

Article

Metabonomics of newborn screening dried blood spot samples – a novel approach in the screening and diagnostics of inborn errors of metabolism

Julia Denes, Eszter Szabo, Steve Robinette, Ildiko Szatmari, Laszlo Szonyi, Joachim G. Kreuder, Ernst W. Rauterberg, and Zoltan Takats

Anal. Chem., **Just Accepted Manuscript** • Publication Date (Web): 24 Oct 2012

Downloaded from <http://pubs.acs.org> on October 29, 2012

Just Accepted

“Just Accepted” manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides “Just Accepted” as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. “Just Accepted” manuscripts appear in full in PDF format accompanied by an HTML abstract. “Just Accepted” manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). “Just Accepted” is an optional service offered to authors. Therefore, the “Just Accepted” Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the “Just Accepted” Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these “Just Accepted” manuscripts.



ACS Publications
High quality. High impact.

Analytical Chemistry is published by the American Chemical Society, 1155 Sixteenth Street N.W., Washington, DC 20036

Published by American Chemical Society. Copyright © American Chemical Society. However, no copyright claim is made to original U.S. Government works, or works produced by employees of any Commonwealth realm Crown government in the course of their duties.

Metabonomics of newborn screening dried blood spot samples – a novel approach in the screening and diagnostics of inborn errors of metabolism

Júlia Dénes¹, Eszter Szabó², Steve Robinette¹, Ildikó Szatmári³, László Szőnyi³, Joachim G. Kreuder⁴, Ernst W. Rauterberg⁴, Zoltán Takáts^{1,3,*}

¹Biomolecular Medicine, Department of Surgery and Cancer, Imperial College London, United Kingdom

²Massprom Ltd. Budapest, Hungary

³Metabolic Diagnostic Laboratory, 1st Department of Pediatrics, Semmelweis University, Budapest, Hungary

⁴Hesse Child Health Centre, Johann Wolfgang Goethe University Hospital, Giessen, Germany

Keywords: neonatal screening, inborn errors of metabolism, high resolution mass spectrometry, metabonomics, principal component analysis, PKU, MCADD

Abstract

A novel, single stage high resolution mass spectrometry-based method is presented for the population level screening of inborn errors of metabolism. The approach proposed here extends traditional electrospray tandem mass spectrometry screening by introducing nanospray ionization and high resolution mass spectrometry, allowing the selective detection of more than 400

individual metabolic constituents of blood including acylcarnitines, amino acids, organic acids, fatty acids, carbohydrates, bile acids, and complex lipids. Dried blood spots were extracted using methanolic solution of isotope labeled internal standards, and filtered extracts were electrosprayed using a fully automated chip-based nanospray ion source in both positive and negative ion mode. Ions were analyzed using an Orbitrap Fourier transformation mass spectrometer at nominal mass resolution of 100,000. Individual metabolic constituents were quantified using standard isotope dilution method. Concentration threshold (cut-off) level-based analysis allows the identification of newborns with metabolic diseases, similarly to traditional ESI-MS/MS method, however the detection of additional known biomarkers (e.g. organic acids) results in improved sensitivity and selectivity. The broad range of detected analytes allowed the untargeted multivariate statistical analysis of spectra and identification of additional diseased states, therapeutic artifacts and damaged samples, besides the metabolic disease panel.

Introduction

Since its introduction, electrospray tandem mass spectrometry (ESI-MS/MS) has become the most important and widely applied method in newborn screening (NBS) programs.¹⁻³ Currently, there are more than 70 different metabolic diseases included in neonatal screening, with approximately 90% of them screened by the ESI-MS/MS method. The standard mass spectrometry newborn screening procedure comprises the extraction of dried blood spot (DBS) samples with a methanolic solution of stable isotope labeled internal standards. Amino acids and acylcarnitines are semi-quantitatively determined by detecting their n-butyl ester derivatives in positive ion mode tandem mass spectrometry. Amino acids are traditionally detected by neutral loss (NL 102) scan and acylcarnitines by precursor ion (PS 85) scan, with quantification based generally on multiple reaction monitoring (MRM) data.

1
2
3
4
5
6 Ionization efficiency and hence sensitivity is routinely increased via butyl esterification by
7
8 eliminating the acidic carboxyl moiety from the analyte molecules. Derivatization is not an
9
10 absolute prerequisite, however, only 30% of screening laboratories use currently derivatization-
11
12 free methods.^{4,5} Elimination of butyl esterification is considered to yield more accurate results by
13
14 avoiding the formation of free amino acids via acidic hydrolysis of protein content.
15
16

17
18 Improving the sensitivity and specificity of screening, with a consequent reduction in the false-
19
20 positive rate is a key area of development in the newborn screening field. Cut-off values in case
21
22 of the tandem mass spectrometry-based technique are established to minimize the false negative
23
24 rate to zero level. As a result, the number of the false-positive cases can occasionally be high,
25
26 especially in case of certain diseases, such as homocystinuria, methylmalonic acidemia and
27
28 propionic acidemia (methionine and propionylcarnitine are used as diagnostic markers for all
29
30 three diseases). Since suspected positive cases require confirmatory testing, false-positive cases
31
32 involve unnecessary stress for the family of affected newborns and increased analytical burden
33
34 on the metabolic diagnostic laboratories.^{6,7} Screening laboratories have used two main strategies
35
36 to reduce the false-positive rate: excluding certain diseases from the screening (especially those,
37
38 which otherwise do not fulfill the criteria for screening postulated by Wilson and Jungner⁸), and
39
40 introducing of so-called second-tier tests⁹ that use the original DBS sample for confirmation of
41
42 the positive result. In these cases the initial positive screening result is not reported until the
43
44 confirmatory test validates this result, decreasing the unnecessary stress on families.¹⁰
45
46
47
48
49

50 Development of second-tier tests by the Mayo Clinic group provided a sufficient solution for
51
52 the problem by minimizing overall false-positive rates to 0.09%.⁹ These tests are highly specific
53
54 compared to the screening methods¹¹⁻¹⁶, but their introduction requires additional
55
56
57
58
59
60

instrumentation, trained personnel, and an extension of the turnover time of screening, which may be critical in the case of several diseases.

