Infer metabolic directions and magnitudes from moment differences of mass-weighted intensity distributions

Johon Tuobang Li (李拓邦)[1,2,3]
[1]Guangdong Technion-Israel Institute of Technology,[2]Technion-Israel Institute of Technology,[3]University of California, Berkeley

Abstract

Metabolic pathways are fundamental maps in biochemistry that detail how molecules are transformed through various reactions. Metabolomics refers to the large-scale study of small molecules. High-throughput, untargeted, mass spectrometry-based metabolomics experiments typically depend on libraries for structural annotation, which is necessary for pathway analysis. However, only a small fraction of spectra can be matched to known structures in these libraries and only a portion of annotated metabolites can be associated with specific pathways, considering that numerous pathways are yet to be discovered. The complexity of metabolic pathways, where a single compound can play a part in multiple pathways, poses an additional challenge. This study introduces a different concept: mass-weighted intensity distribution, which is the empirical distribution of the intensities times their associated m/z values. Analysis of COVID-19 and mouse brain datasets shows that by estimating the differences of the point estimations of these distributions, it becomes possible to infer the metabolic directions and magnitudes without requiring knowledge of the exact chemical structures of these compounds and their related pathways. This method has the potential to bypass the current bottleneck and provide fresh insights into metabolomics studies.

Introduction

Metabolic pathways consist of enzyme-mediated biochemical reactions that are commonly categorized into two main processes within a living organism: biosynthesis (known as anabolism) and breakdown (known as catabolism) of molecules. Many metabolic pathways are still undiscovered or poorly understood. High-throughput mass spectrometry experiments can collect thousands of mass spectra in just minutes, giving mass spectrometry a unique advantage compared to other analytical methods. The fragmentation pattern of a molecule, or the mass spectrum, can provide valuable structural information about the molecule. However, annotation of these spectra is typically restricted to compounds for which reference spectra are present in libraries or databases (1,2,3). Only a small fraction of spectra can be accurately assigned precise chemical structures in nontargeted tandem mass spectrometry studies, a prerequisite for pathway analysis (4). Another challenge arises from the complexity of metabolic pathways, where one compound can be part of several pathways. The change in the amount of certain compounds cannot conclusively determine the metabolic direction of a specific pathway. For example, glucose can be catabolized through glycolysis to produce ATP, or it can be stored as glycogen, or converted to fat. Therefore, a decrease in glucose levels could be due to increased glycolysis, glycogen synthesis, or fat synthesis.

Methods and Materials

The data generated from mass spectroscopy experiments usually consist of two main components: the mass-to-charge ratio (m/z) and its corresponding intensity. The m/z value represents the mass of the ion (when the charge is +1), while the intensity is a measure of the relative abundance of ions present at that specific m/z value in the mass spectrum. Let $C_{1,n}$ represent the first column, which includes the m/z data, and $C_{2,n}$ represent the second column, which includes the corresponding intensity. The mass-weighted intensity distribution of sample A of n molecules of interest is defined as the empirical distribution of $C_{1,n,A}C_{2,n,A}$. The location estimate of $C_{1,n,A}C_{2,n,A}$ is denoted as $L_{1,n,A}$. As the mass-weighted intensity distribution represents the concentrations of molecules of interest in the sample, weighted by their respective masses, in the same study, if sample B contains more low-weight molecules compared to sample A, it is considered that sample B exhibits a catabolic direction compared to sample A with regards to n molecules of interest, the location estimate $L_{n,B}$ is expected to decrease, i.e., $L_{n,A} > L_{n,B}$. Conversely, sample A exhibits an anabolic direction compared to sample B. This provides a mathematical definition for two classic metabolic directions. The absolute difference of $L_{n,A}$ and $L_{n,B}$ is the magnitude of this change. This magnitude can be further standardized by dividing it by $\frac{1}{2}(L_{n,A} + L_{n,B})$. Combing this magnitude with the direction, it is called a metabolic momentum of sample A and B of n molecules of interest with regards to location. Then, further consider a scale estimate of $C_{1,n,A}C_{2,n,A}$, denoted as $S_{n,A}$. If $S_{n,A} > S_{n,B}$, i.e., there is a significant decrease in the scale estimates, the metabolic direction of sample B is considered centrabolic compared to sample A for n molecules of interest. Conversely, sample A is considered duobolic compared to sample B for n molecules of interest. This mathematical approach reveals two new metabolic directions, which have clear biological significance. If the metabolic direction of a sample of n molecules of interest is centrabolic compared to that of another sample of the same n molecules of interest, it indicates that for low molecular weight compounds, the related pathways are generally anabolic, while for high molecular weight compounds, the related pathways are generally catabolic. $|S_{n,A} - S_{n,B}|$ is the magnitude of this change, which can be further standardized by dividing it by $\frac{1}{2}(S_{n,A} - S_{n,B})$. Combing this magnitude with the direction, it is called a metabolic momentum of sample A and B of n molecules of interest with regards to scale.
Due to the extreme heterogeneity of mass spectra data, robust statistics are recommended. In this brief report, Hodges-Lehmann estimator (H-L) (5) and median standard deviation (msd) (6) are used as the location and scale estimators. The overall picture of metabolic momentums of different classes is named as momentome (Table 1).

