

**Untargeted metabolic profiling of saliva by liquid chromatography-mass spectrometry for the identification of potential diagnostic biomarkers of asthma**

Aditya Malkar<sup>1</sup>, Emma Wilson<sup>2</sup>, Tim Harrison<sup>2</sup>, Dominick Shaw<sup>2\*</sup> and Colin S. Creaser<sup>1\*</sup>.

<sup>1</sup>Centre for Analytical Science, Department of Chemistry, Loughborough University, Loughborough, Leicestershire, LE11 3TU, UK.

<sup>2</sup>Respiratory Research Unit, Nottingham City Hospital, Nottingham, NG5 1PB, UK

\*Joint corresponding authors

For correspondence:

E-mail: [C.S.Creaser@lboro.ac.uk](mailto:C.S.Creaser@lboro.ac.uk)

Tel: +44(0)1509 222552

Fax: +44(0)1509 223925

Keywords: Asthma, metabolite profiling, LC-MS, saliva

## **Abstract**

Current clinical tests employed to diagnose asthma are inaccurate and limited by their invasive nature. New metabolite profiling technologies offer an opportunity to improve asthma diagnosis using non-invasive sampling. A rapid analytical method for metabolite profiling of saliva is reported using ultra-high performance liquid chromatography combined with high resolution time-of-flight mass spectrometry (UHPLC-MS). The only sample pre-treatment required was protein precipitation with acetonitrile. The method has been applied to a pilot study of saliva samples obtained by passive drool from well phenotyped patients with asthma and healthy controls. Stepwise data reduction and multivariate statistical analysis was performed on the complex dataset obtained from the UHPLC-MS analysis to identify potential metabolomic biomarkers of asthma in saliva. Ten discriminant features were identified that distinguished between moderate asthma and healthy control samples with an overall recognition ability of 80% during training of the model and 97% for model cross-validation. The reported method demonstrates the potential for a non-invasive approach to the clinical diagnosis of asthma using mass spectrometry-based metabolic profiling of saliva.

## 1. Introduction

Asthma is a heterogeneous disease<sup>1</sup> characterised by symptoms with variable airflow obstruction, bronchial hyper-responsiveness and airway inflammation. The diagnosis of asthma in primary care is based on the presence or absence of airflow limitation, and its reversibility to inhaled bronchodilators, as measured by peak expiratory flow or spirometry. According to the guidelines, using spirometric values or peak flow as the prime outcome of interest is of limited value in patients with normal or near normal pre-treatment lung function since there is little room for measurable improvement<sup>2</sup>; furthermore it is estimated that 30% of patients with a diagnosis of asthma have no evidence of the disease.<sup>3-5</sup> Biomarkers of airway inflammation and airway hyperresponsiveness obtained from various invasive and non-invasive physiological sources such as sputum, blood, serum, urine, and exhaled breath have been used to diagnose asthma.<sup>6</sup> These tests differ in their sensitivity, specificity and patient acceptance. More importantly they cannot be used below certain lung function thresholds or ages, limiting their application and widespread adoption. Consequently there is a need for a simple, non-invasive procedure for the accurate identification of asthma.<sup>7</sup>

Metabolic profiling is the systematic study of chemical fingerprints of low molecular mass endogenous metabolites, known as the metabolome; the complete set of small-molecule chemicals found within a biological sample. The study of metabolic profiles enables the measurement of changes in metabolite concentrations, which in turn reflects the biochemical effects produced in the organism as a function of disease state or therapeutic intervention.<sup>8,9</sup> As a result, metabolic biomarkers can aid early disease detection and provide better understanding of disease incidence and progression. Metabolic profiling has proved to be a valuable tool for diagnosis and personalised medicine in other areas including pharmacology.<sup>10</sup>

The application of metabolic profiling to the discovery of biomarkers of asthma has been explored in biofluids such as urine, plasma and exhaled breath condensate.<sup>11-14</sup> Saliva collection offers distinct advantages as a preferred diagnostic medium; sample collection is easy, rapid and non-invasive and does not require specialised training or equipment. Patients with significant airflow limitation can perform a passive drool sample which is a major advantage in asthma and chronic obstructive pulmonary disease (COPD). Substances present in saliva are biologically active,<sup>15-17</sup> and saliva has been used for metabolomic studies for biomarker discovery in conditions such as physiological stress,<sup>18</sup> periodontal disease<sup>19</sup> and

leukoplakia.<sup>20</sup> We report the development of a rapid method for the metabolic profiling of saliva using ultra performance liquid chromatography-mass spectrometry (UHPLC-MS). The method has been applied to a pilot study to assess the potential as a tool for asthma diagnosis.

## **2. Methods and materials**

### **2.1. Ethical statement**

This research followed the good clinical practices and protocols as approved by the Nottingham Research Ethics committee 2 (REC reference number 10/H0408/34). Informed consent was obtained for all participants.

