Molecular Genetics and Metabolism 121 (2017) 83–90

Contents lists available at ScienceDirect

Molecular Genetics and Metabolism

journal homepage: www.elsevier.com/locate/ymgme

Elucidation of the complex metabolic profile of cerebrospinal fluid using an untargeted biochemical profiling assay

Adam D. Kennedy [a], Kirk L. Pappan [a], Taraka R. Donti [b,1], Anne M. Evans [a], Jacob E. Wulff [a], Luke A.D. Miller [a], V. Reid Sutton [b], Qin Sun [b], Marcus J. Miller [b], Sarah H. Elsea [b,*]

[a] Metabolon Inc., Durham, NC, USA
[b] Dept. of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA

ARTICLE INFO

Article history:
Received 11 January 2017
Received in revised form 7 April 2017
Accepted 8 April 2017
Available online 9 April 2017

Keywords:
Phenotype
Biochemical
Metabolomics
Cerebrospinal fluid
Diagnostic
DHPR deficiency

ABSTRACT

We sought to determine the molecular composition of human cerebrospinal fluid (CSF) and identify the biochemical pathways represented in CSF to understand the potential for untargeted screening of inborn errors of metabolism (IEMs). Biochemical profiles for each sample were obtained using an integrated metabolomics workflow comprised of four chromatographic techniques followed by mass spectrometry. Secondarily, we wanted to compare the biochemical profile of CSF with those of plasma and urine within the integrated mass spectrometric-based metabolomic workflow. Three sample types, CSF (N = 30), urine (N = 40) and EDTA plasma (N = 31), were analyzed from retrospectively collected pediatric cohorts of equivalent age and gender characteristics. We identified 435 biochemicals in CSF representing numerous biological and chemical/structural families. Sixty-three percent (273 of 435) of the biochemicals detected in CSF also were detected in urine and plasma, another 32% (140 of 435) were detected in either plasma or urine, and 5% (22 of 435) were detected only in CSF. Analyses of several metabolites showed agreement between clinically useful assays and the metabolomics approach. An additional set of CSF and plasma samples collected from the same patient revealed correlation between several biochemicals detected in paired samples. Finally, analysis of CSF from a pediatric case with dihydropteridine reductase (DHPR) deficiency demonstrated the utility of untargeted global metabolic phenotyping as a broad assessment to screen samples from patients with undifferentiated phenotypes. The results indicate a single CSF sample processed with an integrated metabolomics workflow can be used to identify a large breadth of biochemicals that could be useful for identifying disrupted metabolic patterns associated with IEMs.

1. Introduction

Identifying biochemical signatures of inborn errors of metabolism (IEMs) can be time consuming due to the number of tests sometimes needed to produce conclusive results. The indistinct (non-specific) phenotypes associated with many IEMs, such as developmental delay and hypotonia, present a challenge to diagnosis and may lead to the need for multiple tests and sample types to screen for the gamut of disorders in the differential diagnosis. Diagnosis is further complicated by the intricate network of pathways that dictate biological processes within the tissues and organs in the human body and the number of biochemicals that accumulate due to perturbed biochemical pathways.

Cerebrospinal fluid (CSF) is in direct contact with the tissues of the central nervous system and has been useful for diagnosis and monitoring several diseases, including IEMs. CSF is a rich source of biochemical information that can be utilized for phenotype analysis during clinical diagnosis [1–4]. As the databases of information and screening technologies become more comprehensive, complete phenotypic assessments of individuals can be made to diagnose disease or to follow treatment of disease. For example, targeted biochemical analysis can identify defects of neurotransmitter biosynthesis or metabolism [5,6], as well as other diseases associated with neurological dysfunction [7–14].

Metabolomics offers the ability to detect the substrates, intermediates, and products of metabolism simultaneously and in multiple different biological matrices. Increased understanding of the biochemical composition of different biological matrices collected from healthy and diseased individuals offers the potential for using samples collected by less invasive means (e.g. plasma or urine) in place of CSF. Further, the ability of untargeted metabolomics to identify multiple biochemicals within a pathway can strengthen the confidence of identifying a potential disease in a patient. IEMs are frequently associated with pathogenic variants in genes encoding metabolic enzymes, but pathogenic variants

* Corresponding author at: Department of Molecular and Human Genetics, One Baylor Plaza, NAB2015, Baylor College of Medicine, Houston, TX 77030, USA.
E-mail address: elsea@bcm.edu (S.H. Elsea).
[1] Current address: Greenwood Genetic Center, Greenwood, SC.

http://dx.doi.org/10.1016/j.ymgme.2017.04.005
1096-7192/

can also target regulatory proteins that affect dysfunctional expression or activity of metabolic enzymes. Genomic-level testing, such as clinical whole exome sequencing (WES), is a well-established means to identify IEM disease variants [15–21], but the application of similar broad-based phenotype screening technologies lag behind genomic and genetic testing approaches. Genomic methods like WES identify variants of unknown significance (VUS) with some frequency, and metabolomic analysis has proven to be a useful source of functional data to determine whether the variant is pathogenic [22]. Newborn screening can identify a small subset of IEMs, but many IEMs are not included in newborn screening panels. Children with neurological symptoms that are undiagnosed, especially if they have seizures or other encephalopathy, are likely to undergo lumbar puncture. Thus, understanding the metabolite composition and characteristics of CSF is critical to apply untargeted metabolic phenotyping for the possible diagnosis of IEMs using CSF samples.

