Quick tips to perform a metabolomics study

Nascent, burgeoning, youngest, ‘connecting link between genotype and phenotype’ are just some of the phrases associated with this ‘omics’ where small molecules (metabolites with molecular weight < 2000 Daltons) are studied in biological systems, i.e., metabolomics. Currently, due to recent advances in the field, metabolomics has demanded large attention from scientists as it has shown tremendous potential in basic research such as the study interaction of ‘omes’ and the discovery of new biochemical pathways. Other areas of impact include metabolic regulation, disease biomarker identification, personalized medicine, clinical trials, toxicology, nutrigenomics, medicine diagnosis, and agriculture. Also included are industrial applications such as metabolic engineering in strain improvement in microbes. Here, the attempt is to get the scientific community interested and excited about the potential use of metabolomics in their existing research areas, by widening its dimensions to another level of data-oriented science leading to better interpretation and understanding of the function and behavior of organisms. Nonetheless, while excellent pioneering reviews and extremely successful studies exist in metabolomics, a summary and/or overview of this area is lacking at this moment. Thus, the target audience are not only researchers who are venturing into metabolomics, but also early career researchers, investigators, students, and individuals interested in implementing metabolomics in their current research and field. Typically, study design to publication can span from months to years in some cases, but without a general grasp of how beneficial and easy to implement metabolomics can be, further inquiry into the field may be mistakenly overlooked. I attempt to summarize and encapsulate the ‘usual’ trends in such efforts, identifying the critical steps to make the trends all fit into these quick tips. The objective of this article is to provide a bird’s eye view of metabolomics research that can be enjoyed over a mug of coffee while simultaneously providing an outline for a variety of audiences interested in incorporating this fascinating field in their work.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2002v1 | CC-BY 4.0 Open Access | rec: 26 Apr 2016, publ: 26 Apr 2016

Quick Tips to Performing a Metabolomics Study

Biswapriya B. Misra[1*]

[1]Independent Researcher, SW 16th Avenue, Gainesville, FL 32601 USA

Short title: Metabolomics education

***Corresponding author:**

E-mail: bbmisraccb@gmail.com

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2002v1 | CC-BY 4.0 Open Access | rec: 26 Apr 2016, publ: 26 Apr 2016

Abstract

Nascent, burgeoning, youngest, 'connecting link between genotype and phenotype' are just some of the phrases associated with this 'omics' where small molecules (metabolites with molecular weight < 2000 Daltons) are studied in biological systems, i.e., metabolomics. Currently, due to recent advances in the field, metabolomics has demanded large attention from scientists as it has shown tremendous potential in basic research such as the study interaction of 'omes' and the discovery of new biochemical pathways. Other areas of impact include metabolic regulation, disease biomarker identification, personalized medicine, clinical trials, toxicology, nutrigenomics, medicine diagnosis, and agriculture. Also included are industrial applications such as metabolic engineering in strain improvement in microbes.

Here, the attempt is to get the scientific community interested and excited about the potential use of metabolomics in their existing research areas, by widening its dimensions to another level of data-oriented science leading to better interpretation and understanding of the function and behavior of organisms. Nonetheless, while excellent pioneering reviews and extremely successful studies exist in metabolomics, a summary and/or overview of this area is lacking at this moment. Thus, the target audience are not only researchers who are venturing into metabolomics, but also early career researchers, investigators, students, and individuals interested in implementing metabolomics in their current research and field. Typically, study design to publication can span from months to years in some cases, but without a general grasp of how beneficial and easy to implement metabolomics can be, further inquiry into the field may be mistakenly overlooked. I attempt to summarize and encapsulate the 'usual' trends in such efforts, identifying the critical steps to make the trends all fit into these quick tips. The objective

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2002v1 | CC-BY 4.0 Open Access | rec: 26 Apr 2016, publ: 26 Apr 2016

of this article is to provide a bird's eye view of metabolomics research that can be enjoyed over a mug of coffee while simultaneously providing an outline for a variety of audiences interested in incorporating this fascinating field in their work.

