

OPEN

# NMR-based newborn urine screening for optimized detection of inherited errors of metabolism

Nieves Embade<sup>1</sup>, Claire Cannet<sup>2</sup>, Tammo Diercks<sup>3</sup>, Rubén Gil-Redondo<sup>1</sup>, Chiara Bruzzone<sup>1</sup>, Sara Ansó<sup>4</sup>, Lourdes Román Echevarría<sup>4</sup>, M. Mercedes Martínez Ayucar<sup>5</sup>, Laura Collazos<sup>5</sup>, Blanca Lodoso<sup>5</sup>, Eneritz Guerra<sup>4</sup>, Izaskun Asla Elorriaga<sup>4</sup>, Miguel Ángel Kortajarena<sup>6</sup>, Alberto Pérez Legorburu<sup>7</sup>, Fang Fang<sup>2</sup>, Itziar Astigarraga<sup>4</sup>, Hartmut Schäfer<sup>2</sup>, Manfred Spraul<sup>2</sup> & Oscar Millet<sup>1</sup>

Inborn errors of metabolism (IEMs) are rare diseases produced by the accumulation of abnormal amounts of metabolites, toxic to the newborn. When not detected on time, they can lead to irreversible physiological and psychological sequels or even demise. Metabolomics has emerged as an efficient and powerful tool for IEM detection in newborns, children, and adults with late onset. In here, we screened urine samples from a large set of neonates (470 individuals) from a homogeneous population (Basque Country), for the identification of congenital metabolic diseases using NMR spectroscopy. Absolute quantification allowed to derive a probability function for up to 66 metabolites that adequately describes their normal concentration ranges in newborns from the Basque Country. The absence of another 84 metabolites, considered abnormal, was routinely verified in the healthy newborn population and confirmed for all but 2 samples, of which one showed toxic concentrations of metabolites associated to ketosis and the other one a high trimethylamine concentration that strongly suggested an episode of trimethylaminuria. Thus, a non-invasive and readily accessible urine sample contains enough information to assess the potential existence of a substantial number (>70) of IEMs in newborns, using a single, automated and standardized <sup>1</sup>H- NMR-based analysis.

Inborn errors of metabolism (IEMs) are genetic disorders, usually monogenic disorders, that produce an abnormal accumulation of metabolites due to a malfunctioning protein (i.e. an enzyme or a membrane transporter), involved in the intermediary metabolism. More than 1000 IEMs have been described so far<sup>1</sup>. While each of these disorders may be considered as a rare disease, in total they constitute a significant socio-economic burden with, for instance, an overall incidence of more than 1 in 1.000 affected newborns in Europe. If these diseases remain undetected and untreated, they can lead to irreversible physical and psychological sequels or even death. Thus, IEMs represent a serious public threat that involves intensive, long, and expensive treatments.

While some of these diseases may emerge only in later stages, with chronic and progressive symptoms, most of the neonates with IEMs will develop symptoms typically within hours or days after birth. In this context, an early diagnosis is crucial to adequately prescribe the right therapies to successfully treat these disorders which, in favorable cases, may grant patients a significantly improved quality of life.

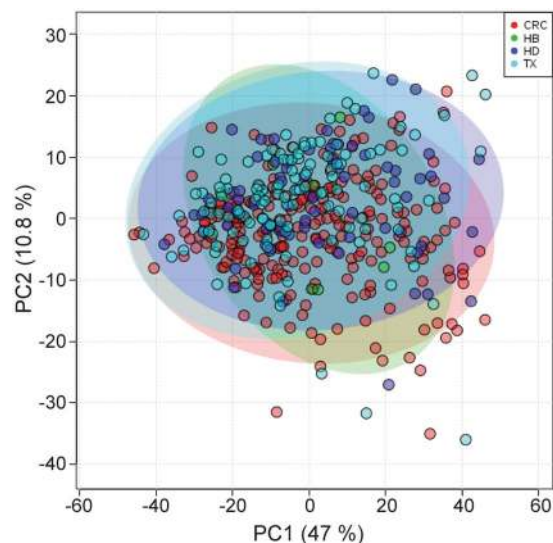
Metabolomics is a powerful tool to study IEMs and, for decades, thousands of neonates have been diagnosed through different newborn screening programs. The standard newborn screening card based on blood spots taken from the heel, is routinely analyzed by conventional mass spectrometry<sup>2</sup>. More sophisticated MS-based techniques such as nanospray ionization with high resolution mass spectrometry (nS-HR-MS)<sup>3</sup> or whole exome sequencing<sup>4,5</sup> are also available, however they are much more time consuming and/or expensive.

<sup>1</sup>Protein Stability and Inherited Disease Laboratory, CIC bioGUNE, Bizkaia Technology Park, Bld. 800, 48160, Derio, Bizkaia, Spain. <sup>2</sup>Bruker Biospin GmbH, Silberstreifen, 76287, Rheinstetten, Germany. <sup>3</sup>NMR Platform, CIC bioGUNE, Bizkaia Technology Park, Bld. 800, 48160, Derio, Bizkaia, Spain. <sup>4</sup>Department of Pediatrics, Cruces University Hospital, Barakaldo, Spain. <sup>5</sup>Department of Pediatrics, Txagorritxu Hospital, Vitoria-Gasteiz, Spain. <sup>6</sup>Department of Nursing, University of the Basque Country, Donostia, Gipuzkoa, Spain. <sup>7</sup>Department of Pediatrics, Basurto University Hospital, Bilbao, Spain. Correspondence and requests for materials should be addressed to O.M. (email: [omillet@cicbiogune.es](mailto:omillet@cicbiogune.es))

Received: 8 May 2019

Accepted: 28 August 2019

Published online: 10 September 2019



**Figure 1.** PLS-DA analysis of urine samples. Two-dimensional PLS-DA score plot for different hospitals. HB: Basurto Hospital, CRC: Cruces Hospital, HD: Donosti Hospital, TX: Txagorritxu Hospital.

