ISSN : 2347 - 2243

Indo - American Journal of Life Sciences and Biotechnology

www.iajlb.com
Email : editor@iajlb.com or iajlb.editor@gamil.com

ISSN 2347-2243 www.iajlb.com

Indo-Am. J. of Life Sc & Bt.2018 Vol.15, Issue3, 2018

Indo-American Journal of Life Sciences and Biotechnology

APPLICATION OF VARIOUS ANALYTICAL TOOLS IN METABOLOMICS

[1]Mr. A. Phanindra,[2]Mrs. M. Spandana,[3]Mrs. V. Naveena
[1,2,3]Assistant Professor
[1]Department of Pharmaceutical Analysis
[2,3]Department of Pharm.Chemistry
Vaagdevi Institute of Pharmaceutical Sciences, Bollikunta, Warangal. Telangana.

ABSTRACT

Context Metabolomics plays a pivotal role in addressing a wide range of biological, medicinal, and environmental inquiries, spanning from drug development and precision medicine to the characterization of the dark chemical space of ecosystems and organisms. Data analytics often co-evolve to match the speed of analytical instruments due to technical advancements in mass spectrometry and spectroscopy platforms that facilitate the creation of complicated big-data sets with a wealth of information.

Databases, solutions, software tools, and resources all assist in using the hidden information found in the produced data to ensure successful translation at the end.

The review's objective The scientific community is exposed to around 85 metabolomics software resources, packages, tools, databases, and other utilities that were released in 2020 via this evaluation.

Important scientific ideas for reviews Table 1 lists the resources according on their usefulness and includes links to download the tools as well as their computational needs. In keeping with efforts made since 2015 to assist the community of metabolomics researchers in finding these resources in one location for future reference and use, the review seeks to provide the community with an up-to-date list of all the resources created in 2020.

Key words: Metabolomics · Instrument · Information base · Software · Labeling · Metabolite · In vitro · Source · Application.

I. INTRODUCTION

The year 2020 has seen an enormous rise in applications of ion mobility mass-spectrometry (IMS), and data-independent acquisition (DIA) methods of analyses in both metabolomics and lipidomics. In terms of application, mass spectrometry as a technology promises advance care for cancer patients in clinical and intraoperative use (J. Zhang, Ge, et al., 2020; Zhang, Sans, et al., 2020), imaging mass spectrometry (MSI) based natural products (NPs) discovery (Spraker et al. 2020), nanoscale secondary ion mass spectrometry (nanoSIMS) usage in subcellular MS imaging and quantitative analysis in organelles (Thomen et al. 2020), capturing urban sources of contamination from high resolution mass spectrometry (HRMS) (Bowen et al., 2020) to detection of COVID-19 disease signatures (Mahmud & Garrett, 2020).

From an analytical method development stand point, interesting developments such as plasma pseudotargeted metabolomics method using

ultra-high-performance liquid chromatography–mass spectrometry (UHPLC-MS) (Zheng et al. 2020) and the need for combined use of nuclear magnetic resonance spectroscopy and mass spectrometry approaches in metabolomics (Letertre et al. 2020) are notable. For volume-limited samples, solutions such as subnanoliter metabolomics via LC–MS/MS such as pulsed MS ion generation method known as triboelectric nanogenerator inductive nanoelectrospray ionization (TENGi nanoESI) MS (Li et al. 2020) was introduced. Flow-injection Orbitrap mass spectrometry (FI-MS) enabled reproducible detection of ~ 9,000 and ~ 10,000 m/z features in metabolomics and lipidomics analysis of serum samples, respectively, with a sample scan time of ~ 15 s and duty time of ~ 30 s; a ~ 50% increase versus current spectral-stitching FI-MS methods (Sarvin et al. 2020). A spatial metabolomics pipeline (metaFISH) that combined fuorescence in situ hybridization (FISH) microscopy and high-resolution atmosphericpressure matrix-assisted laser desorption/ionization mass spectrometry to image host–microbe symbioses and their metabolic interactions (Geier et al. 2020) was also reported. Another study that compared the full-scan, data-dependent acquisition (DDA), and data-independent acquisition (DIA) methods in HR LC–MS/MS based metabolomics to reveal that spectra quality is better in DDA with average dot product score 83.1% higher than DIA and the number of MS2 spectra (spectra quantity) is larger in DIA (Guo & Huan, 2020a). Furthermore, it was shown that DDA mode consistently generated fewer uniquely found signifcant features than full-scan and DIA modes (Guo & Huan, 2020b).

Using with Raman spectroscopy, followed by stimulated Raman scattering (SRS) microscopy and Ramanguided subcellular pharmaco-metabolomics in metastatic melanoma cells revealed intracellular lipid droplets that helped identify a previously unknown susceptibility of lipid mono-unsaturation within de-diferentiated mesenchymal cells with innate resistance to BRAF inhibition (Du et al. 2020). Application of 31P NMR was shown to hold potential of expanding the coverage of the metabolome by detecting phosphorus-containing metabolites (Bhinderwala et al. 2020).