Tip 1: Formulate a robust hypothesis and use a good study design

Of prime importance is to have a good question up your sleeves before initiating a study. This will dictate the entire metabolomics work flow, or pipeline, which will greatly influence the outcome of the study. It could be grant-, program-, or funding-driven, but ultimately the objective of the study must be to solve scientific challenges for future applications and would be as good as the working hypothesis. Moreover, it is important to realize that metabolomics can help both in solving a biological questions in hand and in hypothesis generation as well- a double-edged sword. The hypothesis will dictate which technologies to choose from, the source of the metabolites, such as the appropriate organisms, and which approaches to use (targeted vs. untargeted) to carry on the metabolomics study.

The study design can be challenging, painful, and error-prone, as is the case with any scientific are of research. However, with careful and timely planning, the study can be planned well ahead of time. This would also help in performing effective power analysis to provide necessary dimensionality to the data for statistical validations. Considerations on sample size is thus, very important. For instance, essential to the study design is randomization, a procedure that randomly determines the allocation of the experimental material and the order of experimental runs, or orthogonalization which nullifies the effects from unknown and known sources of variations.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2002v1 | CC-BY 4.0 Open Access | rec: 26 Apr 2016, publ: 26 Apr 2016

Some approaches that can become detrimental issues, like blocking, replication, comparison, and following factorial study designs, can be circumvented. In fact, both free and paid design of experiments (DOEs) software products are available such as SAS (SAS Institute, Cary, NC), JMP (SAS Institute, Cary, NC), Modde (Umetrics, Kinnelon, NJ), Design Expert (StatEase, Minneapolis, MN) and others which provide a compromise between information quality and ease of use.

Tip 2: Choose the right platforms available to you

Typically, users choose either mass spectrometry (MS) or spectroscopy-based platforms to investigate metabolomes. For MS-base metabolomics, gas chromatography (GC), liquid chromatography (LC), and capillary electrophoresis (CE) have been shown to produce the best results. However, for ion mobility mass spectrometry (IMS) [2] and specific applications such as metallomics, inductively coupled plasma mass spectrometry (ICP-MS) [3] has been providing exciting results for various types of studies as well. Among other platforms, Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FT-ICR) with its resolving power values over 1,000,000, a high mass accuracy with less than 1 ppm errors, high limits of detection, a dynamic range of 10,000 better fragmentation, and MS^n capabilities can provide great potential for metabolomics applications [4]. Availability to of an user to desirable platforms either for resolution (such as QToFs) and sensitivities for better quantifications (QTraps), modes of ionization (chemical, electron impact, or electron spray) etc., would help address proper solutions to the study. Non-targeted metabolomics approaches have enormous potentials in discovery science [5].

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2002v1 | CC-BY 4.0 Open Access | rec: 26 Apr 2016, publ: 26 Apr 2016

However, recent integration of multiple platforms such as MS and NMR has shown to have greater promise in expanding the metabolome in a more complimentary manner [6]. A word of caution is that with hyphenation and increased dimensionality or hyphenation, the complexity and enormity of data multiplies and so does the down-stream processing time and expertise in terms of manpower and software requirements. In addition, other methods such as Raman spectroscopy and Fourier Transform Infra-Red (FTIR) spectroscopy have found applications in metabolomics. A balance must be found between identifying a larger number of metabolites and to confidently quantifying a smaller number of metabolites. In addition to choosing the correct platform, skilled manpower, handling skills, and instrument troubleshooting, generating good quality data, to avoid the 'garbage in, garbage out' phenomena, is equally important.