NMR spectroscopy is well suited to characterize biofluids as it is a quantifiable, reproducible, non-selective and non-destructive and it is particularly adequate for the characterization of complex solutions (plasma, serum, urine, etc.)<sup>6–8</sup>, although its sensitivity is lower as compared to mass spectrometry. Consequently, NMR has already been applied to neonates with congenital metabolic diseases<sup>9,10</sup>. <sup>1</sup>H-NMR spectroscopy successfully identified neonates carrying different inborn errors including phenylketonuria<sup>11</sup>, maple syrup urine disease<sup>12</sup> and errors of purine and pyrimidine metabolism<sup>13</sup>.

In 2014 a group of Turkish hospitals carried out a <sup>1</sup>H-NMR based clinical study where urine of more than 900 newborns was analyzed to derive the normal concentration ranges for up to 20 normal metabolites and 45 pathological metabolites involved in IEMs<sup>14</sup>. Here, we have implemented and expanded this methodology to conduct a clinical study on urine samples from 470 newborns, obtained from the four public hospitals with neonatology units in the Basque Country. Data analysis allowed to derive the normal metabolic profile of local newborns and to implement an automated screening routine to identify up to 75 congenital metabolic diseases that can be easily extrapolated to other geographical regions.

## Results

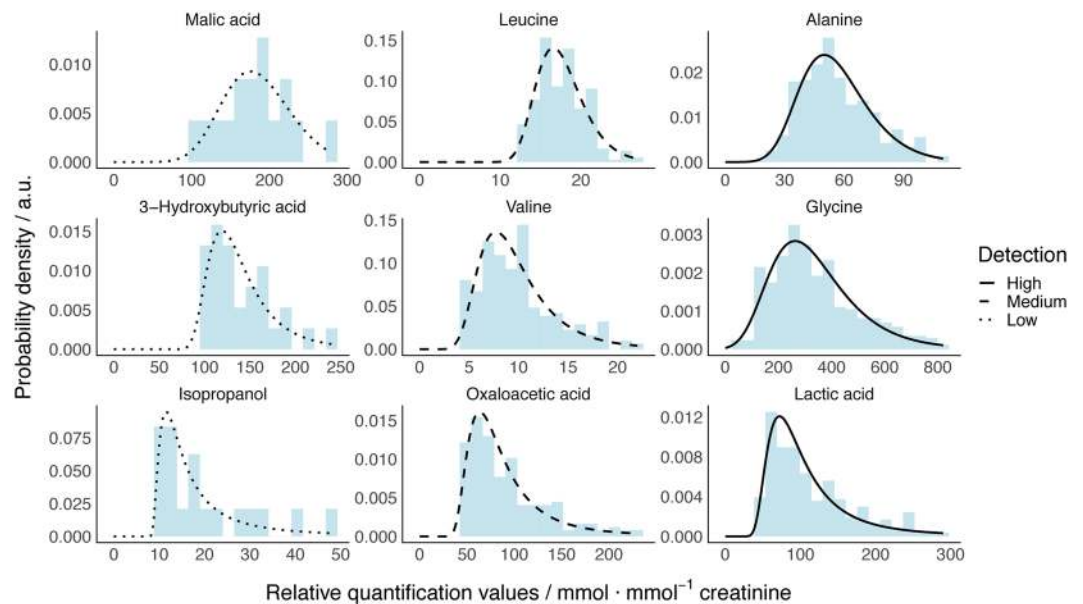
**Description of the data collection.** Complete demographic data was collected for 470 neonates, except for two samples, with no pertaining metadata. After initial quality control, urine samples from 9 newborns were discarded because they showed centrifugation-resistant turbidity, a feature usually associated with bacterial contamination. The final dataset was composed of 461 samples.

More than 50% of the urine samples were collected at 2–3 days age (Table S1) and all samples were obtained from newborns with a birth weight above 2.5 Kg. The dominant ethnic group is Caucasian (>80%) and males and females were equally represented in the study. Gestational age, head circumference and feeding type varied largely (Table S1).

**<sup>1</sup>H-NMR spectra acquisition and multivariate analysis.** Two different <sup>1</sup>H-NMR spectra were collected: a high-resolution 1D <sup>1</sup>H spectrum yielded quantitative metabolite data for statistical analysis and a 2D-*Jres* experiment assisted in peak assignment and metabolite identification. All spectra were collected at 300 K under strict SOPs (see Material and Methods) on either a 600 MHz Bruker AVANCE III (244 samples) or AVANCE IVDr (213 samples) spectrometer, where neither PCA nor PLS analyses revealed any statistical difference between both subsets (Fig. S1).

