https://doi.org/10.1038/s42255-025-01239-4

Discovery of metabolites prevails amid in-source fragmentation

The detectable 'dark metabolome', or unannotated signals in untargeted metabolomics, has become a focal point of debate. The recent Correspondence by Giera et al.[1] published in *Nature Metabolism* argued that most unannotated signals in liquid chromatography–tandem mass spectrometry (LC-MS/MS) data are in-source fragments (ISFs), and hence measurement artefacts rather than new molecules. This claim was based on the observation that 70% of ions detected from 931,000 molecular standards using collision-induced dissociation at 0 eV were ISFs, which led to the interpretation that the dark metabolome may be smaller than previously thought — or even nonexistent[1–3]. However, neither the methodological details nor the data are available to verify this claim. Media headlines, such as "The Dark Metabolome: A Figment of Our Fragmentation?"[2], amplified the sensationalism. However, past research emphasize that such conclusions depend on instrumentation, tuning, sample type, matrix, extraction methods, analyte concentration and data analysis: these studies reported ISF contributions of 2–25% of all detected ions[4,5]. To minimize ISFs, some instruments have dedicated soft ion transfer modes, although this can reduce sensitivity.

The notion that multiple ion forms — different ionizations of the same molecule detectable by mass spectrometry, including ISFs, adducts, isotopes and multimers — arise from a single molecule is not new and well established[4,6]. However, presenting ISFs as evidence that the metabolome is largely known creates a misleading impression, especially for newcomers to this rapidly growing field. Reprocessing of publicly available LC-MS data of ~30,000 chemical standards showed fewer ISFs than the average of 3 reported[1]. As a median, we observed 0–2 ISF fragments, with sodiated ions yielding fewer ISFs than protonated forms and positive ion mode producing more ISFs than negative ion mode (Fig. 1a). However, the observed differences could stem from differences in instrumentation and/or tuning (both in- and post-source). Additionally, our ISF intensity median was just 1% that of the precursor ion, making their prevalence in biological samples unlikely. Intensity varies with structural features: for example, structures with two or more hydroxylations had intensities reaching 5% across all ISFs and 16% when only the most abundant ISF per adduct was considered (Fig. 1b,c). Although untrained or inattentive analysis might misinterpret ISFs as novel molecules, modern data processing workflows account for such errors. Newly discovered molecules, whose structures are reported for the first time, must still be validated using orthogonal techniques such as nuclear magnetic resonance spectroscopy, electronic circular dichroism, X-ray crystallography and/or chemical synthesis.

Abundant evidence highlights the ongoing discovery of new metabolites and biochemical pathways, even in well-studied organisms like *Escherichia coli*, humans and mice[7–9]. Our analysis of a well-studied US National Institute of Standards and Technology (NIST) human fecal reference standard dataset revealed that, even after accounting for all ion forms, 82% of molecules lacked annotations for any ion form when grouped using retention time, MS/MS information and peak shape analysis to estimate the number of molecules they represent (Fig. 1d,e). This result underscores the enduring presence of the dark metabolome, even in thoroughly characterized organisms. The recent and steadily increasing adoption of powerful bioinformatics and data science techniques in metabolomics has propelled the discovery of new metabolites, as shown using rarefaction (Fig. 1f,g), reinforcing the existence of a sizable dark metabolome that awaits discovery.

Although they were in the past considered a nuisance, ion forms, such as ISFs, adducts and multimers, now also present opportunities as computational capabilities improve. For example, ISFs can be used to annotate metabolites[10], give insights into substructures, and improve annotation reliability. ISFs could even be used in the discovery process, as their fragmentation chemistry may parallel certain enzymatic or spontaneous reactions found in biology. Beyond ISFs, adducts and multimeric species could suggest relevant biochemistry: magnesium is needed for ATP stabilization, iron supports haem function, and potassium acts as a counterion for some lipids. Other ionophores bind metals such as copper, zinc or even lanthanides, offering insights into metal transport and homeostasis in biological systems.