Although most of the screened metabolic disorders are caused by the complete or partial absence of a single enzymatic function, the missing functionality results not only in the increase of the level of the specific substrate. Multiple metabolic constituents, both up- and downstream in the metabolic pathways show changes in their concentrations, furthermore, elevation of the concentration of certain metabolites may also activate alternative pathways. As a result, serum profiles should reflect broad perturbations in the metabolome characteristic of individual diseases. While metabolomic studies traditionally employ GC-MS or LC-MS for the untargeted analysis of biological fluids¹⁷⁻¹⁹, the feasibility of direct infusion mass spectrometry analysis has been demonstrated by a number of recent publications.²⁰⁻²² Advantage of the elimination of chromatographic separation include shorter analysis times, simplified data analysis and improved reproducibility. Since the specificity of unit resolution mass analyzers is poor, either high resolution (HR) or tandem mass spectrometry was used, to ensure selective detection of thousands of spectral features. Nevertheless, these improvements do not compensate for the suppression effects of electrospray ionization, which represent the most serious drawback for MS-only approaches.²³ Besides employing chromatographic separation prior to MS detection, suppression effects can partially be eliminated by minimizing the diameter of electrospray emitter tips, i.e. using nano-electrospray (or nanospray) ionization.²⁴ Nanospray has been demonstrated to suffer less from suppression effects, both in case of large proton affinity differences between different analytes and presence of inorganic salts in the sample.

The combination of nanospray ionization with high resolution mass spectrometry (nS-HR-MS) provides metabolome coverage comparable to that of LC-MS approaches as it is demonstrated

by the current study. The concept was implemented by developing a high-throughput nanospray-HR-MS method for mapping the metabolic profile of DBS samples and establishing correlation between metabolite concentrations determined using the new approach and the traditional ESI-MS/MS technique. Resulting metabolic profiles were also correlated with the presence of various diseased statuses of the subjects using multivariate statistical methods. The developed method was tested for the most widely screened metabolic diseases including phenylketonuria (PKU) and medium-chain acyl-coenzyme A dehydrogenase deficiency (MCADD).

Methods and Materials

Samples

Anonymous DBS samples were obtained from the Screening Laboratory at Semmelweis University, Budapest, Hungary and Hesse Screening Centre, Gießen, Germany. 500 healthy and 66 abnormal samples were analyzed; for further details see Supporting Information.

Sample preparation

Dried blood spot samples ($d=0.6$ mm, ~ 10 μ l of whole blood) were placed in 96-well filter plates (Millipore, Billerica, MA, USA) and extracted with 100 μ l methanolic (HPLC grade, Sigma-Aldrich, St. Louis, MO, USA) solution, containing known concentrations of stable isotope labeled standards (Cambridge Isotope Laboratories, Andover, MA, USA), see Table S-1 for details. The samples were extracted for 30 min at room temperature. The methanolic extracts were subsequently filtered into 96-well plates (Greiner, Frickenhausen, Germany), diluted with water (HPLC grade, Sigma-Aldrich, St. Louis, MO, USA) in 1:1 ratio and analyzed by mass spectrometry.

Mass spectrometry

High resolution mass spectrometry was performed using LTQ Orbitrap Discovery and Exactive instruments (Thermo Scientific GmbH, Bremen, Germany). High-throughput screening method was developed using the Exactive instrument equipped with TriVersa NanoMate ion source (Advion, Ithaca, NY, USA). Chip-based infusion mode measurements were performed using 5 μm nominal internal diameter nozzle chips and sample volume of 5 μl was injected. Total data acquisition time was 2 minutes, 1 minute in negative ion mode, 1 minute in positive ion mode with automatic polarity switching. Details about the data analysis are provided in the Supporting Information material.

Results and Discussion

Detection and identification of metabolites

Nanospray–high resolution mass spectrometry analysis of the methanolic extracts of newborn DBS samples results in spectra featuring hundreds of different metabolic constituents both in positive and negative mode. Parameters of the instruments were optimized to achieve the best intensity, sensitivity and mass resolution in the lower mass range of 70-600 m/z (parameters are summarized in Table S-2). Since the mass spectrometry device allows alternating polarity acquisition, positive and negative ion spectra were recorded in the same experiment (Figure S-1). Positive ion mode data features ~ 450 individual ion signals (excluding isotopes and adduct peaks) in the mass range of 70-1000 m/z being present in all 500 samples obtained from healthy newborns and analyzed by high resolution mass spectrometry. Negative ion mode data contains information about the presence of ~ 300 molecular constituents in the identical sample set. In addition to the core set of frequently observed ions, several hundreds of ion signals are present

only in a fraction of the samples. Taking all samples into account, a total of 1150 ion signals were registered (without isotope and adduct signals) using a 10:1 signal-to-noise threshold.

Identification of ion signals was performed using accurate masses of ions determined at 50,000 – 100,000 FWHM (depending on actual m/z of the ion) nominal resolution with mass accuracy window of 1.0 ppm and tandem mass spectrometry. The accurate mass in the given mass range at the given precision generally allows the unequivocal determination of the molecular formula of species detected. Given the constraints set by the human biological origin of the species, one ion signal can be associated with only a few isomeric metabolic constituents. Tandem MS experiments were carried out to clarify the isomeric composition of ion signals corresponding to multiple known metabolites. Proper identification was achieved for 185 species in negative ion mode and 93 species in positive ion mode corresponding to overall 278 identified molecular constituents up to date, however the peak identification process is still ongoing. Results of the identification are summarized in Table S-3.

The observed metabolites include practically all amino acid and acylcarnitine metabolites relevant to newborn screening in addition to metabolites utilized for confirmatory diagnostics of metabolic diseases, including organic acids, carbohydrates, carbohydrate phosphates, bile acids, fatty acids, eicosanoids, phospholipids, and sphingolipids. Although single stage high resolution mass spectrometry does not allow the differentiation of isomers, use of instruments with automated fragmentation of selected peaks (LTQ Orbitrap) enables these distinctions to be made. The resulting MS/MS spectra often provide sufficient information on the isomeric composition of the corresponding peak, as it is demonstrated in case of 4-hydroxyphenyllactic- and homovanillic acid on Figure S-2.

