Matterworks™

Dramatic Efficiency Gains in Untargeted Absolute Quantification of Cellular Metabolites with Advanced Machine Learning

Ana S. H. Costa, Craig Knisley, Devesh Shah, Timothy Kassis, Mimoun Cadosch Delmar, Jennifer M. Campbell, & Jack Geremia. Matterworks, Inc., Somerville, MA, USA.

Introduction

The quantitative profiling of metabolites (i.e., metabolomics) represents the result of genetics, environment, and metabolism and thus provides a valuable biological readout of an organism.[1] Mass spectrometry (MS) has emerged as the method of choice for high-content metabolomics, given its ability to rapidly assess diverse small molecule chemistries over a wide dynamic range.[2] Depending on the goal of a metabolomics study, investigators typically choose between a "targeted" approach to achieve absolute quantitation of a relatively short list of known compounds and an "untargeted" approach to characterize as many known and unknown biochemicals with relative abundance (i.e., fold-change differences among study groups).[3]

Traditional MS methods for targeted metabolomics are laborious, costly, and time-consuming. Isotopically labeled pure standards must be purchased or synthesized for each metabolite under investigation. De novo synthesis significantly ratchets the costs and time to a cost often prohibitive for research labs. Following pure standard procurement, subsequent calibration curves must be generated, requiring the dedication of staff trained in analytical chemistry.

Furthermore, researchers are limited to investigating the biochemical space included in the targeted list of metabolites, prohibiting hypothesis-generating study design and the opportunity for novel discovery. To overcome the limitations of targeted MS-based metabolomics, we developed Pyxis, which eliminates the need for stable isotope-labeled standards, calibration curve preparation, and traditional method development. Pyxis comprises a rapid machine-learning model that uses the signals from a small number of matrix-independent universal calibrators known as StandardCandles™. Data are analyzed by a standardized LC-MS method and processed through cloud-based software to annotate metabolite identities and absolute concentrations directly from the raw MS data.

In this study, we benchmark Pyxis' ability to rapidly analyze several sample matrices and present a case study identifying relevant biomarkers in human biofluids.

	HUMAN BIOFLUIDS						GROWTH MEDIA		CELLS
Sample Matrix	AF	CSF	DBS	NIST 1950	Saliva	Urine	HPLM	CD DH44	CHO Cells (x10^6 cells/mL)
Sample-to-solvent dilutions	5x 15x 30x	5x 15x 30x	NA	5x 15x 30x	5x 15x 30x	5x 15x 30x	5x 15x 30x	5x 15x 30x	0.5 1 2.5 5 10

Table 1. Sample matrices and sample-to-solvent ratio dilutions used to generate analyte concentration ranges. AF=Amniotic Fluid; CSF= Cerebrospinal Fluid; DBS= Dried Blood Spots; NIST 1950= Human plasma (NIST SRM 1950); Saliva= Human Saliva; Urine= Human Urine; CHO Cells= Chinese Hamster Ovary Cells; HPLM= Human plasma-like medium; CD DH44= Cell culture medium (CD DG44)

Materials and Methods

Metabolites were extracted from mammalian cells, cell culture media, dried blood spots, and human biofluids (cerebrospinal fluid, amniotic fluid, urine, saliva, and blood plasma) using an 80% organic solution. Analyte concentration ranges were achieved using different sample-to-solvent ratios (**Table 1**).

An organic solution (methanol:acetonitrile:water, 50:30:20 v/v/v) spiked with 87 internal standards was used to precipitate proteins and isolate metabolites. Extracts were mixed with a StandardCandles™ solution to compare the traditional absolute quantification method and Pyxis (**Figure 1**). Calibration curves comprising mixtures of pure standards were prepared and analyzed in parallel. Four µl of each extract were analyzed on a Transcend LX-2 multichannel UHPLC system coupled to an Orbitrap Exploris 120 mass spectrometer (Thermo Fisher Scientific). HILIC separation was achieved with an Atlantis Premier BEH Z-HILIC column (2.5 mm, 2.1 x 50 mm; Waters Corporation) and a mobile phase consisting of 20 mM ammonium carbonate with 0.25% (v/v) ammonium hydroxide (pH=9.6, solvent A), and acetonitrile (solvent B). High-resolution MS1 spectra were acquired for 6.7 minutes in polarity switching mode.[4]

For the analytical procedure referred to as the "conventional method," TraceFinder™ software (Thermo Fisher Scientific) was used to calculate the absolute quantitation of analytes using internal standards and external calibration curves. Briefly, TraceFinder reports compound quantitation by integrating the area under the peak in the chromatogram for the respective monoisotopic molecular ion. In parallel, the raw MS files were analyzed with Pyxis (version 1.4.1; Matterworks, Inc., Somerville, MA), and absolute metabolite concentrations were reported.

