

1 **GNPS Dashboard: Collaborative Analysis of Mass Spectrometry Data in the 2 Web Browser**

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

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

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

54
55 **Abstract (3 sentences, no more than 70 words)**

56 Access to web-based platforms has enabled scientists to perform research remotely. A critical
57 aspect of mass spectrometry data analysis is the inspection, analysis, and visualization of the raw
58 data to validate data quality and confirm statistical observations. We developed the GNPS
59 Dashboard, a web-based data visualization tool, to facilitate synchronous collaborative
60 inspection, visualization, and analysis of private and public mass spectrometry data remotely.

61
62 **Maintext (1500 words max)**

63 Web-based computing has changed our digital lives. The recent disruptions to office and
64 laboratory workspaces resulting from the COVID-19 pandemic, including campus closures,
65 telework, and stay-at-home orders, all increased the need for the development of novel online
66 approaches to scientific research. In particular, the responses to these needs revealed that near-
67 real-time synchronous interactive web applications used simultaneously by more than one person
68 in the same analysis environment (similar to collaborative text editing in Google Docs or Microsoft
69 Office 365), would enable a level of collaborative research that was simply not possible in
70 traditional work environments.

71 One of the key analytical chemistry techniques applied to the life sciences is mass
72 spectrometry (MS). Although there are numerous software solutions for analysis of MS data, most
73 require the installation of specific software packages, advanced knowledge in command-line
74 based execution, expert knowledge of where data can be found, downloading of data with file
75 transfer protocols to a local drive, and conversion of data into compatible formats¹⁻³. Notably,
76 many MS data analysis software packages suffer from incompatibility with different data formats
77 and poor interoperability. The limited accessibility of analysis software poses a significant barrier
78 for MS data inspection by both experts and non-experts, which prevents collaborators, reviewers,
79 and readers of scientific publications from inspecting the raw MS data to reproduce and verify
80 interpretation or discuss data while inspecting it. Although current web-based visualizations of MS
81 data are available for single spectrum visualization⁴⁻⁷, there are no open browser visualization
82 solutions for full MS datasets. With the rapid growth in MS data availability^{6,8-10} and the potential
83 to leverage large MS datasets to develop novel hypotheses, it is becoming increasingly important
84 for the scientific community to have open and transparent solutions to share, inspect, and
85 reproduce MS data and its analysis.

86 We developed the GNPS Dashboard, a centralized web resource (<https://gnps-lcms.ucsd.edu>), to facilitate the visualization of liquid and gas chromatography-mass
87 spectrometry (LC-MS and GC-MS) data for quality inspection, visualization, sharing, collaborative
88 examination, and hands-on teaching of MS concepts using private and publicly available MS data,
89 including files stored in the MS data repositories GNPS/MassIVE⁶, MetaboLights⁹,
90 ProteomeXchange¹¹, and Metabolomics Workbench⁸ (**Fig. 1**, [Link to Instructions](#)). All publicly
91 shared MS files from compatible repositories can be viewed, selected, and compared in the GNPS
92 dataset explorer (<https://gnps-explorer.ucsd.edu>). Files not deposited in public MS repositories
93 can be visualized through a drag-and-drop option for file transfer. Although .mzXML, .mzML¹²,
94 .CDF, and Thermo .raw file formats are compatible with GNPS Dashboard and can be directly
95 uploaded for analysis (**SI Use Case 9**), GNPS's quickstart interface (<https://gnps-quickstart.ucsd.edu/conversion>) or Proteowizard¹³ should be used to convert files to a compatible
96 format. Via deep linking from the GNPS platform, GNPS Dashboard serves as a data explorer
97 and central hub for further data analysis from Classical Molecular Networking⁶ (**SI Use Case 9**)
98 and Feature-based Molecular Networking¹⁴ (**SI Use Cases 1 and 8**), MSHub GC-MS
99 deconvolution¹⁵, *in silico* spectrum annotation via SIRIUS and CSI:FingerID¹⁶ (**SI Use Case 10**),
100 and MASST¹⁷ (**SI Use Case 5**).

