

Aberystwyth University

Metabolomic-based biomarker discovery for non-invasive lung cancer screening O'Shea, Keiron; Cameron, Simon J. S.; Lewis, Keir E.; Lewis, Paul D.; Lu, Chuan; Mur, Luis A. J.

Published in:

Biochimica et Biophysica Acta (BBA) - General Subjects

DOI:

10.1016/j.bbagen.2016.07.007

Publication date: 2016

Citation for published version (APA):

O'Shea, K., Cameron, S. J. S., Lewis, K. E., Lewis, P. D., Lu, C., & Mur, L. A. J. (2016). Metabolomic-based biomarker discovery for non-invasive lung cancer screening: A case study. *Biochimica et Biophysica Acta (BBA)* - *General Subjects*, *1860*(11 (Part B)), 2682-2687. https://doi.org/10.1016/j.bbagen.2016.07.007

Document License CC BY-NC-ND

General rights

Copyright and moral rights for the publications made accessible in the Aberystwyth Research Portal (the Institutional Repository) are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the Aberystwyth Research Portal for the purpose of private study or research.

- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Aberystwyth Research Portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

tel: +44 1970 62 2400 email: is@aber.ac.uk

Metabolomic-based biomarker discovery for non-invasive lung cancer screening: A case study

Keiron O'Shea^a, Simon J.S. Cameron^{a,b}, Keir E Lewis^c, Chuan Lu^d, Luis AJ Mur^{a,*}

^aInstitute of Biological, Environmental and Rural Studies, Aberystwyth University, Aberystwyth, Wales, SY23 3DA, UK

^bDivision of Computational and Systems Medicine, Department of Surgery and Cancer, Imperial College London, London, W6 8RP, UK

^cDepartment of Respiratory Medicine, Prince Philip Hospital, Llanelli, Wales, SA14 8LY, UK

^dDepartment of Computer Science, Aberystwyth University, Aberystwyth, Wales, SY23 3DA, UK

Abstract

Background: Lung cancer (LC) is one of the leading lethal cancers worldwide, with an estimated 18.4% of all cancer deaths being attributed to the disease. Despite developments in cancer diagnosis and treatment over the previous thirty years, LC has seen little to no improvement in the overall five year survival rate after initial diagnosis.

Methods: In this paper, we extended a recent study which profiled the metabolites in sputum from patients with lung cancer and age-matched volunteers smoking controls using flow infusion electrospray ion mass spectrometry. We selected key metabolites for distinguishing between different classes of lung cancer, and employed artificial neural networks and leave-one-out cross-validation to evaluate the predictive power of the identified biomarkers.

Results: The neural network model showed excellent performance in classification between lung cancer and control groups with the area under the receiver operating characteristic curve of 0.99. The sensitivity and specificity of for detecting cancer from controls were 96% and 94% respectively. Furthermore, we have identified six putative metabolites that were able to

Preprint submitted to BBA General Subjects

^{*}Corresponding author Email address: lum@aber.ac.uk (Luis AJ Mur)

discriminate between sputum samples derived from patients suffering small cell lung cancer (SCLC) and non-small cell lung cancer. These metabolites achieved excellent cross validation performance with a sensitivity of 80% and specificity of 100% for predicting SCLC.

Conclusions: These results indicate that sputum metabolic profiling may have potential for screening of lung cancer and lung cancer recurrence, and may greatly improve effectiveness of clinical intervention.

Keywords:

lung cancer, small vs non-small cell lung cancer, sputum, metabolomics, biomarkers, artificial neural networks

1 1. Introduction

The year 2008 saw an estimated 12.7 million new cases of cancer, and 7.6 million cancer-related deaths worldwide [1]. While the incidence and mortality rates of most cancers is decreasing in the developed world, they are rising in emerging economies such as China and India. Migrant studies have found that cancer rates in the descendent generation of migrants tends to shift toward the host country, suggesting that environmental risk factors such as smoking and weight are responsible for the global variance in cancer rates [2].

