Analytical Science Advances

Chemistry Europe
European Chemical Societies Publishing

EDITORIAL OPEN ACCESS

GinorMass: A Tandem Mass Spectral Repository of 24 Million Spectra From One Molecule

Linh Hoang | Gary Siuzdak

Scripps Center for Metabolomics and Mass Spectrometry, Department of Integrative Structural and Computational Biology, and Department of Chemistry, Scripps Research, La Jolla, California, USA

Correspondence: Gary Siuzdak (siuzdak@scripps.edu)

Received: 2 July 2026 | **Revised:** 4 August 2026 | **Accepted:** 7 August 2026

Mass spectrometry (MS) repositories and reference libraries have become indispensable resources for metabolomics, exposomics, natural products research, and analytical chemistry [1–6]. Recent computational advances now enable searching across billions of tandem mass spectra and spectral relationships within public repositories, representing a remarkable achievement in data indexing and retrieval. These advances have substantially expanded our ability to discover, compare, and reuse mass spectrometry data. However, repository scale should not be conflated with molecular knowledge. Spectral count and authenticated molecular identities represent fundamentally different quantities. Published descriptions of resources such as GNPS, MassBank, and HMDB routinely use the numbers of spectra, entries, datasets, or files to communicate scale [3–6]. These metrics appropriately describe data volume, acquisition activity, and computational capability. However, spectral abundance, now often measured in the billions [6], does not by itself reflect the number, diversity, or confidence of the underlying molecular identities. To make this distinction experimentally explicit, we created GinorMass, a publicly downloadable collection [https://masspec.scripps.edu/] of 23,971,300 MS/MS spectra acquired from a single authenticated molecular standard: arginine.

GinorMass clarifies what each dimension represents. Tandem mass spectra (MS/MS) are analytical observations, whereas authenticated molecular identities represent experimentally anchored molecular knowledge. Replicate measurements may characterise analytical variability and reproducibility, but they do not expand molecular coverage, which grows only when additional compounds are characterised and validated (Figure 1) [7–10].

Arginine was selected as a stable, readily available, biologically relevant and well-characterised reference standard. At 20 eV in positive-ion mode, its spectrum contains five fragment ions, which fall within the typical range of fragments observed across the approximately 960,000 experimentally characterised molecules in METLIN [2]. Arginine therefore provides a reasonable example under typical acquisition conditions. More importantly, GinorMass was designed to distinguish accumulation of spectral observations from expansion of authenticated molecular diversity.

The practical value of additional spectra depends on purpose. Replicates collected across instruments, laboratories, collision energies, or sample environments can improve characterisation of variability, robustness and transferability. For identification of a known compound under a fixed condition, however, additional spectra eventually become increasingly redundant. GinorMass makes both points visible: repeated spectra may deepen analytical characterisation of arginine, but they do not broaden the chemistry contained in the resource. Spectral diversity within a molecule and molecular diversity across a resource should therefore be reported separately.

This distinction raises a broader question: what should the field report when describing spectral resources? Spectral count should remain, but it should be accompanied by complementary measures. At minimum, these should include: (i) the number of unique molecular identities at an explicitly stated structural level; (ii) the number authenticated using reference standards; (iii) chemical-class coverage; (iv) precursor, adduct, polarity, collision-energy and instrument coverage; (v) annotation confi-

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2026 The Author(s). *Analytical Science Advances* published by Wiley-VCH GmbH.

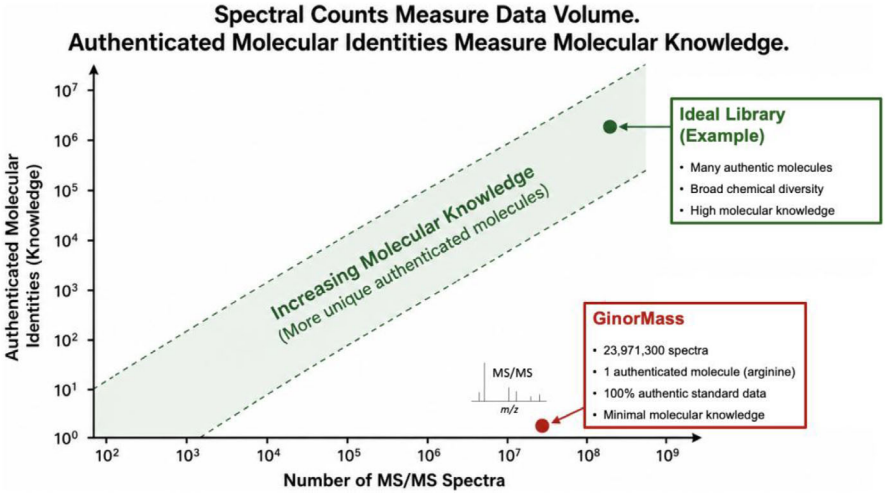


FIGURE 1 | Spectral abundance and authenticated molecular diversity are distinct dimensions of mass spectral resource content. GinorMass (red) contains 23,971,300 experimentally acquired MS/MS spectra from one authenticated compound, arginine. A broadly diverse reference library (green) could contain a comparable number of spectra while representing many authenticated molecular identities. The "Ideal Library" is a conceptual example illustrating a resource that combines high spectral abundance with broad coverage of authenticated molecular identities; its position is illustrative and does not represent a specific repository or quantitatively defined ideal.

dence and provenance. Where extensive replication is present, a measure of spectral redundancy or effective spectral diversity would provide additional context.

