

PAPER • OPEN ACCESS

Exhaled breath analysis in patients with potentially curative lung cancer undergoing surgery: a longitudinal study

To cite this article: Jonas Herth *et al* 2024 *J. Breath Res.* **18** 036003

View the [article online](#) for updates and enhancements.

You may also like

- [Exhaled breath condensate biomarkers for lung cancer](#)
Annalisa Campanella, Simona De Summa and Stefania Tommasi
- [Integration of electronic nose technology with spirometry: validation of a new approach for exhaled breath analysis](#)
R de Vries, P Brinkman, M P van der Schee *et al.*
- [Volatile signature for the early diagnosis of lung cancer](#)
Roberto Gasparri, Marco Santonico, Claudia Valentini *et al.*

Discover breath analysis in

The Breath Biopsy® Guide

Fourth edition

DOWNLOAD THE FREE E-BOOK

BREATH BIOPSY

DOWLSTONE MEDICAL



PAPER

OPEN ACCESS

RECEIVED
4 February 2024REVISED
22 April 2024ACCEPTED FOR PUBLICATION
8 May 2024PUBLISHED
17 May 2024

Original content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.



Exhaled breath analysis in patients with potentially curative lung cancer undergoing surgery: a longitudinal study

Jonas Herth^{2,*} , Felix Schmidt^{1,2}, Sarah Basler^{1,2}, Noriane A Sievi² and Malcolm Kohler^{1,2}¹ Faculty of Medicine, University of Zurich, 8032 Zurich, Switzerland² Department of Pulmonology, University Hospital Zurich, 8091 Zurich, Switzerland

* Author to whom any correspondence should be addressed.

E-mail: jonas.herth@usz.ch**Keywords:** pulmonary oncology, lung cancer, diagnosis, exhaled breath analysis, early detection, volatile organic compoundsSupplementary material for this article is available [online](#)

Abstract

Exhaled breath analysis has emerged as a non-invasive and promising method for early detection of lung cancer, offering a novel approach for diagnosis through the identification of specific biomarkers present in a patient's breath. For this longitudinal study, 29 treatment-naïve patients with lung cancer were evaluated before and after surgery. Secondary electrospray ionization high-resolution mass spectrometry was used for exhaled breath analysis. Volatile organic compounds with absolute \log^2 fold change ≥ 1 and q -values ≥ 0.71 were selected as potentially relevant. Exhaled breath analysis resulted in a total of 3482 features. 515 features showed a substantial difference before and after surgery. The small sample size generated a false positive rate of 0.71, therefore, around 154 of these 515 features were expected to be true changes. Biological identification of the features with the highest consistency (m/z –242.18428 and m/z –117.0539) revealed to potentially be 3-Oxotetradecanoic acid and Indole, respectively. Principal component analysis revealed a primary cluster of patients with a recurrent lung cancer, which remained undetected in the initial diagnostic and surgical procedures. The change of exhaled breath patterns after surgery in lung cancer emphasizes the potential for lung cancer screening and detection.

1. Introduction

Lung cancer is the leading cause of cancer deaths in the world [1]. Mostly, lung cancer is not detected until it is at an advanced stage and more than half of the patients are diagnosed at the most advanced stage IV when curative therapy is usually not possible anymore [2]. Therefore, early detection of lung cancer is of fundamental importance in cancer treatment given a drop in the 5 year survival rate from 70%–90% in stage I to 0%–10% in stage IV [3]. Lung cancer in potentially curative stages does typically not cause symptoms and even if symptoms are present they are non-specific which makes diagnosis challenging [4].

Several lung cancer screening methods were proposed such as high-resolution computed tomography (HRCT), chest x-ray, screening with biomarkers, and biopsy techniques. However, the possible advantages of lung cancer screening have to

be balanced against the potential of inducing harm, over-diagnosing, cost-effectiveness, or the incapability of early diagnosis [5]. Despite the fact that lung cancer screening with computed tomography has been proven to substantially decrease lung cancer specific mortality, its application is challenging in multiple aspects such as variability in radiological standards, the pre-selection of at-risk populations, false positive rates, and over-diagnosing [5–8]. Therefore, new methods for simple, accurate, and non-invasive diagnosis of lung cancer, especially in curative stages, would be highly desirable.