103 The GNPS Dashboard enables rapid inspection of Total Ion Chromatograms (TIC), 2D
104 retention time versus *m/z* heat map for global inspection of all signals, Extracted Ion
105 Chromatograms (XIC) (**SI Use Cases 1, 2, 12, and 15**), and tandem mass spectra (MS/MS) (**SI**
106 **Use Case 13**) for inspection/visualization of individual compounds, as well as quantitative
107 comparison of the peak abundances of two groups as box-plots (**Fig. 2 and SI Use Case 7**).
108 Publication-quality figures of each display item are automatically generated for download in
109 scalable vector graphic (.svg) format. Further, the GNPS Dashboard can aid peer review of
110 scientific manuscripts (**SI Use Case 17**) and inspecting public quantitative proteomics data to
111 validate published results (**SI Use Cases 6 and 13**). Beyond visualization and analysis of MS
112 data, the GNPS Dashboard has been shown to support the development of other bioinformatics
113 tools (**SI Use Case 14**) that may not have their own web-enabled user interfaces.

114 The GNPS Dashboard encodes the full state of the visualization interface in an easily-
115 shareable link, which empowers users to share the exact same visualization with others (e.g.,
116 including pan/zoom configurations), thus reducing miscommunication and improving data
117 transparency, for example during (remote) meetings with collaborators (**SI Use Case 3**). Every
118 visualization and analysis result can be shared via a URL that will re-launch the original data
119 visualization on their device along with the history of the analysis (enabling users to rewind and

120 step forward each discrete analysis step, up to 1000 steps per session). Users can share these
121 links with collaborators and embed them in publications, presentations, or social media posts (e.g.
122 [Example Tweet](#) and **SI Use Case 4**). As with the shared link, a final visualization can also be
123 shared as a Quick Response (QR) code. Anyone with a link or QR code can build upon the
124 analysis and re-share their additions (**SI Use Case 3**). Links and QR codes will remain valid,
125 accessible, and embeddable in publications and presentations for data that has been archived in
126 a public repository.

