Molecular Genetics and Metabolism 119 (2016) 300–306

Contents lists available at ScienceDirect

Molecular Genetics and Metabolism

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

An MRspec database query and visualization engine with applications as a clinical diagnostic and research tool

Filip Miscevic [a,b], Justin Foong [a], Benjamin Schmitt [c,d,e], Susan Blaser [f], Michael Brudno [a,b,e,1], Andreas Schulze [c,e,g,*,1]

[a] Centre for Computational Medicine, Peter Gilgan Center for Research and Learning, The Hospital for Sick Children, Toronto, Canada
[b] Department of Computer Science, University of Toronto, Toronto, Canada
[c] Department of Paediatrics, University of Toronto, Toronto, Canada
[d] Siemens Healthcare, Sydney, Australia
[e] Genetics and Genome Biology, Peter Gilgan Center for Research and Learning, The Hospital for Sick Children, Toronto, Canada
[f] Department of Diagnostic Imaging, The Hospital for Sick Children, Toronto, Canada
[g] Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, Canada

ARTICLE INFO

Article history:
Received 28 September 2016
Received in revised form 7 November 2016
Accepted 8 November 2016
Available online 9 November 2016

Keywords:
MRSpec
Magnetic resonance spectroscopy
Inborn errors of metabolism
Clinical dataset
Computational analysis
Neurochemical profiling

ABSTRACT

Purpose: Proton magnetic resonance spectroscopy (MRspec), one of the very few techniques for in vivo assessment of neuro-metabolic profiles, is often complicated by lack of standard population norms and paucity of computational tools.
Methods: 7035 scans and clinical information from 4430 pediatric patients were collected from 2008 to 2014. Scans were conducted using a 1.5 T ($n = 3664$) or 3 T scanner ($n = 3371$), and with either a long (144 ms, $n = 5559$) or short echo time (35 ms, $n = 1476$). 3055 of these scans were localized in the basal ganglia (BG), 1211 in parieto-occipital white matter (WM). 34 metabolites were quantified using LCModel. A web application using MySQL, Python and Flask was developed to facilitate the exploration of the data set.
Results: Already piloting the application revealed numerous insights. (1), *N*-acetylaspartate (NAA) increased throughout all ages. During early infancy, total choline was highly varied and *myo*-inositol demonstrated a downward trend. (2), Total creatine (tCr) and creatine increased throughout childhood and adolescence, though phosphocreatine (PCr) remained constant beyond 200 days. (3), tCr was higher in BG than WM. (4), No obvious gender-related differences were observed. (5), Field strength affects quantification using LCModel for some metabolites, most prominently for tCr and total NAA. (6), Outlier analysis identified patients treated with vigabatrin through elevated γ-aminobutyrate, and patients with Klippel-Feil syndrome, Leigh disease and L2-hydroxyglutaric aciduria through low choline in BG.
Conclusions: We have established the largest MRSpec database and developed a robust and flexible computational tool for facilitating the exploration of vast metabolite datasets that proved its value for discovering neurochemical trends for clinical diagnosis, treatment monitoring, and research. Open access will lead to its widespread use, improving the diagnostic yield and contributing to better understanding of metabolic processes and conditions in the brain.

1. Introduction

Proton magnetic resonance spectroscopy, also known as MRSpec or ^{1}H-MRS, is increasingly used in conjunction with diagnostic brain magnetic resonance imaging (MRI) to obtain neurometabolic profiles. Changes in these profiles are clinically significant, as they are associated with a variety of morbidities, such as inborn errors of metabolism [1], tumorigenesis [2,3] and neurodegenerative diseases [4]. Furthermore, MRSpec is non-invasive, rapid, and easily integrated as part of an MRI workflow. Here we describe findings from an exploratory analysis of a comprehensive database of 7035 MRSpec brain scans, as well as the computational tool we have developed to enable the exploration of MRSpec databases such as this one.

While a rich source of information, analyzing quantitative MRSpec data is not without its challenges. As our data have been obtained in a clinical setting, standard procedures (though existing) were not always applied, there was no control cohort, and results may be confounded by the presence of underlying conditions at the time of data acquisition. Furthermore, there are many partitions of the data with possibly different trends, such as age, sex, or scan localization.

* Corresponding author at: The Hospital for Sick Children, Peter Gilgan Centre for Research and Learning, 686 Bay Street, Toronto ON M5G 0A4, Canada.
E-mail address: andreas.schulze@sickkids.ca (A. Schulze).
[1] Authors contributed equally to the manuscript.

http://dx.doi.org/10.1016/j.ymgme.2016.11.003
1096-7192/

We sought to address these issues by developing a computational tool for exploring MRSpec data for clinical and research purposes. The tool enables the rapid investigation and validation of trends or research hypotheses concerning patient populations captured in this database. By facilitating the visualization of trends and individual patient data against the whole, the tool is also a promising clinical and diagnostic aide.

2. Methods

2.1. Obtaining MRSpec data

Patients were examined using either a 1.5 T or a 3 T Philips Achieva clinical MRI system (Philips Healthcare, Andover, MA). Point-resolved spectroscopy (PRESS) acquisitions (NEX = 128, repetition time (TR) = 2000 ms) were obtained at two different echo times (TE) (short TE = 35 ms, long TE = 144 ms). Two different locations in the left hemisphere were used: basal ganglia (BG), followed by parieto-occipital white matter (OCC WM). Choice of volume was according to size of the brain area with focus on tissue homogeneity (minimal edge length of a voxel is 1 cm, minimal voxel volume of 1 cm^3). Depending on the clinical indication either a short, 12-min protocol consisting of two scans was used (BG at 144 ms followed by OCC WM at 35 ms), or a 'metabolic', 24-min protocol with four scans (BG at 144 and 35 ms followed by OCC WM at 144 and 35 ms). Raw spectrogram data was quantified using LCModel Version 6.2-1L (LCMODEL Inc., Oakville, ON) [5] without partial volume correction or signal normalization against a standard of reference, i.e. in relative terms using LCModel's inbuilt quantitation routine. Metabolite signals were normalized against the unsuppressed bulk water signal intensity.