The efectiveness of the fow injection analysis-continuous accumulation of selected ions Fourier transform ion cyclotron resonance mass spectrometry (FIA-CASI-FTMS) workfow utilizing isotopic fne structure (IFS) for molecular formula assignment was realized for metabolomics applications (Thompson et al. 2020). A bufer modifcation workfow (BMW) in which the same sample is run by LC–MS in both liquid chromatography solvent with 14NH3–acetate bufer and in solvent with the bufer modifed with 15NH3–formate, resulted in characteristic mass and signal intensity changes for adduct peaks, facilitating their annotation (Lu et al. 2020). Towards reference materials standardization, quantitative measures of approximately 200 metabolites for each of three pooled reference materials (220 metabolites for Qstd3, 211 metabolites for CHEAR, 204 metabolites for NIST1950) were obtained and supported harmonization of metabolomics data collected from 3677 human samples in 17 separate studies analyzed by two complementary HRMS methods (K. H. Liu, Mrzic, et al., 2020; Liu, Nellis, et al., 2020). Another review highlighted the recent progresses (since 2016) in the feld of chemical derivatization LC–MS for

both targeted and untargeted metabolome analysis (Zhao & Li, 2020). The characterization of compounds by the number of labile hydrogen and oxygen atoms in the molecule, which can be measured using hydrogen/deuterium and 16O/18O-exchange approaches allows reduction of the search space by a factor of 10 and considerably increases the reliability of the compound identifcation (Kostyukevich et al. 2020). Preference for monophasic methods that are quicker and simpler than biphasic methods for their amenability and integration into future automation for hydrophilic interaction chromatography (HILIC) ultrahigh-performance liquid chromatography–mass spectrometry (UHPLC–MS) and nonpolar extracts by C18 reversed-phase UHPLC–MS based metabolomics in animal tissues and biofuids (Southam et al. 2020) was also demonstrated. In other innovative applications, use of short columns and direct solvent switches allowed for fast screening (3 min per polarity), where a total of 50 commonly reported diagnostic or explorative biomarkers were validated with a limit of quantifcation that was comparable with conventional LC–MS/MS (van der Laan et al. 2020).

From the stand point of data analysis, metabolomics as a feld is starting to beneft by applying machine learning (ML) (Liebal et al. 2020) and deep learning (DL) (Pomyen et al. 2020; Sen et al. 2020) approaches to address diverse challenges from data preprocessing to biological interpretation. In the context of systems and personalized medicine LIONESS (Linear Interpolation to Obtain Network Estimates for Single Samples) and ssPCC (single sample network based on Pearson correlation) were evaluated and compared in the context of metabolite–metabolite association networks (Jahagirdar & Saccenti, 2020). In annotation domains for low resolution GC–MS data, usage of DL ranking for small molecules identifcation, a deep learning ranking model outperformed other approaches and enabled reducing a fraction of wrong answers (at rank-1) by 9–23% depending on the used data set (Matyushin et al. 2020). In the age of artifcial intelligence, spatial metabolomics and IMS promise to revolutionize biology and healthcare (Alexandrov, 2020). Approaches such as an integrated strategy of fusing features and removing redundancy based on graph density (FRRGD) were proposed that greatly enhanced the metabolome detection coverage with low abundance (Ju et al. 2020).

For a software survey of other mass-spectrometry derived omics tools, packages, resources, softwares and databases, readers can consult other treatise for metaproteomics (Sajulga et al. 2020), data-independent acquisition mass spectrometry-based proteomics (F. Zhang, Ge, et al., 2020; Zhang, Sans, et al., 2020), single cell and single cell-type metabolomics (B. B. Misra, 2020a) among others.

II. PLATFORM-SPECIFC TOOLS

Metabolomics as a discipline depends on mass spectrometry and spectroscopy analytical platforms to generate high through put omics scale data. These include, and are not limited to liquid chromatography-mass spectrometry (LC–MS), gas chromatography-mass spectrometry (GC–MS), capillary electrophoresis-mass spectrometry (CE-MS), and spectroscopic methods such as 1 H-NMR, 13C-NMR, Raman, and Fourier transform infrared (FTIR) among others. In this section, I discuss all the tools that appeared in 2020 for analyses of datasets that

ISSN 2347-2243 www.iajlb.com

Indo-Am. J. of Life Sc & Bt.2018 *Vol.15, Issue3, 2018*

Indo-American Journal of Life Sciences and Biotechnology

are specifc to a metabolomics platform or technology, i.e., LC–MS, GC–MS, and NMR.