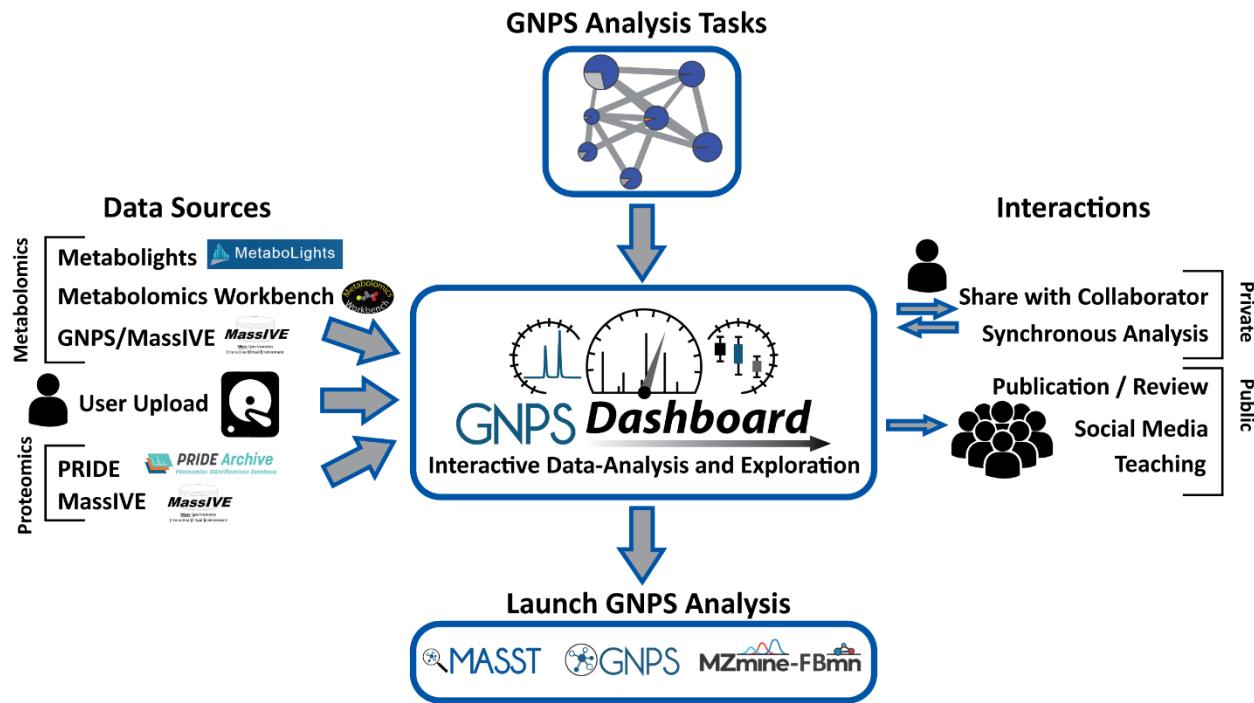
127 To further and uniquely enhance remote collaborations or classroom teaching, the GNPS
128 Dashboard includes leader-follower synchronization (real-time updates from one user) and fully
129 collaborative synchronization (real-time updates from multiple users). The leader-follower mode
130 enables followers to mirror a leader's analysis in real-time. The followers can then disconnect the
131 synchronization at any time to continue the analysis from where the leader left off without needing
132 to reload the data. Synchronization modes facilitate collaborations and remote as well as in-
133 person teaching. The GNPS Dashboard leader-follower paradigm has already been used in
134 remote classroom teaching by at least five institutions, including undergraduate institutions, with
135 up to 50 students per classroom (**SI Use Case 5**). The fully collaborative synchronization enables
136 multi-user simultaneous shaping of the visualization and data exploration in a manner similar to
137 online synchronous collaborative document editing (**SI Use Case 18**). For example, users can
138 initiate a collaborative session with two or more people on any web-accessible device and can
139 edit simultaneously with this link (GNPS Collab Start [Link](#) and [Instructions](#)). In these
140 synchronization and standard analysis modes, not only is the final state of the analysis saved but
141 so is every discrete action, enabling users to follow advancement or review past evolution of data
142 analysis. A snapshot and history of the collaborative work can be created and shared.

143 While the GNPS Dashboard is accessible as a free public web service, it is possible to
144 locally install the GNPS Dashboard to function with local data sources, making collaborative
145 analysis and sharing possible, privately, within an institution when necessary (e.g. government
146 agencies and clinical laboratories). Although only recently introduced, the GNPS Dashboard has
147 already supported the visualization of 8,144 mass spectrometry files from a worldwide user base
148 (October 2020 - April 2021). We envision that over time new features will be added to the GNPS
149 Dashboard to support new mass spectrometry data types (e.g., ion mobility) and visualizations,
150 in collaboration with the mass spectrometry community.

151 Overall, the GNPS Dashboard facilitates the visualization and exploration of mass
152 spectrometry data online, which significantly lowers the barriers to entry for data analysis, and
153 introduces new modes of collaborative data analysis. Tying the capabilities of the GNPS