The dark metabolome remains a frontier for discovery, with countless unannotated features potentially representing novel molecules or biochemical insights. The human microbiome, with its tens to hundreds of millions of protein-coding genes far exceeding the ~20,000 in the human genome, adds metabolic diversity to these processes. By leveraging innovative approaches, including those that use alternative ion forms, researchers can reveal entirely new dimensions of the metabolome and expand our understanding of the molecular factors shaping human health.

Data availability

The datasets re-analysed here were downloaded from Metabolomics Workbench (study ST002336), MassIVE/Global Natural Product Social Molecular Networking 2 (GNPS accession codes MSV000080673 and MSV000093526) and Zenodo (accession code 13890851).

Code availability

Annotated code used for data analyses performed in this Correspondence, with instructions for use, can be found on GitHub[11,12].

Yasin El Abiead[1,25], Adriano Rutz[2,25], Simone Zuffa[1], Bashar Amer[3], Shipei Xing[1], Corinna Brungs[4], Robin Schmid[4], Mario S. P. Correia[2], Andres Mauricio Caraballo-Rodriguez[1,5], Amir Zarrinpar[6,7,8,9], Helena Mannochio-Russo[1], Michael Witting[10,11], Ipsita Mohanty[1], Tomáš Pluskal[4], Wout Bittremieux[12], Rob Knight[8,13], Andrew D. Patterson[14], Justin J. J. van der Hooft[15,16], Sebastian Böcker[17], Warwick B. Dunn[18], Roger G. Linington[19], David S. Wishart[20], Jean-Luc Wolfender[21,22], Oliver Fiehn[23], Nicola Zamboni[2] & Pieter C. Dorrestein[1,5,13,24] ✉