10 1.1. Lung cancer

Lung cancer is a major cause of death in the developed and developing 11 worlds. It is the leading cause of cancer-related deaths in men, and second 12 only to breast cancer in women. There was an estimated 1.6 million new 13 cases of lung cancer and 1.4 million deaths in 2008. This accounts for 12.6%14 of all cancer incidence and a staggering 18.4% of all cancer-related deaths 15 [2]. This can be attributed to its poor prognosis, with the five-year survival 16 rate being a mere 15%. Despite recent advances in lung cancer treatment, 17 survival rates are low when compared to other forms of cancer [3]. How-18 ever, improvements in surgical techniques and chemotherapy over the past 19 twenty years has resulted in one-year lung cancer survival rates drastically 20 improving. Despite this, the overall five-year lung cancer survival rates have 21 remained stagnant at 6% for small cell lung cancer and 18% for non-small cell 22 lung cancer. Unfortunately the vast majority (85%) of lung cancer cases are 23

diagnosed at advanced stages, heavily reducing the effectiveness of treatment
[1].

This can be attributed to the difficulty of effectively diagnosing cancer of 26 the lung at stage early enough to make a real impact. One of the main diffi-27 culties is that symptoms of the conditions are often identical to less serious 28 conditions. This makes the pre-clinical diagnosis of lung cancer particularly 29 problematic as the observed symptoms are often confused with other respi-30 ratory conditions. Prognostic factors may help diagnose patients who show 31 symptoms of a disease, or have an increased chance of recurrence or progres-32 sion to advanced disease which should support clinicians in the creation of 33 appropriate treatment plans. The World Health Organisation (WHO) have 34 set out ten key principles to be met by an effective screening procedure in 35 order for it to be beneficial and cost effective [4]. Currently there are no lung 36 cancer screening techniques of which meet all of the ten conditions laid out 37 by the WHO. 38

³⁹ 1.2. Metabolomic insights into lung cancer

An emerging screening methodology to other traditional screening meth-40 ods is the utilisation of molecular biomarkers in biofluids. The ease of analy-41 sis of biofluids using mass spectrometry (MS) or nuclear magnetic resonance 42 (NMR) makes metabolomics a well-suited methodology for the non-invasive 43 detection of biomarkers in lung cancer. Current focus of metabolomics in lung 44 cancer has been on the exploitation of serum, urine and tumour biopsies. For 45 example, the analysis of serum using liquid chromatography (LC-MS) and 46 gas chromatography (GC-MS) approaches have suggested a potential use for 47 biomarkers of lung cancer. A small-scale pilot study sampling lung cancer pa-48 tients before and after surgical intervention, alongside patients without lung 49 cancer has suggested ten candidate biomarkers for lung cancer, including 50 sphingosine, oleic acid and serine [5]. 51

Sputum has been suggested as a potential biofluid source of biomarkers 52 in lung cancer [3, 6]. Recent work has used Fourier Transform Infra-Red 53 (FTIR) spectroscopy as a non-invasive method to detect lung cancer in spu-54 tum samples. This work concluded that FTIR was able to sufficiently dis-55 tinguish between lung cancer and control samples, and effectively act as a 56 non-invasive, high-throughput and cost-effective method for screening spu-57 tum samples from high-risk patients. Furthermore, it further validated the 58 use of sputum as an effective biofluid for lung cancer screening [7]. 59

60 1.3. Artificial Neural Networks

Artificial neural networks (ANNs) are a class of sophisticated computational modelling structures that are inspired by biological neurological systems, regarding how they are able to learn and process highly non-linear information [37]. The past three decades have seen ANNs being widely used for biomedical decision support systems [8, 9, 10, 11, 12].

In general, an artificial neural network is formed of interconnected pro-66 cessing units, commonly referred to as neurons. Each neuron applies an ac-67 tivation function over the weighted sum of the incoming stimuli (or inputs), 68 and generate an output signal, which could be the input signal for other neu-69 rons. Many different neural network architectures exist, in this paper we will 70 focus on the popular feed-forward artificial neural network, in particular the 71 multi-layer perceptron (MLP) [13], that usually consists of multiple layers 72 of neurons - the input layer, one or more hidden layers and the final output 73 layer, as illustrated in Figure 1.