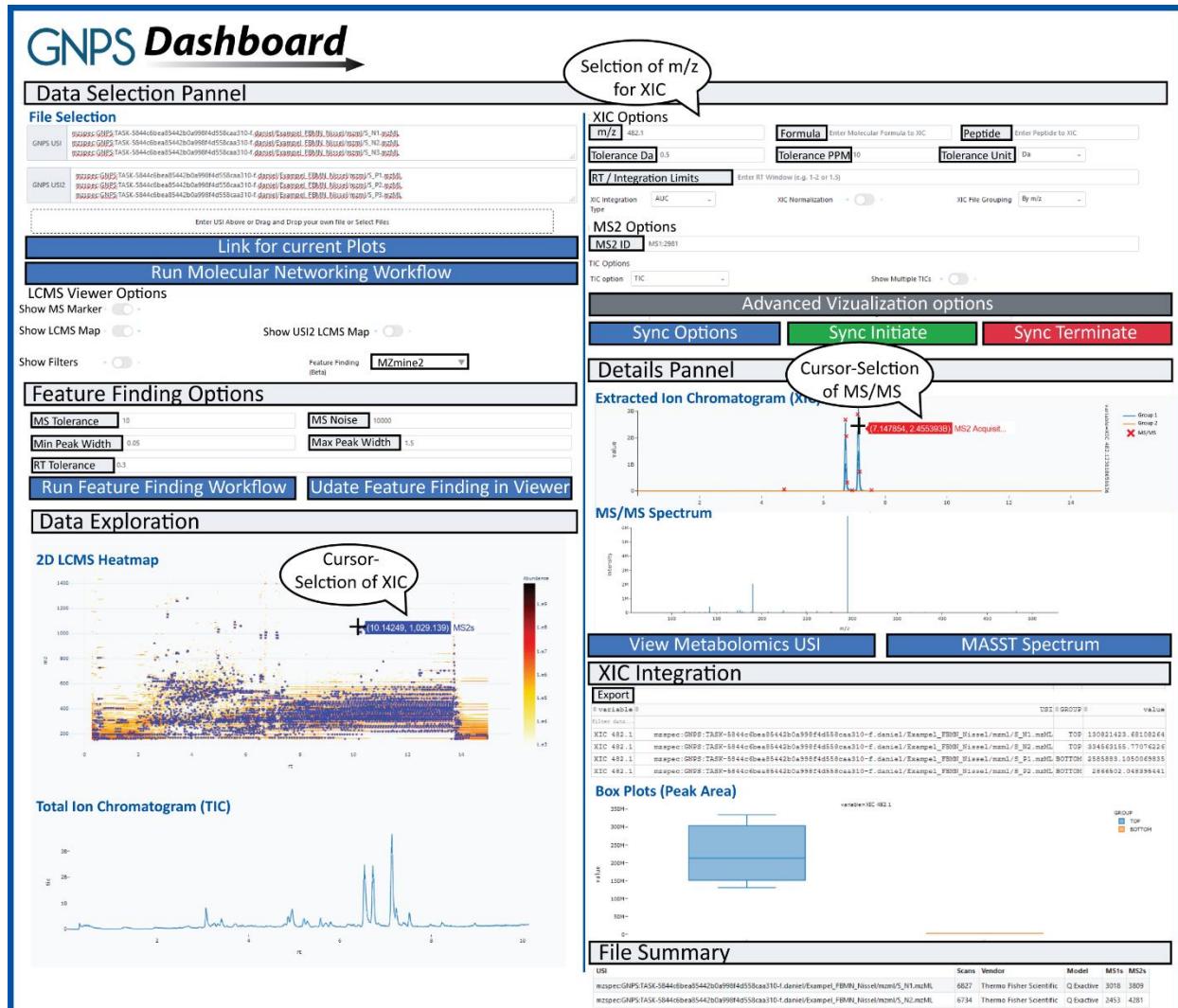
154 Dashboard together with public data repositories will improve public data analysis, promote re-
155 analysis, encourage data transparency and sharing, and strengthen the reproducibility of data
156 analysis.

157
158



159
160 **Figure 1: Overview of the GNPS Dashboard in the Data/Analysis Ecosystem.** The GNPS Dashboard
161 has been integrated into the web-accessible mass spectrometry ecosystem. The top and left panels show
162 how the GNPS Dashboard can ingress datasets from public metabolomics resources: MetaboLights,
163 Metabolomics Workbench, and GNPS/MassIVE, proteomics resources: PRIDE and MassIVE, private user
164 uploads, and private GNPS analysis tasks. The bottom panel shows that directly out of the GNPS
165 Dashboard, users can launch downstream analysis on their data. Finally, the right panel shows how the
166 GNPS Dashboard and its visualizations can interact with the wider community with reproducible URL or
167 QR code link outs from any analysis and teaching/synchronous collaboration modes for real-time
168 interactivity.