The 1D <sup>1</sup>H-NMR spectra were then automatically processed and integrated over segments of 0.01 ppm spectral width to obtain bin intensities for statistical analysis. A visual comparison identified six urine samples with poor spectral quality and water suppression that were discarded from further analysis. Using multivariate analysis, 7% of the samples presented multiple regions of the spectra with deviations from normality (Fig. S2) and were no longer considered. This value is lower than previous equivalent studies<sup>11</sup>, and we attribute the discrepancy to the different number of engaged clinical partners and to an improved coordination with the hospitals.

For the remaining set of 437 urine samples, an initial PCA of bin intensities provided a first survey of the metabolomic data and unbiased rough sample clustering. Neither PCA nor PLS revealed any statistically significant differences between sample subsets from the different hospitals (Fig. 1). Likewise, no correlation with other metadata was observed by PLS-DA except for gender and age at sample collection, for which a metabolic fingerprint is observed (Fig. S3) as reported previously<sup>14</sup>. A slight trend for clustering was observed as a function of head circumference (Fig. S3E), consistent with previous results<sup>15</sup>.



**Figure 2.** Example of probability densities from 9 representative GEV models. Densities calculated from models are represented as black lines with the following legend: dotted for low, dashed for intermediate and solid for high percentage of detection for the metabolite. Light blue bars plot experimental real data, where values below LOD were discarded.

**Targeted analysis: metabolite identification and quantification.** Targeted analysis aims at quantifying a given set of metabolites. To that end, the chemical information associated with the bins in the 1D  $^1\text{H-NMR}$  spectrum in combination with the multiplicity information extracted from the 2D J-resolved spectrum was combined to identify up to 150 metabolites in the urine spectrum (Table S2). For each metabolite, the concentration was calculated both, absolute or relative to creatinine. Absolute quantification was referred to the TSP signal in the urine spectra, calculating a signal-intensity per-proton-ratio after the correction for several effects including molecular mass, number of protons, relaxation time, chemical shift, signal multiplicity and coupling constants, line width and Gauss–Lorentz ratio. To that end, signal was fitted using a simplex algorithm also looking for the ranges for signal detection and quantification. Quantification relative to creatinine (in mmol/mol creatinine) used the intensity ratio between a specified metabolite signal and the methyl signal of creatinine.

For a given metabolite, the limit of detection (LOD, Table S2) was determined via spiking experiments from an equivalently measured reference sample database and by simulating spectra after different concentrations of the analyte were added<sup>14</sup>. Averaging over the total cohort allows calculating the probability of obtaining a value above the LOD for the metabolite.

**Probabilistic models for the quantified metabolites.** Of 150 metabolites considered, 58 always remained below their LOD (see Table S2) and where, therefore, never detected. Of the remaining 92 metabolites, 66 had a detection incidence of 2.5% or more, allowing to fit a distribution model. As a metabolite's LOD implies incomplete sampling of the underlying distribution with a natural skew towards higher metabolite concentrations, a generalized extreme value (GEV) distribution model was chosen for its inherent adaptability<sup>16</sup>. GEV is a probability distribution based on extreme value analysis (i. e. extreme deviations from the median of probability distributions) and it seeks to assess the probability of extreme events such as the skewed distributions for the quantified metabolites that considers the LOD values.

Table S2 reports the derived distribution parameters for the 150 metabolites considered and the Q2.5 and Q97.5 quantiles (which delimit the range where 95% of samples are found) obtained from a GEV distribution model (calculated with Eq. 4) or directly from the data histogram. The quantile values obtained by both methods are in excellent agreement except for acetic acid, where the Q97.5 value from the GEV model (332) is larger than that from the data histogram (240) presumably due to the slightly more permissive ranges obtained with the model (Fig. S4).

Figure 2 shows 9 representative examples of metabolite concentration histograms along with modeled probability densities from a GEV distribution that fit well to the experimental data regardless of the detection frequency and the histogram shape (the data for all traceable 66 metabolites with a detection frequency larger than 2.5% is shown in Fig. S4). Thus, the protocol for model building presented here can be applied to derive the normal metabolite concentrations in newborn urine samples for the general population in the Basque Country and to reveal significant deviations as a strong indication for a genetic disorder associated to the pertaining metabolite.

**Diagnostic value of the investigated metabolites.** Out of the 150 metabolites determined in the urine samples (Table S2), 7 derive from contaminations from the matrix used (diaper and the plastic consumables),

19 metabolites are associated with bacterial contamination, likely during the sample collection, 14 metabolites derive from drug catabolism (active principles or excipients) and 12 metabolites are associated to the mother's diet (i. e. caffeine). These metabolites contain little diagnostic nor therapeutic information, but provide metadata information and potential insight on the bacterial gut microbiota of the newborn.

The remaining set of 98 relevant metabolites have been associated with inborn errors of metabolism (IEMs) and, therefore, have potential diagnostic value for the 75 diseases and unhealthy conditions summarized in Table 1. The metabolites are numbered as shown in Table S2. Most commonly, an IEM produces alterations for several metabolites where the “reference metabolites” listed in Table 1 are those required to unambiguously identify an IEM, according to the seminal work of Wevers<sup>17</sup> and others<sup>18</sup>. Yet, as many metabolites can show abnormal concentrations due to different IEMs, their specific range values (if available, e.g., from HMDB, Metagene) are reported in Table S2. These “support metabolites” are also related to a disorder and can, thus, further confirm an IEM or help to differentiate between related IEMs. Finally, for metabolites that occur in the urine of both healthy and IEM affected newborns, the modeled GEV distributions can be used to determine the probability for our <sup>1</sup>H NMR based test to indicate a potential IEM case in the Basque Country ( $F_{global}$  listed in Table 1). Such value corresponds to the frequency probability of finding a newborn with a concentration above the pathogenic threshold value. In cases where more than one metabolite showed up with a diagnostic distribution for a given IEM, the larger value of  $F_{global}$  was used for indication.