Figure 1: Illustration of a three-layered feed-forward artificial neural network, where each neuron in one layer has connections to the subsequent layer.

The design of network architectures involves setting the number of hidden layers, the number of neurons within each layer, the connections between them, and the type of activation function to use. The connection weights in the network could be adjusted through a learning algorithm that minimises the amount of error in the outputs compared to the true ones. Generalisation, and to avoid overfitting the training data would be a central issue both in network design and training.

⁸² 2. Case Study

We have recently developed this approach by employing flow-infusion 83 electrospray-Mass Spectrometry (FIE-MS) to evaluate the potential of spon-84 taneous sputum as a source of non-invasive metabolomic biomarkers for LC 85 status [14]. Spontaneous sputum was collected and processed from 34 pa-86 tients suspected of having LC, and 33 healthy controls. Of the 34 patients, 87 23 were subsequently diagnosed with LC (LC+) at various stages of disease 88 progression. The clinical characteristics of all samples taken are summarised 80 in Table 1. 90

Characteristics	Lung cancer	Symptoms	Healthy con-
	(LC+)	(LC-)	trols
Number	23	11	33
Age (mean \pm SD)	66.6 ± 8.1	66.5 ± 14.3	55.3 ± 14.6
Gender (Male/Female)	11/12	10/1	20/13
Smoking (Current / Ex / Never)	10/10/3	3/0/8	15/18/0
Previous cancer (Yes/No)	3/20	N/A	N/A
Final clinical diagnosis	5/17/1	N/A	N/A
(SCLC/NSCLC/Radiological)			
CO level (ppm)	4.2 ± 2.8	3.7 ± 1.3	N/A

Table 1: Summary of clinical characteristics of patients with Lung Cancer (LC+), Symptoms (LC-) and Healthy Controls

In these preliminary analyses, discriminatory metabolites were identified 91 using ANOVA and Random Forest and included Ganglioside GM1 which has 92 previously been linked to lung cancer [15]. This suggested that the use of 93 sputum as a non-invasive source of metabolite biomarkers may aid in the 94 development of an at-risk population screening programme for lung cancer 95 or enhanced clinical diagnostic pathways. We now demonstrate how further 96 data-mining of the FIE-MS data has revealed further metabolite biomark-97 ers, and evaluate further the use of metabolomics to yield biomarkers for 98 distinguishing lung cancer type. 99

100 2.1. Ethics statement

¹⁰¹ The MedLung observational study (UKCRN ID 4682) received loco-regional ¹⁰² ethical approval from the Hywel Dda Health Board (05/WMW01/75). Written informed consent was obtained from all participants at least 24 hours
before sampling, at a previous clinical appointment, and all data was link
anonymised before analysis.

106 2.2. Mass spectrometry

Frozen sputum samples were thawed before being exposed to 0.5 mL of dithiothreitol (DTT) to isolate sputum cells. Each sample was mixed using a vortex mixer for 15 minutes before being centrifuged at 1800g for 10 minutes before removing the supernatant. Sputum pellets were then analysed using Flow Infusion Electrospray Ion Mass spectrometry (FIE-MS).

Signals identified under 50 m/z were removed, and the resulting FIE-MS data matrix contains 2,582 m/z values after binning. The data was further preprocessed by total ion count (TIC) normalisation (to ensure the intensity values for each spectrum sum up to one), followed by log10 transformation prior to further data analysis.

117 2.3. Effect of clinical characteristics on the metabolic profiles

To explore the possible effects of the clinical characteristics on the global 118 metabolic profiles, we conducted the so called 50-50 MANOVA test, which is 119 essentially a variant of classical MANOVA that can handle multiple highly 120 correlated responses [16]. We found no significant effect of age, gender or 121 the CO level on the preprocessed metabolomic data (with p-values of 0.4, 122 0.08, and 0.8, respectively), whilst the effect of disease status (LC+/LC-123 /Control) is really strong (*p*-value = 1e-12). And perhaps not surprisingly, 124 as tobacco smoking is an important risk factor for lung cancer, there is indeed 125 a significant effect of the smoking pack numbers per year over the metabolic 126 profiles for the patients of LC- and LC+ (p-value = 2e-5). 127

¹²⁸ 2.4. Diagnostic modelling with artificial neural networks

To find discriminatory m/z features, Welch's unequal variance t-test have been performed using the pre-processed intensity values after log-transformation. Random forests have also been tried (results not shown), and the top ranked features identified for both methods are quite similar.