### **2.2. Participants**

The participants, 9 people with asthma (3 male and 6 female) and 21 healthy individuals (7 male and 14 female), were identified from an asthma database held in Nottingham City Hospital (Supplementary Fig. S1.). The mean age of the people with asthma was 53 (range 28-66) and of the healthy controls 35 (range 21-65). (Supplementary Table S1). The classification of asthma was based on established diagnostic tests for asthma including forced expiratory volume (FEV1%), sputum eosinophil count and methacholine challenge test<sup>21</sup> Participants all had well controlled asthma diagnosed by a consultant respiratory physician with a normal asthma control questionnaire<sup>22</sup> score over the last 3 months, and had not required oral steroids within a year. Tests of airway function, including methacholine PC<sub>20</sub> and induced sputum had all been performed between 1 and 6 months prior to saliva collection (Supplementary Fig. S1 and Table S1.).

### **2.3. Collection of saliva samples**

Sampling of whole saliva was carried out by the passive drool approach<sup>23</sup>. Each participant was seated with his or her head tilted forward causing saliva to pool in front of his or her mouth. Participants were asked not to use their asthma medications or consume alcohol for 12 hours prior to saliva collection. Participants were also asked to refrain from brushing their teeth, having a large meal, drinking tea or coffee and smoking for an hour prior to the sample collection. Saliva drooled from the mouth was collected in a glass vial (30mL Chromacol, UK) and aliquoted (600 µL) into micro-centrifuge tubes (2.0 mL, LoBind Eppendorf, UK) for storage. Equal volume sub-aliquots (600 µL) of saliva from all the participants were pooled together, mixed and re-aliquoted as 600 µL aliquots to be used for method development and quality control (QC). Samples were stored at -80 °C immediately after aliquoting.

## 2.4. Chemicals and reagents

Acetonitrile (analytical grade), water (analytical grade) and formic acid (>99%) were purchased from Fisher Scientific (Loughborough, UK). L-carnitine (>99%) and hydrocortisone (>99%) were obtained from Sigma-Aldrich (Gillingham, UK). A standard mixture of 0.005 mg/mL L-carnitine and hydrocortisone was prepared in water:acetonitrile (95:05) (v/v) with 0.1% formic acid. A ToF Reference Mass Solution kit (Agilent Technologies, Santa Clara, CA, USA) consisting of purine (m/z 121.0509) and Hexakis(1H, 1H, 3H-tetrafluoropropoxy)phosphazine (m/z 922.0098) was used to prepare a reference mass solution for acquiring accurate mass data.

## 2.5. Sample pre-treatment

Samples stored at -80 °C were thawed to room temperature before sample clean-up by protein precipitation. Each thawed saliva sample was vortexed for 30 seconds followed by ultra-sonication for 1 minute to breakdown mucous substances in saliva and to improve homogeneity. Protein precipitation was achieved by the addition of 1.2 mL acetonitrile to 600 µL of saliva i.e. 2:1 ratio of precipitant to saliva <sup>24</sup>. The mixture was vortexed for 30 seconds followed by ultra-sonication for 1 minute and then subjected to centrifugation at 10,000 g for 10 minutes at ambient temperature. Precipitated proteins from the sample were removed as a pellet at the bottom of the microcentrifuge tube. The supernatant was transferred to a fresh microcentrifuge tube and evaporated to near dryness (~5 µL) using Turbovap LV concentration workstation (Caliper Life Sciences, Hopkinton, MA, USA) prior to being reconstituted in 300 µL water/acetonitrile (95:05) (v/v) with 0.1% formic acid. The reconstituted saliva extract was placed in a polypropylene micro-insert (Supelco, UK) in an autosampler vial (2 mL) for UHPLC-MS analysis.

## 2.6. Instrumentation

LC-MS analyses were carried out using an Agilent 1200 series HPLC interfaced with an Agilent 6230 time-of-flight (TOF) mass spectrometer fitted with a JetStream ESI source operated in positive ion mode (Agilent Technologies, Santa Clara, CA, USA). UHPLC separation was carried out using a Poroshell 120 EC-C18 column (2.1 mm x 75 mm, 2.7 µm); Agilent Technologies, Santa Clara, CA, USA). Poroshell particles have a solid silica core (1.7 µm) and a porous silica outer layer (0.5 µm thickness), with an end capped C18 bonded phase is coated onto the porous outer layer of the particle (Kirkland et al., 2000).

Mobile phase A consisted of 0.1% aqueous formic acid (v/v) and mobile phase B was 0.1% formic acid in acetonitrile (v/v). The chromatographic gradient programme was optimised using a solution of L-carnitine and hydrocortisone (0.005 mg/mL). The mobile phase flow rate was set to 0.5 mL/min. Saliva extracts (10  $\mu$ L injected) were analysed using an optimised chromatographic gradient: 5% B (0-1 min), increased to 30% B (1-10 min). This was followed by a column clean-up phase, built into the method to reduce carry over and condition the column for analysis of subsequent samples, in which the mobile phase was increased to 95% B (10-11 min) and maintained at 95% B (11-12 min) before returning to initial conditions (13-15 min).