**Metabolic disorders detected within the cohort.** In at least two newborn urine samples, we found abnormal concentrations of metabolites related to an alteration in the ketonic bodies. The first sample showed concentrations of 3-hydroxybutyric acid (240 mmol/mol creat), acetone, (210 mmol/mol creat), and acetoacetic acid (62 mmol/mol creat) far above their upper limits derived from the modeled distributions, with probabilities of 0.002, 0.0012, and <0.0001, respectively. Moreover, these metabolite concentrations were also higher than states as normal in metabolomics databases (HMDB, Metagene). The second sample likewise showed abnormally high concentrations for 3-hydroxybutyric acid (170 mmol/mol creat, probability 0.0124) and acetoacetic acid (49 mmol/mol creat, probability 0.0004), which were accompanied by an elevated glucose concentration (560 mmol/mol creat, probability 0.0083). Taken together, these experimental results strongly suggest two cases of ketosis that might be related to a transition state of the newborn.

Finally, a third newborn presented high levels of trimethylamine (24 mmol/mol creatinine; probability 0.0014) suggesting the IEM Trimethylaminuria (FMO3) since the normal concentrations in newborns, as determined in our <sup>1</sup>H NMR study (Fig. 2), are not higher than 6 mmol/mol creat. Yet, this finding should be accompanied by elevated concentrations TMAO, currently unavailable within our method.

## Discussion

In this contribution we have investigated novel strategies for an early identification of congenital metabolopathies from <sup>1</sup>H NMR data from newborn urine samples. Even though IEMs are individually considered as rare disorders, their collective incidence is considerable and represents a serious public health problem. An early diagnosis of these pathologies is crucial to avoid serious acute symptoms and/or to avoid a lifelong treatment of patients. In the context of personalized medicine, even a negative (unsuspicious) result of such IEM tests still produces relevant data that can be integrated in a personal database together with molecular data which is expected to gain diagnostic value in the future.

In the last years had been increasing the number of IEM (around 50) that can be detected using mass spectrometry<sup>19</sup>, even though that the high sensitivity of the technique suggests that this number will increase. Moreover, MS requires the collection of dried blood spots (DBS), which is done by pricking the heel of newborns. It is easy to perform but the process is in a certain way invasive. In NMR spectroscopy, currently 75 IEMs may be detected and this number is also increasing over time. In here, up to 1000 urine metabolites are analyzed simultaneously in a measurement that avoids pricking and takes only minutes. In summary, compared to other techniques such as MS, NMR spectroscopy has great potential for newborn screening, as it is able to sample the metabolic profile and detect both known and unknown metabolites in a quantitative and non-targeted way. Moreover, it is reproducible, and the quantification of urine metabolites is independent on the instrument configuration as long as the SOPs are followed. Finally, the measurement is economically advantageous since the comprehensive metabolite set is already obtained. All these aspects position the technique in a leading place for a potential implementation of the methodology in hospitals and health centers for the screening of large population segments.

Using the metabolomics data provided from the whole cohort of newborns along with statistical analysis, we were able to create a healthy metabolome profile at a high degree of precision. A subset of 66 of the analyzed metabolites were found present in at least 2.5% of all the newborns. Normal ranges for these common metabolites have been established using automatic quantification, as shown in Figs S4–S6, and Table S2, also compared to the normal and pathogenic levels reported in the literature. Some metabolites showing close proximity (or partial overlap) between normal and pathogenic concentrations reflect a heterogenic origin for the metabolite or an incomplete understanding of the relationship between the metabolite and the disease.

IEM incidence strongly depends on geographical and ethnic background of a population<sup>20</sup> and it is instructive to compare our model with the Turkish model previously reported<sup>14</sup>. The metabolites found in our study enclose 18 out of the 20 found in the Turkish work, while myo-Inositol and trimethylamine are found very often only in the Turkey model. We attribute such discrepancy to an improved identification of the two metabolites in our deconvolution algorithm.

Another 84 metabolites that are typically not found at NMR detectable level in healthy newborns were validated for their absence. Finally, we were also able to observe metabolites from different sources of contamination. Propylene glycol was present in about half of the urine samples likely from external manipulation. This aliphatic