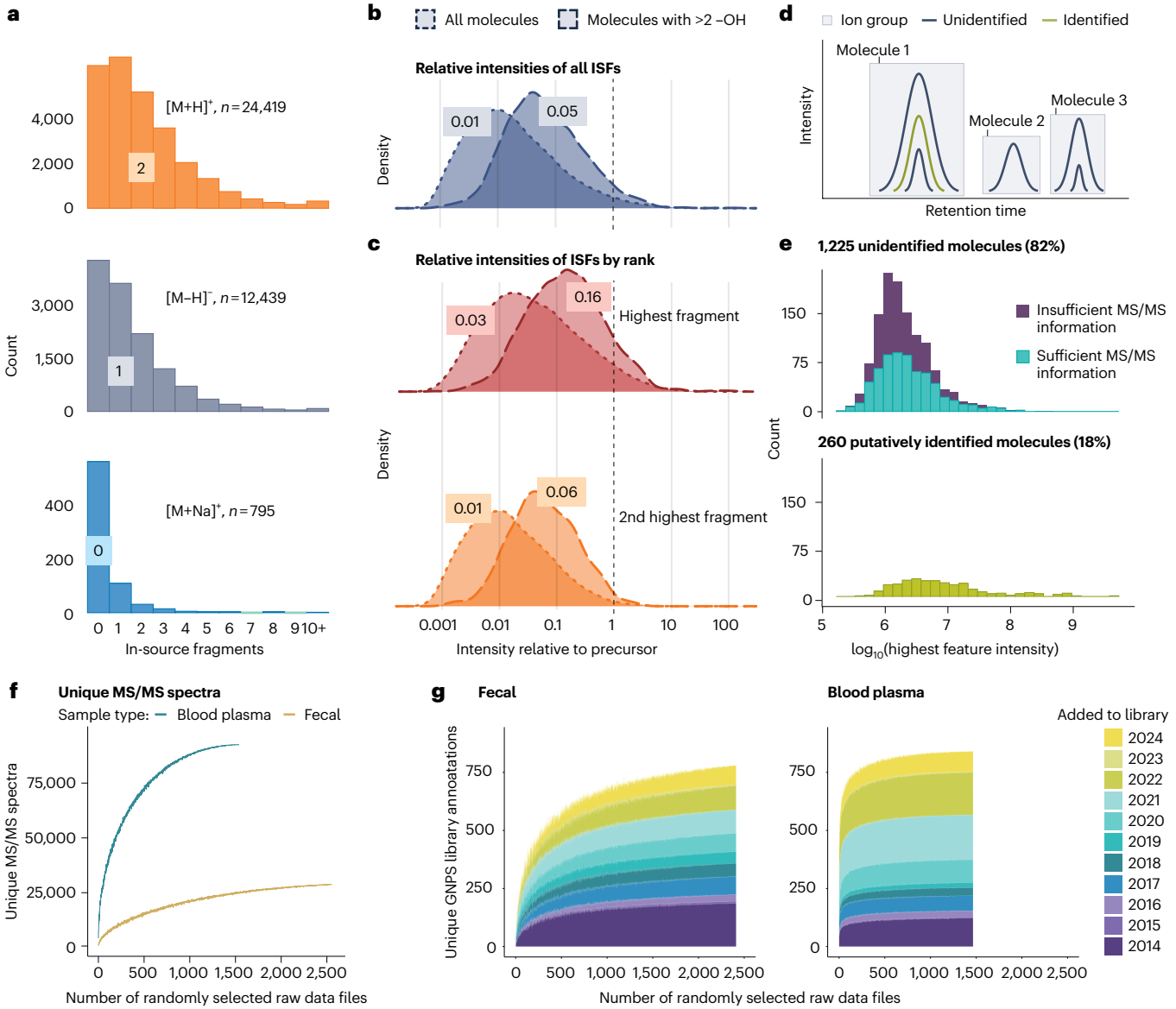

Fig. 1 | ISF distributions, effect on the dark metabolome, and yearly growth of new annotations. a, The number of fragment ions observed for $[M+H]^+$, $[M-H]^-$ and $[M+Na]^+$ adducts, with medians given in boxes. **b**, Distribution and medians of relative intensities for all in- or post-source fragments of ~30,000 standard molecules with no or two or more hydroxyl groups, as these are very prone to ISF. **c**, Relative intensity distributions of the most intense and second most abundant fragment ions. **d**, Features of adducts and in- or post-source fragments can be grouped into ion groups potentially derived from the same molecule through peak shape correlation cliques and MS/MS signals overlapping with co-eluting features. If a feature in an ion group is annotated, the molecule is considered annotated. **e**, Histograms of the intensity distributions for 1,485 feature groups representing unique molecules. They were grouped on the basis of retention time and peak shape analysis. 82% of molecules remain unannotated, with 48% containing at least one feature with no or low-quality MS/MS spectra (data used are from human fecal material MSV000093526). **f**, Rarefaction analysis of unique MS/MS spectra as samples are added for fecal and blood plasma LC-MS/MS data. Differing counts between sample types are due to differing instrumentation. **g**, The growth of unique metabolite annotations coloured by the year the annotation became available. Different adducts for the same molecule were counted as one metabolite annotation. Blood plasma, ST002336; fecal, MSV000080673.

[1]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA, USA. [2]Institute for Molecular Systems Biology, ETH Zurich, Zurich, Switzerland. [3]Thermo Fisher Scientific, San Jose, CA, USA. [4]Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Prague, Czech Republic. [5]Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA, USA. [6]Division of Gastroenterology, University of California, San Diego, La Jolla, CA, USA. [7]Division of Gastroenterology,

