



Exhaled breath profiling in diagnosing wheezy preschool children

Kim D.G. van de Kant^{*}, Joep J.B.N. van Berkel[#], Quirijn Jöbsis^{*}, Valéria Lima Passos[†], Ester M.M. Klaassen^{*}, Linda van der Sande^{*}, Onno C.P. van Schayck⁺, Johan C. de Jongste[§], Frederik Jan van Schooten[#], Eduard Derks[‡], Edward Dompeling^{*} and Jan W. Dallinga[#]

ABSTRACT: Although wheeze is common in preschool children, the underlying pathophysiology has not yet been disentangled. Volatile organic compounds (VOCs) in exhaled breath may serve as noninvasive markers of early wheeze. We aimed to assess the feasibility of VOC collection in preschool children, and to study whether a VOC profile can differentiate between children with and without recurrent wheeze.

We included children (mean (range) age 3.3 (1.9–4.5) yrs) with (n=202) and without (n=50) recurrent wheeze. Exhaled VOCs were analysed by gas chromatography–time-of-flight mass spectrometry. VOC profiles were generated by ANOVA simultaneous component analysis (ASCA) and sparse logistic regression (SLR).

Exhaled breath collection was possible in 98% of the children. In total, 913 different VOCs were detected. The signal-to-noise ratio improved after correction for age, sex and season using ASCA pre-processing. An SLR model with 28 VOCs correctly classified 83% of the children (84% sensitivity, 80% specificity). After six-fold cross-validation, 73% were correctly classified (79% sensitivity, 50% specificity).

Assessment of VOCs in exhaled breath is feasible in young children. VOC profiles are able to distinguish children with and without recurrent wheeze with a reasonable accuracy. This proof of principle paves the way for additional research on VOCs in preschool wheezing.

KEYWORDS: Asthma, volatile organic compounds, wheezing

In preschool children, wheeze is a frequent reason for consulting a doctor. The group of wheezing preschool children is diverse, including children with asthma and children with transient, virus-induced symptoms. Due to this heterogeneity, different wheezing phenotypes have been defined [1]. These phenotypes are mainly based on the course of symptoms and their triggers. They do not represent biologically plausible classifications, since the underlying pathophysiology of preschool wheeze has not yet been disentangled.

High-throughput techniques, such as proteomics, genomics and metabolomics, have become of increasing importance in elucidating the pathways of multiple diseases [2]. In recent years, interest has emerged in applying the “omics” techniques in a noninvasive mode by using exhaled breath as a medium to assess biomarker profiles. Exhaled breath consists of hundreds of volatile organic

compounds (VOCs), which are formed by various inflammatory and metabolic pathways [3]. During the inflammatory process, reactive oxygen species (ROSs) are formed by inflammatory cells. ROSs react with lipid membrane structures, leading to the degradation of polyunsaturated fatty acids. Consequently, several stable breakdown products (VOCs) are formed [3, 4]. After entering the bloodstream, VOCs are excreted into breath due to their low solubility. In adults and older children, the study of metabolomics using exhaled VOCs is known to be a safe and noninvasive procedure to evaluate ongoing processes of inflammation and oxidative stress [3, 5]. Profiling exhaled VOCs has been successfully applied in the diagnosis of inflammatory lung diseases, such as cystic fibrosis [6], asthma [5] and chronic obstructive pulmonary disease (COPD) [7, 8].

There is increasing evidence that wheezing children have enhanced airway inflammation

AFFILIATIONS

^{*}Dept of Paediatric Pulmonology, School for Public Health and Primary Care (CAPHRI), Maastricht University Medical Centre (MUMC),

[#]Dept of Toxicology, Research Institute NUTRIM, MUMC,

[†]Dept of Methodology and Statistics, CAPHRI, MUMC, and

⁺Dept of General Practice, CAPHRI, MUMC, Maastricht,

[§]Dept of Paediatrics, Division Paediatric Respiratory Medicine, Erasmus University Medical Centre, Rotterdam, and

[‡]Dept of Analytics and Statistics, DSM Resolve, Geleen, The Netherlands.