169
170
171
172
173
174



175
176
177 **Figure 2: Overview of the GNPS Dashboard User Interface:** The GNPS Dashboard's user interface
178 includes data selection and data exploration visualization via 2D LC-MS heatmaps, total ion
179 chromatogram, extracted ion chromatogram (XIC) and tandem mass spectra as well as simple feature
180 finding options and manual XIC integration. Any analysis and visualization that is created within the
181 GNPS Dashboard can be shared as a URL or quick response (QR) code for the community to visualize,
182 reproduce, and improve the data analysis.
183
184
185
186

187 **Online Methods**

188 The GNPS Dashboard itself is the intersection of several web and mass spectrometry
189 technologies and standards, including the Universal Spectrum Identifier (USI)⁷, pymzML¹⁸,
190 ProteoWizard¹³, ThermoRawFileParser¹⁹, Dinosaur²⁰, and MZmine². The dashboard is built on
191 web technologies with a Python backend and the inclusion of mass spectrometry-specific tools to
192 enable automatic conversion of data and access to the data. The open source data science tools
193 DataShader (Holoviz) and Plotly/Dash (Plotly Inc.) are used for visualization. Finally, the
194 dashboard is integrated into public mass spectrometry resources with the ability to resolve,
195 translate, and acquire data via the mass spectrometry USI.

196 The run environment and deployments are handled via Docker and Docker-Compose. A
197 full set of dependencies and configurations are written into the accompanying source files for the
198 GNPS Dashboard. Detailed documentation on how to use the GNPS dashboard can be found
199 here ([Link](#)).

200

201 **Data availability**

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

210

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

213

214 **Code availability**

215 The GNPS Dashboard and GNPS Dataset Explorer and their source code are available through
216 the GNPS web environment and GitHub, enabling quick installation on local servers.

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

219 The GNPS Dataset Explorer source code can be found on GitHub:

220 https://github.com/mwang87/GNPS_DatasetExplorer under an MIT License.

221

222 **Acknowledgments**

223 This work was, in part, supported by the National Institutes of Health (NIH) with grant numbers
224 U19AG063744, U2CDK119886, OT2 OD030544, GM107550, R03CA211211, R24GM127667,
225 1R01LM013115 and P41GM103484, the National Science Foundation (NSF) with grant IOS-
226 1656475 and ABI 1759980, and the Gordon and Betty Moore Foundation (GBMF7622). VVP was
227 supported by the L.S. Skaggs Professorship and Therapeutic Innovation Award from the ALSAM
228 Foundation and NIH R35GM128690. NB and BP were supported by NIH P41 GM103484. MTM
229 was supported by NSF grant CHE-1845230. DB was supported by NSF grant DUE 16-25354. The
230 I.M laboratory was supported by grants from the European Research Council (No. 640384) and
231 from the Israel Science Foundation (ISF No. 1947/19). TRN, BB, and KL were supported by the
232 U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, which
233 is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-
234 AC02-05CH11231.

235 Furthermore, we would like to thank Tristan de Rond, Laura-Isobel McCall, Wout
236 Bittremieux, Kelly Weldon, and Emily Gentry for testing the software and suggesting updates. We
237 would like to thank Vagisha Sharma, for working on the standalone MSView app during her
238 masters degree research 12 years ago, which provided some visualization inspiration for the
239 GNPS Dashboard. We thank Claire O'Donovan and the team at MetaboLights for developing a
240 well-documented API. AWP and DACJ thank N. Cecilia Martinez-Gomez (UC Berkeley) for
241 *Methylococcus extorquens* PA1 Δ cel strain CM2730, and Ming Hammond (University of Utah) for
242 use of her LC-MS system. Lastly, we thank all members of the research community who make
243 their data publicly accessible, which contributes to open, transparent, and reproducible science.