Exhaled breath analysis is considered a promising method for lung cancer screening. The concentrations of volatile organic compounds (VOCs) in exhaled breath has shown to be representative of its blood concentration, reflecting ongoing internal metabolic processes [9]. In pathological conditions, VOCs in exhaled breath compounds are altered due to the shift in cell metabolism and can

be directly linked to diseases [10]. Various potential VOCs from breath analysis in lung cancer screening are postulated, however, high sensitivity and specificity has to be reached and validation studies are missing. Furthermore, most studies used methods such as proton transfer reaction mass spectrometry or gas chromatography, which are less attractive for clinical practice due to the need for sample preparation or limited sensitivity and specificity. Therefore, the translation of research into the clinical environment has not taken place yet. The most recent developed on-line breath measurement is secondary electrospray ionization high-resolution mass spectrometry (SESI-HRMS) [11–13] which enables breath measurements in a clinical setting without sample preparation and offers one of the largest range of detectable masses, thus allowing biomarker identification [14].

The objective of this study was to obtain lung cancer specific VOCs with SESI-HRMS in patients with potentially curative lung cancer undergoing surgery. By using a longitudinal study design with consistent measurement conditions, we aim to enhance the likelihood of detecting cancer-related VOCs following tumor removal. We hypothesize that the complete removal of lung cancer will lead to alterations in lung cancer specific VOCs potentially enabling improved cancer screening.

2. Methods

2.1. Study design and participants

For this longitudinal study, 29 treatment-naive patients with lung cancer were recruited during hospital stay at the University Hospital of Zurich between March 2020 and January 2023.

For inclusion, patients had to (i) have a high suspicion for lung cancer (ii) undergo a lung cancer surgery with curative intention (UICC stage I and II) (iii) be ≥ 18 years old (iiii) have signed the informed consent. Exclusion criteria were (i) other active carcinoma (ii) other lung diseases except of chronic obstructive pulmonary disease (iii) renal failure (glomerular filtration rate ≤ 15 ml min^{-1}). Online real-time exhaled breath measurement was performed for each cancer patient at two study visits. The first visit was within one week before surgery and the second visit was performed no earlier than one month after the surgery (figure 1). The local ethical committee approved the study protocol (KEK-ZH 2015–0607). The experiments were conducted in accordance with the Declaration of Helsinki, principles of Good Clinical Practice and written informed consent was obtained from all participants before participation. The clinical trial was registered at ClinicalTrials.gov (NCT02781857). Results were reported according to STROBE guidelines.

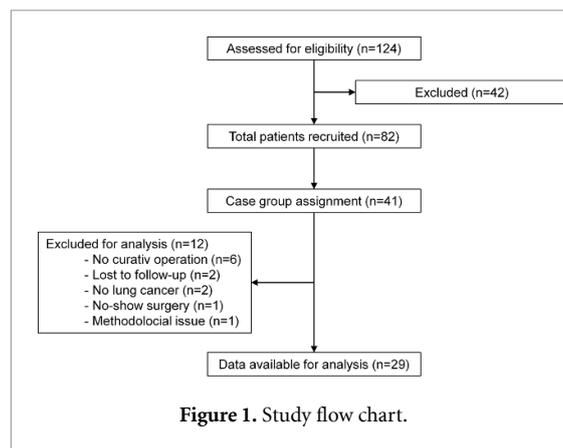


Figure 1. Study flow chart.

2.2. Study design and participants

To ensure proper measurement, participants were asked to abstain from drinking (except water), eating, chewing gum, brush their teeth, and smoking for at least one hour prior to the measurement. The analytical platform utilized in this study consisted of a SUPER-SESI source (Fossil Ion Technology, Spain) coupled to a Q Exactive Plus HRMS (Thermo Fisher Scientific, Germany). To standardize exhaled breath volume and flow, a capnograph (Exhalion, Fossil Ion Technology, Spain) was used. Full scan mass spectra were recorded in both positive and negative ion mode with an accumulation time of 1.2 s within the 70–1000 Dalton range. The patients set upright in front of the mass spectrometer and underwent twelve consecutive exhaled breath measurements using a sterile filter (MicroGard II, Vyair, Germany). To ensure data comparability throughout the study all measurements were taken in the same environment, daily mass spectrometer calibrations were performed, and instrument operators adhered to internal standard operating procedures.

2.3. Data processing and analysis

The dataset utilized in this analysis consisted of clinical metadata and raw mass spectrometry breath data obtained from 29 participants, resulting in a total of 58 breath measurements. Each measurement contained a mass spectrometry file in both positive and negative mode, as well as corresponding Exhalion measurement files capturing carbon dioxide data. The breath measurements were categorized into two groups within the study setup (i) breath measurements from lung cancer patients before surgery (ii) breath measurements from the same lung cancer patients after surgery. The mass spectra data were directly retrieved from the RAW files using in-house software based on RawFileReader from Thermo Fisher Scientific. The measurement data were preprocessed using deep breath intelligences (DBI) proprietary pipeline, resulting in a data