One of the main concerns regarding quantitative determination using FT-MS technique is the narrow dynamic range provided by these instruments. Limit of detection (LOD) and linear range for a number of key metabolites was determined by extracting parallel samples (discs punched from the same NBS card, containing DBS from a single subject taken at the same time) with methanol containing various concentrations of isotope labeled internal standards. Identical DBS sample is expected to contain reproducible amount of the analyte, so in this study the native metabolite was used as internal standard. LOD values were defined as the concentration which yields 3:1 signal-to-noise ratio. Linear range was defined as the widest logarithmic concentration range where linear regression to the data gives $R^2 > 0.990$. Results are shown in Table S-4 for 12 amino acids and 8 acylcarnitines generally used in newborn screening programs. As it is demonstrated, the LOD values are always lower than the physiological (or pathological if disease results in lower than physiological concentration) levels of the metabolites. Calibration curves exceeding beyond the linear range always show saturation phenomenon at higher concentrations (i.e. no breakdown phenomenon was observed). This guarantees that all pathological values will be detected as diseased, even if the determined level falls outside of the linear range.

Validation against traditional ESI-MS/MS method

Both the traditional ESI-MS/MS method and the nS-HR-MS method presented here were able to quantify a set of 20 amino acid and 20 acylcarnitine species. These analytes were used as a benchmark to validate the nS-HR-MS method against the ESI-MS/MS techniques for a set of 500 newborn samples analyzed with both methods. Correlation plots generally reveal a linear relationship with correlation coefficients (Pearson product-moment correlation coefficient, r) in the range of 0.4 to 0.92. The slope of functions obtained by linear regression was < 1 when nS-HR-MS data is plotted as a function of values measured by ESI-MS/MS, indicating that the

former technique tends to yield equal or lower values compared to the latter one (see Figure S-3). The systematic deviation was tentatively associated with the improved selectivity of the high resolution method, but we could not confirm this assumption due to the inability of the Orbitrap mass spectrometer to isolate parent ions at arbitrarily narrow m/z windows. Furthermore, while the HR-MS method detects the unmodified molecule, the traditional method detects its butyl-ester, raising further complications for the comparison. The butyl-esterification procedure may also induce the hydrolysis of proteins, which shifts the determined amino acid concentrations further towards higher values. Nevertheless, both nS-HR-MS and ESI-MS/MS were validated against a quantification of amino acids by LC-MS/MS, and the results for a limited number of samples clearly showed that nS-HR-MS results gives better correlation with those of LC-MS/MS (0.87 vs. 0.68 for Phe, 50 samples). These results indicate that the traditional method gives systematically higher concentrations caused either by interferences with similar molecular weight and fragmentation pattern or identical metabolite produced during sample preparation. However, this systematic shift does not necessarily compromise the diagnostic value of concentrations obtained by the traditional method; while the healthy and diseased range may be different for the nS-HR-MS and the traditional ESI-MS/MS approach, the separation of the ranges is comparable or better for metabolites determined by both methods.

Screening for PKU and MCADD

Although the set of diseases included in newborn screening programs varies from country to country (in some cases even between regions of the same country), phenylketonuria and medium-chain acyl-coenzyme A dehydrogenase deficiency are included in all programs where mass spectrometry is used for screening. Phenylketonuria is generally screened using the phenylalanine concentration as primary parameter and Phe/Tyr ratio (and occasionally Fischer

1
2
3
4
5
6 quotient) as secondary/confirmatory parameters.²⁵ Box plots representing the distribution of key
7
8 metabolite concentrations for healthy population and diseased population using the two methods
9
10 are shown on Figure 1 for MCAD deficiency and PKU, since samples from sufficiently large
11
12 number of freshly diagnosed patients were available in these two cases. The box plot diagrams
13
14 indicate that these analytes differentiate between healthy and diseased population using the nS-
15
16 HR-MS method.
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