Then an ANN was used as a diagnostic model for various binary classification problems, taking the selected m/z signals (the preprocessed intensity values after log-transform) as the inputs and estimating the probability for individual classes. The activation function was set to hyperbolic tangent for both hidden and output layer; and the number of hidden layers was set to two for all problems based on our initial analysis. Regularisation techniques
such as weight decay [17] have been employed to control the complexity of
the model parameters in order to avoid overfitting the models to the training
data.

The predictive power of neural network classifiers was evaluated using Receiver Operating Characteristic (ROC) analysis and through leave-one out (LOO) cross-validation (CV) - the overall ROC curve and the area under the curve (AUC) from CV was obtained using the pooled test examples from CV. For each binary classification problem and within each round of training, a *t*-test would be performed on the training set, only the m/z signals with resulting *p*-values < 0.05 were selected as input features for ANN modelling.

¹⁴⁹ 3. Results and Discussion

Representative spectra of the samples from the LC+, LC- and healthy control sample groups are shown in Figure 2. FIE-MS profiles were analysed using principal component analysis (PCA) (Figure 4). One can observe no clear separation between clinically collected sample groups (LC+ and LC-) if all m/z signals were used for PCA.

Welch t-tests provided 445 distinctive m/z values for LC+ versus healthy 155 controls, and 90 significant m/z values for LC+ versus LC- from our pooled 156 leave one out cross validation *t*-tests. PCA of both stratifications showed 157 good discriminative ability when using the pooled features, as shown in Fig-158 ure 4. The number of neurons in each hidden layer was chosen through grid 150 search [18]. The best-performing models and their diagnostic performance 160 from LOOCV can be found in Table 2. And Figure 6 shows the resulting 161 ROC curves. 162

Classification	Mean no.	No. of	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	AUC
	of inputs	nidden					
		neurons					
LC+ (vs Control)	1730.2	100, 50	96%	94%	92%	97%	0.99
LC+ (vs $LC-$)	71.9	40, 10	100%	91 %	96%	100%	1.00
SCLC (vs NSCLC)	77.8	50, 20	80%	100%	100%	94%	1.00

Table 2: Results of cross-validation prediction performance of our ANN models. The diagnostic performance (except for AUC) was obtained by using the default probability cutoff value of 0.5 to determine the class labels.



Figure 2: Typical FIE-MS spectra of sputum obtained from sputum obtained from (a) a patient with lung cancer, (b) a patient with symptoms of lung cancer and (c) a healthy control sample.



Figure 3: PCA score plots of FIE-MS data using all normalised m/z intensity values in negative ionisation mode, showing no clear separation between LC+ and LC- samples.

¹⁶³ 3.1. Analysis of small-cell lung cancer and non-small cell lung cancer

Determining the type of lung cancer that has developed in a patient is a key component of determining the correct treatment and management pathway. For lung cancer, two broad classes of classification exist: non-



Figure 4: PCA score plots of the FIE-MS data with (a) only 970 selected m/z signals for LC+ and healthy controls, and (b) 125 selected m/z signal for LC+ and LC-. Using m/z features taken from t-tests clearly differentiates between relevant classes.

small-cell lung cancer (NSCLC) and small-cell lung cancer (SCLC). Patients 167 with NSCLC are usually classified as one of three main subtypes: adenocar-168 cinoma, squamous-cell carcinoma and large-cell carcinoma. Of these, ade-169 nocarcinoma is the most common and is characterised by overproduction of 170 mucin. Squamous-cell carcinoma is the second most common form of lung 171 cancer and typically occurs in the centre of the lungs. Large-cell carcinoma 172 is less common and is characterised by cancerous cells that are large, with 173 excess cytoplasm and large nuclei. The extent of NSCLCs is reported using 174 the TNM format, which is important for prognosis and treatment planning. 175 The TNM format ranges from Stage 0 to Stage IV, with the relevant stage 176 determined through assessment of the primary tumour, involvement of re-177 gional lymph nodes, and the extent of distant metastasis against set criteria 178 [19]. 179