Results

The conventional method, based on spiked-in isotopically labeled standards, quantified the 87 endogenous metabolites over a concentration range of 0.05 to 30 µM. These endogenous metabolite concentrations were used to benchmark Pyxis

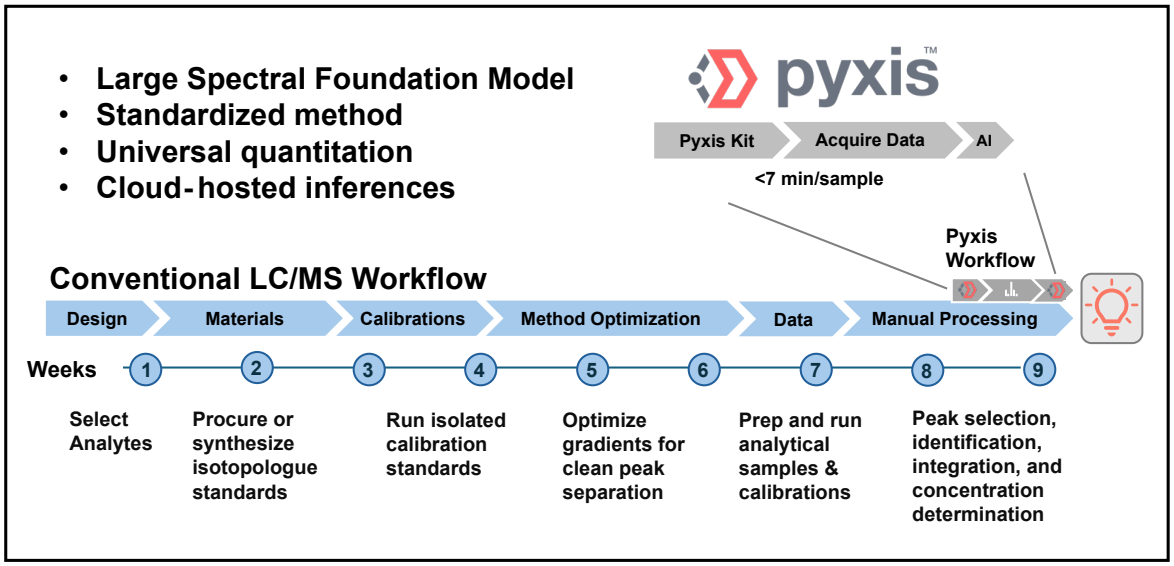

Figure 1. Data acquisition and analysis steps used for both traditional and Pyxis–based absolute metabolite quantitation. Pyxis standardizes LC-MS and reduces weeks of method development, calibration, and data analysis to minutes.

predictions among 27 samples across nine sample matrices. Pyxis successfully quantified all 87 of these biochemicals, ranging from 23 metabolites in the fresh cell culture media to 73 in the CHO cell pellets (**Table 2**).

	HUMAN BIOFLUIDS						GROWTH MEDIA		CELLS
Sample Matrix	AF	CSF	DBS	NIST 1950	Saliva	Urine	HPLM	CD DH44	CHO Cells
Analytes (no.)	55	49	46	51	63	55	50	23	73
Median slope	0.94	0.98	1.38	1.15	0.81	0.76	0.96	0.92	0.81
Median R^2	0.87	0.71	0.60	0.85	0.78	0.81	0.86	0.62	0.87

Table 2. Number of metabolites and linear regression analysis for Pyxis concentration compared to the conventional method concentration. AF=Amniotic Fluid; CSF= Cerebrospinal Fluid; DBS= Dried Blood Spots; NIST 1950= Human plasma (NIST SRM 1950); Saliva= Human Saliva; Urine= Human Urine; CHO Cells= Chinese Hamster Ovary Cells; HPLM= Human plasma-like medium; CD DH44= Cell culture medium (CD DG44)

A linear regression analysis was applied to determine how closely Pyxis predicted the metabolite absolute concentrations within each sample matrix. Pyxis' results were compared with the concentrations determined using the conventional stable isotope results. A slope of 1 indicates perfect 1:1 alignment, while an R2 of 1 represents perfect linear correlation. A summary of the analysis is presented in **Table 2**. Overall, Pyxis predictions achieved median slopes ranging from 0.76 for urine to 1.38 for dried blood spots and median R2 ranging from 0.60 for dried blood spots and 0.87 for amniotic fluid (**Figure 2A**).