Automated spectraL processing system for NMR (AlpsNMR), is an R-package that provides automated signal processing for untargeted NMR metabolomics datasets by performing region exclusion, spectra loading, metadata handling, automated outlier detection, spectra alignment and peak-picking, integration and normalization (Madrid-Gambin et al. 2020). The tool can load Bruker and JDX samples and can preprocess them for downstream statistical analysis.

Signature mapping (SigMa), developed as a standalone tool using MATLAB dependencies, for processing raw urine 1 H-NMR spectra into a metabolite table (Khakimov et al. 2020). SigMa relies on the division of the urine NMR spectra into Signature Signals (SS), Signals of Unknown spin Systems (SUS) and bins of complex unresolved regions (BINS), thus allowing simultaneous detection of urinary

Table 1 The entire list of reviewed tools is organized by important analytical steps in metabolomics data analysis and includes details regarding their platform dependency, and implementation, e.g., programming language (R, Python, Java, C/C++, etc.) or web browser based and their availability

Name of the Software Tool	Category	Platform dependency	Implementation/ use dependency	Software availability	References
AlpsNMR	Platform	NMR	R	https://github.com/sipss/AlpsNMR	(Madrid-Gambin et al. 2020)
SigMa	Platform	NMR	MATLAB, Standalone	https://github.com/BEKZODKHAKIMOV/SigMa_Ver1	(Khakimov et al. 2020)
NMRfilter	Platform	NMR	NA	https://github.com/stefhk3/nmrfilterprojects	(Kuhn et al. 2020)
MSHub + EI-GNPS	Platform	GC–MS	GNPS, Web	https://bitbucket.org/iAnalytica/mshub_process/src/master/; https://github.com/CCMS-UCSD/GNPS_Workflows/tree/master/mshub-gc/tools/mshub-gc/proc	(Aksenov et al. 2020)
RGCxGC toolbox	Platform	GCXGC-MS	R, TeX	https://github.com/DanielQuiroz97/RGCxGC	(Quiroz-Moreno et al. 2020)
CROP	Preprocessing	LC–MS/MS	R	https://github.com/rendju/CROP	(Kouřil et al. 2020)
ncGTW	Preprocessing	LC–MS/MS	R, C++	https://github.com/ChiungTingWu/ncGTW	(Wu et al. 2020)
TidyMS	Preprocessing	LC–MS/MS	Python	https://github.com/griquelme/tidyms	(Riquelme et al. 2020)
AutoTuner	Preprocessing	LC–MS/MS	R	https://github.com/crmclean/Autotuner	(McLean & Kujawinski, 2020)
hRUV	Preprocessing	LC–MS/MS	R	https://shiny.maths.usyd.edu.au/hRUV/	(Kim et al. 2020)
MetumpX	Preprocessing	Any	R	https://github.com/hasaniqbal777/MetumpX-bin	(Wajid et al. 2020)
MetaQuac	QC	Targeted LC–MS	R	https://github.com/bihealth/metaquac	(Kuhring et al. 2020)
dbnorm	QC	Any	R	https://github.com/NBDZ/dbnorm	(Bararpour et al. 2020)
MetaClean	QC	LC–MS/MS	R	https://cran.r-project.org/web/packages/MetaClean/index.html	(Chetnik et al. 2020)
NeatMS	QC	LC–MS/MS	Python	https://github.com/bihealth/NeatMS	(Gloaguen et al. 2020)
MESSAR	Annotation	LC–MS/MS	Web	https://messar.biodatamining.be/	(Liu, Mrzic, et al., 2020; Liu, Nellis, et al., 2020)
SMART 2.0	Annotation	2D NMR	Web	https://smart.ucsd.edu/classic	(Reher et al. 2020)
MetFID	Annotation	MS/MS data	NA	NA	(Fan et al. 2020)

Table 1 (continued)