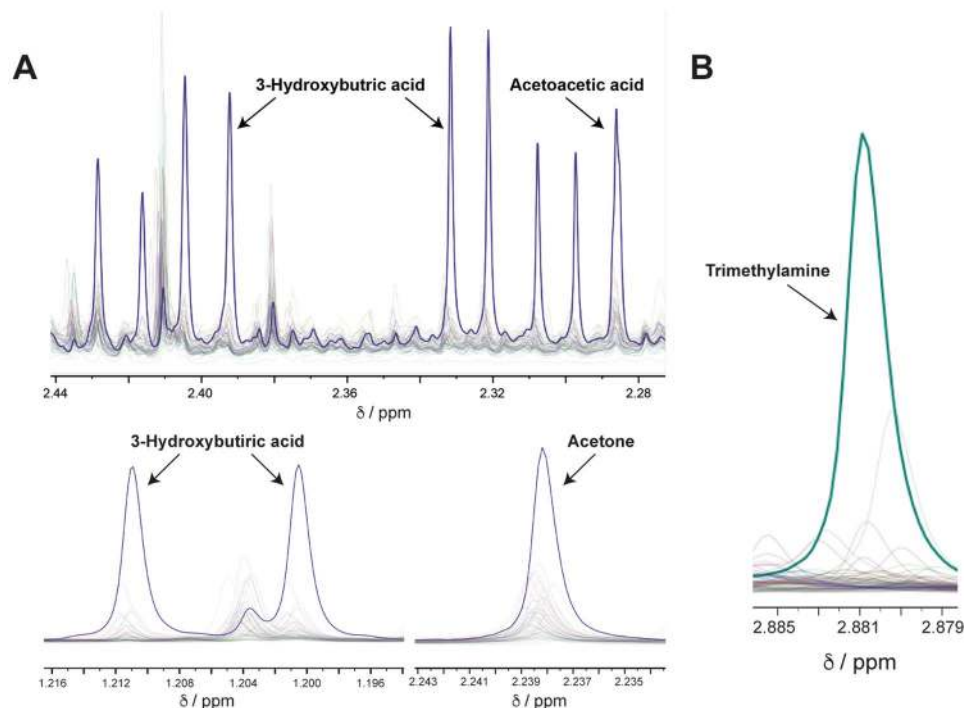
IEM	Reference metabolites	Support Metabolites	$\overline{F}_{global}$
<b>Organic Acidurias</b>			
2-ketoadipic acidemia	80, 105, 135		1.82e-6
3-Hydroxy-3-Methylglutaryl-CoA-lyase Deficiency	17, 20, 25	24, 38, 84	1.74e-04
3-Hydroxyisobutyric aciduria	8, 20, 21, 102		n. a.
3-Methyl-crotonyl-glycinuria	20, 24	13	1.61e-3
3-Methyl-glutaconic acidurias	20, 25, 103	54, 132, 135	1.58e-2
Biotinidase deficiency	20, 24, 102	21	2.90e-2
$\beta$ -Ketothiolase deficiency	25, 143	18, 20, 21, 24, 38	>7.14e-2
Canavan disease	110	135	>7.14e-4
Cobalamin malabsorption	108		4.13e-6
Ethylmalonic encephalopathy	43, 78, 80, 102, 135	12, 114	2.34e-3
Fumaric aciduria	80, 135		7.14e-4
Glutaric aciduria type I	19, 76, 84	18	5.38e-3
Glutaric aciduria type I (low excretor)	19, 76, 84		5.38e-3
Glutaric aciduria type I (non-excretor)	19, 76		n. a.
Glutaric aciduria type II	12, 78, 84, 92	19, 76, 114, 134	2.34e-3
Glutaric aciduria type II (late onset)	20, 78, 84		2.34e-3
Glutaric aciduria type III	84		n.a.
Hyperoxaluria type II	87		>3.09e-1
Isovaleric aciduria	20, 114	12, 18, 86, 108	n. a.
Malonyl-CoA decarboxylase deficiency	78, 108, 135		7.14e-4
Methylmalonic aciduria	16, 75, 91, 129		3.19e-2
Methylmalonate semialdehyde dehydrogenase deficiency	21, 43, 102, 108		2.00e-4
MMA Cbl A deficiency	108	22, 86, 89, 99	4.13e-6
Propionic acidemia	18, 21, 38, 40, 143	3, 22, 86, 89, 99, 128, 129	6.66e-2
Pyroglutamic acidemia	99		n. a.
Transcobalamin II deficiency	108		n. a.
Trimethylaminuria	145		4.25e-3
<b>Amino Acidurias</b>			
Argininemia	117, 147	88, 89	4.67e-3
Argininosuccinic aciduria	47, 117, 147	86, 89	4.37e-3
Cystinuria	46, 58		4.01e-5
Dicarboxylic aminoaciduria	82		n. a.
Dimethylglycine dehydrogenase deficiency	49, 115		5.91e-12
Hartnup disease	43, 83, 98, 103, 149	75, 101	4.39e-4
Hawkinsinuria	32, 99		8.54e-2
Holocarboxylase syntetase deficiency	20, 21, 22		2.07e-1
Homocystinuria	97	49, 107	n.a.
Kyureninase deficiency	74, 150		n. a.
Maple Sirup Urine disease	9, 15, 23	3, 8, 14, 18, 98, 103, 149	8.46e-8
Methionine malabsorption	11		n. a.
Mild Phenylketonuria	116, 124		n. a.
Phenylketonuria	10, 26, 32, 112, 123, 124, 125	37, 65	n. a.
Saccharopinuria	55		n. a.
Sarcosinemia	134		4.56e-1
Tyrosinemia type I	31, 32, 33, 75, 136	35, 113, 147	5.83e-3
Tyrosinemia type II	31, 33	113	3.51e-2
Tyrosinemia type III	31, 32, 33		>7.62e-2
Transient newborn tyrosinemia	31, 33		3.51e-2
Valinemia	149		4.39e-4
<b>Fatty Acid Oxidation Disorders</b>			
Continued			

