TITLE

Metabolomics differentiate cancer from non-cancer pleural effusions based on steroid lipids and acyl carnitines

Katie Love¹, Ricardo MF da Costa¹, Manfred Beckmann¹, Anna Morley², Rahul Bhatnagar², Nick Maskell², Luis AJ Mur¹ and Keir E Lewis^{3,4 *}

¹ Institute of Biological, Environmental and Rural Sciences, Edward Llywd Building, Penglais Campus, Aberystwyth University, Ceredigion SY23 3FG, UK

² Academic Respiratory Unit, University of Bristol, Learning and Research Building, Southmead Hospital, BS10 5NB

³ Department of Respiratory Medicine, Prince Phillip Hospital, Llanelli SA14 8LY, UK

⁴ School of Medicine, Swansea University, Swansea SA2 8PP, UK

*CORRESPONDING AUTHOR

Keir Lewis, Tel: 44-1554783133

k.e.lewis@swansea.ac.uk

1 AUTHOR CONTRIBUTIONS

2

3 KEL had the original idea, led on study design and write-up and is guarantor of the
4 paper.

5

6 KL and RdC conducted the practical work described in this manuscript and are joint
7 first authors.

8

9 MB advised on the metabolomic analyses.

10

11 AM advised on sample metadata.

12

13 RB and NM collected and provided the pleural effusions samples and advised on
14 programme.

15

16 LAJM supervised the metabolomic analyses, contributed to the write up
17 and is joint senior author.

18

19 Supported by grants from Knowledge Economy Skills Scholarships
20 (KESS) and TENOVUS Cancer Charity UK. Funders had no involvement
21 in the study design, collection, analysis or interpretation of data. Nor in
22 the writing of the report, or in the decision to submit the article for
23 publication.

24

25

1

2 **Abbreviations**

3

4 CCF Congestive Cardiac Failure

5 COPD Chronic Obstructive Pulmonary Disease

6 CPPE Complicated Parapneumonic Effusion

7 FIE-MS Flow Infusion Electrospray ion Mass Spectrometry

8 IPF Idiopathic Pulmonary Fibrosis

9 PCA Principal Component Analyses

10 ROC Receiver Operating Characteristic

11 AUC Area under the Curve

12 VIP Variable Importance in Projection

13 RF Random Forest

14

1 **Abstract**

2 **Objective:** To identify and detect metabolomic markers to differentiate cancer
3 related from other pleural effusions.

4 **Material and methods:** Un-targeted flow infusion electrospray mass spectrometry
5 was used on a cohort of 100 samples from benign and malignant pleural effusions
6 (12 primary lung cancers, 14 mesotheliomas, 24 other cancers, 25 congestive
7 cardiac failure, 22 parapneumonic effusion, 3 empyema). Standard metabolomic
8 statistical models for analysis were performed.

9 **Results:** Five novel markers were identified using univariate and multivariate
10 receiving operator curve analyses using five discriminant features yielding a
11 diagnostic accuracy of above 0.91 (95% CI: 0.839-0.976) with a sensitivity of 0.97
12 (95% CI 0.952-0.979) and specificity 0.87 (95% CI 0.852-0.894).

13 **Conclusion:** Individual biomarkers with the highest accuracy (>0.84) were all raised
14 in cancer and are involved in metabolism of sterol-lipids and acyl carnitines linked to
15 β -oxidation. Our markers performed better than those previously published. Further,
16 these published markers failed to achieve their stated accuracies in our data.

17

18

19 **Key words:**

20 Pleural effusions; Malignancy; Metabolomics; Biomarkers; Mass Spectrometry;
21 Diagnostics

22

1 Introduction

2 Accumulation of pleural fluid is a common pathology in over 50 recognised
 3 aetiologies ranging from treatable infections, to malignant mesothelioma [1]. For
 4 more than 40 years 'Light's criteria' has been the gold standard for the diagnosis of
 5 pleural effusions, able to differentiate effusions with a diagnostic accuracy of 93-96%
 6 [2]. However, they are known to mis-classify some transudates as exudates e.g.
 7 especially in patients using diuretics [3], which may have significant implications for
 8 clinical management. Further, cytological confirmatory tests for malignancy through
 9 pleural aspiration account for an accuracy of only 60 %, with a sensitivity of 40-87 %
 10 [4]. Better diagnostics can be developed with new technologies, such as
 11 metabolomics, and have yielded accurate, reliable, and sensitive tests for lung
 12 cancer using other non-invasive samples such as saliva and urine [5]–[7].