Name of the Software Tool	Category	Platform dependency	Implementation/ use dependency	Software availability	References
CPVA	Annotation	Any	Web	https://github.com/13479776/cpvamscpva.med.sustech.edu.cn	(Luan et al. 2020)
NRPro	Annotation	LC–MS/MS	Java, Web	https://bioinfo.cristal.univ-lille.fr/nrpro/	(Ricart et al. 2020)
MetENP/MetENPWeb	Annotation	LC–MS/MS	R, Web	https://www.metabolomicsworkbench.org/data/analyze.php	(Choudhary et al. 2020)
CANOPUS	Annotation	LC–MS/MS	Standalone	https://bio.informatik.uni-jena.de/software/canopus/	(Dührkop et al. 2020)
MolDiscovery	Annotation	LC–MS/MS	Python	https://github.com/mohimanilab/molDiscovery	(Cao et al. n.d.)
MetIDfyR	Annotation	LC–MS/MS	R	https://github.com/agnesblch/MetIDfyR	(Delcourt et al. 2020)
Qemistree	Annotation	LC–MS/MS	Python	https://github.com/biocore/q2-qemistree	(Tripathi et al. 2020)
IIMN	Annotation	LC–MS/MS	GNPS, Web	https://ccms-ucsd.github.io/GNPSDocumentation/fbmn-iin/	(Robin Schmid, Daniel Petras, Louis-Félix Nothias, Mingxun Wang, Allegra T. Aron, Annika Jagels, Hiroshi Tsugawa, Johannes Rainer, Mar Garcia-Aloy, Kai Dührkop, Ansgar Korf, Tomáš Pluskal, Zdeněk Kameník, Alan K. Jarmusch, Andrés Mauricio Caraballo-Rodrígu 2020)
FOBI	Annotation	Any	R, Web	https://github.com/pcastellanoescuder/FOBI_Visualization_Tool	(Castellano-Escuder et al. 2020)
Biodendro	Annotation	LC–MS/MS	Python	https://github.com/ccdmb/BioDendro	(Rawlinson et al. 2020)
AllCCS atlas	Annotation	IM-MS	Web	https://github.com/ZhuMetLab/AllCCS; http://allccs.zhulab.cn/	(Zhou et al. 2020)
Binner	Annotation	LC–MS/MS	Java	https://binner.med.umich.edu/	(Kachman et al. 2019)
MS-CleanR	Annotation	LC–MS/MS	R	https://github.com/eMetaboHUB/MS-CleanR	(Fraisier-Vannier et al. 2020)
Retip	Annotation	LC–MS/MS	R	https://www.retip.app/	(Bonini et al. 2020)
QSRR Automator	Annotation	LC–MS/MS	Python	https://github.com/UofUMetabolomicsCore/QSRR_Automator/releases/tag/v1_exe	(Naylor et al. 2020)

Table 1 (continued)

Name of the Software Tool	Category	Platform dependency	Implementation/ use dependency	Software availability	References
MFAssignR	Annotation	LC–MS/MS	R, HTML	https://github.com/skschum/MFAssignR	(Schum et al. 2020)
McSearch	Annotation	LC–MS/MS	R	https://github.com/HuanLab/McSearch; http://cloudmetabolomics.ca/mcsearch	(Xing et al. 2020)
REDU	Annotation	LC–MS/MS	GNPS, Web	https://redu.ucsd.edu/	(Jarmusch et al. 2020)
MASST	Annotation	LC–MS/MS	GNPS, Web	https://masst.ucsd.edu/	(Wang, Jarmusch, et al., 2020; Wang, Leber, et al., 2020)
NPClassifier	Annotation	Any	Web	https://npclassifier.ucsd.edu/	(kim et al. 2020)
patRoon	Annotation	HR MS/MS	R	https://github.com/rickhelmus/patRoon	(Helmus et al. 2021)
LipidLynxX	Annotation	LC–MS/MS	Python, Standalone	http://lipidmaps.org/lipidlynxx/	(Ni & Fedorova, 2020)
Skyline	Multifunctional	Any	Standalone	https://skyline.ms/project/home/software/Skyline/begin.view	(Adams et al. 2020)
NoTaMe	Multifunctional	LC–MS/MS	R, Web	https://github.com/antonvsdata/notame	(Klåvus et al. 2020)
BALSAM	Multifunctional	IMS, GC–MS, LC–MS	Web, Python, HTML, Java	https://exbio.wzw.tum.de/balsam/, https://github.com/philmaweb/balsam_django	(Weber et al. 2020)
MRMkit	Multifunctional	Targeted LC–MS	Python, R	https://github.com/csblab/MRMkit	(Teo et al. 2020)
MetaboShiny	Multifunctional	Any	R	https://github.com/joannawolthuis/MetaboShiny	(Wolthuis et al. 2020)
SmartPeak	Multifunctional	Many	C#, Python	https://github.com/AutoFlowResearch/SmartPeak	(Kutuzova et al. 2020)
MS-DIAL 4.0	Multifunctional	LC–MS/MS, GC–MS, IMS	Standalone	http://prime.psc.riken.jp/compms/msdial/main.html	(Tsugawa et al. 2020)
IP4M	Multifunctional	LC–MS/MS	Java, Perl, R, Standalone	https://IP4M.cn	(Liang et al. 2020)
DropMS	Multifunctional	HR MS	Web	http://www.dropms.online/	(Rosa et al. 2020)
Epimetal	Statistics, visualization	Any	JavaScript, Web	https://github.com/amergin/epimetal; http://epimetal.computationalmedicine.fi/	(Ekholm et al. 2020)
Metabolite AutoPlotter	Statistics, visualization	Quantitative metabolomics data, any	R, Web	https://mpietzke.shinyapps.io/AutoPlotter/	(Pietzke & Vazquez, 2020)
Metabolite-Investigator	Statistics, visualization	LC–MS	R, Web	https://github.com/cfbeuchel/Metabolite-Investigator; https://apps.health-atlas.de/metabolite-investigator/	(Beuchel et al. 2020)
VIIME	Statistics, visualization	Any	Web	https://viime.org/#/	(Choudhury et al. 2020)