IEM	Reference metabolites	Support Metabolites	$\overline{F}_{global}$
3-Oxoacid CoA transferase deficiency	18, 38		5.38e-3
Long Chain 3-hydroxyacyl-CoA dehydrogenase deficiency	18		6.66e-2
Short Chain Acyl-CoA dehydrogenase deficiency	12, 78		4.92e-5
Short Chain 3-hydroxyacyl-CoA dehydrogenase deficiency	18		6.66e-2
<b>Urea Cycle Disorders</b>			
Carbamoyl Phosphate synthetase I deficiency	86, 99, 147		4.37e-3
Citrullinemia	55, 117, 147, 148	47, 61, 89	3.47e-1
Neonatal intrahepatic cholestasis	32		n. a.
Ornithine carbamoyltransferase deficiency	117, 147, 148	86, 89, 99	4.67e-3
<b>Creatine Deficiencies</b>			
Creatine transport deficiency	56, 57		n. a.
Guanidinoacetate Methyltransferase deficiency	88		n. a.
<b>Purine/Pyrimidine Disorders</b>			
Dihydropyrimidine Dehydrogenase deficiency	141, 147		4.37e-3
Dihydropyrimidinase deficiency	70, 71	141, 146, 147	4.37e-3
Orotic aciduria	117		n. a.
<b>Carbohydrate Disorders</b>			
Fructose-1,6-bisphosphatase Deficiency	18, 38, 40	85	>7.14e-2
Fanconi-Bickel syndrome	63	61, 81	>3.47e-1
Galactosemia	61, 81	60	3.47e-1
<b>Lysosomal Disorders</b>			
Tay-Sachs disease	99		n. a.
<b>Mitochondrial Disorders</b>			
Dihydrolipoyl dehydrogenase E3	14, 23		2.29e-4
Lactic acidemia	43, 102	21	2.9e-2
Pyruvate carboxylase deficiency	18, 40, 102	43, 135	5.38e-3
<b>Porphyrias</b>			
Acute Intermittent Porphyria	35		1.49e-5
Delta-aminolevulinic acid dehydratase deficiency	35		1.49e-5
<b>Other Disorders</b>			
Asphyxia	23, 102		4.03e-2
Aminoacylase I deficiency	111		n. a.
GABA transaminase deficiency	27, 55		>1.24e-1
Molybdenum cofactor deficiency	139		1.42e-2
Ketosis	18, 38		6.66e-2

**Table 1.** IEMs that can be associated to the NMR-based urine analysis. List of different inborn errors of metabolism that can be identified by NMR. Reference metabolites are required to unambiguously identify a given IEM. Support Metabolites add value to confirm a specific IEM and/or help to discriminate between related IEMs. All numbers are related to the metabolites listed in Table S2.  $\overline{F}_{global}$  describes the probability for the test to identify a potential IEM case in the Basque Country.

alcohol is a chemical included in cosmetics, skin conditioning also found in medication<sup>21</sup>. Moreover, the metabolite acetoin (present in perfumes) was observed in a few spectra from the same hospital (Fig. S2), while unknown metabolites could be detected in a subset of the spectra as well. A future goal is to proceed with the identification of unknown peaks, to determine whether they contain potentially diagnostic information.

Deviations from normality in some of those common metabolites provide very useful information about rare and prevalent metabolic disorders. Up to 75 IEMs may potentially be detected based on the quantification of a single metabolite or the combination of several of them (Table 1). Such disorders can be divided into urea cycle disorders, organic acidurias, purine/pyrimidine disorders, amino acidurias, carbohydrate disorders, creatine deficiencies, fatty acid oxidation disorders and other diseases and syndromes. For example, elevated concentrations of 3-hydroxybutyric acid, acetone and acetoacetic acid revealed a couple of ketosis cases (Fig. 3). Ketosis is a metabolic state where the newborns use fat as fuel in preference to carbohydrates. The body makes ketones from fat, when dietary glucose (from carbohydrates and sugar) is low. Acetoacetate is one of the products of fatty acid oxidation and acetone and hydroxybutyrate are formed from acetoacetate<sup>22</sup>. In one case, the metabolic alteration was accompanied by high levels of glucose, an indicator of Insuline-dependant diabetes Mellitus. In another unrelated sample, a case of trimethylaminuria was also potentially identified. Hence, this approach has proven very useful for the discrimination of pathological samples by identifying abnormal metabolite concentrations or



**Figure 3.** Newborn diseases identify by NMR metabolomics. **(A)** Three different markers (3OH-butyric acid, acetoacetic acid and acetone) in an NMR spectrum of a urine sample from a newborn, show high concentrations as compared to the normal ranges of intensity for other urine samples, suggesting ketosis. **(B)** The metabolite trimethylamine shows also an extremely high concentration in one sample, strongly indicating the presence of trimethylaminuria for that neonate.

patterns after comparison with the model. Unfortunately, the samples were codified and anonymized so clinicians did not have the opportunity to follow the babies that showed these abnormal metabolites.

The number of samples analyzed constitute a limitation in our study: approximately 1:1000 neonates may be affected by a congenital metabolic disorder worldwide. In this context, analyzing less than 500 samples, we were able to detect up to three different metabolopathies, showing the method's potential. Yet, a higher number of samples will be needed to further validate the analytical technique.

