

ORIGINAL ARTICLE

Airways Disease

WILEY

Wheezing preschool children with early-onset asthma reveal a specific metabolomic profile

Silvia Carraro¹  | Sara Bozzetto¹ | Giuseppe Giordano^{1,2} | Dania El Mazloum¹ | Matteo Stocchero^{1,2} | Paola Pirillo^{1,2} | Stefania Zanconato¹ | Eugenio Baraldi¹

¹Department of Women's and Children's Health, University of Padova, Padova, Italy

²Città della Speranza Institute of Pediatric Research (IRP), Padova, Italy

Correspondence

Silvia Carraro, Department of Women's and Children's Health, Unit of Pediatric Allergy and Respiratory Medicine, University of Padova, Padova, Italy.
Email: silvia.carraro@unipd.it

Funding information

Fondazione Cassa di Risparmio di Padova e Rovigo, Italy

Abstract

Background: Many children of preschool age present with recurrent wheezing. Most of them outgrow their symptoms, while some have early-onset asthma. Aim of this prospective preliminary study was to apply a metabolomic approach to see whether biochemical-metabolic urinary profiles can have a role in the early identification of the children with asthma.

Methods: Preschool children with recurrent wheezing were recruited and followed up for 3 years, after which they were classified as cases of transient wheezing or early-onset asthma. A urine sample was collected at recruitment and analyzed using a metabolomic approach based on UPLC mass spectrometry.

Results: Among 34 children aged 4.0 ± 1.1 years recruited, at the end of the 3-year follow-up, 16 were classified as having transient wheezing and 16 as cases of early-onset asthma. Through a joint multivariate and univariate statistical analyses, we identified a subset of metabolomic variables that enabled the 2 groups to be clearly distinguished. The model built using the identified variables showed an AUC = 0.99 and an AUC = 0.88 on sevenfold full cross-validation ($P = .002$).

Conclusions: Metabolomic urinary profile can discriminate preschoolers with recurrent wheezing who will outgrow their symptoms from those who have early-onset asthma. These results may pave the way to the characterization of early non-invasive biomarkers capable of predicting asthma development.

KEYWORDS

asthma, metabolomics, pediatrics, recurrent wheezing, urine

1 | INTRODUCTION

Recurrent wheezing is a heterogeneous condition and a very common problem in children of preschool age. A minority of these children will continue to experience wheezing episodes as they grow older, and will eventually be classified as asthmatic. The remainder will outgrow their symptoms.¹

Unfortunately, at preschool age, these 2 groups of children are indistinguishable from a clinical standpoint. The early identification

of children who will subsequently develop asthma is therefore a key field of research, given the potential for targeted preventive and therapeutic measures for this condition.

A number of methods have been proposed for guiding the early identification of asthmatic children, such as the Asthma Predictive Index (API),² the Isle of Wight index,³ and the Prevention and Incidence of Asthma and Mite Allergy (PIAMA) index.⁴ Their clinical utility has been questioned, however, and it has been suggested that finding new biomarkers related to the pathophysiologic mechanisms

underlying asthma might significantly improve our ability to identify children bound to become asthmatic early on.⁵

Metabolomic analysis is guided by no a priori hypothesis. The overall metabolic picture of a sample is analyzed; then, appropriate statistical methods are applied to identify the metabolic profile associated with a given condition.⁶

The aim of the present study was to apply the metabolomic approach to analyze urine samples collected in preschoolers with recurrent wheezing to see whether their metabolomic profile can predict their clinical outcome, as ascertained after a 3-year follow-up.

2 | METHODS

2.1 | Study subjects and study design

This was a preliminary prospective case-control study conducted at the Women's and Children's Health Department of the University of Padova. Preschool age children with recurrent wheezing, referred to our outpatient clinic, were recruited between October 2012 and April 2013.

Inclusion criteria included children aged 2 to 5 years who had experienced at least 3 episodes of wheezing diagnosed by a physician during the previous 12 months, and who were fully potty-trained. Exclusion criteria included children born preterm, or with a history of respiratory distress at birth, or suffering from chronic respiratory or other diseases.

A group of healthy children with no history of allergic or respiratory diseases was recruited as a control.

All the recruited children had been free of any acute respiratory symptoms for at least 1 month before they were enrolled for the study.

At recruitment, a medical history was obtained and a physical examination was conducted on all the children. All the visits were in the morning. After the visit, they were asked to collect the first part of the urine stream (at least 6 mL), and their samples were immediately stored in 2 aliquots at -80°C until analysis. In children with recurrent wheezing, a skin prick test was performed for the most common aeroallergens (Lofarma, SpA, Milan, Italy) and the children's modified Asthma Predictive Index (mAPI) was calculated.⁷

The children with recurrent wheezing were then followed up regularly at our outpatient clinic for 3 years, being seen every 3 months. Their assessment at the end of the follow-up included skin prick test and spirometry with reversibility test.⁸

At the end of the follow-up, the children with recurrent wheezing were classified in the "early-onset asthma" group if they had experienced recurrent asthma-like symptoms during the previous 12 months, and if they had been prescribed daily ICS for at least 6 months in accordance with the GINA guidelines.⁹

The study was approved by the Ethics Committee of the Azienda Ospedaliera di Padova (Prot. 1503P), and parents provided their informed consent.

2.2 | Chromatographic and mass spectrometry analyses

The urine samples were analyzed at the Mass Spectrometry Laboratory of the Institute for Pediatric Research (IRP), at the Women's and Children's Health Department of Padova University Hospital.

The metabolic analysis on the urine samples was conducted with a high-resolution mass spectrometer (Q-ToF Synapt G2; Waters) interfaced with an ultra-performance liquid chromatography (UPLC) system (Acquity; Waters). A full description of the analytical procedure is provided in the Supporting information.

2.3 | Statistical data analysis

A multivariate data analysis based on projection methods was used to model the data collected. Specifically, an exploratory data analysis was performed using principal component analysis (PCA), and projection to latent structures discriminant analysis based on VIP (variable influence on projection) selection (VIP-based PLS-DA)¹⁰ was applied to identify the differences between the groups under investigation. Stability selection based on Monte-Carlo sampling¹¹ was used to highlight the subset of relevant variables characterizing the differences between the groups and to estimate the predictive power of the models.

The main available metabolomic databases (Human Metabolome DataBase [HMDB] and METLIN) were searched to identify the relevant variables with a plausible biologic significance characterizing each group.

3 | RESULTS

Thirty-four children with recurrent wheezing and 13 healthy controls were recruited. Two children were lost to follow-up, so 32 children with recurrent wheezing (aged 4.0 ± 1.1 years) and 13 healthy controls (aged 3.6 ± 1.3 years) were included in the final analysis. The 2 groups were matched for age (P -value = .29) and sex (P -value = .09).

At the end of the 3-year follow-up, 16 children (age at recruitment 4.3 ± 1.1 years) with recurrent wheezing were classified in the "early-onset asthma" group, while the other 16 (age at recruitment 3.75 ± 0.94 years) were classified in the "transient wheezing" group. At the time of their recruitment, the 2 groups did not differ significantly in terms of age (t test; P -value = .13), sex (Fisher's exact test; P -value = 1.00), or medication taken (Fisher's exact test P = .69 and P = 1.00 for inhaled steroids and montelukast, respectively). The distribution of the relevant clinical variables at recruitment in the 2 groups is reported in Table 1. Among them, only the allergic sensitization resulted significantly more frequent in the early-onset asthma group (P = .018).