matrix comprising 3482 features. The features corresponded to the signal intensity of 2067 positive and 1415 negative mode mass-to-charge ratios (m/z) derived from the mass spectrum. A fold-change analysis was conducted to examine the differences in feature intensities before and after surgery. One- and two-sample t -tests (within and between groups) were employed to assess whether the mean of the sample significantly differ from the expected mean. Subsequent estimate of the false discovery rate for multiple hypothesis testing was performed by computing q -values for each p -value as instructed by Storey [15]. Statistical significance level was set to $\alpha < 0.05$. The features with absolute \log^2 fold change greater than 1 and q -values greater than 0.71 (equivalent to p -value < 0.20 , since the small samples size prohibited q -values < 0.05 after multiple testing) were selected as features of interest. These potentially relevant features were displayed in a volcano plot. To quantify the consistency of changes, a score for each feature of interest was calculated by taking the \log^2 transformation of the ratio between the count of patients with increased feature intensity values at visit 2 and the count of patients with decreased feature intensity values at visit 2. Features with higher number of patients showing increased values in visit 2 have higher positive consistency scores, while those with more patients displaying decreased values have higher negative consistency scores. Features with an equal number of patients showing both increased or decreased values have scores close to zero, suggesting a lack of consistency in observed changes. Principal component analysis (PCA) of the features of interest with \log^2 fold changes ≥ 1 and q -value > 0.71 was performed to identify a potential underlying structure of the data. To assess the potential correlation between the features of interest and the operational procedure itself, a correlation of time since surgery and \log^2 fold changes was conducted for each feature. This analysis involved categorizing patients into three groups based on the time intervals between their first and second visit: less than 6 weeks, less than 12 weeks, and over 12 weeks, respectively. MetaboAnalystR (version 2.0.4) was used to translate all the measured features into metabolic pathways and a biological context. A dedicated SESI-HRMS database from DBI was used to assign possible sum formulas to exact mass features. Derivatized formulas are computationally generated and ranked based on valence and elemental ratio check in combination with the original MS-FINDER formula scoring algorithm. This database is based on Fiehn's seven golden rules [16].

A power analysis was conducted to determine the sample size needed to achieve the desired power level. To generate simulated datasets, the mean and standard deviation of a specific feature out of this study was utilized, generating 1000 datasets from a normal distribution. This simulation was repeated 146

times, gradually increasing the sample size from 5 to 150. For each generated dataset of every feature, the one-sample t -test was employed to calculate the p -value, maintaining the null hypothesis that there is no significant difference between the sample mean and the expected population mean. C#, MATLAB, and Python were used for data processing and data analysis.

3. Results

3.1. Patient characteristics

A total of 29 participants were included in this study. All subjects were of Caucasian ethnicity and the mean (SD) age of patients was 65 (9.8) years. The proportions of active smokers, former smokers, and never-smokers were 30%, 53%, and 14% (9, 16, 4 patients, respectively) and mean (SD) pack years was 32 (22) (table 1). The mean (SD) duration between visit 1 and visit 2 was 83.7 (51.5) days.

The most common histological type of cancer found in the cancer patients was adenocarcinoma (22/30 patients), followed by squamous cell carcinoma (5/30). Most of the cancers were found in the upper lobe of the lung and the most common Union for International Cancer Control (UICC) classification was IAIII (table 2). Three patients developed a recurrent or secondary lung cancer (LC001, LC176, and LC185) at the time of data analysis.

3.2. Potential change of exhaled breath in patients after lung cancer surgery

Exhaled breath data, obtained from 29 patients before and after lung cancer surgery, resulted in a total of 3482 breath features. The p -values and q -values for the difference between pre- and post-surgery of each feature can be found in the supplements (table S1) as well as the distribution of fold-changes and one-sample t -test (figures S1 and S2). The lowest q -value obtained in our analysis was 0.7082, which exceeded the standard cut-off value of 0.05. The distribution of the q -values revealing a peak of 828 features with q -values around 0.70. (figure S3) Among these 828 features, 248 features (30%) were suggested to be true positive discoveries. The volcano plot with the 3482 breath features plotted showed 515 features with a fold change greater than or equal to 2 and q -value of ≤ 0.71 (figure 2).

These highlighted features (features of interest) represented a substantial fold change before and after surgery and a relevant difference between pre- and post surgery greater than the selected cut-off q -value of 0.71. Due to the false positive rate of 0.71, around 154 of these features were expected to show true changes after lung cancer surgery. None of these features showed a correlation with time between cancer surgery and follow-up.

Table 1. Patients characteristics.