Results

The study by Yang et al. compares the plasma metabolome of ordinary convalescent patients with antibodies (CA), convalescents with rapidly faded antibodies (CO), and healthy subjects (H) (7). For both CA and CO, purine-related metabolism significantly towards anabolism and duobolism compared to the healthy volunteers (Table 1), aligned with a previous study that showed purine metabolism is significantly up-regulated after SARS-CoV-2 infection (8). Acylcarnitine-related pathways exhibit a significant inclination towards catabolism and centrabolism (Table 1). This conclusion, which does not require knowledge of individual compounds within the acylcarnitine class, was also emphasized by Yang et al. (7). It was observed that long-chain acylcarnitines were generally lower in both convalescent groups, while medium-chain acylcarnitines displayed the opposite pattern (7). Bile acid-related pathways leaned towards anabolism and duobolism in CA group, while bile acids have been reported to be immunomodulatory (9,10). For both CA and CO, metabolism related to carbohydrates significantly shifts towards anabolism and duabolism compared to that of healthy volunteers. This might be due to the dysregulated glucose metabolism (11,12). Because the mass-weighted intensity distribution is the product of the concentration of the molecules and their mass, if the mass shrinks to half during a reaction but the concentration doubles, the location of the mass-weighted intensity distribution should generally remain the same. In addition, many intermediates in the glycolytic pathway have higher molecular weights than glucose, e.g., glucose-6-phosphate. Therefore, the breakdown of glucose into two molecules of pyruvate theoretically should increase the location of the mass-weighted intensity distribution. This is a limitation of metabolic momentums as they can only accurately reflect the directions of chemical reactions that have two or more distinct major compounds as substrates or products. If one compound and its related products dominant the metabolites of interest and have such issues, the metabolic vectors might not be accurate.

Table 1. Estimates of mass weighted intensity distributions of two compound groups in Yang et al.'s UHPLC-MS dataset.

	Group	H-L	msd
Purines	H	633.40	355.87
Purines	CO	1430.75	1035.89
Purines	CA	996.7	678.51
Acylcarnitine	H	114.84	77.80
Acylcarnitine	CO	80.53	50.96
Acylcarnitine	CA	93.45	61.75

Contact

Johon Tuobang Li (李拓邦)
Guangdong Technion-Israel Institute of Technology
Email: tuobangli@berkeley.edu
Website: https://www.researchgate.net/profile/Tuobang-Li-2
Phone: +86 13059397095

References

1. M Wang, et al., Mass spectrometry searches using masst. Nat. Biotechnol. 38, 23–26 (2020).
2. T Kind, et al., Fiehnlib: Mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. Anal. Chem. 81, 10038–10048 (2009).
3. K Dührkop, H Shen, M Meusel, J Rousu, S Böcker, Searching molecular structure databases with tandem mass spectra using csi:fingerid. Proc. Natl. Acad. Sci. United States Am. 112, 12580–12585 (2015).
4. K Dührkop, et al., Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra. Nat. Biotechnol. 39, 462–471 (2020).
5. J Hodges Jr, E Lehmann, Estimates of location based on rank tests. The Annals Math. Stat. 34, 598–611 (1963).
6. PJ Bickel, EL Lehmann, Descriptive statistics for nonparametric models iv. spread in Selected Works of EL Lehmann. (Springer), pp. 519–526 (2012).
7. Z Yang, et al., Plasma metabolome and cytokine profile reveal glycylproline modulating antibody fading in convalescent covid-19 patients. Proc. Natl. Acad. Sci. 119, e2117089119 (2022).
8. N Xiao, et al., Integrated cytokine and metabolite analysis reveals immunometabolic reprogramming in covid-19 patients with therapeutic implications. Nat. Commun. 12 (2021).
9. F Kong, LJ Saif, Q Wang, Roles of bile acids in enteric virus replication. Animal Dis. 1 (2021).
10. A Visekruna, M Luu, The role of short-chain fatty acids and bile acids in intestinal and liver function, inflammation, and carcinogenesis. Front. Cell Dev. Biol. 9 (2021).
11. AC Codo, et al., Elevated glucose levels favor sars-cov-2 infection and monocyte response through a hif-1alphaglycolysis-dependent axis. Cell Metab. 32, 437–446.e5 (2020).
12. K Khunti, et al., Covid-19, hyperglycemia, and new-onset diabetes. Diabetes Care 44, 2645–2655 (2021).