In summary, the characterization of a large set of samples prompted us to develop a statistical model for several metabolites applicable to the healthy Basque Country newborn population. The model allows the identification of 75 different IEMs based on the comparative analysis of the absolute concentration of metabolites from a test sample in the context of the normal values obtained by the model which, in correlation with clinical parameters, allows the diagnosis and eventually the monitoring of patients throughout their illness<sup>23</sup>. The reasonable agreement with the number of metabolites and the concentration ranges found in other regions (i. e. Turkey) underline the idea that the model can be easily extrapolated to multiple regions, ultimately constituting a reliable alternative to the heel test for IEMs premature detection.

## Methods

**Study design.** Urine samples were collected from 470 individuals that were born in one of the four public hospitals of the Basque Country with neonatology units: Cruces (Barakaldo, Bizkaia), Txagorritxu (Vitoria, Araba), Basurto (Bilbao, Bizkaia) & Donostia (Donostia, Guipuzkoa). All samples were codified and anonymized to protect the confidentiality of individual participants.

**NMR measurements.** For each sample, a one-dimensional (1D) <sup>1</sup>H-NMR spectrum with water peak suppression<sup>24</sup>, and a 2D *J*res experiment were collected at 300 K, using the Standard Operating Procedures described in the Supplementary Materials and Methods.

**Statistical analysis.** Outliers were properly identified and eliminated from the dataset. Multivariate and univariate statistics was applied to the dataset as explained in the Supplementary Materials and Methods.

**Generalized extreme value distributions.** The GEV distributions can be described by three parameters (location  $\mu$ , scale  $\sigma$ , and shape  $\xi$ ) and provide different levels of information. As usual in probability distributions, the cumulative distribution function (CDF) gives the probability of obtaining a value (concentration in this case) that is lower or equal than a specific threshold  $x$ . The CDF ( $F_{\text{GEV}}$ ) for a GEV distribution can be calculated according to the following expression

$$F_{GEV}(s; \xi) = \begin{cases} \exp(-(1 + \xi s)^{-1/\xi}) & \xi \neq 0 \\ \exp(-\exp(-s)) & \xi = 0 \end{cases} \quad (1)$$

where  $s$  is the scaled value of  $x$  ( $s = (x - \mu)/\sigma$ ). Similarly, it is possible to calculate the probability of obtaining a concentration higher than  $x$  using the complementary CDF ( $\overline{F}_{GEV} = 1 - F_{GEV}$ ). Due to LOD,  $F_{GEV}$  was in fact incomplete because the fraction of samples below LOD had to be considered and the global CDF ( $F_{global}$ ) was calculated with the following expression

$$F_{global} = (1 - \lambda) + \lambda F_{GEV}(s; \xi) \quad (2)$$

where  $\lambda$  is the fraction of samples above LOD. From this expression is straightforward to deduce that

$$\overline{F}_{global} = \lambda \overline{F}_{GEV}(s; \xi) \quad (3)$$

Moreover, since the CDF is invertible it is also possible to calculate specific concentrations associated to quantiles using function  $Q_{GEV}$ , as shown in Eq. 4:

$$Q(p; \mu, \sigma, \xi) = \begin{cases} \mu + \sigma((- \log(p))^{-\xi} - 1)/\xi & \xi > 0 \text{ and } p \in [0, 1]; \xi < 0 \text{ and } p \in (0, 1] \\ \mu - \sigma \log(- \log(p)) & \xi = 0, p \in (0, 1), \end{cases} \quad (4)$$

where  $p$  is the quantile for which is wanted to obtain its associated concentration. It is also affected by incompleteness of CDF due to LOD. For this reason, quantile  $q$  must be adjusted as follows:

$$p' = \frac{p - (1 - \lambda)}{\lambda} \quad (5)$$

where the numerator calculates the amount of quantile  $p$  that is not covered by the fraction below LOD ( $1 - \lambda$ ) and the numerator scales this result in the context of the modeled GEV.

**Ethical approval.** All procedures followed were in accordance with the Helsinki Declaration of 1975 and under the guidance of the Good Clinical Practice (GCP). The study was approved by the Basque Country's Clinical Research Ethics Committee (CEIC) with the title "Metabolomic analysis of newborn urine for the improvement of the diagnosis of congenital pathologies". Informed consent was obtained from parents of all the newborns included in the study. All samples were codified and anonymized to protect the confidentiality of individual participants. The sample collection and the measurement of the urines by NMR was carried out from 2015 to 2017.