Jennifer Moreno Department of Veterans Affairs Medical Center, La Jolla, CA, USA. [8]Shu Chien-Gene Lay Department of Bioengineering, University of California, San Diego, La Jolla, CA, USA. [9]Institute of Diabetes and Metabolic Health, University of California, San Diego, La Jolla, CA, USA. [10]Metabolomics and Proteomics Core, Helmholtz Zentrum München, Neuherberg, Germany. [11]Chair of Analytical Food Chemistry, TUM School of Life Sciences, Technical University of Munich, Freising-Weihenstephan, Germany. [12]Department of Computer Science, University of Antwerp, Antwerp, Belgium. [13]Center for Microbiome Innovation, University of California San Diego, La Jolla, CA, USA. [14]Department of Biochemistry and Molecular Biology and Department of Veterinary and Biomedical Sciences, the Pennsylvania State University, University Park, PA, USA. [15]Bioinformatics Group, Wageningen University, Wageningen, the Netherlands. [16]Department of Biochemistry, University of Johannesburg, Johannesburg, South Africa. [17]Institute of Computer Science, Friedrich Schiller University Jena, Jena, Germany. [18]Department of Biochemistry, Cell and Systems Biology, Centre for Metabolomics Research, Institute of Systems, Molecular, and Integrative Biology, University of Liverpool, Liverpool, UK. [19]Department of Chemistry, Simon Fraser University, Burnaby, British Columbia, Canada. [20]Department of Biological Science, University of Alberta, Edmonton, Alberta, Canada. [21]Institute of Pharmaceutical Sciences of Western Switzerland, University of Geneva, Geneva, Switzerland. [22]School of Pharmaceutical Sciences, University of Geneva, Geneva, Switzerland. [23]West Coast Metabolomics Center, University of California, Davis, Davis, CA, USA. [24]Department of Pharmacology, University of California San Diego, La Jolla, CA, USA. [25]These authors contributed equally: Yasin El Abiead, Adriano Rutz.
✉e-mail: pdorrestein@ucsd.edu

Published online: 28 February 2025

References

1. Giera, M., Aisporna, A., Uritboonthai, W. & Siuzdak, G. *Nat. Metab.* **6**, 1647–1648 (2024).
2. Strachan, J. The dark metabolome: a figment of our fragmentation? *Analytical Scientist* https://theanalyticalscientist.com/fields-applications/the-dark-metabolome-a-figment-of-our-fragmentation (2024).
3. Lowe, D. Phantom metabolites? *Science* https://www.science.org/content/blog-post/phantom-metabolites (2024).
4. Schmid, R. et al. *Nat. Commun.* **12**, 3832 (2021).
5. Lu, W. et al. *Anal. Chem.* **92**, 11573–11581 (2020).
6. Xu, Y.-F., Lu, W. & Rabinowitz, J. D. *Anal. Chem.* **87**, 2273–2281 (2015).
7. Agongo, J. et al. *Anal. Chem.* **96**, 11639–11643 (2024).
8. Qiang, H. et al. Preprint at *bioRxiv* https://doi.org/10.1101/2024.11.13.623458 (2024).
9. Gentry, E. C. et al. *Nature* **626**, 419–426 (2024).
10. Baygi, S. F., Kumar, Y. & Barupal, D. K. *Anal. Chem.* **95**, 9480–9487 (2023).
11. Rutz, A. et al. *GitHub* https://github.com/zamboni-lab/ion-type-analysis/tree/correspondence (2024).
12. El Abiead, Y. et al. *GitHub* https://github.com/YasinEl/DarkMetabolome_figures/blob/correspondence (2024).

Acknowledgements

We acknowledge the NIST Complex Microbial Systems Group for providing the NIST material. A.M.C.-R. and P.C.D. were supported by the Gordon and Betty Moore Foundation, GBMF12120.

Competing interests

P.C.D. is an advisor to and holds equity in Cybele, BileOmix and Sirenas and is a scientific co-founder of, is advisor to, holds equity in and/or received income from Ometa, Enveda and Arome with prior approval by the University of California, San Diego. P.C.D. also consulted for DSM Animal Health in 2023. J.J.J.v.d.H. is member of the scientific advisory board of Naicons Srl., Milano, Italy and consults for Corteva Agriscience, Indianapolis, IN, USA. R.S. and T.P. are co-founders of mzio GmbH. A.Z. is a co-founder of and equity holder in Endure Biotherapeutics. R.K. is a scientific advisory board member and consultant for BiomeSense, Inc., has equity and receives income. He is a scientific advisory board member and has equity in GenCirq. He is a consultant for DayTwo and receives income. He has equity in and acts as a consultant for Cybele. He is a co-founder of Biota, Inc., and has equity. He is a cofounder of Micronoma, has equity and is a scientific advisory board member. The terms of these arrangements have been reviewed and approved by the University of California, San Diego in accordance with its conflict of interest policies. The other authors declare no competing interests. S.B. is co-founder of Bright Giant.