13 The metabolome describes the large numbers of metabolites within a given sample.
 14 Untargeted metabolomic profiling, aims to comprehensively detect, quantify, and
 15 identify metabolites in a given sample. Flow Infusion Electrospray-Mass
 16 Spectrometry (FIE-MS) is one technique that allows high-throughput and highly
 17 sensitive detection of metabolites. Metabolomics increases the fundamental
 18 understanding of biochemical pathways in normal and abnormal pathologies.

19 Although metabolomics could be of major benefit to clinical medicine [8], it has yet to
 20 deliver on this potential. Despite having a comprehensive list of processing and
 21 reporting standards for metabolomic studies, many published papers fail to adhere to
 22 them – resulting in a lack of reproducible findings and a scarcity of translating
 23 metabolomics into clinical practice [9].

1 Herein we use FIE-MS to generate metabolomic data from cancerous and non-
2 cancerous pleural effusions and compare its accuracy with standard approaches
3 including Lights criteria and other published pleural fluid biomarkers. We also aim to
4 use our data set as a rudimentary external validation cohort for previously published
5 biomarker studies.

6

7 **Methods**

8 In this pilot, proof of concept study, 100 pleural effusion samples were selected from
9 a UK, cohort-based repository. Samples were obtained from patients attending a
10 tertiary clinical centre between 2015-2018. A cancer diagnosis was confirmed
11 cytologically or histologically and non-cancer diagnosis confirmed either cytologically
12 or histologically from the pleural space, in line with usual clinical standards.

13 Samples were chosen to consist of an equal number of 50 cancers (12 primary (LC),
14 14 mesothelioma (MPM) 24 other secondary cancers), and 50 samples from non-
15 cancer (25 congestive cardiac failure (CCF), 22 Complicated Parapneumonic
16 Effusion (CPPE) and 3 with empyema).

17 18 patients were diagnosed on their first pleural tap (x4 LC, 1 mesothelioma, 13
18 other cancer), with 32 samples obtained from further tests (CT-guided biopsy, video
19 assisted thoracoscopy, repeat pleural taps, surgery). Patient metadata are
20 summarised in Supplementary Table 1.

21

22 **Processing of pleural effusions**

23 All samples were processed for metabolomics in a blinded manner. A two-step
24 extraction was used for metabolomic analysis. Aliquots were thawed on ice then

1 vortexed, followed by centrifugation (1800 $\times g$ for 10 min). 100 μ L of supernatant was
2 added to 50mg of acetone-washed glass beads ($\leq 106\mu$ m), with 1mL of 4:1 (v/v) mix
3 of MeOH:chloroform (HPLC grade). Samples were vortexed, shaken (60 min at 4°C),
4 left to precipitate (-80°C for 20 min), then centrifuged (1800 $\times g$ for 10 min). 100 μ L of
5 the upper layer was re-extracted following the same procedure.

6 7 **Flow infusion electrospray mass spectrometry (FIE-MS)**

8
9 60 μ L of the supernatant was transferred to a glass vial with insert. Sample injection
10 order was randomised to reduce batch effects and run on an Exactive™ Orbitrap
11 Mass Spectrometer (Thermo Scientific). Data were acquired in alternating positive
12 and negative ionisation modes over 4 scan ranges (15–110, 100–220, 210–510, and
13 500–1200 m/z) 5 min acquisition time. Following the infusion of each sample, mass
14 spectra were derived from the mean of 20 scans about the apex of the peak signal,
15 Data were normalised against the total ion count and the intensity value for each
16 individual m/z (ranging from 55.93187 to 1178.146 Da) and expressed as a
17 percentage of the total ion count. Data are given in Supplementary Table 2.

18

19 **Data and statistical analysis**

20 Disease classifications were known to the data analyst, and missing values were
21 assigned as 'NA'. Data were Bonferroni corrected for multiple comparisons. Principal
22 Component Analyses (PCA), Partial Least Square Discriminant Analysis (PLS-DA),
23 ANOVA, Receiver Operating characteristic (ROC), and Area Under Curve (AUC)
24 analyses were applied using the MetaboAnalyst 4.0 platform [10]. Our ROC curves
25 were generated by Monte-Carlo cross validation using subsampling (2/3 to evaluate

1 feature importance, and 1/3 to validate the models). ROC-AUC analyses were
 2 conducted on metabolites with a Variable Importance in Projection (VIP) score of >2.
 3 The 5 metabolites were chosen based on VIP scores (>2) and statistical significance
 4 ($P<0.01$) (corrected for false discovery rates). Metabolites were identified based on
 5 PubChem [<https://pubchem.ncbi.nlm.nih.gov/>], HMDB [<http://www.hmdb.ca/>], KEGG
 6 [<https://www.genome.jp/kegg/compound/>] and CheBI databases
 7 [<https://www.ebi.ac.uk/chebi/>] which were interrogated using MZedDB
 8 [<http://maltese.dbs.aber.ac.uk:8888/hrmet/index.html>] and DimeDB
 9 [<https://dimedb.ivers.aber.ac.uk/>] interfaces.