## References

- Pampols, T. Inherited metabolic rare disease. *Adv Exp Med Biol* **686**, 397–431, [https://doi.org/10.1007/978-90-481-9485-8\\_23](https://doi.org/10.1007/978-90-481-9485-8_23) (2010).
- Sun, A., Lam, C. & Wong, D. A. Expanded newborn screening for inborn errors of metabolism: overview and outcomes. *Adv Pediatr* **59**, 209–245, <https://doi.org/10.1016/j.yapd.2012.04.013> (2012).
- Denes, J. *et al.* Metabonomics of newborn screening dried blood spot samples: a novel approach in the screening and diagnostics of inborn errors of metabolism. *Analytical chemistry* **84**, 10113–10120, <https://doi.org/10.1021/ac302527m> (2012).
- Yu, J. H., Jamal, S. M., Tabor, H. K. & Bamshad, M. J. Self-guided management of exome and whole-genome sequencing results: changing the results return model. *Genet Med* **15**, 684–690, <https://doi.org/10.1038/gim.2013.35> (2013).
- Yu, H. C. *et al.* An X-linked cobalamin disorder caused by mutations in transcriptional coregulator HCFC1. *Am J Hum Genet* **93**, 506–514, <https://doi.org/10.1016/j.ajhg.2013.07.022> (2013).
- Embade, N. *et al.* Metabolic Characterization of Advanced Liver Fibrosis in HCV Patients as Studied by Serum 1H-NMR Spectroscopy. *PLoS One* **11**, e0155094, <https://doi.org/10.1371/journal.pone.0155094> (2016).
- Fan, T. W. & Lane, A. N. Applications of NMR spectroscopy to systems biochemistry. *Prog Nucl Magn Reson Spectrosc* **92–93**, 18–53, <https://doi.org/10.1016/j.pnmr.2016.01.005> (2016).
- Nagana Gowda, G. A., Rafferty, D. & Can, N. M. R. Solve some significant challenges in metabolomics? *J Magn Reson* **260**, 144–160, <https://doi.org/10.1016/j.jmr.2015.07.014> (2015).
- Engelke, U. F. *et al.* NMR spectroscopy of aminoacylase 1 deficiency, a novel inborn error of metabolism. *NMR Biomed* **21**, 138–147, <https://doi.org/10.1002/nbm.1170> (2008).
- Moolenaar, S. H., Engelke, U. F. & Wevers, R. A. Proton nuclear magnetic resonance spectroscopy of body fluids in the field of inborn errors of metabolism. *Ann Clin Biochem* **40**, 16–24, <https://doi.org/10.1258/000456303321016132> (2003).
- Constantinou, M. A. *et al.* 1H NMR-based metabonomics for the diagnosis of inborn errors of metabolism in urine. *Analytica Chimica Acta* **542**, 169–177 (2005).
- Holmes, E. *et al.* 750 MHz 1H NMR spectroscopy characterisation of the complex metabolic pattern of urine from patients with inborn errors of metabolism: 2-hydroxyglutaric aciduria and maple syrup urine disease. *J Pharm Biomed Anal* **15**, 1647–1659 (1997).
- Wevers, R. A. *et al.* 1H-NMR spectroscopy of body fluids: inborn errors of purine and pyrimidine metabolism. *Clin Chem* **45**, 539–548 (1999).
- Aygen, S. *et al.* NMR-Based Screening for Inborn Errors of Metabolism: Initial Results from a Study on Turkish Neonates. *JIMD Rep* **16**, 101–111, [https://doi.org/10.1007/8904\\_2014\\_326](https://doi.org/10.1007/8904_2014_326) (2014).
- Spraul, M. *et al.* In *Webinar Bruker BioSpin* (2013).
- Coles, S. *An Introduction to Statistical Modeling of Extreme Values.* (Springer-Verlag., 2001).
- Engelke, U. F. *et al.* *Body Fluid NMR Spectroscopy and Brain NMR Spectroscopy.* 3rd Edition edn, (Radboucumc, 2014).
- Gilbert-Barness, E. & Farrell, P. M. Approach to diagnosis of metabolic diseases. *Translational Science of Rare Diseases* **1**, 3–22 (2016).
- Ozben, T. Expanded newborn screening and confirmatory follow-up testing for inborn errors of metabolism detected by tandem mass spectrometry. *Clin Chem Lab Med* **51**, 157–176, <https://doi.org/10.1515/cclm-2012-0472> (2013).



20. Gupta, A. To err is genetic: diagnosis and management of inborn errors of metabolism. *Inherited Errors of Metabolism* **32**, 415–423 (2007).
21. Fiume, M. M. *et al.* Safety assessment of propylene glycol, tripropylene glycol, and PPGs as used in cosmetics. *International journal of toxicology* **31**, 245s–260s, <https://doi.org/10.1177/1091581812461381> (2012).
22. Fritz, I. B. Factors influencing the rates of long-chain fatty acid oxidation and synthesis in mammalian systems. *Physiol Rev* **41**, 52–129, <https://doi.org/10.1152/physrev.1961.41.1.52> (1961).
23. Mussap, M. & Fanos, V. Reducing neonatal mortality and expenditure in the era of health care crisis: is it possible? *J Matern Fetal Neonatal Med* **25**, 1–3, <https://doi.org/10.3109/14767058.2012.715434> (2012).
24. Mackay, R. T. How the 1D-NOESY suppresses solvent signal in metabolomics NMR spectroscopy: an examination of the pulse sequence components and evolution. *Concepts in Magnetic Resonance Part A* **38A**, 197–200 (2011).

## Acknowledgements

The authors thank the collaboration of the Basque Biobank/BioCruces Node [www.biobancovasco.org](http://www.biobancovasco.org) for collecting the samples and data from the neonates included in this study. Support was provided from The Department of Industry, Tourism and Trade of the Government of the Autonomous Community of the Basque Country (Elkartek BG2015 & BG2017); grant from Agencia Estatal de Investigación (Spain) CTQ2015-68756-R and for the Severo Ochoa Excellence Accreditation (SEV-2016-0644).

## Author Contributions

N.E. Design, data collection, analysis and interpretation of data. Drafting the article. T.D. Data collection and processing. C.C., R.G. Data processing, analysis and interpretation. C.B. Data collection. F.F., H.S. and M.S., Review of the article critically for important intellectual content. L.R.E., S.A., M.M.A., L.C, B.L., E.G., I.A.E., M.A.K., A.P.L. and I.A., Selection and acquisition of the samples used in the study. O.M. Conception and design. Critically for important intellectual content. Writing the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-019-49685-x>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019