None of the control children developed wheezing symptoms during the 3-year follow-up.

TABLE 1 Clinical variable distribution in the 2 groups

	Early-onset asthma (n = 16)	Transient wheezing (n = 16)	Fisher's exact test P-value
Males	69% (11/16)	69% (11/16)	1.00
Presence of allergic sensitization ^a	69% (11/16)	25% (4/16)	.018
Presence of atopic dermatitis	25% (4/16)	13% (2/16)	.65
Presence of food allergy	19% (3/16)	6% (1/16)	.60
Presence of allergic rhinitis	19% (3/16)	0% (0/16)	.23
Positive mAPI	69% (11/16)	44% (7/16)	.18
Inhaled steroids	31% (5/16)	19% (3/16)	.69
Montelukast	19% (3/16)	13% (2/16)	1.00
Spirometric values at the end of follow-up^b			t test P-value
FVC	94 ± 11%pred	90 ± 8%pred	.25
FEV ₁	91 ± 12%pred	94 ± 9%pred	.43
FEV ₁ /FVC	89 ± 10%	92 ± 7%	.33
FEF ₂₅₋₇₅	92 ± 32%pred	98 ± 18%pred	.52

^aSensitization for the following aeroallergens was checked: dust mite, grass pollen, tree pollen, alternaria, dog, and cat dander.

^bValues are expressed as mean ± SD.

Lung function assessment at the end of the 3-year follow-up revealed no significant differences between the 2 groups (Table 1). The reversibility test result was positive in 6 of 16 children with early-onset asthma and in none of the children with transient wheezing. One child with early-onset asthma became sensitized to aeroallergens during the follow-up.

3.1 | Metabolomics: recurrent wheezing vs controls

The MS analysis of the urine samples led to the generation of 2 data sets. The T3 NEG data set was obtained in negative ionization mode and included 1484 RT_{mass} variables, the T3 POS data set in positive ionization mode, and comprised 1927 RT_{mass} variables. A preliminary exploratory data analysis was performed by PCA on each data set to identify outliers and specific patterns in the data collected. No outliers were detected on the basis of the DModX test, or Hotelling's T2 test (at 95% level) performed on the PCA models of each group.

As a first step in our data analysis, the control group was compared with the children with recurrent wheezing. We obtained reliable PLS-DA models capable of distinguishing between recurrent wheezers and controls. For the T3 NEG data set, we obtained a model showing A = 2 components, AUC = 1.00, AUC during sevenfold full cross-validation = 0.95 (P-value < .001), and median of the AUC_{pred} calculated by stability selection = 0.71 (with the 5th percentile >0.50). For the T3 POS data set, we obtained A = 2 components, AUC = 1.00, AUC during sevenfold full cross-validation = 0.93 (P-value < .001), and median of the AUC_{pred} = 0.78.

The score scatter plots¹² of the 2 models are shown in Figure 1 (panels A and B).

3.2 | Metabolomics: early-onset asthma vs transient wheezing

The differences between the metabolomic profiles of the early-onset asthma and transient wheezing groups were investigated by PLS-DA. A stability selection procedure was applied, based on Monte-Carlo sampling and VIP-based PLS-DA. Two hundred random subsamples were extracted by Monte-Carlo sampling from the samples collected, and VIP-based PLS-DA was applied to each subsample. All 200 VIP-based PLS-DA models passed the permutation test on the class response ($P < .05$), and the median of the AUC_{pred} was 0.65 for the T3 NEG data set, and 0.65 for the T3 POS data set (with the 5th percentile >0.50 in both cases). Two subsets of 24 and 40 relevant variables were selected for the T3 NEG and T3 POS data sets, respectively.

Considering only the variables with P-value below .05 in the t test, q-value below 0.15 for the false discovery rate, and a lower border of the 95% confidence interval for the AUC higher than 0.50, univariate data analysis led us to select 41 variables for the T3 NEG data set and 73 for the T3 POS data set.

The results of the multivariate analysis were combined with the results of the univariate analysis to obtain the set of relevant variables to be investigated (51 for the T3 NEG data set, and 80 for the T3 POS data set, Tables S1 and S2 available online). On searching available online metabolite databases, we found a plausible biologic meaning for a subset of these variables (putative biomarkers), as listed in Tables 2 and 3.