10 A literature search was conducted via PubMed, into pleural effusion biomarkers
 11 using a variety of analytical platforms. 7 studies which reported sufficient details to
 12 identify metabolites were included in our analyses [11]–[17].

13

14 **Results**

15 PCA show that patients clustered between the two broad primary pathologies of
 16 cancer and non-cancer (Fig. 1A). The major sources of variation were investigated,
 17 and 5 metabolites increased in cancer with AUC accuracies of 0.914 (95% CI: 0.839-
 18 0.976) with a sensitivity of 0.97 (95% CI 0.952-0.979) and specificity 0.87 (95% CI
 19 0.852-0.894). (Fig. 1B). These metabolites were lipid steroids; hydroxyprogesterone
 20 (AUC 0.85, 95% CI: 0.767-0.913), stigmastentriol (AUC 0.838, 95% CI: 0.773-0.901),
 21 hydroxycholesterol (AUC 0.84, 95% CI: 0.773-0.897) and acylcarnitines; hexadec-2-
 22 enoyl carnitine (AUC 0.83, 95% CI: 0.739-0.898), hydroxyhexadecanoylcarnitine
 23 (AUC 0.81, 95% CI: 0.726-0.873).

24 Due to limitations in using multivariate models with relatively small sample sizes, we
 25 employed another multivariate classifier approach, Random Forest (RF). This

1 confusion matrix showed a class error for the cancer group of only 0.12 and for non-
2 cancer of only 0.03. The out-of-bag, error was used as a test sample to obtain an
3 unbiased estimate of the classification error, which was only 0.066.

4 Our data was interrogated for previously published pleural fluid biomarkers. These
5 were identified on accurate mass (5 ppm) and predicted positive and negative
6 ionisation patterns. Where metabolites could also be found in our data, pairwise
7 comparisons were conducted, and AUC values were generated (Supplementary
8 Figure 1). Univariate or multivariate assessments of these previously published
9 biomarkers did not perform as well as the five that we targeted in our data.

10

11 **Discussion**

12 This study demonstrates that metabolomic approaches can be applied to stored
13 pleural fluid to differentiate malignancies from non-cancer aetiologies, with a high
14 sensitivity and specificity of 97% and 87% respectively. No other study has reported
15 untargeted FIE-MS to pleural fluid. Due to the exploratory nature of this pilot study,
16 no external validation set was available. However, our data was used as a validation
17 set for previously published pleural fluid biomarkers.

18 The metabolites we discovered included hydroxyprogesterone, stigmastentriol, and
19 hydroxycholesterol. These indicate an upregulation of lipid-related metabolic
20 pathways in pleural effusions caused by cancer. This was further indicated by the
21 acyl carnitines which is the form that long-chain fatty acids take as they cross the
22 inner mitochondrial membrane for subsequent β -oxidation [18]. Increased levels of
23 these metabolites are consistent with the increased fatty acid synthase expression
24 which occurs during carcinogenesis [19]. Further, our results compliment the findings
25 of Ziananro et al., who linked the differentiation of cancer from non-cancer pleural

1 effusions using proton magnetic resonance (^1H -NMR) techniques, to increased lipid
 2 metabolism, along with increased expression of branched amino acids and lactate
 3 metabolism [17]. Che et al., also applied metabolomics to pleural fluid using
 4 metabolomics to differentiate the underlying causes of pleural fluid accumulation.
 5 Their study focused on tuberculous versus cancer effusions, achieving a sensitivity
 6 of 93% and specificity of 86% [16]. Both our studies indicate that metabolomics
 7 offers a potential real-world benefit in the diagnosis and differentiation of pleural
 8 effusion cause.

9 Other genetic and protein biomarkers have been reported to differentiate the cause
 10 of pleural effusions but very few are accurate enough to be accepted into clinical
 11 practice yet. Metabolomics does offer potential to identify a suite of new potential
 12 biomarkers. Our biomarkers appear to be more accurate than previously reported
 13 biomarkers for differentiating between Cancer and Non-cancer causes (Table 1).
 14 However, this observation highlights the requirement for a larger, multisite, blinded
 15 study to validate all the pleural biomarkers that have been defined to date.