Tip 3: Collect and handle the samples following the metabolomics standard initiative (MSI) guidelines

Sample selection, preparation or processing, quenching, storage, handling, and extraction are all important steps in obtaining quality metabolomics data. Variations induced in each of the steps would impact huge differences in acquired data sets. Including appropriate internal standards, which can be instrument-or chemical class dependent, retention index standards, metabolomic test mixtures, either commercial or pooled samples, and quality controls (QCs) are important to have. The factors check on data acquisition, batch variations, and instrument health among other issues that may arise in the study. Many times, at this step of sample collection, isotopic labeling needs to be performed. For processing live organisms to bio fluids, whole lysate, or extracts,

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2002v1 | CC-BY 4.0 Open Access | rec: 26 Apr 2016, publ: 26 Apr 2016

approaches can be different. Quenching of metabolism to storage of thermo-labile intermediates can be really challenging. At this stage, a thought must be given to either have technical replicates (instrument or sample wise) or conditional (i.e., biological) replicates to allow understanding of variability in population, independent sampling, and enough replicates for statistical power analysis. Helpful guidelines in the form of metabolomics standard initiative (MSI) proposed by the Metabolomics Society (http://metabolomicssociety.org/) are summarized [7, 8]. Even organizing the meta-data in the form of Investigation/Study/Assay (tab-delimited (ISA-TAB) format [9] would help follow most basic requirements while conducting a metabolomics investigation. These standards are still evolving with community-driven initiatives [10].

Tip 4: Choose your tools for data pre (processing) and annotation

Depending on whether the datasets are obtained from MS, tandem MS, or spectroscopy platforms, the software tools can be the similar or strikingly different. For instance, ProteoWizard Tools (http://proteowizard.sourceforge.net/tools.shtml) offers a command line tool, msConvert for converting between various file formats in MS to more popular and convenient formats such as .mzZML, .mzML, and .cdf when vendor-specific commercial tools do not provide acquired data sets in universally readable formats. A plethora of curated tools can be obtained from OMICTools (http://omictools.com/), Fiehn lab resources (http://fiehnlab.ucdavis.edu/), and Metabolomics Society's Resources pages (http://metabolomicssociety.org/resources/metabolomics-databases) among others.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2002v1 | CC-BY 4.0 Open Access | rec: 26 Apr 2016, publ: 26 Apr 2016

This software can be either commercial, supplied by vendors of the platforms or otherwise, or free-sources. Webservers such as MetaboAnalyst [11], XCMS (http://metlin.scripps.edu/xcms/) [12] and its online version XCMS Online [13], and tools such as mzMINE (http://mzmine.sourceforge.net/) for LC-MS data sets [14] are just some of the free-access software programs available for data processing and annotation. For GC-MS based data sets, the National Institute of Standards and Technology (NIST) provides free, downloadable deconvolution software AMDIS (Automated Mass spectral Deconvolution and Identification System) is a good starting point. Other processing and annotation tools are either implemented in C++, .NET platform, Python, Java, or in other commercial tools such as Matlab and are operable on Linux or Windows platforms - which is another level of challenge that can have an impact on the data analysis and the analyst. For instance, for metabolite identification, the chemical and mass spectral databases such as METLIN, HMDB, ChemSpider, PubChem, and KEGG are immensely valuable. Guidelines set by the 'seven golden rules' [15] and fragmentation tree approaches such as mzCloud (https://www.mzcloud.org/) are indispensable tools for MS/MS-based data annotations.

R (https://www.r-project.org/) has taken a center stage as free software for statistical computing and graphics with numerous packages developed for data analysis and visualization; this software complies with and runs on a wide variety of UNIX platforms, as well as on Windows and MacOS systems. With advances in analytical platforms, mass-spectrometric techniques, data acquisition quality, and data volume, new resources and tools in metabolomics tools are being developed [16]. Nonetheless, commercial tools developed by vendors can be extremely useful in data collection and meta-analysis as all the features are included in one software suite, but these may have a higher price cap to consider.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2002v1 | CC-BY 4.0 Open Access | rec: 26 Apr 2016, publ: 26 Apr 2016

Post-data acquisition, data transformation, normalization, centering, and scaling are usually needed to make data suitable for further downstream statistical analysis [17] as well as to get rid of batch variations, analytical and other systematic variations induced during sample preparation and data collection.

Tip 5: Do your statistical analysis and interpretation well or consult!