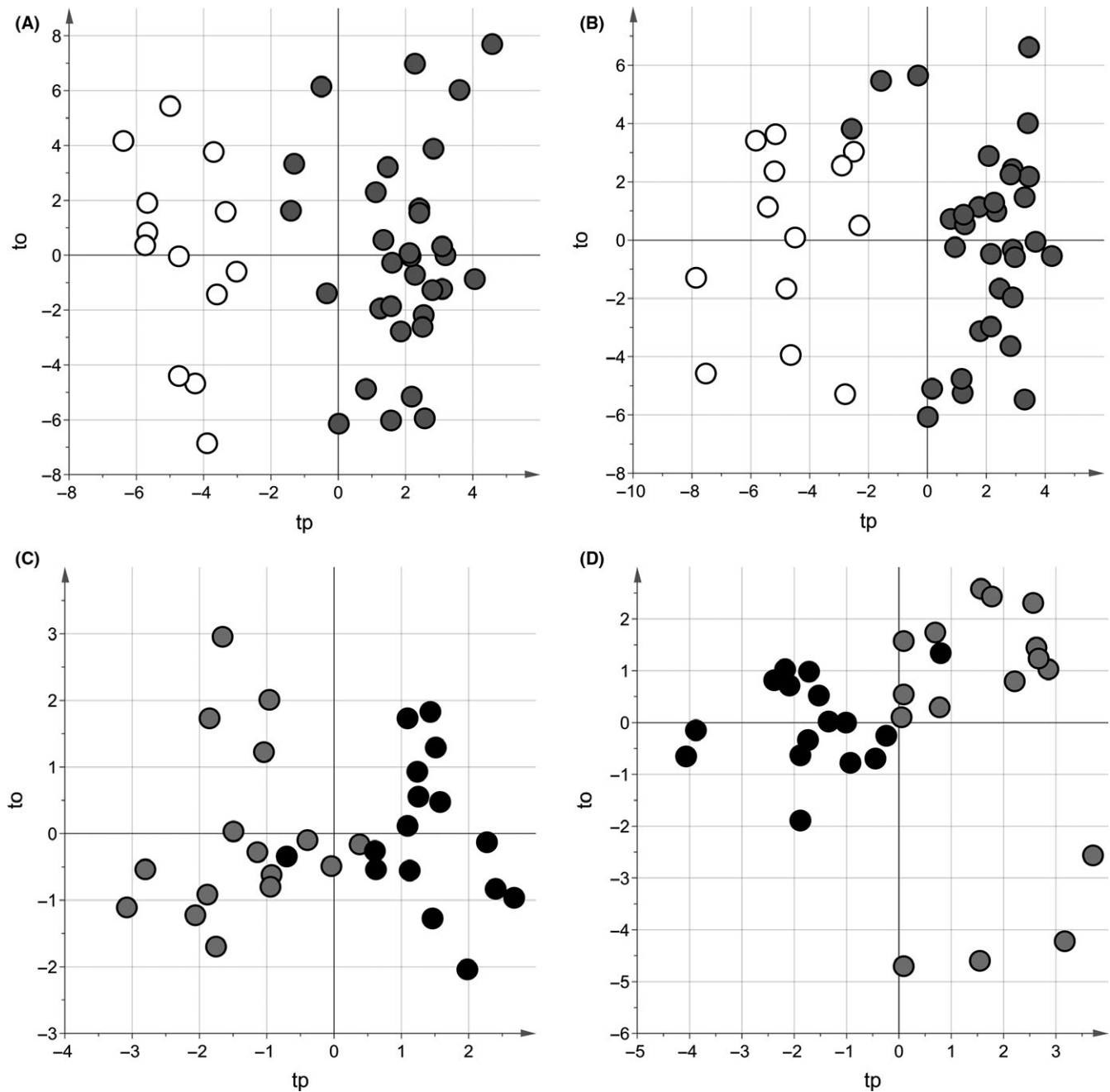


FIGURE 1 Score scatter plots after post-transformation of the PLS-DA models obtained comparing the recurrent wheezing group (gray circle) and the healthy control group (empty circle) for the T3 NEG data set (A) and the T3 POS data set (B). Score scatter plots after post-transformation of the PLS-DA models comparing the “transient wheezing” group (light gray circle) and the “early-onset asthma” group (black circle) for the T3 NEG data set (C) and the T3 POS data set (D). tp and to are the predictive and orthogonal components of the PLS-DA models obtained for each comparison

The relevant variables were used to build PLS-DA models capable of distinguishing between the “transient wheezing” and the “early-onset asthma” groups. We thus obtained reliable models showing $A = 2$ components, $AUC = 0.99$ and AUC during sevenfold full cross-validation = 0.88 (P -value = $.002$) for the T3 NEG data set, and $A = 2$ components, $AUC = 0.99$ and AUC during sevenfold full cross-validation = 0.93 (P -value < $.001$) for the T3 POS data set. Figure 1 (panels C and D) shows the score scatter plots of the 2 models.

Using only the relevant variables with a plausible biologic significance (putative biomarkers), we obtained a PLS-DA model with $A = 1$ component, $AUC = 0.98$ and AUC during sevenfold full cross-validation = 0.94 (P -value = $.002$).

3.3 | Metabolomics and allergic sensitization

Metabolomic analysis can be used to discriminate between children with and without allergies. The PLS-DA models obtained $A = 1$

TABLE 2 Putative biomarkers with higher levels in the “transient wheezing” than in the “early-onset asthma” group

RT	m/z	Ionization mode	AUC ROC (95%)	t test P-value	Putative biomarker
1.046	159.0297	NEG	0.591-0.947	4.2E-03	Oxoadipic acid
4.597	182.0812	NEG	0.612-0.942	6.0E-03	(-)-epinephrine
4.109	180.0653	NEG	0.540-0.912	1.6E-02	L-tyrosine
2.165	194.0449	NEG	0.563-0.913	1.7E-02	3-hydroxyhippuric acid
3.168	121.0286	NEG	0.504-0.878	4.6E-02	Benzoic acid
4.916	217.1070	NEG	0.532-0.897	5.0E-02	3-hydroxy-sebacic acid
4.101	275.0225	NEG	0.489-0.877	5.3E-02	Dihydroferulic acid 4-sulfate
4.590	109.0654	POS	0.662-0.970	9.5E-04	p-cresol
4.585	206.0798	POS	0.576-0.939	1.8E-02	Indolelactic acid
4.760	208.0953	POS	0.535-0.910	2.7E-02	N-acetyl-L-phenylalanine
2.086	175.1081	POS	0.537-0.900	3.9E-02	N2-acetyl-ornithine

RT = retention time; m/z = mass-to-charge ratio; AUC ROC (95%) = confidence interval of the area under the ROC curve at 95% level.

TABLE 3 Putative biomarkers with higher levels in the “early-onset asthma” than in the “transient wheezing” group

RT	m/z	Ionization mode	AUC ROC (95%)	t test P-value	Putative biomarker
0.711	351.0559	NEG	0.575-0.924	1.2E-02	4-(4-deoxy- α -D-gluc-4-enuronosyl)-D-galacturonate
4.362	131.0337	NEG	0.607-0.946	2.9E-02	Glutaric acid
5.433	155.1069	NEG	0.529-0.907	3.4E-02	4-hydroxynonenal
0.604	245.0424	NEG	0.495-0.871	4.9E-02	Phosphatidyl glycerol
1.374	257.0773	NEG	0.482-0.869	5.5E-02	3-methyluridine
4.730	335.1334	NEG	0.485-0.873	7.3E-02	Steroid O-sulfate
2.387	221.0926	POS	0.551-0.917	8.0E-03	5-hydroxy-L-tryptophan
4.339	176.0711	POS	0.603-0.950	1.4E-02	3-indoleacetic acid
2.558	158.0819	POS	0.523-0.898	1.5E-02	Tiglylglycine
4.058	118.0658	POS	0.582-0.925	1.6E-02	Indole
1.222	112.0512	POS	0.520-0.893	2.1E-02	Cytosine
0.692	131.1189	POS	0.545-0.907	2.4E-02	N-acetylputrescine
2.391	175.0873	POS	0.511-0.887	2.4E-02	Indole-3-acetamide
3.147	150.0780	POS	0.553-0.907	2.8E-02	6-methyladenine
0.835	126.0667	POS	0.547-0.906	3.4E-02	5-methylcytosine
4.408	130.0507	POS	0.509-0.880	4.0E-02	N-acryloylglycine
3.940	183.0659	POS	0.539-0.898	5.8E-02	Hydroxyphenyllactic acid

RT = retention time; m/z = mass-to-charge ratio; AUC ROC (95%) = confidence interval of the area under the ROC curve at 95% level.

component, AUC during sevenfold full cross-validation = 0.97 (P -value < .01) for the T3 NEG data set, and A = 1 component, AUC during sevenfold full cross-validation = 0.87 (P -value < .01) for the T3 POS data set. In other words, metabolomic analysis is capable of revealing metabolic differences both between “transient wheezing” and “early-onset asthma,” and between allergic and non-allergic children.

In addition, metabolomics outperformed allergic sensitization in distinguishing cases of “transient wheezing” from children with “early-onset asthma,” as shown in Figure 2. We ran an exploratory investigation to test whether the discriminatory power of the “allergic sensitization” variable could be improved by combining it with the results of metabolomic analysis. For example, combining the “allergic sensitization”