16 In conclusion, we have further demonstrated the promise of untargeted metabolomic
 17 based discovery of biomarkers, for differentiating cancer vs non-cancer pleural
 18 effusions. We report the importance of lipid biomarkers and fatty acid synthase
 19 expression. The increased lipid turnover is biologically plausible and reported in
 20 cancer metabolism. However, future validation on larger and different clinical sets is
 21 required.

22

23 **Acknowledgments**

1 We appreciate the technical support provided by Helen Phillips with mass
2 spectrometry (Aberystwyth University, UK).

3

4 **Financial support**

5 RdC was funded by a Tenovus iGrant TIG2016-26. KL is supported by a Knowledge
6 Economy Skills Scholarship (KESS 2) and further supported by European Social
7 Funds (ESF).

8

9

10

11

1 **Legend**

2

3 **Figure 1. Metabolomics can distinguish between malignant and non-malignant**
4 **causes of pleural effusions.**

5 **(A)** Principal Component Analysis (PCA) the larger circles in the PCA represent 95%
6 CI for each sample class) and **(B)** multivariate ROC curve comparison evaluated the
7 performance of CANCER and non-cancer (NON) samples

8

Table 1: Assessment of putative PE biomarkers from the literature

2

Name	AUC	T-test (<i>P</i>)	Log ₂ FC	Reference
R-2-Hydroxyhexadecanoic acid	0.7313	7.52E-10	0.77869	Ho et al., (2016)
Choline	0.7162	4.17E-08	0.73284	Wang et al., (2017)
Oleic acid	0.679	2.16E-06	0.75869	Lam & Law (2014)
Linoleic acid	0.6742	6.38E-06	0.79921	Lam & Law (2014)
Ceramide (d18:1/16:0)	0.6718	0.002699	0.44135	Lam & Law (2014)
Docosapentaenoic acid	0.6714	6.29E-04	0.59109	Ho et al., (2016)
Eicosatrienoic acid	0.6635	0.00262	0.49347	Ho et al., (2016)
Ornithine	0.6623	2.24E-04	-0.56498	Wang et al., (2017)
Linolenic acid	0.6546	4.54E-05	0.52851	Lam & Law (2014)
Tetradecadienoic acid	0.6431	0.001177	0.46765	Ho et al., (2016)
LysoPEtn(P16:0)	0.6376	0.001096	0.78928	Ho et al., (2016)
Eicosapentaenoic acid	0.6253	0.026597	0.3196	Ho et al., (2016)
Adrenic acid	0.6151	0.11403	0.32006	Ho et al., (2016)
Tyrosine	0.596	0.013926	-0.38366	Wang et al., (2017)
SM(d42:2)	0.59435	0.011886	-0.50853	Ho et al., (2016)
Creatinine	0.5898	0.071719	0.58439	Wang et al., (2017)

3

4

5

6

REFERENCES

- [1] R. W. Light, 'Pleural Effusions', *Med. Clin. North Am.*, vol. 95, no. 6, pp. 1055–1070, Nov. 2011, doi: 10.1016/j.mcna.2011.08.005.
- [2] R. W. Light, 'Pleural Effusions: The Diagnostic Separation of Transudates and Exudates', *Ann. Intern. Med.*, vol. 77, no. 4, p. 507, Oct. 1972, doi: 10.7326/0003-4819-77-4-507.
- [3] S. Romero-Candeira, C. Fernández, C. Martín, J. Sánchez-Paya, and L. Hernández, 'Influence of diuretics on the concentration of proteins and other components of pleural transudates in patients with heart failure', *Am. J. Med.*, vol. 110, no. 9, pp. 681–686, Jun. 2001, doi: 10.1016/S0002-9343(01)00726-4.
- [4] Emmet. E. McGrath and Paul. B. Anderson, 'Diagnosis of Pleural Effusion: A Systematic Approach', *Am. J. Crit. Care*, vol. 20, no. 2, pp. 119–128, Mar. 2011, doi: 10.4037/ajcc2011685.
- [5] M. Haznadar *et al.*, 'Urinary Metabolite Risk Biomarkers of Lung Cancer: A Prospective Cohort Study', *Cancer Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol.*, vol. 25, no. 6, pp. 978–986, 2016, doi: 10.1158/1055-9965.EPI-15-1191.
- [6] D. X. Assad *et al.*, 'Salivary metabolites to detect patients with cancer: a systematic review', *Int. J. Clin. Oncol.*, Mar. 2020, doi: 10.1007/s10147-020-01660-7.
- [7] L. Deng *et al.*, 'Development and Validation of a High-Throughput Mass Spectrometry Based Urine Metabolomic Test for the Detection of Colonic Adenomatous Polyps', *Metabolites*, vol. 7, no. 32, Jun. 2017, doi: 10.3390/metabo7030032.