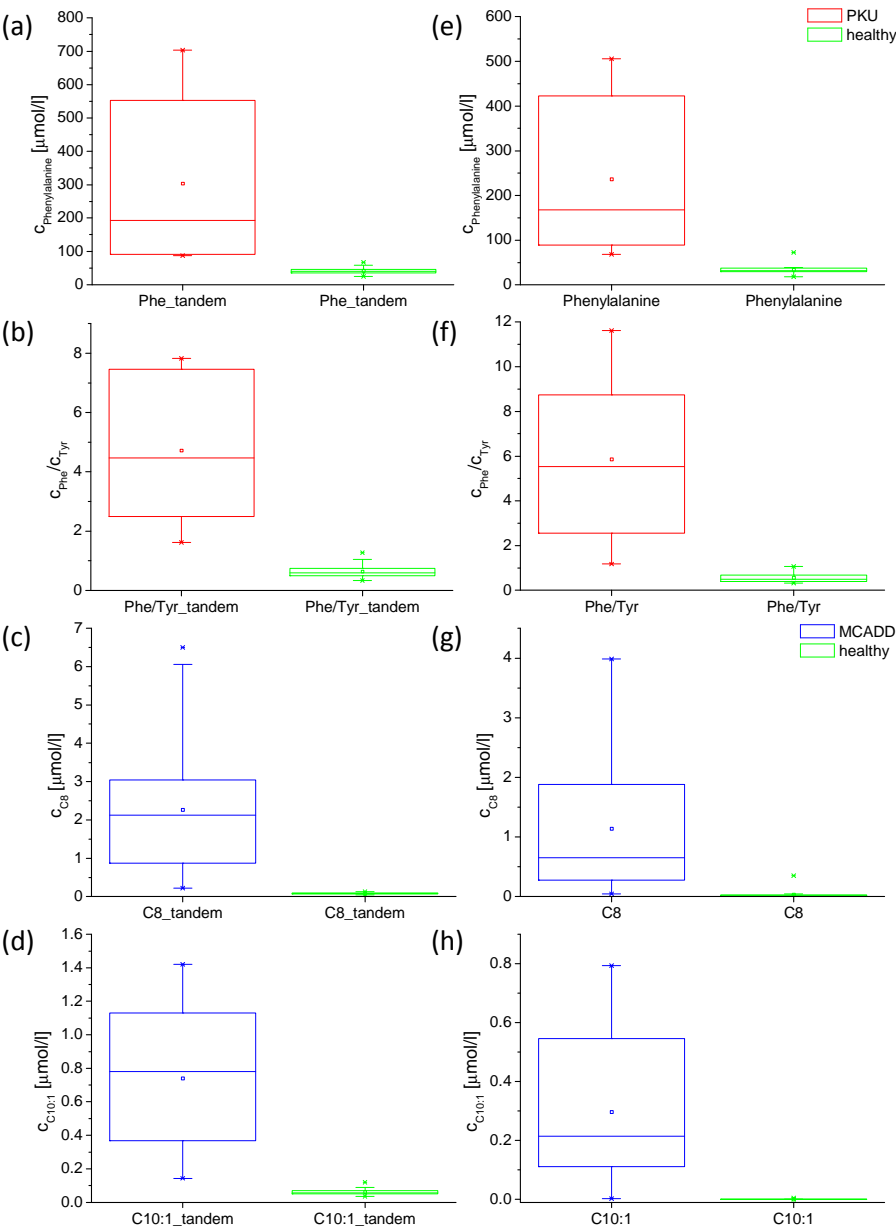


Figure 1. Box charts for PKU (a,b,e,f) and MCADD markers (c,d,g,h). (a, e) Phenylalanine, (b, f) Phenylalanine and Tyrosine ratio, (c, g) Octanoylcarnitine (C8) and (d, h) Decenoylcarnitine (C10:1). Tandem mass spectrometry screening results are shown on the left, nS-HR-MS results are on the right, from the same dataset (500 healthy, 21 PKU, 21 MCADD).

While both, sensitivity and specificity of phenylalanine level-based phenylketonuria screening is excellent, the rare false-positive cases are generally associated with a transient high phenylalanine level caused by total parenteral nutrition or immature metabolism of prematurely born babies. Both of these statuses are indicated in principle on the screening cards, however on one hand this information is often missing, on the other hand these babies require further examination anyway, according to currently existing protocols. In contrast, the novel method offers the utilization of a few additional parameters which are analytically not associated with the phenylalanine concentration. Phenyllactic acid and phenylpyruvic acid are metabolites along an alternative catabolic pathway of phenylalanine. Since phenylalanine is normally converted to tyrosine by the phenylalanine hydroxylase (PAH) enzyme, loss of this enzymatic function (i.e. in case of phenylketonuria) dramatically increases the concentration of the metabolites of the alternative pathway, e.g. that of phenyllactic and phenylpyruvic acid.²⁶ These two metabolites along with a few others are readily detected by the nS-HR-MS method in negative ion mode, using the same nanospray emitter. Separation of healthy and diseased population is even better in case of phenylpyruvic acid, than in the case of phenylalanine. DBS samples from 12 false-positive cases (babies screened out with Phe levels higher than 99 percentile, but having normal PAH activity) were analyzed using nS-HR-MS method and the phenylpyruvic acid levels were found to be $< 0.25 \mu\text{M}$ (cf. Figure 2a and b) in all cases. In 5 of these cases the elevated Phe level was associated with amino acid infusion, while in 3 cases the gestational age of babies was < 36 weeks. Although our results do not fully validate phenylpyruvic acid as a selective marker for PKU in cases of premature babies and babies received amino acid infusion, the observations strongly support its testing at population level as an additional/confirmatory marker for PKU.

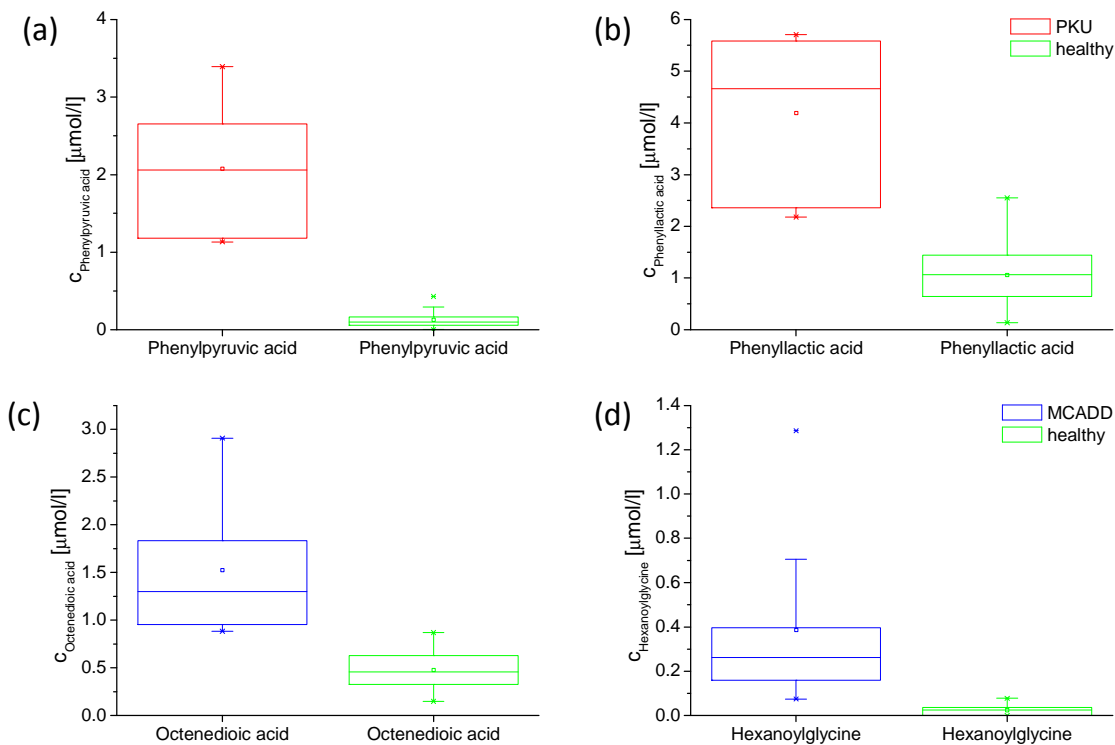


Figure 2. Box charts for possible markers of PKU (a) Phenylpyruvic acid and (b) Phenyllactic acid. Measurements were performed in negative ion mode for 500 healthy, 21 PKU and 12 false-positive PKU samples. Possible markers of MCADD (c) Octenedioic acid and (d) Hexanoylglycine. Measurements were performed in negative ion mode for 500 healthy, 21 MCADD samples.