Age, years	65 (9.8)
Male, N (%)	13 (45)
BMI, kg m ²	26.6 (5.7)
Smoking history, N (%)	
Never smoker	4 (14)
Ex-smoker	16 (53)
Current smoker	9 (30)
Pack years, py	32 (22)
Lung diseases, N (%)	
COPD	8 (28)
OSA	3 (10)
Other	2 (7)
CVD, N (%)	10 (34)
GERD, N (%)	0 (0)
CKD, N (%)	3 (10)
Diabetes mellitus, N (%)	1 (3)

Values are presented as mean (standard deviation) unless otherwise stated. Abbreviations: N, number; COPD, chronic obstructive pulmonary disease; OSA, obstructive sleep apnea; CVD, cardiovascular disease; GERD, gastroesophageal reflux disease; CKD, chronic kidney disease.

Table 2. Lung cancer characteristics.

Type, N (%)	
Adenocarcinoma	21 (72)
Squamous cell carcinoma	5 (17)
NUT carcinoma	2 (7)
NET	1 (3)
UICC classification, N (%)	
IAI	4 (14)
IAII	9 (31)
IAIII	2 (7)
IIA	3 (10)
IBI	7 (24)
IIBI	4 (14)
Localization, N (%)	
RUL	10 (35)
RML	1 (3)
RLL	6 (21)
LUL	9 (31)
LLL	3 (10)

Abbreviations: NET, lung neuroendocrine tumor; UICC, Union internationale contre le cancer; RUL, right upper lobe; RML, right middle lobe; RLL, right lower lobe; LUL, left upper lobe; LLL left lower lobe.

3.3. Consistency of change in features

Consistency of change analysis was performed across all patients for each feature of interest (table S1). The analysis revealed the increasing features m/z -95.98022 and -242.18428 and the decreasing features m/z -117.0539 and 122.01998 with the highest consistency scores (figures 3(A)–(D)).

Biological identification of these features revealed m/z -242.18428 to potentially be 3-Oxotetradecanoic acid and m/z -117.0539 to be indole (table 3). No sum formula for m/z -95.98022 and m/z 122.01998 could be determined.

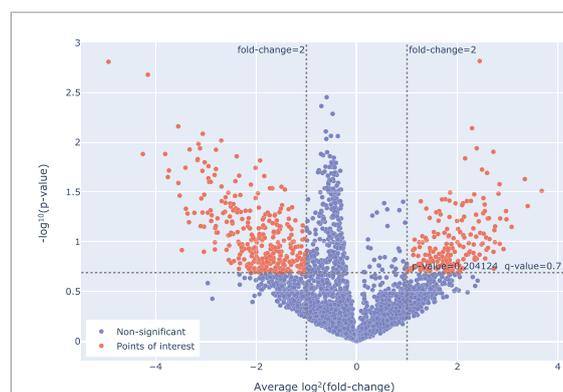


Figure 2. Volcano plot of high potential features. Each point represents a feature. The x-axis is the mean of the \log^2 (fold-change) of a feature, the y-axis is the $-\log^{10}$ (p -value) of the corresponding fold-change from one-sample's t -test. The vertical cut off is a two-fold change. The horizontal cut off is q -value of 0.71 (or the corresponding p -value of 0.204). The orange points are the potential features of interest, with a change in signal intensity more than two-folds between pre- and post-surgery and a q -value for multiple testing greater or equal to 0.71. The blue points are considered noise or non-significant.