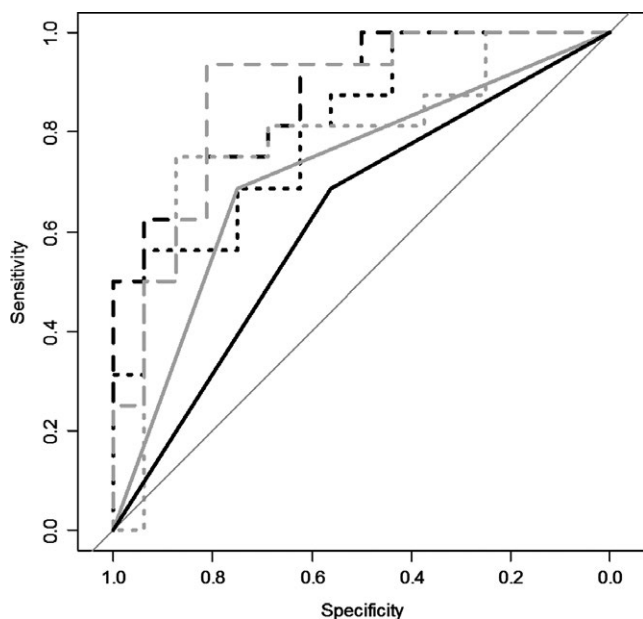


FIGURE 2 ROC curves arising from the metabolomic analysis and from the clinical variables; black dashed line: relevant variables for the T3 POS data set (AUC sevenfold full cross-validation = 0.93); black dotted line: “allergic sensitization,” hydroxyphenyllactic acid, and 5-hydroxy-L-tryptophan (T3 POS data set, AUC sevenfold full cross-validation = 0.80); gray dashed line: relevant variables for the T3 NEG data set (AUC sevenfold full cross-validation = 0.88); gray dotted line: “allergic sensitization,” 3-hydroxyhippuric acid, and dihydroferulic acid 4-sulfate (T3 NEG data set, AUC sevenfold full cross-validation = 0.79); gray line: “allergic sensitization” (AUC = 0.72); black line: mAPI (AUC = 0.63)

variable with the 3-hydroxyhippuric acid and dihydroferulic acid 4-sulfate variables derived from the T3 NEG data set, we obtained a multivariate biomarker by PLS-DA with an AUC during sevenfold full cross-validation = 0.79. On the other hand, by combining “allergic sensitization” with hydroxyphenyllactic acid and 5-hydroxy-L-tryptophan derived from the T3 POS data set, for instance, we obtained a multivariate biomarker with an AUC during sevenfold full cross-validation = 0.80.

4 | DISCUSSION

In the present study, we found that, already at preschool age, children with transient wheezing have a different urinary metabolome from children with early-onset asthma. We also identified some putative biomarkers characteristic of each of the 2 conditions.

The relationship between recurrent wheezing in preschool age and asthma later in life is close and complex. We know from epidemiologic studies that asthma-like symptoms begin at preschool age in most patients.¹³ On the other hand, recurrent wheezing is a very common problem in preschoolers, although only a minority of the children affected will subsequently be classified as asthmatic.¹⁴ The early identification of the subgroup of children with asthma is crucial inasmuch as it might enable early interventions capable of changing the natural history of this condition.

A number of cohort studies conducted in the past 20 years have investigated the natural history of wheezing in childhood. Based on the collected data, some prediction rules have been proposed for the early identification of individuals with asthma,^{2-4,15} but their clinical value might be somewhat limited.⁵ Integrating clinical data with biomarkers relating to the pathophysiologic mechanisms underlying asthma might significantly improve the predictive capability of such rules. To this end, Klassen et al¹⁶ have recently demonstrated that integrating the API with data on exhaled VOCs (markers of airway inflammation) at preschool age significantly improves the accuracy of asthma prediction at 6 years of age.

In the present study, we applied the untargeted metabolomic approach to investigate the biochemical-metabolic processes sustaining the link between recurrent wheezing at preschool age and asthma in later life.

The main outcome of our study lies in the finding of an overall metabolic profile capable of discriminating children with transient wheezing from children with early-onset asthma. This result strongly suggests that a different biochemical-metabolic arrangement underlies these 2 conditions already at preschool age.

Among clinical characteristics evaluated at recruitment (Table 1), only allergic sensitization was significantly more common in the early-onset asthma group than in the transient wheezing group—a finding in line with the literature supporting the key role of early allergic sensitization in the onset of pediatric asthma.¹⁷⁻²⁰ In this regard, we acknowledge that the metabolic differences identified between the 2 groups might relate partially to differences in atopic status. The strong relationship between pediatric asthma and allergic sensitization makes it impossible to disentangle these 2 aspects, even using such a powerful untargeted approach as metabolomics.

Searching the identified discriminating metabolites in the available databases led us to hypothesize the biologic identity of some of them. It is worth noting that, among the variables characterizing the early-onset asthma group, some are attributable to metabolites related to tryptophan (ie, glutaric acid, 5-hydroxy-L-tryptophan, indole-3-acetamide, and 3-indoleacetic acid). Increased plasma levels of tryptophan have already been reported in asthmatic patients.^{21,22} Although the role of tryptophan metabolism in asthma pathogenesis is not fully understood, the idea of this metabolic pathway having an early pathogenic role in the natural history of asthma is intriguing.

In addition, the finding of phosphatidyl glycerol and 2 metabolites related to fatty acid metabolism (N-acryloylglycine and tiglylglycine) is consistent with data suggesting a lipid compound dysregulation in asthma.²³ At this regard, recent data have suggested a possible relationship between blood levels of n-3 and n-6 PUFAs at age 8, and allergic conditions at age 16.²⁴

Finally, a few metabolites that emerged for both groups are probably products of gut microbiota metabolism, that is 3-hydroxyhippuric acid, p-cresol, indolelactic acid, and benzoic acid in the transient wheezing group, and 4-(4-deoxy-(D-gluc-4-enuronosyl)-D-galacturonate, indole, and hydroxyphenyllactic acid in the early-onset asthma group. A promising field of research nowadays concerns the role of the cross-talk between microbiota and immune

system in the development of allergic diseases.²⁵ Though preliminary, our findings suggest that a different microbial arrangement may have a role in driving children with recurrent wheezing toward a solution of their breathing problem or toward the onset of asthma. This is in line with recent evidence that point to the role of the "microbial pathways" as one of the major determinants of asthma risk.²⁶

It seems interesting that, although the metabolomic approach outperformed "allergic sensitization" in discriminating between transient wheezing and early-onset asthma, we demonstrated that such clinical variable might gain in discriminatory power if considered together with certain metabolomic variables. Such an exploratory analysis, if confirmed and quantified using targeted approaches, suggests that putative biomarkers could become part of an efficient, cheaply, and easily assessable, multidimensional (clinical and metabolic) biomarker for the early differentiation between cases of transient wheezing and early-onset asthma.

The strengths of our study include its prospective design and the use of stringent clinical criteria for identifying children with early-onset asthma.

Our results were internally validated using a robust approach, though we recognize that the lack of an external validation of the models, and the small sample size are limitations of the study. Because of these limitations, our study must be considered preliminary and further studies are needed to confirm our discriminating models in larger independent groups of children and to validate, also using targeted approaches, the identified metabolites.

In conclusion, we found that metabolomic profiling can non-invasively distinguish children with recurrent wheezing in preschool age who will outgrow their symptoms from those who have early-onset asthma. For each of these 2 conditions, some putative biomarkers were identified that, if confirmed, might have a role as early biomarkers of asthma.