1
2
3
4
5
6 Samples of patients suffering from other aminoacidopathies including tyrosinemia I,
7
8 citrullinemia and homocystinuria have also been tested with similar results. Markers found in
9
10 case of each disease are summarized in Table S-5.
11
12

13 MCAD deficiency is one of the most abundant mitochondrial fatty acid beta oxidation
14
15 disorders and it is included in most of the newborn screening programs worldwide. The
16
17 deficiency of medium chain acyl-coenzyme A dehydrogenase activity interrupts the stepwise
18
19 beta oxidation of long and medium chain fatty acids, causing problems in the general energy
20
21 (ATP) production of the body. The condition also results in the accumulation of fatty acid
22
23 metabolites produced by alternative (i.e. to mitochondrial beta oxidation) metabolic pathways,
24
25 e.g. peroxysomal beta oxidation or omega oxidation by cytochrome P450. The disease is
26
27 traditionally screened by ESI-MS/MS method, monitoring the elevated level of octanoylcarnitine
28
29 (C8). While an elevated octanoylcarnitine level is a hallmark of the disease, the sole use of this
30
31 metabolite results in fairly high (up to 3 %) false-positive rates. False-positives are generally
32
33 associated with the overlap of healthy (comprising also heterozygous cases) and diseased
34
35 population and administration of valproic acid to infants with epileptic symptoms. Since valproic
36
37 acid is an isomer of caprylic acid and conjugates with carnitine, the resulting valproylcarnitine
38
39 cannot be distinguished from C8. Further problem, pertaining to the specificity of the method is
40
41 that other disorders of mitochondrial fatty acid beta oxidation (e.g. Glutaric acidemia type 2) also
42
43 involve the elevation of C8 levels. Hence, additionally to C8, other acylcarnitine concentrations
44
45 (C2, C6, C10, C10:1 or their concentration ratios) are used as additional or confirmatory
46
47 markers.^{27,28} The nS-HR-MS method is also able to quantify these acylcarnitine levels and ratios,
48
49 as it is demonstrated by Figure 1.
50
51
52
53
54
55
56
57
58
59
60

The standard, generally accepted confirmatory test for newborns screened with suspected MCAD deficiency is urinary organic acid analysis, followed by genetic testing for the G985 hotspot mutation. In case of urinary GC-MS analysis, dicarboxylic acids, acylglycines and fatty acid metabolites are determined for the confirmation of the diagnosis. Unlike the standard positive ion ESI-MS/MS method, nS-HR-MS technique is able to detect and quantify these metabolites from the dried blood spot extracts. Since generally the urinary concentration (normalized to creatinine level) of these metabolites is used for diagnostics^{29,30}, it was a question whether the concentration determined in DBS samples also serves as an acceptable marker for the disease. As the box charts depicted on Figure 2 demonstrate, octenedioic acid and hexanoylglycine levels determined by the nS-HR-MS technique can serve as analytically independent markers for the MCAD deficiency. Similarly to PKU, we have studied the utility of these markers for improving the specificity of the results. 21 newborn DBS samples were selected which were found to be positive for MCAD deficiency, however the results of the confirmatory test did not support the diagnosis. For the traditional ESI-MS/MS method, C8, C6, C10:1-acylcarnitines and C10:1/C2-acylcarnitine ratio were used as screening parameters, and result was accepted as positive where at least 3 of these parameters were higher than pre-defined cutoff value. The nS-HR-MS analysis of these samples yielded confirmatory metabolite concentrations lower than the 1st percentile of diseased population, between 3rd and 44th percentile of healthy population for octenedioic acid, between 65th and 89th percentiles for hexanoylglycine and between 40 and 51st percentile for suberic acid.

The results indicate that using these set of markers as confirmatory markers, the false-positive cases found by acylcarnitine analysis can be correctly classified as healthy newborns. Using octanoylcarnitine level combined with these markers and accepting a case positive if all of the

parameters exceed the estimated 1 percentile value of diseased population, 100% selectivity and sensitivity was achieved for the detection MCAD deficiency in the test set, however the tested population (500 healthy and 21 diseased) is not sufficiently large to draw far-fetched conclusions regarding specificity and sensitivity of the method in a population level screening program.

As it has already been pointed out, the nS-HR-MS method reveals the presence of more than a thousand molecular species, with currently 278 compounds already identified. The group of identified compounds includes a number of species which are known markers of metabolic and non-metabolic diseases as it is discussed in details in the Supporting Information.

Statistical data analysis

While the detection and inclusion of alternative biomarkers into the mass spectrometry screening method already provides an incremental improvement in the specificity of the process, the untargeted nature of presented nS-HR-MS method also allows a multivariate statistical data interpretation approach, similar to the method recently described for ESI-MS/MS³¹. Since the removal of a specific enzymatic function of key importance from the metabolic machinery is expected to result in a global change in the metabolome (vide supra), this has to be traceable by pattern recognition approaches.³² Mass spectra can be treated as n-dimensional data vectors where n equals to the number of possible m/z values, i.e. the peak capacity of the spectrum. This value in the mass range of 70-600 m/z is 106,000 at high resolution (0.01 Da bin size) or 1060 at low resolution (1 Da bin size) for the combined positive/negative ion mode data. Individual spectra appear as points in this 106,000 dimensional space, potentially forming groups according to the specific conditions. However, since in most of these dimensions the data cloud has no extension and many of these dimensions are not linearly independent, it is reasonable to reduce the dimensions by appropriate mathematical tools, e.g. principal component analysis (PCA).