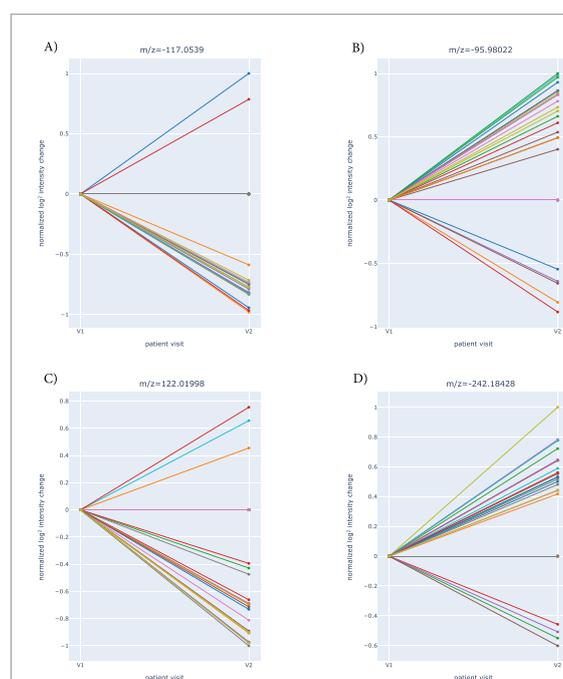


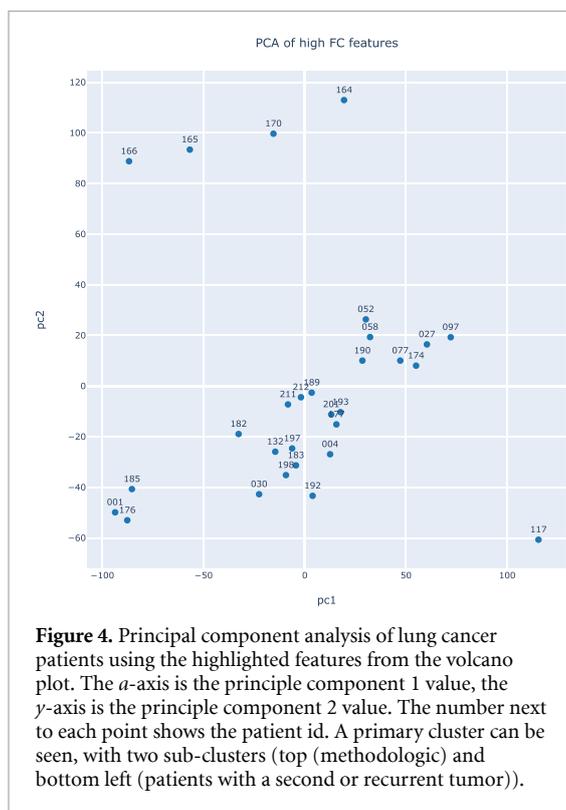
Figure 3. Change of (A) $m/z = -117.0539$ (B) $m/z = -95.98022$ (C) $m/z = 122.01998$ (D) $m/z = -242.18428$ intensity between visit 1 and visit 2. The y-axis represents the min-max normalized \log^2 transformation of the visit 2 values subtracted by the visit 1 values. The transformation range is set between 0 and 1, with negative values on the y-axis indicating a decrease in value.

3.4. Principle component analysis (PCA)

PCA of the features of interest from the volcano plot showed that the first two principal components (PC) explain 34% of the variance (figure 4). Pre- and post-surgery status cannot be differentiated by inclusion of all features of interest. However, a main cluster can

Table 3. Chemical compound identification.

Mass	Change	Sum formula	Putative Molecule
-117.0539	Down	[13C]C7H7N	Indole
-95.98022	Up	Not available	
122.01998	Down	Not available	
-242.18428	UP	[13C]C13H26O3	3-Oxotetradecanoic acid



be observed with two additional distinct clusters separated from the main cluster (Cluster 1 with patient 166, 165, 170, and 164; Cluster 2 with patient 001, 176, and 185). UICC classification and cancer type for each patient are shown in table S2.

3.5. Power analysis

A post-hoc power analysis was conducted to determine the sample size needed to achieve statistical significance for future studies. Therefore, a power analysis plot was generated by computing the mean across all 515 features identified in the volcano plot analysis. It revealed that a dataset comprising at least 120 patient pairs is necessary to achieve a desired power level of 0.9 (figure S4).

4. Discussion

In this study, we were able to measure 154 breath features that change upon lung cancer surgery and thus may be related to lung cancer itself.

Detection of lung cancer poses a critical clinical challenge, with HRCT, magnetic resonance imaging (MRI), and positron emission tomography-computed tomography (PET-CT) being the primary imaging modalities used. Despite the fact that the detection rate of curative lung cancer stages and the sensitivity of used methods improved over the last decade, there are still numerous challenges faced. HRCT offers high resolution but struggles with distinguishing benign from malignant lesions [17]. MRI, with its superior soft tissue contrast, faces challenges in detecting smaller lesions, while PET-CT may yield a substantial rate of false positive results due to underlying inflammatory processes. Furthermore, there are challenges related to tumor heterogeneity, overdiagnosis, motion artifacts, radiation exposure, costs, accessibility, and the integration of multi-modal data [18–20].

On-line mass spectrometry with SESI-HRMS could be a promising method fulfilling the challenged points and might be implemented in clinical practice improving early detection, precise diagnosis, and ultimately, better patient outcomes. It is non-invasive, real-time, and its high sensitivity and specificity have already been shown in various diseases such as chronic obstructive pulmonary disease and breast cancer [21, 22]. The analysis of VOCs seems to be promising since they emanate from cancer cells and their micro-environment and represent alterations in cell metabolism that lead to local and systematic shifts in the production of VOCs. Therefore, it seems reasonable to hypothesize that these shifts could be detected in potentially curative cancer stages [23, 24]. Previous studies have indeed established correlations between VOCs and the pathophysiology of lung cancers [10, 23]. Wang *et al* found 16 VOCs in a targeted analysis during peri-operative exhaled breath analysis in patients with lung cancer with high-pressure photon ionization time-of-flight mass spectrometry and validation confirmed good performance [25]. In a multicenter, multi device study including 575 subjects, Kort *et al* could distinguish patients with lung cancer from subjects without lung cancer using an electronic nose, with an area under the receiver-operating characteristic curve of 0.86 [26]. Despite decades of research, variability in sample collection, testing conditions, and data analysis poses significant challenges among these studies to the widespread adoption of exhaled breath analysis

in clinical settings. However, SESI-MS shows great potential in simplifying exhaled breath analysis processes and enhancing the reliability of exhaled breath analysis outcomes.