ACKNOWLEDGMENT

We thank Giulia Mazzarolo (Department of Women's and Children's Health, University of Padova) for her help in sample collection. This work was partly supported by "Bando Ricerca Pediatrica 2012-2014" Fondazione Cassa di Risparmio di Padova e Rovigo, Italy. The funding source had no role in study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

ORCID

Silvia Carraro  <http://orcid.org/0000-0003-2284-1225>

REFERENCES

- Martinez FD. What have we learned from the Tucson children's respiratory study? *Paediatr Respir Rev*. 2002;3:193-197.
- Castro-Rodriguez JA, Holberg CJ, Wright AL, Martinez FD. A clinical index to define risk of asthma in young children with recurrent wheezing. *Am J Respir Crit Care Med*. 2000;162(4 Pt 1):1403-1406.
- Kurukulaaratchy RJ, Matthews S, Holgate ST, Arshad SH. Predicting persistent disease among children who wheeze during early life. *Eur Respir J*. 2003;22:767-771.
- Caudri D, Wijga A, Schipper C, et al. Predicting the long-term prognosis of children with symptoms suggestive of asthma at preschool age. *J Allergy Clin Immunol*. 2009;124:903-910.e1-e7.
- Savenije OE, Kerkhof M, Koppelman GH, Postma DS. Predicting who will have asthma at school age among preschool children. *J Allergy Clin Immunol*. 2012;130:325-331.
- Carraro S, Giordano G, Reniero F, Perilongo G, Baraldi E. Metabolomics: a new frontier for research in pediatrics. *J Pediatr*. 2009;154:638-644.
- Guilbert TW, Morgan WJ, Krawiec M, et al. Prevention of early asthma in kids study, Childhood Asthma Research and Education network. The prevention of early asthma in kids study: design, rationale and methods for the Childhood Asthma Research and Education network. *Control Clin Trials*. 2004;25:286-310.
- Miller MR, Hankinson J, Brusasco V, et al. ATS/ERS task force. Standardisation of spirometry. *Eur Respir J*. 2005;26:319-338.
- Global Initiative for Asthma (GINA). Pocket guide for asthma management and prevention in children 5 years and younger, updated April 2015. <http://ginasthma.org/pocket-guide-for-asthma-management-and-prevention-in-children-5-years-and-younger/>. Accessed on November 2015.
- Chong I-G, Jun C-H. Performance of some variable selection methods when multicollinearity is present. *Chemomet Intel Lab Syst*. 2005;78:103-112.
- Arredouani A, Stocchero M, Culeddu N, et al. Metabolomic profile of low copy-number carriers at the salivary alpha-amylase gene suggests a metabolic shift towards lipid-based energy production. *Diabetes*. 2016;65:3362-3368.
- Stocchero M, Paris D. Post-transformation of PLS2 [ptPLS2] by orthogonal matrix: a new approach for generating predictive and orthogonal latent variables. *J Chemometrics*. 2016;30:241-252.
- Yunginger JW, Reed CE, O'Connell EJ, Melton LJ 3rd, O'Fallon WM, Silverstein MD. A community-based study of the epidemiology of asthma. Incidence rates, 1964-1983. *Am Rev Respir Dis*. 1992;146:888-894.
- Martinez FD, Wright AL, Taussig LM, Holberg CJ, Halonen M, Morgan WJ. Asthma and wheezing in the first six years of life. *N Engl J Med*. 1995;332:133-138.
- Chang TS, Lemanske RF Jr, Guilbert TW, Martinez FD. Evaluation of the modified asthma predictive index in high-risk preschool children. *J Allergy Clin Immunol Pract*. 2013;1:152-156.
- Klaassen EM, van de Kant KD, Jöbsis Q, et al. Exhaled biomarkers and gene expression at preschool age improve asthma prediction at 6 years of age. *Am J Respir Crit Care Med*. 2015;191:201-207.
- Simpson A, Tan VY, Winn J, et al. Beyond atopy: multiple patterns of sensitization in relation to asthma in a birth cohort study. *Am J Respir Crit Care Med*. 2010;181:1200-1206.
- Martinez FD. The origins of asthma and chronic obstructive pulmonary disease in early life. *Proc Am Thorac Soc*. 2009;6:272-277.
- Kusel MM, de Klerk NH, Keadze T, et al. Early-life respiratory viral infections, atopic sensitization, and risk of subsequent development of persistent asthma. *J Allergy Clin Immunol*. 2007;119:1105-1110.
- Illi S, von Mutius E, Lau S, et al. Perennial allergen sensitisation early in life and chronic asthma in children: a birth cohort study. *Lancet*. 2006;368:763-770.
- van der Sluijs KF, van de Pol MA, Kulik W, et al. Systemic tryptophan and kynurenine catabolite levels relate to severity of rhinovirus-induced asthma exacerbation: a prospective study with a parallel-group design. *Thorax*. 2013;68:1122-1130.
- Collipp PJ, Chen SY, Sharma RK, Balachandar V, Maddaiah VT. Tryptophan metabolism in bronchial asthma. *Ann Allergy*. 1975;35:153-158.

23. Kang YP, Lee WJ, Hong JY, et al. Novel approach for analysis of bronchoalveolar lavage fluid (BALF) using HPLC-QTOF-MS-based lipidomics: lipid levels in asthmatics and corticosteroid-treated asthmatic patients. *J Proteome Res*. 2014;13:3919-3929.
24. Magnusson J, Ekström S, Kull I, et al. Polyunsaturated fatty acids in plasma at 8 years and subsequent allergic disease. *J Allergy Clin Immunol*. 2017. <https://doi.org/10.1016/j.jaci.2017.09.023>. [Epub ahead of print].
25. Simonyte Sjödin K, Vidman L, Rydén P, West CE. Emerging evidence of the role of gut microbiota in the development of allergic diseases. *Curr Opin Allergy Clin Immunol*. 2016;16:390-395.
26. Martinez FD, Guerra S. Early origins of asthma: role of microbial dysbiosis and metabolic dysfunction. *Am J Respir Crit Care Med*. 2018;197:573-579.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Carraro S, Bozzetto S, Giordano G, et al. Wheezing preschool children with early-onset asthma reveal a specific metabolomic profile. *Pediatr Allergy Immunol*. 2018;29:375-382.
<https://doi.org/10.1111/pai.12879>

1 **ONLINE SUPPLEMENT**

2 **WHEEZING PRESCHOOL CHILDREN WITH EARLY-ONSET ASTHMA REVEAL A**
3 **DISTINCTIVE METABOLOMIC PROFILE**

4
5 **Silvia Carraro¹, Sara Bozzetto¹, Giuseppe Giordano^{1,2}, Dania El Mazloun¹, Matteo**
6 **Stocchero^{1,2}, Paola Pirillo^{1,2}, Stefania Zanconato¹, Eugenio Baraldi¹**

7
8 ¹ Department of Women's and Children's Health, University of Padova, Italy

9 ² Città della Speranza Institute of Pediatric Research (IRP), Padova, Italy

10

11 **METHODS**

12 **Spirometry**

13 Lung function parameters were measured with a 10-liter bell spirometer (Biomedin,
14 Padova, Italy). Spirometry was performed at baseline and 15 minutes after administering
15 400 mcg of salbutamol; the reversibility test was considered positive for a FEV1 change $\geq$
16 12% from baseline.

17 **Urine sample preparation**

18 The analysis of urine samples was performed at the Laboratory of Mass Spectrometry of
19 Institute of Pediatric Research (IRP) Women's and Children's Health Department of
20 University-Azienda Ospedaliera of Padova. Urine samples, after being thawed at room
21 temperature, were stirred for 30 seconds in a vortex mixer, then were centrifuged at 6000
22 g for 10 minutes to remove the sediment present in the urine. 200 μ L of the supernatant
23 were added to 900 μ L of a solution of water with 0.1% formic acid (FA) to obtain a 1:5
24 dilution, then the diluted sample was transferred into a glass vial suitable for the auto-
25 sampler, kept at 5°C.