MS data were acquired with the instrument mode set to extended dynamic range (2 GHz). Mass spectra were acquired in the range  $m/z$  50-1200. The scan rate of the MS was 10 scans/s. Electrospray source conditions for the mass spectrometer were: sheath gas temperature and flow, 350  $^{\circ}$ C and 12 L/min; drying gas temperature and flow, 150  $^{\circ}$ C and 10 L/min; nebuliser pressure, 35 psig; transfer capillary, 4000 V; skimmer voltage, 65 V; fragmentor voltage, 150 V. A standard solution consisting of purine and hexakis(1H, 1H, 3H-tetrafluoropropoxy) phosphazine] was infused *via* the reference channel of the mass spectrometer. The responses from the ions at  $m/z$  121.0509 and  $m/z$  922.0098 were used to apply internal mass calibration to the mass spectrum for the analyte channel.

## **2.7. Method development and quality control**

The method workflow is shown in Supplementary Fig. S2. A series of five consecutive pooled saliva extracts were analysed prior to the analysis of the control and asthmatic saliva to condition the chromatographic column. The metabolomic sample list was randomised prior to analysis. A quality control (QC) sample was analysed with every batch of five samples to monitor the system stability and data quality. Method blanks were prepared by subjecting 500  $\mu$ L of water to the entire sample pre-treatment process and were analysed to assess contamination arising from polypropylene microcentrifuge tubes (LoBind Ependorf, UK) and polypropylene micro-inserts used during storage and sample pre-treatment. Method blanks demonstrated no contamination due to the polypropylene micro-centrifuge tubes and UHPLC micro-inserts used for storage, pre-treatment and analysis.

## 2.8. Data processing

Mass Hunter Qualitative Analysis software version B.05.00 (Agilent Technologies, Santa Clara, CA, USA) was used for processing raw data. Molecular feature extraction (MFE) algorithm was used to extract unknown molecular features (MF) and compile a data matrix for each sample.

The generated data matrix consisting of the molecular features from all the samples were exported into Mass Profiler Professional software Version 12.05 (Agilent Technologies, Santa Clara, CA, USA) for metabolomic data analysis including chromatographic peak alignment across multiple LC-MS data files and multivariate statistical analysis. Molecular features were extracted using following thresholds:  $m/z$  range 50-600 Da; retention time range 0.01-10.00 min and mass spectral peak height of ions (centroid data) 100 cps. The abundance value for each molecular feature was calculated by MassHunter software as the sum of intensities of corresponding isotopic and adducts peaks. The data was normalised to the total abundance value within a particular sample for each feature by the software. Data filtering and alignment was carried out using following parameters: minimum absolute abundance of molecular features 7000 cps, mass alignment window 10 ppm and retention time alignment window 0.15 min.

Stepwise reduction of the number of molecular features was performed based on their abundance values and frequency of occurrence in sample classes. Aligned molecular features were filtered to reduce the dimensionality of the dataset prior to principal component analysis (PCA). In the first instance, only the molecular features present in at least 85% of samples of at least one group (i.e. asthma or healthy control) were retained for further processing. In the next step, molecular features were filtered based on p-value calculated for each using a student's T-test. Lower p-value signifies higher significant difference between the conditions and  $P \leq 0.1$  was chosen as the filtering criterion to ensure that only the molecular features, which differ in their respective conditions with a statistical significance of 90%, were retained. The final criterion for filtration of molecular features was the fold change (FC). Molecular features, which satisfied the threshold of  $FC \geq 2.0$ , were retained for further processing by PCA and to build predictive models.



### 3. Results and Discussion

#### 3.1. UHPLC-MS analysis

A method has been developed for the rapid metabolic profiling of saliva using ultra performance liquid chromatography combined with high resolution mass spectrometry. The workflow for the method (Supplementary Fig. S2) and data reporting are consistent with proposed standards for metabolomics studies.<sup>25,26</sup> The method has been applied to a pilot study of saliva samples collected by passive drool from asthma patients and controls. Saliva extracts were analysed by UHPLC-MS following a protein precipitation step using acetonitrile. Protein precipitation was the only sample clean-up process employed in order to minimise the effects of discrimination and unpredictable behaviour of unknown metabolites during sample preparation.<sup>27</sup>

A typical LC-MS profile for a saliva extract overlaid with a method blank is shown in Fig. 1 demonstrating the complexity of the dataset. The chromatographic gradient was adjusted using a solution of L-carnitine and hydrocortisone (0.005 mg/mL) such that highly polar L-carnitine was slightly retained on the column whereas less polar hydrocortisone eluted towards the end of the 10 minute chromatographic gradient (data not shown). This enabled all the metabolites from saliva extract to be separated within 10 minutes with a typical peak width at base of 9 seconds; hence a total analysis time of 15 minutes was achieved, which is suitable for a metabolomic study.

The reproducibility of the method was determined as a part of method development using the retention times and chromatographic peak areas of randomly selected ions across the chromatographic gradient (Supplementary Table S2 and Fig. S3). The effect of injection volumes in the range of 1-20  $\mu$ L on chromatographic peak areas was analysed using pooled saliva extract samples, which were analysed in triplicate. The chromatographic peak areas of all the extracted endogenous metabolite ions showed linear responses ( $R^2 > 0.989$ ). An injection volume of 10  $\mu$ L was selected from these experiments to be used for the analysis of saliva extract, which ensured that the analytical method was capable of detecting small positive or negative changes in the metabolite concentrations.