To our knowledge, this is the first study measuring VOCs with untargeted exhaled breath analysis in lung cancer patients showing 515 features with relevant alterations before and after surgery. Of those features 154 are expected to be possibly lung cancer related features. However, the sample size is too low for identification of these 154 features and exact biological interpretation remains challenging. Therefore, consistency of change analysis was used to identify features that display both relevant changes and consistent patterns. It revealed m/z -95.98022 , m/z -242.18428 , m/z -117.0539 and m/z 122.01998 as the most consistent features. Using a specialized SESI-HRMS database one can assume that m/z -117.0539 is Indol. Indoles are products from tryptophan metabolism and important endogenous ligands for the aryl hydrocarbon receptor (AhR), enables AhR to go into the nucleus, where it combines with the AhR nuclear translocator (Arnt). Together, they attach to specific DNA sequences known as xenobiotic response elements (XRE) in the regulatory areas of certain genes. These genes, particularly those linked to dealing with foreign substances (xenobiotics), get activated. Some of these genes are responsible for producing enzymes that help break down and eliminate potential carcinogens or toxins [27, 28]. This is in line with our data since indol was upregulated at baseline (before cancer surgery). Over 90% of tryptophan degradation takes place through the kynurenine pathway, where it undergoes downstream metabolism into kynurenine, 3-hydroxyanthranilic acid (3-HAA), quinolinic acid (QA) and nicotinicamide adenine dinucleotide [29]. Interestingly 3-HAA is reported to be increased in non-small-cell lung cancer patients. This finding is supported by our data, as we identified the precise mass of 3-HAA and QA in both positive and negative modes, observing $[M + H]^+$ and $[M - H]^-$ adducts. The intensities of 3-HAA exhibit an increase prior to surgery and a subsequent decrease post-surgery, aligning with previous research findings [30]. The literature provides substantial evidence for the connection between cancer and tryptophan metabolites such as 3-HAA and QA. Both of these metabolites have been reported to contribute to immune evasion in lung cancer by inducing apoptosis in T-cells, thereby promoting immune tolerance in cancer cells [30, 31]. The feature with the mass m/z -242.18428 may well be 3-Oxotetradecanoic acid, a byproduct of fatty acid oxidation, also known as beta-oxidation [32]. Beta-oxidation is a fundamental process in cellular metabolism, contributing to the efficient utilization of fatty acids for energy production. Cancer cells, however, reactivate de-novo lipogenesis, which removes their reliance on externally derived lipids and

allows them to proliferate at a faster rate. Our results are in line with this mechanism, as the decrease in beta-oxidation at baseline correlates with a reduction in the levels of 3-Oxotetradecanoic acid in exhaled breath, and these levels subsequently increase following tumor removal [33].

PCA of all features of interest was performed to identify the underlying structure of the data and potential subgroups and outliers within the group of patients. Although we were unable to differentiate between pre- and post-surgery status it revealed a primary cluster, with two notable sub-clusters (top and bottom left) and one outlier situated at the bottom right, distinctly separated from the primary cluster (figure 4). It is remarkable that all patients in the lower left cluster had developed a recurrent or secondary lung cancer in the meantime, which remained undetected in the initial diagnostic and surgical procedures. Patients might already have had secondary undetected lung cancer during the measurements, recurring cancer following surgical resection, or tumor cells spread through the peripheral blood as circulating tumor cells and therefore might still be detected in exhaled breath [34]. The pattern noticed in the upper sub-group and the outlier in the lower right corner could be linked to methodological factors, as they were all assessed during the same month.

Although our results suggest a difference in molecular breath pattern before and after lung cancer surgery, the differences did not reach statistical significance due to the small sample size. Therefore, we generated simulated datasets out of the present study to determine the sample size needed to achieve a power level of 90% that resulted in a sample size of at least 120 patient pairs. Therefore, further studies with larger sample sizes are needed to confirm our preliminary findings and identify the features with a true positive change as relevant biomarkers for lung cancer. In our study, we did not differentiate between various cancer histologies and phenotypes, which may have limited the findings. Therefore, a viable approach could be to narrow the focus to specific histological types and phenotypes in further research. Another limitation is that we cannot conclude with absolute certainty that none of the molecular changes are related to the surgery itself. However, as we did not observe any correlation over time, we assume no direct impact of the surgery procedure. Furthermore, as we observed no elevated inflammatory parameters as markers of long-lasting processes released by surgery, we do not conclude that the detected changes in VOCs are linked to inflammation.