Dimension reduction by PCA results in a still orthogonal data space determined by the linearly independent principal components. Three or two dimensional projection of the data in the PCA space is generally used for the demonstration of similarities and dissimilarities. Separation of assigned data groups (i.e. condition specific data groups) can further be improved by linear discriminant analysis (LDA), which is a supervised linear transformation resulting in a non-orthogonal data space where the distances among data groups are maximized. Similarly to PCA, three dimensional projection of LDA is used for the demonstration of dissimilarity among assigned data groups. Besides demonstration of the similarities and dissimilarities, an assigned LDA space can also be used for the identification of an unknown data point by localizing it in the multidimensional data space and determining its distance from the assigned data groups. In the optimal case the unknown falls into or close to one of the clearly separated condition-specific data clouds. Since the peaks of the important biomarker molecules have typically low absolute intensity values, application of normalization is necessary during the data analysis. While a number of multivariate statistical tools suitable for the analysis of this type of data exist, in the present study we employed only consecutive PCA and LDA analyses. Peaks in mass spectra were reduced to the list of the identified metabolite molecules shown in the Table S-3, and peak-specific PCA/LDA was performed on this reduced data set. The influence of the intensive contaminant peaks – especially in positive ion mode – and non-metabolite peaks can be avoided using this approach.

The results of PCA and LDA analysis of the first 60 principal components corresponding to newborn cohort comprising 500 healthy newborns, 21 PKU, 21 MCADD, 4 homocystinuria (HCY) patients and 6 cholestatic liver diseases (CLD) is shown in Figure 3. Figure 3a represents the data obtained from low resolution measurements, Figure 3b shows the high resolution

analysis and Figure 3c demonstrates the peak-specific data analysis method. Each plot reveals complete separation of the healthy population and individual metabolic disorder-specific data groups.

In the model presented here, it is seemingly not critical to use high resolution mass spectrometry since unambiguous identification can be obtained even using data binned to unit resolution. However, when statistically relevant number of healthy samples (in the order of some 10,000s vs. 500 in the current study) and more conditions (all currently screened inborn errors of metabolism including sample mistreatment conditions etc.) are examined, a considerable spreading of the data within individual subject groups is expected. Thus, the compactness of data groups will become more critical and chance of unambiguous identification of the conditions will be improved by high mass resolution.

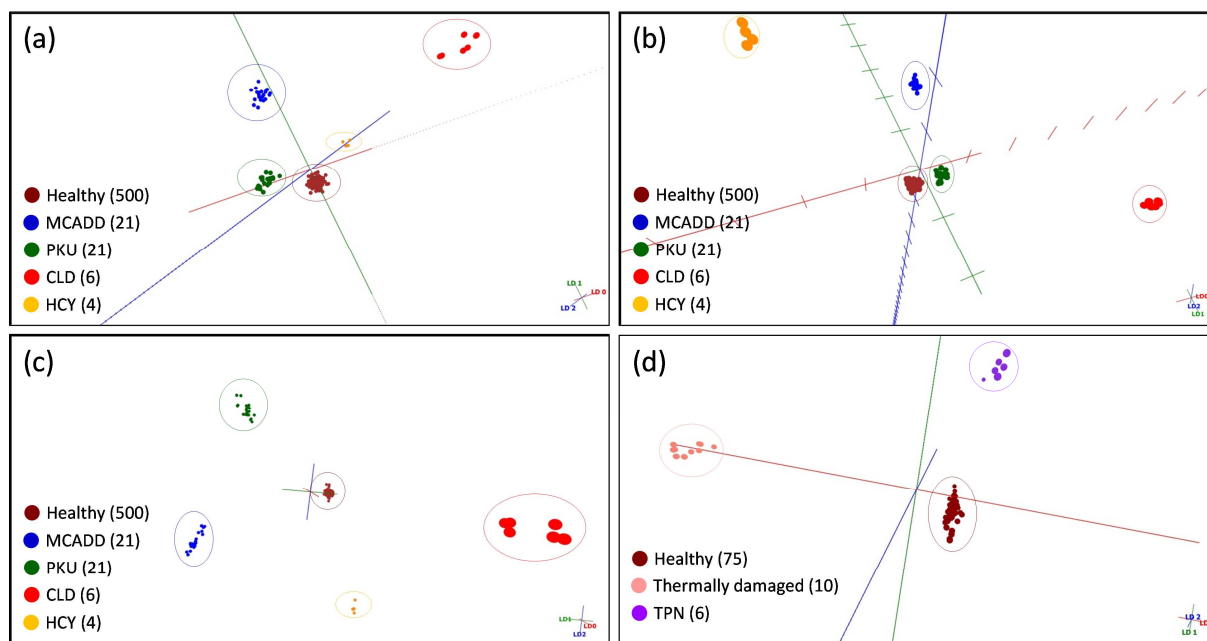


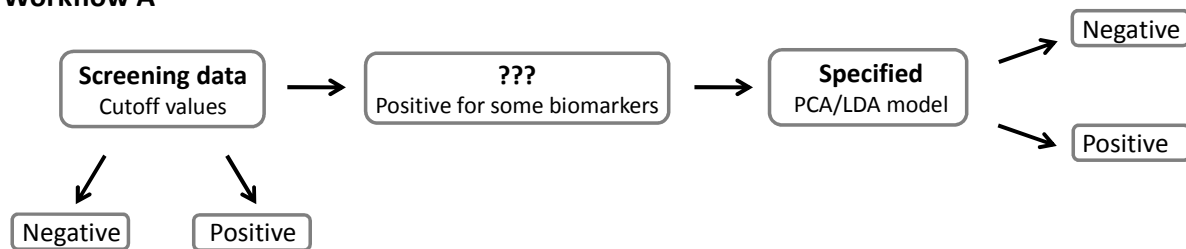
Figure 3. Three dimensional LDA models representing the 60 dimensional PCA/LDA analyses of the sample sets. (a) Model of low resolution (1 Da bin size) measurements, the first 60 components represent 96.52% of the total variance. (b) Model of high resolution (0.01 Da bin size) measurements, the first 60 components represent 98.22% of the total variance. (c) Model for identified peaks, the first 60 components represent 98.81% of the total variance. (d) Model for low resolution measurements on different types of false-positive samples.