Small-cell lung cancers are less common than NSCLC, with approximately 10% of all lung cancers classified as SCLC. These lung cancers are characterised by their small cells, with minimal cytoplasm, and poorly-defined cell borders. Cancerous cells are usually rounded, oval and spindle-shaped. Typically, patients with SCLC present when the disease has metastasised from the lungs and symptoms frequently this, such as issues with bone marrow and the liver because of metastasis. Small-cell lung cancers are staged differ-



Figure 5: Mean FIE-MS spectra illustrating five key distinguishable metabolites between patients with lung cancer (LC) and healthy controls.



Figure 6: Receiver operating characteristic curves obtained from ANN models using leave one out cross-validation for classifying: (a) LC+ against LC-, and (b) LC against healthy controls.

ently to NSCLC. Although the TNM format can be used, it does not predict
survival and other outcomes well. Typically, SCLC is staged as either lim-

ited or extensive disease, with the latter equivalent to Stage IV of the TNM
staging format for NSCLC [19].

A total of nine m/z values provided strong differentiation between SCLC 191 and NSCLC with p-values less than 0.05 from Welch t-tests. Out of these 192 9 m/z values, 6 were identified and brought forward for further analysis as 193 potential biomarkers of NSCLC and SCLC. Their relative values and p-values 194 are shown in Table 3. Furthermore, ranges of each metabolite shown as box-195 and-whisker plots can be found in Figure 7. These both indicate that the 6 196 biomarker candidates show that the median levels of the 6 markers are higher 197 in NSCLC samples to patients with SCLC. 198

Metabolite	m/z	Normalised In	<i>p</i> -value	
		NSCLC	SCLC	
Phenylacetic acid	137.09	-3.31 ± 0.21	-2.85 ± 0.42	0.001
L-Fucose	165.09	-3.19 ± 0.12	-2.86 ± 0.30	0.001
Caprylic acid	145.18	-2.81 ± 0.44	-2.12 ± 0.31	0.001
Acetic acid	61.09	-2.96 ± 0.19	-2.56 ± 0.40	0.002
Propionic acid	75.09	-1.91 ± 0.18	-1.56 ± 0.34	0.003
Glycine	76.09	-3.17 ± 0.12	-2.91 ± 0.21	0.004

Table 3: Identified metabolites that are significantly different between patients with Non-Small Cell Lung Cancer (NSCLC) and Small Cell Lung Cancer (SCLC) using Welch t-tests.

PCA analysis (Figure 8a) showed good separation capabilities between NSCLC and SCLC samples, 6 metabolites were selected as input features to build a second ANN. Due to the small sample size, leave-one-out cross validation was performed to estimate the generalisation performance of the model. Our MLP model was able to distinguish between NSCLC and SCLC with a sensitivity of 80% and a specificity of 100% for predicting SCLC from cross-validation (see Table 2 and Figure 8b).

206 4. Biomarker analysis

Metabolic profiling recognised and provided identifications of 6 candidate metabolites that offered superb predictive values. Amongst the targeted metabolites are examples which have already been linked to lung cancer. The enzyme glycine decarboxylase (GLDC) is involved in the degradation of



Figure 7: Box-and-whisker plots of the candidate metabolite biomarkers for discrimination between NSCLC and SCLC. The y axis represents the normalised intensity of each metabolite. Horizontal lines in the middle portion of the box illustrates the median, bottom and top boundaries of boxes represent the lower and upper quartile, whiskers depict the 5th and 95th percentiles and plus signs depicts the outliers.

glycine which is coupled to the generation of methylgroups which can be used in (for example) purine biosynthesis. GLDC expression was increased in cells isolated from NSCLC tumours with concomitent decreases in glycine [20]. These authors showed that GLDC expression could serve as a biomarker, we now provide evidence that relative decreases in glycine is a feature of NSCLC in sputum. This biofluid represents a less invasive and potentially cost-effective means of clinically assessing patient LC status.