The reproducibility of the retention times and chromatographic peak areas of extracted metabolite ions were assessed from replicate analyses of pooled saliva samples (n=6), using the optimised injection volume. Based on replicate measurements (n=6) of five endogenous

metabolites, the reproducibility of the retention time and chromatographic peak area was determined to be less than 0.4% and 4% respectively (Supplementary Table S2).

The method developed was applied to the analysis of asthma patient and control saliva samples. The resulting dataset were pre-processed and then compared by unsupervised multivariate statistical analysis using principal component analysis to identify discriminant molecular features.

### **3.2. Data reduction and multivariate analysis**

A molecular feature extractor (MFE) algorithm was employed, because of the complexity of the acquired datasets, which enabled the automated extraction of ions corresponding to compounds present in saliva samples derived from extracted ion chromatograms. This algorithm combines ion species such as multiply charged species, isotopes and adducts generated from a single compound into a single molecular feature (MF) to be utilised as a single variable in subsequent analysis.

A total of 35011 molecular features were extracted from all the samples. Out of these, 14267 were aligned across all the samples from the study. Principal component analysis (PCA) was carried out using these 14267 molecular features on all the sample groups from the study including conditioning runs and quality control (QC) samples. Clustering of QC samples (circled in Fig. 2) provides evidence of consistent and reliable data quality across all the samples demonstrating the suitability of the method for metabolic profiling studies. However, Fig. 2 shows no clear separation of asthmatic and healthy population; hence, further data reduction was required to identify discriminant metabolites between control and asthmatic samples.

Participants with asthma (n=9) were compared against healthy individuals (n=21). The dimensionality of the dataset was reduced prior to PCA by filtering the data based on frequency of occurrence of molecular features and outcomes of t-test (summarised in a Venn diagram in Fig. 3 (a)). In the first step, all molecular features which were present in at least 85% of samples of at least one group (i.e. asthmatic or healthy) were retained. This reduced the number of MFs from 14627 to 188. In the next phase, the MFs were filtered based on their p-values obtained from a t-test and fold change analysis. MFs, which satisfied the p-value cut-off value of 0.1 and fold change value of  $FC \geq 2.0$  were retained as illustrated from the volcano plot in Fig 3 (b). This reduced the number of MFs from 188 to 10

(Supplementary Table S3). A PCA plot was constructed based on the filtered 10 molecular features for moderate asthmatic and healthy control individuals (Fig. 4). The PCA model shows clear separation of moderate asthmatics from healthy control samples, with one asthma patient sample falling within the control domain. The reasons for this outlier are unknown, but this may reflect the known heterogeneity in asthma or be an artefact of the relatively small sample set.

The correlation and covariance of the 10 discriminant features was assessed from an s-plot (Supplementary Fig S4). MFs characteristic of asthmatics (i.e. up-regulated) are represented in the first quadrant of the plot and the MFs characteristic of healthy control individuals (i.e. down-regulated) are shown in the third quadrant of the plot. Table 1 summarises the biomarker ions identified from this study. The box and whisker plots for these metabolites are provided in Supplementary Figs. S5 and S6. Accurate mass measurement data and isotopic ratio data on the discriminant metabolite ions were obtained by analysing a pooled saliva extract sample using high resolution mass spectrometry and utilised to obtain tentative elemental compositions on the discriminant metabolite ions with a mass window of 3 ppm (Table 1). Putative annotations<sup>25</sup> for some of the discriminant ions based on physicochemical properties (5 ppm mass window and isotope abundance) and spectral similarity with metabolome databases are given in supplementary Table S3.<sup>28-30</sup> Identification of these ions requires further investigation involving tandem mass spectrometry and comparison with reference standards. Tandem mass spectrometry was not possible with the time-of-flight mass analyser used in this work, but the method could be readily adapted for use with a tandem mass spectrometer in combination with data dependent or data independent acquisition.

### **3.3. Predictive models**

The discrimination potential offered by the data was evaluated using a widely utilised supervised pattern recognition method of partial least squares discriminant analysis (PLS-DA). A statistical model based on PLS-DA was constructed and validated for classification of asthmatic samples. The results of sample classification are presented in Table 2 in terms of recognition and prediction abilities, representing the percentage of the samples correctly classified during model training and cross-validation. The overall recognition ability during training of the model was 80% and for model cross-validation was 96.7%. The model cross validation methodology used was k-fold cross-validation with 10 iterations of 3-folds (i.e. k=3). These training and cross validation data compare favourably to current tests used

for the diagnosis of asthma.<sup>31,32</sup> More invasive tests such as bronchial challenge or induced sputum are available, but are limited by their invasive nature, cost and need for expertise. They also only reflect one aspect of the asthmatic process.<sup>33</sup> Despite these tests a proper diagnosis remains a challenge even in specialist asthma clinics, emphasising the need for a non-invasive saliva-based test.<sup>34</sup> However, further longitudinal studies are required to confirm the diagnostic utility of LC-MS in this variable heterogeneous condition.