5. Conclusions

Our findings indicate a variance in the molecular breath pattern before and after lung cancer surgery. PCA reveal a cluster of patients with recurrent or

secondary lung cancer, which remained undetected in the initial diagnostic and surgical procedures. The further investigation of exhaled breath analysis in lung cancer patients is worthwhile, as our findings have revealed that exhaled breath patterns undergo alterations even in potentially curative stages of lung cancer, offering promising possibilities for early detection and improved clinical care.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Author contributions

Conceptualization, Malcolm Kohler and Felix Schmidt.; methodology, Noriane Sievi; software, Felix Schmidt.; validation, Felix Schmidt, Sarah Basler and Jonas Herth; formal analysis, Noriane Sievi.; investigation, Sarah Basler, Felix Schmidt, Jonas Herth.; resources and data curation, Noriane Sievi, Jonas Herth and Sarah Basler; writing—original draft preparation, Jonas Herth.; writing—review and editing Jonas Herth, Noriane Sievi, Sarah Basler and Malcolm Kohler.; visualization, Sarah Basler.; supervision, Malcolm Kohler.; project administration, Malcolm Kohler.; funding acquisition, Malcolm Kohler and Noriane Sievi. All authors have read and agreed to the published version of the manuscript.”

Funding

This research was funded by the Lotte und Adolf Hotz-Sprenger Stiftung.

Institutional review board statement

The local ethical committee approved the study protocol (KEK-ZH 2015-0607). The experiments were conducted in accordance with the Declaration of Helsinki, principles of Good Clinical Practice and written informed consent was obtained from all participants before participation. The clinical trial was registered at ClinicalTrials.gov (NCT02781857). Results were reported according to STROBE guidelines.

Informed consent statement

Informed consent was obtained from all subjects involved in the study.

Conflict of interest

Malcolm Kohler is Board member of Deep Breath Intelligence AG, a company that provides services in the field of breath analysis. Felix Schmidt is employee

of Deep Breath Intelligence AG. J H, S B, N S declare no conflict of interest.

ORCID iD

Jonas Herth  <https://orcid.org/0000-0002-3390-5741>

References

- [1] Chen H H, Zhang T-N, Wu Q-J, Huang X-M and Zhao Y-H 2021 Circular RNAs in lung cancer: recent advances and future perspectives *Front. Oncol.* **11** 664290
- [2] Walters S et al 2013 Lung cancer survival and stage at diagnosis in Australia, Canada, Denmark, Norway, Sweden and the UK: a population-based study, 2004–2007 *Thorax* **68** 551–64
- [3] Torre L A, Bray F, Siegel R L, Ferlay J, Lortet-Tieulent J and Jemal A 2015 Global cancer statistics, 2012 *CA Cancer J. Clin.* **65** 87–108
- [4] Ruano-Ravina A et al 2020 Lung cancer symptoms at diagnosis: results of a nationwide registry study *ESMO Open* **5** e001021
- [5] Vansteenkiste J, Dooms C, Mascaux C and Nackaerts K 2012 Screening and early detection of lung cancer *Ann. Oncol.* **23** x320–7
- [6] Welch H G and Black W C 2010 Overdiagnosis in cancer *J. Natl Cancer Inst.* **102** 605–13
- [7] Gao W, Wen C P, Wu A and Welch H G 2022 Association of computed tomographic screening promotion with lung cancer overdiagnosis among Asian women *JAMA Intern. Med.* **182** 283–90
- [8] Migliore M, Palmucci S, Nardini M and Basile A 2020 Imaging patterns of early stage lung cancer for the thoracic surgeon *J. Thorac. Dis.* **12** 3349–56
- [9] Wallace M A G and Pleil J D 2018 Evolution of clinical and environmental health applications of exhaled breath research: review of methods and instrumentation for gas-phase, condensate, and aerosols *Anal. Chim. Acta* **1024** 18–38
- [10] Tsou P H et al 2021 Exploring volatile organic compounds in breath for high-accuracy prediction of lung cancer *Cancers* **13** 1431
- [11] Singh K D et al 2019 Standardization procedures for real-time breath analysis by secondary electrospray ionization high-resolution mass spectrometry *Anal. Bioanal. Chem.* **411** 4883–98
- [12] Hengerics Szabo A et al 2015 The analysis of linear and monomethylalkanes in exhaled breath samples by GCxGC-FID and GC-MS/MS *J. Chromatogr. B* **978–979** 62–69
- [13] Fernandez Del Rio R, O'Hara M E, Pemberton P, Whitehouse T and Mayhew C A 2016 Elimination characteristics of post-operative isoflurane levels in alveolar exhaled breath via PTR-MS analysis *J. Breath Res.* **10** 046006
- [14] Zeng J, Christen A, Singh K D, Frey U and Sinues P 2022 Comparison of plasma ionization- and secondary electrospray ionization high-resolution mass spectrometry for real-time breath analysis *Chimia* **75** 127–32
- [15] Storey J D 2002 A direct approach to false discovery rates *J. R. Stat. Soc. B* **64** 479–98
- [16] Kind T and Fiehn O 2007 Seven golden rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry *BMC Bioinform.* **8** 105
- [17] Li F, Aoyama M, Shiraishi J, Abe H, Li Q, Suzuki K, Engelmann R, Sone S, MacMahon H and Doi K 2004 Radiologists' performance for differentiating benign from malignant lung nodules on high-resolution CT using computer-estimated likelihood of malignancy *AJR Am. J. Roentgenol.* **183** 1209–15