Statistical consultation is imperative and unavoidable. Researchers are recommended to perform an *a priori* power calculation as failures can result in disappointing results [18]. Depending on samples, the researcher can choose parametric (one and two-sample t-tests) or non-parametric (Wilcoxon sign-rank or Mann Whitney tests) tests for comparing two samples. For more than two samples, researchers/you may use a one-way ANOVA or multifactor ANOVAs or a bootstrap analyses with a false discovery rate (FDR) and Bonferroni or Benjamini-Hochberg (BH) corrections. For dimensionality reduction, use multivariate tools such as principal component analysis (PCA) or orthogonal partial linear discriminant analysis (O-PLSDA). Interestingly, a recently updated status of the above computational and statistical approaches is presented [19]. These tools can be useful for detection and removal of out-liers as well. For grouping the samples and metabolites, and hierarchical clustering, K-means clustering, and self-organizing (SOM) maps can be used. In addition, there have been guidelines discussing as to how to avoid false discoveries in metabolomics [20]. Statistical programming languages such as R (https://www.r-project.org/), more specifically Bioconductor (https://www.bioconductor.org/) and R Studio (https://www.rstudio.com/), have been particularly instrumental in changing the

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2002v1 | CC-BY 4.0 Open Access | rec: 26 Apr 2016, publ: 26 Apr 2016

landscape of statistical computing in metabolomics in recent times. As newer programming languages find their way into statistical computing, metabolomics data analysis and interpretation continue to become easier to perform.

Tip 6: The more the better: So integrate other "omics" too

Although this integration of omics data, popularly called as "transomics" or "integromics", can be challenging, is complimentary, useful, and more informative than if it stood alone. Recently, these tools aiding in integration of a diverse –omics datasets have been reviewed [21]. With correct choice of (a) empirical-correlation-based methods, (b) ontology, (c) enrichment and over representation analysis, and (d) network and pathway scale modeling, such integration is helpful for datasets ranging from genomics, epigenetics, transcriptomics, proteomics, fluxomics, and phenomics with metabolomics. Towards this end, pathway enrichment analysis tools like IMPALA (http://impala.molgen.mpg.de/) and iPEAP (http://www.tongji.edu.cn/~qiliu/ipeap.html) and biological network analysis tools like pwOmics (http://www.bioconductor.org/packages/release/bioc/html/pwOmics.html), MetaMapR (http://dgrapov.github.io/MetaMapR/), and MetScape (http://metscape.ncibi.org/) are worth mentioning. With advances in the areas of computational modeling, biomarker discovery, genome-scale modelling, and flux analysis, metabolomics is going to carve out a bigger niche than ever before towards understanding of the 'trans-omes' or the 'globalomes'.

Tip 7: Choose what best represents your data: Pathway or network view

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2002v1 | CC-BY 4.0 Open Access | rec: 26 Apr 2016, publ: 26 Apr 2016

Depending on the goals of the study and the platforms used, the user has a choice to present their data in either a pathway view, a more orthodox approach, or in a network view. Both approaches will allow seamless integration of other "omics" data sets, albeit with differential considerations for data normalization and statistical pre-processing. For example, KEGG (Kyoto Encyclopedia of Genes and Genomes)-based pathways have formed the basis on uncountable studies over time [22]. However, with these pathways being organism-specific and not undergoing frequent updates, other databases such as MetaCyc [23] have come into the picture. However, for untargeted and unbiased analysis, where a more metabolome-wide picture is anticipated, the network view allows connecting more metabolites (nodes or dots) using biological (KEGG), chemical (Tanimoto), functional, structural, or other connectivity (edges) such as positive or negative correlations. Moreover, with versatile and consistently enriched tools such as Cytoscape and a huge list of plugins that come with it, metabolomic networking opportunities are immense. Very versatile mass spectral molecular networking approaches for MS/MS-based metabolomics datasets seem to have found abundant applications in metabolomics studies of microbiomes [24, 25].

