorobenzene (150°C)	0.021 ± 0.094	0 ± 0
6	Trifluoromethane (225°C)*	0.242 ± 1.012	0 ± 0
7	Menthol (150°C)	0.153 ± 0.859	0 ± 0
8	n-Undecane (300°C)†	0.770 ± 2.720	0 ± 0
9	Diethyl ketone (150°C)	0.397 ± 1.713	0 ± 0
10	2,3-Dimethylbutane (225°C)	1.569 ± 5.064	0.200 ± 1.180
11	2-Nonanone (75°C)†	0.150 ± 0.931	1.319 ± 4.315
12	n-Tetradecane (75°C)	0.132 ± 1.036	1.473 ± 4.351
13	Formic acid (75°C)	0.014 ± 0.089	0.054 ± 0.190
14	Furfural (300°C)	0.186 ± 0.561	0.054 ± 0.319
15	tert-Butyl alcohol (150°C)*	0.135 ± 0.245	0.072 ± 0.215
16	3-Carene (75°C)	0.076 ± 0.366	0 ± 0
17	Methanol (300°C)†	0.065 ± 0.174	0.511 ± 3.572
18	1-Undecene (150°C)	0.132 ± 0.444	0.025 ± 0.149
19	Ethyl butyrate (300°C)	0.274 ± 0.837	0.055 ± 0.280
20	Ethyl tert-butyl ether (225°C)*	0.034 ± 0.136	0.002 ± 0.016
21	Isopropanol (75°C)*†	1.667 ± 1.994	0.914 ± 1.767
22	Ethyl tert-butyl ether (75°C)*	0.323 ± 0.782	0.110 ± 0.385
23	2-Methyl-1-propanal (75°C)	0 ± 0	0.147 ± 0.581
24	1,2,3,5-Tetramethylbenzene (75°C)	0.458 ± 2.577	0 ± 0
25	Isopropylamine (75°C)†	0.638 ± 1.891	0.194 ± 1.178
26	Hexanal (300°C)†	0 ± 0	0.245 ± 1.145
27	Propylene sulfide (225°C)	0.034 ± 0.162	0 ± 0
28	1-Hexanoic acid (75°C)	0.152 ± 1.117	0.361 ± 1.050
29	Trifluoromethane (75°C)*	0.119 ± 0.935	1.218 ± 7.274
30	tert-Butyl alcohol (75°C)*	0.098 ± 0.197	0.037 ± 0.120

* VOC selected twice from different temperature spectra
† VOC found to differ between lung cancer and control subjects in other works

Conclusions

These preliminary results show promise for a novel exhaled breath-based screening system using infrared CRDS technology and spectral VOC fitting. We have previously built a model for this data using spectral patterns as classification features (rather than VOC concentrations), which achieved similar accuracy. Future work will compare the two approaches.

Reference

Reiman et al. Analysis of exhaled breath of lung cancer patients using infrared spectroscopy [abstr.]. In: *American Society of Clinical Oncology*. 2020.

Acknowledgements

Funding Partners

New Brunswick Health Research Foundation
Fondation de la recherche en santé du Nouveau-Brunswick
Mitacs Accelerate

Saint John Regional Hospital
Robert Thompson, Farah Naz, James Charles Roger Michael, Trisha Daigle-Maloney, Amanda Caissie, Holly Anne Campbell, Erin Brooks, Donna Fairweather

Dr. Everett Chalmers Hospital
Stella Park

Moncton Hospital
Joseph Ojah, Maged Salem, Judit Kokai, Ian Chute, Trina Mullin