Table 1 (continued)

Name of the Software Tool	Category	Platform dependency	Implementation/ use dependency	Software availability	References
struct	Statistics, visualization	Any	R	http://bioconductor.org/packages/struct	(Lloyd et al. 2020)
lipidr	Statistics, visualization	LC–MS/MS	R	https://www.lipidr.org/	(Mohamed et al. 2020)
NOREVA	Statistics	Any	Web, R, Standalone	http://idrblab.cn/noreva/	(Yang et al. 2020)
%polynova_2way	Statistics	Processed data	SAS	https://doi.org/10.1371/journal.pone.0244013.s002	(Manjarin et al. 2020)
rawR	Visualization	LC–MS	R, C++	https://github.com/fgcz/rawR	(Kockmann & Panse, 2020)R
Metaboverse	Visualization	Any	Java, HTML, Standalone	https://github.com/Metaboverse	(Jordan A. Berg, Youjia Zhou, T. Cameron Waller, Yeyun Ouyang, Sara M. Nowinski, Tyler Van Ry, Ian George, James E. Cox, Bei Wang 2020)
JS-MS 2.0	Visualization	LC–MS/MS	Java, JavaScript, HTML	https://github.com/optimusmoose/jsms	(Henning & Smith, 2020)
COCONUT	Database	Any	Web	https://coconut.naturalproducts.net/	(Sorokina et al. n.d.)
METLIN MS2 molecular standards database	Database	LC–MS/MS	Web	http://metlin.scripps.edu/	(Xue et al. 2020)
CSMDB	Database	NMR	MATLAB	https://github.com/cibionmrlab/CSMDB-with-ConQuer-ABC	(Charris-Molina et al. 2020)
EMBL-MCF	Database	LC–MS	NA	https://curatr.mcf.embl.de/	
MIAMI	Isotopic	GC–MS	C++	http://miami.tu-bs.de/	(Dudek et al. 2020)
isoSCAN	Isotopic	GC–MS	R	https://github.com/jcapelladesto/isoSCAN	(Capellades et al. 2020)
LiPydomics	Lipidomics	Ion Mobility, Lipidomics	Python, HTML	https://github.com/dylanhross/lipydomics	(Ross et al. 2020)
LipidCreator	Lipidomics	LC–MS	C#, HTML, Skyline plugin	https://github.com/lifs-tools/lipidcreator	(Peng et al. 2020)
Lipid Annotator	Lipidomics	LC–MS/MS	NA	NA	(Koelmel et al. 2020)
Raman2imzML	MSI	Raman	C++, R	https://github.com/LlucSF/Raman2imzML	(Iakab et al. 2020)
SUMMER	Multiomics	Any	R, Web	http://igc1.salk.edu:3838/summer/ and https://bitbucket.org/salkigc/summer/src/master/	Huang et al. (2020)
metpropagate	Analysis, visualization	Untargeted LC–MS/MS	R, Python	https://github.com/emmagraham/metPropagate	Graham Linck et al. (2020)

The tools generally follow their order of appearance in the manuscript text

metabolites in large-scale NMR metabolomics studies using a SigMa chemical shift library and a new automatic peak picking algorithm. NMR flter, is a stand-alone interactive software for highconfdence NMR compound identifcation that runs NMR chemical shift predictions and matches them with the experimental data, where it defnes the identity of compounds using a list of matching rates and correlating parameters of accuracy together with fgures for visual validation (Kuhn et al. 2020). MSHub/ electron

ionisation (EI)-Global Natural Product Social (GNPS) Molecular Networking analysis, as a platform enables users to store, process, share, annotate, compare and perform molecular networking of both unit/ low resolution and GC–HRMS data (Aksenov et al. 2020). GNPS-MassIVE is a public data repository for untargeted MS2 data, EI-MS data, with sample information (metadata) and annotated MS2 spectra (Aron et al. 2020). MSHub performs the auto-deconvolution of compound fragmentation patterns via unsupervised non-negative matrix factorization and quantifes the reproducibility of fragmentation patterns across samples, followed by GNPS molecular networking analyses. RGCxGC toolbox, is an R-package that aids in analysis of two dimensional gas chromatography-mass spectrometry (2D GC–MS) data by ofering pre-processing algorithms for signal enhancement, such as baseline correction based on asymmetric least squares, smoothing based on the Whittaker smoother, and peak alignment 2D Correlation Optimized Warping and multiway principal component analysis (Quiroz-Moreno et al. 2020).