Fucose (6-deoxy-L-galactose) is N- and O-linked to a range of glycolipids and glycopeptides produced by mammalian cells. Increases in fucoslylated



Figure 8: PCA score plot of the 6 identified metabolites for (a) NSCLC and SCLC, and (b) the ROC curve obtained from ANN LOOCV.

proteins, for example, a-fetoprotein are used in the diagnosis of hepatocel-220 lular carcinoma [21]. Fucosylation is dependent on the availability of the 221 substrate guanosine 5'-diphospho-fucose (GDP-Fucose) and associated gly-222 cosyltransferases to transfer the fucose motif on to the protein / lipid. Fucose 223 is a precursor to GDP-Fucose production [22] so that increases in fucoslyation 224 could lead to a relative depletion in fucose as noted in our study. Increased 225 fucosylation has been previously linked to NSCLC biopies [23] but ours is the 226 first suggestion that decreases in fucose pools in sputum could be clinically 227 suitable marker. 228

Other key metabolites were volatile short chain fatty s acetic (C2), propionic (C3) and caprylic [octanoic] (C8), These could be derived as a result of lipid peroxidation [24] but irrespective of their means of generation would provide further support for efforts that are attempting to sample breath as an non-invasive method for lung cancer detection [24, 25, 26].

A somewhat surprising observation was the detection of phenylacetic. 234 This is classically associated with phenylketonuria (PKU); an inherited dis-235 order of amino metabolism. PKU arises from a deficiency of the liver enzyme 236 phenylalanine-4-hydroxylase which production of tyrosine from phenylala-237 nine. If this enzyme is non-functional a range of alternative metabolites are 238 produced, including phenylacetic [27]. However, beyond PKU, phenylacetate 239 accumulates in patients with chronic kidney disease and during renal failure 240 where it can inhibit nitric oxide generation [28] and macrophage intracellular 241

killing of bacteria [29]. Further, disease-associated phenylacetate accumula-242 tion can contribute to an inflammatory responses [30]. To our knowledge, 243 phenylacetate has not been associated with cancer; indeed quite the opposite 244 it has a history of being tested for its anti-tumour properties [31]. However, 245 it may be that NSCLC has particularly phenotypic / biochemical features 246 which lead to altered phenylalanine metabolism leading to the accumulation 247 of phenylacetate to contribute to inflammatory events and a reduced ability 248 to deal with the lung microbiome. This is currently under investigation in 249 our group. 250

²⁵¹ 5. Conclusions

A metabolomics approach based on FIE-MS coupled with univariate 252 Welch *t*-test based feature selection and artificial neural networks provides 253 an efficient methodology for metabolomic-based profiling of sputum to dif-254 ferentiate between non-small cell and small-cell lung cancer. This paper has 255 identified 6 candidate metabolites markers, including L-fucose, phenylacetic, 256 caprylic, acetic, propionic acid, and glycine, which were found to have good 257 discriminatory abilities and low *p*-values. Excellent sensitivity and specificity 258 was also shown using these markers through leave-one-out cross validation, 250 which further indicates the promise of metabolomic analysis of sputum for 260 non-invasive screening for LC. Further analysis involving a larger number of 261 samples is required to determine both the precision and applicability of this 262 approach in guiding the diagnosis and treatment of LC and respective forms. 263

264 Acknowledgements

The work described in this paper was funded by two Aberystwyth University PhD scholarships to Keiron O'Shea and Simon Cameron. Keiron O'Shea would like to thank Nicholas Dimonaco and Manfred Beckmann of Aberystwyth University for their helpful support offered during this project.

269 **References**

[1] J. Ferlay, H.-R. Shin, F. Bray, D. Forman, C. Mathers, D. M. Parkin,
Estimates of worldwide burden of cancer in 2008: Globocan 2008, International journal of cancer 127 (2010) 2893–2917.