26 **Chromatographic and mass spectrometry analysis**

27 The metabolic analysis of the collected urine samples was performed by the high
28 resolution mass spectrometer Q-ToF Synapt G2 (Waters) interfaced with the
29 chromatographic system UPLC (Ultra Performance Liquid Chromatography) (Acquity-
30 Waters), characterized by high chromatographic resolution, short analytical time and
31 enhanced sensitivity. Solvents and reactants with LC-MS ULTRA grade (Sigma St. Luis,
32 MO, USA) and water type MilliQ with 1 grade of purity were used for the analysis.

33 The chromatographic analysis was performed through the reverse phase column HSS T3
34 1,8 μ m 2,1mm and 100mm (Acquity HSS T3, Waters co., Miliford, MA USA) at 50°C. The
35 solvents used for eluent A were H₂O and 0,1% of FA , and methanol:acetonitrile 90:10 and
36 0,1% FA for eluent B. Percentage of eluent B varied through time in the current way: at

37 time 0minutes 5%, at 3.5minutes 30%, at 6.5minutes 95%, at 11minutes 5%, with a
38 constant flow of 500 μ L/min. The MS analysis was conducted with an Electrospray source
39 (ESI) in both positive (ESI+) and negative (ESI-) ionization mode. The settings of the ESI
40 were arranged at a capillary voltage of 3KV in the positive polarity and 1.5KV in the
41 negative, the source temperature being of 110°C and the desolvation temperature of
42 350°C with nitrogen gas flow at 600L/h. The mass scan was set in a range between 20
43 and 1200amu in 0.3sec.

44 All the samples were injected in triplicate with random sequences in order to prevent any
45 analytical artifact from conditioning the samples classification.

46 In order to verify the analytical reproducibility during the chromatographic analysis, quality
47 control (QC) samples, made by mixing 20 μ L of each urine samples, were injected every
48 ten injections.

49 A standard mix of known compounds (Sulfadimethoxine, Cholic acid, Leucine-Enkephalin,
50 Erythromycin, Hippuric acid, Acetaminophene, Sulfaguanidine, Verapamil, Terfenadine and
51 Caffeine) were analyzed in the sequence to verify the mass accuracy and the process
52 efficiency during the analytical investigation.

53 The urine samples were analyzed in both LC-MS and LC-MS^E for structural hypothesis.
54 This last is a non-selective fragmentation technique based on the passage of ionized
55 molecules through the collision cell at High and Low collision energy during the
56 chromatography.

57 **Data processing and pre-treatment**

58 UPLC-MS data were transformed into data tables by the software MarkerLynx (Waters co.,
59 Miliford, MA USA). The parameters used for data extraction were optimized through the
60 preliminary analysis of the QC samples. As a result, the so called (RT_{mass}) variables
61 (where RT is the retention time and mass is the mass to charge ratio m/z of the chemical
62 compound) were generated. For each sample, variables were calculated as the median

63 mass intensity of the 3 replicates. The application of two ionization modes (positive and
64 negative) led to the generation of 2 data sets, identified as T3 POS and T3 NEG.
65 Variables with more than 20% of missing data were eliminated to avoid spurious statistical
66 models generated by unrealistic combinations of the measured variables. For each
67 variable passing such filter, missing data were filled with the minimum value measured for
68 the variable. Median Fold Change normalization was applied in order to take into account
69 dilution effects, and log-transformation was performed to reduce the scale effects of the
70 variables on the models. The obtained data sets were mean centred before statistical data
71 analysis.

72 **Statistical data analysis**

73 Clinical qualitative variables were compared using Fisher's exact test, and spirometric
74 parameters were compared with the t-test for independent samples.

75 A multivariate data analysis based on projection methods was used to model the data
76 collected. Specifically, an exploratory data analysis was performed using Principal
77 Component Analysis (PCA), and Projection to Latent Structures Discriminant Analysis
78 based on VIP (Variable Influence on Projection) selection (VIP-based PLS-DA) [E1] was
79 applied to identify the differences between the groups under investigation. Stability
80 selection based on Monte-Carlo sampling [E2] was used to highlight the subset of
81 relevant variables characterizing the differences between the groups and to estimate the
82 predictive power of the models. Two hundred random subsamples of the samples
83 collected were extracted by Monte-Carlo sampling (with a prior probability of 0.70), then
84 VIP-based PLS-DA was applied to each subsample, obtaining a set of 200 discriminant
85 models. The relevant discriminant variables were those selected in more than 50% of all
86 the random models. The predictive performance of each model was estimated by means
87 of a ROC (Receiver Operating Characteristic) analysis on the outcome prediction of the
88 samples excluded during Monte-Carlo sampling, and expressed in terms of the Area

89 Under the ROC curve in prediction (AUC_{pred}). The VIP threshold to use in the VIP-based
90 PLS-DA was ascertained by maximizing the Q^2 value calculated by means of a seven-fold
91 full cross-validation (Q^2 is the cross-validated R^2 , and it is an estimate of the predictive
92 power of the model, whereas R^2 indicates the goodness of fit) under the constraint that the
93 model had to pass the permutation test on the class response (p-value less than 0.05).
94 Since multivariate analysis explores the correlation structure of the data collected, while
95 univariate analysis investigates the properties of single variables, we ran the univariate
96 data analysis using the t-test and ROC analysis in order to complete the results of the
97 multivariate data analysis.

98 VIP-based PLS-DA with stability selection and the univariate data analysis were
99 implemented on the R 3.1.2 platform (R Foundation for Statistical Computing), while the
100 PCA was done with SIMCA 14 (MKS Data Analytics Solutions).

101 The main available metabolomic databases (Human Metabolome DataBase [HMDB] and
102 METLIN) were searched to identify the relevant variables with a plausible biological
103 significance characterizing each group.

104 **Identification of relevant variables**

105 To identify the relevant variables, characteristic of each clinical group, we searched the
106 main available metabolomics databases (Human Metabolome DataBase (HMDB) and
107 METLIN). These databases enable the comparison between the spectroscopic
108 characteristics of the variables and those of known metabolites. This process is the first of
109 different steps that eventually lead to the identification of potential biomarkers, and it
110 permits to hypothesize both the chemical structure and the biological activity of the
111 putative molecules.

112 The METLIN database has provided the most useful information for the characterization of
113 the relevant variables. The research has been set with the subsequent parameters: m/z
114 ratio of the analyzed variable; charge of the mass (depending on the ionization mode);

ionization type $[M+H]^+$ or $[M+H]^-$ (as the electrospray ionization is a soft technique which usually produces molecular ion m/z (charge molecules)); error tolerance of less than 10 ppm (where $\text{ppm} = ((M_{\text{measured}} - M_{\text{expected}}) / M_{\text{expected}}) * 10^6$). In order to obtain additional structural information, the metabolites were analyzed by their molecular ion to get the Elemental Composition, and by the MS^E fragmentation spectra (using the Mass Fragment tools of Masslynx software) to confirm the hypothetical structure. These approaches help establishing the chemical structure of the hypothetical metabolite derived from the initial variable.

123

124 RESULTS

125 **Table S1.** Relevant variables for the T3 NEG data set (n=51) obtained merging the results
126 of the multivariate and univariate data analysis.