A diverse array of analytes representing multiple major biochemical pathways was chosen to demonstrate Pyxis' flexibility. The selected metabolites were grouped into nine primary pathways, and metabolite detection above the LoQ varied depending on the sample matrix. Pyxis exhibited high accuracy in identifying and quantitatively determining metabolites across the different metabolic pathways (**Figure 2B**).

These results demonstrate that Pyxis can annotate analyte concentrations in several human biofluids, CHO cells, and cell growth medium in minutes without tedious and expensive stable isotope-based methodology. Pyxis offers the absolute quantitation of diverse metabolites and delivers results from a sample set within days rather than weeks. A comparable conventional targeted quantitative method costs an order of magnitude more and can take a month or longer to deliver results.

Biomarkers in human health studies

Metabolites uniquely report on genetic function and environmental influences, including diet, microbiome, and exposure.[7] Diverse coverage of the biochemical space offered by traditional untargeted metabolomics affords scientists the best opportunity to identify biomarkers of interest in human populations. While relative abundance measurements can provide clues as to the significance of potential biomarkers, they necessarily require a control or time-zero cohort for comparison, which is not always available in a research study, drives up costs, and extends timelines. Absolute concentrations of metabolite biomarkers hasten their adoption in translational and clinical medicine. For example, “normal” concentration windows of blood metabolic markers (e.g., glucose, bilirubin, creatinine, etc.) and complete blood counts drive their clinical utilization and further enable individual health assessments.[5,6]

We evaluated Pyxis' ability to identify and quantify metabolites in several human biospecimen types routinely used for biomarker studies (**Table 2**). Among all the human matrices, Pyxis quantified all 20 nominal amino acids. To simplify the presentation