#### 4. Conclusions

The rapid saliva metabolite profiling method developed in this work allowed discrimination between asthma and healthy controls with good sensitivity and reproducibility. To our knowledge, this is the first study to employ a passive drool saliva test combined with UHPLC-MS as a potential discriminatory tool for asthma. A model derived from 10 discriminant features classified asthma with an accuracy of  $\geq 80\%$  for model training and cross-validation. These figures compare favourably to current tests used for the diagnosis of asthma.

The benefits of a simple, non-invasive sampling technique for asthma diagnosis are obvious and a great deal of work has focussed on the development of breath sampling, including exhaled nitric oxide<sup>35</sup>, measurement of volatile organic compounds<sup>13,36</sup> and the e-nose<sup>37</sup>. The discriminatory ability of these tests varies. Importantly the passive drool technique can be used in conjunction with or as a preliminary alternative to the existing diagnostic tests such as FEV1%, has good discriminatory ability, and provides metabolic information on the status of the airway. The identification, validation and the biological significance of the discriminant molecular features identified in this study also need to be validated and quantified using a larger cohort of moderate and/or severe asthmatics versus healthy control populations using a targeted metabolite analysis approach. However, the passive drool technique combined with UHPLC-MS offers a novel non-invasive approach to differentiating patients with asthma from healthy controls.

## **Acknowledgements**

The authors wish to acknowledge the Mass Profiler Professional software technical support provided by Agilent Technologies, and wish to specially thank Libin Joy at Strand Life Sciences Pvt. Ltd. for his help with the software. Authors would also like to thank William EP Greenland for training on the MPP software, Wendy Gerrard-Tarpey and Glenn Hearson at Nottingham City Hospital for help with sample collection.

Study funding- The study was funded by an internal grant for the Universities of Loughborough and Nottingham.

Conflicts of interest- The authors declare no conflict of interest.

## References

1. G. P. Anderson, *Lancet*, 2008, **372**, 1107–1419.
2. British Thoracic Society, Scottish Intercollegiate Guidelines Network. British Guideline on the Management of Asthma: A National Clinical Guideline, 2012.
3. D. Shaw, R. Green, M. Berry, S. Mellor, B. Hargadon, M. Shelley *et al.*, *Prim. Care Respir. J.*, 2012, **21**, 283–2877.
4. L-P. Boulet. *Curr. Allergy Asthm. R.*, 2003, **3**, 166–171.
5. B. Marklund, A. Tunsäter and C. Bengtsson, *Fam. Pract.*, 1999, **16**, 112–116.
6. S. Wadsworth, D. Sin and D. Dorscheid, *J. Asthma Allergy*, 2011, **4**, 77–86.
7. S. J. Szeffler, S. Wenzel, R. Brown, S. C. Erzurum, J. V. Fahy, R.G. Hamilton, *et al.*, *J. Allergy Clin. Immunol.*, 2012, **129**, S9–23.
8. J. K. Nicholson and J. C. Lindon, *Nature*. 2008, **455**, 1054–1056.
9. A. Smolinska, L. Blanchet, L. M. C. Buydens and S. S. Wijmenga, *Anal. Chim. Acta*, 2012, **750**, 82–97.
10. A. Nordström and R. Lewensohn. *J. Neuroimmune Pharmacol.*, 2010, **5**, 4–17.
11. E. J. Saude, C. D. Skappak, S. Regush, K. Cook, A. Ben-Zvi, Becker A, *et al.*, *J. Allergy Clin. Immunol.*, 2011, **127**, 757–764.
12. E. Mattarucchi, E. Baraldi and C. Guillou, *Biomed. Chrom.*, 2012, **26**, 89–94.
13. B. Ibrahim, M. Basanta, P. Cadden, D. Singh, D. Douce, A. Woodcock *et al.*, *Thorax* 2011, **66**, 804–809.
14. S. Carraro, S. Rezzi, F. Reniero, K. Héberger, G. Giordano, S. Zanconato *et al.*, *Am J Respir Crit Care Med.*, 2007, **175**, 986–990.
15. R. Pink, J. Simek, J. Vondrakova, E. Faber, P. Michl, J. Pazdera *et al.*, *Biomed. Pap. Med. Fac. Univ. Palacky Olomouc Czech Repub.*, 2009, **153**, 103–110.
16. A. Zhang, H. Sun and X. Wang. *Appl. Biochem. Biotech.*, 2012, **168**, 1718–1727.
17. R. G. Schipper, E. Silletti and M. H. Vingerhoeds. *Arch. Oral Biol.*, 2007, **52**, 1114–1135.
18. A. Malkar, N. A. Devenport, H. J. Martin, P. Patel, M. A. Turner, P. Watson *et al.*, *Metabolomics*, 2013, **9**, 1192–1201.