- 1 [8] L. Deng *et al.*, 'Clinical validation of a novel urine-based metabolomic test for the
2 detection of colonic polyps on Chinese population', *Int. J. Colorectal Dis.*, vol. 32,
3 no. 5, pp. 741–743, May 2017, doi: 10.1007/s00384-016-2729-9.
- 4 [9] R. A. Spicer, R. Salek, and C. Steinbeck, 'Compliance with minimum information
5 guidelines in public metabolomics repositories', *Sci. Data*, vol. 4, Sep. 2017, doi:
6 10.1038/sdata.2017.137.
- 7 [10] J. Xia, I. V. Sinelnikov, B. Han, and D. S. Wishart, 'MetaboAnalyst 3.0—
8 making metabolomics more meaningful', *Nucleic Acids Res.*, vol. 43, no. Web
9 Server issue, pp. W251–W257, Jul. 2015, doi: 10.1093/nar/gkv380.
- 10 [11] C.-W. Lam and C.-Y. Law, 'Untargeted Mass Spectrometry-Based
11 Metabolomic Profiling of Pleural Effusions: Fatty Acids as Novel Cancer
12 Biomarkers for Malignant Pleural Effusions', *J. Proteome Res.*, vol. 13, no. 9, pp.
13 4040–4046, Sep. 2014, doi: 10.1021/pr5003774.
- 14 [12] Y. S. Ho *et al.*, 'Lipidomic Profiling of Lung Pleural Effusion Identifies Unique
15 Metabotype for EGFR Mutants in Non-Small Cell Lung Cancer', *Sci. Rep.*, vol. 6,
16 p. 35110, 14 2016, doi: 10.1038/srep35110.
- 17 [13] C. Wang, J. Peng, Y. Kuang, J. Zhang, and L. Dai, 'Metabolomic analysis
18 based on ¹H-nuclear magnetic resonance spectroscopy metabolic profiles in
19 tuberculous, malignant and transudative pleural effusion', *Mol. Med. Rep.*, vol.
20 16, no. 2, pp. 1147–1156, Feb. 2017, doi: 10.3892/mmr.2017.6758.
- 21 [14] C.-W. Lam and C.-Y. Law, 'Pleural Effusion Lipoproteins Measured by NMR
22 Spectroscopy for Diagnosis of Exudative Pleural Effusions: A Novel Tool for
23 Pore-Size Estimation', *J. Proteome Res.*, vol. 13, no. 9, pp. 4104–4112, Sep.
24 2014, doi: 10.1021/pr5004856.

- 1 [15] H. Liu, H. Wang, C. Li, L. Wang, Z. Pan, and L. Wang, 'Investigation of volatile
2 organic metabolites in lung cancer pleural effusions by solid-phase
3 microextraction and gas chromatography/mass spectrometry', *J. Chromatogr. B*,
4 vol. 945–946, pp. 53–59, Jan. 2014, doi: 10.1016/j.jchromb.2013.11.038.
- 5 [16] N. Che *et al.*, 'Integrated semi-targeted metabolomics analysis reveals distinct
6 metabolic dysregulation in pleural effusion caused by tuberculosis and
7 malignancy', *Clin. Chim. Acta Int. J. Clin. Chem.*, vol. 477, pp. 81–88, Feb. 2018,
8 doi: 10.1016/j.cca.2017.12.003.
- 9 [17] L. Zennaro, P. Vanzani, L. Nicolè, R. Cappellesso, and A. Fassina,
10 'Metabonomics by proton nuclear magnetic resonance in human pleural
11 effusions: A route to discriminate between benign and malignant pleural effusions
12 and to target small molecules as potential cancer biomarkers', *Cancer*
13 *Cytopathol.*, vol. 125, no. 5, pp. 341–348, May 2017, doi: 10.1002/cncy.21832.
- 14 [18] N. Longo, C. A. di San Filippo, and M. Pasquali, 'DISORDERS OF
15 CARNITINE TRANSPORT AND THE CARNITINE CYCLE', *Am. J. Med. Genet.*
16 *C Semin. Med. Genet.*, vol. 142C, no. 2, pp. 77–85, May 2006, doi:
17 10.1002/ajmg.c.30087.
- 18 [19] J. A. Menendez and R. Lupu, 'Fatty acid synthase and the lipogenic
19 phenotype in cancer pathogenesis', *Nat. Rev. Cancer*, vol. 7, no. 10, Art. no. 10,
20 Oct. 2007, doi: 10.1038/nrc2222.

21

(A) PCA

(B) ROC-AUC based on five variables