- [2] A. Jemal, R. Siegel, J. Xu, E. Ward, Cancer statistics, 2010, CA: a cancer journal for clinicians 60 (2010) 277–300.
- [3] L. Guzmán, M. S. Depix, A. M. Salinas, R. Roldán, F. Aguayo, A. Silva,
 R. Vinet, Analysis of aberrant methylation on promoter sequences of
 tumor suppressor genes and total dna in sputum samples: a promising
 tool for early detection of copd and lung cancer in smokers, Diagn Pathol
 7 (2012) 1596-7.
- [4] A. Andermann, I. Blancquaert, S. Beauchamp, V. Déry, Revisiting
 wilson and jungner in the genomic age: a review of screening criteria
 over the past 40 years, Bulletin of the World Health Organization 86
 (2008) 317–319.
- [5] Y. Chen, Z. Ma, A. Li, H. Li, B. Wang, J. Zhong, L. Min, L. Dai,
 Metabolomic profiling of human serum in lung cancer patients using liquid chromatography/hybrid quadrupole time-of-flight mass spectrometry and gas chromatography/mass spectrometry, Journal of cancer
 research and clinical oncology 141 (2015) 705-718.
- [6] A. Hubers, C. Prinsen, G. Sozzi, B. Witte, E. Thunnissen, Molecular
 sputum analysis for the diagnosis of lung cancer, British journal of
 cancer 109 (2013) 530-537.
- [7] P. D. Lewis, K. E. Lewis, R. Ghosal, S. Bayliss, A. J. Lloyd, J. Wills,
 R. Godfrey, P. Kloer, L. A. Mur, Evaluation of ftir spectroscopy as a
 diagnostic tool for lung cancer using sputum, BMC cancer 10 (2010) 1.
- [8] P. J. Lisboa, A. F. Taktak, The use of artificial neural networks in
 decision support in cancer: a systematic review, Neural networks 19
 (2006) 408-415.
- [9] J. Khan, J. S. Wei, M. Ringner, L. H. Saal, M. Ladanyi, F. Westermann,
 F. Berthold, M. Schwab, C. R. Antonescu, C. Peterson, et al., Classification and diagnostic prediction of cancers using gene expression profiling
 and artificial neural networks, Nature medicine 7 (2001) 673–679.
- [10] Y. Wu, M. L. Giger, K. Doi, C. J. Vyborny, R. A. Schmidt, C. E.
 Metz, Artificial neural networks in mammography: application to decision making in the diagnosis of breast cancer., Radiology 187 (1993) 81-87.

- [11] C. Lu, J. De Brabanter, S. Van Huffel, I. Vergote, D. Timmerman, Using artificial neural networks to predict malignancy of ovarian tumors, in: Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE, volume 2, IEEE, pp. 1637–1640.
- ³¹¹ [12] R. Dybowski, V. Gant, Artificial neural networks in pathology and ³¹² medical laboratories, The Lancet 346 (1995) 1203–1207.
- [13] D. E. Rumelhart, G. E. Hinton, R. J. Williams, Learning internal representations by error propagation, Technical Report, DTIC Document, 1985.
- [14] S. J. Cameron, K. E. Lewis, M. Beckmann, G. G. Allison, R. Ghosal,
 P. D. Lewis, L. A. Mur, The metabolomic detection of lung cancer
 biomarkers in sputum, Lung Cancer 94 (2016) 88–95.
- [15] T. Brezicka, B. Bergman, S. Olling, P. Fredman, Reactivity of monoclonal antibodies with ganglioside antigens in human small cell lung
 cancer tissues, Lung Cancer 28 (2000) 29–36.
- ³²² [16] Ø. Langsrud, 5050 multivariate analysis of variance for collinear re³²³ sponses, Journal of the Royal Statistical Society: Series D (The Statis³²⁴ tician) 51 (2002) 305–317.
- [17] R. C. S. L. L. Giles, Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping, in: Advances in Neural Information
 Processing Systems 13: Proceedings of the 2000 Conference, volume 13, MIT Press, p. 402.
- [18] J. Bergstra, Y. Bengio, Random search for hyper-parameter optimiza tion, The Journal of Machine Learning Research 13 (2012) 281–305.
- [19] W. D. Travis, C. Harris, Pathology and genetics of tumours of the lung,
 pleura, thymus and heart, Feance: IARC Press, 2004, 2004.
- [20] W. C. Zhang, N. Shyh-Chang, H. Yang, A. Rai, S. Umashankar, S. Ma,
 B. S. Soh, L. L. Sun, B. C. Tai, M. E. Nga, et al., Glycine decarboxylase activity drives non-small cell lung cancer tumor-initiating cells and
 tumorigenesis, Cell 148 (2012) 259–272.