III. PREPROCESSING AND QUALITY CONTROL (QC) TOOLS

In untargeted metabolomics workflows that use either LC–MS/MS, GC–MS or NMR, depend a lot on pre-processing of the acquired raw datasets prior to statistical analyses and interpretation. Preprocessing typically involves tools that aid in the detection of masses (as m/z's) from mass spectra (i.e., feature detection), construct and display extracted ion chromatograms, detect chromatographic peaks, deconvolution, peak alignment, data matrix curation steps such as batch and blank corrections to fltration and normalization steps, and quality assessments. Though, there are decade old popular preprocessing tools available to the community in the form of xcms (Tautenhahn et al. 2008), MZmine 2 (MZmine Development Team 2015), MS-DIAL (Tsugawa et al. 2015) there is a consistent efort to improve the workfows- from reducing computational time, to developing graphical user interfaces (GUIs) for users to render them user friendly to addressing challenges associated with interpretation of data from advanced platforms such as HRMS data or those from IMS, MSI etc. In fact, a recent comparative efort (among software tools such as software packages MZmine 2, enviMass, Compound Discoverer™, and XCMS Online) demonstrated a low coherence between the four processing tools, as overlap of features between all four programs was only about 10%, and for each software between 40 and 55% of features did not match with any other program (Hohrenk et al. 2020). Moreover, quality control (QC) tools are important to take care of systematic and random variations/ errors induced during experimental and analytical workfows. Batch efects can pose a lot of challenges, i.e., introduction of experimental artifacts that can interfere with the measurement of phenotype-related metabolome changes in metabolomics data (Han & Li, 2020), and data normalization strategies, tools, and software solutions available are reviewed to circumvent some of these challenges (B. B. Misra, 2020b). In this section, I cover the preprocessing and the QC tools that appeared in 2020. Correlation-based removal Of multiPlicities (CROP), implemented as an R-package is a visual post-processing tool that removes redundant features from LC–MS/MS based untargeted metabolomic data sets (Kouřil et al. 2020), where it groups highly correlated features within a defned retention time (RT)

window avoiding the condition of specifc m/z diference making it a second-tier strategy for multiplicities reduction. The output is a graphical representation of correlation network allowing a good understanding of the clusters composition that can aid in further parameter tuning. neighbor-wise compound-specific Graphical Time Warping (ncGTW), is an integrated reference-free profle alignment method, implemented as an R-package and is available as a plugin for xcms that aids in detecting and fxing the bad alignments (misaligned feature groups) in the LC–MS data to render accurate grouping and peak-flling (Wu et al. 2020). TidyMS, is a Python package for preprocessing of untargeted LC–MS/MS derived metabolomics data that reads raw data fro-m a .mzML fle format, generates spectra and total ion chromatograms (TICs), allows peak picking, feature detection, reads processed data from xcms, MZmine 2 among others, ofers functionalities for data matrix curation, normalization, imputation, scaling, quality metrics, QC-based batch corrections and interactive visualization of results (Riquelme et al. 2020). AutoTuner, available as an R-package, is a parameter optimization algorithm that obtains parameter estimates from raw data in a single step as opposed to many iterations in a data-specifc manner to generate robust features from untargeted LC–MS/MS runs (McLean & Kujawinski, 2020). For input, AutoTuner requires at least 3 samples of raw data converted from proprietary instrument formats (e.g. .mzML, .mzXML, or .CDF).

IV. ANNOTATION TOOLS

Metabolite annotation remains a critical step that defnes the success or failure of untargeted metabolomics eforts. With newer technologies such as collision cross section (CCS) data for ion mobility, high resolution mass spectra from Orbitrap, direct injection data, data independent acquisition (DIA)/ all ion fragmentation (AIF), imaging MS and multi-dimensional chromatography the annotation results have gained additional impetus in compound identifcation, but these methods have ofered newer challenges in themselves for tool development. False discovery rates (FDRs) of annotations indicate that low FDRs yield low number yet reliable annotations, whereas higher FDR report high number of annotations by those of poor-quality annotations. Though metabolite annotation eforts can beneft from RT as an orthogonal information, eforts for combining RT predictions with MS/MS data is currently lacking (Witting & Böcker, 2020). Clearly reference spectra and spectral DBs/ libraries are not enough to annotate roughly 5–30% of the total features captured (depending on the environmental/ biological matrices in question) in a given mass spectrometry-based metabolomics dataset. Though experimentally obtained MS/MS data and NMR data on pure standards are precious, and aid in development of computational solutions for compound identifcation, they do not sufce at their current numbers, accessibility, and availability. Moreover, in 2020, the Metabolite Identifcation Task Group of the International Metabolomics Society assessed and proposed a set of revised reporting standards for metabolite annotation/ identifcation and requested community feedback for levels from A-G, from defning an enantiomer or a chiral metabolite (level A) (to unknown molecular formula with specifc spectral features (G). Once formalized, these would positively afect and improve reporting standards in studies and the publication landscape in metabolomics research. In Fig. 1, 2,

Indo-American Journal of Life Sciences and Biotechnology

3, shown are the software interfaces and analysis outputs for some of the annotation tools discussed in the following sections.