1
2
3
4
5
6
7 Sample conditions associated with false-positive cases have also been studied using the
8
9 multivariate statistical approach. Total parenteral nutrition (TPN) cases or thermally damaged
10
11 samples obtained from healthy subjects have shown complete separation, already in the three
12
13 dimensional projection of the 60 dimensional LDA space (Figure 3d). Although single stage
14
15 mass spectrometry is not suitable for identification of associated metabolites, the differences of
16
17 the spectral pattern enable the distinction between disorders and these conditions. Selective
18
19 identification of these conditions is extremely important, since most of the false-positives are
20
21 associated with these or other similar causes.
22
23

24
25 The approach offers two potential workflows depicted in Figure 4. In case of workflow A, the
26
27 screening data is evaluated using cut-off values established for a number of parameters in case of
28
29 each condition. If the results are unambiguous (all markers are either in the healthy or in the
30
31 diseased range) then the result is established. However, if some values are in the healthy range
32
33 while others are in the diseased range, then PCA/LDA analysis is performed using a model
34
35 which comprises all relevant conditions. For instance, if octanoylcarnitine is found to be
36
37 elevated, then a model comprising MCADD, Glutaric acidemia type 2 cases, false-positives,
38
39 babies received valproate, and damaged cards would be used.
40
41
42

43
44 In case of workflow B, the data points corresponding to screened newborns are localized in the
45
46 universal model (containing all true positives, normal healthy newborns and non-metabolic
47
48 conditions) and evaluation is performed in a single step. While workflow B seems to be simpler,
49
50 the large number of assigned data groups may compromise selectivity, which makes the
51
52 utilization of workflow A safer.
53
54
55
56
57
58
59
60

Workflow A



Workflow B

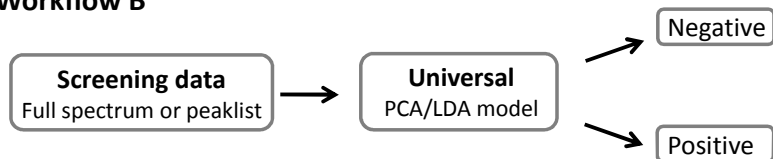


Figure 4. Different workflows of the newborn screening process suggested for the application of the nS-HR-MS approach.

Due to the limited number of positive samples, the system was tested only in case of PKU, MCADD and HCY using both workflows and also the ESI-MS/MS method used for population level screening in western and central Hungary. The cut-off values used in the national screening program are harmonized with those established by the international collaborative study coordinated by Rinaldo et al. at Mayo Clinic.³³ Performance of the three approaches was compared and it was concluded that both workflow A and B operate flawlessly using these limited number of samples and conditions, while the traditional method yields a low, but still significant number of false-positives (~0.65 %). The required analysis time is comparable for each method, but significant growth of overall turnaround time occurs, when second-tier tests are included. The necessity of second-tier tests in the case of the ESI-MS/MS method is well expressed through the reduced false-positive rate (0.09 %), which is expected to reach even lower values in the herein reported method (estimated value: 0.05 %).

Conclusions

The presented nS-HR-MS method offers a promising alternative to traditionally used ESI-MS/MS approach at the population level screening of newborn babies for inborn errors of metabolism. Thereby the chip-based nano-electrospray system offers several advantages compared to an ESI autosampler, most notably the lack of cross-contamination resulting in a significantly reduced number of false-positive results. Besides the detection of amino acids and acylcarnitines, the novel method is also capable of the semi-quantitative determination of highly specific disease markers including various organic acids, acylglycines or even carbohydrates. The untargeted nature of the novel method also allows multivariate pattern recognition analysis of the resulting spectral information, which results in improved specificity at the recognition of

1
2
3
4
5
6 screened conditions, and also the identification of interferences caused by the medical treatment
7
8 of babies or damaged DBS samples. Routine application of the method is expected to decrease
9
10 the number of second samples requested as well as the number of second-tier or confirmatory
11
12 tests (amino acid chromatography, GC-MS profiling of urinary organic acids, succinylacetone,
13
14 etc).
15
16

17
18 In principle, the method has also potential in the screening of galactosemia (monitoring hexose
19
20 and hexose-phosphate), biotinidase deficiency, congenital hypothyroidism, lysosomal storage
21
22 disorders, certain hemoglobinopathies and a number of other diseases involving global changes
23
24 in the metabolome.
25
26

27 The sample set used for this pilot study was a limited collection of healthy and diseased
28
29 neonates, which represents neither the statistical variance of healthy population, nor the
30
31 prevalence of the examined disorders yet. The data presented here already shows significant
32
33 differences between normal and abnormal samples, indicating the suitability of the proposed
34
35 methodology for the neonatal screening of DBS samples at proof-of-principle level. The study is
36
37 being extended for a larger, statistically relevant set of samples and the validation at population-
38
39 level is already in progress.
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Supporting Information

Additional information as noted in text. This material is available free of charge via the Internet at <http://pubs.acs.org>.

Corresponding Author

* Address correspondence to this author at: Biomolecular Medicine, Department of Surgery and Cancer, Imperial College London, South Kensington Campus, Sir Alexander Fleming Building, SW7 2AZ, London, United Kingdom. Phone: +44 20 7594 2760. E-mail: z.takats@imperial.ac.uk.

Acknowledgments

The Exactive instrument was generously provided by Thermo Fisher Scientific for the project and the TriVersa NanoMate ion source was kindly provided by Advion for the high-throughput analysis of the samples. The work was funded by European Research Council under Starting Grant scheme (Contract No: 210356) and Hessisches Ministerium für Wissenschaft und Kunst (LOEWE Schwerpunkt “AmbiProbe”).