127

Rt	m/z	type	AUC ROC (95%)	specificity	sensitivity	t-test p-value	putative biomarker
5.478	400.1421	>T	0.668 - 0.964	0.81	0.75	1.3E-03	
1.046	159.0297	>T	0.591 - 0.947	0.88	0.69	4.2E-03	oxoadipic acid
4.597	182.0812	>T	0.612 - 0.942	0.69	0.75	6.0E-03	(-)-epinephrine
5.998	442.1892	>T	0.595 - 0.943	0.81	0.75	8.4E-03	
6.138	481.2073	>T	0.598 - 0.948	0.81	0.75	1.1E-02	
4.109	180.0653	>T	0.540 - 0.912	0.63	0.88	1.6E-02	L-tyrosine
2.165	194.0449	>T	0.563 - 0.913	0.69	0.75	1.7E-02	3-hydroxyhippuric acid
2.162	295.9833	>T	0.579 - 0.928	0.88	0.63	1.8E-02	
5.301	231.0322	>T	0.467 - 0.860	1.00	0.44	2.0E-02	
4.029	178.0494	>T	0.555 - 0.921	0.75	0.69	2.0E-02	
4.774	184.0969	>T	0.557 - 0.919	0.88	0.69	2.5E-02	
1.476	159.0658	>T	0.564 - 0.928	0.81	0.69	2.7E-02	
5.432	365.0670	>T	0.535 - 0.924	0.75	0.75	3.0E-02	
4.291	210.0753	>T	0.528 - 0.901	0.69	0.81	3.1E-02	
3.395	251.1124	>T	0.485 - 0.874	0.69	0.69	3.3E-02	
1.153	341.0383	>T	0.544 - 0.901	0.56	0.81	3.5E-02	
4.524	255.0877	>T	0.531 - 0.897	0.69	0.75	3.6E-02	
0.648	211.0271	>T	0.501 - 0.889	0.75	0.69	4.0E-02	
4.241	339.1650	>T	0.471 - 0.864	0.56	0.81	4.1E-02	
0.769	350.9718	>T	0.512 - 0.885	0.94	0.50	4.2E-02	
3.168	121.0286	>T	0.504 - 0.878	0.81	0.56	4.6E-02	benzoic acid
2.921	429.0367	>T	0.428 - 0.837	0.81	0.50	4.7E-02	

0.571	108.9010	>T	0.498 - 0.876	0.81	0.56	4.8E-02	
4.916	217.1070	>T	0.532 - 0.897	0.69	0.69	5.0E-02	3-hydroxy-sebacic acid
4.101	275.0225	>T	0.489 - 0.877	0.81	0.56	5.3E-02	dihydroferulic acid 4-sulfate
0.623	293.0680	>T	0.527 - 0.902	0.81	0.63	5.9E-02	
2.964	307.0056	>A	0.634 - 0.951	0.81	0.75	3.2E-03	
4.362	113.0235	>A	0.630 - 0.947	0.75	0.75	5.1E-03	
0.640	177.0394	>A	0.568 - 0.923	0.81	0.63	5.5E-03	
4.982	436.2193	>A	0.618 - 0.952	0.81	0.81	5.6E-03	
4.362	350.0870	>A	0.607 - 0.946	0.88	0.63	8.0E-03	
6.084	435.2221	>A	0.573 - 0.926	0.94	0.56	8.7E-03	
4.098	193.0344	>A	0.570 - 0.929	0.69	0.75	9.2E-03	
4.361	723.1638	>A	0.602 - 0.937	1.00	0.56	1.0E-02	
0.711	351.0559	>A	0.575 - 0.924	0.69	0.75	1.2E-02	4-(4-deoxy- α -D-gluc-4-enuronosyl)-D-galacturonate
1.183	418.1095	>A	0.585 - 0.930	0.88	0.63	1.2E-02	
5.262	485.1655	>A	0.571 - 0.928	0.63	0.88	1.6E-02	
4.363	174.0534	>A	0.615 - 0.947	0.81	0.69	1.8E-02	
5.987	481.2430	>A	0.547 - 0.906	0.75	0.69	2.6E-02	
4.362	131.0337	>A	0.607 - 0.946	0.88	0.69	2.9E-02	glutaric acid
2.354	180.0124	>A	0.580 - 0.934	0.69	0.81	3.0E-02	
1.119	174.0161	>A	0.534 - 0.903	0.69	0.81	3.0E-02	
4.359	701.1819	>A	0.516 - 0.890	0.88	0.56	3.4E-02	
5.433	155.1069	>A	0.529 - 0.907	0.88	0.63	3.4E-02	4-hydroxy nonenal
4.362	466.0558	>A	0.519 - 0.894	0.69	0.63	3.7E-02	
4.061	261.0436	>A	0.521 - 0.908	0.75	0.69	4.2E-02	
6.200	435.2225	>A	0.518 - 0.895	0.63	0.69	4.2E-02	
1.475	147.0657	>A	0.517 - 0.896	0.75	0.63	4.6E-02	
0.604	245.0424	>A	0.495 - 0.871	0.63	0.69	4.9E-02	phosphatidyl glycerol
1.374	257.0773	>A	0.482 - 0.869	0.75	0.63	5.5E-02	3-methyluridine
4.730	335.1334	>A	0.485 - 0.873	0.69	0.75	7.3E-02	steroid O-sulfate

128

129

130

131

132

133

Rt=Retention time; m/z=mass to charge ratio; AUC ROC (95%)=confidence interval of the Area Under the ROC curve at the level of 95%; >A=variables with higher levels in the early onset asthma group than in the transient wheezing group; >T=variables with higher levels in the transient wheezing group than in the early onset asthma group.

134 **Table S2.** Relevant variables for the T3 POS data set (n=80) obtained merging the results
 135 of the multivariate and univariate data analysis.