V. DATABASES

In this section, I discuss the databases (both spectral and structural) that have appeared or updated in 2020. COlleCtion of Open Natural prodUcTs (COCONUT), is available as a webserver (with downloadable structural data on NPs) an aggregated dataset of NPs from diferent open resources and ofers a subsequent web interface to browse, search and easily and quickly download NPs (Sorokina & Steinbeck, 2020). The DB contains structures and sparse annotations for over 400,000 non-redundant NPs. METLIN MS2, is chemical standards spectral DB that is well annotated and structurally diverse database consisting of over 850,000 chemical standards with MS/MS data generated in both positive and negative ionization modes at multiple collision energies (CEs), collectively containing over 4,000,000 curated HR MS/MS data that covers almost 1% of PubChem's 93 million compounds (Xue et al. 2020). EMBL-MCF, is an open LC–MS/MS spectral library that currently contains over 1600 fragmentation spectra obtained from 435 authentic standards of endogenous metabolites and lipids (Phapale et al. 2021). The EMBL-MCF spectral library is created and shared using an in-house developed web-application. The Wake Forest CPM GC–MS spectral and RT libraries consist of HR EI-MS and HR chemical ionization (CI)- MS/MS spectra obtained from silylated chemical standards obtained from the Mass Spectrometry Metabolite Library of Standards (MSMLS Kit™) (B. B. Misra & Olivier, 2020). Chemical Shift Multiplet Database (CSMDB), is a database that uses JRES spectra obtained from the Birmingham Metabolite Library (BML), to provide scores by accounting for both matched and unmatched peaks from a query list and the database hits (Charris-Molina et al. 2020). This input list is generated from a projection of a 2D statistical correlation analysis on the J-RESolved (JRES) spectra, p-[JRESStatistical TOtal Correlation SpectroscopY (STOCSY)], being able to compare the multiplets for the matched peaks. The CSMDB is complemented with "consecutive queries to assess biological correlation" (ConQuer ABC), a simple inspection of peaks left unmatched from the query list and consecutive queries to assign all (or most) peaks in the original query list.

VI. OTHER SPECIALIZED TOOLS

This section covers numerous tools that did not quite fall into the six categories listed above, and are developed with a purpose to address a specialized application to facilitate metabolomics data analysis. These tools include the ones developed for isotopic data analysis in stable isotope labelling experiments, softwares for analysis of lipidomics data, mass spectrometry imaging data, and multiomics/ integrated omics analysis. Mass isotopolome analysis for mode of action identifcation (MIAMI), is a tool that uses MetaboliteDetector (https://md.tu-bs.de/) and non-targeted tracer fate detection (NTFD) libraries (http://ntfd.mit.edu/), combines the strengths of targeted and non-targeted eforts for estimation of metabolic fux changes in GC–MS datasets (Dudek et al. 2020). In stable isotope labeling experimental data, MIAMI determines a mass isotopomer distribution-based (MID) similarity network and incorporates the data into metabolic reference networks and aids in the identifcation of MID variations of all labeled metabolites

across conditions, targets of metabolic changes are detected. isoSCAN, is an R-package that automatically quantifes all isotopologues of intermediate metabolites of glycolysis, tricarboxylic acid (TCA) cycle, amino acids, pentose phosphate pathway, and urea cycle, from low resolution (LR) MS and HRMS data (i.e., GC-chemical ionization -MS) in stable isotope labeling experiments (Capellades et al. 2020). LiPydomics, is available as a Python package which performs statistical and multivariate analyses ("stats" module), generates informative plots ("plotting" module), identifes lipid species at diferent confdence levels ("identifcation" module), and performs a text-based interface ("interactive" module) aiding in further interpretation (Ross et al. 2020). LipidCreator, is available both as a Skyline plugin and a standalone/command-line operation, is a lipid building block-based workbench and knowledgebase for semi-automatic generation of targeted lipidomics MS assays and in silico spectral libraries (Peng et al. 2020). It can support diverse lipid categories, allows SRM/ parallel reaction monitoring (PRM) assay generation for both labeled and unlabeled lipid species and their derived fragment ions, allows in silico spectral library generation and CEs optimization and the entire workfow can be integrated into Konstanz Information Miner (KNIME™) and Galaxy workfows as a native node. Lipid Annotator, is a standalone software for lipidomic analysis of data collected by HR LC–MS/MS (Koelmel et al. 2020). Lipid Annotator algorithm, intended for lipid annotation based on in-silico libraries, consists of fve general steps: feature fnding, association of MS/MS scans with features, annotation of possible lipids for each feature, calculation of the percent abundance of each fatty acyl constituent under a single chromatographic peak in the case of mixed spectra, and fltration of fnal annotated features. Lipid Annotator can be used on large datasets for rapid annotation, relative quantifcation, and statistics (using a downstream workfow with commercial tools such as MassHunter Profnder (Agilent Technologies) and MassHunter Mass Profler Professional softwares (Agilent Technologies).