Tip 8: Visualization matters

We often hear that an appealing picture conveys more than a thousand words. An overwhelming number of graphical tools are available to maneuver study results to create effective presentations. Oftentimes, a tool is chosen based on whether or not the product is financially feasible and on what purpose they will serve in a particular study. Hierarchical heat maps, Venn diagrams, volcano plots, along with PCA score plots and scree plots or S-plots have all found their worthiness in underscoring the similarities and dissimilarities within metabolomics datasets.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2002v1 | CC-BY 4.0 Open Access | rec: 26 Apr 2016, publ: 26 Apr 2016

Software such as XCMS can create very informative cloud plots [13], while Voronoi tree maps [26] can be useful as well. Hive plots (http://www.hiveplot.net/) have been used to provide a framework for the interactive visualization of a network relationship or the correlation among datasets [27]. Noteworthy mention goes to the cloud plots functionality available with XC-MS for differential global metabolomics data sets [28], or for that matter, interactive PCA available with XCMS online [29]. For GC-MS datasets, MetabolomeExpress can provide a variety of interactive visualizations such as heat maps, chromatogram viewers, and 3D PCA plots [30]. With new visualization tools added every year [16], it may seem challenging to choose the right tool for right kind of interpretation of the metabolomics datasets but a study-dependent approach to present the results in the most suitable manner would guide your search.

Tip 9: Make your data publicly available and do publish!

Data storage can be a challenging task too, yet it can be locally addressed by using encrypted hard drives to shared local networks. For public data sharing, however, there are recommended databases for storage of all metadata. The most prominent ones are Metabolights at http://www.ebi.ac.uk/metabolights/ [31] and Metabolomics Workbench at http://www.metabolomicsworkbench.org/ [32]. For instance, 150+ studies have been archived at Metabolights, and the numbers have shot up in recent times. In addition, websites maintained by research groups i.e., Data Resources of Plant Metabolomics (DROP Met) at Platform for RIKEN Metabolomics (PRIMe) can be good platforms for data disbursal in a timely fashion, making them available to developers for generation of novel algorithms and workflows for metabolomics datasets. Data archiving on public domains are not only a requirement for all funding agencies, scientific journals, and scientific societies, but it is considered ethical for public interest and a

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2002v1 | CC-BY 4.0 Open Access | rec: 26 Apr 2016, publ: 26 Apr 2016

very basic step in open science. This tip is not only essential for the entire scientific community, but it is also a responsibility since most of these studies are funded with taxpayer money.

In addition, many subscriptions and Open-Access journals, do publish standalone metabolomics studies while almost all journals tend to accept metabolomics experiments alongside other -omics datasets for thematic fits. In fact, a PubMed search at the National Center for Biotechnology Information (NCBI) for the term 'metabolomics' (http://www.ncbi.nlm.nih.gov/pubmed/?term=metabolomics) will reveal majority of the scientific journals publishing metabolomics-oriented investigations.

Conclusions

Having written this 'quick tips' article, the author does not assume that either conducting an error-free metabolomics study and then writing a metabolomics stand-alone manuscript are easy whatsoever. Challenges lay in every step of the study, starting from the sample collection to acceptance of the manuscript, but this article has only attempted to summarize the basic framework which can be generalized to some extent and prove beneficial for those recently entering into the metabolomics field. Having summarized above, still the 'big elephants in the room' may remain invisible and studies lead to failures- irrespective of all measures taken to avoid pitfalls.

Acknowledgements

This work is an *ode* to all pioneers and leaders in metabolomics through whose seminal works the author has learned the very basics in metabolomics.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2002v1 | CC-BY 4.0 Open Access | rec: 26 Apr 2016, publ: 26 Apr 2016

Funding

None.

Competing Interests

The author has read the journal's policy and have no conflicts of interest.