136

Rt	m/z	type	AUC ROC (95%)	specificity	sensitivity	t-test p-value	putative biomarker
4.590	109.0654	>T	0.662 - 0.970	0.88	0.69	9.5E-04	p-cresol
6.017	426.1943	>T	0.673 - 0.982	0.81	0.81	1.9E-03	
6.069	315.1215	>T	0.646 - 0.963	0.69	0.94	3.0E-03	
2.091	239.1004	>T	0.616 - 0.946	0.75	0.75	3.5E-03	
6.134	505.2030	>T	0.643 - 0.973	0.94	0.69	4.0E-03	
5.484	384.1469	>T	0.628 - 0.949	0.88	0.63	4.2E-03	
3.552	316.1759	>T	0.600 - 0.946	0.81	0.75	6.3E-03	
6.253	299.1258	>T	0.543 - 0.901	0.81	0.56	9.2E-03	
4.233	341.1808	>T	0.589 - 0.925	0.69	0.75	1.0E-02	
5.637	472.2195	>T	0.592 - 0.930	0.75	0.63	1.0E-02	
4.231	379.1359	>T	0.557 - 0.918	0.63	0.88	1.2E-02	
6.266	519.2189	>T	0.544 - 0.908	0.69	0.69	1.4E-02	
1.142	343.0534	>T	0.543 - 0.901	0.56	0.75	1.5E-02	
5.768	334.1661	>T	0.582 - 0.940	0.75	0.88	1.6E-02	
6.133	289.1802	>T	0.622 - 0.955	0.81	0.75	1.6E-02	
2.754	346.2226	>T	0.540 - 0.905	0.69	0.69	1.6E-02	
3.810	404.2276	>T	0.599 - 0.939	0.75	0.75	1.6E-02	
4.863	146.0604	>T	0.584 - 0.923	0.63	0.75	1.7E-02	
6.253	259.1334	>T	0.578 - 0.921	0.56	0.88	1.8E-02	
4.585	206.0798	>T	0.576 - 0.939	0.69	0.81	1.8E-02	indolelactic acid
6.253	277.1435	>T	0.577 - 0.922	0.63	0.81	2.0E-02	
5.635	245.0493	>T	0.592 - 0.946	0.75	0.75	2.0E-02	
4.760	186.1129	>T	0.573 - 0.926	0.88	0.69	2.2E-02	
3.846	297.1546	>T	0.498 - 0.884	0.56	0.88	2.2E-02	
3.151	389.0814	>T	0.562 - 0.917	0.75	0.63	2.5E-02	
4.760	208.0953	>T	0.535 - 0.910	0.69	0.75	2.7E-02	N-acetyl-L-phenylalanine
4.404	293.1162	>T	0.545 - 0.907	0.88	0.56	2.8E-02	
1.151	245.1609	>T	0.548 - 0.904	0.75	0.56	3.3E-02	
3.463	360.2018	>T	0.503 - 0.879	0.69	0.63	3.3E-02	
5.441	231.0333	>T	0.590 - 0.956	0.81	0.81	3.4E-02	
4.720	446.2597	>T	0.512 - 0.885	0.94	0.50	3.6E-02	
3.156	225.0344	>T	0.511 - 0.886	0.63	0.69	3.7E-02	
6.220	376.2122	>T	0.510 - 0.887	0.69	0.69	3.8E-02	
4.243	349.1836	>T	0.437 - 0.835	0.44	0.88	3.8E-02	
4.233	363.1623	>T	0.456 - 0.856	0.56	0.81	3.8E-02	
2.086	175.1081	>T	0.537 - 0.900	0.69	0.69	3.9E-02	N2-acetyl-ornithine
1.138	102.0555	>T	0.521 - 0.899	0.81	0.63	3.9E-02	
6.253	231.1385	>T	0.548 - 0.911	0.56	0.88	4.1E-02	
4.762	83.0860	>T	0.547 - 0.929	0.75	0.75	4.2E-02	
4.903	137.0968	>T	0.528 - 0.900	0.75	0.69	4.3E-02	

0.954	240.1354	>T	0.500 - 0.881	0.88	0.56	4.7E-02	
4.719	429.2327	>T	0.460 - 0.852	0.63	0.69	4.7E-02	
4.523	407.1883	>T	0.492 - 0.874	0.63	0.75	4.9E-02	
4.533	409.1833	>T	0.416 - 0.818	0.69	0.50	5.0E-02	
5.533	167.1072	>T	0.520 - 0.909	0.69	0.75	5.0E-02	
4.340	130.0657	>A	0.618 - 0.943	0.69	0.81	4.1E-03	
0.531	205.1552	>A	0.609 - 0.937	0.81	0.69	5.2E-03	
2.387	221.0926	>A	0.551 - 0.917	0.56	0.94	8.0E-03	5-hydroxy-L-tryptophan
4.951	438.2349	>A	0.564 - 0.912	0.56	0.75	1.0E-02	
4.340	374.0834	>A	0.615 - 0.947	0.69	0.88	1.2E-02	
5.302	374.2536	>A	0.549 - 0.919	0.88	0.63	1.4E-02	
4.339	176.0711	>A	0.603 - 0.950	0.75	0.81	1.4E-02	3-indoleacetic-acid
2.558	158.0819	>A	0.523 - 0.898	0.94	0.56	1.5E-02	tiglylglycine
1.199	157.0723	>A	0.581 - 0.926	0.75	0.69	1.6E-02	
4.058	118.0658	>A	0.582 - 0.925	0.75	0.69	1.6E-02	indole
4.340	390.0580	>A	0.559 - 0.932	0.75	0.81	1.8E-02	
4.951	394.2448	>A	0.504 - 0.878	0.75	0.56	1.8E-02	
1.222	112.0512	>A	0.520 - 0.893	0.75	0.69	2.1E-02	cytosine
3.953	292.1576	>A	0.560 - 0.931	0.81	0.69	2.1E-02	
4.338	218.0818	>A	0.577 - 0.930	0.75	0.75	2.2E-02	
5.843	504.2800	>A	0.526 - 0.894	0.69	0.75	2.3E-02	
0.874	364.0636	>A	0.530 - 0.922	0.63	0.94	2.4E-02	
0.692	131.1189	>A	0.545 - 0.907	0.81	0.63	2.4E-02	N-acetylputrescine
2.391	175.0873	>A	0.511 - 0.887	0.75	0.69	2.4E-02	indole-3-acetamide
6.196	189.1644	>A	0.542 - 0.910	0.69	0.69	2.5E-02	
1.089	223.0224	>A	0.547 - 0.905	0.69	0.69	2.8E-02	
3.147	150.0780	>A	0.553 - 0.907	0.56	0.81	2.8E-02	6-methyladenine
4.197	316.2119	>A	0.565 - 0.919	0.63	0.81	2.9E-02	
4.753	299.0859	>A	0.543 - 0.909	0.69	0.75	3.0E-02	
0.883	247.1293	>A	0.528 - 0.900	0.69	0.81	3.1E-02	
0.835	126.0667	>A	0.547 - 0.906	0.63	0.69	3.4E-02	5-methylcytosine
3.708	579.7748	>A	0.513 - 0.900	0.75	0.69	3.5E-02	
1.200	185.0676	>A	0.550 - 0.910	0.69	0.69	3.9E-02	
4.406	243.1343	>A	0.520 - 0.893	0.69	0.75	3.9E-02	
6.276	532.3122	>A	0.500 - 0.898	0.63	0.88	4.0E-02	
4.408	130.0507	>A	0.509 - 0.880	0.56	0.75	4.0E-02	N-acryloylglycine
1.125	256.0929	>A	0.595 - 0.989	0.75	1.00	4.3E-02	
1.201	181.0724	>A	0.545 - 0.915	0.69	0.75	4.3E-02	
1.455	195.0771	>A	0.511 - 0.886	0.63	0.81	4.3E-02	
3.940	183.0659	>A	0.539 - 0.898	0.63	0.69	5.8E-02	hydroxyphenyllactic acid

137

138 Rt=Retention time; m/z=mass to charge ratio; AUC ROC (95%)=confidence interval of the
139 Area Under the ROC curve at the level of 95%; >A=variables with higher levels in the early

140 onset asthma group than in the transient wheezing group; >T=variables with higher levels
141 in the transient wheezing group than in the early onset asthma group.

142

143 **REFERENCES**

144 E1. Chong I-G, Jun C-H: Performance of some variable selection methods when
145 multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems*
146 2005;78:103-112.

147 E2. Arredouani A, Stocchero M, Culeddu N, El-Sayed Moustafa J, the D.E.S.I.R. study
148 group, Tichet J, Balkau B, Brousseau T, Manca M, Falchi M. Metabolomic profile of low
149 copy-number carriers at the salivary alpha-amylase gene suggests a metabolic shift
150 towards lipid-based energy production. *Diabetes* 2016;65:3362-3368.

151