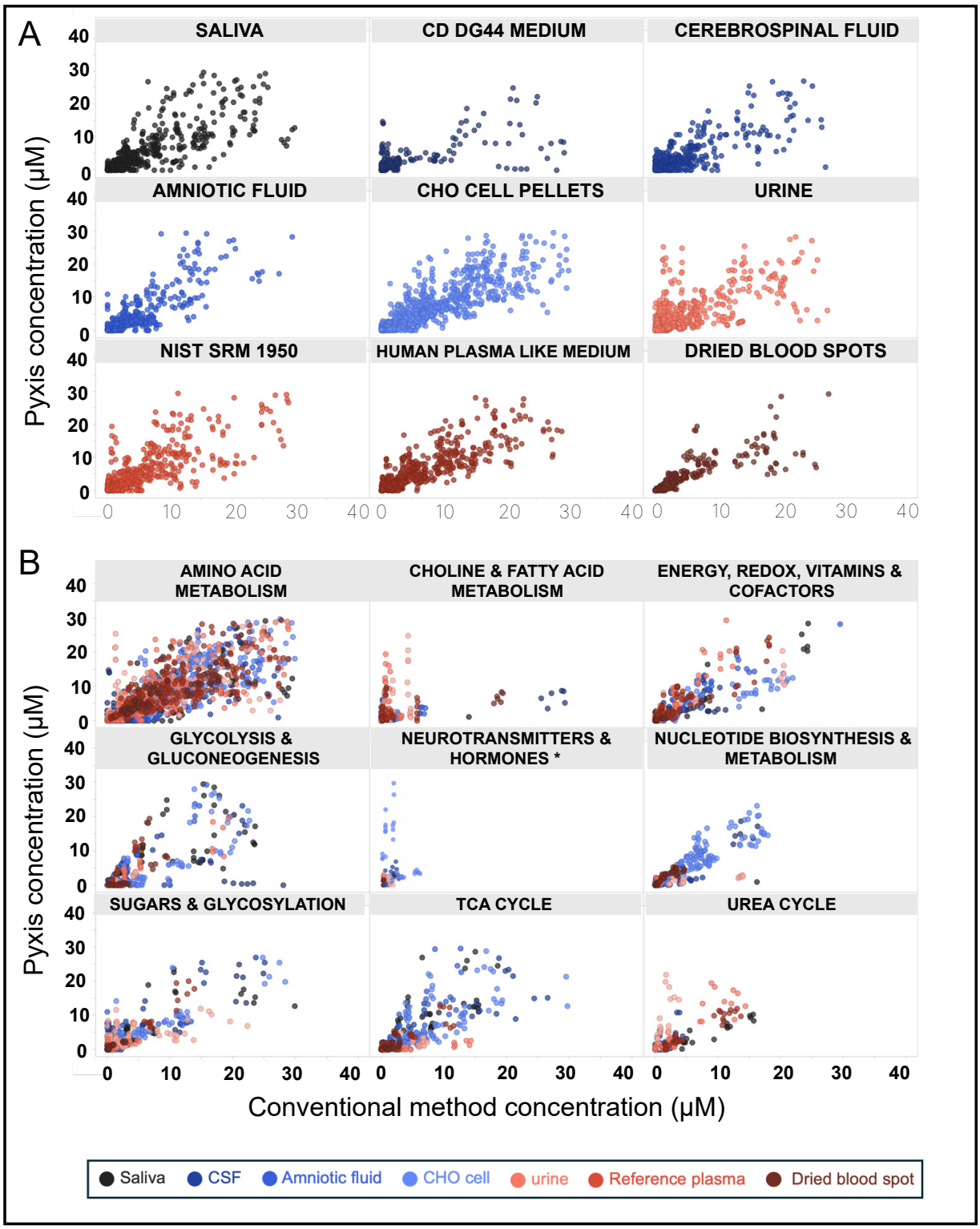

Figure 2. Overview of Pyxis predictions versus the conventionally determined analyte concentrations among (A) nine evaluated matrices and (B) nine grouped metabolic pathways. Sample matrices are colored according to the legend.
**Note the analyte concentrations of "Neurotransmitters & Hormones" are less abundant and thus an order of magnitude lower than the indicated axes.*

of the results, we focused on amino acids quantified in standard reference plasma, amniotic fluid, and urine.

Amino acids and their secondary metabolites inform on nutritional status, and abnormal levels are associated with several chronic and cardiovascular diseases. For example, higher circulating glycine levels may protect against developing coronary heart disease and insulin resistance, the latter of which may lower the risk of type 2 diabetes.[7,8] Pyxis-predicted absolute concentrations of glycine were in good agreement with the conventional method among the standard reference plasma, amniotic fluid, and urine sample matrices (**Figure 3A**).

Tryptophan is an essential amino acid that must be consumed in the human diet. The metabolism of tryptophan to kynurenine and melatonin by human enzymes and indole-related catabolites by bacteria are well-documented mechanisms involved in immune modulation, sleep cycles, and microbiome-potentiated health effects.[9,10] For standard reference plasma, amniotic fluid, and urine biospecimens, Pyxis predicted the concentration dilutions of tryptophan (**Figure 3B**) and kynurenine (**Figure 3C**) in good agreement with conventional method data.

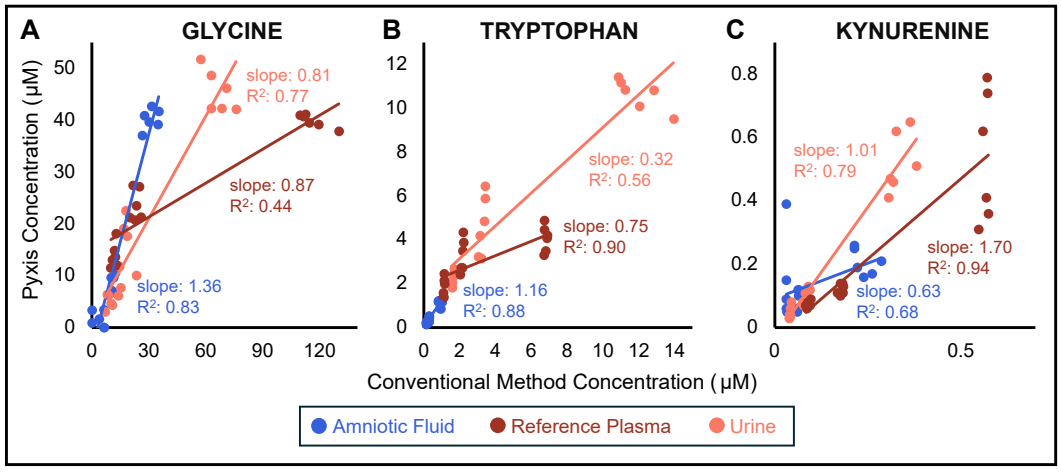

Figure 3. Linear regression analysis for selected amino acids in human biofluids. (A) Glycine, (B) Tryptophan, (C) Kynurenine. Amniotic fluid (blue), standard reference plasma (red), and urine (orange) samples are indicated with associated linear regression fit statistics.

Kynurenine levels were at relatively low abundance, particularly in amniotic fluid, yet Pyxis robustly quantified the amino acid in every dilution of these three matrices.

These results indicate that the Pyxis methodology represents a feasible avenue for determining the levels of physiologically essential biochemicals in human biomarker studies.