CORRESPONDENCE

K.D.G. van de Kant
Maastricht University Medical Centre
Dept of Paediatric Pulmonology
PO Box 5800
6202 AZ Maastricht
The Netherlands
E-mail: kim.vande.kant@mumc.nl

Received:

July 18 2011

Accepted after revision:

March 28 2012

This article has supplementary material available from www.erj.ersjournals.com

and oxidative stress [9–11]. Therefore, we hypothesise that wheezing children and children without wheeze have different VOC profiles in their exhaled breath. However, no studies in preschool children have been performed so far. In this proof-of-principle study, we aimed to assess whether collection of exhaled VOCs is feasible in preschool children and whether a profile of VOCs can differentiate preschool children with recurrent wheeze from healthy controls.

METHODS

Study population

We included children from the ADEM (Asthma Detection and Monitoring) study in the Province of Limburg, the Netherlands (registered at www.clinicaltrials.gov with identifier number NCT00422747). In this study, children with recurrent wheezing symptoms ($n=202$) and control subjects with no wheezing symptoms ($n=50$) were included at the age of 2–4 yrs and were studied prospectively until 6 yrs of age. Recurrent wheeze was assessed using the ISAAC (International Study of Asthma and Allergy in Childhood) questionnaire and was defined as at least two episodes of wheeze during life [12]. The study protocol has been described in detail previously [13]. Ethical approval was obtained from the Dutch National Medical Ethical Committee (the Hague, the Netherlands) (identifier number CCMO:NL17407.000.07/2007-001817-40). All parents gave written informed consent.

Study design

For the current study, we analysed the data at inclusion. During the initial visit, measurements of exhaled breath and clinical patient characteristics were assessed. Inhaled corticosteroids (ICS) were stopped ≥ 4 weeks before this visit. Measurements were postponed for 4 weeks in the case of parentally reported symptoms of an airway infection, such as runny nose, sneezing, sore throat, cough, malaise or fever. Solid foods and exercise were not allowed within 1 h of the measurements. The presence of atopy was determined using a Phadiatop Infant test (Phadia, Uppsala, Sweden).

Exhaled breath analysis

Sample collection and analysis

Children breathed tidally through a facemask connected to a non-rebreathing valve system, while watching cartoons. On the expiratory port of the valve, a 1-L polycarbonate bag (Tedlar® bag; SKC Ltd, Blandford Forum, UK) was connected to collect exhaled breath. When the bag was filled, it was emptied across a stainless steel, two-bed sorption tube, filled with Carbograph 1 TD/Carbopack™ X (Markes International, Llantrisant, UK) for rapid adsorption and stabilisation of volatile compounds. The tubes were airtight capped and stored at room temperature until analysis. VOCs were released from the tube using thermal desorption (Unity desorption unit; Markes International). Next, the sample was injected into the gas chromatography (GC) capillary (Trace GC; ThermoFischer Scientific, Austin, TX, USA). In the GC capillary, VOCs were separated and subsequently detected and identified using time-of-flight (TOF) mass spectrometry (MS) (Tempus Plus; ThermoFischer Scientific). Detailed information about the conditions and settings of the GC–TOF–MS measurements has been provided previously [14].

Data pre-processing

The pre-processing of the raw GC–TOF–MS output files included automatic peak detection, baseline correction and normalisation of retention times [14]. Thereafter, identical compounds in all corrected files were combined into a large database file. Complementary compounds in different samples were linked based on similarity of retention times and mass spectra. Similarity of mass spectral data was based on calculating match factors [14]. Finally, normalisation of the peak area was performed to compare the different peak areas and accompanying intensities of the different compounds. We previously tested our analytical procedure (including sampling, storage and instrumental analysis) for reproducibility in both adults and school children [6, 14]. We demonstrated that instrumental, long-term and short-term intra-individual reproducibility was high [6, 14].