- [18] Philip B, Jain A, Wojtowicz M, Khan I, Voller C, Patel R S K, Elmahdi D and Harky A 2023 Current investigative modalities for detecting and staging lung cancers: a comprehensive summary *Indian J. Thorac. Cardiovasc. Surg.* **39** 42–52
- [19] Jassim M M and Jaber M M 2022 Systematic review for lung cancer detection and lung nodule classification: taxonomy, challenges, and recommendation future works *J. Intell. Syst.* **31** 944–64
- [20] Bak S H, Kim C, Kim C H, Ohno Y and Lee H Y 2022 Magnetic resonance imaging for lung cancer: a state-of-the-art review *Precis. Future Med.* **6** 49–77
- [21] Martinez-Lozano Sinues P, Meier L, Berchtold C, Ivanov M, Sievi N, Camen G, Kohler M and Zenobi R 2014 Breath analysis in real time by mass spectrometry in chronic obstructive pulmonary disease *Respiration* **87** 301–10
- [22] Martinez-Lozano Sinues P, Landoni E, Miceli R, Dibari V F, Dugo M, Agresti R, Tagliabue E, Cristoni S and Orlandi R 2015 Secondary electrospray ionization-mass spectrometry and a novel statistical bioinformatic approach identifies a cancer-related profile in exhaled breath of breast cancer patients: a pilot study *J. Breath Res.* **9** 031001
- [23] Hakim M, Broza Y Y, Barash O, Peled N, Phillips M, Amann A and Haick H 2012 Volatile organic compounds of lung cancer and possible biochemical pathways *Chem. Rev.* **112** 5949–66
- [24] Nakhleh M K *et al* 2017 Volatolomics of breath as an emerging frontier in pulmonary arterial hypertension *Eur. Respir. J.* **49** 1601897
- [25] Wang P, Huang Q, Meng S, Mu T, Liu Z, He M, Li Q, Zhao S, Wang S and Qiu M 2022 Identification of lung cancer breath biomarkers based on perioperative breathomics testing: a prospective observational study *EClinicalMedicine* **47** 101384
- [26] Kort S *et al* 2023 Diagnosing non-small cell lung cancer by exhaled breath profiling using an electronic nose: a multicenter validation study *Chest* **163** 697–706
- [27] Hubbard T D, Murray I A and Perdeu G H 2015 Indole and tryptophan metabolism: endogenous and dietary routes to Ah receptor activation *Drug. Metab. Dispos.* **43** 1522–35
- [28] Li Y, Li X and Sarkar F H 2003 Gene expression profiles of I3C- and DIM-treated PC3 human prostate cancer cells determined by cDNA microarray analysis *J. Nutrition* **133** 1011–9
- [29] Modoux M *et al* 2020 Tryptophan metabolism as a pharmacological target *Cell* **42** 60–73
- [30] Li C and Zhao H 2021 Tryptophan and its metabolites in lung cancer: basic functions and clinical significance *Front. Oncol.* **11** 707277
- [31] Abd El-Fattah E E 2022 IDO/kynurenine pathway in cancer: possible therapeutic approaches *J. Transl. Med.* **20** 347
- [32] Colepicolo P *et al* 2019 Piper anisum as a promising new source of bioactive metabolites *Chem. Pap.* **74** 1505–15
- [33] Schilero C and Firestein B L 2021 Mechanisms of metabolic reprogramming in cancer cells supporting enhanced growth and proliferation *Cells* **10** 1056
- [34] Choi H and Hwang W 2022 Perioperative inflammatory response and cancer recurrence in lung cancer surgery: a narrative review *Front. Surg.* **9** 888630