References

1. Fiehn O (2002) Metabolomics–the link between genotypes and phenotypes. Plant Mol Biol 48: 155-171. doi: 10.1023/A:1013713905833
2. May JC, Goodwin CR, McLean JA (2015) Ion mobility-mass spectrometry strategies for untargeted systems, synthetic, and chemical biology. Curr Opin Biotechnol 31: 117-121. doi:10.1016/j.copbio.2014.10.012
3. Vogiatzis CG, Zachariadis GA (2014) Tandem mass spectrometry in metallomics and the involving role of ICP-MS detection: A review. Anal Chim Acta 819: 1-14. doi:10.1016/j.aca.2014.01.029
4. Brown SC, Kruppa G, Dasseux JL (2005) Metabolomics applications of FT-ICR mass spectrometry. Mass Spectrom Rev 24: 223-231. doi: 10.1002/mas.20011
5. Sévin DC, Kuehne A, Zamboni N, Sauer U (2015) Biological insights through nontargeted metabolomics. Curr Opin Biotechnol 34: 1-8. doi:10.1016/j.copbio.2014.10.001

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2002v1 | CC-BY 4.0 Open Access | rec: 26 Apr 2016, publ: 26 Apr 2016

6. Bingol K, Brüschweiler R (2015) Two elephants in the room: new hybrid nuclear magnetic resonance and mass spectrometry approaches for metabolomics. Curr Opin Clin Nutr Metab Care 18: 471. doi: 10.1097/MCO.0000000000000206
7. Fiehn O, Robertson D, Griffin J, van der Werf M, Nikolau B, et al. (2007) The metabolomics standards initiative (MSI). Metabolomics 3: 175-178. doi: 10.1007/s11306-007-0070-6
8. Fiehn O, Wohlgemuth G, Scholz M, Kind T, Lee DY, et al. (2008) Quality control for plant metabolomics: reporting MSI-compliant studies. Plant J 53: 691-704. doi: 10.1111/j.1365-313X.2007.03387.x
9. Rocca-Serra P, Brandizi M, Maguire E, Sklyar N, Taylor C, et al. (2010). ISA software suite: supporting standards-compliant experimental annotation and enabling curation at the community level. Bioinformatics 26: 2354-2356. doi: 10.1093/bioinformatics/btq415
10. Abate-Pella D, Freund DM, Ma Y, Simón-Manso Y, Hollender J, et al. (2014) Retention projection enables accurate calculation of liquid chromatographic retention times across labs and methods. J Chromatogr A 1412: 43-51. doi:10.1016/j.chroma.2015.07.108
11. Xia J, Sinelnikov IV, Han B, Wishart DS (2015) MetaboAnalyst 3.0—making metabolomics more meaningful. Nucleic Acids Res 43(W1): W251-257. doi: 10.1093/nar/gkv380
12. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem 78: 779-787. doi: 10.1021/ac051437y

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2002v1 | CC-BY 4.0 Open Access | rec: 26 Apr 2016, publ: 26 Apr 2016

296 13. Tautenhahn R, Patti GJ, Rinehart D, Siuzdak G (2012) XCMS Online: a web-based
297 platform to process untargeted metabolomic data. Anal Chem 84: 5035-5039. doi:
298 10.1021/ac300698c
299 14. Pluskal T, Castillo S, Villar-Briones A, Orešič M (2010) MZmine 2: modular framework
300 for processing, visualizing, and analyzing mass spectrometry-based molecular profile
301 data. BMC Bioinformatics 11: 395. doi: 10.1186/1471-2105-11-395
302 15. Kind T, Fiehn O. (2007) Seven golden rules for heuristic filtering of molecular formulas
303 obtained by accurate mass spectrometry. BMC Bioinformatics 8: 105. doi: 10.1186/1471-
304 2105-2-5.
305 16. Misra BB, der Hooft JJ (2016) Updates in metabolomics tools and resources: 2014–2015.
306 Electrophoresis 37: 86-110. doi: 10.1002/elps.201500417
307 17. van den Berg RA, Hoefsloot HC, Westerhuis JA, Smilde AK, van der Werf MJ (2006)
308 Centering, scaling, and transformations: improving the biological information content of
309 metabolomics data. BMC Genomics 7: 142. doi: 10.1186/1471-2164-7-142
310 18. Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, et al. (2013) Power failure: why
311 small sample size undermines the reliability of neuroscience. Nat Rev Neurosci 4: 365-
312 376. doi:10.1038/nrn3475
313 19. Ren S, Hinzman AA, Kang EL, Szczesniak RD, Lu LJ (2015) Computational and
314 statistical analysis of metabolomics data. Metabolomics 11: 1492-1513. doi:
315 10.1007/s11306-015-0823-6
316 20. Broadhurst DI, Kell DB (2006) Statistical strategies for avoiding false discoveries in
317 metabolomics and related experiments. Metabolomics 2: 171-196. doi: 10.1007/s11306-
318 006-0037-z