Statistical analysis

Data were analysed using R (version 2.10.1; University of Vienna, Vienna, Austria) and GLMnet toolbox (version 1.4) [15]. Clinical variables are presented as mean \pm SEM or mean (range). Categorical variables are presented as n (%). Differences in clinical characteristics between children with and without recurrent wheeze were evaluated by means of independent t -tests, Mann–Whitney U -tests and Chi-square tests for continuous parametric and nonparametric, and categorical variables, respectively. A significance level of $p < 0.05$ was used. Our dataset is characterised by a large number of variables measured compared to the number of subjects in the study (VOCs, $n=913$; children, $n=252$). This raises multicollinearity and singularity problems, which cannot be adequately solved by conventional statistical methods (neither multivariable nor multivariate). Therefore, we applied modern techniques to handle the high dimensionality of the data: ANOVA simultaneous component analysis (ASCA) and sparse logistic regression (SLR). ASCA was used as pre-processing tool. ASCA worked by splitting up the total VOC variation and assigning it to known sources unrelated to wheezing, *i.e.* the factors sex, height, weight, age, season of measurement and their interactions. The VOC residuals obtained by ASCA were subsequently entered as classifiers in a SLR model to study their effectiveness in predicting recurrent wheeze. The objective was to select a subset of VOCs that is highly discriminative between recurrent and nonrecurrent wheezers. SLR selected the most predictive VOCs by suppressing the noisy correlated VOCs, *i.e.* by pruning them from the model and, therefore, protecting against high false positive rates (fig. S1). In this way, the most informative VOCs were selected. SLR is a classification method that gives a sparse solution with high accuracy. More details about SLR are described by HASTIE *et al.* [16]. Only VOCs with a prevalence of $\geq 7\%$ in the population were included in the SLR analysis [14, 17]. Weights were applied to correct for class imbalance (202 wheezers and 50 nonwheezers). Moreover, a six-fold cross-validation was applied. The most predictive VOCs for recurrent wheeze were identified by using the National Institute of Standards and Technology library as described previously [14].

RESULTS

Clinical characteristics and feasibility

A total number of 258 children were invited to participate in the study. We were able to successfully collect exhaled breath

in 98% (n=252) of the children. Sudden anxiety was the major reason for drop-out of the remaining six children. Of the included children, 202 children had recurrent wheezing symptoms (mean (range) age 3.2 (1.9–4.5) yrs) while 50 children did not experience recurrent wheezing (age 3.3 (2.2–4.1) yrs). The clinical characteristics are described in table 1. Eczema was more frequent in children with recurrent wheeze compared with nonwheezers ($p < 0.05$). Around 40% of the wheezing children used a short-acting β_2 -agonist as rescue medication and 19% of the wheezing children were on maintenance treatment with ICS. ICS were stopped ≥ 4 weeks before the measurements and short-acting β_2 -agonists ≥ 8 h before the measurements. In seven patients, stopping of ICS was not permitted by the responsible physician because of severe asthma symptoms.

Exhaled VOCs

VOC profiles could be determined in all samples. In total, 913 different VOCs with a prevalence of $\geq 7\%$ could be detected. The number of exhaled VOCs per child was, on average, 350.

Covariates

An initial principal component analysis demonstrated that the data were divided in clear clusters irrespective of wheezing. We performed ASCA to determine which covariates were predictive of this clustering but were not predictive of recurrent wheeze. The responsible covariates included sex, age and season, of which only season mildly correlated with wheezing ($r = 0.125$, $p = 0.05$). The variance explained by sex, age and season varied between VOCs. The explained variance of season was $< 23\%$ for 99% of the VOCs (fig. 1). This percentile limit (1%) was 7% for age and 6% for sex (figs S2 and S3). Data variability explained by these variables (and their interaction) was subtracted from the VOC data in order to improve the signal-to-noise ratio.

Sparse logistic regression

SLR was applied to the ASCA residuals (VOCs data corrected for sex, age and seasonal effects). Linear weights were

computed to correct for class imbalance (202 wheezers, weight 0.62; 50 nonwheezers, weight 2.52). The SLR model for recurrent wheeze including the 28 most discriminative VOCs (with a lambda of 0.135) demonstrated a classification error of 0.17, with a sensitivity of 84% and a specificity of 80%. The receiver operating characteristic curve of this SLR model is presented in figure 2. Internal cross-validation of the data had a limited effect on sensitivity (79%) but diminished the specificity (50%), with an overall cross-validated classification error of 0.27. The most discriminative VOCs of the SLR model are presented in table 2.

DISCUSSION

This is the first study that explored the discriminative value of exhaled VOCs in preschool children with respiratory symptoms. We demonstrated that VOCs can be easily and safely collected and determined in the exhaled breath of preschool children. Moreover, a profile of exhaled VOCs was fairly able to differentiate children with recurrent wheeze from children without wheeze with an acceptable cross-validated classification error and sensitivity, though with a limited specificity.