We recently demonstrated the utility of metabolomics to identify biochemical signatures of disease in plasma [23] and urine [24] for diverse classes of IEMs. This approach identified and assisted in the diagnosis of aromatic amino acid decarboxylase deficiency through the metabolomic and genomic analyses of plasma [22]. Prior to untargeted metabolomic analysis, the child was subjected to an extensive series of tests including (but not limited to) very long-chain fatty acid profiling, lysosomal storage disorders panel, urine mucopolysaccharide screening, chromosomal microarray, CSF amino acid analysis, urine organic acid profiling, plasma acylcarnitine determination, and serial MRIs – all of which failed to reveal the underlying diagnosis. Compound heterozygous variants of unknown significance were revealed in the *DDC* gene by WES. Follow-up neurotransmitter testing of CSF from a subsequent lumber puncture showed a prominent elevation of the L-DOPA metabolite 3-methoxytyrosine and undetectable levels of homovanillic acid and 5-hydroxyindoleacetic acid that confirmed the homozygous VUS present in the *DDC* gene caused genuine disease and confirmed the diagnosis of AADC deficiency [22]. Subsequent metabolomic profiling of plasma revealed that 3-methyoxytyrosine levels were markedly elevated (Z-score +6.1) and pointed to the utility of untargeted global metabolomic phenotyping to aid in the identification and accelerated diagnosis of IEMs.

In the present study, we sought to expand this approach to CSF through determination of the relative biochemical composition of CSF and comparison of the CSF profile to those of plasma and urine from two similar, but unpaired cohorts of samples. Further, we analyzed additional paired CSF and EDTA plasma samples from the same patient to determine if levels of molecules between the two matrices correlate. Finally, as a proof-of-concept demonstration, we analyzed CSF from a patient with confirmed DHPR deficiency and identified a metabolic signature indicative of disrupted tetrahydrobiopterin metabolism.

2. Methods

2.1. Sample collection

All procedures were in accordance with the ethical standards of the U.S. Department of Health and Human Services and were approved by the Baylor College of Medicine Institutional Review Board. This study was approved with a waiver of informed consent.

Specimens used in metabolomic testing were collected from residual patient samples in the Baylor clinical biochemical genetics laboratory. All samples were stored at −20 °C for 1–9 months prior to metabolomic testing. The average age for the patients was 6.7 years of age (range from 22 days to 20 years of age, median of 4.5 years of age) and 60% female (n = 18). Urine (40 samples) and EDTA plasma (31 samples) samples contained equivalent numbers of age-matched male and female subjects.

2.2. Metabolomic analysis

2.2.1. Sample preparation

Metabolomics was performed as described previously [25,26]. One hundred microliters of sample were utilized for each analysis. Small molecules were extracted in an 80% methanol solution containing recovery standards, outlined below, and used to monitor extraction efficiency [27]. The resulting clarified supernatant extract was divided into five aliquots, one for each of the individual LC/MS analyses, briefly evaporated to remove the organic solvent and stored overnight under nitrogen before preparation for analysis.

2.2.2. LC/MS/MS^n analyses

All methods utilized a Waters ACQUITY ultra-performance liquid chromatography (UPLC) and a Thermo Scientific Q-Exactive high resolution/accurate mass spectrometer interfaced with a heated electrospray ionization (HESI-II) source and Orbitrap mass analyzer operated at 35,000 mass resolution [26]. On the day of analysis, the dried sample extract aliquots were reconstituted in solvents compatible to each of the four methods and one for a spare. Each reconstitution solvent contained a series of standards (isotopically labeled compounds) at fixed concentrations to monitor injection and chromatographic consistency and to align chromatograms during data processing. Separate aliquots were separated by two reverse phase positive ion methods, one reverse phase negative ion method, and one hydrophilic interaction liquid chromatographic method [26]. All of the methods alternated between full scan MS and data-dependent MS^n scans using dynamic exclusion. The scan range varied slightly between methods but generally covered 70–1,000 *m/z*. Raw data files were archived and data extracted as described below.

Metabolites were identified by matching the ion chromatographic retention index, accurate mass, and mass spectral fragmentation signatures with a reference library consisting of over 4,000 entries created from authentic standard metabolites under the identical analytical procedure as the experimental samples [25]. With the exception of compounds marked with an asterisk after their name (Supplemental Tables 1–4), identification of compounds was based on the match of its retention time, parent ion accurate mass, and MS/MS fragmentation spectrum to an authentic standard. This represents Tier 1 identification (highest level) as defined by the Sumner publication from the Metabolomics Standards Initiative (MSI) [28]. Compounds whose name is marked by an asterisk also were identified based on parent ion accurate mass and MS/MS fragmentation mass spectral data but no reference standard currently exists to make a library entry (Tier 2).

CSF, EDTA plasma, and urine were run as independent sample sets. Small aliquots of each of the clinical samples were pooled and run as 6 technical replicates at randomly spaced intervals throughout the entire sample set to monitor process variability and quality control for the performance of each batch. The median relative standard deviation was calculated for all spiked standards and endogenous biochemicals using median scaled values. The median relative standard deviations (RSDs) for the internal standards and endogenous biochemicals for each of the samples matrices were as follows: 4% and 11% for CSF, 6% and 6% for EDTA plasma, and 4% and 7% for urine, respectively.

2.2.3. Data analysis and statistics

Raw ion intensity values from mass spectrometry analysis were median scaled followed by imputation of any missing values with a value based on the minimum detected value, and, finally, were natural log-transformed on a per biochemical basis. Imputation was based on a random uniform variable with a range between 0.99 and 1.00 times the observed minimum. Z-scores were calculated by comparing the log transformed median scaled biochemical values to the associated mean and standard deviation, for a given biochemical, found in the reference population for the respective biological matrix.