These measures can be implemented using transparent, reproducible definitions. Unique identities may be counted at the connectivity level, for example using the first block of the InChIKey, while stereochemical identities can be reported separately. Chemical-class diversity can be summarised using a documented chemical ontology and reported as both the number and distribution of represented classes. Biological relevance can be described through stated cross-references to curated pathway, organism, endogenous-metabolite, drug, natural-product or exposure resources. Annotation confidence should follow an established framework, and provenance should link each spectrum to its molecular standard or source, acquisition conditions and processing history [7–13]. No single metric captures every purpose, but explicit reporting makes resources comparable without conflating data volume with chemical breadth.

This is not an argument against large repositories or extensive replication. The accumulation and sharing of spectra have transformed analytical science. Rather, GinorMass illustrates why the dimensions of resource value should be named accurately. Spectral count usefully measures data volume and computational scale, but it should be reported alongside authenticated molecular identities, chemical and analytical coverage, annotation confidence, provenance and spectral redundancy. GinorMass expresses the distinction in its simplest form: 23,971,300 experimentally acquired tandem mass spectra from one authenticated molecule.

Author Contributions

Linh Hoang: data curation, investigation, visualization, validation and formal analysis. **Gary Siuzdak:** conceptualisation, investigation, funding acquisition, writing – original draft, methodology, validation, visualisation, writing – review and editing, project administration, formal analysis, supervision and resources.

Acknowledgements

The authors appreciate the support of the US National Institutes of Health grants R35 GM130385 (Gary Siuzdak), U01 CA235493 (Gary Siuzdak) and U01 CA305256 (Gary Siuzdak) during the writing of this manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data supporting this article, comprising 23,971,300 arginine MS/MS spectra, are publicly available through Scripps Research at https://masspec.scripps.edu/. Spectra were acquired from an authenticated arginine standard in positive-ion mode using a Bruker Impact II QTOF. Data were collected at an acquisition time of 20–50 ms/spectrum and a collision energy of 20 eV. The arginine standard at 10 µg/mL was introduced into the instrument via syringe pump at a flow rate of 3 µL/min.

References

1. C. A. Smith, G. O'Maille, E. J. Want, et al., "METLIN: A Metabolite Mass Spectral Database," *Therapeutic Drug Monitoring* 27 (2005): 747–751, https://doi.org/10.1097/01.ftd.0000179845.53213.39.

2. W. Uritboonthai, A. E. Aisporna, L. Hoang, et al., "METLIN 960K: An Empirical Tandem Mass Spectrometry Data Resource," *Analytical Chemistry* 98 (2026): 12498–12507, https://doi.org/10.1021/acs.analchem.5c08031.

3. M. Wang, J. J. Carver, V. V. Phelan, et al., "Sharing and Community Curation of Mass Spectrometry Data With Global Natural Products Social Molecular Networking," *Nature Biotechnology* 34 (2016): 828–837, https://doi.org/10.1038/nbt.3597.

4. H. Horai, M. Arita, S. Kanaya, et al., "MassBank: A Public Repository for Sharing Mass Spectral Data for Life Sciences," *Journal of Mass Spectrometry* 45 (2010): 703–714, https://doi.org/10.1002/jms.1777.

5. D. S. Wishart, A. Guo, E. Oler, et al., "HMDB 5.0: The Human Metabolome Database for 2022," *Nucleic Acids Research* 50 (2022): D622–D631, https://doi.org/10.1093/nar/gkab1062.

6. Y. El Abiead, J. I. Seo, V. Charron-Lamoureux, et al., "Structure-centric Searching Enables Global Mapping of the Public Metabolome," *Nature Biotechnology* (2026), https://doi.org/10.1038/s41587-026-03082-8.

7. H. Oberacher, M. Sasse, J.-P. Antignac, et al., "A European Proposal for Quality Control and Quality Assurance of Tandem Mass Spectral Libraries," *Environmental Sciences Europe* 32 (2020): 1, https://doi.org/10.1186/s12302-020-00314-9.

8. L. Sumner, A. Amberg, D. Barrett, et al., "Proposed Minimum Reporting Standards for Chemical Analysis," *Metabolomics* 3 (2007): 211–221, https://doi.org/10.1007/s11306-007-0082-2.

9. E. Schymanski, J. Jeon, R. Gulde, et al., "Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence," *Environmental Science & Technology* 48, no. 4 (2014): 2097–2098, https://doi.org/10.1021/es5002105.

10. G. Theodoridis, O. Fiehn, M. Holčapek, et al., "What's in a Name? Metabolite Identification: Challenges and Pitfalls in Untargeted Metabolomics," *Metabolomics* 22 (2026): 22, https://doi.org/10.1007/s11306-025-02387-0.

11. L. M. S. Khoo and R. Barzilay, "Why Machine Learning Fails at Mass Spectrometry for Small Molecules," *Nature Metabolism* 8 (2026): 1247–1249, https://doi.org/10.1038/s42255-026-01544-6.

12. M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, et al., "The FAIR Guiding Principles for Scientific Data Management and Stewardship," *Scientific Data* 3 (2016): 160018, https://doi.org/10.1038/sdata.2016.18.

13. V. Turkina, Q. van Erp, A. Petrignani, G. Corthals, J. W. O'Brien, and S. Samanipour, "From Spectra to Local Networks: Evaluating MS^2 Similarity Metrics," *Analytical Chemistry* 98 (2026): 17937–17947, https://doi.org/10.1021/acs.analchem.6c01202.