Preschool wheeze is a common, though complex, symptom. The underlying causes for preschool wheezing are heterogeneous and are not sufficiently understood [1]. There is increasing evidence that airway inflammation plays a significant role in preschool wheeze. However, data are scarce since the current techniques to assess airway inflammation (*e.g.* biopsy) are too invasive for routine use in children. The need for noninvasive assessment of airway inflammation has led to increasing interest in breath analysis including exhaled VOCs. In this study, we demonstrated that measuring VOCs in exhaled breath is feasible in preschool children. Our high success rate was the consequence of three main factors. First, we achieved a high success rate for the collection of exhaled breath by using tools to make the children feel at ease during the measurements. Secondly, we used the GC-TOF-MS technique for VOC analysis. This is a highly sensitive technique capable of detecting a wide range of VOCs. Thirdly, we used the raw mass spectra to match similar compounds in all children. In most studies, compounds

TABLE 1 Clinical characteristics of the study population

	With recurrent wheeze	No recurrent wheeze
Subjects n	202	50
Age yrs mean (range)	3.2 (1.9–4.5)	3.3 (2.2–4.1)
Males/females n/n	112/90	24/26
Atopy [#]	47/191 (25)	11/45 (24)
Eczema*	77/197 (39)	11/50 (22)
Airway resistance kPa·s·L ⁻¹ median (IQR)	1.4 (1.3–1.7)	1.4 (1.2–1.7)
Exhaled nitric oxide fraction ppb median (IQR)	7.1 (3.9–13.3)	8.2 (3.4–13.0)
Use of ICS*	38/202 (19)	0/50 (0)
Use of short-acting β_2 -agonists*	77/202 (38)	0/50 (0)
Parental asthma*	70/201 (35)	8/50 (16)
Parental smoking	75/202 (37)	13/50 (26)
Season of measurement* winter/spring/summer/autumn %	25/24/32/19	24/40/34/2

Data are presented as n/N (%), unless otherwise stated. IQR: interquartile range; ICS: inhaled corticosteroids. [#]: defined as a positive Phadiatop Infant test (Phadia, Uppsala, Sweden); numbers may not sum to 202 or 50 in wheezers and nonwheezers, respectively, due to missing values. *: $p < 0.05$.

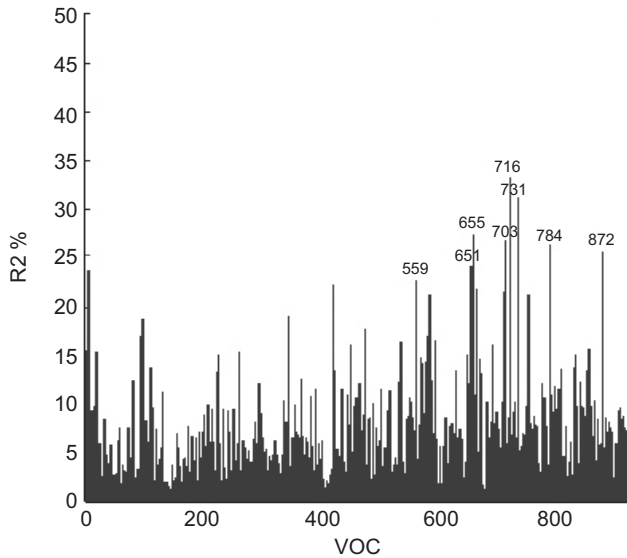


FIGURE 1. Explained variance of season for each volatile organic compound (VOC). The percentages of explained variance (R²) for season for each VOC are presented. Labels are displayed for VOCs with the 1% highest R² scores. Based on visual inspection, the explained variance of season was <23% for 99% of the VOCs.

are first identified in a library and, thereafter, compared between subjects. However, this approach is vulnerable to mistakes and improper identification of compounds can hamper the quality of databases. Our innovative procedure that matches compounds on the basis of the raw mass spectra and retention time creates a more reliable database [8].