19. M. Aimetti, S. Cacciatore, A. Graziano and L. Tenori, *Metabolomics*, 2011, **8**, 465–474.
20. J. Wei, G. Xie, Z. Zhou, P. Shi, Y. Qiu, X. Zheng *et al.*, *Int. J. Cancer*, 2010 **129**, 2207–2217.
21. L. J. Fetters and J. I. Matthews, *Arch. Intern. Med*, 1984, **144**, 938–940.
22. E. F. Juniper, P. M. O’Byrne, G. H. Guyatt, P. J. Ferrie and D. R. King, *Eur. Respir. J.*, 1999, **14**, 902–907.
23. S. Chiappin, G. Antonelli, R. Gatti and E. F. De Palo, *Clin. Chim. Acta*, 2007, **383**, 30–40.
24. C. Polson, P. Sarkar, B. Incledon, V. Raguvaran and R. Grant, *J. Chromatogr. B*, 2003, **785**, 263–275.
25. L. W. Sumner, A. Amberg, D. Barrett, M. H. Beale, R. Beger *et al.*, *Metabolomics*, 2007, **3**, 211–221.
26. R. M. Salek, C. Steinbeck, M. Viant, R. Goodacre and W. B. Dunn, *GiggaScience*, 2013, **2**, 13–15.
27. B. Álvarez-Sánchez, F. Priego-Capote and M. D. Luque de Castro, *J. Chromatogr. A*, 2012, **1248**, 178–181.
28. D. S. Wishart, D. Tzur, C. Knox, *et al.*, HMDB: the Human Metabolome Database. *Nucleic Acids Res.* 2007, **35**, D521–526.
29. D. S. Wishart, C. Knox, A. C. Guo, *et al.*, HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.* 2009, **37**, D603–610.
30. D. S. Wishart, T. Jewison, A. C. Guo, M. Wilson, C. Knox *et al.*, HMDB 3.0 — The Human Metabolome Database in 2013. *Nucleic Acids Res.* 2013, **41**, D801–817.
31. C. J. Hunter, C. E. Brightling, G. Woltmann, A. J. Wardlaw and I. D. Pavord. *Chest*, 2002, **121**, 1051–1057.
32. V. P. Luks, K. L. Vandemheen and S. D. Aaron, *Eur. Respir. J.*, 2010, **36**, 255–260.
33. C. L. Grainge, L. C. K. Lau, J. A. Ward, V. Dulay, G. Lahiff, Wilson S, *et al.*, *N. Engl. J. Med.*, 2011, **364**, 2006–2015.
34. D. S. Robinson, D. A. Campbell, S. R. Durham, J. Pfeffer, P. J. Barnes and K. F. Chung, *Eur. Respir. J.*, 2003, **22**, 478–483.
35. D. E. Shaw, M. Berry, M. Thomas, R. H. Green, C. E. Brightling, Wardlaw AJ, *et al.*, *Am J Respir Crit Care Med.*, 2007, **176**, 231–237.



36. P. Montuschi, M. Santonico, C. Mondino, G. Pennazza, G. Mantini, E. Martinelli, *et al.*, *Chest*, 2007, **137**, 790–796.
37. S. Dragonieri, O. Tongoussouva, A. Zanini, A. Imperatori and A. Spanevello, *Monaldi Arch. Chest Dis.*, 2009, **71**, 119–126.

**Table 1. Discriminant molecular features (MFs) obtained from moderate asthmatics versus control samples.**

<b>Marker No.</b>	<b>Retention Time (min)</b>	<b><i>m/z</i></b>	<b>Regulation</b>	<b>Proposed elemental composition</b>
1	0.4	116.0699	Up	C <sub>5</sub> H <sub>9</sub> NO <sub>2</sub>
2	2.2	261.1446	Up	C <sub>12</sub> H <sub>23</sub> N <sub>3</sub> O <sub>5</sub>
3	1.1	290.1711	Up	C <sub>11</sub> H <sub>20</sub> N <sub>2</sub> O <sub>5</sub>
4	3.6	316.2217	Up	C <sub>15</sub> H <sub>29</sub> N <sub>3</sub> O <sub>4</sub>
5	4.4	439.4569	Up	Unidentified <sup>a</sup>
6	3.4	227.1268	Down	C <sub>12</sub> H <sub>18</sub> O <sub>4</sub>
7	4.7	573.2561	Down	C <sub>26</sub> H <sub>44</sub> N <sub>4</sub> O <sub>4</sub> S <sub>3</sub>
8	0.5	596.3253	Down	C <sub>28</sub> H <sub>41</sub> N <sub>11</sub> O <sub>2</sub> S
9	1.7	345.1163	Down	C <sub>22</sub> H <sub>17</sub> ClN <sub>2</sub>
10	5.5	227.1396	Down	C <sub>11</sub> H <sub>18</sub> N <sub>2</sub> O <sub>3</sub>

<sup>a</sup> No elemental composition computed within  $\pm 3$  ppm mass window.