Conclusions

In this study, we evaluated the ability of Pyxis, an ML-based cloud platform, to annotate metabolite identity and absolute concentrations in diverse sample matrices using conventional stable isotope-labeled standard methodology as a benchmark. Overall, Pyxis successfully annotated the identity and concentrations of the metabolites in all nine sample types measured, including human-derived matrices, cell pellets, and fresh cell media. The predicted concentrations agreed with the conventional method based on laborious, technically demanding, and expensive isotope labeling and data processing. Furthermore, the Pyxis metabolite identifications and concentrations were available within minutes of uploading the raw data to the platform. Pyxis' annotated metabolite identities and concentrations offer a novel and rapid metabolomics workflow applicable to various biomedical applications, including bioprocess optimization, drug discovery, and translational and clinical biomarker studies.

References

1. Yurkovich, J.T., Evans, S.J., Rappaport, N., Boore, J.L., Lovejoy, J.C., Price, N.D., and Hood, L.E. (2023). The transition from genomics to phenomics in personalized population health. Nature Reviews Genetics 2023 25:4 25, 286–302. https://doi.org/10.1038/s41576-023-00674-x.
2. Lai, Y., Koelmel, J.P., Walker, D.I., Price, E.J., Papazian, S., Manz, K.E., Castilla-Fernández, D., Bowden, J.A., Nikiforov, V., David, A., et al. (2024). High-Resolution Mass Spectrometry for Human Exposomics: Expanding Chemical Space Coverage. Environ Sci Technol 58, 12784–12822. https://doi.org/10.1021/ACS.EST.4C01156/ASSET/IMAGES/LARGE/ES4C01156_0008.JPEG.
3. Beger, R.D., Goodacre, R., Jones, C.M., Lippa, K.A., Mayboroda, O.A., O'neill, D., Lukas Najdekr, ·, Ntai, I., Wilson, I.D., Warwick, ·, et al. (123AD). Analysis types and quantification methods applied in UHPLC-MS metabolomics research: a tutorial. Metabolomics 20, 95. https://doi.org/10.1007/s11306-024-02155-6.
4. Matterworks, I. (2024). Application Note: Simple, Scalable Absolute Concentrations in Untargeted Metabolomics. https://www.matterworks.ai/resources.
5. Foy, B.H., Petherbridge, R., Roth, M.T., Zhang, C., De Souza, D.C., Mow, C., Patel, H.R., Patel, C.H., Ho, S.N., Lam, E., et al. (2024). Haematological setpoints are a stable and patient-specific deep phenotype. Nature 2024, 1–9. https://doi.org/10.1038/s41586-024-08264-5.
6. Overbey, E.G., Kim, J.K., Tierney, B.T., Park, J., Houerbi, N., Lucaci, A.G., Garcia Medina, S., Damle, N., Najjar, D., Grigorev, K., et al. (2024). The Space Omics and Medical Atlas (SOMA) and international astronaut biobank. Nature 2024 632:8027 632, 1145–1154. https://doi.org/10.1038/s41586-024-07639-y.
7. Wittemans, L.B.L., Lotta, L.A., Oliver-Williams, C., Stewart, I.D., Surendran, P., Karthikeyan, S., Day, F.R., Koulman, A., Imamura, F., Zeng, L., et al. (2019). Assessing the causal association of glycine with risk of cardio-metabolic diseases. Nature Communications 2019 10:1 10, 1–13. https://doi.org/10.1038/s41467-019-08936-1.
8. Julkunen, H., Cichońska, A., Tiainen, M., Koskela, H., Nybo, K., Mäkelä, V., Nokso-Koivisto, J., Kristiansson, K., Perola, M., Salomaa, V., et al. (2023). Atlas of plasma NMR biomarkers for health and disease in 118,461 individuals from the UK Biobank. Nature Communications 2023 14:1 14, 1–15. https://doi.org/10.1038/s41467-023-36231-7.
9. Agus, A., Planchais, J., and Sokol, H. (2018). Gut Microbiota Regulation of Tryptophan Metabolism in Health and Disease. Preprint at Cell Press, https://doi.org/10.1016/j.chom.2018.05.003 https://doi.org/10.1016/j.chom.2018.05.003.
10. Routy, J.P., Routy, B., Graziani, G.M., and Mehraj, V. (2016). The kynurenine pathway is a double-edged sword in immune-privileged sites and in cancer: Implications for immunotherapy. International Journal of Tryptophan Research 9, 67–77. https://doi.org/10.4137/IJTR.S38355.

Inquiries

Please direct inquiries to:
Matterworks, Inc.
444 Somerville Ave., Somerville, MA 02143
info@matterworks.ai
https://www.matterworks.ai