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2002v1 | CC-BY 4.0 Open Access | rec: 26 Apr 2016, publ: 26 Apr 2016

21. Wanichthanarak K, Fahrmann JF, Grapov D (2015) Genomic, proteomic, and metabolomic data integration strategies. Biomarker insights 10: 1. doi: 10.4137/BMI.S29511

22. Okuda S, Yamada T, Hamajima M, Itoh M, Katayama T, et al. (2008) KEGG Atlas mapping for global analysis of metabolic pathways. Nucleic Acids Res 36: W423-W426. doi: 10.1093/nar/gkn282

23. Caspi R, Foerster H, Fulcher CA, Kaipa P, Krummenacker M, et al. (2008) The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Res 36: D623-D631. doi: 10.1093/nar/gkm900

24. Watrous J, Roach P, Alexandrov T, Heath BS, Yang JY, et al. (2012) Mass spectral molecular networking of living microbial colonies. Proc Natl Acad Sci U S A 109: E1743-E1752. doi: 10.1073/pnas.1203689109

25. Garg N, Conrad D, Dorrestein P (2015) Metabolomics by mass spectrometry based molecular networking and spatial mapping. The FASEB J 29: 369-371.

26. Nocaj A, Brandes U (2012) Computing Voronoi Treemaps: Faster, Simpler, and Resolution-independent. In *Computer Graphics Forum* (31: 3pt1, 855-864). Blackwell Publishing Ltd. doi: 10.1111/j.1467-8659.2012.03078.x

27. Krzywinski M, Birol I, Jones SJ, Marra MA (2012) Hive plots—rational approach to visualizing networks. Brief Bioinform 13: 627-644. doi: 10.1093/bib/bbr069

28. Patti GJ, Tautenhahn R, Rinehart D, Cho K, Shriver LP, et al. (2012) A view from above: cloud plots to visualize global metabolomic data. Anal Chem 85: 798-804. doi: 10.1021/ac3029745

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2002v1 | CC-BY 4.0 Open Access | rec: 26 Apr 2016, publ: 26 Apr 2016

342 29. Gowda H, Ivanisevic J, Johnson CH, Kurczy ME, Benton HP, et al. (2014) Interactive
343 XCMS Online: simplifying advanced metabolomic data processing and subsequent
344 statistical analyses. Anal Chem 86: 6931-6939. doi: 10.1021/ac500734c
345 30. Carroll AJ, Badger MR, Millar AH (2010) The MetabolomeExpress Project: enabling
346 web-based processing, analysis and transparent dissemination of GC/MS metabolomics
347 datasets. BMC Bioinformatics 11: 376. doi: 10.1186/1471-2105-11-376
348 31. Haug K, Salek RM, Conesa P, Hastings J, de Matos P, et al. (2012) MetaboLights—an
349 open-access general-purpose repository for metabolomics studies and associated meta-
350 data. Nucleic Acids Res gks1004. doi: 10.1093/nar/gks1004
351 32. Sud M, Fahy E, Cotter D, Azam K, Vadivelu I, et al. (2016) Metabolomics Workbench:
352 An international repository for metabolomics data and metadata, metabolite standards,
353 protocols, tutorials and training, and analysis tools. Nucleic Acids Res 44(D1): D463-70.
354 doi: 10.1093/nar/gkv1042

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2002v1 | CC-BY 4.0 Open Access | rec: 26 Apr 2016, publ: 26 Apr 2016