We demonstrated that a profile of VOCs was fairly able to distinct wheezing children from healthy controls. We achieved an acceptable prediction model after deducting the data variability explained by sex, age and season. Despite the fact that season was slightly related to wheeze and, therefore, some

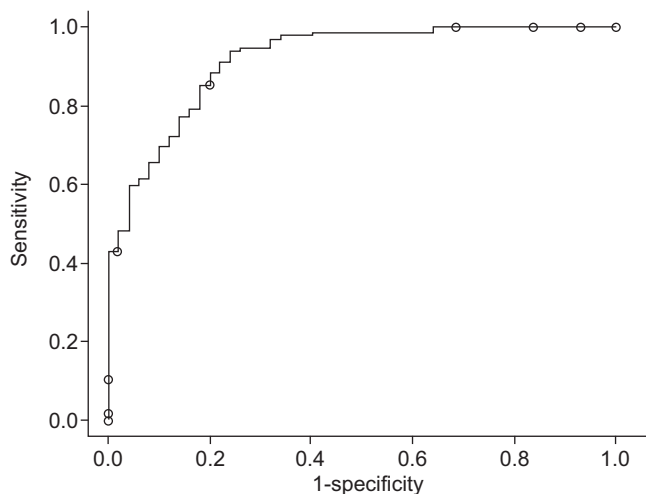


FIGURE 2. Receiver operating characteristic curve of the most discriminative compounds between wheezers and nonwheezers before internal cross-validation. The area under the curve is 0.92. After internal cross-validation, the sensitivity was 79% and the specificity was 50%.

TABLE 2 Most discriminative volatile organic compounds (VOCs) for recurrent wheeze

VOC	Recurrent wheeze	No wheeze	Total
2-methyl-1-butene	2	28	7
6-methyl-2-heptanone	6	16	8
2,4,6-trimethyl-decane	4	32	9
2,4-octadiene	5	28	9
1,4-dichloro-benzene	6	22	10
1,2-propanediol	9	12	10
Unknown	6	24	10
2,4-dimethyl-1-heptene [#]	8	26	12
2,4,4-trimethyl-1-pentene	9	30	13
Unknown	10	34	15
Acetophenone	17	44	22
2-propen-1-ol	27	30	27
1,2,3-trimethylbenzene	18	6	15
Pentanoic acid	17	8	15
Unknown	15	4	13
Unknown	18	6	16
Nitrocyclohexane	24	14	22
4-methyl-1-decene	15	6	14
3-methyl-1-butene	13	4	11
Hexadecan-2-ol	19	2	15
Unknown	18	6	15
Naphthalene	25	16	23
Unknown	8	2	7
2-ethyl-1,3-butadiene	9	0	7
2,4-dimethyl-1-heptene [#]	7	4	7
Unknown	10	2	8
1-methylethynylbenzene	18	0	15
4-methyl-2-pentenal	9	0	7

Data are presented as percentage of subjects (wheezers, nonwheezers and the total group) in which the compound was detected; [#]: difference in retention time ~3 s, probably isomeric compounds.

relevant variance explained by wheeze might have been subtracted, the predictive value of our model was improved by performing ASCA pre-processing. The fact that season of measurement, sex and age are major sources of variability in our data stresses the importance of further exploring these, and possible other, confounding factors in future studies. The most important discriminative compounds in our model for recurrent wheeze mainly included hydrocarbons. Around 40% of these hydrocarbons were less prevalent in the exhaled breath of children with recurrent wheeze compared with children without wheezing symptoms. This is comparable to our previous findings in COPD patients and children with asthma [5, 8]. The complicated biological equilibrium of formation and removal of VOCs in the human body might be a possible explanation for this finding. As VOCs are believed to be formed during processes of inflammation and oxidative stress, the relative composition of VOCs in exhaled breath can change due to a disease, and this change can be either an increase or a decrease of certain compounds [5]. For example, it is conceivable that children with wheezing symptoms have enhanced oxidative

stress. This might lead to enhanced lipid peroxidation in which certain VOCs, especially long-chain hydrocarbons, are further oxidised into smaller components and, consequently, are exhaled in lower concentrations.

One of the strengths of our study is that we have studied the entire range of exhaled VOCs instead of pre-selected inflammatory markers. This enables us to explore all potential compounds of interest. When analysing the entire range of exhaled VOCs, hundreds of components can be identified, leaving a challenging task regarding the statistical analysis to interpret and understand the data. We analysed our data using SLR. Conventional logistic regression can easily be applied on small-scale and well-designed problems but will generally fail for megavariable, collinear data. SLR can be considered as a biomarker selection method that suppresses noisy correlated VOCs (GC–MS peaks; fig. S1). This results in more sparse models that provide protection against high false positive rates. Despite this advanced statistical technique, the results are still easy to interpret. Other modelling techniques that were previously applied to VOC data included discriminant analysis (either with or without principal component analysis as a pre-processing tool) [5–7, 18], support vector machine analysis [8, 19] and fuzzy logic models [20]. Currently, there is no consensus about the optimal statistical technique to analyse VOCs. Comparing the performance of different methods should be assessed in future studies.