**Table 2. Classification results obtained by PLS-DA model for asthma patients and controls.**

	Predicted Asthmatic	Predicted Controls	Accuracy (%)
Model Training			
True Moderate Asthmatic	8	1	88.9
True Controls	5	16	76.2
Recognition ability			80.0
Model cross-validation			
True Moderate Asthmatic	9	0	100.0
True Controls	1	20	95.2
Recognition ability			96.7

## 5. List of Figures

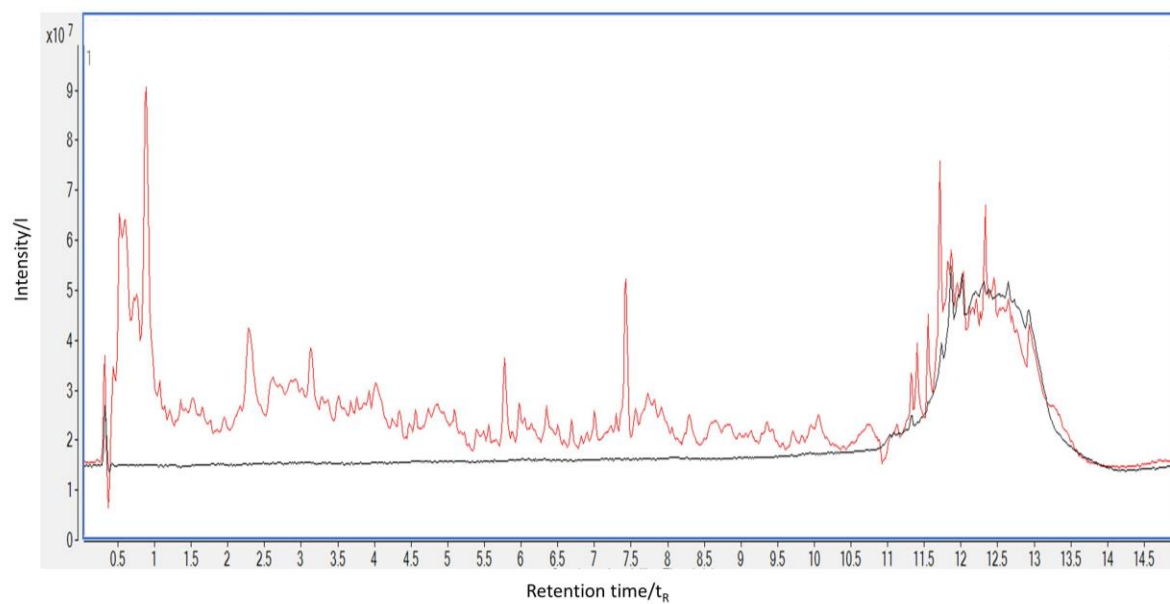
**Fig 1.** LC-MS analysis of a saliva extract showing the total ion chromatogram (TIC) for the saliva extract (red trace) and a method blank (black trace).

**Fig 2.** Principal component analysis 3D plot of saliva samples from the study coloured according to sample groups. X-Axis: Component 1 (20.07%), Y-Axis: Component 2 (13.26%) and Z-Axis: Component 3 (6.44%).

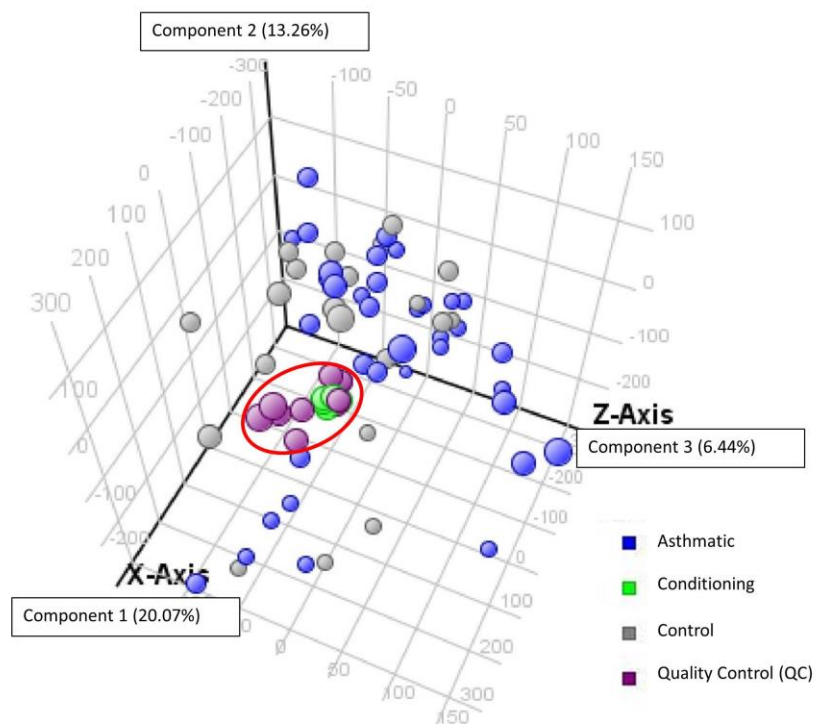
**Fig 3. a)** Summary of data reduction based on frequency of molecular features in samples, **b)** volcano plot illustrating filtering of molecular features based on the results of t-test and fold change analysis (P value cut off 0.1 and fold change cut off 2.0)

**Fig 4.** Principal component analysis 3D plot of saliva samples from moderate asthmatics (red) and healthy control samples (blue). X-Axis: Component 1 (49.08%); Y-Axis: Component 2 (15.47%) and Z-Axis: Component 3 (12.02%).