244

245 **Author Contributions**

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

254 the project. PCD, MW, VVP, and DP wrote the draft manuscript. All authors edited and approved
255 the final manuscript.

256

257 **Competing Interests**

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

261

262

263 **References - 20 references allowed, currently 20.**

264

265 1. Tsugawa, H. *et al.* MS-DIAL: data-independent MS/MS deconvolution for comprehensive
266 metabolome analysis. *Nat. Methods* **12**, 523–526 (2015).

267 2. Pluskal, T., Castillo, S., Villar-Briones, A. & Orešič, M. MZmine 2: Modular framework for
268 processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC
269 Bioinformatics* **11**, 395 (2010).

270 3. Röst, H. L. *et al.* OpenMS: a flexible open-source software platform for mass spectrometry
271 data analysis. *Nat. Methods* **13**, 741–748 (2016).

272 4. Kolbowski, L., Combe, C. & Rappsilber, J. xiSPEC: web-based visualization, analysis and
273 sharing of proteomics data. *Nucleic Acids Res.* **46**, W473–W478 (2018).

274 5. Huang, Y.-C., Tremouilhac, P., Nguyen, A., Jung, N. & Bräse, S. ChemSpectra: a web-based
275 spectra editor for analytical data. *J. Cheminformatics* **13**, 8 (2021).

276 6. Wang, M. *et al.* Sharing and community curation of mass spectrometry data with Global
277 Natural Products Social Molecular Networking. *Nat. Biotechnol.* **34**, 828–837 (2016).

278 7. Deutsch, E. W. *et al.* Universal Spectrum Identifier for mass spectra. *bioRxiv*
279 2020.12.07.415539 (2020) doi:10.1101/2020.12.07.415539.

280 8. Sud, M. *et al.* Metabolomics Workbench: An international repository for metabolomics data
281 and metadata, metabolite standards, protocols, tutorials and training, and analysis tools.
282 *Nucleic Acids Res.* **44**, D463–D470 (2016).

283 9. Haug, K. *et al.* MetaboLights: a resource evolving in response to the needs of its scientific
284 community. *Nucleic Acids Res.* **48**, D440–D444 (2020).

285 10. Perez-Riverol, Y. *et al.* The PRIDE database and related tools and resources in 2019:
286 improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2019).

287 11. Vizcaíno, J. A. *et al.* ProteomeXchange provides globally coordinated proteomics data
288 submission and dissemination. *Nat. Biotechnol.* **32**, 223–226 (2014).

289 12. Martens, L. *et al.* mzML--a community standard for mass spectrometry data. *Mol. Cell. Proteomics MCP* **10**, R110.000133 (2011).

290 13. Chambers, M. C. *et al.* A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **30**, 918–920 (2012).

291 14. Nothias, L.-F. *et al.* Feature-based molecular networking in the GNPS analysis environment. *Nat. Methods* **17**, 905–908 (2020).

292 15. Aksenov, A. A. *et al.* Auto-deconvolution and molecular networking of gas chromatography–mass spectrometry data. *Nat. Biotechnol.* **39**, 169–173 (2021).

293 16. Dührkop, K. *et al.* SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nat. Methods* **16**, 299–302 (2019).

294 17. Wang, M. *et al.* Mass spectrometry searches using MASST. *Nat. Biotechnol.* **38**, 23–26 (2020).

295 18. Bald, T. *et al.* pymzML—Python module for high-throughput bioinformatics on mass spectrometry data. *Bioinformatics* **28**, 1052–1053 (2012).

296 19. ThermoRawFileParser: Modular, Scalable, and Cross-Platform RAW File Conversion | Journal of Proteome Research. <https://pubs.acs.org/doi/10.1021/acs.jproteome.9b00328>.

297 20. Teleman, J., Chawade, A., Sandin, M., Levander, F. & Malmström, J. Dinosaur: A Refined Open-Source Peptide MS Feature Detector. *J. Proteome Res.* **15**, 2143–2151 (2016).

298

299

300

301

302

303

304

305

306

307

308

309

310