VII. CONCLUSION

In conclusion, it is evident that a great deal of tools were created in 2020 alone, either entirely from scratch or as an evolution of earlier iterations. Certain techniques and instruments discovered new uses, as GNPS in the field of GC–MS-based metabolomics (Aksenov et al. 2020), or were made available as a beta or enhanced version, like MS-DIAL for lipidomics (Tsugawa et al. 2020) workflows.

Which of these 2020 tools survives another year in terms of usefulness or applicability, is kept up to date and accessible, is enhanced, and is embraced by the metabolomics research community will depend only on what comes next. Whatever the case, these tools are all helpful in comprehending metabolomics data from many perspectives and are valuable contributions to the community as we go into the big data-driven precision medicine age. Generally speaking, the tendency is to create robust, user-friendly, open-source, quick, and computationally light tools that can follow the findable, accessible, interoperable, and repeatable (FAIR) principles. The metabolomics research community surely needs more of these enhanced tools, and in the next years, more and better tools, resources, and databases will be made available.

REFERENCES

1. Adams, K. J., Pratt, B., Bose, N., Dubois, L. G., St. John-Williams, L., Perrott, K. M., et al. (2020). Skyline for small molecules: A unifying software package for quantitative metabolomics. Journal of Proteome Research, 19(4), 1447–1458. https://doi.org/10.1021/acs.jproteome.9b00640
2. Aksenov, A. A., Laponogov, I., Zhang, Z., Doran, S. L. F., Belluomo, I., Veselkov, D., et al. (2020). Auto-deconvolution and molecular networking of gas chromatography–mass spectrometry data. Nature Biotechnology. https://doi.org/10.1038/s41587-020-0700-3
3. Alexandrov, T. (2020). Spatial metabolomics and imaging mass spectrometry in the age of artifcial intelligence. Annual Review of Biomedical Data Science, 3, 1. https://doi.org/10.1146/annurev-biodatasci-011420-031537
4. Aron, A. T., Gentry, E. C., McPhail, K. L., Nothias, L. F., NothiasEsposito, M., Bouslimani, A., et al. (2020). Reproducible molecular networking of untargeted mass spectrometry data using GNPS. Nature Protocols, 15(6), 1954–1991. https://doi.org/10.1038/ s41596-020-0317-5
5. Bararpour, N., Gilardi, F., Carmeli, C., Sidibe, J., Ivanisevic, J., Caputo, T., Augsburger, M., Grabherr, S., Desvergne, B., Guex, N., Bochud, M., Thomas, A. (2020). Visualization and normalization of drift efect across batches in metabolome-wide association studies. biorx, 914051. https://doi.org/10.1101/2020.01.22.914051
6. Berg, J. A., Zhou, Y., Cameron Waller, T., Ouyang, Y., Nowinski, S. M., Van Ry, T., George, I., Cox, J. E., Wang, B., Rutter, J. (2020). Gazing into the Metaboverse: Automated exploration and contextualization of metabolic data. bioRxiv, 171850. https://doi.org/10.1101/2020.06.25.171850
7. Beuchel, C., Kirsten, H., Ceglarek, U., & Scholz, M. (2020). MetaboliteInvestigator: An integrated user-friendly workfow for metabolomics multi-study analysis. Bioinformatics. https://doi.org/10.1093/bioinformatics/btaa967
8. Bhinderwala, F., Evans, P., Jones, K., Laws, B. R., Smith, T. G., Morton, M., & Powers, R. (2020). Phosphorus NMR and its application to metabolomics. Analytical Chemistry, 92(14), 9536–9545. https://doi.org/10.1021/acs.analchem.0c00591
9. Bonini, P., Kind, T., Tsugawa, H., Barupal, D. K., & Fiehn, O. (2020). Retip: Retention time prediction for compound annotation in untargeted metabolomics. Analytical Chemistry, 92(11), 7515–7522.

Indo-American Journal of Life Sciences and Biotechnology

https://doi.org/10.1021/acs.analchem.9b05765

10. Bowen, Du., Tian, Z., Peter, K. T., & Kolodziej, E. P. (2020). Developing unique nontarget high-resolution mass spectrometry signatures to track contaminant sources in urban waters. Environmental Science & Technology Letters, 7(12), 923–930. https://doi.org/10.1021/acs.estlett.0c00749