An additional strength of our study is that we have used GC–TOF-MS for the detection of VOCs. This sensitive technique is able to detect single compounds both quantitatively and qualitatively, and therefore provides detailed and complete information on the different exhaled VOCs. Another potential technique to detect VOCs is the electronic nose [7, 21]. Although this technique is rapid, easy to use and not very expensive, it is unable to identify single compounds in exhaled breath.

Some limitations of our study can be noted. First, we classified children based on parental-reported wheeze in order to simulate clinical practice. We did not confirm wheeze with a doctor. Therefore, it is possible that some children were misclassified and, thereby, have weakened the discriminative value of our model. It can be expected that a more stringent classification using doctor-confirmed wheeze will result in less misclassification and, subsequently, to a better discriminative model for preschool wheeze. Moreover, the group of wheezing children that we have studied is still heterogeneous. While some of the children wheeze due to recurrent viral respiratory tract infections, in other children, asthma is the underlying cause of their symptoms. Differentiating these wheezing phenotypes will be possible after our 4-yr follow-up.

Despite the potential use of VOCs in paediatric lung diseases, multiple constraints need to be resolved before the analysis of VOCs can be applied in clinical practice. First, future efforts should be directed towards standardisation of the collection and analysis of data. The development of international recommendations for standardised procedures, as developed for exhaled nitric oxide and markers in exhaled breath condensate, would be useful in order to enhance the interlaboratory comparability [22, 23]. Secondly, validation is important. Although most studies include cross-validation, the next step to be taken is the

use of an external validation set. Thirdly, more insight is needed into the physiological meaning and biochemical origin of VOCs. As described before, VOCs can be formed during lipid peroxidation as the result of airway inflammation. Since VOCs are blood borne, it is conceivable that they are also formed during other processes and have a more systemic origin as well. Therefore, in future, clinical studies should be combined with animal and *in vitro* studies to assess a more profound biological interpretation of discriminative VOC models. Finally, the refinement of sampling techniques would enhance the potential use of VOCs in the diagnosis and monitoring of pulmonary diseases in the future. Once these steps are taken and the discriminative VOC profiles have been validated in large-scale clinical studies, technologies such as dedicated GC devices or other sensor technologies will speed up analysis and become cost-effective as a diagnostic tool.

In conclusion, a profile of VOCs in exhaled breath was fairly able to discriminate between preschool children with and without recurrent wheeze. Although our model showed an acceptable classification error and sensitivity, it had a limited specificity. However, in this proof-of-principle study, we demonstrated the potential of VOC analysis as a technique that can provide valuable information about disease status in wheezing children. Hopefully, this will pave the way for additional research on VOCs in the field of preschool wheezing. Potential future applications of VOCs include objectively defining preschool wheezing phenotypes (based on underlying pathophysiology), assessing an early asthma diagnosis and evaluating treatment.

SUPPORT STATEMENT

This study was supported by the Dutch Asthma Foundation, Stichting Astma Bestrijding and Maastricht University Medical Centre.

CLINICAL TRIAL

The ADEM study is registered at www.clinicaltrials.gov with identifier number NCT00422747.

STATEMENT OF INTEREST

None declared.

ACKNOWLEDGEMENTS

The authors wish to thank their research nurse K. Groot, and medical students A. Nijhuis, B. Thönissen, E. Kalicharan and E. Cohen for their outstanding assistance during the measurements; E. Moonen for his skilled assistance with the laboratory procedures (all Maastricht University Medical Center, Maastricht, the Netherlands); the Dutch Asthma Foundation, Stichting Astma Bestrijding and Maastricht University Medical Centre for their financial support; and the parents and children for their participation.

REFERENCES

- 1 Henderson J, Granell R, Sterne J. The search for new asthma phenotypes. *Arch Dis Child* 2009; 94: 333–336.
- 2 Vinayavekhin N, Homan EA, Saghatelian A. Exploring disease through metabolomics. *ACS Chem Biol* 2010; 5: 91–103.
- 3 Buszewski B, Keszy M, Ligor T, *et al.* Human exhaled air analytics: biomarkers of diseases. *Biomed Chromatogr* 2007; 21: 553–566.
- 4 Miekisch W, Schubert JK, Noeldge-Schomburg GF. Diagnostic potential of breath analysis—focus on volatile organic compounds. *Clin Chim Acta* 2004; 347: 25–39.