For measurements taken with the 1.5 T scanner, the metabolites available for analysis were guanidinoacetate (Gua),[2] total *N*-acetylaspartate (tNAA), aceto-acetate (AcAc), lactate (Lac), glutamine (Gln), taurine (Tau), glucose (Glc), scyllo-inositol (Scyllo), acetone (Acn), total choline (tCho), myo-inositol (Ins), glutamate (Glu), aspartate (Asp), L-alanine (Ala), *N*-acetylaspartylglutamate (NAAG), total creatine (tCr), γ-aminobutyrate (GABA) as well as the lipid resonances (Lip) at 2.0 ppm (Lip20), two species at 1.3 ppm (Lip13a/b), one at 0.9 ppm (Lip09). Macromolecule resonances (MM), MM09, MM12, MM14, MM17 and MM20, were also recorded [5].

For measurements taken with the 3 T scanner, metabolites analyzed included those above as well as glycerophosphocholine (GPC), phosphocholine (PCh), phosphocreatine (PCr), and creatine (Cr).

tNAA, tCho, tCr and Glx were calculated, where applicable, by taking the sum of PCr and Cr for tCr; NAA and NAAG for tNAA; Cho, GPC and PCh for tCho; and Gln and Glu for Glx.

The 7035 scans of 4430 unique patients were dated February 2008 to October 2014. 3664 scans were conducted using 1.5 T scanner, while 3371 scans were conducted using 3 T. 5559 of the scans were conducted using a 144 ms echo time; 1476 scans were conducted using a short echo time. 3055 of these scans were localized in the basal ganglia (BG), 1211 in parieto-occipital white matter (OCC WM), 55 in the cerebellum and 26 in fronto-temporal white matter (FT WM). Location data was unavailable for 958 of the scans. Research ethics board approval was obtained for retrospective chart review to record patient histories, medication, and diagnoses, which was transcribed manually and deidentified prior to being added to the database. For 6555 of these scans, the indication for the scan and/or a working or final diagnosis was available.

Mean patient age was 4.6 years, (0–18 years, median 3.3 years). The male-to-female ratio was 1.4:1. 4931 scans were known to have been conducted under some form of sedation, with 4468 of these conducted under general anesthesia. Average age of anesthetization was 4.4 years (median 3.2 years).

[2] More precisely, it is a simulated singlet to account for an occasional significant signal at 3.78 ppm.

2.2. The MRSpec database query and visualization engine

Patient health information was deidentified, assigned a study ID, and imported into a MySQL 5.6 database. The database query and visualization tool was implemented in Python 2.7 and a lightweight Flask 0.10 web application. The full source code is available at https://github.com/compbio-UofT/mrspec.

2.3. Data preprocessing

Because the data are noisy, we excluded certain results from analysis with user-definable thresholds for the confidence of measurement determined by LCModel when fitting data to the raw spectrograph curve (see [5]). The thresholds used are available on the project page at https://github.com/compbio-UofT/mrspec/blob/master/config/metabolite_thresholds.txt.

We assumed that although each patient underwent ^{1}H-MRS for a particular indication, the severity of the condition and the likelihood that it would systematically affect metabolite measurements was distributed normally. This would mean that while the data would potentially have greater variance, it would still be most dense around the population norm, and so it would be analytically useful to compare individual patients to overall trends in the data.

2.4. Age-adjusted standard deviation score

Given that metabolite concentrations can vary with age, we found it useful to compare a given data point with a moving 'average' of an age-matched cohort rather than the population average. For every metabolite for each patient, we calculated the expected metabolite value at that age using a linear regression of the 50 closest patients by age on either side of the current patient being considered, for a total of 100 patients. The standard deviation of this patient's metabolite with respect to the regression was then calculated, hereafter referred to as the standard deviation score. Since standard deviation is a unitless measure that indicates where the measurement falls in the spread of the data, it is appropriate for comparing differences in concentration between metabolites.

2.5. Examining trends using linear regression

Partitions of the database were examined for trends using the multiple linear regression model

$$y = \beta_0 + \beta_1 x + \beta_2 I x + \beta_3 I$$

where y is metabolite concentration, x is age, and I is a binary indicator variable for whether a particular measurement belongs to one partition or the other (i.e., 1 for male, 0 for female). The statistical significance of coefficients of regression β_2 and β_3 was examined to determine whether this partition revealed trends explainable above and beyond chance, using a Bonferroni correction at an $\frac{\alpha}{m}$ level of significance, where $\alpha = 0.05$ and m refers to the number of hypotheses that were tested (i.e., number of partitions that were examined). Since 3 partitions were examined, trends were only considered significant if $p < 0.0167$. Refer to the tables in the Supplementary material for the p values of these coefficients for all metabolites.

Certain metabolites followed a clear sublinear trend, for which we used a zero-shifted logarithm of age, $\log(1 + x)$, in place of the age if the adjusted R^2 was better. Residual plots were examined in determining the appropriateness of the log-linear model over the linear model.

In comparing trends that were not different partitions of the same data (i.e. trends between different metabolites), we calculated the Z