- E. Miyoshi, K. Moriwaki, N. Terao, C.-C. Tan, M. Terao, T. Nakagawa,
 H. Matsumoto, S. Shinzaki, Y. Kamada, Fucosylation is a promising
 target for cancer diagnosis and therapy, Biomolecules 2 (2012) 34–45.
- [22] D. J. Becker, J. B. Lowe, Fucose: biosynthesis and biological function
 in mammals, Glycobiology 13 (2003) 41R-53R.
- [23] X. Zeng, B. L. Hood, M. Sun, T. P. Conrads, R. S. Day, J. L. Weissfeld,
 J. M. Siegfried, W. L. Bigbee, Lung cancer serum biomarker discovery
 using glycoprotein capture and liquid chromatography mass spectrometry, Journal of proteome research 9 (2010) 6440–6449.
- ³⁴⁶ [24] C. Wang, R. Dong, X. Wang, A. Lian, C. Chi, C. Ke, L. Guo, S. Liu,
 ³⁴⁷ W. Zhao, G. Xu, et al., Exhaled volatile organic compounds as lung can³⁴⁸ cer biomarkers during one-lung ventilation, Scientific reports 4 (2014).
- [25] M. Phillips, K. Gleeson, J. M. B. Hughes, J. Greenberg, R. N. Cataneo,
 L. Baker, W. P. McVay, Volatile organic compounds in breath as markers
 of lung cancer: a cross-sectional study, The Lancet 353 (1999) 1930–
 1933.
- [26] M. Phillips, R. N. Cataneo, A. R. Cummin, A. J. Gagliardi, K. Gleeson,
 J. Greenberg, R. A. Maxfield, W. N. Rom, Detection of lung cancer with
 volatile markers in the breath, Chest Journal 123 (2003) 2115–2123.
- ³⁵⁶ [27] A. Bajtarevic, C. Ager, M. Pienz, M. Klieber, K. Schwarz, M. Ligor,
 T. Ligor, W. Filipiak, H. Denz, M. Fiegl, et al., Noninvasive detection
 of lung cancer by analysis of exhaled breath, BMC cancer 9 (2009) 1.
- J. Jankowski, M. Van Der Giet, V. Jankowski, S. Schmidt, M. Hemeier,
 B. Mahn, G. Giebing, M. Tölle, H. Luftmann, H. Schlüter, et al., Increased plasma phenylacetic acid in patients with end-stage renal failure
 inhibits inos expression, The Journal of clinical investigation 112 (2003)
 256–264.
- S. Schmidt, T. H. Westhoff, P. Krauser, R. Ignatius, J. Jankowski,
 V. Jankowski, W. Zidek, M. Van der Giet, The uraemic toxin phenylacetic acid impairs macrophage function, Nephrology Dialysis Transplantation 23 (2008) 3485–3493.

- [30] G. Cohen, J. Raupachova, W. H. Hörl, The uraemic toxin phenylacetic
 acid contributes to inflammation by priming polymorphonuclear leucocytes, Nephrology Dialysis Transplantation 28 (2013) 421–429.
- [31] A. Thibault, D. Samid, M. R. Cooper, W. D. Figg, A. C. Tompkins,
 N. Patronas, D. J. Headlee, D. R. Kohler, D. J. Venzon, C. E. Myers,
 Phase i study of phenylacetate administered twice daily to patients with
 cancer, Cancer 75 (1995) 2932–2938.