References

- (1) Rashed, M. S. *J. Chromatogr. B* **2001**, 758, 27-48.
- (2) Chace, D. H.; Kalas, T. A.; Naylor, E. W. *Clin. Chem.* **2003**, 49, 1797-1817.
- (3) Chace, D. H. *J. Mass Spectrom.* **2009**, 44, 163-170.
- (4) Piraud, M.; Vianey-Saban, C.; Petritis, K.; Elfakir, C.; Steghens, J. P.; Morla, A.; Bouchu, D. *Rapid Commun. Mass Spectrom.* **2003**, 17, 1297-1311.
- (5) De Jesus, V. R.; Chace, D. H.; Lim, T. H.; Mei, J. V.; Hannon, W. H. *Clin. Chim. Acta* **2010**, 411, 684-689.
- (6) Waisbren, S. E.; Albers, S.; Amat, S.; Ampola, M.; Brewster, T. G.; Demmer, L.; Eaton, R. B.; Greenstein, R.; Korson, M.; Larson, C.; Marsden, D.; Msall, M.; Naylor, E. W.; Pueschel, S.; Seashore, M.; Shih, V. E.; Levy, H. L. *JAMA-J. Am. Med. Assoc.* **2003**, 290, 2564-2572.
- (7) Hewlett, J.; Waisbren, S. E. *J. Inherit. Metab. Dis.* **2006**, 29, 677-682.
- (8) Wilson, J. M. G.; Jungner, G. *WHO Chronicle* **1968**, 22, 473.
- (9) Matern, D.; Tortorelli, S.; Oglesbee, D.; Gavrilov, D.; Rinaldo, P. *J. Inherit. Metab. Dis.* **2007**, 30, 585-592.
- (10) Chace, D. H.; Hannon, W. H. *Clin. Chem.* **2010**, 56, 1653-1655.
- (11) Oglesbee, D.; Sanders, K. A.; Lacey, J. M.; Magera, M. J.; Casetta, B.; Strauss, K. A.; Tortorelli, S.; Rinaldo, P.; Matern, D. *Clin. Chem.* **2008**, 54, 542-549.
- (12) Chace, D. H.; Lim, T.; Hansen, C. R.; De Jesus, V. R.; Hannon, W. H. *Clin. Chim. Acta* **2009**, 407, 6-9.
- (13) Schwarz, E.; Liu, A. P.; Randall, H.; Haslip, C.; Keune, F.; Murray, M.; Longo, N.; Pasquali, M. *Pediatr. Res.* **2009**, 66, 230-235.
- (14) Forni, S.; Fu, X. W.; Palmer, S. E.; Sweetman, L. *Mol. Genet. Metab.* **2010**, 101, 25-32.

- (15) Turgeon, C. T.; Magera, M. J.; Cuthbert, C. D.; Loken, P. R.; Gavrilov, D. K.; Tortorelli, S.; Raymond, K. M.; Oglesbee, D.; Rinaldo, P.; Matern, D. *Clin. Chem.* **2010**, *56*, 1686-1695.
- (16) Shigematsu, Y.; Hata, I.; Tajima, G. *J. Inherit. Metab. Dis.* **2010**, *33*, S283-S288.
- (17) Pasikanti, K. K.; Ho, P. C.; Chan, E. C. Y. *J. Chromatogr. B* **2008**, *871*, 202-211.
- (18) Theodoridis, G. A.; Gika, H. G.; Want, E. J.; Wilson, I. D. *Anal. Chim. Acta* **2012**, *711*, 7-16.
- (19) Dettmer, K.; Aronov, P. A.; Hammock, B. D. *Mass Spectrom. Rev.* **2007**, *26*, 51-78.
- (20) Koulman, A.; Tapper, B. A.; Fraser, K.; Cao, M. S.; Lane, G. A.; Rasmussen, S. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 421-428.
- (21) Lokhov, P. G.; Kharybin, O. N.; Archakov, A. I. *Int. J. Mass Spectrom.* **2012**, *309*, 200-205.
- (22) Schuhmann, K.; Almeida, R.; Baumert, M.; Herzog, R.; Bornstein, S. R.; Shevchenko, A. *J. Mass Spectrom.* **2012**, *47*, 96-104.
- (23) Tang, L.; Kebarle, P. *Anal. Chem.* **1993**, *65*, 3654-3668.
- (24) Wilm, M.; Mann, M. *Anal. Chem.* **1996**, *68*, 1-8.
- (25) Chace, D.; Millington, D.; Terada, N.; Kahler, S.; Roe, C.; Hofman, L. *Clin. Chem.* **1993**, *39*, 66-71.
- (26) Kaufman, S. *The Journal of Pediatrics* **1989**, *114*, 895-900.
- (27) Chace, D. H.; Hillman, S. L.; Van Hove, J. L. K.; Naylor, E. W. *Clin. Chem.* **1997**, *43*, 2106-2113.
- (28) Clayton, P. T.; Doig, M.; Ghafari, S.; Meaney, C.; Taylor, C.; Leonard, J. V.; Morris, M.; Johnson, A. W. *Arch. Dis. Child.* **1998**, *79*, 109-115.

- (29) Bennett, M. J.; Coates, P. M.; Hale, D. E.; Millington, D. S.; Pollitt, R. J.; Rinaldo, P.; Roe, C. R.; Tanaka, K. *J. Inherit. Metab. Dis.* **1990**, *13*, 707-715.
- (30) Bonafe, L.; Troxler, H.; Kuster, T.; Heizmann, C. W.; Chamoles, N. A.; Burlina, A. B.; Blau, N. *Mol. Genet. Metab.* **2000**, *69*, 302-311.
- (31) Marquardt, G.; Currier, R.; McHugh, D. M. S.; Gavrilov, D.; Magera, M. J.; Matern, D.; Oglesbee, D.; Raymond, K.; Rinaldo, P.; Smith, E. H., et al. *Genet. Med.* **2012**, *14*, 648-655.
- (32) Nicholson, J. K.; Lindon, J. C.; Holmes, E. *Xenobiotica* **1999**, *29*, 1181-1189.
- (33) McHugh, D. M. S.; Cameron, C. A.; Abdenur, J. E.; Abdulrahman, M.; Adair, O.; Al Nuaimi, S. A.; Ahlman, H.; Allen, J. J.; Antonozzi, I.; Archer, S., et al. *Genet. Med.* **2011**, *13*, 230-254.

For TOC only:

