GNPS Dashboard: Collaborative Analysis of Mass Spectrometry Data in the Web Browser

Daniel Petras [1,2], Vanessa V. Phelan [7], Deepa Acharya [15], Andrew E. Allen [2,12], Allegra T. Aron [1], Nuno Bandeira [13], Benjamin P. Bowen [25], Deirdre Belle-Oudry [16], Simon Boecker [17], Dale A. Cummings Jr. [10, 11], Jessica M Deutsch [9], Eoin Fahy [14], Neha Garg [9], Rachel Gregor [18], Jo Handelsman [15], Mirtha Navarro-Hoyos [24], Alan K. Jarmusch [1], Scott A. Jarmusch [23], Katherine Louie [25], Katherine N. Maloney [6], Michael T. Marty [16], Michael M. Meijler [19, 21], Itzhak Mizrahi [20, 21], Rachel L Neve [26], Trent R. Northen [25], Carlos Molina-Santiago [8], Morgan Panitchpakdi [1], Benjamin Pullman [13], Aaron W. Puri [10, 11], Robin Schmid [1], Shankar Subramaniam [14], Monica Thukral [2,12], Felipe Vasquez-Castro [22], Pieter C Dorrestein [1,2,3,4,5], Mingxun Wang [1,4]

1. Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA, USA
2. Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA
3. Center for Microbiome Innovation, University of California San Diego, La Jolla, CA, USA
4. Department of Pharmacology, School of Medicine, University of California San Diego, La Jolla, CA, USA
5. Department of Pediatrics, University of California San Diego, La Jolla, CA, USA
6. Department of Chemistry, Point Loma Nazarene University, San Diego, CA, USA
7. Department of Pharmaceutical Sciences, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of Colorado, Anschutz Medical Campus, CO, USA
8. Instituto de Hortofruticultura Subtropical y Mediterránea, Universidad de Málaga-Consejo Superior de Investigaciones Científicas (IHSM-UMA-CSIC), Departamento de Microbiología, Universidad de Málaga, Bulevar Louis Pasteur 31, 29071, Málaga, Spain
9. School of Chemistry and Biochemistry, Center for Microbial Dynamics and Infection, Georgia Institute of Technology, Atlanta, Georgia, USA
10. Department of Chemistry, University of Utah, Salt Lake City, UT, USA
11. Henry Eyring Center for Cell & Genome Science, University of Utah, Salt Lake City, UT, USA
12. Environmental Genomics, J. Craig Venter Institute, La Jolla, CA, USA
13. Center for Computational Mass Spectrometry, Department of Computer Science and Engineering, University of California, San Diego, La Jolla, CA, USA
14. Department of Bioengineering, University of California San Diego, La Jolla, CA, USA
15. Wisconsin Institute for Discovery and Department of Plant Pathology, University of Wisconsin-Madison, Madison, Wisconsin, USA
16. Department of Chemistry and Biochemistry, University of Arizona, Tucson, AZ, USA
17. Analysis and Redesign of Biological Networks, Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany
18. Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
19. Department of Chemistry, Ben-Gurion University of the Negev, Be'er Sheva, Israel
20. Department of Life Sciences, Ben-Gurion University of the Negev, Be'er Sheva, Israel
21. National Institute of Biotechnology in the Negev, Ben-Gurion University of the Negev, Be'er Sheva, Israel.
22. Centro Nacional de Innovaciones Biotecnologicas (CENIBiot), CeNAT-CONARE, 1174-1200 San Jose, Costa Rica
23. Department of Biotechnology and Biomedicine, Technical University of Denmark, Søltofts Plads 221, DK-2800 Kongens Lyngby, Denmark
24. Bioactivity for Sustainable Development Group (BIODESS), Department of Chemistry, University of Costa Rica, San Jose, Costa Rica

25. DOE Joint Genome Institute and Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA
26. Department of Immunology and Microbiology, University of Colorado, Anschutz Medical Campus, CO, USA

Main Text

The recent disruptions to office and laboratory workspaces resulting from the COVID-19 pandemic, including campus closures, telework, and stay-at-home orders, increased the need for online approaches to scientific research. Most solutions for the analysis of mass spectrometry (MS) data require specific software packages, expertise in command-line based execution, knowledge of where data can be found, familiarity with file transfer protocols, and conversion of data into compatible formats[1–4]. Together, these bottlenecks prevent collaborators, reviewers, and readers of scientific publications from inspecting the raw MS data to reproduce and verify interpretations. It is becoming increasingly important for the scientific community to have open and transparent solutions to share, inspect, and reproduce MS data analysis.

To address these challenges, we developed a synchronous interactive web application, the GNPS dashboard (https://gnps-lcms.ucsd.edu), to be used simultaneously by more than one person in the same analysis environment (similar to collaborative text editing in Google Docs) for MS data analysis to enable a level of collaborative research that was not possible in traditional work environments. The GNPS Dashboard, facilitates the visualization of liquid and gas chromatography-mass spectrometry (LC-MS and GC-MS) data for quality inspection, visualization, sharing, collaborative examination, and hands-on teaching of MS concepts using private and publicly available MS data, including files stored in the MS data repositories GNPS/MassIVE[5], MetaboLights[6], ProteomeXchange[7], and Metabolomics Workbench[8] (**Fig. 1,** Link to Instructions). All public data from compatible repositories can be selected, using the GNPS dataset explorer (https://gnps-explorer.ucsd.edu/). Files not deposited in public MS repositories can be visualized through a drag-and-drop option for file transfer. .mzXML, .mzML, .CDF, and Thermo .raw file formats are compatible with GNPS Dashboard (**SI Note 9**). Via deep linking from the GNPS platform, GNPS Dashboard serves as a data explorer and hub for further data analysis including Molecular Networking[5] (**SI Note 1, 8, 9**), GC-MS deconvolution[9], *in silico* annotation (**SI Note 10**), and MASST[10] (**SI Note 5**).

The GNPS Dashboard enables rapid inspection of Total Ion Chromatograms (TIC), 2D retention time versus *m/z* heat map, extracted ion chromatograms (**SI Note 1, 2, 12, and 15**), and tandem mass spectra (MS/MS) (**SI Note 13**) for inspection/visualization of individual compounds, as well as quantitative comparison of the peak abundances of two groups as box-plots (**SI Fig. 1 and SI Note 7)**. Publication-quality figures of each display item are generated for download in

.svg format. The dashboard can aid peer review of scientific manuscripts (**SI Note 17**) and inspecting public quantitative proteomics data to validate published results (**SI Note 6 and 13**). Beyond visualization and analysis of MS data, the dashboard has been shown to support the development of other bioinformatics tools (**SI Note 14**) that may not have their own web-enabled user interfaces.

The GNPS Dashboard encodes the visualization in a shareable link, which empowers users to share the exact same visualization with others, thus reducing miscommunication and improving data transparency, for example during (remote) meetings with collaborators (**SI Note 3**). Every visualization and analysis result can be shared via a URL that will re-launch the original data visualization on their device along with the history of the analysis (up to 1000 steps per session). Users can share these links with collaborators and embed them in publications, presentations, or social media posts (e.g. Example Tweet and **SI Note 4**). The final visualization can also be shared as a Quick Response (QR) code. Anyone with a link or QR code can build upon the analysis and re-share their additions (**SI Note 3**). Links and QR codes will remain valid for data that has been archived in a public repository.

To uniquely empower remote collaborations or classroom teaching, the GNPS Dashboard includes leader-follower synchronization (real-time updates from one user) and fully collaborative synchronization (real-time updates from multiple users). Followers can disconnect the synchronization to continue the analysis from where the leader left off without needing to reload the data. The GNPS Dashboard leader-follower paradigm has already been used in classroom teaching by at least five institutions, including undergraduate institutions, with up to 50 students per classroom (**SI Note 5**). The fully collaborative synchronization enables multi-user visualization and data exploration similar to online synchronous collaborative document editing (**SI Note18**). For example, users can initiate a collaborative session with two or more people on any web-accessible device with this link (Link and Instructions). Not only is the final state of the analysis saved but so is every discrete action, enabling review past evolution of data analysis. A snapshot and history of the collaborative work can be created and shared with collaborators or publications. Interfaced with the major public MS repositories, the capabilities of the GNPS Dashboard will improve public data analysis, promote re-analysis, encourage data transparency and sharing, and strengthen the reproducibility of MS data analysis.

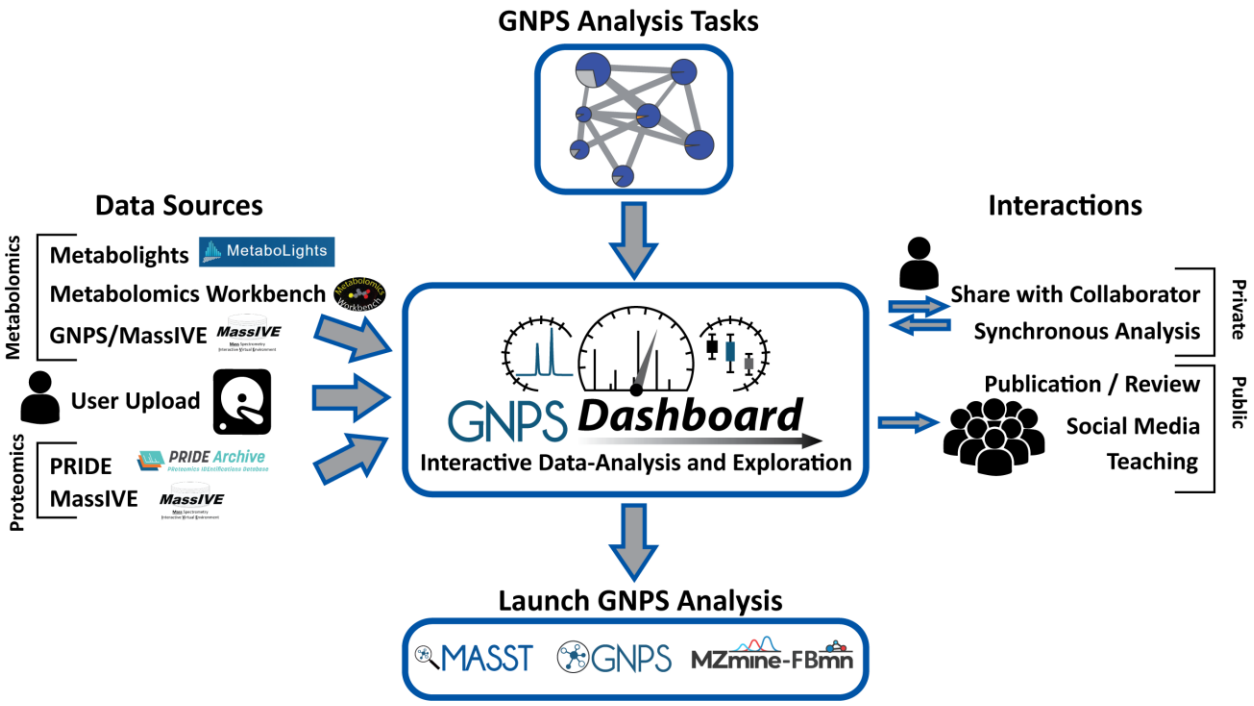

Figure 1: Overview of the GNPS Dashboard in the Data/Analysis Ecosystem. The GNPS Dashboard has been integrated into the web-accessible mass spectrometry ecosystem. The top and left panels show how the GNPS Dashboard can ingress datasets from public metabolomics resources: MetaboLights, Metabolomics Workbench, and GNPS/MassIVE, proteomics resources: PRIDE and MassIVE, private user uploads, and private GNPS analysis tasks. The bottom panel shows that directly out of the GNPS Dashboard, users can launch downstream analysis on their data. Finally, the right panel shows how the GNPS Dashboard and its visualizations can interact with the wider community with reproducible URL or QR code link outs from any analysis and teaching/synchronous collaboration modes for real-time interactivity.

Data availability

All data used within this manuscript and supplemental use cases is publicly available through the MassIVE Repository (massive.ucsd.edu) and proteomeXchange (proteomexchange.org) under the following accession numbers: MSV000086206, MSV000086834, MSV000082378, MSV000086996, MSV000086079, MSV000082493, MSV000085618, MSV000085852, MSV000086584, MSV000086453, MSV000085974, MSV000085376, MSV000086729, MSV000081885, MSV000085070, MSV000086092, MSV000079843/PXD015300, MSV000083508/PXD010154, MSV000087056, MSV000087075, MSV000083859, MSV000087157.

The processed data highlighted in each use case can be downloaded and visualized in the included urls in each example in the supplemental information.

Code availability

While the GNPS Dashboard is accessible as a free public web service, it is possible to locally install the GNPS Dashboard to function with local data sources, making collaborative analysis and sharing possible, privately, within an institution when necessary (*e.g.* government agencies and clinical laboratories).The GNPS Dashboard and GNPS Dataset Explorer and their source code are available through the GNPS web environment and GitHub, enabling quick installation on local servers.

The GNPS Dashboard source code can be found on GitHub: https://github.com/mwang87/GNPS_LCMSDashboard under a modified UCSD BSD License.

The GNPS Dataset Explorer source code can be found on GitHub: https://github.com/mwang87/GNPS_DatasetExplorer under an MIT License.

Acknowledgments

This work was, in part, supported by the National Institutes of Health (NIH) with grant numbers U19AG063744, U2CDK119886, OT2 OD030544, GM107550, R03CA211211, R24GM127667, 1R01LM013115 and P41GM103484, the National Science Foundation (NSF) with grant IOS-1656475 and ABI 1759980, and the Gordon and Betty Moore Foundation (GBMF7622). VVP was supported by the L.S. Skaggs Professorship and Therapeutic Innovation Award from the ALSAM Foundation and NIH R35GM128690. NB and BP were supported by NIH P41 GM103484. MTM was supported by NSF grant CHE-1845230. DB was supported by NSF grant DUE 16-25354.The I.M laboratory was supported by grants from the European Research Council (No. 640384) and from the Israel Science Foundation (ISF No. 1947/19). TRN, BB, and KL were supported by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Furthermore, we would like to thank Tristan de Rond, Laura-Isobel McCall, Wout Bittremieux, Kelly Weldon, and Emily Gentry for testing the software and suggesting updates. We would like to thank Vagisha Sharma, for working on the standalone MSView app during her masters degree research 12 years ago, which provided some visualization inspiration for the GNPS Dashboard. We thank Claire O'Donovan and the team at MetaboLights for developing a well-documented API. AWP and DACJ thank N. Cecilia Martinez-Gomez (UC Berkeley) for *Methylorubrum extorquens* PA1 ∆*cel* strain CM2730, and Ming Hammond (University of Utah) for

use of her LC-MS system. Lastly, we thank all members of the research community who make their data publicly accessible, which contributes to open, transparent, and reproducible science.

Author Contributions

MW and DP conceived the project. MW developed the software for GNPS Dashboard and GNPS Dataset Explorer. MW, DP, VVP, and PCD provided guidance and supervision. MW, DP, VVP, KNM, ATA, AKJ, BP, DACJ, AWP, CMS, MT, NG, NB, and PCD tested and offered user feedback for the GNPS Dashboard. SS and EF implemented the APIs in Metabolomics Workbench to support GNPS Dashboard. TN, BB, KL facilitated making metabolomics data in Genome Portal available in GNPS Dashboard. MW, DP, VVP, KNM, ATA, AKJ, SAJ, DA, BP, DACJ, AWP, CMS, MT, AEA, RS, NB, SB, MTM, DBO, MNH, RG, MMM, IM, JH, FVC, JMD, NG, NB, RLN, and PCD wrote use cases for the GNPS Dashboard Supplemental Information. PCD provided funding for the project. PCD, MW, VVP, and DP wrote the draft manuscript. All authors edited and approved the final manuscript.

Competing Interests

PCD is a scientific advisor of Sirenas, Galileo, Cybele, and scientific advisor and co-founder of Ometa Labs LLC and Enveda with approval by the UC San Diego. MW is a founder of Ometa Labs LLC. TRN is an advisor of Brightseed Bio.

References – 10 references allowed

1. Pluskal, T., Castillo, S., Villar-Briones, A. & Orešič, M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11**, 395 (2010).
2. Tsugawa, H. *et al.* MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat. Methods* **12**, 523–526 (2015).
3. Röst, H. L. *et al.* OpenMS: a flexible open-source software platform for mass spectrometry data analysis. *Nat. Methods* **13**, 741–748 (2016).
4. Huang, Y.-C., Tremouilhac, P., Nguyen, A., Jung, N. & Bräse, S. ChemSpectra: a web-based spectra editor for analytical data. *J. Cheminformatics* **13**, 8 (2021).
5. Wang, M. *et al.* Sharing and community curation of mass spectrometry data with Global

Natural Products Social Molecular Networking. *Nat. Biotechnol.* **34**, 828–837 (2016).

6. Haug, K. *et al.* MetaboLights: a resource evolving in response to the needs of its scientific community. *Nucleic Acids Res.* **48**, D440–D444 (2020).
7. Vizcaíno, J. A. *et al.* ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* **32**, 223–226 (2014).
8. Sud, M. *et al.* Metabolomics Workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. *Nucleic Acids Res.* **44**, D463–D470 (2016).
9. Aksenov, A. A. *et al.* Auto-deconvolution and molecular networking of gas chromatography–mass spectrometry data. *Nat. Biotechnol.* **39**, 169–173 (2021).
10. Wang, M. *et al.* Mass spectrometry searches using MASST. *Nat. Biotechnol.* **38**, 23–26 (2020).