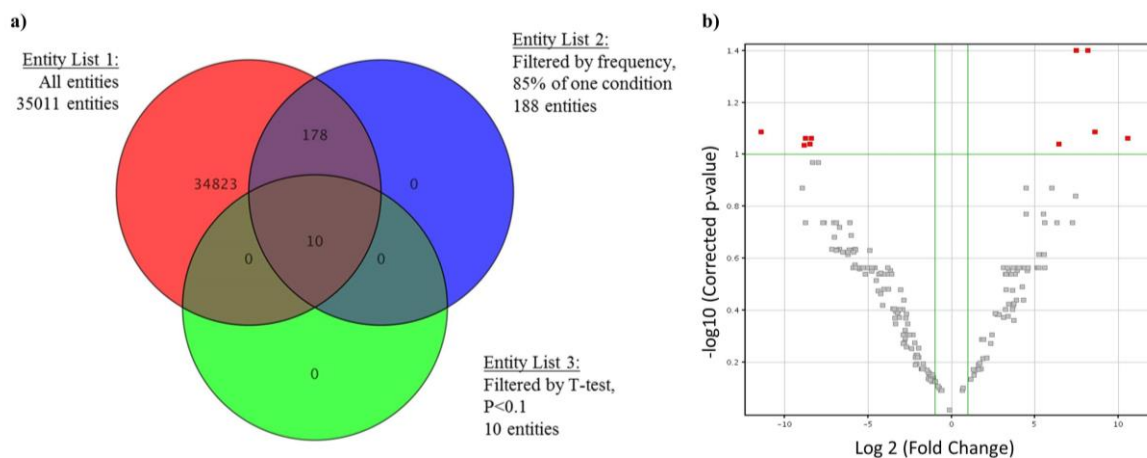
## 6. FIGURES



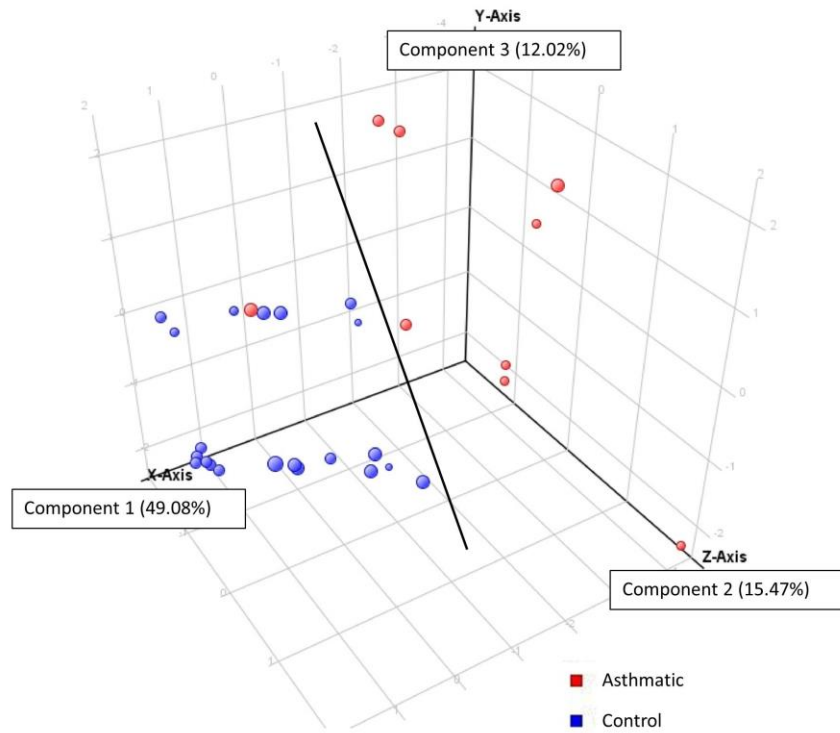
**Fig. 1.** LC-MS analysis of a saliva extract showing the total ion chromatogram (TIC) for the saliva extract (red trace) and a method blank (black trace).



**Fig 2. Principal component analysis 3-D plot of saliva samples from the study coloured according to sample groups. X-Axis: Component 1 (20.07%), Y-Axis: Component 2 (13.26%) and Z-Axis: Component 3 (6.44%).**



**Fig 3. a) Summary of data reduction based on frequency of molecular features in samples, b) Volcano plot illustrating filtering of molecular features based on the results of T-test and fold change analysis (P value cut off 0.1 and fold change cut off 2.0)**



**Fig 4. Principal component analysis 3-D plot of saliva samples from moderate asthmatics (red) and healthy control samples (blue). X-Axis: Component 1 (49.08%); Y-Axis: Component 2 (15.47%) and Z-Axis: Component 3 (12.02%).**