- 5 Dallinga JW, Robroeks CM, van Berkel JJ, *et al.* Volatile organic compounds in exhaled breath as a diagnostic tool for asthma in children. *Clin Exp Allergy* 2010; 40: 68–76.
- 6 Robroeks CM, van Berkel JJ, Dallinga JW, *et al.* Metabolomics of volatile organic compounds in cystic fibrosis patients and controls. *Pediatr Res* 2010; 68: 75–80.
- 7 Fens N, Zwinderman AH, van der Schee MP, *et al.* Exhaled breath profiling enables discrimination of chronic obstructive pulmonary disease and asthma. *Am J Respir Crit Care Med* 2009; 180: 1076–1082.
- 8 van Berkel JJ, Dallinga JW, Moller GM, *et al.* A profile of volatile organic compounds in breath discriminates COPD patients from controls. *Respir Med* 2010; 104: 557–563.
- 9 Oommen A, Patel R, Browning M, *et al.* Systemic neutrophil activation in acute preschool viral wheeze. *Arch Dis Child* 2003; 88: 529–531.
- 10 Marguet C, Jouen-Boedes F, Dean TP, *et al.* Bronchoalveolar cell profiles in children with asthma, infantile wheeze, chronic cough, or cystic fibrosis. *Am J Respir Crit Care Med* 1999; 159: 1533–1540.
- 11 Saglani S, Payne DN, Zhu J, *et al.* Early detection of airway wall remodeling and eosinophilic inflammation in preschool wheezers. *Am J Respir Crit Care Med* 2007; 176: 858–864.
- 12 The International Study of Asthma and Allergies in Childhood (ISAAC) Steering Committee. Worldwide variations in the prevalence of asthma symptoms: the International Study of Asthma and Allergies in Childhood (ISAAC). *Eur Respir J* 1998; 12: 315–335.
- 13 van de Kant KD, Klaassen EM, Jöbsis Q, *et al.* Early diagnosis of asthma in young children by using non-invasive biomarkers of airway inflammation and early lung function measurements: study protocol of a case-control study. *BMC Public Health* 2009; 9: 210.
- 14 van Berkel JJ, Dallinga JW, Moller GM, *et al.* Development of accurate classification method based on the analysis of volatile organic compounds from human exhaled air. *J Chromatogr B Analyt Technol Biomed Life Sci* 2008; 861: 101–107.
- 15 Friedman JH, Hastie T, Tibshirani R. Regularisation paths for generalised linear models *via* coordinate descent. *J Stat Soft* 2010; 33: 1–22.
- 16 Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning. Data Mining, Inference and Prediction.* Philadelphia, Springer, 2009.
- 17 Penn DJ, Oberzaucher E, Grammer K, *et al.* Individual and gender fingerprints in human body odour. *J R Soc Interface* 2007; 4: 331–340.
- 18 Dragonieri S, Schot R, Mertens BJ, *et al.* An electronic nose in the discrimination of patients with asthma and controls. *J Allergy Clin Immunol* 2007; 120: 856–862.
- 19 Machado RF, Laskowski D, Deffenderfer O, *et al.* ; Detection of lung cancer by sensor array analyses of exhaled breath. *Am J Respir Crit Care Med* 2005; 171: 1286–1291.
- 20 Phillips M, Altorki N, Austin JH, *et al.* Prediction of lung cancer using volatile biomarkers in breath. *Cancer Biomark* 2007; 3: 95–109.
- 21 Dragonieri S, Annema JT, Schot R, *et al.* An electronic nose in the discrimination of patients with non-small cell lung cancer and COPD. *Lung Cancer* 2009; 64: 166–170.
- 22 ATS/ERS recommendations for standardized procedures for the online and offline measurement of exhaled lower respiratory nitric oxide and nasal nitric oxide, 2005. *Am J Respir Crit Care Med* 2005; 171: 912–930.
- 23 Horvath I, Hunt J, Barnes PJ, *et al.* Exhaled breath condensate: methodological recommendations and unresolved questions. *Eur Respir J* 2005; 26: 523–548.